Ethical Data Visualization: Taming Treacherous Data

Chris Church
Katherine Hepworth
Welcome to DHSI 2019!

Thanks for joining the DHSI community!

In this booklet, you will find essential course materials prefaced by some useful information about getting settled initially at UVic, finding your way around, getting logged in to our network (after you’ve registered the day before our courses begin), and so on.

Given our community’s focus on things computational, it will be a surprise to no one that we might expect additional information online for some of the classes - your instructors will let you know - or that the most current version of all DHSI-related information may be found on our website at dhsi.org.

Do check in there first if you need anything that’s not in this coursepak.

To access the DHSI wifi network, simply go into your wireless settings and connect to the “DHSI” network and enter the password “dhsi2019”.

And please don’t hesitate to be in touch with us at institut@uvic.ca or via Twitter at @AlyssaA_DHSI or @DHInstitute if we can be of any help ....
DHSI Wi-Fi

Network name: DHSI
Passkey: dhsi2019
The 2019 schedule is just taking shape nicely! A very few things to confirm, add, etc, still but this is the place to be to find out what is happening when / where ...

### Sunday, 2 June 2019 [DHSI Registration + Suggested Outings]

If you’re here a day or two before we begin, or staying a day or two afterwards, here are a few ideas of things you might consider doing ....

**Suggested Outing 1, Botanical Beach (self-organised; car needed)**

A self-guided visit to the wet, wild west coast tidal shelf (and historically-significant former research site) at Botanical Beach; we recommend departing early (around 8.00 am) to catch low tide for a better view of the wonderful undersea life! Consider bringing a packed lunch to nibble-on while looking at the crashing waves when there, and then have an afternoon drink enjoying the view from the deck of the Port Renfrew Hotel.

**Suggested Outing 2, Butchart Gardens (self-organised)**

A shorter journey to the resplendently beautiful Butchart Gardens and, if you like, followed by (ahem) a few minutes at the nearby Church and State Winery, in the Saanich Penninsula. About an hour there by public bus from UVic, or 30 minutes by car.

**Suggested Outing 3, Salt Spring Island (self-organised; a full day, car/bus + ferry combo)**

Why not take a day to explore and celebrate the funky, laid back, Canadian gulf island lifestyle on Salt Spring Island. Ferry departs regularly from the Schwartz Bay ferry terminal, which is about one hour by bus / 30 minutes by car from UVic. You may decide to stay on forever ....

**Suggested Outing 4, Paddling Victoria’s Inner Harbour (self-organised)**

A shorter time, seeing Victoria’s beautiful city centre from the waterways that initially inspired its foundation. A great choice if the day is sunny and warm. Canoes, kayaks, and paddle boards are readily rented from Ocean River Adventures and conveniently launched from right behind the store. Very chill.

And more!

Self-organised High Tea at the Empress Hotel, scooter rentals, visit to the Royal BC Museum, darts at Christies Carriage House, a hangry breakfast at a local diner, whale watching, kayaking, brew pub sampling (at Spinnaker's, Swans, Moon Under Water, and beyond!), paddle-boarding, a tour of used bookstores, and more have also been suggested!

### Monday, 3 June 2019

Your hosts for the week are Alyssa Arbuckle, Ray Siemens, and Jannaya Friggstad Jensen.
### Tuesday, 4 June 2019

**8:30 to 10:00**
- Welcome, Orientation, and Instructor Overview ([MacLaurin A144](#))
  - Welcome to the Territory
  - Welcome to DHSI: Ray Siemens, Alyssa Arbuckle
  - Welcome from UVic: Jonathan Bengtson (University Librarian), Alexandra D'Arcy (Associate Dean Research, Humanities)

**Classes in Session (click for details and locations)**

1. [Foundations] Digitisation Fundamentals and their Application ([Clearihue A103, Lab](#))
2. [Foundations] Introduction to Computation for Literary Criticism ([Clearihue A102, Lab](#))
4. [Foundations] DH For Department Chairs and Deans ([David Strong Building C124, Classroom](#))
5. [Foundations] Developing a Digital Project (With Omeka) ([Clearihue A031, Lab](#))
8. [Foundations] Fundamentals of Programming/Coding for Human(s|ists) ([Clearihue A108, Lab](#))
9. Out-of-the-Box Text Analysis for the Digital Humanities ([Human and Social Development A160, Lab](#))
10. Sound and Digital Humanities ([Cornett A120, Classroom](#))
11. Critical Pedagogy and Digital Praxis in the Humanities ([Clearihue D132, Classroom](#))
12. Digital Humanities for Japanese Culture: Resources and Methods ([McPherson Library A003, Classroom](#))
13. Conceptualising and Creating a Digital Edition ([McPherson Library 210, Classroom](#))
14. Retro Machines & Media ([McPherson Library 129, Classroom](#))
15. Geographical Information Systems in the Digital Humanities ([Clearihue A105, Lab](#))
16. Introduction to IIIF: Sharing, Consuming, and Annotating the World’s Images ([Cornett A121, Classroom](#))
17. Web APIs with Python ([Human and Social Development A170, Lab](#))
18. Ethical Data Visualization: Taming Treacherous Data ([Cornett A128, Classroom](#))
19. Linked Open Data and the Semantic Web ([Cornett A132, Classroom](#))
20. Palpability and Wearable Computing ([McPherson Library A025, Classroom](#))
22. Information Security for Digital Researchers ([David Strong Building C114, Classroom](#))

**10:15 to Noon**

1. [Foundations] Digital Humanities for Japanese Culture: Resources and Methods ([McPherson Library A003, Classroom](#))
2. Conceptualising and Creating a Digital Edition ([McPherson Library 210, Classroom](#))
3. Conceptualising and Creating a Digital Edition ([McPherson Library 210, Classroom](#))
4. Geographical Information Systems in the Digital Humanities ([Clearihue A105, Lab](#))
5. Introduction to IIIF: Sharing, Consuming, and Annotating the World’s Images ([Cornett A121, Classroom](#))
6. Web APIs with Python ([Human and Social Development A170, Lab](#))
7. Ethical Data Visualization: Taming Treacherous Data ([Cornett A128, Classroom](#))
8. Linked Open Data and the Semantic Web ([Cornett A132, Classroom](#))
11. Information Security for Digital Researchers ([David Strong Building C114, Classroom](#))

**12:15 to 1:15**

- Lunch break / Unconference Coordination Session ([MacLaurin A144](#))
  - (Grab a sandwich and come on down!)
  - Discussion topics, scheduling, and room assignments from among all DHSI rooms will be handled at this meeting.

**1:30 to 4:00**

- Classes in Session

**4:10 to 5:00**

- Institute Lecture: Jacqueline Wernimont (Dartmouth C): "Sex and Numbers: Pleasure, Reproduction, and Digital Biopower"
  - Chair: Anne Cong-Huyen (U Michigan) ([MacLaurin A144](#))

**4:10 to 5:00**

- Abstract: Drawing from Numbered Lives (MIT 2018), this talk will consider a long history of sex-number entanglement in Anglo-American Cultures. Drawing on historical and contemporary objects and practices, Wernimont will ask "in what ways do theories of biopower, critical gender and critical race studies, and media studies" suggest that we can understand this set of entanglements and their impacts. NB: While relevant, this talk will not include discussions of sexual trauma or violence. It will include frank discussion of sex acts and various ways of translating sexual behavior into numbers.

**5:00 to 6:00**

- Opening Reception ([University Club](#))

**9:00 to Noon**

- Classes in Session

**12:15 to 1:15**

- Lunch break / Unconference

**1:30 to 4:00**

- Classes in Session

---

DHSI Conference and Colloquium Lightning Talk Session 1 ([MacLaurin A144](#))
Wednesday, 5 June 2019

4:15 to 5:15

- Marion Grant (Ryerson U), "Visualizing Networks: Yellow Nineties Print and Performance"
- Megan Perram (U Alberta), "Configuring the Postdigital Body Through the Digital Illness Narratives of Women with Polycystic Ovarian Syndrome"
- Kristen Starkowski (Princeton U), "Mapping Minor Characters: Quantifying and Visualizing Character Space in Dickens's Novels and in their Adaptations"
- Leah Henrickson (Loughborough U), "Who is the author of the computer-generated text?"

6:00 to 8:00

DHSI Newcomer's Gathering (Grad House Restaurant, Graduate Student Centre)
Come down, buy meal and a beverage, and make some new friends!

9:00 to Noon

Classes in Session

Lunch break / Unconference

"Mystery" Lunches

Presentation: An Introduction to Scholarly Publishing with Manifold (MacLaurin A144)
Lunch included for those who [register here]

This presentation introduces Manifold Scholarship, a Mellon-funded digital publishing platform developed by the CUNY Graduate Center, The University of Minnesota Press, and Cast Iron Coding. Manifold allows you to create beautiful, dynamic open access projects that can include text, images, video, embedded resources, and social annotation. We will provide an overview of Manifold and demonstrate how faculty, students and staff in the digital humanities can use Manifold to publish open access scholarly works, conduct and participate in peer review, and create custom edited versions of public domain course texts and OER.

1:30 to 4:00

Classes in Session

DHSI Conference and Colloquium Lightning Talk Session 2 (MacLaurin A144)
Chair: Kim O'Donnell (Simon Fraser U)

- Catherine Ryu (Michigan State U), "Tone Perfect: Developing a Multimodal Audio Database for Mandarin Chinese as an Open Source"
- Kenzie Burchell (U Toronto Scarborough), "Making Responsible Reporting Practices Visible: Comparing newswire coverage of humanitarian crises in Syria"
- Jessica Linzel (Brock U), "The Shopkeeper Aristocracy: Mapping Trade Networks in Colonial Niagara"
- Kirsten Painter (U Washington), "From Bogatyr to Bread: Digitization & Online Exhibition of Rare Russian Children's Books at the U Washington"
- John Barber (Washington State U), "A Mighty Span"

6:00 to 7:00

"Half Way There!" [An Informal, Self-Organized Birds of a Feather Get-Together] (Felicitas, Student Union Building)
Bring your DHSI nametag and enjoy your first tipple on us! [A great opportunity for an interest group meet-up ....]
Friday, 7 June 2019 [DHSI; ADHO Pedagogy SIG Conference Opening]

9:00 to Noon
Classes in Session

12:15 to 1:15
Lunch Reception / Course E-Exhibits (MacLaurin A100)

1:30 to 1:50
Remarks, A Week in Review (MacLaurin A144)

2:00 to 3:00
Joint Institute Lecture (DHSI and ADHO Pedagogy SIG Conference):
Matt Gold (CUNY Graduate Center and Association for Computers and the Humanities): “Thinking Through DH: Proposals for Digital Humanities Pedagogy”
Chair: Diane Jakacki (Bucknell U) (MacLaurin A144)

Abstract: How do we teach digital humanities, and how should DH be taught? What, indeed, should we teach when we teach DH? This talk will present a proposal for grounding digital humanities pedagogical practice in the research interests of our students and the epistemological foundations of our methods rather than through an approach grounded more central in data and methods.

3:30 to 5:00
Joint Reception: DHSI and ADHO Pedagogy SIG Conference (University Club)
E-Poetry Event (Chris Tanasescu)
Watch this space for details, including how to participate!

DHSI Conference and Colloquium Poster/Demo Session
- Pia Russel (U Victoria); Emily Streml (U Victoria), “British Columbia’s Historical Textbooks Digital Library”
- Cody Hennesey (U Minnesota); Rachael Samberg (U California, Berkeley); Stacy Reardon (U California, Berkeley), “Finding the Haystack: Literacies for Accessing and Using Text as Data”
- Paula Johanson (ETCL; Independent Scholar), “Proving Seahorses and Juan de Fuca’s Travels in The Curve of Time”
- Tara Baillargeon (Marquette U); Elizabeth Wawrzyniak (Marquette U), “FellowsHub: J. R. R. Tolkien Fanzine Portal”
- Caterina Agostini (Rutgers U), “Art at the Time of Syphilis: A First-Person Medical Narrative in Benvenuto Cellini’s Vita”
- Lauren Elle DeGaine (ETCL; U Victoria), “Women at the Front: A Digital Exhibit of Victorian Frontpiece Illustrations”
- Adam Griggs (Mercer U); Kathryn Wright (Mercer U); Christian Pham (Mercer U); Gail Morton (Mercer U); Stephanie Miranda (Mercer U), “Digitizing Middle Georgia’s History of Slavery”

Saturday, 8 June 2019 [Conference, Colloquium, and Workshop Sessions]

8:00 to 9:00
Conference / Workshop Registration (MacLaurin A100)

The day’s events are included with your DHSI registration. If you’re not registered in DHSI, you’re very welcome to join us by registering here as a Conference / Colloquium / Workshop participant. We’ll have a nametag waiting for you!

Coffee, Tea, &c?
Looking for some morning coffee or tea, or a small nibble? Options and hours of operation for weekend campus catering are available here. Mystic Market usually opens around 10.00.

9:00 to 4:00
DHSI Conference and Colloquium Sessions
ADHO Pedagogy SIG Conference Sessions
Right2Left Workshop Sessions

9:00 to 4:00
All Day DHSI Workshop Session (click for workshop details and free registration for DHSI participants)
- 55. Introduction to Machine Learning in the Digital Humanities [8-9 June; All day, each day] (David Strong Building C124, Classroom)

9:00 to 9:10
Informal Greetings, Room Set-up (Lobby, outside Hickman 105)

9:00 to 9:30
Session 1
DHSI Colloquium and Conference (Hickman 105)
Digital Humanities & Literature, Chair: Kim O'Donnell (Simon Fraser U)
- Youngmin Kim (Dongguk U), “Transdiscursivity in the Convergence of Digital Humanities and World Literature”
- Caroline Winter (U Victoria), “Digitizing Adam Smith’s Literary Library”
- Kaitlyn Fralick (U Victoria); Kailey Fukushima (U Victoria); Sarah Karlson (U Victoria), “Victorian Poetry
9:10 to 10:30  ADHO Pedagogy SIG Conference (Hickman 110)
Chair: Katherine Faull (Bucknell U)
- Aaron Tucker and Nada Savicvetic (Ryerson U), “Write Here, Right Now: An Open Source eTextbook for the Flipped Classroom”
- Heather McAlpine (U Fraser Valley), “Digital Meters: Using Text Encoding to Teach Literature in the Undergraduate Classroom”
- Tiina H. Airaksinen (U Helsinki), “Digital Humanities in Cultural Studies: Creating a MOOC course for University Students and A-Level Students”

Right2Left Workshop (Hickman 116)

10:30 to 10:40  Break

10:40 to Noon  Session 2
DHSI Colloquium and Conference (Hickman 105)
Digital Humanities & Society, Chair: Eleanor Reed (Hastings C)
- Joel Zapata (Southern Methodist U), “Uncovering the Southern Plains’ Mexican American Civil Rights Movement”
- Brendan Mackie (U California, Berkeley), “Visualizing Long-Term Cultural Change: An Example From The Birth of Civil Society”

ADHO Pedagogy SIG Conference (Hickman 110)
Chair: Laura Estill (St Francis Xavier U)
- Jane Jackson (Chinese U of Hong Kong), “Interrogating digital spaces for intercultural meaning-making”
- Christopher Church, Katherine Hepworth (U Nevada, Reno), “We’re STEAMed! A call for balancing technical instruction and disciplinary content in the digital humanities”

Right2Left Workshop (Hickman 116)
- Edward “Eddie” Surman (Claremont Graduate U), “Qualitative Digital Text Analysis and #Right2Left Languages: A Demonstration of Atlas.ti using the Hebrew Bible”

Noon to 1:10  Lunch (We recommend Mystic Market on weekends!)

1:10 to 2:30  Session 3
DHSI Colloquium and Conference (Hickman 105)
Digital Humanities & Community, Chair: Claire Carlin (U Victoria)
- Pia Russel (U Victoria); Emily Stremel (U Victoria), “Mentorship and disability: Supporting disabled employees in digital humanities”
- Amy Lueck (Santa Clara U), “Virtually Emplacing Indigenous Memory”
- Md. Shehabul Alam (National U Bangladesh), “Integrating Library Service with Union Information and Service Center: A Joint Initiative towards Digital Bangladesh”
- Veronica Gomez (Instituto de Humanidades y Ciencias Sociales (HuSo) - UNL-CONICET), “Latin American E-literature and Location: The Nation Revisited in Electronic Literature Organization (ELO)”

ADHO Pedagogy SIG Conference (Hickman 110)
Chair: Chris Tănăsescu (UC Louvain)
- Laura Estill (St Francis Xavier U), “One Assignment, Three Ways: Assessing DH Projects in a Literature Course”
- Shu Wan (U Iowa), “A digital ‘historical gaze’ of Chinese students in Iowa, 1911-1930”
- Francesca Giannetti (Rutgers U, New Brunswick), ““So near while apart: Correspondence Editions as Critical Library Pedagogy and Digital Humanities Methodology”

Right2Left Workshop (Hickman 116)
- Najla Jarkas (American U Beirut) and David Joseph Wrisley (NYU Abu Dhabi), “RTL Software Localization and Digital Humanities: the Case Study of Translating Voyant Tools into Arabic”
2:30 to 2:40  
**Break**

2:40 to 4:00  
**Session 4**  
- **DHSI Colloquium and Conference (Hickman 105)**
  - Ashleigh Casserme-Stanfield (U Chicago), “Sonifying Hamlet and Reading the Room”

- **ADHO Pedagogy SIG Conference (Hickman 110)**
  - Chair: Aaron Tucker (Ryerson U)
  - Youngmin Kim (Dongguk U), “Teaching Digital Humanities and World Literature in Class”
  - Alice Fleerackers, Juan Pablo Alperin, Esteban Morales, Remi Kalir (Simon Fraser U, U Colorado Denver), “Online annotations in the classroom: How, why, and what do students learn from annotating course material?”
  - Andie Silva (York C and Graduate Center, CUNY), “Keeping it Local: Undergraduate DH as Feminist Practice”

- **Right2Left Workshop (Hickman 116)**
  - Joanna Byszuk (Institute of Polish Language, Polish Academy of Sciences, Warsaw/Computational Stylistics Group) and Alexey Khismatulin (Institute of Oriental Manuscripts, Russian Academy of Sciences, Saint Petersburg), “Attribution of Authorship for Medieval Persian Quasidas with Stylometry”
  - Ilan Benattar (New York U), “#Right2Left Biblical Translations in Jewish Textual History: Case Studies in Judeo-Arabic and Judeo-Spanish”

---

**Sunday, 9 June 2019 [Workshop Sessions]**

8:00 to 5:00  
The day's events are included with your DHSI registration. If you're not registered in DHSI, you're very welcome to join us by registering here as a Conference / Colloquium / Workshop participant. We'll have a nametag waiting for you!

---

Coffee, Tea, &c?  
Looking for some morning coffee or tea, or a small nibble? Options and hours of operation for weekend campus catering are available here. Mystic Market usually opens around 10.00.

---

9:00 to 4:00  
**All Day Workshop Sessions**  
- 55. Introduction to Machine Learning in the Digital Humanities [8-9 June; All day, each day] (David Strong Building C124, Classroom)
- 56. Pedagogy of the Digitally Oppressed: Anti-Colonial DH Methods and Praxis [9 June; All Day] (Hickman 115, Classroom)
- 57. Natural Language Processing and Network Coding Apps for Text & Textual Corpus Analysis in the Humanities [9 June; All Day] (David Strong Building C114, Classroom)

---

9:00 to Noon  
**AM Workshop Sessions**
- 59. 3D Visualization for the Humanities [9 June; AM] (Cornett A229, Classroom)
- 60. It’s All Relational: AbTeC’s Indigenous Video Game Workshops as Storytelling Praxis [9 June; AM] (Cornett A121, Classroom)
- 61. Spatial DH: De-Colonizing Cultural Territories Online [9 June; AM] (Clearihue D130, Classroom)
- 63. Creating a CV for Digital Humanities Makers [9 June; AM] (David Strong Building C108, Classroom)

---

Noon to 1:00  
**Lunch**  
(We recommend Mystic Market on weekends!)

---

1:00 to 4:00  
**PM Workshop Sessions**
- 65. Indigenous Futures in the Classroom and Beyond [9 June; PM] (Cornett A123, Classroom)
- 66. DHSI Knits: History of Textiles and Technology [9 June; PM] (Fine Arts 109, Classroom)
- 68. Linked Open Datafication for Humanities Scholars [9 June; PM] (McPherson Library A003, Classroom)
- 69. Stylo - WYSIWYM Text Editor for Humanities Scholars [9 June; PM] (McPherson Library A025, Classroom)

---

After the day, many will wander to Cadboro Bay and the pub at Smuggler’s Cove OR the other direction to Shelbourne Plaza and Maude Hunter’s Pub OR even into the city for a bite to eat.

---

Monday, 10 June 2019
Your hosts for the week are Ray Siemens and Jannaya Friggstad Jensen.

7:45 to 8:15  
DHSI Last-minute Registration (MacLaurin A100)

8:30 to 10:00  
Welcome, Orientation, and Instructor Overview (MacLaurin A144)

Classes in Session (click for details and locations)
- 29. [Foundations] Understanding The Predigital Book: Technologies of Inscription (McPherson Library A003, Classroom)
- 30. [Foundations] Databases for Digital Humanists (McPherson Library 210, Classroom)
- 33. Digital Storytelling (Cornett A120, Classroom)
- 34. Text Mapping as Modelling (Clearihue D131, Classroom)
- 35. Stylometry with R: Computer-Assisted Analysis of Literary Texts (Clearihue A102, Lab)
- 36. Open Access and Open Social Scholarship (Clearihue D130, Classroom)
- 37. Digital Games as Tools for Scholarly Research, Communication and Pedagogy (Cornett A229, Classroom)
- 38. Queer Digital Humanities (David Strong Building C114, Classroom)
- 40. Introduction to Electronic Literature in DH: Research and Practice (Cornett A128, Classroom)
- 41. Surveillance and the Critical Digital Humanities (David Strong Building C108, Classroom)
- 42. Text Analysis with Python and the Natural Language ToolKit (Clearihue A103, Lab)
- 43. Creating LAMP Infrastructure for Digital Humanities Projects (Human and Social Development A170, Lab)
- 44. Digital Humanities Pedagogy: Integration in the Curriculum (Cornett A121, Classroom)
- 45. Accessibility & Digital Environments (Priestly Law Library 265, Classroom)
- 46. Digital Humanities Pedagogy: Integration in the Curriculum (Cornett A121, Classroom)
- 47. Agile Project Management (Cornett A132, Classroom/Lab)
- 48. XPath for Processing XML and Managing Projects (Clearihue A105, Lab)
- 49. Endings: How to End (and Archive) your Digital Project (Priestly Law Library 192, Classroom)
- 50. Text Processing - Techniques & Traditions (McPherson Library A025, Classroom)
- 51. Introduction to Humanities Data Analysis & Visualization in R (HDA) (Human and Social Development A160, Lab)
- 52. Introduction to Network Analysis in the Digital Humanities (Clearihue D132, Classroom)

12:15 to 1:15  
Lunch break / Unconference Coordination Session (MacLaurin A144)  
(Grab a sandwich and come on down!)

"Mystery" Lunches

1:30 to 4:00  
Classes in Session

Institute Lecture: Angel David Nieves (San Diego State U): "3D Mapping and Forensic Traces of Testimony: Documenting Apartheid-Era Crimes Through the Digital Humanities"  
Chair: Constante Crompton (U Ottawa) (MacLaurin A144)

Abstract: In 1989 the killing of a queer, 14-year-old youth in Winnie Mandela’s house named Stompie Seipei (an event that few in South Africa are willing to recall, let alone discuss, in any detail) -- is perhaps one of the most glaring examples where the queer and activist community was suppressed or erased from anti-apartheid/liberation histories. Digital humanities may actually help both reconstruct and recover a history that is still very early in the telling, despite what is commonly believed about the liberation struggle and the contributions of queer activists in the dismantling of apartheid. Perhaps it could explain why a youth such as Seipei was killed -- or at the very least, provide a more complex and messy narrative that permits one to know more how the history of queer anti-apartheid activists was suppressed. This talk outlines a methodology for "messy thinking and writing" in the digital humanities that -- through a queer and feminist intersectional framework -- permits a more complex layering of oral histories and 3D historical reconstructions.

5:00 to 6:00  
Reception (University Club)

Tuesday, 11 June 2019

9:00 to Noon  
Classes in Session

12:15 to 1:15  
Lunch break / Unconference

"Mystery" Lunches
1:30 to 4:00  Classes in Session

4:15 to 5:15  DHSI Conference and Colloquium Lightning Talk Session 4 (MacLaurin A144)
Chair: Lindsey Seatter (U Victoria)

- Ashley Caranto Morford (U Toronto); Kush Patel (U Michigan); Arun Jacob (McMaster U), “#OurDHSIs anti-colonial: Questions and challenges in dismantling colonial influences in digital humanities pedagogy”
- Luis Meneses (ETCL; U Victoria), “Identifying Changes in the Political Environment in Ecuador”
- Laura Horak (Carleton U), “Building the Transgender Media Portal”

6:00 to 8:00  DHSI Newcomer’s Gathering (Grad House Restaurant, Graduate Student Centre)
Come down, buy meal and a beverage, and make some new friends!

9:00 to Noon  Classes in Session

12:15 to 1:15  Lunch break / Unconference

1:30 to 4:00  DHSI Conference and Colloquium Lightning Talk Session 5 (MacLaurin A144)
Chair: Lindsey Seatter (U Victoria)

- Calin Murgu (New College of Florida), “Putting local metadata to strategic use: A Dashboard for visualizing 60 years of theses metadata”
- Jason Lajoie (U Waterloo), “Queer Critical Making and the Logic of Control”
- John Barber (Washington State U), “Zambezi River Bridge”

6:00 to 7:00  “Half Way There (yet again)!” [An Informal, Self-Organized Birds of a Feather Get-Together] (Felicitas, Student Union Building)
Bring your DHSI nametag and enjoy your first tipple on us! [A great opportunity for an interest group meet-up ...]

9:00 to Noon  Classes in Session

12:15 to 1:15  Lunch break / Unconference

1:30 to 4:00  Institute Lecture: Karina van Dalen-Oskam (Huygens Institute and U Amsterdam; Alliance of Digital Humanities Organizations): “The Riddle of Literary Quality: Some Answers”
Chair: Aaron Mauro (Penn State, Behrend C) (MacLaurin A144)

Abstract: What is literature, and can you measure it? That is the key question of the project The Riddle of Literary Quality. “The Riddle” is a research project of the Huygens Institute for the History of the Netherlands (Amsterdam) in collaboration with the Fryeke Akademy (Leeuwarden) and the Institute for Logic, Language and Computation (University of Amsterdam). The Riddle combines computational analysis of writing style with the results of a large online survey of readers, completed by almost 14,000 participants. In my talk, I will go into...
some of the main results of the project.

Friday, 14 June 2019

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 to Noon</td>
<td>Classes in Session</td>
<td></td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>Lunch Reception / Course E-Exhibits (MacLaurin A100)</td>
<td></td>
</tr>
<tr>
<td>1:30 to 2:00</td>
<td>Closing, DHSI in Review (MacLaurin A144)</td>
<td></td>
</tr>
</tbody>
</table>

Contact info:  
institut@uvic.ca  P: 250-472-5401  F: 250-472-5681
Ethical Data Visualization:
Taming Treacherous Data Coursepak
Digital Humanities Summer Institute
3–7 June, 2019
University of Victoria, Vancouver Island
British Columbia, Canada

Instructors
Christopher Church & Katherine Hepworth
About Chris: http://www.christophermchurch.com/
Contact Chris: christopherchurch@unr.edu
About Katherine: https://kathep.com
Contact Katherine: khepworth@unr.edu
Table of Contents

- **Course description**
- **Syllabus**
- **Detailed agenda**
  - **Day 1: Overview — Introducing Treacherous Data**
    - Slides
    - Readings
    - **Documentation**
      - Sample R Codes and Sample Datasets
      - Ggplot cheat sheets
  - **Day 2: Racist Historical Data**
    - Slides
    - Readings
    - **Documentation**
      - Sample R Codes and Sample Datasets
  - **Day 3: Contemporary Problematic Data — Politics and Culture**
    - Slides
    - Readings
  - **Day 4: Ideologically Laden Data — Propaganda and Polemics**
    - Readings
  - **Day 5: Defining Position Statements — Taming Treacherous Data**
    - Readings

- **Extra materials**
  - **Optional extra readings**
  - **Space for notes**
Course Description

This course teaches participants how to use ethical visualization principles and practices to visualize treacherous, or culturally problematic, data. Such data includes racist historical documents, ideologically laden materials, culturally controversial texts, politically charged topics, gendered works, etc. Aimed at people who work with culturally sensitive datasets, and those who are interested in critical reflection on visualization practice, the course will combine hands-on activities and discussion. Participants will create data visualizations using R and instructor-provided stock code, and then interrogate their visualizations, identifying the extent and severity of the ethical pitfalls they inevitably contain. By the end of the week, participants will have produced several visualizations and prepared a position statement on ethical visualization appropriate for their own cultural and disciplinary contexts. No previous knowledge in coding, R, or visualizations is required. Participants are welcome to bring their own treacherous data, or they may use sample projects provided by the instructors. If you are unsure as to whether your data will work in this class, please feel welcome to contact the instructors in advance. Though not required, this course would be an ideal follow-up to “Visualizing Information: Where Data Meets Design” as well as the “Introduction to Javascript and Data Visualization.”
Syllabus

**Day 1 Overview: Introducing Treacherous Data**
Morning 1: First Principles and Visualization Examples
10:15 to Noon (1.75 hours)
- **Class introductions**
- **Lecture:** First principles — Visual hierarchy, legibility, and color
- **Examples and discussion:** Exploring Sample Visualizations — Best and worst practices
- **Two readings and discussion:**

Afternoon 1: Building Your First Treacherous Visualization
1:30 to 4:00 (2.5 hours)
- **Activity overview**
- **Documentation:** Brief crash course in R Studio for visualization purposes
- **Activity:** Creating basic best and worst ethical visualization plots by editing custom, pre-made R codes

**Day 2: Racist Historical Data**
Morning 2: Visualization and Graphic Conventions
9:00am - 12pm (3 hours)
- **Lecture:** Historicizing visualization — narrative and graphic visualization conventions through history
- **Examples and discussion:** Exploring Sample Visualizations — Racist historical visualizations
- **Reading:** Forgotten Dead, Introduction

Afternoon 2: Visualizing Race
1:30pm to 4:00pm (2.5 hours)
- Reading and discussion: Hepworth, K. J., Church, C., “Racism in the Machine: Visualization Ethics in Digital Humanities Projects”. Digital Humanities Quarterly. [under review]

- Visual exploration activity & discussion: Exploring Racial Terror Lynchings map and Map of White Supremacist Mob Violence

- Documentation: Ethical Visualization Workflow

- Visualization activity: Creating visualization plots (using provided, pre-made R codes) based on the mapping projects using the following data: https://www.kaggle.com/rtatman/historical-american-lynching

Day 3: Contemporary Problematic Data: Politics and Culture
Morning 3: Creating Culturally and Politically Sensitive Visualizations
9:00am to 12pm (3 hours)
- Lecture: Problematizing visualizations — The visual rhetoric of graphic elements and their impact on audiences

Afternoon 3: Discussion and Project Time
1:30pm to 4:00pm (2.5 hours)
- Two activity options
  Students will do prepared activities on the day’s theme, or they can work on their own datasets / projects
- Prepared activity and discussion: Explore the data at the following sites — Fatal Encounters, Fatal Force, and The Counted — which one presents the data in the most ethical way? Why? Choose one of these data sets to make a visualization from. Think of a target audience you would like to understand this information better. Choose a visualization format and graphic elements that are sensitive to the represented material and appropriate for the intended audience.
- Project time: apply principles and practices learned so far to your own treacherous data

Day 4: Ideologically Laden Data: Propaganda and Polemics
Morning 4: Becoming Acquainted with the Mores of Data Collection and Visualization
9:00am to 12pm (3 hours)

- Activity: Creating Visualizations from Census Data
Afternoon 4: Discussion and Project Time
1:30pm to 4:00pm (2.5 hours)

Two activity options
Students will do prepared activities on the day’s theme, or they can work on their own datasets / projects

- Activity: Students will create two visualizations of the same data: one propagandized and one ethical. These visualizations can be created using provided sample data or their own treacherous data.

- Activity: Students will present either their propagandized or ethical visualization to the rest of the class, and fellow students will identify whether the visualization can be trusted based on what they have learned in the class so far.

Day 5: Defining Position Statements: Taming Treacherous Data
Morning 5: Framing your own visualization practise through position statements
9:00am to 12:00pm (3 hours)

- Reading and discussion: Kirk St. Amant. 2016. "Introduction to the special issue: Cultural considerations for communication design: integrating ideas of culture, communication, and context into user experience design". Communication Design Quarterly Review. 4, no. 1: 6-22.

- Activity: Writing 500-1000 word position statements on your own visualization practise, based on socio-cultural considerations of their subject and audience.
Detailed Agenda

The following pages contain a detailed breakdown of topics, activities, readings and slides that will be covered each day in the Ethical Data Visualization course.

Day 1 Overview: Introducing Treacherous Data

Morning 1: First Principles and Visualization Examples

10:15 to Noon (1.75 hours)

- Class introductions
- Lecture: First principles — Visual hierarchy, legibility, and color
- Examples and discussion: Exploring Sample Visualizations — Best and worst practices
- Two readings and discussion:

Afternoon 1: Building Your First Treacherous Visualization

1:30 to 4:00 (2.5 hours)

- Activity overview
- Documentation: Brief crash course in R Studio for visualization purposes
- Activity: Creating basic best and worst ethical visualization plots by editing custom, pre-made R codes
DAY 1 SLIDES

DATA VISUALIZATION INTRODUCTION SLIDES

STATISTICAL LITERACY SLIDES and LYING WITH STATISTICS

SAMPLE DATA VISUALIZATION SLIDES (GOOD, BAD, and UGLY)
Introduction to Data Visualization
Information visualization is a set of technologies that use visual computing to amplify human cognition with abstract information. [ . . . ]

Information visualizations should do for the mind what automobiles do for the feet.

- Stuart Card, The Human-Computer Interaction Handbook, 539-542
A FEW DEFINITIONS

The graphic design community mostly uses two terms for the visual displays of information: **infographics** and **information design**. In a nutshell, **infographics** stand for visual displays in which graphics (illustrations, symbols, maps, diagrams, etc) together with verbal language communicate information that would not be possible otherwise. Infographics can range from early scientific illustrations of the human body to modern representations of how the brain functions, from early route maps and train schedules to the emblematic London subway map. Journalism as well as technical and pedagogical books employ established practices that traditionally have used infographics to explain complex information and tell stories. From the familiar weather map to visual explanations of natural phenomena and recent facts, infographics help us better understand the news around us.

*Isabel Meirelles, “Design for Information”*
Visual displays of information can be considered cognitive artifacts, in that they can complement and strengthen our mental abilities. I examine the visualizations in relation to the cognitive principles underlying them, which can be a combination of the following:

- to record information;
- to convey meaning;
- to increase working memory;
- to facilitate search;
- to facilitate discovery;
- to support perceptual inference;
- to enhance detection and recognition;
- to provide models of actual and theoretical worlds;
- to provide manipulation of data.

Isabel Meirelles, “Design for Information”
“Effective analytic designs entail turning thinking principles into seeing principles.”

— Edward Tufte
Why Visualize Data?
“Getting information from a table is like extracting sunbeams from a cucumber”

misquote of A. Farquar and H. Farquar, Economic and Industrial Delusions, 1891
Vocabulary of Forms
Lines
The Long-term Future of Hits from 2013
GROWTH, DECAY-RATE OF 2013-RELEASED TRACKS ON SPOTIFY
TRACKS WERE OFFICIALLY RELEASED IN 2013.

FIND A TRACK:
Find an artist or track...
Clear All

- Lorde - Royals
- Daft Punk - Get Lucky...
- OneRepublic - Counting...
- Lana Del Rey - Young a...
- Avicii - Wake me up...
- Robin Thicke - Blurred...
- Pink - Just Give Me a...
- Calvin Harris - I Need...
- Justin Timberlake - MIL...
- Bastille - Pompeii...
- Bruno Mars - When I Wa...
- Mike Curb - We Can't...

http://poly-graph.co/timeless/future.html
The Jobless Rate for People Like You
Not all groups have felt the recession equally.

Distribution of Average Income Growth During Expansions

source: Pavlina R. Tcherneva calculations based on Piketty/Saez data and NBER

- bottom 90%
- top 10%
Pies
http://www.datavis.ca/gallery/evil-pies.php
Scatterplots
UNITED STATES ECONOMIC ACTIVITY, SPLIT IN HALF

GDP

50% 50%
Small Multiples
Recently Voted On

<table>
<thead>
<tr>
<th>Bill</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>H R 4901</td>
<td>Passed the House</td>
</tr>
<tr>
<td>S 1177</td>
<td>Passed the House</td>
</tr>
<tr>
<td>H R 10</td>
<td>Passed the House</td>
</tr>
<tr>
<td>S 1177</td>
<td>Bill Passed the Senate</td>
</tr>
<tr>
<td>H R 5</td>
<td>Passed the House</td>
</tr>
<tr>
<td>H R 1020</td>
<td>Passed the House</td>
</tr>
</tbody>
</table>
Sense-making and communication
Data Visualization is....

“graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication”

– Stephen Few, *Data Visualization for Human Perception*
sense-making

communication
Sense-Making (Data Analysis)
Sense-Making (Data Analysis)

Geo-located donations plotted within France using Gephi
Sense-Making (Data Analysis)
Sense-Making (Data Analysis)
Communication (Story-telling)

Paris, 20 November 1869

The number of men present at any given time is represented by the width of the grey line; one mm. indicates ten thousand men. Figures are also written besides the lines. Grey designates men moving into Russia; black, for those leaving. Sources for the data are the works of messrs. Thiers, Segur, Fontanes, Chambry and the unpublished diary of Jacob, who became an Army Pharmacist on 28 October. In order to visualize the army's losses more clearly, I have drawn this as if the units under prince Jerome and Marshall Davoust (temporarily separated from the main body to go to Minsk and Mthlou, which then joined up with the main army again) had stayed with the army throughout.
## Damage by Commune (Top 10)

Martinique, 1891 Hurricane

<table>
<thead>
<tr>
<th>Commune</th>
<th>Francs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saint Pierre</td>
<td>12,191,800</td>
</tr>
<tr>
<td>Lamentin</td>
<td>6,569,658</td>
</tr>
<tr>
<td>Le Francois</td>
<td>5,665,200</td>
</tr>
<tr>
<td>Le Robert</td>
<td>5,407,100</td>
</tr>
<tr>
<td>Fort-de-France</td>
<td>4,392,320</td>
</tr>
<tr>
<td>La Trinite</td>
<td>4,166,756</td>
</tr>
<tr>
<td>Gros-Morne</td>
<td>2,903,266</td>
</tr>
<tr>
<td>Marin</td>
<td>2,658,720</td>
</tr>
<tr>
<td>Vauclin</td>
<td>2,386,000</td>
</tr>
<tr>
<td>Morne-Rouge</td>
<td>2,357,000</td>
</tr>
</tbody>
</table>
Total Damage by Commune (1891 Hurricane)
Communication (Story-telling)

Geo-located donations plotted within France using Gephi
Communication (Story-telling)
Perception
Chunking
18596746321475030608030504090
70502769843010215346748950213
06057204020503090845064201040
70204070835061305080239245798

Mirielles, Ware
Communication Strategies
Visual hierarchy

Organizing information so the most important details are read first
Position and size
Positioning an element in the entry hotspot, while also altering its size and introducing spacing, establishes its dominance in the hierarchy.

Position, size and emphasis
A final technique is to add extra emphasis to an element to cement its position at the top of the hierarchy – as seen in the use of colour above.
Weight vs Horsepower of Cars

Models built between 1952-1992
Visual consistency

Repeated use of positions, sizes, and colors makes information easier to understand
Rivals go for the kill as Bush falters

Over the top again in Armentas as the French retake battle of the Somme

Still for new weapons may cut strength of US army

Government is 'pressing Waldheim to resign'

Soviet women turned off by their dull-witted husbands
Visual conventions

Follow visual standards of the format you are creating
All About Fujisan
Fujisan, Sacred Place and Source of Artistic Inspiration

- 22nd June 2013
- 3776m
- 11 m/s
- -6°C
- Average wind speed at highest peak
- 21.2°F
- Average temperature at highest peak

25 Sites which reflect the essence of Fujisan's sacred landscape

119 Species of butterfly

180 Species of wild birds

5 Number of lakes
Proportional ink

Visualized data points do not take more ink proportionally than the numbers they represent
Simplicity

As few visual details as necessary

Less colors, less lines, less effects
Lying with Statistics

Drawing by Ben Orlin, www.mathwithbaddrawings.com
Mean  Median
Mode  Range

Statistics
Distribution Matters!

Mode
Median
Mean

Mode is most frequent number
Median is middle number in list
Mean is average of all numbers

Symmetrical distribution
Asymmetrical distribution
What would my starting salary be?

I'll put it this way: our average starting salary is $80,000!
Average: $80,000.

- You: $30,000
- All your coworkers: $30,000 each
- CEO's son: $430,000

Drawing by Ben Orlin, www.mathwithbaddrawings.com
So, why should I invest with you?

Well, not to brag, but my fund has a median gain of 8% per year!
How are you doing on your tests?

My modal category is 70-80%!
<table>
<thead>
<tr>
<th>Score Category</th>
<th>Number of Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>90s</td>
<td>0</td>
</tr>
<tr>
<td>80s</td>
<td>0</td>
</tr>
<tr>
<td>70s</td>
<td>2</td>
</tr>
<tr>
<td>60s</td>
<td>1</td>
</tr>
<tr>
<td>50s</td>
<td>1</td>
</tr>
<tr>
<td>40s</td>
<td>1</td>
</tr>
<tr>
<td>30s</td>
<td>1</td>
</tr>
<tr>
<td>20s</td>
<td>1</td>
</tr>
</tbody>
</table>

Drawing by Ben Orlin, www.mathwithbaddrawings.com
Our students come from a wide range of socioeconomic backgrounds...
Correlation Coefficient

Try our energy drink — it's highly correlated with performance!
Correlation Coefficient
Shows Strength & Direction of Correlation

Strong ← Weak

Weak ← Strong

Negative Correlation
Zero
Positive Correlation
athletic performance

professional athletes are paid to guzzle the stuff

amount of drink consumed
Variance

Larger Variation

Smaller Variation
Variance

These results are a disaster!

Sure, they look bad, but there's a lot of variance!
Don't rush to judgment.
BY ACTUAL TEST

Science proves that tossed pennies come up heads 80 per cent of the time.
Changing the Base

Last year: 50% pay cut

This year: 50% raise

We’re even, Steven!
50% PAY CUT

$100 \div 50\% = \frac{-50}{50}$
50% Pay Cut Restored

\[
\begin{array}{c}
\frac{50}{50\%} \\
\frac{50}{25} \\
\frac{25}{75}
\end{array}
\]
2017 profits: 3%
2018 profits: 6%

Our profits increased 3 percentage points.

vs

Our profits increased 200 percent!
Semiattached Statistic

Survey: “Do you have any complaints about your union?”
RESULTS
--------
200 workers total

156 workers have one or more complaints about their union
REPORTING

78% of workers are opposed to the union

The vast majority of workers do not like the union!
Does this square with the question?

“Do you have any complaints about your union?”
2D Stats in the 3\textsuperscript{rd} Dimension!

USA \\
----- \\
$60/week \\

ROTUNDIA \\
---------- \\
$30/week \\

The Third Dimension!
ROTUNDIA  USA
ROTUNDIA

USA
Best & Worst Visualization Practices
Defense Against the Dark Arts
2012 PRESIDENTIAL RUN
GOP CANDIDATES

BACK PALIN 70%
BACK HUCKABEE 63%
BACK ROMNEY 60%

SOURCE: OPINIONS DYNAMIC
PLANNED PARENTHOOD FEDERATION OF AMERICA:
ABORTIONS UP — LIFE-SAVING PROCEDURES DOWN

SOURCE: AMERICANS UNITED FOR LIFE
Planned Parenthood Federation of America: Abortion vs. Cancer and Prevention Services

Source: http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-choffetz/chart-shown-planned-parenthood-hearing-misleading/
Planned Parenthood changes in services, year-over-year

- Abortion procedures
- Cancer screening / preventative services
Services Provided by Planned Parenthood

- Abortions
- Cancer Screenings & Prevention Services
- STI/STD testing and treatment
- Contraception
- Pregnancy tests and prenatal services
- Other services

Source: http://www.politifact.com/truth-o-meter/statements/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading/
**UNDER PRESIDENT OBAMA,**

MORE STUDENTS ARE EARNING THEIR HIGH SCHOOL DIPLOMAS THAN EVER BEFORE

HIGH SCHOOL GRADUATION RATE

- 2007-08: 75%
- 2008-09: 75%
- 2009-10: 78%
- 2010-11: 79%
- 2011-12: 80%
- 2012-13: 81%
- 2013-14: 82%

SOURCE: U.S. DEPARTMENT OF EDUCATION, NATIONAL CENTER FOR EDUCATION STATISTICS

#LeadOnEducation
High school graduation rates in the US

ATLAS | Data: White House

37
High school graduation rates in the US, 1975 to 2012

- Class of '75: 70%
- Carter: -2.2 ppt. points
- Reagan: +1.2
- Bush: +0.4
- Clinton: -2.2
- Bush: +3.8
- Obama: +5.3
- '10: 80.8%

Source: National Center for Education Statistics
Women in the U.S. Congress
1971-2013

Percentage of Women

© Center for American Women and Politics
Updated 1/8/13

RUTGERS
Center for American Women and Politics
N-Gram, Congressman vs. Congresswoman as a total percentage of mentions (congressman + congresswoman)
The only #climatechange chart you need to see. natl.re/wPKpro
Average global temperature, 1880 to 2014
Activity:
Gun Deaths in Florida
Create a Line Chart

1. Open RStudio
2. Install ggplot package
3. Paste chart code
4. Change variables
5. Screenshot your final chart
6. Compare with class
Gun Deaths in Florida
Number of murders committed using firearms

2005: Florida enacted its 'Stand Your Ground' law
Gun deaths in Florida

Number of murders committed using firearms

2005 Florida enacted its ‘Stand Your Ground’ law

Source: Florida Department of Law Enforcement
C. Chan 16/02/2014
John Self @john_self · Apr 15
@ChristineHHChan Hi, just quick question about this chart from Feb 14:
graphics.thomsonreuters.com/14/02/US-FLORI... Why invert the y-axis? Seems counterintuitive!

Christine Chan @ChristineHHChan
@john_self Thanks for the feedback. I prefer to show deaths in negative terms (inverted). It's a preference really, can be shown either way.
Gun deaths in Florida

Number of murders committed using firearms

Source: Florida Department of Law Enforcement
Gun deaths in Florida

Number of murders committed using firearms

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014
DAY 1 READINGS

Howard Wainer, *Trout in the Milk*

[INSERT R STUDIO INSTALLATION AND BASIC INSTRUCTIONS, 9 PAGES]
Editorial
Big Data Visualization: Promises & Pitfalls

Katherine Hepworth
University of Nevada, Reno
khepworth@unr.edu

A few weeks ago, I was having dinner with a friend when a controversial subject came up. My friend had an extremely strong opinion about the harm caused by vaccination, and his argument went something like this: “I’ve seen the data. There was an infographic laying it all out.” He couldn’t remember specific numbers from the visualization he’d seen or the author of the article. He couldn’t even remember the name of the publication, but the data visualization’s overall argument was firmly lodged in his mind. His situation is not unique, and it provides telling insights on how we, as humans, perceive and respond to big data visualization.

SEEING IS BELIEVING BECAUSE SEEING IS SEDUCTION

People want to believe. From faith practices and fortune-tellers to science and data, the desire to believe in grand visions, and small facts that support those visions, is undeniably human. And we have a tendency to believe what we see more than what we hear, read, or feel. No matter how much evidence we have that seeing is not necessarily believing, the experience of seeing is strongly correlated with truth.

This human quality is as true of the most rigorous researchers, as it is of small children. Seeing things concretizes them in our mind’s eye. It makes them seem more permanent and more real. We’re all less skeptical of information presented in a visualization than information presented entirely as numbers and text. No matter how abstract, complex, or multifaceted we know a given research problem is, information distilled into a simplified visualization is
seductive. Among researchers there is little acknowledgement of or reflection upon the seductive quality of data visualization. This oversight has dangerous implications for research quality, and the human subjects represented through research data visualizations.

**DATA VISUALIZATION DEFINED**

Data visualization, in sum, is the reduction and spatial representation of datasets in such a way as to make them more intelligible than in their pre-visualization, tabular format. Data visualization includes broad categories of spatial representation such as maps, charts, tables, and infographics. In the realm of research tools, data visualization is uniquely persuasive. It’s a fundamentally political practice, one that constantly molds beliefs, behaviors, and emotions, predominantly at a subconscious level. Sociologist Nikolas Rose has called them “little machine[s] for producing conviction in others” (Rose, 2008, p. 36). This persuasive quality of data visualization, is sometimes portrayed as inherently bad. A recent National Geographic article described it as “weaponized data visualization” as if your next chart could be hiding a secret cache of weapons (McGhee, 2015).

This militaristic language is representative of our certainty that “show me, don’t tell me” must be true and the extreme suspicion with which subjectivity in data visualization is held. We rely on data visualization for so much: reporting on sensor networks, displaying critical data, facilitating communication, and helping us make good decisions, both in research and in our everyday lives. Consider, for a moment, life without Google Maps or GPS navigators. While geographic maps are a convenience that we would miss, data visualization also affects our democratic rights in fundamental ways. In the 2000 presidential race, for example, the poor visual organization of the butterfly ballot in Palm Beach County in Florida is said to have caused thousands of voters to vote for someone they did not intend to, costing Al Gore the Presidency (Wand et al., 2001). The thought that so many these valuable functions depend on a subjective, fallible medium is uncomfortable.

I’ll explain the persuasive power of visualization in order to demystify it. This power relates to the potency of immersive
communication as compared with literate communication. Literate communication is of course related to language, reason, and critical thinking. It encompasses all forms of communication that require fluency in written, spoken, or read language. While most communication requires some degree of literacy, certain forms of communication depend on it much more than others. Academic articles, long-form journalistic writing, and detailed instructional manuals are all forms of communication that depend primarily on literate communication. This concept of literacy is the one communication researchers thrive on, and feel most comfortable with.

**IMMERSIVE, EXPERIENTIAL COMMUNICATION**

Then there is immersive communication, the experiential quality. It’s related to aesthetics and emotion. All communication leaves us feeling a certain way, due to many small details that affect our emotional reception; this emotional response is the immersive component. When listening to someone, the cadence, tone, and volume of the words spoken all affect our emotional response to those words. When reading highly literacy-dependent communications such as academic articles, immersive aspects of the communication include the quality of the paper the text is written on, or the brightness of the screen, the font style, size, and weight, and the compositional layout of the article. Text that is published in a hard-to-read font, too-small font, or laid out with lines too close together tend to illicit negative responses irrespective of the content. Generally, the more visually or aurally dependent a communication is, the more it relies on immersive communication. Data visualizations, in-person debates, and video footage are all communication formats that rely heavily on immersive communication.

The potency of such immersive communication is connected to the seductive power of emotions and the ancient fight or flight responses in our brains. Strong emotions trigger what psychologist Daniel Goleman refers to as “emotional hijack” (2012) in which the reasoning parts of our brains are temporarily overridden by emotional responses, which come with overwhelming feelings of righteousness and certainty. The triggering of this physiological
response by immersive communication is the reason advertising is effective, the reason we have art, and the reason we love music. Immersive communication is uncomfortable to contemplate for most researchers. Some may even consider the phrase “immersive communication” a contradiction. But its existence is undeniable.

In the anecdote I relayed earlier, my friend had no recollection of the literate aspects of the data visualization he saw on the subject we were discussing. But he clung firmly to the visceral experience of the immersive communication it conveyed. Immersive communication in research is unavoidable simply because the immersive aspect of communication is unavoidable. No matter how clinical or remote a communicative exchange is, we’re left with an experience of it.

Most research processes and outputs focus on the literate aspects of communication, with little to no attention given to immersive aspects. For example, visual presentation of research findings in academic articles is most commonly kept to a minimum and used exclusively for illustrative purposes. It is possible to communicate complex research findings and arguments in visual formats such as maps, infographics, and charts. But primarily visual formats are not considered acceptable research outputs by most academic publications unless they are presented as supplements to an argument written out in standard academic language. This situation results in research findings being experienced as flat, and boring.

**WHY DATA VISUALIZATIONS ARE ENJOYABLE**

Data visualizations are so compelling because the medium demands equal footing between literate and immersive communication. Without significant attention given to immersive visual elements, such as color, symbols, and spatial relationships, data visualizations are unintelligible. Individual colors, color combinations, line thicknesses, patterns, icons, font styles, weights, and sizes all affect our emotional response to communication. Consider, for example, the differences between presenting statistical information about US states in a black and white table in an academic article, and that same information presented online in
a geographical map with statistical significance represented by color tints. The considered selection and spatial arrangement of these immersive elements contributes to data visualizations being inherently more pleasant than most research outputs, both to look at, and to work with. And so, we like them. It’s this liking that makes data visualizations inherently more persuasive. More productive of experiential, visceral responses that sway our beliefs in ways we’re either completely unaware of, or only vaguely realize.

Subjectivity and persuasion are inevitable in big data visualization. But persuasive doesn’t necessarily mean inaccurate or immoral. Immersive communication is inherently persuasive simply because it is emotionally, rather than rationally, parsed. Despite what common terms such as “visual language,” “visual grammar,” and “visual rhetoric” suggest, visual elements cannot be parsed in the same way as language. Visual presentation of information is far more complex, and far less rule-bound, than language.

In the mid-twentieth century, communication research was dominated by the idea that all communication, including visual communication, was as rule-bound as language. This idea was epitomized in semiotics, a theory that envisaged all communication as one-way transactions involving three elements: sender, message, and receiver. While a few researchers still hold this view, many others have repeatedly shown it to be false. Across many fields—anthropology, communication studies, graphic design, and sociology—the prevailing understanding is now that communication depends on interpretation and mediation. A vast array of individual, cultural, and environmental factors contribute to visual meaning in any given context. These include associations with particular visual elements stemming from cultural and subcultural, community-wide, societal, and national traits. There are also important institutional contexts.

**COLLABORATION IS KEY**

No one, no matter how expert in a particular visual discipline, can understand all uses of particular visual conventions. No amount of visual comprehension at one particular moment in time guarantees
continued comprehension. Visual meaning changes at the same pace, and as intricately, as human culture and physical environment do. For example, two visualization strategies in common usage today, fever charts and donut charts, were unknown only decades ago, while many other charts have fallen out of common usage.

The changing nature of visual meaning affects individual elements within visualizations too. The symbolism of color, for example, is different and at times opposite, from culture to culture and also changes over time within cultures. For example, red is regularly used in so-called “universal” signage worldwide — in airports, train stations, and many public buildings — to indicate danger. While this makes sense in a western context, where red has traditionally been identified with anger and danger it is not so intuitive from a Chinese perspective since red is more commonly associated in Chinese culture with good fortune and luck.

These varied and shifting meanings are part of the reason why data visualization is such a collaborative and interdisciplinary field. Collaboration allows multiple perspectives as well as multiple skill sets and disciplinary backgrounds. Data visualizations draw from the disciplines of graphic design (my area), interaction design, technical communication, data analytics, statistics, math, psychology, and computer science. Collaborations between artists, designers, historians, journalists, scientists, and technical communicators who draw from these disciplines make best practice big data visualizations.

Working with big data shifts the role of data visualization in research. It moves from an optional research output to a necessity for data exploration. With small datasets of say, hundreds or even thousands of data points, it is possible for many researchers to navigate the data in a numerical format to identify patterns or answers to research questions. In a big data context, where data points are at least in the millions, this kind of scanning of numerical data becomes a physiological impossibility. Instead, researchers must rely on visualizations to query and summarize data in order to answer research questions. The subjective, persuasive practice of data visualization becomes an essential part of the research process.
TWO DIRECTIONS FOR BIG DATA VISUALIZATION

Working with big data also changes the nature of visualization. Because big data sets are, by definition, harder to comprehend than small data, big data visualizations need to be better at making data comprehensible. Two trends in best practice in big data visualization, bear this out: interactivity, and real-time updating.

Trend 1: Interactivity

Interactive elements help researchers explore their data sets at various levels of complexity and in various spatial configurations. An increasing number of data analysis tools — NVivo and R, for example — contain interactive visualization modules that allow researchers to visualize a particular set of data points in many different spatial organizations. By hovering or clicking on data groupings and individual data points, the researcher is lead to other visualizations of his or her data, providing new perspectives that may not have been arrived at without this interactivity. Being able to switch between so many views quickly and easily allows researchers to explore their data in previously unimaginable ways.

Still other tools have been created — Graphiq, Immersion, and Tableau, for example — that use interactive visualizations as the primary form of data analysis. These tools hide the numerical complexity of data, instead presenting interactive visualizations as the sum of research findings. Each of the sometimes thousands of individual perspectives enabled by these interactive data visualizations temporarily reduces complexity, cutting through otherwise impenetrable fogs of data, and each providing a unique perspective (Salvo, 2012). This increase in data visualization interactivity has been made possible through a combination of advancements in interaction design, software usability, and Web 2.0 technologies.

Trend 2: Real-Time Updating

The second trend in big data visualization is real-time updating. These interactive visualizations incorporate new data added to datasets instantaneously, through API access to publicly available online databases. This technology allows visualizations to become
integrated with data generating systems, extending the value and therefore shelf-life of visualizations well past their development stage. These are the visualizations that become valuable exploratory research tools.

Interactivity and real-time updating transform visualization, which has traditionally been a two dimensional exercise using static graphics. These traditional, static graphs and maps are snapshots of data frozen in time. They can be very effective research findings; they can even be beautiful. But best practice big data visualization has moved beyond the static page to applications and websites that display live feeds of data continuously in four dimensions, the fourth dimension being time.

At their best, big data visualizations put data into a human context by relating scientific and statistical insights to environmental and social contexts. They highlight perspectives about our world, and our societies, that we can collectively benefit from. Most importantly, they build understanding between users and the people who interact with the data generating systems we study by fostering respect and empathy for people and situations of which we would otherwise be unaware.

**Example 1: Histography**

The “Histography” project by interactive designer Matan Stauber (see Figure 1), for example, visualizes the entire contents of Wikipedia as an interactive timeline. Shaped like a reflected histogram, it records every page on Wikipedia as a single dot.
Figure 1: Histography big data visualization site, organizing. Top image shows one event on rollover, bottom image shows extra information on click.

Videos and photos are pulled from Wikipedia, and converted into enticing visual rollover effects for each documented event. Histography encourages us to choose pages randomly, and we’re rewarded with surprising footage of people, experiments, and events from the past. Clicking on a rollover image brings up both more details from Wikipedia and the option to read the entire article within Histography. As new articles are added to Wikipedia, Histography automatically updates. Upon selecting a date range,
Histography highlights the events within that range that are mentioned in the greatest number of Wikipedia pages.

By presenting such a comprehensive timeline, Histography contextualizes important historical events and scientific discoveries in a way that easily counteracts revisionist histories based on religious beliefs and pseudo-science. It debunks myths at the same time as being entertaining, thoughtful, and fun. Histography also presents a completely original way to investigate historical events, one that gets around a significant challenge in historical investigation: avoiding confirmation bias, the unconscious tendency to look for information that confirms one’s preconceived ideas. Histography is by no means perfect. Given that Wikipedia, on which it’s based, is renowned for unreliable content, Histography can’t be used as an authoritative source. But it does show the potential of big data visualization to be accessible, enjoyable, beautiful, and in the public interest.

**Example 2: The Drone Papers**

At their worst, big data visualizations help a few corporations make vast profit off of human suffering. The big data of modern surveillance and warfare is also big business. This is the side of big data revealed in the Drone Papers, obtained by investigative news outlet *The Intercept*. The “Finish ops” represented on this map (see Figure 2) are operations to assassinate individuals using missiles. This is the big data of dehumanization.

![Figure 2: Images from the Drone Papers, US drone program documents obtained by The Intercept. This image shows drone strikes in Yemen.](image
In another image from the Drone Papers (see Figure 3), a similar kind of dehumanization is created visually.

Figure 3: Images from the Drone Papers, US drone program documents obtained by The Intercept. This image shows deaths and injuries from drone-delivered missiles.

In this image, E.K.I.A. stands for “enemy killed in action.” This is the default government classification for everyone killed by United States-operated, drone-delivered missiles. The word “enemy” is used despite the US military’s own finding that only 10% of the people killed in this way are their intended targets. Civilians, bystanders, ordinary people going about their everyday activities, just like you and I, are listed as “enemies killed in action.” It bears repeating because it’s so unbelievable. These big data visualizations are shocking only in their banality. The language is ambiguous at best, and intentionally misleading at worst. The visualizations are crude and simplistic, obfuscating the big data underpinnings of the drone program. And I argue that they’re intentionally so, reducing such large scale loss of life, to administrative details.

Most big data visualization falls between these two extremes. Frequently for profit, but without causing harm. One growing trend, in big data visualizations falling between the extremes, is combining interactive technology with design and gaming principles to incorporate big data into interactive, entertaining narratives.
Example 3: Interactive Wall, US Tennis Open 2013

For example, design agency Hush made a fifteen-foot interactive wall for attendees of the US Tennis Open to enjoy. The interactive wall allowed tennis fans to test their knowledge of player statistics in gameplay that is something like Wii Fit meets Tinder. This beautiful, fun interface wove vast sums of tennis data, that would otherwise be overwhelming, into an inviting and enjoyable experience.

The interactive wall epitomizes the pleasurable potential, as well as the economic realities, of big data visualization. Created as an elaborate promotional tool for IBM and the US Open, the interactive wall was only enjoyed by those who could afford expensive tickets to the exclusive sporting event. This one example is representative of the elitism and profit motive currently driving corporate research into big data visualization (Schwartz, 2015).

A CALL FOR EMPATHETIC BIG DATA VISUALIZATIONS

While big data visualization has the potential to benefit large amounts of people and tackle global issues, current development and spending indicates this won’t be its main use. Instead, big data visualization is likely to continue to be developed for the use, and benefit, of relatively few, privileged communities. For better or worse, we are one of those communities.

In this piece, I’ve provided examples of a few different uses of big data visualizations. To keep the inevitable persuasion in big data visualizations within moral bounds, we must apply the same ethical rigor to our visualizations as we do to other aspects of our research. Renowned information designer Alberto Cairo has advocated for development of a field of “ethics of data visualization” (Cairo, 2014). He argues that the fundamental goal of data visualization should be to make people better informed. I would add to this that big data visualization should help people gain empathy for each other’s situations. Big data visualizations can only fulfill this aim when the teams working on them have large enough, and critical enough, understandings of the
persuasive, visual elements from which they are built. When we’re well informed, we can use the persuasive qualities of big data visualization to better inform others and to foster empathy.

We have a responsibility to educate ourselves about the literate and immersive aspects of various design elements within the communities where we want to share data. We owe this to ourselves, the agencies who fund our research, and the public. We also have a responsibility to gain enough understanding of the practical and ethical considerations of visualization in those communities, to wield the persuasive power of big data visualization for the common good. An important step on this path is acknowledging the persuasive nature of big data visualization and the inherent risks in its use. To understand, and take seriously, how visualizations affect both our conscious and unconscious judgments is to reduce the risk of misuse.

REFERENCES


Introduction

Let me begin with a few kind words about the bubonic plague. In 1538, Thomas Cromwell, the Earl of Essex (1485–1540),* issued an injunction (one of seventeen) in the name of Henry VIII that required the registration of all christenings and burials in every English parish. The London Company of Parish Clerks compiled weekly Bills of Mortality from such registers. This record of burials provided a way to monitor the incidence of plague within the city. Initially, these Bills were circulated only to government officials, principal among them the Lord Mayor and members of the King’s Council.† They were first made available to the public in 1594, but were discontinued a year later with the abatement of the plague. However, in 1603, when the plague again struck London, their publication resumed on a regular basis.

The first serious analysis of the London Bills was done by John Graunt in 1662, but in 1710, Dr. John Arbuthnot, a physician to Queen Anne, published an article that used the christening data to support an argument (probably tongue-in-cheek) for the existence of God. These data also provide supporting evidence for the lack of existence of graphs at that time.

Figure 1 is a simple plot of the annual number of christenings in London from 1630 until 1710. The preparation of such a plot is straightforward, certainly requiring no more complex apparatus than was available to Dr. Arbuthnot in 1710. Moreover, as we will see in a moment, it is quite informative. Yet it is highly unlikely that Arbuthnot, or any of his contemporaries, ever made such a plot.

The overall pattern we see in figure 1 is a trend over these eighty years of an increasing number of christenings, almost doubling from 1630 to 1710. A number of fits and starts manifest themselves in substantial jiggles. Yet each jiggle, save one, can be explained. Some of these explanations are written on the plot. The big dip that began in 1642 can only partially be explained by the onset of the English Civil War. Surely the chaos common to civil war can explain the initial drop, but the war ended in 1649 with the beheading of Charles I at Whitehall, whereas the christenings did not return to their earlier levels until 1660.‡ Graunt offered a more complex explanation that involved the distinction between births and christenings, and the likelihood that Anglican priests would not enter children

---

* Thomas Cromwell was an English statesman who had a successful career as an administrator and advisor to the king. Among his other accomplishments, he arranged Henry VIII’s divorce from Catherine of Aragon (and was largely responsible for the beheading of Sir Thomas More [1478–1535] in the process). Ironically, five years later, he was done in by Henry’s aversion to Anne of Cleves, a consort of Cromwell’s choosing, when he was consequently sent to the Tower of London and beheaded. Perhaps specifying the consequences of failure in this way would provide a workable pathway toward helping to improve the efficacy of computer dating services.

† This exposition is heavily indebted to the scholarly work of Sandy Zabell, to whose work the interested reader is referred for a much fuller description (Zabell 1976). It was Zabell who first uncovered Arbuthnot’s clerical error.

‡ The year 1660 marked the end of the protectorate of Oliver Cromwell and the beginning of the restoration.
born to Catholics or Protestant dissenters into the register.

Many of the other irregularities observed are explained in figure 1, but what about the mysterious drop in 1704? That year has about four thousand fewer christenings than one might expect from observing the adjacent data points. What happened? There was no sudden outbreak of war or pestilence, no great civil uprising, nothing that could explain this enormous drop.

The plot not only reveals the anomaly, it also presents a credible explanation. In figure 2, I have duplicated the christening data and drawn a horizontal line across the plot through the 1704 data point. In doing so we immediately see that the line goes through exactly one other point—1674. If we went back to Arbuthnot’s table we would see that in 1674 the number of christenings of boys and girls were 6,113 and 5,738, exactly the same number as he had for 1704. Thus the 1704 anomaly is likely to be a copying error! In fact, the correct figure for that year is 15,895 (8,153 boys and 7,742 girls), which lies comfortably between the christenings of 1703 and 1705 as expected.

It seems reasonable to assume that Arbuthnot, upon seeing such an unusual data point, would have investigated and, finding a clerical error, would have corrected it. Yet he did not. He did not, despite the fact that when graphed the error stood out, literally, like a sore thumb. Thus we must conclude that he never graphed his data. Why not?

The remarkable answer to this question occupies most of part I of this book. In brief,
it is that the very idea of graphing data was not yet invented. The story of its invention and inventor is a fascinating one that mixes science and politics, intrigue and scandal, revolution and shopping. The principal actors in this drama include some of the leading scientists of the day as well as a scoundrel of the first order. It also includes a French physician, a scientist of great renown, two signers of the American Declaration of Independence, and America’s third president.

Part II illustrates the power of this marvelous invention to help us understand the modern world. It contains examples as disparate as a murder trial in Connecticut, the effect of the Vietnam war on college admissions, faxing policies in Canada, preposterous sports cars, the Boston Marathon, and the perambulations of the stock market over the past century. In each of these situations we see how a graphic depiction of data can help us see things that otherwise would have been as invisible to us as Arbuthnot’s clerical error was to him.

Part III of this book looks to the future. I begin with a brief examination of the life and work of John Wilder Tukey, one of the twentieth century’s great geniuses. I then describe some of the sorts of graphical methods, closely tied to modern computing, that are likely to become the bedrock of the way that we will be able to understand the torrent of data that has become the standard in our information-laden society.

Superficially, this is a book celebrating data
display. But that phrase has two parts: data and display. Fine display formats are worthless without important data, and so in celebrating the marvelous invention of data-based graphics we must also save some kudos for those who undertook the often thankless task of gathering the data.* Where would the Johannes Keplers of history have been without the Tycho Brahes?

Graphics evolved because (i) there was a growing recognition that important questions could be answered with data,† (ii) data were being gathered to aid in the quest for such answers, and (iii) graphs were the best way to find both the structure and the surprises hidden in data. The London Bills of Mortality were one of the earliest organized efforts to gather extensive longitudinal data in the hope of aiding the government in its search for an effective policy for dealing with public catastrophes such as the bubonic plague. Thus, let me muster a weak thank-you to the plague for providing the immediate motivation for the sort of data gathering that has been the backbone of public health policy in the four hundred years since.

* It is critical, in our efforts to gain a glimpse of the cosmos, that we do not forget the importance of the demanding and sometimes boring intellectual regimen without which discovery is impossible. The famous conductor Pierre Montreux, rich in years and reputation, was once told by a young conductor in one of his master classes at Tanglewood that he, the student, was desperately seeking the “meaning” of the “ineffable essence of Mozart.” Montreux congratulated him on his high aim, and then said it would do him no harm, meanwhile, to “learn how to keep the beat.”

† In the mid-eighteenth century, Samuel Johnson pointed this out when he said, “To count is modern practice, the ancient method was to guess,” but he was far from the first to recognize this, as Seneca too was aware of the difference: “Magnum esse solvem philosophus probabit, quantus sit mathematicus.” (“The philosopher says the sun is large, but the mathematician measures it.” Epistulae 88.27)
The graphic explosion of the nineteenth century that manifested itself in the publication of atlases in all aspects of the observational sciences had its origin in the intellectual turbulence of the eighteenth century. The hundred-year span between 1750 and 1850 saw a shift in the language of science from words to pictures.*

This shift began with the historical time charts of Jacques Barbeau-Dubourg (see chapter 7), which were followed closely by similar efforts from the Scottish philosopher Adam Ferguson and twelve years later from Joseph Priestley (chapter 5). Indeed, this idea proved so useful that Thomas Jefferson even used it to keep track of the price of vegetables in the Washington market (chapter 6). The scientific use of graphic displays had its origin with the Dutch polymath Christiaan Huygens (1629–1693), who developed the first survival chart in 1669, and accelerated when Martin Lister provided graphical summaries of weather data before the Oxford Philosophical Society on March 10, 1683 (chapter 1), which was quickly followed by many others who wanted to understand better the weather data that were now available with the invention of the barometer. All of the plots of these worthy scholars were precursors to the work of the Scot William Playfair (1759–1823), whose Commercial and Political Atlas of England and Wales contained no maps but instead beautifully polished versions of most of the common graphical forms in use today. This atlas, in which modern techniques of the graphical presentation of quantitative phenomena emerged fully developed, earned Playfair my nomination for the title of father of modern graphical display. (See chapters 1–4 for more about Playfair’s contributions and his remarkable life.) The traditions of graphical display and weather were brought into a wonderful consilience by Francis Galton, who mapped weather reports of December 1861, drawn from all over western Europe, to show for the first time phenomena that had previously only been the hinted at (chapter 8). With Galton, the case for a transition to graphical representation of scientific phenomena was complete.

But others before Galton had already been converted. In 1878, the French physiologist Etienne Marey understood the value of graph-

* After the perfection of engraving in the fifteenth century, scientific documents often included lots of pictures, including abstract diagrams. What did not appear until after Playfair (with a few exceptions such as musical notation and those examples I give in chapter 1) was the diagramming of variates from empirical observations that went beyond the scale translations of space and time as movement in space (Biderman 1990).
His graphic schedule of all the trains between Paris and Lyons (see figure I.1) provides a powerful illustration of the breadth of value of this approach. And, on the off chance that someone might have missed the point, he provided an explicit conclusion: “There is no doubt that graphical expression will soon replace all others whenever one has at hand a movement or change of state—in a word, any phenomenon. Born before science, language is often inappropriate to express exact measures or definite relations.”

Marey was also giving voice to the movement away from the sorts of subjectivity that had characterized prior science in support of the more modern drive toward objectivity. Although some cried out for the “insights of dialectic,” “the power of arguments,” “the insinuations of elegance,” and the “flowers of language,” their protestations were lost on Marey, who dreamed of a wordless science that spoke instead in high-speed photographs and mechanically generated curves, in images that were, as he put it, in the “language of the phenomena themselves.”

Historians have pointed out that “Let nature speak for itself” was the watchword of the new brand of scientific objectivity that emerged at the end of the nineteenth century. “At issue was not only accuracy but morality as well: the all-too-human scientists must, as a matter of duty, restrain themselves from imposing their hopes, expectations, generalizations, aesthetics, and even their ordinary language on the image of nature.”

Mechanically produced graphic images would take over when human discipline failed. Marey and his contemporaries turned to mechanically produced images to eliminate human intervention between nature and representation. “They enlisted polygraphs, photographs, and a host of other devices in a near-fanatical effort to produce atlases—the bibles of the observational sciences”—documenting birds, fossils, human bodies, elementary particles, flowers, and economic and social trends that were certified free of human interference.

The problem for nineteenth-century atlas makers was not a mismatch between world and mind, as it had been for seventeenth-century epistemologists, but rather a struggle with inward temptation. The moral remedies sought were those of self-restraint: images mechanically reproduced and published, warts and all; texts so laconic that they threatened to disappear entirely. Seventeenth-century epistemology aspired to the viewpoint of angels; nineteenth-century objectivity aspired to the self-discipline of saints. The precise observations and measurements of nineteenth-century science required taut concentration endlessly repeated. It was a vision of scientific work that glorified the plodding reliability of the bourgeois rather than the moody brilliance of the genius.

The graphical representation of scientific phenomena served two purposes. Its primary function was standardizing phenomena in visual form, but it also served the cause of publicity for the scientific community. It preserved what was ephemeral and distributed it to all who would purchase the volume, not just the lucky few who were in the right place at the right time with the right

---

* Marey’s view that graphical display could avoid the problems of language echoed the insight of his ill-fated countryman, Louis XVI, who expressed this view after receiving and reading a copy of Playfair’s Atlas (see chapter 2).

† Although with such contributors as Condorcet (1743–1794), von Humboldt (1769–1859), and Florence Nightingale (1820–1910), there was certainly room for genius in the eighteenth and nineteenth centuries. Indeed Galton’s weather maps, developed in the middle of the nineteenth, show how plodding reliability, when adjoined with moody brilliance, can yield especially fruitful results (chapter 8), yet no one would doubt that Robert Plot was a plodding plotter.
Figure I.1. Etienne Marey's (1878) "graphical train schedule" showing the daily passage of all of the trains between Paris and Lyons. This version, prepared by Edward Tufte (1983), uses a gray grid and is reproduced with his permission.
equipment. And it served the cause of memory, for images are more vivid and indelible than words.

But the graphical display of natural phenomena was viewed as yet more. Marey, in an accompanying note to his design of a portable polygraph, which automatically registered a variety of measures, suggested that through the use of graphics scientists could reform the very essence of scientific research and scientific evidence. “The graphic method translates all these changes in activity of forces into an arresting form that one could call the language of the phenomena themselves, as it is superior to all other modes of expression” (p. iv). Such a language was, for Marey, universal in two senses. Graphical representation could cut across the artificial boundaries of natural languages to reveal nature to all people, and it could cut across disciplinary boundaries to capture phenomena as diverse as the pulse of a heart and the downturn of an economy. Pictures became more than merely helpful tools: they were the words of nature herself.*

* This simple? Perhaps not. An alternative thesis to the one that characterizes science’s task as capturing the glorious revelations by nature of her sublime design is one that sees man imposing the order of his senses and his arts upon the unheavenly disorder amidst which he finds himself.
“Getting information from a table is like extracting sunbeams from a cucumber” (Farquhar and Farquhar).* This evocative indictment of data tables by two nineteenth-century economists comes as no great insight to anyone who has ever tried to draw inferences from such a data display. For most purposes we almost always prefer a graphical representation. Indeed, graphs are ubiquitous now; hence it is hard to imagine a world before they existed. Yet data graphs are a human invention, indeed a relatively modern one. Data-based graphics began to make an appearance in the mid-seventeenth century but their full value and great popularity can be traced to a single event and a single person. The event was the publication, in 1786, of a small atlas describing the imports and exports of England and Wales with their various trading partners. The atlas contained forty-four graphs and no maps. Its author was a twenty-seven-year-old Scot named William Playfair, and his Commercial and Political Atlas forever changed the way that we look at data.

William Playfair (1759–1823) worked as a draftsman for James Watt and was the ne’er-do-well younger brother of the well-known scientist John Playfair (1748–1814).† William Playfair is often credited with being the progenitor of modern statistical graphics. Most histories of statistical graphics give him huge credit while acknowledging important graphical work that preceded him.1 A balanced summary is that he invented many of the currently popular graphical forms,‡ improved the few that already existed, and broadly popularized the idea of graphic depiction of quantitative information. Before Playfair, statistical graphics were narrowly employed and even more narrowly circulated. After him, graphs popped up everywhere, being used to convey information in the social, physical, and natural sciences.

The title comes from Albert Biderman’s private characterization of the question that has intrigued him for more than a decade. The intellectual contents of this chapter come principally from two sources: Biderman’s 1978 talk at the Leesburg, Virginia, conference on Social Graphics (and its published elaboration, Biderman 1990) and Patricia Costigan-Eaves and Michael Macdonald-Ross’s extensive, but as yet unpublished, history of early graphic developments. Some of their material is in Costigan-Eaves and Macdonald-Ross 1990.

* This well-known quotation, though pithy, is somewhat inaccurate. What the brothers Farquhar actually wrote (1891, p. 55) was, “The graphical method has considerable superiority for the exposition of statistical facts over the tabular. A heavy bank of figures is grievously wearisome to the eye, and the popular mind is as incapable of drawing any useful lessons from it as of extracting sunbeams from cucumbers.”

† John Playfair’s activities were remarkably varied: minister, geologist, mathematician, and professor of natural philosophy at Edinburgh University. In fact, in 1805 William thanked his brother for the idea of using “lines applied to matters of finance” that William used in his 1786 book. We can only speculate why it took him nineteen years to give his brother some credit.

‡ Invented is perhaps too strong a term. It may be more accurate to refer to him as an important deployer of graphical forms. He did not invent so much as he permuted and manip-
Before we meet Playfair, however, it is worthwhile to step back a century or so and examine the attitudes that prevailed in scientific investigations. Because natural science originated within natural philosophy, it favored a rational rather than an empirical approach to scientific inquiry. Such an outlook was antithetical to the more empirical modern approach to science, which does not disdain the atheoretical plotting of data points with the goal of investigating suggestive patterns. Graphs that were in existence before Playfair (with some notable exceptions that I will discuss shortly) grew out of the same rationalist tradition that yielded Descartes’s coordinate geometry—that is, the plotting of curves on the basis of an a priori mathematical expression. For example, Oresme’s “pipes” on the first page of the Padua edition of his 1486 Tractatus de latitudinibus formarum (figure 1.1) is often cited as an early example.*

This notion is supported by statements such as that of Luke Howard, a prolific graphe r of data in the late eighteenth and early nineteenth centuries who, as late as 1844, apologized for his methodology and referred to it as “an autograph of the curve . . . confessedly adapted rather to the use of the dilettanti in natural philosophy than that of regular students.”

All the mechanical pieces necessary for data-based graphics were in place long before Playfair. For example, a primitive coordinate system of intersecting horizontal and vertical lines that enable a precise placement of data points was used by surveyors of the Nile flood basin as early as 1400 B.C. A more refined coordinate system was used by Hipparchus (ca. 140 B.C.), whose terms for the coordinate axes translate into Latin as *longitudo* and *latitudo*.

* Clagett (1968) argued convincingly that this work was not written by Oresme, but probably by Jacobus de Sancto Martino, one of his followers, in about 1390—yet another instance of how surprisingly often eponymous referencing is an indication only of who did not do it (Stigler 1980).
latitudo, to locate points in the heavens. Somewhat later, Roman surveyors used a coordinate grid to lay out their towns on a plane that was defined by two axes. The decimani were lines running from east to west, and the cardin ran north to south. Many other special-purpose coordinate systems were in wide use before Playfair: for example, musical notation placed on horizontal running lines was in use as early as the ninth century, and the chessboard was invented in seventh-century India.

One of the earliest examples of printed graph paper dates from about 1680. Large sheets of paper engraved with a grid were apparently printed to aid in designing and communicating the shapes of the hulls of ships. Many historians describe Descartes’s 1637 development of a coordinate system as an important intellectual milestone in the path toward statistical graphics. More recent work interprets this in exactly the opposite way—as an intellectual impediment that took a century and a half and Playfair’s eclectic mind to overcome.

Although the use of coordinate grids is very old, graphic encoding of information is older still. Paleolithic cave art provides an early and very striking example of graphic display. Some Ice Age bone carvings of animals are intermixed with patterns of dots and strokes that some archeologists have interpreted as a lunar notation system related to the animals’ seasonal appearance. These are almost identical in structure, as well as degree of detail, to the engraving on the hull of the Pioneer 10 spacecraft that shows a drawing of a man and a woman along with a simple plotting of the Earth’s location by dotted pulsar beams.

* This material is classed in the “collection” category of the British Library with the entry, “A collection of engraved sheets of squared paper, whereon are traced in pencil or ink the curves or sweeps of the hulls of sundry men-of-war.”

So we have the ideas of graphic encoding of information and a coordinate grid system. Why not the plotting of data? Well, some data were plotted. Let us consider three examples.

**Example 1. Pliny’s Ninth-Century Astronomical Charts**

Pliny’s (ca. 810) astronomical data were plotted in a roughly circular form (see figure 1.2) corresponding to the varying locations of the bodies in the heavens. But these graphs (astronomical maps) were not as useful as they might have been, because locating a body in the circular path required visually tracking the
complete cycle. A manuscript originating in Auxerre toward the end of the ninth century contained a scheme that remedied this by transforming the circular grid into a rectangular one (see figure 1.3). The cyclic nature of the orbit is less apparent, but by making explicit the time (horizontal axis) and height above the horizon (vertical axis) it made locating and identifying a heavenly body somewhat easier.*

* There is actually less here than meets the eye. The horizontal axis is not particularly well defined. It is really an unfold-

Example 2. Christiaan Huygens's Seventeenth-Century Survival Charts

On October 30, 1669, the Dutch polymath Christiaan Huygens (1629–1693) received a letter from his brother Lodewijk containing some interpolations of life expectancy data taken from John Graunt's 1662 book Natural and Political Observations on the Bills of Mortality. Christiaan responded in letters
dated November 21 and 28, 1669, with graphs of those interpolations. Figure 1.4 contains one of those graphs showing age on the horizontal axis and number of survivors of the original birth cohort on the vertical axis. The curve drawn was fitted to his brother’s interpolations.* The letters on the chart are related to an associated discussion on how to construct a life expectancy chart from this one—that is, analyzing a set of data to gain deeper insights into the subject. Christiaan constructed such a chart and indicated that it was more interesting from a scientific point of view; figure 1.4, he felt, was more helpful in wagering.†

Example 3. Robert Plot’s Seventeenth-Century Plots of Barometric Pressure

Good graphs can make difficult problems trivial. We have all become used to weather forecasts that are very accurate and detailed for a

* Huygens’s twenty-two-volume Oeuvres complètes (1888–1950) contains many other graphical devices to be explored by anyone with fluency in ancient Dutch, Latin, and French. Incidentally, Huygens’s graphical work on the pendulum proved to him that a pendulum’s oscillations would be isochronic regardless of its amplitude. This discovery led him to build the first clock based on this principle.

† This scooped a 1976 paper by the Chicago statistician Sandy Zabell, whose graphical analysis of the Bills of Mortality found inconsistencies, clerical errors, and a remarkable amount of other information. “much of it unappreciated at the time of their publication” (Zabell 1976, p. 27). Zabell’s point, though implicit in his paper, is important in this discussion. As we illustrated in one situation in the introduction to this book, this was strong evidence that graphic display was not widely available. For had they been seen, these errors, which could not be missed with any sort of competent display, would have been discovered and eliminated.
Figure 1.5. Robert Plot’s (1685) “History of the Weather” recording of the daily barometric pressure in Oxford for the year 1684. Appears in Philosophical Transactions and is based on the original work of Martin Lister. Photo © The Royal Society.
day or two and pretty good for as far in advance as a week. I used to think that this was due to the increasing sophistication of complex prediction models.* But then I noticed the weather maps shown on every news broadcast. Using a model of no greater sophistication than that employed by Benjamin Franklin (weather generally moves from west to east), I was able to predict that the area of precipitation currently over Ohio would be hitting New Jersey by tomorrow and would stay over us until the weekend. Any fool could see it. The improvement in forecasting has not been entirely due to improvements in the mathematical models of the weather. The enormous wealth of radar and satellite data summarized into a multicolored and dynamic graphic can turn anyone into an expert.

The path to modern weather graphs is more than three hundred years long. The barometer was developed in 1665. Robert Plot recorded the barometric pressure in Oxford every day in 1684 and summarized his findings in a remarkably contemporary graph (figure 1.5) that he called a “History of the Weather.” He sent a copy of this graph with a letter to Martin Lister† in 1685 with a prophetic insight:

For when once we have procured fit persons enough to make the same Observations in many foreign and remote parts, how the winds stood in each, at the same time, we shall then be enabled with some grounds to examine, not only the coastings, breadth, and bounds of the winds themselves, but of the weather they bring with them; and probably in time thereby learn, to be forewarned certainly, of divers emergencies (such as heats, colds, dearths, plague, and other epidemical distempers) which are not unaccountable to us; and by their causes be instructed for prevention, or remedies: thence too in time we may hope to be informed how far the positions of the planets in relation to one another, and to the fixed stars, are concerned in the alterations of the weather, and in bringing and preventing diseases and other calamities . . . we shall certainly obtain more real and useful knowledge in matters in a few years, than we have yet arrived to, in many centuries.

With so many predecessors, why have I chosen Playfair as my candidate for the father of modern graphical display? Although the arguments will build in subsequent chapters, a visual comparison of a sampling of Playfair’s plots with any of those that came before him makes clear the qualitative jump that Playfair’s work represents. Ian Spence points out that Playfair was the first to use hachure, color, and area to represent quantities in a systematic way. Moreover, Playfair published these forms in widely circulated volumes. Most of all, however, Playfair’s graphs provided proof that the presentation of evidence could be beautiful.

As but one such example, compare Playfair’s harmonious depiction of more than a century of data on England’s national debt summaries of weather data before the Oxford Philosophical Society on March 10, 1683, and later in the same year presented a modified version to the Royal Society. Plot was not the only one enthusiastic about Lister’s graphical methods. William Molyneux was so taken that he had an engraving made of the grid, and he faithfully sent a “Weather Diary” monthly to William Musgrave. One of Molyneux’s charts was reproduced in Gunther (1968).

* It is true that models are more sophisticated than they were in the past. I was enormously impressed when some surprising turns in a hurricane’s path were predicted well in advance, but such models seem to be needed no more often than seldom.
† The origin of the graphical depiction of weather data, sadly, for the obvious eponymous glory, rests not with Plot but rather with Lister, who presented various versions of graphical

DHSI | ETHICALVIZ | 159
(figure 1.6) with any of the graphs produced previously by others. The viewer’s eye is drawn from the soaring debt to the vertical lines that communicate the events that presaged a change in the debt. The viewer’s mind cannot avoid making the causal inference suggested. Nothing that had been produced before was even close. Even today, after more than two centuries of graphical experience, Playfair’s graphs remain exemplary standards for clear communication of quantitative phenomena.

It now seems wise to recapitulate the argument: Graphical forms were available before Playfair, but they were rarely used to plot empirical information. I argue that this was because there was an antipathy toward the empirical approach. This suggestion is supported by statements such as that made by Luke Howard. But at least sometimes when data were available (for example, Pliny’s astronomical data, Graunt’s survival data, Plot’s weather data, and several other admirable uses), they were plotted. Could it be that the exponential increase in the use of graphics after the publication of Playfair’s Atlas was merely concomitant to the exponential growth in the availability of data? Or did the availability of graphic devices for analyzing data encourage data gathering? And why Playfair? Was he merely at the cusp of an explosion in data gathering and so his graphic efforts appear causal? Or did he play an important role in that explosion?

The consensus of scholars is that until Playfair “many of the graphic devices used were the result of a formal and highly deductive science. . . . This world view was more comfortable with an arm-chair, rationalistic approach to problem-solving which usually culminated in elegant mathematical principles” often associated with elegant geometrical diagrams. The empirical approach to problem solving, a critical driving force for data collection, was slow to emerge. But the empirical approach began to demonstrate remarkable success in solving problems, and with improved communications, the news of these successes and hence the popularity of the associated graphic tools began to spread quickly.

So the picture is almost complete. The fundamental tools for the graphical display of data were available; there was an increase in the acceptance of an empirical approach to science as an important part of the scientific process; data were being gathered in greater and greater quantities; and the success of empirical procedures in solving important practical problems was being more widely communicated. This explains the growth of the graphical method, but still leaves the initial question, “Why Playfair?”

We are accustomed to intellectual diffusion taking place from the natural and physical sciences into the social sciences; certainly that is the direction taken for both calculus and the scientific method. But statistical graphics in particular, and statistics in general, went the reverse route. Although, as we have seen, there were applications of data-based graphics in the natural sciences, only after Playfair applied them in the social sciences did their popularity begin to accelerate. Playfair should be credited with producing the first chartbook of social statistics; indeed, publishing an Atlas that contained not a single map is one indication of his belief in the methodology (to say nothing of his chutzpah). Playfair’s work was immediately admired, but emulation, at least new journals appeared, and between 1790 and 1800 twenty-five more (McKie 1972)

* The first encyclopedia in English appeared in 1704. The number of scientific periodicals began a rapid expansion at the end of the eighteenth century; between 1780 and 1789 twenty-five new journals appeared, and between 1790 and 1800 twenty-five more (McKie 1972)
Figure 1.6. A close facsimile of William Playfair’s plot of England’s national debt from 1688 until 1786. It appeared in his *Commercial and Political Atlas* and accompanied his discussion arguing against the British government’s policy of financing its colonial wars through debt.
in Britain, took a little longer (graphics use started on the continent a bit sooner). Interestingly, one of Playfair’s earliest emulators was the banker S. Tertius Galton (the father of Francis Galton, and hence the biological grandfather of modern statistics), who in 1813 published a multilane time-series chart of the money in circulation, rates of foreign exchange, and prices of bullion and of wheat.* The relatively slower diffusion of the graphical method back into the natural sciences provides additional support for the hypothesized bias against empiricism there. The newer social sciences, having no such tradition and faced with both problems to solve and relevant data, were quicker to see the potential of Playfair’s methods.

Playfair’s graphical inventions and adaptations look contemporary. He invented the statistical bar chart out of desperation, because he lacked the time-series data required to draw a line showing the trade with Scotland and so used bars to symbolize the cross-sectional character of the data he did have. Playfair acknowledged Priestley’s priority in this form, although Priestley used bars to symbolize the life spans of historical figures in a time line.† (See chapters 5, 6, and 7 for more on the fascinating history of time lines and graphical display of historical data.)

Playfair’s role was crucial for several reasons, but it was not for his development of the graphical recording of data; others preceded him in that. Indeed, in 1805 he pointed out that as a child his brother John had him keep a graphical record of temperature readings. But Playfair was in a remarkable position. Because of his close relationship with his brother and his connections with James Watt, he was on the periphery of science. He was close enough to know of the value of the graphical method, but sufficiently detached in his own interests to apply them in a very different arena—that of economics and finance.

* Biderman (1978) pointed out that, ironically, Galton’s chart predicted the financial crisis of 1831 that created a ruinous run on his own bank.
† Priestley’s use of the bar as a metaphor is somewhat different than Playfair’s in that the data were not really statistical. A much earlier precedence has been recovered from its resting place in the Qumran caves abutting the Dead Sea. The graphic dates from approximately 1400 B.C. and was prepared as a summary of population changes in the twelve tribes of Israel as they emerged from their almost four decades of wandering in the Sinai after their exodus from Egypt, which began in April 1446 B.C. A faithful copy of this bar chart, with the captions and legends translated from Aramaic, is reproduced here as figure 1.7. Some aspects of this historic figure have been computer-enhanced for better reproduction. Note that it presages Huygens in subject matter and Playfair in form.

Figure 1.7. A translated and computer-enhanced reproduction of perhaps the earliest statistical graphic yet uncovered. It was apparently constructed about 1400 B.C. and was preserved in a sealed ceramic container in the Qumran caves. It was purchased by the author from Moishe the mapman at his Dead Sea antiquities stall in 1991.
These areas, then as now, tend to attract a larger audience than matters of science, and Playfair was adept at self-promotion.

In a review of his 1786 Atlas that appeared in the Political Herald, Dr. Gilbert Stuart wrote,

The new method in which accounts are stated in this work, has attracted very general notice. The propriety and expediency of all men, who have any interest in the nation, being acquainted with the general outlines, and the great facts relating to our commerce are unquestionable; and this is the most commodious, as well as accurate mode of effecting this object, that has hitherto been thought of. . . . To each of his charts the author has added observations [that] . . . in general are just and shrewd; and sometimes profound. . . . Very considerable applause is certainly due to this invention; as a new, distinct, and easy mode of conveying information to statesmen and merchants.

Such wholehearted approval rarely greets any scientific development. Playfair’s adaptation of graphic methods to matters of general interest provided an enormous boost to the popularity of statistical graphics. His energy and artistic sense showed themselves in the forty color charts in his initial Atlas. The size of the undertaking required to produce such a book indicates Playfair’s deep understanding of the power of the graphical method. His energy and skill as a draftsman, coupled with that understanding, led him to communicate his enthusiasm both widely and effectively. However, to be able to focus on graphics when the prevailing view of science looked upon such an approach as generally illegitimate requires a willingness to go against the tide—indeed, perhaps even taking joy in being an iconoclast. The events described in the next two chapters illuminate this aspect of Playfair’s personality.

In Kagemusha, a film by the great Japanese director Akira Kurosawa, a legendary warlord is mortally wounded. The warlord’s staff finds a petty thief, who bears a remarkable physical resemblance to the fallen leader, to substitute for him. With the substitute in place, the political strategy evolved by the dead warlord succeeds in his absence. In this examination of the question “the man or the moment?” Kurosawa clearly favors the latter. The Playfair enigma represents another instance of this great theme, although unlike Kurosawa’s fictional situation, the more limited information available to us does not allow unambiguous conclusions.
Sample R Code and Sample Datasets

All R code and sample datasets for this course will be available on Github at the following repository:

https://github.com/cmchurch/DHSI-ethical-dataviz

Git is a software package used for version control: the tracking and reporting of changes to files (text files, software code, etc). Github hosts git repositories, or collections of code, and makes it easy to share these collections with others. By using Git and Github, you will have access to the latest, up-to-date versions of the sample R code.

You can read more about Git and Github at the following link:
http://dh.obdurodon.org/github.xhtml
RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

How to Install R Studio

In order to run R and R-studio on your system, you need to follow the following three steps in the same order.

1. Install R
2. Install R-Studio
3. Install R-Packages (If needed)

1. Install R

Follow the steps below with respect to the operating system you are using.

To download R, please visit https://cran.r-project.org/

For Windows:
   1. Download the binary setup file for R from the following link. ( R for Windows )
   2. Open the downloaded .exe file and Install R

For Mac:
   1. Download the appropriate version of .pkg file form the following link. ( R for Mac )
   2. Open the downloaded .pkg file and Install R

For Linux:
   1. For complete R System installation in Linux, follow the instructions on the following link ( Link )
   2. For Ubuntu with Apt-get installed, execute `sudo apt-get install r-base` in terminal.

2. Install R Studio

On the following link Download R Studio choose the appropriate installer file for your operating system, download it and then run it to install R-studio.
3. Install the packages (Optional)

If your need to use R requires a particular package/library to be installed in R-studio. You can follow the instructions below to do so

1. Run R studio
2. Click on the Packages tab in the bottom-right section and then click on install. The following dialog box will appear
3. In the Install Packages dialog, write the package name you want to install under the Packages field and then click install. This will install the package you searched for or give you a list of matching packages based on your package text.

This completes the installation procedure for R Studio. If you want to continue with the Basic R tutorial click on the Basic Tutorial button in the left column.
Basic Data Analysis through R/R Studio

In this tutorial, I'll design a basic data analysis program in R using R Studio by utilizing the features of R Studio to create some visual representation of that data. Following steps will be performed to achieve our goal.

1. Downloading/importing data in R
2. Transforming Data / Running queries on data
3. Basic data analysis using statistical averages
4. Plotting data distribution

Let's go over the tutorial by performing one step at a time.

1. Importing Data in R Studio

For this tutorial we will use the sample census data set ACS. There are two ways to import this data in R. One way is to import the data programmatically by executing the following command in the console window of R Studio.

```r
acs <- read.csv(url("http://stat511.cwic.kco.nz/homeworks/acs_or.csv"))
```

Once this command is executed by pressing Enter, the dataset will be downloaded from the internet, read as a csv file and assigned to the variable name `acs`. 
The second way to import the data set into R Studio is to first download it onto your local computer and use the **import dataset** feature of R Studio. To perform this follow the steps below.

1. Click on the **import dataset** button in the top-right section under the environment tab. Select the file you want to import and then click open. The Import Dataset dialog will appear as shown below.
2. After setting up the preferences of separator, name and other parameters, click on the Import button. The dataset will be imported in R Studio and assigned to the variable name as set before.

Any dataset can be viewed by executing the following line:

```r
View(acs)
```

where `acs` is the variable dataset is assigned to.

## 2. Transforming Data

Once you are done with importing the data in R Studio, you can use various transformation features of R to manipulate the data. Let's learn few of the basic data access techniques

To access a particular column, Ex. `age_husband` in our case.

```r
acs$age_husband
```

To access data as a vector

```r
acs[1:3]
```
To run some queries on data, you can use the `subset` function of R. Let's say I want those rows from the dataset in which the `age_husband` is greater than `age_wife`. For this we'll run the following command in console:

```r
a <- subset(acs, age_husband > age_wife)
```

The first parameter to the `subset` function is the dataframe you want to apply that function to and the second parameter is the boolean condition that needs to be checked for each row to be included or not. So the above statement will return the set the rows in which the `age_husband` is greater than `age_wife` and assign those rows to `a`.

### Getting Statistical Averages from data

Following functions can be used to calculate the averages of the dataset:

1. For mean of any column, run: `mean(acs$age_husband)`
2. Median, run: `median(acs$age_husband)`
3. Quantile, run: `quantile(acs$age_wife)`
4. Variance, run: `var(acs$age_wife)`
5. Standard Deviation, run: `sd(acs$age_wife)`
You can also get the statistical summary of the dataset by just running on either a column or the complete dataset summary(acs)

4. Plotting Data

A very liked feature of R studio is its built in data visualizer for R. Any data set imported in R can visualized using the plot and several other functions of R. For Example

To create a scatter plot of a data set, you can run the following command in console

```r
plot(x = s$age_husband , y = s$age_wife, type = 'p')
```

Where s is the subset of the original dataset and type 'p' set the plot type as point. You can aslo choose line and other change type variable to 'L' etc.

For data distribution plots, there are several features tools and packages available in R that you can use to draw any kind of distribution. For example

To draw a Histogram of a dataset, you can run the command

```r
hist(acs$number_children)
```
Similarly for Bar Plots, run the following set of commands:

```r
counts <- table(acs$bedrooms)
barplot(counts, main="Bedrooms Distribution", xlab="Number of Bedrooms")
```
I hope this will give you a basic idea on how to do simple statistics in R.

**Note**

For any documentation or usage of the function in R Studio, just type the name of the function and then press `ctrl+space` to get the auto completion window.
You can use `>` before any function name to view the official documentation
[add rstudio data visualization cheat sheet (ggplot) here]

[add rstudio dates and times cheat sheet (lubridate) here]
PARSE DATE-TIMES (Convert strings or numbers to date-times)
1. Identify the order of the year (y) and second (s) elements in your data.
2. Use the function below whose name replicates the order. Each accepts a wide variety of input formats.

```
d <-
```

GET AND SET COMPONENTS
Use an accessor function to get a component.
Assign into an accessor function to change a component in place.

```
date(x) Day component. date(dt)
year(x) Year. year(dt)
isoyear(x) The ISO 8601 year.
epiyear(x) Epidemiological year.
month(x), label, abbr) Month.
month(dt)
day(x) Day of month. day(dt)
day(x) Day of year.
hour(x) Hour.
hours = 2)
minute(x) Minutes.
second(x) Seconds.
isoweek(x) The ISO 8601 week.
epiweek(x) Epidemiological week.
quarter(x) Quarter.
qday(x) Day of quarter.
semester(x) Semester.
semester(dt)
week(x) Week of the year. week(dt)
iso_week(x) ISO 8601 week.
epiweek(x) Epidemiological week.
```

Round Date-times
floor_date(x, unit = "second") Round down to nearest unit.
floor_date(dt, unit = "month")
round_date(x, unit = "second") Round to nearest unit.
round_date(dt, unit = "month")
rollback(x, roll_to_first = FALSE, preserve_hms = TRUE) Roll back to last day of previous month.rollback(dt)

Stamp Date-times
```
stamp() Derive a template from an example string and return a new function that will apply the template to date-times. Also
stamp_date() and stamp_time().
1. Derive a template, create a function
df <- stamp("Created Monday, Apr 05, 2010 00:00")
2. Apply the template to dates
```

Time Zones
R recognizes ~600 time zones. Each encodes the time zone, Daylight Savings Time, and historical calendar variations for an area. Assigns one time zone per vector.

Use the UTC time zone to avoid Daylight Savings.
OlsonNames() Returns a list of valid time zone names. OlsonNames()

```
f with_tz(time, tzname = "UTC") Get the same date-time in a new time zone (a new clock time).
with_tz(dt, "US/Pacific")
```

DHSI | ETHICALVIZ | 175

RStudio® is a trademark of RStudio, Inc. • CC BY SA RStudio • info@rstudio.com • 844-448-1212 • rstudio.com • Learn more at lubridate.tidyverse.org • lubridate 1.6.0 • Updated: 2017-12
Math with Date-times

Math with date-times relies on the timeline, which behaves inconsistently. Consider how the timeline behaves during:

A normal day
nor <- ymd_hms("2018-01-01 01:30:00" ,tz="US/Eastern")

The start of daylight savings (spring forward)
gap <- ymd_hms("2018-03-11 01:30:00" ,tz="US/Eastern")

The end of daylight savings (fall back)
lap <- ymd_hms("2018-11-04 00:30:00" ,tz="US/Eastern")

Leap years and leap seconds
leap <- ymd("2019-03-01")

Periods track changes in clock times, which ignore time zone irregularities.

\[
\begin{align*}
\text{nor} + \text{minutes}(90) \\
\text{gap} + \text{minutes}(90) \\
\text{lap} + \text{minutes}(90) \\
\text{leap} + \text{years}(2)
\end{align*}
\]

Durations track the passage of physical time, which deviates from clock time when irregularities occur.

\[
\begin{align*}
\text{nor} + \text{dminutes}(90) \\
\text{gap} + \text{dminutes}(90) \\
\text{lap} + \text{dminutes}(90) \\
\text{leap} + \text{dyears}(2)
\end{align*}
\]

Intervals represent specific intervals of the timeline, bounded by start and end dates.

\[
\begin{align*}
\text{interval(nor, nor + minutes(90))} \\
\text{interval(gap, gap + minutes(90))} \\
\text{interval(lap, lap + minutes(90))} \\
\text{interval(leap, leap + years(1))}
\end{align*}
\]

Durations are stored as seconds, the only time unit with a consistent length. Add or subtract durations to model physical processes, like battery life. Durations are stored as seconds, the only time unit with a consistent length.

\[
\text{"1209600s (~2 weeks)"}
\]

Make a period with the name of a time unit pluralized, e.g.

\[
\begin{align*}
\text{years}(x = 1) & \quad x \text{ years}, \\
\text{months}(x) & \quad x \text{ months}, \\
\text{weeks}(x) & \quad x \text{ weeks}, \\
\text{minutes}(x = 1) & \quad x \text{ minutes}, \\
\text{seconds}(x = 1) & \quad x \text{ seconds}. \\
\end{align*}
\]

Add a date to a period to specify the start of the period.

\[
\begin{align*}
\text{dd} & \quad \text{days(14)} \\
\text{dd} & \quad \text{"1209600s (~2 weeks)"}
\end{align*}
\]

Durations track the passage of physical time, which deviates from clock time when irregularities occur.

\[
\begin{align*}
\text{nor} + \text{dminutes}(90) \\
\text{gap} + \text{dminutes}(90) \\
\text{lap} + \text{dminutes}(90) \\
\text{leap} + \text{dyears}(2)
\end{align*}
\]

Intervals represent specific intervals of the timeline, bounded by start and end dates.

\[
\begin{align*}
\text{interval(nor, nor + minutes(90))} \\
\text{interval(gap, gap + minutes(90))} \\
\text{interval(lap, lap + minutes(90))} \\
\text{interval(leap, leap + years(1))}
\end{align*}
\]

Leap years and leap seconds
leap <- ymd("2019-03-01")

Add or subtract durations to model physical processes, like battery life. Durations are stored as seconds, the only time unit with a consistent length.

\[
\text{"1209600s (~2 weeks)"}
\]

Make a period with the name of a time unit pluralized, e.g.

\[
\begin{align*}
\text{years}(x = 1) & \quad x \text{ years}, \\
\text{months}(x) & \quad x \text{ months}, \\
\text{weeks}(x) & \quad x \text{ weeks}, \\
\text{minutes}(x = 1) & \quad x \text{ minutes}, \\
\text{seconds}(x = 1) & \quad x \text{ seconds}. \\
\end{align*}
\]

Add a date to a period to specify the start of the period.

\[
\begin{align*}
\text{dd} & \quad \text{days(14)} \\
\text{dd} & \quad \text{"1209600s (~2 weeks)"}
\end{align*}
\]

Add or subtract durations to model physical processes, like battery life. Durations are stored as seconds, the only time unit with a consistent length.

\[
\begin{align*}
\text{nor} + \text{dminutes}(90) \\
\text{gap} + \text{dminutes}(90) \\
\text{lap} + \text{dminutes}(90) \\
\text{leap} + \text{dyears}(2)
\end{align*}
\]

Intervals represent specific intervals of the timeline, bounded by start and end dates.

\[
\begin{align*}
\text{interval(nor, nor + minutes(90))} \\
\text{interval(gap, gap + minutes(90))} \\
\text{interval(lap, lap + minutes(90))} \\
\text{interval(leap, leap + years(1))}
\end{align*}
\]

Leap years and leap seconds
leap <- ymd("2019-03-01")

Add or subtract durations to model physical processes, like battery life. Durations are stored as seconds, the only time unit with a consistent length.

\[
\begin{align*}
\text{nor} + \text{dminutes}(90) \\
\text{gap} + \text{dminutes}(90) \\
\text{lap} + \text{dminutes}(90) \\
\text{leap} + \text{dyears}(2)
\end{align*}
\]

Intervals represent specific intervals of the timeline, bounded by start and end dates.

\[
\begin{align*}
\text{interval(nor, nor + minutes(90))} \\
\text{interval(gap, gap + minutes(90))} \\
\text{interval(lap, lap + minutes(90))} \\
\text{interval(leap, leap + years(1))}
\end{align*}
\]
Data Visualization with ggplot2 :: CHEAT SHEET

**Basics**

**ggplot2** is based on the grammar of graphics, the idea that you can build every graph from the same components: *data*, *coordinates*, and *geoms*—visual marks that represent data points.

**ggplot()** creates a complete plot with given data, geom, and mappings. Supplies many useful defaults.

**last_plot()** returns the last plot.

---

**Geoms**

Use a `geom` function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

**Graphical Primitives**

- **Two Variables**
  - Continuous x, continuous y
  - `geom_contour(aes(x=cty, y=hwy))`
- **Discrete x, continuous y**
  - `geom_bar(aes(x=x, y=y))`
  - `geom_col(aes(x=x, y=y))`
  - `geom_point(aes(x=x, y=y), size=size)`
  - `geom_tile(aes(x=x, y=y))`
- **Continuous x, Discrete y**
  - `geom_bar(aes(x=x, y=y))`
  - `geom_histogram(aes(x=x))`
  - `geom_density2d(aes(x=x, y=y))`
  - `geom_hex(aes(x=x, y=y))`
- **Three Variables**
  - `geom_contourf(aes(x=cty, y=hwy))`
  - `geom_tile(aes(x=x, y=y))`

---

**Line Segments**

- **One Variable**
  - Continuous
  - `geom_line(aes(x=x, y=y))`
  - `geom_point(aes(x=x, y=y), size=size)`
- **Discrete**
  - `geom_step(aes(x=x, y=y))`
  - `geom_blank()`

---

**Shapes**

- `geom_blank()`
- `geom_point()`
- `geom_crossbar()`
- `geom_errorbarh()`
- `geom_boxplot()`
- `geom_linerange()`
- `geom_pointrange()`
- `geom_raster()`
- `geom_tile()`
- `geom_raster()`
- `geom_errorbarh()`

---

**Coordinate Systems**

- `ggplot()`
- `ggplot(data = map)`
- `ggplot2()`

---

**Aesthetics**

- `aes()`
- `aes(x=x, y=y, fill=z)`
- `aes(group=group, size=size)`
- `aes(color=color, alpha=alpha)`
- `aes(shape=shape, linetype=linetype)`

---

**Themes**

- `ggplot()`
- `ggplot() + theme()`
- `ggplot() + theme(panel.background=element_rect(fill=color), panel.grid=element_blank(), axis.title=element_blank(), axis.text=element_blank(), axis.line=element_blank(), legend.position="bottom")`

---

**Data Mapping**

- `aes()`
- `aes(x=x, y=y, fill=z)`
- `aes(group=group, size=size)`
- `aes(color=color, alpha=alpha)`
- `aes(shape=shape, linetype=linetype)`

---

**Examples**

- **Jittered Points**
  - `ggplot(mpg, aes(displ, displ/100)) + geom_jitter(position=position_jitter())`
- **Bar Chart**
  - `ggplot(mpg, aes(displ, displ/100)) + geom_bar(stat="identity")`
- **Histogram**
  - `ggplot(mpg, aes(displ)) + geom_histogram()`
- **Density Plot**
  - `ggplot(mpg, aes(displ)) + geom_density()`
- **Box Plot**
  - `ggplot(mpg, aes(displ)) + geom_boxplot()`
- **2D Density Plot**
  - `ggplot(mpg, aes(displ)) + geom_density2d()`
- **Hexagon Plot**
  - `ggplot(mpg, aes(displ)) + geom_hex()`
- **Violin Plot**
  - `ggplot(mpg, aes(displ)) + geom_violin()`
- **Contour Plot**
  - `ggplot(mpg, aes(displ, displ/100)) + geom_contour()`
- **Contour Fill**
  - `ggplot(mpg, aes(displ, displ/100)) + geom_contour_filled()`
- **Ribbon Plot**
  - `ggplot(mpg, aes(displ, displ/100)) + geom_ribbon()`
- ** elections Data**
  - `ggplot(elections, aes(x=year, y=percent, fill=party)) + geom_bar(stat="identity")`
Stats

An alternative way to build a layer

A stat builds new variables to plot (e.g., count, prop).

\[
\text{data} \rightarrow \text{geom} \rightarrow \text{coordinate system} \rightarrow \text{stat} \rightarrow \text{element} \rightarrow \text{layer} \rightarrow \text{plot}
\]

Visualize a stat by changing the default stat of a geom function, `geom_bar(stat="count")` or by using a stat function, `stat(count=geom_bar)`, which calls a default stat to make a layer (equivalent to a geom function).

Use `...` syntax to map stat variables to aesthetics.

```
+ stat dysfunction(aes(fill = ..y..), geom = "bar")
```

Scales

Scales map data values to visual values of an aesthetic. To change a mapping, add a new scale.

```
+ scale_color_manual(values = c("red", "blue", "green"))
```

Position Adjustments

Position adjustments determine how to arrange geoms that would otherwise occupy the same space.

```
+ geom_point(position = "jitter")
```

Color and Fill Scales

```
+ scale_fill_manual(values = c("red", "blue", "green"))
```

Legend

```
+ scale_fill_manual(values = c("red", "blue", "green"))
```

Themes

```
+ theme_minimal()
```

 Coordinate Systems

```
+ coord_polar(theta = "x", direction = 1)
```

Faceting

```
+ facet_wrap(~ fl)
```

Legends

```
+ scale_fill_discrete(name = "Legend")
```

Zooming

```
+ coord_cartesian(xlim = c(0, 5), ylim = c(0, 10))
```
Day 2: Racist Historical Data

Morning 2: Visualization and Graphic Conventions
9:00am - 12pm (3 hours)

- Lecture: Historicizing visualization — narrative and graphic visualization conventions through history
- Examples and discussion: Exploring Sample Visualizations — Racist historical visualizations
- Reading: Forgotten Dead, Introduction

Afternoon 2: Visualizing Race
1:30pm to 4:00pm (2.5 hours)

- Reading and discussion: Hepworth, K. J., Church, C., "Racism in the Machine: Visualization Ethics in Digital Humanities Projects". Digital Humanities Quarterly. [under review]
- Visual exploration activity & discussion: Exploring Racial Terror Lynchings map and Map of White Supremacist Mob Violence
- Documentation: Ethical Visualization Workflow
- Visualization activity: Creating visualization plots (using provided, pre-made R codes) based on the mapping projects using the following data: https://www.kaggle.com/rtatman/historical-american-lynching
DAY 2 SLIDES

[add slides here]
Lying with Statistics

Drawing by Ben Orlin, www.mathwithbaddrawings.com
Distribution Matters!

- **Mode** (most frequent number)
- **Median** (middle number in list)
- **Mean** (average of all numbers)

Symmetrical distribution:
- Mode = Median
- Mean = Mode

Asymmetrical distribution:
- Mode ≠ Median
- Mean ≠ Mode
What would my starting salary be?

I’ll put it this way: our average starting salary is $80,000!
Average: $80,000.
So, why should I invest with you?

Well, not to brag, but my fund has a median gain of 8% per year!
Mode

How are you doing on your tests?

My modal category is 70-80%!
<table>
<thead>
<tr>
<th>Score Category</th>
<th>Number of Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>90s</td>
<td>0</td>
</tr>
<tr>
<td>80s</td>
<td>0</td>
</tr>
<tr>
<td>70s</td>
<td>2</td>
</tr>
<tr>
<td>60s</td>
<td>1</td>
</tr>
<tr>
<td>50s</td>
<td>1</td>
</tr>
<tr>
<td>40s</td>
<td>1</td>
</tr>
<tr>
<td>30s</td>
<td>1</td>
</tr>
<tr>
<td>20s</td>
<td>1</td>
</tr>
</tbody>
</table>
Our students come from a wide range of socioeconomic backgrounds...
Correlation Coefficient

Try our energy drink—it's highly correlated with performance!
Correlation Coefficient
Shows Strength & Direction of Correlation

-1.0  -0.5  0.0  +0.5  +1.0

Strong ← Weak ← Weak → Strong

Negative Correlation  Zero  Positive Correlation
professional athletes were paid to guzzle the stuff

athletic performance

amount of drink consumed
Variance

Larger Variation

Smaller Variation
These results are a disaster!

Sure, they look bad, but there's a lot of variance! Don't rush to judgment.
terrible results

outlier

great results
Figure 1: History of data visualization timeline
The Earl of Arundel points to a colonial venture in Madagascar
William Playfair
1759–1823
Exports and Imports to and from Denmark & Norway from 1700 to 1780.

The bottom line is divided into years, the right hand line into £10,000 each.

Published at the Act Offices, 10th May 1786. by W. Kippenley.

Made and sold by J. B. Grand, London.
CHART
Shewing the Value of the Quarter of Wheat
In Shillings &c in Days Wages
Of a Good Mechanic
From 1563 to 1821

1821
John Snow

1813–1858
Florence Nightingale
1820–1910
The areas of the blue, red, & black wedges are each measured from the centre as the common vertex. The blue wedges represent the deaths from Preventable or Mitigable Zymotic diseases, the red wedges represent the deaths from wounds, & the black wedges represent the deaths from all other causes. The blue line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during that month. In Oct. 1854 & Apr. 1855 the black area coincides with the red, in Jan.–Feb. 1855 the blue coincides with the black. The entire areas may be compared by following the blue, the red & the black lines enclosing them.

http://www.york.ac.uk/depts/maths/histstat/small.htm
US Census Office
Bi-polar charts, 1874
49. COMPOSITION OF THE FOREIGN-BORN POPULATION: 1890.
Otto & Marie Neurath
1882–1945
1898–1986
THE WORLD'S MOTOR CAR INDUSTRY IN 1929

North and South America  
Europe

Each figure represents 100,000 workers employed in the motor car industry. Each car represents a production of 100,000 cars.
Home and Factory Weaving in England

1820

1830

1845

1860

1880

Each blue symbol represents 50 million pounds total production
Each black man symbol represents 10,000 home weavers
Each red man symbol represents 10,000 factory weavers
DAY 2 READINGS

[add readings here]
Racism in the Machine: Visualization Ethics in Digital Humanities Projects

Katherine Hepworth <khepworth_at_unr_dot_edu>, University of Nevada, Reno
Christopher Church <christopherchurch_at_unr_dot_edu>, University of Nevada, Reno

Abstract

Data visualizations are inherently rhetorical, and therefore bias-laden visual artifacts that contain both explicit and implicit arguments. The implicit arguments depicted in data visualizations are the net result of many seemingly minor decisions about data and design from inception of a research project through to final publication of the visualization. Data workflow, selected visualization formats, and individual design decisions made within those formats all frame and direct the possible range of interpretation, and the potential for harm of any data visualization. Considering this, it is imperative that we take an ethical approach to the creation and use of data visualizations. Therefore, we have suggested an ethical data visualization workflow with the dual aim of minimizing harm to the subjects of our study and the audiences viewing our visualization, while also maximizing the explanatory capacity and effectiveness of the visualization itself. To explain this ethical data visualization workflow, we examine two recent digital mapping projects, Racial Terror Lynchings and Map of White Supremacy Mob Violence.

Introduction

In March 2016, Microsoft's Technology and Research division released an artificial intelligence chatbot known as TayTweets under the handle @TayandYou on Twitter. Microsoft had programmed the chatbot to interact with the online community and subsequently learn from the tweets of users mentioning it, in order to improve its natural language communication abilities. Though unlikely to pass the Turing Test any time soon, TayTweets was nevertheless able to generate its own internet memes, crack its own jokes, and participate in the regular back-and-forth of online conversation.

Microsoft presented TayTweets as a value-neutral project that would showcase the technological achievements of its research division, and early on the AI was seemingly innocuous, posting anodyne responses to tweets welcoming it to Twitter and asking it mundane questions. With its first tweet of “hellooooooo w¿¿¿¿rld!!!” — a riff on the traditional output of a computer programmer’s first program — TayTweets had virtually stepped onto the world stage and greeted the Twitterverse in a seemingly innocent, even youthfully naive, way.

Within 16 hours, however, the experiment had to be shut down, as TayTweets had algorithmically learned to be racist. Posting inflammatory tweets championing white supremacy and denigrating racial minorities and marginalized groups, the artificial intelligence program had become an online menace and in fact a cyberbully. TayTweets’s rapid descent into racist demagoguery serves as a harrowing reminder that our digital productions are not free from the cultural assumptions and prejudices that shape everyday human experience. Microsoft’s designers had a blind spot about the depth and breadth of American racism, which allowed TayTweets to replicate deep-seated preconceptions and prejudices without critically examining them.

For our purposes, what is interesting about TayTweets is not its experimentation with artificial intelligence, but the project’s apparent assumption that an algorithm — the set of rules by which any programmatic approach operates, whether digital or analog — is value neutral and divorceable from human prejudice and malice. On the contrary, humans are at the center of algorithms, not only as their creators but, in the case of data-driven algorithms, as the producers of the content they shape and present. Though an extreme example, TayTweets clearly demonstrates how wider cultural assumptions, prevalent political ideologies, and public discourses shape the output of our algorithmic productions, potentially replicating what we
Ethical visualization is essentially a human-centric approach to algorithmic production, considering the underlying biases, ideologies, and beliefs that animate algorithms as they structure and reproduce past inequities and harmful realities. All data visualizations — whether static or interactive, printed or digital, computer-generated or hand-drawn — are algorithmic by nature, in the sense that they solve complex problems of representation and require a set of tree-based actions and decisions for their success. Ethical visualization practices sit at the intersection of humanistic inquiry, ethics, and communication design. We define ethical visualization as the presentation of visualized information in ways that acknowledge and mitigate the potential for harm engendered within the visualization form and content. While good design practice forms the backbone of ethical visualizations, ethical visualization practice goes one step further to consider the ultimate societal impact of such design choices: do such choices cause harm or mislead, either intentionally or unintentionally? Do they result in a net societal benefit, or do they prove deleterious to marginalized individuals? These questions must be brought to the forefront when considering good design, because a visualization can follow good design practices and consequently be easy to understand, but still produce a negative societal impact for its subject matter all the same.

Considering the object lesson of TayTweets, this article proposes that digital humanists must adhere to a form of visualization ethics that considers how both choices about working with data and the rhetorical qualities of communication elements — color, composition, line, symbols, type, and interactivity — shape users’ understandings of represented people and places. Given the “racism in the machine” — the ways in which our digital tools can inadvertently recreate the latent racism, underlying prejudices, and cultural blind spots in our society — the goal of this visualization ethics should be “increasing understanding [for users] while minimizing harm” to represented people and places [Cairo 2014].

To propose a methodology for visualization ethics, this article examines two projects that visualize the horrific history of racial lynching in the United States, Lynching in America by the Equal Justice Initiative and Monroe Work Today by Autt Studio, in order to show that no visualization is ideologically neutral, but is instead part of an argument that must be critically examined. In contrast to TayTweets’s outright bigotry, Lynching in America and Monroe Work Today both present visualizations that demonstrate to varying degrees ethical visualization practices, though one succeeds in producing an ethical visualization to a greater degree than the other. This paper showcases ethical visualization in practice to varying degrees in these two digital mapping projects, demonstrating how choices about representation, interaction, and annotation in their data visualizations either do harm in the sense described above, or challenge dominant narratives. In comparing these two projects, the article outlines a workflow that can ensure that data visualizations adhere to best practices in visualization ethics and thereby present opportunities for more inclusive and critical interaction with represented data.

**Lynching in America (EJI) vs Monroe Work Today**

**Example 1: Racial Terror Lynchings Map**

Lynching in America (https://lynchinginamerica.eji.org/), a promotional website made by Google for the Equal Justice Initiative, a mass incarceration not-for-profit organization, contains an interactive map titled Racial Terror Lynchings (https://lynchinginamerica.eji.org/explore). Drawing on a dataset compiled by the EJI for their eponymous report of “4075 racial terror lynchings of African-Americans,” this interactive choropleth map of the United States purports to depict “Reported lynchings by county”, occurring between 1877 and 1950 [“Lynching in America”]. Overall, the map has a minimalist aesthetic, reminiscent of Google’s Material Design visual style (https://material.io/) that depicts topographical elevation as well as country, state, and county borders, but not cities, towns, roads, rivers, lakes, or landmarks (see Figure 1). The data visualized is also minimalist in presentation, and is tied directly to the state and county borders depicted, as is typical of choropleth maps.

The map is focused on the southeastern United States. On first page load in screen widths below 1500px, the map centers on the United States below the Mason-Dixon line, while in larger screens the entire contiguous United States is shown. However, in both cases, the American south grabs attention with many southern counties highlighted. Users can click on any county or state, and are taken to a zoomed-in view of the state, with the total reported lynchings in that state displayed in large letters. The user can then hover over individual counties to find out how many lynchings were reported in that county (see Figure 2). This zoomed-in view focuses on the number of lynchings per state and county — represented by polygons shaped according to county boundaries — creating a visual argument that encourages measurement of states...
and counties as more or less reprehensible in terms of number of lynchings. It is focused on lynchings in the context of political boundaries, and thereby presents a strong geopolitical argument about recorded lynchings that appears definitive and damning, particularly of the states with the most counties marked in bright red: Alabama, Florida, Louisiana, and Mississippi. There is a straightforwardness about this visual argument: racial terror was and is morally wrong, and its contours are plain to any competent observer.

In terms of color, the map features a dark, almost monochrome color palette, with a dark grey United States segmented by black state lines, placed in a dark blue sea. This mostly dark color scheme is dramatically interrupted with many counties highlighted in various shades of red, with bright red indicating 20 or more lynchings recorded in the county (see Figure 1). The color scheme of red on an otherwise dark, monochrome palette compounds the visual argument described above, as it references the brutality and violence inflicted upon African Americans in those locations, recalling the blood stains on United States history as it relates to racial violence and white supremacy.

![Figure 1. Figure 1](image)
Figure 2. Racial Terror Lynchings map in Lynching in America website. Top image (Figure 1) shows initial view of map on page load. Bottom image (Figure 2) shows the zoomed-in view on Georgia, with mouse hover over Richmond County. With 590 reported lynchings in Georgia, Richmond County’s 2 reported lynchings appear significantly less reprehensible than the record of other counties in Georgia.

Example 2: Map of White Supremacy Mob Violence

The second example, Map of White Supremacy Mob Violence (http://www.monroeworktoday.org/explore/), is a far more complex visualization than the minimalist Racial Terror Lynchings. It is an interactive map within Monroe Work Today (http://www.monroeworktoday.org), a website dedicated to publicizing the research of sociologist Monroe Work, who systematically documented lynchings in the United States. Created by education-focused digital agency Aut Studio, this interactive map of the United States depicts lynching records in the context of historical racial violence and public discourses of white superiority, and is consequently subtitled “The lynchings and riots to enforce racial superiority in the US”. Upon first page load, Map of White Supremacy Mob Violence stands in stark contrast to Racial Terror Lynchings, presenting the entire contiguous United States center of screen, irrespective of screen size, and representing each recorded case of lynching as a single grey dot on a white national landmass background (see Figure 3). State borders are not visible in this view, and a single, bright, attention grabbing color is only found on instructions for users. This initial map view — including map framing, choice of colors, and the carefully worded title and subtitle — provides a strong and markedly different message than the Racial Terror Lynchings map. Collectively, these elements present a strong visual argument that discourses of white supremacy are a nationwide reality in the United States, one that has historically been enforced through lynchings and riots by mobs. The geopolitical framing of lynching as a nation-wide reality is balanced with the acknowledgement of a crucial aspect of the historically pervasive intellectual climate, one of white superiority, and an important aspect of the social behavioral climate, mob violence. At the center bottom of the page is a key annotation: a label that reads “Should I trust this? Find out.” If users click this message, they are taken to a plethora of information about the veracity of the data used, and a discussion of the importance of thinking critically about the visualized data.

Although it is not entirely clear from this initial page view what the grey dots represent, the user is presented with instructions to zoom in and click on individual points. After the user follows these instructions, the map gains more color, and individual dots become more visible, along with state and county boundaries (see Figure 4). In this view, state and county lines are visible, as are the boundaries of Native American reservations and areas that historically were Spanish colonies. However, these geographic lines are indicated with subtlety, using pale colors. In contrast, the dots representing each individual lynching are highlighted using bright colors and an onclick interaction effect. Marking geopolitical boundaries but deemphasizing them, combined with emphasizing each individual lynching with bright colors, makes the
visual argument that lynchings occurred in the context of geopolitical boundaries, but that the individual deaths are of greater significance than the boundaries themselves.

Along the bottom of the screen, supplemental information and interactive features appear, including a timeline, a label stating the time span represented in the current view, and a color-coded legend of dot colors. Six different dot colors are used: five to represent races of lynched people, and one to represent “other”, where the race recorded in the historical records does not fall into one of the main five. Notably, the “other” category includes lynchings of white abolitionists, thereby demonstrating the complexity of the history of lynching in the United States, and that white people were also, however rarely, victims of lynching in the name of white supremacy.

Both the timeline and the legend serve two purposes: one functional, the other persuasive. Users can select a time period using the timeline, which then alters data presented on the map, so they can see how many lynchings occurred in an area over a specific timeframe. Secondly, the timeline provides a persuasive visual cue that racial superiority-motivated lynchings occurred continually over a long timespan, with some time periods seeing evidence of significantly more racial superiority-motivated lynchings. The legend also works on these two levels. It firstly allows the user to identify the race of a lynched person based on the color of the dot used to represent them, and secondly provides a strong visual counterargument to the widespread public assumption that lynchings were perpetrated exclusively on African Americans. These two elements, the timeline and the legend, confront the user with the temporal and racial extent of white superiority-motivated lynchings, both qualities that are absent from the Racial Terror Lynching map.

Most strikingly, the zoomed-in view of Map of White Supremacy Mob Violence contains a list, in the bottom left hand corner of the screen, of the name (where available) of every lynched person represented by a dot in the current map view, and the year they were lynched. Clicking on any individual dot brings up a callout box containing extra details of the lynching available in the historical record, including the county in which it occurred, the details of mob violence in which the lynching occurred, the accusation made before the lynching, and links to every available historical source that verifies the record (see Figure 5). Naming individuals who were lynched, and providing circumstances surrounding their death, focuses attention on the humanity of victims of lynching and on the social circumstances in which lynching was a viable possibility. In the present day context of many white supremacists denying their racism, it is worth noting that these historical records rarely mention race as a motivator for lynching. For example, William B. Willis of Richmond County, Georgia, was accused of murder before being lynched. The Map of White Supremacy Mob Violence does an exemplary job of demonstrating that racism was indeed the motivating factor in lynchings, and also that racism was largely hidden in the official historical record by documenting other, non-racial reasons. Providing links to multiple historical sources within the interactive map increases trust in the veracity of data, while at the same time giving users the opportunity to investigate the historical evidence themselves. Map of White Supremacy Mob Violence uses multiple compelling strategies to both humanize the data it represents, and to contextualize it in the societal racism and discourses of white supremacy in the United States.
Figure 5. Map of White Supremacy Mob Violence in Monroe Work Today website. Top image (Figure 3) shows the initial view of the map on page load. Center image (Figure 4) shows a zoomed-in view on Georgia. Richmond County is highlighted in green by the authors, for comparison with the Racial Terror Lynching map in Figures 1 and 2. Bottom image (Figure 5) shows a view of clicking over an individual lynching record in Richmond County. Naming individuals who were lynched and providing circumstances of their death focuses attention on the humanity of victims of lynching, the cover up of racist motivations for lynching documented in the historical record, and deemphasizes geopolitical comparisons. Additionally, the context of a large number of lynchings close across the border in South Carolina suggests that the geographic area including Richmond county was no less immune to racial superiority-motivated violence than the surrounding areas, as the presentation of information in the Racial Terror Lynchings map visually argues.

Comparison of both maps’ depictions of the West

The contrast between these two maps is even more striking when looking at the west coast of the United States. For example, Figure 6 shows a marked difference between the recorded lynchings in California in the Racial Terror Lynchings map and the Map of White Supremacy Mob Violence respectively. The difference in the representation can be accounted for by the fact that the former only depicts lynchings of African Americans, whereas the latter depicts lynchings of African Americans, Native Americans, Latinos, Italians, and other races. The view of the American West depicted in Map of White Supremacy Mob Violence in Figure 6b provides a compelling visual narrative that lynchings were common across California in the name of enforcing white superiority. The view of Racial Terror Lynchings Map depicted in Figure 6a, by its lack of clarity about exactly which data is being represented (i.e. historical records of lynchings of African Americans only), its use of a bold and expansive title that suggests comprehensive coverage of lynchings of all races (i.e. Racial Terror Lynchings), and its tonal emphasis on the south (California appearing grey, while visual attention is drawn to the large amounts of red in the bottom right hand corner of the map), makes a visual argument that California had few instances of racial terror or lynchings. This is particularly problematic because the Equal Justice Initiative’s stated goal is to challenge black incarceration, and a key part of their organizational message is that this goal is urgent and directly related to the history of lynchings motivated by white supremacy. Black incarceration rates in California are among the highest in the nation today: according to the U.S. Bureau of Justice, California imprisons blacks 8.8 times more frequently than whites, well above the national average of 5.5:1 ["The Sentencing Project"]. Consequently EJI’s visual argument in the Racial Terror Lynchings Maps inadvertently breaks down in advocacy regarding California prisons.
Figure 6. Contrasting views of California in Racial Terror Lynching Map (figure 6a, left) and Map of White Supremacy Mob Violence (figure 6b, right). The map framing on the left suggests lynching occurred rarely in California, whereas the map framing on the right indicates that lynching was widespread in California.

Discussion

The emphases of these two maps are necessarily different because of the different purposes of the sites in which they are situated. Lynching in America is a promotional and advocacy tool for the Equal Justice Initiative, primarily created to visualize data within (and thereby promote) the report “Lynching in America”, which records lynchings of African Americans and frames lynching as a societal tool — enabled through mob violence and discourses of white superiority — to subjugate African Americans between slavery and mass incarceration. It is in the Equal Justice Initiative’s interests to visualize historical lynching data in a way that draws attention to geopolitical divides, so that clear links can be made between historical lynching events and present-day constituencies of sitting politicians as well as county and state local governments. The map provides compelling visual evidence for the organization’s present-day advocacy work regarding the inequitable mass incarceration of black Americans, the case of California notwithstanding. The problematic aspect of this is that the Equal Justice Initiative’s website, report, and Racial Terror Lynching map are unfortunately named to suggest that they cover all historical records of lynchings in the United States. The Lynching in America report includes instructions for educators who wish to use it as a teaching resource. In this context, the geopolitical emphasis, use of color, use of summary data, and lack of links to sources give a concerning impression that African-American lynchings were the complete record of lynchings in the United States for the purposes of racial terror.

The majority of both scholarship and public attention regarding lynchings centers on the experience of African Americans in the Southern United States, and for good measure: for blacks in the Jim Crow South, lynchings represented a terrifying aspect of everyday life. In the grand scheme of racial violence, however, lynchings represented one small piece in a complex puzzle of individual, institutional, and structural racism. The conflation of lynching with the full extent of racialized violence in United States history obscures the historic depth and breadth of the oppression of people of color. A black individual was far more likely to suffer public humiliation, assault, rape, and murder than a public lynching. While lynchings do not represent the totality of racial violence in America, they come to the fore because they were highly symbolic affairs: gruesome spectacles of white supremacy, racial violence, and bodily mutilation meant to suppress and intimidate as much as they were meant to kill [Wood 2009, 1–4]. Due to their highly symbolic nature and the lasting implications of racist attitudes, policies, and actions for African Americans today, lynchings have become synonymous with racial hatred in the postbellum American south.

However, mob murder historically extended well beyond Dixie, representing a form of prejudicial frontier “justice” in the Midwest, West, and Southwest against minorities and immigrants of various backgrounds [Pfeifer 2006] [Pfeifer 2013]. Outside of the latest scholarship, such victims, whom Carrigan and Webb describe as the “forgotten dead,” are largely overshadowed or overlooked in the public sphere, missing an opportunity to explore the structural, cross-regional, and transethnic dimensions of American lynching. In fact, lynch mobs murdered hundreds of Mexicans between 1848 and 1928 in the American Southwest [Carrigan and Webb 2017].
Moving beyond EJI’s limited focus on African-American populations in the American South, Auut Studio acknowledges the historic violence committed against Native American populations, which is noticeably absent from the vast majority of lynching data sets. This general oversight plays into present-day blind spots regarding violence against native populations, who, despite suffering more state violence and community disruption than any other minority group in 2016 according to data collected by the Centers for Disease Control and Prevention, rarely garner the public spotlight [“The Counted”]. To counteract the public oversight of this “forgotten minority,” Auut Studio included the boundaries of Native American reservations as “sovereignt[ies] deserving of equal visual treatment on the map” [Ramey 2017]. Similarly, Auut Studio included lynchings of the Chinese along the frontier, namely in California, as well as Mexicans in the Southwest, keeping in line with recent historical scholarship.

In order to ethically represent historical subjects, data visualization techniques, particularly those geared toward public consumption, must remain abreast of the insights made in two areas of the scholarly literature: debates on the subjects they depict, and debates on design and representational considerations in the ethical visualization literature. Present-day academic debates on lynching challenge the widely accepted notion that lynching was exclusively a Southern phenomenon, excusing those regions of the United States outside of the South of their own racist heritage. The small but growing ethical visualization literature emphasizes the need for acknowledging and mitigating the potential for harm inherent in visualizing data, particularly when it comes to selection of design elements, visual style, and selections of data to annotate and visually emphasize [Cairo 2014] [Hepworth 2016] [Kostelnick 2016] [Skau, Harrison, and Kosara 2015].

Collectively, we tend to visualize old arguments, as visualization practices have not kept pace with dominant arguments in the digital humanities literature about the importance of critical practices. Visualization practice in the digital humanities runs the risk of following a functionalist methodological approach that assumes visualization to be an impartial medium. This illusory functionalism has led others to charge that data collection, processing, and visualization practices constitute mere “janitorial work” in the service of “real” humanities scholarship, ignoring the important decisions made during such processes that critically shape historical narratives. Construing digital humanities practice as a “support field” has led to further accusations that the digital humanities simply show to us what we already know, rather than challenging us to think critically about historical topics in new, interesting, and socially responsible ways [Allington et al. 2016].

Following Alan Liu’s charge that digital humanists have ignored cultural criticism, which in turn has blocked “the digital humanities from becoming a full partner of the humanities,” data visualization practitioners need to critically engage with the ways in which digital tools can “communicate humanity” rather than relegating it to the margins, or worse, obscuring the human stories essential to understanding structural racism today [Liu 2012].

Visualizing data that exclusively focuses on the African-American experience in the Southern United States provides an important argument about the nature of Jim Crow racism. However, purporting such data to be an inclusive representation constitutes a harm in the sense that it perpetuates common narratives of racial violence as a southern exception to an otherwise inclusive nation. After publishing their interactive map, the Equal Justice Initiative itself recognized this oversight, acknowledging the 300 lynchings of African Americans outside of the American south, though notably leaving aside other ethnicities like Native Americans, Mexicans, and the Chinese that fell victim to much of Western and Southwestern mob violence [“EJI Releases New Data” 2017]. This overlooks the depth of structural racism and its support of white supremacy, thereby denying the experience of millions of present-day Americans. Historical information has the capacity to legitimize or delegitimize present-day experience, and visualizations of historical data are a particularly compelling and resonant medium through which such information can either harm or help.

**Ethical Visualization Workflow**

We call for critical and practical analysis of the entire endeavor of data collection and visualization in the digital humanities. Humanities scholars in recent decades have critically examined the categories of scientific analysis inherited from the enlightenment that presuppose essential differences (based on sex, race, age etc) acknowledging that such presuppositions frame and ultimately determine scholarly insight [Knorr-Cetina 1981]. However, digital humanists and data scientists rely heavily on these categories in their visualization practices precisely because they animate the entirety of the scientific endeavor.

Similarly, visual communication has been studied for decades in terms of its highly rhetorical qualities [Barton and Barton 1985] [Gallagher et al. 2011] [Tapia and Hodgkinson 2003]. Despite early interventions by journalist Darrell Huff and statistician Howard Wainer, data visualization literature and practice seldom focus on the argument-altering, persuasive
qualities of individual design decisions or visualization conventions to a degree that allows for effectively mitigating the harmful potential consequences of visualization [Huff 1954] [Wainer 1984]. One notable exception to this overall trend is the work of cartography scholar Mark Monmonier, who has long advocated for acknowledging the complexity and nuance inherent in the minutiae of visualization design decisions [Monmonier 1991] [Monmonier 1995]. Huff, Monmonier, and Wainer can be seen as the grandfathers of a small, interdisciplinary body of work on ethical visualization practices that directly tackles the challenge of mitigating the potential for harm inherent in data visualization [Cairo 2014] [Hepworth 2016] [Kostelnick 2007]. We argue that there is an urgent need for this ethical visualization literature to grow in detail and scope, particularly with regard to digital humanities projects. It is imperative that ethical data visualizers evaluate not only the rhetorical decisions of the analysis, but also critically examine the entire process of working with data from collection to final visualization and publishing.

Visual theorist Johanna Drucker offers one proposal to address the challenge of ethical representation. She argues that the humanities need their own forms of visualization, distinct from those developed for administrative and scientific purposes [Drucker 2011, 1]. She does this for good reason: the standard visualization conventions that we are most familiar with — bar charts, line charts, pie charts — were all created in European countries at the height of their colonial expansion and industrial transformation. They were created to track demographics, trade, war, and debt; all the trappings of their growing empires [Wainer 2013] [Cole 2000]. These visualization conventions carry this history, and these associations, with them.

However, in his work on the role of charts in the social sciences, historian Howard S. Becker reminds us that “if we invent a new format every time we have something to say, we risk alienating users” [Becker 2007, 169]. Finding the right balance between visualization innovation and working within established conventions is a complex procedure that demands a combination of high visual literacy, advanced visualization production skills, intimate understanding of the visualization context, and a critical perspective on the entire data collection and visualization process. Much valuable work has been done by geographers in terms of working critically with established visualization formats, in the form of critical GIS [Harvey et al. 2005] [Thatcher 2016].

We argue that for pragmatic reasons, humanists must work with the visualization formats that are familiar to their audiences much of the time. We encourage innovation in visualization practices only insofar as innovations are both intelligible to the intended audience, and that they foster consideration of the dignity of the represented subjects. Therefore, we propose an ethical visualization workflow (see Figure 7) that operates within existing data collection and information design frameworks but ensures that any given visualization’s argument provides a compelling yet ethical and accurate representation of historical subjects.

Prior to creating a data visualization, a scholar following our ethical visualization workflow would complete several critical steps: defining, reviewing, collecting, pruning, describing, surveying, and pre-visualizing. These steps involve processes that many digital humanities scholars will be familiar with, with the important difference that they are suggested here with alterations that we believe will result in an ethical data visualization. The steps can be grouped into three standard digital humanities practice phases: pre-data collection (defining, reviewing); data collection and curation (collecting, pruning, describing); and data visualization and argumentation (surveying, pre-visualizing, visualizing, publishing).

Pre-Data Collection

In the pre-data collection phase, the first step involves clearly defining the subject area that the data visualization will depict, while the second step involves reviewing the latest secondary literature on the topic at hand. Reviewing subject area literature would inform the remaining steps in the ethical data visualization workflow, inviting the researcher to compensate for the data set’s shortcomings by seeking out and including new information, or to limit the scope of the visual argument to be produced with said data. Doing so would avoid the glaring oversights and interpretive overreach that plagued the EJI’s Racial Terror Lynchings Map.

Data Collection and Curation

The second phase, data collection and curation, is perhaps most crucial in producing an ethical data visualization, precisely because it is so frequently overlooked. The third step in the ethical visualization workflow involves collecting primary documents, artifacts, and datasets, as well as secondary datasets of potential relevance, while the fourth step involves checking the appropriateness, authenticity, veracity, and feasibility of use of collected primary and secondary materials, and
pruning those that don’t hold up under scrutiny. Once these two critical data collection steps have been finished, the researcher completes the fifth step, describing, by creating their own dataset that combines the collected materials.

Creating a custom dataset for the researcher’s visualization in this way eliminates other people’s and institutions’ biases from the data, ensuring erroneous arguments are not unintentionally added through using unaltered historical datasets. This process of collecting, pruning, and describing data sets was undertaken by Aaut Studios for ten years before visualization, contributing to the particularly considerate treatment of ethical factors in the Map of White Supremacy Mob Violence. Similarly, EJI created an extensive dataset of over 4,000 public lynchings based on work done by Tuskegee University and the research of E.M. Beck and Stewart E. Tolnay that compellingly shows the long legacy of terroristic violence in the American south. Nevertheless, decisions made about what counts or does not count as “racial terror violence” made during the data collection phase — namely to exclude the American frontier — ultimately shaped EJI’s visual argument, making it overreach in its claims and thereby creating a narrative around racial violence that excludes other minorities and other geographic locales.

<table>
<thead>
<tr>
<th>PHASES</th>
<th>STEPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-data collection</td>
<td>1. Defining field of inquiry</td>
</tr>
<tr>
<td></td>
<td>2. Reviewing latest subject scholarship</td>
</tr>
<tr>
<td>Data curation and collection</td>
<td>3. Collecting primary documents and artifacts</td>
</tr>
<tr>
<td></td>
<td>4. Pruning non-viable primary documents</td>
</tr>
<tr>
<td></td>
<td>5. Describing primary data in custom data set</td>
</tr>
<tr>
<td>Data visualizing and argumentation</td>
<td>6. Surveying ethical visualization literature</td>
</tr>
<tr>
<td></td>
<td>7. Pre-visualization context consideraton</td>
</tr>
<tr>
<td></td>
<td>8. Visualizing data</td>
</tr>
<tr>
<td></td>
<td>9. Publishing visualization</td>
</tr>
</tbody>
</table>

**Figure 7.** The Ethical Visualization Workflow we propose for producing visualizations that minimize harm to three groups: people using visualizations, people represented in visualizations, and people personally affected by the represented material.

**Data Visualization and Argumentation**

The third and final phase of the visualization workflow, data visualization and argumentation, is the main one associated with the ethics of data visualization, and this phase involves four steps: surveying, pre-visualizing, visualizing, and publishing.

**Surveying & Pre-visualizing**

Once the custom data set has been created, the researcher moves onto the sixth step: surveying the latest literature from the small but growing interdisciplinary field of ethical visualization. Surveying this literature allows the researcher to keep
abreast of ethical visualization innovations and recommended best practices. The seventh step, pre-visualizing, involves considering the contextual factors around the visualization: normalizing representations of the data; selecting the publishing medium; identifying intended and potential unintended audiences based on that medium; and discerning between visualization formats possible in that medium. When making decisions about the argument produced by the visualization, it is imperative to consider the larger contextual framing and the story that it tells. For instance, in the case of EJI’s Racial Terror Lynchings, the data isn’t normalized against census population data. Normalizing the data in this way would show how frequent these lynchings were in a population of people rather than a geographic space, thus showing the relative societal impact of a lynching of two people in a county of 100 rather than 5 people in a county of 10,000.

In considering these contextual factors, the researcher can then select the most appropriate visualization format, creating test visualizations (these are more rudimentary than, and distinct from, alpha prototypes) and performing any necessary re-structuring of the dataset based on findings of this prototyping. In the pre-visualizing step, the selection of visualization format is particularly important. The mechanics and conventions of specific visualization formats contribute to determining meaning. For example, Mercator geographic projections have been criticized for privileging Northern Europe and underemphasizing the Global South [Monmonier 2010]. While they receive less attention for their distortional effects, pie charts encourage a comparison between visualized elements as if they make up a unified whole, whether or not they actually do so in reality [Tufte 1998]. It is not that such visualization formats are by definition unethical to use, but that a critical perspective can unearth the ways in which they limit or close off possibilities of argumentation.

Critically examining each of the areas involved in pre-visualizing, and making careful, considered choices on each area, can dramatically change the ethical implications of a project. For example, if the Lynching in America report and Racial Terror Map were printed documents, they would have a much smaller audience than they have as web-based documents. This would result in less reach for the organizations, but also less potential harm in terms of the Racial Terror Lynching map being used in contexts outside of lobbying for prison reform, and therefore giving an erroneous presentation of the history of lynching in the United States.

The pre-visualizing step provides an opportunity to acknowledge the prior understanding and cultural frame of the intended and unintended potential audiences. The persuasive and culturally bound associations those audiences necessarily have with design elements, explanatory text, headers, legends and interaction experiences need to be considered. The choice of colors and color ramps, as well as graphic or cartographic elements like political boundaries, invariably influence the argument produced by the visualization, as do map default views at certain screen widths, and zoom options. To be ethical, these choices must be made with the scholarly literature and the ethical visualization literature in mind, as well as a critical perspective on the power of individual design elements. To maintain visualization ethics, they should strive to minimize harm while increasing understanding, and this can only be done when the latest ethical developments in the field are factored into the visualization at hand.

For instance, given the research on the dehumanization of deaths by aggregating them into faceless statistics and the resulting inability to enact meaningful social change [Du Wors et al. 1960] [Bernard et al. 1971] [Slovic 2007] [Katz 2011], Monroe Work Today created an individual marker for every person killed, so that their name and details could be uncovered by the viewer. While more visually complex than a choropleth map, as was used in Racial Terror Lynchings, creating individual markers conveys the sheer gravity of the violence while not losing sight of the individuals who suffered at the hands of white supremacy. As the director of Auut Studio, RJ Ramey, explains,

   “it was a conscious decision to include every person killed as their own marker on the map — so that their name could be discovered. I have received suggestions that a more “efficient” visualization would have been a choropleth map or graduated symbol size — and visualizers at EJI and the NYT have utilized these methods. However, for purposes of respecting the gravity of the violence and the humanity of the victims, I very much believed it was most appropriate to provide the audience a census of the lynching record, not a visual or numerical analysis.” [Ramey 2017]

In contrast, EJI replicates the conventions, and thereby the limitations, of how the Tuskegee Institute has visualized their data by state and county since 1931 (see, for example: Lynchings by States and Counties, 1931, https://www.loc.gov/resource/g3701e.ct002012).

The visual argument produced must be a bounded one, explaining what it does show while inviting the user to interrogate and explore what it does not show. Effacing the data, along with its assumptions, within the visualization itself, as Map of
White Supremacy Mob Violence so admirably does, invites the user to judge the veracity and scope of the data themselves, providing an opportunity for users to build informed trust in the visualized data. Auut Studio makes plain the decisions made during all of these pre-visualization steps, encouraging the user to take a critical stance toward the argument ultimately put forth by the visualization itself. On the other hand, while EJI does provide a link to its full lynching report, which goes into detail about how the dataset was collected and why it was bounded in such a way as to exclude the American frontier, this link stands apart from the visualization itself, buried within the website’s navigation drawer. Moreover, the report does not address the translation of those choices into the rhetorical qualities of the visualization itself, which purports to show the entirety of American Racial Terror Lynchings. As the argument presenting the data, an ethical visualization should provide clear and apparent options for users to investigate sources, so that the user will be able to get a critical sense of the underlying data set. Showing the data effectively equates to showing your work for the user, but it does not necessitate providing the user with what amount to false choices that obfuscate the rhetorical value of the visualization in the first place.

Visualizing and Publishing

The eighth step involves development of the visualization itself. To be as ethical as possible, this needs to be an iterative process, beginning with an alpha prototype, ending with a final visualization, and including rounds of user testing with intended and unintended audiences after each round of iteration. While user testing is not commonplace in the digital humanities, it is a long-established practice in the allied fields of computer science and visual communication design [Nielsen 1993] [Sanders and Stappers 2012]. Such rigorous prototyping and testing ensures the research mitigates harm to audiences that may result from drawing associations and conclusions that are false, and that the researcher could not predict. Lastly, the ninth step involves publishing the visualization, the culmination of a thorough and considered ethical research practice. Finally, in the interest of reproducibility and data interrogation, it is vital to publish alongside the visualization its underlying datasets or, in the case of lacking the requisite rights, to provide ample documentation and citation of those datasets.

Feasibility of Proposed Workflow

Digital humanities teams can implement this workflow by centering their activity on ethical questions around their subject area and the technology used to present it. This can be accomplished using a twofold approach: firstly, by familiarizing themselves with the latest research in the content field and adjacent fields; secondly, by including team members familiar with the entirety of the data pipeline from collection to cleaning to presentation, as well as communication design principles and user experience methods. User experience design is particularly important for evaluating the interpretive intervention made by the visualization and mitigating harm caused by the final visualization.

We recognize that the combination of skills we advocate is not the norm in digital humanities projects and that it is rare for any one single humanist to possess all of these skills. However, these are skills that are common in several disciplines, particularly in the humanities (source interrogation and intellectual framing); social sciences (data collection, curation, and analysis); and communication design departments (interface design and user experience). Interdisciplinary collaboration in teams that contain — or consult with — humanists, social scientists, and communication designers is the most feasible way to implement the ethical visualization workflow. There is strong interest in digital humanities collaborations in the field of communication design, as evidenced by a recent special issue of Visible Language (see volume 49, issue 3), and there are exemplary collaborations between social scientists and humanists that can serve as models for such teams (see the cooperation between the Digital Humanities and the Social Science D-Lab at University of California, Berkeley).

Ethical data visualization is much more about priorities and planning in digital projects than about increasing the amount of resources, or using the latest technology. Whereas EJI’s Lynching in America had the full support of Google Labs, Auut Studio’s Monroe Work Today resulted from the careful consideration of a single individual working in consultation with domain-area experts. Auut Studio began as a single-person operation, but through proper planning, interdisciplinary collaboration, and attention to historiographical implications, it was able to present a more ethically minded visualization of lynching data sets.

Digital humanists need to bring their skepticism toward their source material to bear on visualizations themselves, and consultation with communication design faculty will illuminate the design and interaction elements that need to be interrogated. The medium matters as much as the content in shaping the message, and thus it is essential to understand the medium in order to fully appreciate a visualization’s societal intervention. In other words, as we stated at the outset of this paper, the technology is not itself value-neutral.
Conclusion

Data visualizations are inherently rhetorical, and therefore bias-laden visual artifacts that contain both explicit and implicit arguments. The implicit arguments depicted in data visualizations are the net result of many seemingly minor decisions about data and design from inception of a research project through to final publication of the visualization. Data workflow, selected visualization formats, and individual design decisions made within those formats all frame and direct the possible range of interpretation, and the potential for harm of any data visualization.

Considering this, it is imperative that we take an ethical approach to the creation and use of data visualizations. Therefore, we have suggested an ethical data visualization workflow — defining, reviewing, collecting, pruning, describing, surveying, pre-visualizing, visualizing, and publishing — with the dual aim of minimizing harm to the subjects of our study and the audiences viewing our visualization, while also maximizing the explanatory capacity and effectiveness of the visualization itself. To arrive at our ethical data visualization workflow, we have examined two recent digital mapping projects, Racial Terror Lynchings and Map of White Supremacy Mob Violence, to demonstrate the potential pitfalls of data visualization, as well as suggest ethical ways to avoid such pitfalls.

While EJI’s Racial Terror Lynchings is an admirable project, it nevertheless presents an incomplete picture of racial lynchings in the American South in a way that forecloses meaningful discussions about racism and white supremacy in the American North and West, leaving out the “forgotten dead” among frontier populations. By contrast, Monroe Work Today’s Map of White Supremacy Mob Violence is an example of the possibilities of ethical visualization, precisely because of the extended period in which the data were interrogated in line with the latest scholarship in the field. Auut Studio brought a critical eye to the topic at hand, acknowledging the shortcomings of the data and investing the time to ultimately create a more inclusive picture of white supremacist violence.

Monroe Work Today’s Map of White Supremacy Mob Violence is an exemplar of ethical visualization, not because it is free of two centuries of baggage and biases, but because it acknowledges them, while also acknowledging the potential pitfalls of the very endeavor of transforming human beings into visualized historical data. Unlike TayTweets, who consumed the racist prejudices and ideologies of our modern society and became an intolerable digital bigot, Monroe Work Today gives real thought to the quality and veracity of the data, the ways in which that data represented (and did not fully stand in for) marginalized persons, and the design decisions taken in visually representing that data. In so doing, it presents an accessible, nuanced and compelling account of America’s sordid history of lynching those on the margins. Likewise, by following our process, readers can create similarly critical and self-effacing visualizations that make apparent the argument, assumptions, and inherent flaws that animate their digital humanities projects.

Notes

[1] Both authors contributed equally to this article.

Works Cited


Ethical Data Visualization Workflow

Proposed by Katherine Hepworth and Christopher Church, 2018

1. Defining - define field of inquiry
2. Reviewing - find and study latest scholarship in subject secondary literature
   a. Selecting
   b. Reading
   c. Associating - making connections between the current literature and the defined field of inquiry
3. Collecting - collect primary documents/artifacts/datasets and secondary datasets of potential relevance. Three sub-steps:
   a. Finding
   b. Capturing
   c. Formatting
4. Pruning - critically evaluate collected primary materials for authenticity, veracity, appropriateness, and feasibility and remove documents/artifacts/datasets that don’t hold up under scrutiny. Ask three questions:
   a. Is this source relevant?
   b. Is this source usable (technically fit and ethically viable)?
   c. Does this source contain duplicate data?
5. Describing - collect data from remaining primary materials and secondary datasets (creating your own instead of using others eliminates other people’s biases in data collection and cleaning) into unique, custom dataset for the visualization project
   a. Identifying data
   b. Scraping/storing data in dataset
   c. Structuring dataset
   d. Cleaning dataset - critically examine the data while cleaning the dataset
   e. Classifying data
   f. Interpreting data
6. Surveying - survey ethical visualization literature, find best practices and current debates
7. Pre-visualizing -
   a. Choose publishing medium (ie html, ) and consequent privacy/public level (ie a print book will have a far smaller readership potential than a website, and therefore less potential to harm)
   b. Collecting - Research available visualization formats, design elements and interactivity for that medium
   c. Identifying Audience - considering design element meaning for that audience; considering audience’s prior exposure to data and subject
d. Pruning - Debate potential visualization formats and narrow down to viable and ethical options

e. Visualization for analysis - rapid prototyping and iteration of viable visualization formats, as an aid to interpreting data

f. Re-structure dataset as appropriate based on findings

8. Visualizing -
   a. Alpha prototyping
   b. Critical reviewing of alpha prototypes to ascertain effectiveness and potential for harm
   c. User testing - preferably with both intended audience

9. Publishing -

   1 = project conception
   2 = literature review
   3-5 = data workflow - creating a custom dataset
   6-8?=? = visualization workflow
Introduction

In January 2000, the doors opened on a new exhibit of photographs at the Roth Horowitz Gallery in New York City. As visitors surveyed the show, they viewed the disquieting images of lynching victims. Many photographs captured dead corpses hanging from trees, bridges, and telephone poles. Some showed mutilated or burned bodies. A number of photographs included the crowds—filled with men, women, and children posing for the camera lens—that attended these killings. A few of the exhibit images were actually picture postcards sold as souvenirs by entrepreneurial photographers. For most visitors touring the exhibit, the images were shocking and disturbing, revealing a troubling chapter in American history. Widely acclaimed, the exhibit was later toured, and its photographs would be published in James Allen’s *Without Sanctuary: Lynching Photography in America*.

The exhibit and the book, however, obscured as well as enlightened. Even for those visitors strong enough to gaze upon every image, one of the most important stories in the history of lynching in the United States was missing. The vast majority of images on display were quite rightly those of African Americans, the group that suffered more than any other at the hands of lynching mobs in the United States. Yet the exhibit failed to include any Mexican victims of lynching. Although there were numerous images of white, or Anglo, victims as well as photographs of a Jewish victim and an artifact related to a Chinese victim, none of the fifty-four images or items displayed in the exhibit (and the ninety-eight plates later published in *Without Sanctuary*) captured an image of a Mexican lynching victim.

From the California Gold Rush to the last recorded instance of a Mexican lynched in public in 1928, vigilantes hanged, burned, and shot thousands of persons of Mexican descent in the United States. The scale of mob violence against Mexicans is staggering, far exceeding the violence exacted on any other immigrant group and comparable, at least on a per capita basis, to the mob violence suffered by African Americans. Yet despite its importance and pervasiveness, mob violence against Mexicans has never been fully studied. More than almost all other victims of lynching, Mexican victims have been the “forgotten dead.”

This book is, in part, an attempt to figure out who these men and women were. Where, when, by whom, and why were they lynched? What did their deaths mean? What was the scale and significance of mob violence against Mexicans? How did their fellow Mexicans respond to these killings and how did they attempt to protect themselves from similar acts of mob violence?

But this book is about more than the lynching victims. It is also about the deeper question hinted at in the title: Why were their deaths forgotten by so many? For those who did not forget, why didn’t they, or why couldn’t they, share their memories with others so that the lynched would not be so forgotten? In short, the book is about both the actual mob violence that claimed the lives of so many persons of Mexican descent in the United States and the reactions by Mexicans, whites, and blacks—at the time and over the past one hundred and fifty years—to that violence.

The study of Mexican victims of vigilantism presents some startling new perspectives on patterns of mob violence. The lynching of blacks and its significant place in American history is not undermined or lessened but instead revealed in a new way when compared to the lynching of Mexicans. The similarities and the differences between mob violence against these two minority groups illuminate larger questions of racial and ethnic conflict in American history.

To give just one example, the study of anti-Mexican mob violence sheds new light on the relative cultural distance between whites and blacks. Most studies of the lynching of African Americans in the late nineteenth and early twentieth centuries portray whites and blacks as alien from one another. Historian Joel Williamson famously wrote that the era of lynching in the American South coincided with “the crystallization of a separate and viable black culture.” Yet, when compared to Mexicans, the gulf between blacks and whites in the South seems less vast. Despite the violent brutality that characterized the American South, by the mid-nineteenth century both blacks and whites shared a common cultural connection in their language, religion, folkways, and food. By contrast, Mexicans spoke a different language and practiced a different religion than the vast majority of Anglo settlers in the West. This created a sense among whites that Mexicans were strange and alien, fueling suspicion and mistrust.
Introduction

The contrasting legal cultures of the United States and Mexico were equally significant in distancing Anglos from Mexicans. Mexico, a country long under European colonial rule and subsequently under dictatorship, based its system of jurisprudence on the Napoleonic code. By contrast, the United States, the first colony in the New World to secure its independence, boasted a federal legal system immersed in notions of popular sovereignty and shaped by English common law traditions. Although whites often restricted access to the courts, African Americans had a long history of successfully utilizing the legal system to push for freedom, beginning with freedom suits during the American Revolution and continuing through Dred Scott and into the twentieth century with the great lawsuits of the Civil Rights Movement. Comparatively, Mexicans were at a much greater legal disadvantage than blacks. This was true especially for those Mexicans who were not citizens but mere residents with limited legal rights in the United States, but it was also the case for many of those Mexicans who became American citizens.

These comparisons help place the African American experience in context and suggest fertile lines for new research. Such comparisons do not detract from the heroism of African American activists, who were terribly disadvantaged in their struggle for equal rights, but it does help explain why different groups took different approaches to the problems besetting racial and ethnic minorities in the United States. For example, if a comparison with Mexicans reveals a relative advantage for African Americans with regard to the American judicial system, that same comparison demonstrates that Mexicans possessed different resources unavailable to blacks in the United States. Despite the advantages of the Northern states and Canada for enslaved runaways and postbellum black migrants, African Americans had no easily accessible refuge in the New World. Mexicans could, and did, flee across the border from Anglo oppression. While Mexico, with its widespread poverty and often negligent government, was no Eden, it was “home” to Mexicans in the United States, not only for those born there but even to the many ethnic Mexicans born or naturalized in the United States. Perhaps more importantly, Mexicans in the United States could urge diplomats from their native country to intervene in cases of mob violence against Mexican nationals in the United States.

Mexican resistance to mob violence is one of the central stories of this book, and it is a narrative that sometimes parallels but often diverges from the experience of African American antilynching activists. The pages that follow highlight acts of heroic resistance by a number of individuals whose names are largely unknown. They include José T. Canales, the state representative whose protests led to a public investigation into the actions of the Texas Rangers, a group long regarded by Mexicans as an instrument of racial oppression and terror. For his efforts, Representative Canales was stalked and threatened with death. In dramatically publicizing mob violence against Mexicans, Canales served a role similar to that of African American activist Ida B. Wells in her crusade against the lynching of African Americans. Opponents threatened to kill both of them for their efforts to expose lynching and end mob violence. Yet, Canales possessed direct access to political power that was denied Wells, a significant difference that helps explain their tactical choices.

Even less well-known than Canales are diplomats such as Manuel Téllez, Ignacio Mariscal, Manuel de Zamacona, and especially Matías Romero, who tirelessly petitioned the State Department in Washington to protect the rights of Mexican nationals in the United States. Other activists include courageous journalists such as Carlos I. Velasco, Nicasio Idar, Francisco P. Ramírez, and Práxedes G. Guerrero, who editorialized against the brutal mistreatment of their people.

This study stretches over eight decades, beginning in 1848, the year that the United States won the US-Mexican War, secured the contested annexation of Texas, and forced Mexico to sign the Treaty of Guadalupe Hidalgo. This treaty transferred to the United States a half million acres of land in what is today the American Southwest. Under the terms of the treaty, the residents of this territory became US citizens, thus introducing into the United States a large number of persons of Mexican descent. Mob violence does not, of course, follow political turning points such as this one. There are examples of Americans who, prior to 1848, exacted mob justice against Mexicans in Texas or along the border. And the first confirmed case of mob violence against a Mexican after the signing of the Treaty of Guadalupe Hidalgo did not take place until 1849. Nevertheless, 1848 marked the American government’s extension of citizenship rights to Mexicans in the expanded American West and is a logical starting date.

The end point is 1928, the year that a mob lynched Rafael Benavides in Farmington, New Mexico. Benavides’s hanging was the last known case in which a mob publicly executed a Mexican in the United States. Although violence against Mexicans continued for decades after 1928, Benavides’s death was the last of its kind and a turning point in the history of mob violence against Mexicans. Extralegal executions of Mexicans after 1928 have become shrouded in secrecy and were never carried out in public.
without fear of legal punishment. Benavides’s murderers were known and continued to live in the community for some time afterward without fear of arrest. Subsequent murderers of Mexicans were not so bold and took greater precautions or faced legal action.

In 1999, an episode typical of this new, more mysterious type of violence took place in a remote part of New Mexico. A maintenance worker discovered the decomposing corpse of a man chained to an electric pole near Deming. A preliminary examination suggested that the man had been dead for one or two months and that he was not immediately executed but subjected to a slow, tortuous death. The reasons for the killing are as elusive as his murderers. Was he killed by drug lords or as part of a hate crime? With such deaths, historians simply have too little information to place them in any kind of comparison with public Lynchings that took place in the late nineteenth and early twentieth centuries. For these reasons, the book concludes the systematic study of anti-Mexican mob violence in 1928 and includes more recent materials primarily when it relates to the memory of that earlier violence.

Between 1848 and 1928, mobs killed an unknown number of Mexicans. Conservative estimates place the number in the thousands. This study is not, however, based upon a collection of estimates but instead utilizes a set of data compiling cases of actual individuals murdered by mobs. This inventory, incomplete as it is, contains data on 547 victims and can be found in the appendix.

Some historians suggest foregoing the compiling of any systematic data out of concern for how such numbers will be perceived by a public that resists statistics. Readers might conclude that this list of victims somehow equals the actual number of Mexicans killed by mobs in the United States, when in fact they are a fraction of the actual number of Mexicans lynched, a total that will never be known because it is impossible to recover the names and dates and places of all those killed by mobs.

Despite the difficulties of parsing all of the cases of mob violence against Mexicans, we have persevered in compiling our list of Mexican victims for two reasons. First, such a list can be a stimulus to further research on the subject. The inventories compiled by the NAACP and other civil rights groups, flawed as they were, clearly galvanized sociological and historical research on African American lynching. Second, as long as the subject of anti-Mexican mob violence rests upon anecdotal evidence alone, without an actual count of victims, it will continue to be received skeptically by both scholars and the general public. Many people tend to disbelieve in
great tragedies until forced to face overwhelming evidence to the contrary. Determining the precise number of people killed in the Holocaust, for example, is impossible, just as it is for Mexican victims of mob violence. Yet, in both cases, numbers matter.

With hundreds of specific victims in our list, mob violence against Mexicans cannot be dismissed as a rare occurrence. This violence was, not surprisingly, concentrated heavily in those states bordering Mexico (see Table 0.1). As a consequence, this book largely focuses on the four southwestern states of Arizona, California, New Mexico, and Texas (see Figure 0.1).

The number of Mexicans executed by vigilantes compels us to reconsider the geography of mob violence as a whole. A standard lynching map of the United States depicts mob violence as being predominantly a phenomenon of the Deep South. By including data on mob violence against Mexicans, it can be seen as much more common occurrence in the southwestern states than has been previously understood. This changed perspective will be even clearer when there are more studies of mob violence in the American West that detail vigilantism against the Chinese, Native Americans, and men and women of European descent.

In 1949, progressive journalist and bestselling author Carey McWilliams wrote in North from Mexico, his classic survey of Mexicans in the United States, that “vast research would be required to arrive at an estimate of the number of Mexican lynchings.” In researching this book, we have come to a fuller appreciation of the truth of this statement. We have seen only a few
mentions of Mexican lynchings in the traditional sources used by lynching scholars, such as the archival records of the National Association for the Advancement of Colored People (NAACP). Had we known then what we had embarked upon, we might have deemed the project unfeasible. Given the number of Mexicans killed by mobs in too many places, it is impossible to uncover every surviving document related to the lynching of Mexicans in the United States. Nevertheless, we have pored over a wide variety of sources to create the first systematic study of Mexican victims of mob violence. One narrative episode constructed from our research may help illustrate our methodological approach.

At two o’clock in the morning of May 3, 1877, a mob seized Francisco Arias and José Chamales from their jail cells and hanged them from the Upper San Lorenzo Bridge in Santa Cruz, California. The mob alleged that the two men had killed and robbed a carpenter named Henry De Forest two days earlier. When the mob attempted to wrangle a confession from the prisoners, according to one newspaper report, each man denied culpability and indicted the other as the murderer. The two men, both natives of California, were widely reported to be ex-convicts. Their obfuscation over who killed De Forest meant little to the mob. Observers noted that the mob consisted largely of men from the vicinity of De Forest’s home in Felton. The Lynchers placed the two men in a small wagon and transported them to the bridge. There, they tied ropes around the necks of the prisoners, and then drove the wagon away from the bridge. One account stated that the region experienced relief after the hanging of these “desperate assassins.” None of those involved in the murder were ever indicted or prosecuted. The brief investigation into the affair concluded that “parties unknown” had caused the deaths of the two men.

The sources that helped us reconstruct this narrative are the most numerous documents available for the study of lynching in the United States: newspaper accounts. Newspapers often provide detailed information on the crime allegedly committed by the mob’s victims, descriptions of the lynching itself, and editorial commentary on the episode’s aftermath. In the case of Francisco Arias and José Chamales, the coverage of the Sacramento Union, one of at least a half dozen newspapers that reported the story, is illuminating. First, the Union stated the motive for the killing of De Forest was robbery and that he was targeted at random. Second, the paper noted that the mob broke open the jail yard door and forced the jailer and the deputy sheriff to turn over the keys to the cells holding Arias and Chamales. The two men were then taken from the jail, their hands and feet bound with hay ropes, and transported to the bridge where they were hanged. Third, the Sacramento Union concluded with an editorial endorsement of the affair: “we really do not see how such wretches could have been more satisfactorily disposed of than upon the gallows.”

Newspapers were not objective reporters, but they are critical for the reconstruction of mob violence in the nineteenth and early twentieth centuries. Many small communities supported weekly newspapers, making it possible to recover local details on lynchings in remote locations. Even when the last copies of those smaller newspapers have disappeared, their reporting often survives because it was clipped and copied in larger urban newspapers.

Several Spanish-language press sources, such as Francisco P. Ramírez’s El Clamar Público of Los Angeles, have also survived. Though
Spanish-language newspaper accounts do not exist for most cases of mob violence against Mexicans, taken collectively, these sources reveal much about Mexican reactions. Displaying the same wide range of attitudes as the Anglo press, these newspapers defended and at other times condemned lynching. They often provide differing interpretations of cases involving Mexicans and were far less sympathetic, in general, to contemporary defenses of mob violence.

Another invaluable category of sources is government documents: county level criminal court proceedings, prison records, death and birth certificates, the files of state agencies and police authorities, federal census records, diplomatic materials received and produced by the US State Department, and the correspondence of diplomats, governors, adjutants general, and other officials. Any particular case is unlikely to be recorded in very many, if any, of these sources, but the few instances where such documents can be tied to lynching victims are valuable.

Although nothing could be found on José Chamales in California’s penitentiary records, the file for Francisco Arias tells us that he was born in 1852 in California when it was still part of Mexico. He stood nearly five feet nine inches tall. In 1859, when the prison at Santa Quitenia admitted him for the first time, officials listed him as a laborer. Convicted of grand larceny, he also bore evidence of a difficult life. The prison records indicate that scars covered his body, including his ears, wrist, arms, and shoulders. His thumb was crooked from being broken at some point in his life. Nothing is known of his life after his release from San Quentin in 1866 until the time he returned to prison on conviction of assault to do bodily harm on March 6, 1871. He served eighteen months of a twenty-four-month sentence and was released on November 20, 1872. He again disappears from the historical record until his fateful encounter with De Fores.17

Most of the time our searches for prison records turned up nothing because nineteenth-century penitentiary documents from the American West are so incomplete. We were fortunate to find Arias’s prison record, but the most unusual source uncovered related to the lynching of Arias and Chamales was a photograph (see Figure 0.2). For years, there was very little attention given to the photographic record of lynching victims in general and virtually no analysis of images of Mexican victims of mob violence.18

As with most lynching photographs, “Hanged at the Water Street Bridge” was taken after the lynching. It was clearly shot during daylight hours, whereas Arias and Chamales were hanged at two o’clock in the morning, meaning that their corpses had been suspended for many hours when they were discovered. The men who committed the lynching are almost surely not pictured, having departed the scene. The suit-wearing men and the barefooted boys photographed were spectators. That they did not cut down the bodies but instead called upon and then posed for a photographer says much about the culture of lynching in Santa Cruz in 1877.19

Our use of photographs emphasizes the value of sources other than the traditional written record to reconstruct the stories of Mexican lynching.
victims. Oral testimony may be even more significant. Many Mexicans living in the nineteenth and early twentieth centuries could neither read nor write, preserving stories of mob violence not in written sources but rather through oral tradition and folklore. There is little doubt that Mexicans in the Santa Cruz region talked about the lynchings of Arias and Chamales and fashioned their own interpretation of that event’s meaning. Indeed, the episode must have been particularly painful because newspapers reported that Arias still had relatives in the area. What was said at the time and later, however, has not been preserved. Still, whatever oral tradition did exist combined with the widely reproduced photograph to keep alive the memory of the lynchings among Latinos in the United States. In 2002, the Latino poet Martin Espada published a poem inspired by the photograph of the lynched Arias and Chamales. 13

Two Mexicanos lynched in Santa Cruz, California, May 3, 1877

More than the moment
when forty gringo vigilantes
cheered the rope
that snapped two Mexicanos
into the grimacing sleep of broken necks,
more than the floating corpses,
trusted like cousins of the slaughterhouse,
dangling in the bowed mute humility
of the condemned
more than the Virgin de Guadalupe
who blesses the browskinned
and the crucified,
or the guitar-plucking skeletons
they will become
on the Dia de los Muertos,
remain the faces of the lynching party;
faded as pennies from 1877, a few stunned
in the blur of the execution,
a high collar boy smirking, some peering
from the shade of bowler hats, but all
crowding into the photograph.

We have not emphasized such contemporary references in our research, but we have sought out folktales and oral records of Mexicans

from the late nineteenth and early twentieth centuries. Spanish-language ballads called corridos were very popular forms of conveying narratives and interpretations of Mexican American history in the Southwest, and we have relied upon these sources for inklings of Mexican attitudes toward the violence that was so prevalent in the Borderlands. In addition to corridos, we have also learned much from commonly repeated legends and myths. Finally, we have been fortunate to find that local historians conducted numerous oral interviews with Mexicans, especially in Texas, that allow us to recover, however imperfectly, an important dimension of mob violence in the United States. 14

Studying lynching means contending with the scattered nature of the sources. Although the lynching of Arias and Chamales was relatively well documented, there are several significant categories of primary sources relied upon in this study for which there are no records in this particular case. For example, among the most important sources for this study are diplomatic records maintained by both the United States and Mexico. We have not exhausted these sources, to be sure, but the correspondence, investigative files, and newspaper clippings collected by US and Mexican diplomats were absolutely crucial to our research. Unlike materials Mexican diplomats kept and later deposited in the Archives of the Secretaría de Relaciones Exteriores in Mexico City, Mexican officials did not investigate the lynchings of Arias or Chamales because both men were identified in the earliest accounts of the lynching as having been born in the United States. Indeed, Chamales was said to have been born three hundred yards from where he was hanged. Yet, the fact that Arias and Chamales were not Mexican nationals means that they produced no diplomatic records for historians to consult.

At the opposite end of the spectrum from diplomatic records were the sources created by ordinary men and women in the Borderlands. Particularly during the Gold Rush era, thousands of miners wrote letters, made entries in diaries and journals, and found time later in life to compose their memoirs. While no such sources could be located for the Arias and Chamales case, such documents were essential to our research of mob violence in the 1850s and, upon occasion, very helpful in later years. These sources, especially memoirs, are irreplaceable documents revealing the inner thoughts of the men and women living amidst the violence of the American West in the nineteenth and early twentieth centuries.

Almost all studies of lynching in the United States have relied upon the files of several civil rights organizations, notably the records of the
NAACP, the lynching clipping files collected by sociologist Monroe Work at Tuskegee Institute, and the archival collections of the Association of Southern Women for the Prevention of Lynching. These sources proved much less complete and useful for studying Mexican lynching victims than they are for studying African American lynching victims (or even Anglo lynching victims in the American South). To begin with, the published summaries of the data collected by these organizations divided lynching victims into two racial categories, black and white. No allowance was made for the fact that the “white” category included Mexicans, Naive Americans, Chinese, and a host of other ethnic minorities who were not considered fully “white” by the Anglo mobs that lynched them. For example, Tuskegee reported that mobs lynched thirty-six people in New Mexico between 1882 and 1968. Thirty-three of these victims were listed as “white” and the other three were listed as “black.” Our investigation of these records indicates that nine of the thirty-three whites were Mexicans and that one was a Native American. The pattern of ethnic misidentification is prevalent throughout the data on the western states. We have partially overcome this difficulty by our direct inspection of the archival materials, but we were often forced to use surname, an imperfect instrument to be sure, to identify potential Mexican victims of mob violence. We then tracked down the individuals in other sources, such as newspapers, to find clues as to the ethnic identity of the victim.

Beyond the binary racial categorization, there is another serious problem with relying upon these inventories for data on Mexican lynching victims. They seriously undercount victims of western vigilantism in general and Mexican victims in particular. According to Tuskegee’s archival records, the states of Texas, New Mexico, Arizona, and California played host to fifty lynchings involving Mexican victims during the same period that our research has turned up over two hundred Mexican lynching victims. Even more limiting is the fact that none of the major inventories of lynching victims begins earlier than 1882, and more Mexican victims of mob violence died prior to that date than after. Nonetheless, these inventories and collections were still helpful. Organizations like the NAACP did not intentionally refrain from collecting material on Mexican victims of mob violence, and they have preserved important materials in several instances.°

It is ultimately impossible to recover from the obscurity of history every Mexican murdered by vigilantes in the United States. While this study represents the most exhaustive treatment of the subject to date, some of the victims remain unknown.

The book begins with an exploration of causes and origins of mob violence suffered by Mexicans in the United States and ends with a discussion of the reasons for the post-1928 decline in Anglo tolerance for public executions of Mexicans. It is organized thematically, with the first half exploring characteristics and patterns evident in the history of anti-Mexican mob violence, and the second half turning to the history of Mexican resistance to lynching.

The book balances the social scientist’s desire to generalize broad patterns from particular data with the historian’s understanding of the powerful impact of time and place on people and events. Thus, the book cites evidence drawn from multiple eras and regions to buttress arguments while also deducing sections to the variations that come from studying a region as diverse as the American West.

It begins by focusing on some of the most basic questions surrounding Mexican victims of lynching mobs. The first chapter aims to chart the scale, scope, and general characteristics of mob violence against Mexicans, explaining why so many persons of Mexican descent were killed by lynching mobs from 1848 to 1928. While Chapter One underscores the degree to which lynching mobs targeting Mexicans were often motivated by the same compulsions that drove vigilantes more broadly, the second chapter analyzes the lynching of Mexicans from a comparative perspective, highlighting those aspects of anti-Mexican violence that diverged from more general patterns, especially the patterns found in African American lynching.

The second half of this book is dedicated to the struggle against lynching by Mexicans and their allies. Chapter Three examines the reaction of Mexicans resident in the United States to lynching and vigilantism. Due to the proximity of the border, many Mexicans fled from such violence, returning to the shelter of their ancestral homes and kin networks to the south. Others chose not to flee but to protect themselves and their families by aligning with sympathetic Anglos and assimilating to American culture to a greater or lesser degree. Numerous Mexicans, however, responded in kind to the attacks and violence they suffered. Such men became “bandits” in the eyes of Anglos but folk heroes in the songs and oral tradition of the Mexican people. Finally, Mexican civic leaders from newspaper editors to politicians consistently voiced strident criticism of mob violence and vigilantism.

Despite the impressive resistance orchestrated by Mexicans in the United States, the most effective form of opposition to this violence in
the United States came from Mexico City. Chapter Four explores diplomatic protest against the abuse of Mexican nationals living in the United States. Such protests waxed and waned according to the internal politics of Mexico. Furthermore, officials in Washington often ignored the pleas of Mexican diplomats. Yet, few opponents of lynching in the United States equaled in resources those Mexican diplomats who protested anti-Mexican mob violence. In the late nineteenth and early twentieth centuries Mexico City could claim real progress relative to opponents of African American lynching in the American South.

For all of their political power, those diplomats who sought to end abuse of Mexicans by mobs had very little control over how the media and popular histories would frame the memory of anti-Mexican violence. In the decades that followed the lynching of Rafael Benavides, historical memory of Mexican lynching largely faded from public consciousness.

Not everyone, of course, forgot that Mexicans were lynched in the United States. Even when confined to the margins, native Spanish speakers in the Southwest held onto their stories through a powerful oral and written tradition. As the Latino presence in the United States increased during the twentieth century, they were increasingly able to raise voices of protest in the mainstream culture. The Conclusion traces the reasons for the decline of lynching of Mexicans, the limited attention paid to this history in the mainstream culture, and how Mexicans in the United States, despite this neglect, preserved their memories of what took place in the late nineteenth and early twentieth centuries.

In the first decade of the twenty-first century, a new vigilante movement emerged in Arizona targeting Mexicans. In 2000, an organization known as Neighborhood Ranch Watch launched a campaign against Mexican migrants along the border between Sonora and Arizona. According to its leader, Roger Barnett, the ranchers had arrested thousands of illegal immigrants and said that a serious accident “with me or somebody else” was inevitable. Throughout the decade, numerous Mexicans were found dead in the Arizona desert. Some had clearly been murdered, whether by Anglo vigilantes or by Mexican criminals was not clear. While the murderers of these men remain unknown, there is little doubt that some Anglos along the border see Mexican immigrants as hostile invaders who are an unacceptable threat to traditional American culture. In the words of Glenn Spencer of American Border Patrol, there is a “wholesale invasion” of illegal migrants from Mexico that is being orchestrated “with hostile intent” to subvert the United States.

Mainstream media discussions of these border killings have rarely placed them in the long history of conflict and violence between Anglos and Mexicans in the United States. By contrast, contemporary discussions of hate crimes against blacks have almost always situated such crimes in the long history of antiblack mob violence in the United States. To be clear, contemporary violence against Mexicans is not identical to the violence that Mexicans suffered in the nineteenth and early twentieth centuries, just as contemporary hate crimes against blacks are not identical to the earlier violence to which African Americans were subjected. Yet, to many Mexicans, contemporary violence between Anglos and Mexicans can never be divorced from the bloody history of the Borderlands. They remember even if the rest of the country does not.

This persistence of racial conflict along the border was not the impetus for our writing this book, but it does underline the importance of a history of mob violence against Mexicans in the United States. Such a study provides a clearer contextual understanding of modern-day hate crimes, placing them within the longer history of Anglo violence against Mexicans while also helping better illuminate what makes contemporary violence ultimately different than that of the nineteenth and early twentieth centuries. This book not only seeks to recover the “forgotten dead” but also offers lessons from the past for those concerned with persisting conflicts arising out of race and immigration.
Day 3:
Contemporary Problematic Data —
Politics and Culture

Morning 3:
Creating Culturally and Politically Sensitive Visualizations
9:00am to 12pm (3 hours)

- Lecture: Problematizing visualizations — The visual rhetoric of graphic elements and their impact on audiences

Afternoon 3: Discussion and Project Time
1:30pm to 4:00pm (2.5 hours)

Two activity options
Students will do prepared activities on the day’s theme, or they can work on their own datasets / projects
- Prepared activity and discussion: Explore the data at the following sites — Fatal Encounters, Fatal Force, and The Counted — which one presents the data in the most ethical way? Why? Choose one of these data sets to make a visualization from. Think of a target audience you would like to understand this information better. Choose a visualization format and graphic elements that are sensitive to the represented material and appropriate for the intended audience.
- Project time: apply principles and practices learned so far to your own treacherous data
Discussion: Challenging Racism with Visualizations
THE RACIAL ELEMENTS OF THE POPULATION OF THE UNITED STATES.
Problematizing Visualizations
Design elements are seen diversely

Perception of color, line, shape, movement and interactivity is broad
What you design is not what all your users experience
Design elements are rhetorical

Every design element contributes to arguments within your visualizations
Explicit rhetoric
Framing

Use of images, words, or other content to manipulate how people perceive data and/or information
MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions.
Iraq's bloody toll

The biggest killers

Designer: S Scarr
Implicit rhetoric
Priming

Activation of specific concepts in memory to influence interpretation and understanding
Gun deaths in Florida

Number of murders committed using firearms

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

Designer: CH Chan
The Amalgamation of the White and Black elements of the population in the United States.

Amalgamation des elements blanc et noir pendant la population americaine.

Done by Anna Wightman.
William B Willis

Also recorded as: William W Wilson

:: Black male lynched in May 1906
+ Richmond Co.

He was lynched by a mob after being accused of murder.

Source [1][2][3][4]
2010 Census Block Data

1 Dot = 1 Person

- White
- Black
- Asian
- Hispanic
- Other Race / Native American / Multi-racial

What am I looking at?...

Designer:
D Cable
PBS The Effect of Color

https://www.youtube.com/embed/nX0DHd5QNS8
Design elements are discursive

Every design element prescribes the bounds of meaning users can gain from your visualizations
Governmentality, Technologies, & Truth Effects in Communication Design

Katherine Hepworth

Abstract This chapter argues that communication design knowledge and artifacts are inherently governmental. As a means of communication that combines aesthetics and function, communication design knowledge is a product and producer of a uniquely pervasive form of governance that has seldom been studied. While several researchers and philosophers have expressed interest in the relationship between power, communication design knowledge and communication design artifacts, the governance inherent in communication design has yet to be seriously investigated. Building on the author’s PhD research, this chapter extends Foucault’s theories of
DAY 3 READINGS

[add readings here]
The Visual Rhetoric of Data Displays:
The Conundrum of Clarity

—CHARLES KOSTELNICK

Abstract — The visual rhetoric of data displays (e.g., charts, graphs, maps) has changed profoundly over the past 50 years as a result of research in display techniques, the application of traditional and emerging rhetorical approaches, and the democratising effects of data design technology. Perhaps in no other visual realm than data design is the notion of clarity more critical or more contested. Indeed the ascendancy of rhetorical approaches was initiated by the perceptual/cognitive science of data design, which in seeking to identify optimal display techniques, fostered a concern for ethics and evoked the universality and minimalism of modernist aesthetics. The rhetoric of adaptation, which emphasizes the variability of audiences, purposes, and situational contexts, rendered clarity contingent and mutable—a moving target that requires constant attention. Social rhetoric considered data design as a collective construct, tethering clarity to visual discourse communities, convention-building, cultural values, and power. The concept of clarity has been further reoriented by the rhetoric of participation, which is fostered by interactive digital design that enables users to adapt displays according to their needs and interests.

Index Terms — Data displays, graphics, information design, information visualization, visual rhetoric.

Charts and graphs appear today nearly everywhere—technical reports, research articles, and annual reports as well as less formal documents such as fact sheets, brochures, newsletters, and even monthly power bills. And as the internet has grown, the ubiquity of data displays online has vastly accelerated. Now, in the first decade of the 21st century, consumers of information are immersed in data visualization. Advances in the technology used to create and display charts, which are increasingly interactive and on a screen rather than on static paper, ensure that this trend will only continue. We are already inundated with data displays, and a deluge is heading our way.

In order to glimpse that future, I will outline several approaches to the visual rhetoric of data design over the past half century, using clarity as a touchstone by which to compare and contrast these approaches. Because data displays are quintessentially utilitarian in nature, the rhetoric of data displays begins with the issue of clarity and the injunction that they must facilitate the reader’s comprehension of the data. We are immediately drawn to Tufte’s maxim that “graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency,” that it “gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space” and that it “requires telling the truth about the data” [1, p. 51]. Who would dare contest such timeless, universal, and self-evident maxims that coalesce around the gold standard of data design—clarity? What reader would expect anything less?

However, from the standpoint of readers who individually and collectively interpret the billions of data displays produced annually—and especially now in an environment of fast-moving technology with a global reach—how exactly do we define clarity? When viewed through the reader-oriented lens of rhetoric, and through the multiple lenses of several rhetorical approaches that have evolved over the past 50 years, the concept of clarity can be multifaceted, complex, and even contradictory—in short, something of a conundrum.

To define and explore that conundrum, I begin with the rhetoric of science, which is built on perceptual/cognitive models and empirical research. Although the advocates of this approach are not rhetoricians per se, how they conceptualize readers certainly has rhetorical implications. Next, I examine clarity through two other lenses: the rhetoric of adaptation, which extends the rhetorical tradition of tailoring communications to specific situations and where clarity is contingent on audience, purpose, and context; and social rhetoric, which views data design as a process of communal convention-building whereby readers interpret displays through their collective learning, experience, and values. I then examine the emergence of digital data design and the rhetoric of participation, in which readers actively manipulate displays, digging beneath the surface and exploring and exploiting them according to
their own interests and interpretive preferences. Although I discuss these four approaches in the rough sequence in which they have developed, they all have their advocates today. Moreover, the rhetorical approaches are not absolutely distinct from each other by any means, and so they cannot be discussed in total isolation. And although data design is a rhetorical process between designer and reader, I focus primarily on the reader’s perspective.

Along the way, I assess the profound impact of technology on visual rhetoric, both in terms of interpreting computer-generated print displays and interacting with them online. Technology has always played a role in data visualization, and over the past two decades it has revolutionized its design, production, and reception—particularly the emerging technology of digital data display. The rhetoric of data design parallels the rhetoric of new media, which has both “centripetal” and “centrifugal” aspects, whereby it pulls to the center certain conventional practices while at the same time it spins outward as both designers and users chart their own ways [2, pp. 331–332].

THE RHETORIC OF SCIENCE: THE NEXUS OF SCIENCE, PERCEPTION, AND AESTHETICS

For some time, particularly in the field of statistics, researchers have studied the perceptual effects of data displays in an attempt to identify the most effective practices—in other words, displays that exemplify clarity. The rhetoric of clarity is rooted in what Brasseur calls the “perceptual cognitive-based school of thought” [3, p. 4], the rational, research-intensive approach that she claims dominates the field of data design [3]. The objective of this approach is to identify principles of design that will ensure the optimal transmission of data from designer to user. In static print displays, this approach extends the so-called “transmission” model of communication pioneered by Shannon and Weaver [4]. A belief in the shared rational thinking of users is its creed, and nature and science are its guides.

The cognitive perceptual school derives its principles from empirical research into effective data display, which goes back far longer than 50 years, as MacDonald-Ross documents it [5], and is exemplified in the information-processing models of Bertin [6] and the research of Cleveland and McGill [7], who studied graphical forms based on perceptual and cognitive criteria. Empirical research like Cleveland and McGill’s sought to address the question of which display techniques (e.g., area, lines, grayscale, color) enable optimal interpretation (i.e., those that engender the greatest clarity). For example, according to Cleveland and McGill’s research, readers can compare data on bar graphs better than on pie charts because bar charts adhere more closely to the optimal configuration: “Position along a common scale” [7, p. 536]. Areas, volumes, and grayscale, on the other hand, yield less accurate results for readers (see also Cochran, Albrecht, and Green [8]). Cleveland’s famous “dot chart,” which is illustrated in Fig. 1, applies this research by adhering strictly to the “position along a common scale” principle [7, pp. 545, 547–548], [9, pp. 12–15, 34, 302–340].

This approach to data visualization finds its intellectual roots in the work of Neurath [10], who in the early 20th century invented his Isotype pictographic displays, a system that was founded on philosophical positivism and gestalt principles of perception and that was intended to make statistical data accessible to a wide public audience [11]. This rational, scientific approach is exemplified by many recent scholars and practitioners in the field, chief among whom Bertin, whose semiotic model of data display is based on cognitive and perceptual principles of “information processing” [6, pp. 180–184] (see also [12]). In the area of computer-generated and online data design, both Spence’s Information Visualization [13] and Card, Mackinlay, and Shneiderman’s Readings in Information Visualization [14] (a collection of papers on cutting-edge concepts and techniques) have a cognitive-perceptual orientation. Although Tufte’s design philosophy is too complex to be easily categorized, he also largely advocates this approach [3], given his emphasis on universal principles of design and his ideal of having data displays emulate nature [15].
One could argue that this rational, cognitive-perceptual approach is essentially non-rhetorical because it assumes that readers are universal—that they are largely undifferentiated from each other and share a common interpretive framework, in that their brains are hard-wired to process some design elements better than others. Ideally, however, the tight and abiding focus on audience of this approach yields rhetorical benefits—mainly, the efficiency and predictability of readers’ responses.

Researchers can identify forms with optimal clarity, and designers can apply these findings by calculating the interpretive gain or loss in deploying certain forms. Readers are well served because their visual processing is maximized, they consequently have a high degree of confidence in the data, and they are spared perceptual disasters. This rhetoric of clarity with its emphasis on the rational efficiency of language finds its modern antecedent in the 17th-century rhetoric of the Royal Society of London, which eschewed verbal embellishments that impeded the transmission of fact and truth. And indeed experimental graphs that appeared in the early Philosophical Transactions of the Royal Society modeled this standard visually [16].

This rational, efficient rhetoric of data design embodies an intrinsic ethical component because it implies that readers deserve a full, unadulterated disclosure of the data and that designers have a moral imperative to provide it. This imperative was illustrated at least as early as 1914 by Brinton’s Graphic Methods for Presenting Facts [17], in which he exposed deceptive practices of displaying data by explaining the underlying flaws and how readers could identify them to avoid being hoodwinked. The moral imperative of clarity culminated in the early 1950s with Huff and Geis’s classic How to Lie with Statistics [18], which demonstrated how designers, unconstrained by graphical standards or professional oversight, can manipulate charts and graphs for their own ends. Caveat rhetor: Unwitting readers need to protect themselves from these practices and arm themselves with the interpretive tools to unmask deceptive data design.

Tufte extends this tradition by aiming to protect visual consumers from deceptive and abominable design practices, which are particularly prevalent in the popular media. Beginning with The Visual Display of Quantitative Information over two decades ago, Tufte has emerged as one of the prominent figures in contemporary data design and his influence spans many disciplines. For Tufte, “telling the truth about the data” [1, p. 51] requires the designer to avoid the “Lie Factor” [1, p. 57], whereby the size of the data fails to match its visual representation. In addition, credible displays should be structurally transparent, avoid clutter (what Tufte calls “chartjunk” [1, p. 107]), and be visually concise (Tufte’s “Data-ink ratio” and “data density” [1, pp. 93, 162]). All of these prescriptions ensure that readers are protected from the excesses, malpractice, and duplicity of designers.

This heightened vigilance about ethical issues in data visualization is highly warranted, largely because data can be visualized in so many different forms, like the shifting sands on a beach. The design templates and wizards of graphing software make readers particularly vulnerable. Three-dimensional displays often hide data or impede the reader’s ability to make comparisons; pie, donut, and mountain charts require readers to compare oddly shaped areas; and options for filling bars, plot frames, and other forms can result in eye-straining clashes of colors and patterns. By undermining the clarity of the display, these perceptual faux pas also weaken the ethos, or credibility, of the designer. Readers may wonder, for instance, why the designer is hiding data in a forest of bars in the 3-D multiple bar chart shown in Fig. 2. What the designer probably intended as a way to bolster ethos—visualizing the data in the seemingly more complex, sophisticated, and labor-intensive realm of perspective—may very well have the opposite rhetorical effect, with the perceptual challenge undermining ethos by degrading clarity.
The rhetoric of science also embodies an aesthetic element that both belies and complements its rationalism. Brasseur claims members of the “graphing culture,” particularly mathematicians, often celebrate the aesthetics of data design [3, pp. 27–28]. Although he derides graphics that foreground the artistic element, Tufte the social scientist waxes aesthetic about the “wonder,” “beauty,” and “graphical elegance” of well-designed displays [1, pp. 121, 137, 177]. By eliciting a subjective response from readers, this ancillary aesthetic element engenders two rhetorical effects. First, it makes displays more inviting to readers because readers are naturally drawn to elegant displays; and second, it bolsters their credibility because beauty and truth are cognate qualities.

The rhetoric of science also embodies a broader and more pervasive aesthetic element that closely dovetails with 20th-century modernism. Modernists advocated minimalist, high-contrast displays guided by perceptual principles—particularly gestalt principles—so that they could appeal to large public audiences, often across cultures and national borders, as exemplified by Neurath’s Isotype [10]. Modernists emphasized direct, unmediated communication that was objective, perceptually pure, and unburdened by past conventions, resulting in what Kinross calls the “rhetoric of neutrality”—an attempt to trump rhetoric but which itself engendered a rhetorical stance [19, pp. 22–29]. With its emphasis on objective, universal, and culturally neutral design, modernism complements and indeed fosters the rhetoric of science.

So if we examine the display in Fig. 3 through the lens of the rhetoric of science, several questions come to the fore about its clarity. Does the display enable optimal perceptual uptake of the data? The high figure-ground contrast of the bars, the display of data points “along a common scale,” and the gridlines will all help here. Is the display lean and efficient? Bars are less so than lines or dots, but they are at least thin, and they don’t contain distracting color or patterns. Does it “tell the truth” about the data? The X-axis begins at zero, and the plot frame does not stretch inordinately in either direction, so it probably meets this criterion. Is it elegant? Its lack of complexity and sophistication probably prevent it from rising to that level, but its stoic simplicity might be pleasing, if not inviting, to readers.

**Fig. 3. Sample bar chart.**

**Rhetorical Adaptation: Contingencies of Data Visualization**

Clarity and ethos, however, can be contingent and even elusive when data displays are interpreted, as they always are, in specific rhetorical situations; that is, they are interpreted relative to the audiences for the display, its purposes, and the contexts in which readers encounter it [20], [21]. While the rhetoric of science aspires to a Platonic-like, universal ideal that supersedes time and place, the rhetoric of adaptation finds it rooted largely in traditional rhetoric and hinges on the opportunities and constraints of communicative interactions. Although the philosophy of rhetorical adaptation is now common in the field of professional communication, and increasingly so in graphic design, it is still relatively foreign to science and social science fields like statistics and economics, which rely heavily on data design.

That the rhetoric of adaptation remains obscure in many disciplines is not surprising, given its variable and relativistic nature, which eludes scientific description and lacks certainty. For one, the rhetoric of adaptation accepts—even celebrates—that different readers have different interpretive frameworks that profoundly influence what they find clear and credible in data displays. A highly technical chart that appears in a journal for high-energy physicists may be perfectly clear to that audience, but that same chart may be inscrutable to lay readers. A simple, low-key display (like the one in Fig. 3) that readers are highly motivated to explore may engage them quite successfully, but for indifferent readers, clarity may be irrelevant if the display fails to capture their attention. As shown in Fig. 4, the display may need more energy—lightly shaded bars against a rich pattern in the plot frame—to spark the reader’s interest. If that version is still too low key for the rhetorical situation, the designer may even have to
In short, the clarity of a data display depends on its fit for a given rhetorical situation. A good match for one rhetorical situation may be a disaster in another, and vice versa. Of course, however skillful and well-intentioned a designer may be about analyzing a rhetorical situation and adapting displays to it, the match with a given situation can remain at risk, especially without direct feedback from the audience. By studying design in context and by eliciting holistic feedback directly from readers, designers can interrogate the clarity of a display for a given rhetorical situation using context-specific research which entails a variety of methods—testing, interviews, questionnaires, observations, focus groups, and the like [20]. Schriver, for example, extends rhetorical analysis to action by envisioning design as an iterative, interactive process of radical audience adaptation, whereby designers constantly solicit, analyze, and implement reader feedback [22].

The opportunities for implementing rhetorical adaptation have increased exponentially over the past 15 years by the widespread access to computer-generated graphics. In the history of data design over the past 350 years, no tool has more profoundly revolutionized data visualization by drastically reducing the labor to create and modify charts and graphs. A decade and a half ago, virtually anyone who had the skill to interpret data displays was suddenly empowered to design them as well, thus democratizing data design in the same way that laser printers democratized text design. In doing so, computer-generated displays initiated a critical shift in the rhetorical process by substantially altering reader expectations. Given the templates and wizards at their fingertips, professional communicators are now expected to visualize data, and because the technology offers the opportunity for rhetorical adaptation, readers expect that designers will seize that opportunity. To the extent that designers do, a well-executed display will enhance its clarity and its ethos, as readers recognize the designer’s ability to manipulate the technology for the reader’s benefit.

The effects of these technological tools, however, have been mixed, partly because not every professional communicator has the skills or motivation to master them, and partly because the defaults, templates, and wizards can actually undermine, rather than foster, rhetorical adaptation.
by impeding the designer’s inclination to consider optimal design solutions, as Brasseur shows in her study of fill patterns [23]. The array of design options in software like Microsoft Excel and PowerPoint creates the illusion of flexibility. With a few clicks, virtually any designer can completely transform the appearance of a graph—for example, by altering its genre, adding 3-D effects and colorful bars, and stretching the plot frame or filling it with a gradient texture. So marvelously malleable are these graphical effects—but for whom and to what end? Paradoxically, then, even as the technology for visualizing data has become more sophisticated, it does not necessarily engender rhetorically sensitive design. However, as we will see with the participatory rhetoric of digital data design, that dynamic will be redefined because adaptation is controlled partially by the readers themselves.

SOCIAL RHETORIC: WHAT READERS BRING TO THE INTERPRETIVE ACT

Clarity may lie in the eye of the beholder, as the rhetoric of adaptation tells us, but oftentimes it lies simultaneously in the eyes of many beholders. Because data design occurs in a particular historical and cultural context shared by large numbers of readers, its interpretation is a highly social act. Given that data design is very conventional and that readers learn how to interpret displays (and often different genres of displays) in a variety of ways, the social approach finds problematic the assumption that the reader’s perceptual system or situational variables dominate interpretation. Simply put, from the perspective of social rhetoric, nature may play a role in interpretation, but nurture really matters.

Examining data displays through the lens of social rhetoric complicates how designers deploy and readers interpret displays. Clarity remains an issue, but how it is achieved in the minds of readers becomes murkier, more variegated, and less visible. Analyzing a rhetorical situation tells us a great deal about what’s explicit and idiosyncratic about the interaction between designer and reader, but the outcome of that interaction also relies on what lies beneath the surface, what knowledge readers implicitly share that they bring to that interaction. From a social perspective, clarity depends on a process of learning and enculturation, and that perspective is particularly relevant to data displays, given that they are intrinsically artificial and conventional.

We are not born with the capacity to read charts and graphs, so we acquire this skill, to varying degrees, depending on the visual discourse communities we participate in, including those communities that coalesce around organizations, disciplines, and more broadly, nations and cultures [16]. Disciplinary knowledge is typically acquired through formal training. For example, sociologists learn how to read scatterplots and population distribution charts, economists learn how to read candle-stick graphs, geologists learn how to read topographical maps, nurses and doctors learn how to read cardiograms, and so on. Drawing on the work of sociologists such as LaTour and Woolgar, Brasseur observes that the data displays used by scientists are shaped by the collective social processes in which these forms originate [3]. The interpretive skills of members of a given discipline can be highly specialized and sophisticated—for example, a geologist who reads seismic diagrams to locate or predict an earthquake or a structural engineer who reads graphs about the strength of building materials to determine their suitability for various applications.

Organizations may also develop their own methods of data visualization (e.g., charts or diagrams displaying hourly energy use), or they may adapt existing forms to meet their own needs (e.g., line graphs with a consistent design in monthly production reports). In these ways, organizations provide a reliable framework for their members to interpret data, and clarity is enabled, or at least greatly enhanced, as members become accustomed to these practices.

So if we regard data displays as socially constructed conventions, clarity depends on readers’ experiences in disciplines and organizations but also as members of public discourse communities that enculturate readers in forms of data visualization through schools, popular culture, and news media. As a result, visual literacy varies across national boundaries (Tufte, for example, claims that displays are more complex in Japan than in the US [1, pp. 82–84]). The effects of a collective, national experience on visual literacy are illustrated by the series of statistical atlases of the United States, which first appeared in 1874 [24]. The atlases contained hundreds of displays of census data in a variety of forms, such as pie charts, population distribution charts, bar charts, percent charts, and many others. So novel were most of these forms to the public that the early atlases provided detailed explanations on how to interpret them; however, as Americans collectively
began accustomed to reading data displays, the explanations disappeared in later editions because the public had achieved a level of visual literacy that rendered the explanations unnecessary [25].

Technology has also influenced the social aspects of visual rhetoric. Popular data design software, like Microsoft Excel and PowerPoint, reproduces and reifies a small set of conventional genres, as shown in Fig. 5. Circumscribing the designer’s available design choices helps enculturate readers into a finite group of familiar forms and thereby reduces the reader’s interpretive stress. In this way, data design software both enhances and constrains visual literacy, ensuring that a wide range of readers become enculturated in certain conventional genres—from scatterplots to donut charts—while at the same time limiting data visualization to those forms. Of course, popular graphing software also has the rhetorical drawbacks of suppressing less well-known conventions and inhibiting the invention of new ones.

Interpreting data displays can also be filtered by the reader’s cultural background, though this factor might be less influential than for other forms of visual communication such as illustrations, symbols, or icons. Nevertheless, a color scheme for bars on a chart—say red and blue—might mean something quite different to an American reader than to a Chinese reader. A newspaper display with pictorial elements—say a line graph about oil inscribed in a 10-gallon Texas hat—might produce a similar interpretive divergence between American readers and Saudi readers. Cultural knowledge also includes aesthetics. Modernism profoundly affected the way we look at any designed artifact, including data displays. Fig. 3 certainly invokes the lean, efficient, high-contrast aesthetic of modernism, which for many contemporary readers may bolster its ethos; to other readers, however, its stark nakedness may make it look dated and passé.

The social aspect of conventions is further illuminated by theorists like Barton and Barton, who scrutinize the purposes data displays fulfill through the power that they embody [26], [27]. That power can be measured in part by whose interest those displays serve, how designers control what information they include or exclude in a display, and the “naturalizing” effect displays create when readers interpret them uncritically rather than as artificial constructs [27]. In these ways, readers collectively delegate power to data displays that may serve their own self-interests (e.g., by perpetuating the status quo), or they might unwittingly cede their own power by acquiescing to forms that they lack the pragmatic, intellectual, or cultural authority to resist.

Because readers’ encounters with data displays do not occur in a social or cultural vacuum, achieving clarity is a collective, though often invisible, effort. Readers are not naïve noble savages who gaze innocently; rather, they are members of discourse communities—large and small, public and specialized—that foster their interpretive skills. So when readers encounter a display, they bring with them experiences and expectations that define their sense of clarity.

THE RHETORIC OF PARTICIPATION: THE POSTMODERN FRONTIER OF DYNAMIC DISPLAYS

Visual rhetoric is being transformed by another wave of technological advances that has occurred with the proliferation of the World Wide Web. Digital data design, whereby the display is produced and interpreted entirely on a screen (and typically on the web), has opened up opportunities to invent novel designs and to widen the pool of conventional forms, including some from the past, and thereby socialize design by redefining our concept of visual literacy. Because many digital online displays are interactive, they allow readers (or users, given their redefined role) to adapt them to their varying needs and interests, fulfilling traditional rhetorical goals. Zappen claims that digital media, broadly construed, is transforming rhetoric through “self-expression, participation, and creative collaboration” [28, p. 321]. Although self-expression and creative collaboration are facets of the emerging rhetoric of digital data visualization,
active reader participation stands at the heart of this enterprise.

This shift from print to screen has enormous consequences for clarity because print limits the amount of data that can be visualized and reasonably processed in a single static display, whatever its genre. Interactive displays give readers access to complex data sets,¹ which enhance clarity through the close proximity of displays that would otherwise—in print form—be scattered across multiple surfaces. Interactive displays also enhance clarity by empowering readers to visualize data themselves by controlling the size, scope, and genre of the display. Conceptually, then, in the realm of interactive digital design, Fig. 3 is a mere surface chart that readers can metamorphose in a variety of ways, with a rich underlayer of possibilities waiting for the reader to discover.

Audience Adaptation Online data design shifts the user’s interpretive act from a passive to an active, participatory role. Clarity becomes a moving, self-adjusting target whereby users shape (and reshape) data displays according to their needs and preferences. In static print displays, the burden is placed entirely on the designer to adapt the display to the rhetorical situation by choosing the data set for the display, the genre in which the data appear (bar chart, line graph, scatterplot, etc.), and the design features of the display. Print limits the amount of data that can be visualized and that users can reasonably process in a single display. With interactive displays, on the other hand, users participate in adapting the display, which empowers them to visualize data on their own terms and which places them in a quasi-collaborative relationship with the designer. Moreover, interactive displays give users access to complex data sets by allowing them to mine them for information they want to visualize. For example, investment performance charts typically allow users to display data in different time zones, enabling users to visualize both short- and long-term performance. The investment performance charts in the TIAA-CREF website, for example, allow users to display data in several time zones, ranging from 30 days (top of Fig. 6) to 10 years (bottom of Fig. 6) [29]. In this instance, each of these displays yields a very different picture of the “Stock” fund, enabling users to visualize both short- and long-term performance.

Interactive displays with additional options can be found at the Bankrate.com website, where within the same plot frame users can graph interest rates for mortgages, both nationally and in a specific state of their choosing, as well as over several time spans [30]. Users can also choose the genre—line, scatter, area, or bar graph—so if one form of graph doesn’t provide adequate clarity or perceptual efficiency or meet the reader’s expectations, readers can transform the display into another genre, adapting the variables (e.g., time, place, genre of display) to their own informational needs and preferences. Unlike a static graph, readers can generate several views of the same data or, more importantly, mine and visualize more complex data sets by digging beneath the surface display. Because they place users in a position of authority by giving them control over how data is visualized, interactive displays are highly postmodern in their rhetorical and aesthetic orientation. Interactive

¹I would like to acknowledge my colleague Heike Hofmann for this observation.
displays are multilayered and flexible, and they trump modernist assumptions about the authoritative relationship between designer and user. In interactive displays, meaning lies beneath the surface, waiting for the user to discover—unlike modernism which promoted authoritative transparency and designer control (epitomized by Neurath’s high-contrast Isotype displays, where meaning was accessed explicitly on the surface). By contrast, the surface image of interactive displays belies its complexity, engendering a rhetoric of accessibility that enables readers to extract the information they need from complex data sets without being overwhelmed by them.

Visualizing Macro-/Micro-Levels Unlike a static graph, interactive displays enable users with different motivations and information needs to explore data at their own pace and level and, more importantly, mine and visualize complex data sets. By zooming in and out of the display, users can visualize what Tuftu calls the “macro” and “micro” views [32, pp. 37–51]. Users of the TIAA-CREF and Bankrate.com charts have this flexibility by being able to select a spectrum of displays ranging from the micro to the macro view. Of course, interactivity does not guarantee rhetorical adaptability to every situation. The TIAA-CREF and Bankrate.com charts are not detailed and precise enough for readers to make fine-grained comparisons, which limits their rhetorical flexibility. Investors and mortgage shoppers will likely find them helpful for long-term decision-making, but users who need more micro-level precision, like day traders and lenders, probably will not.

An imaginative graphing system that gives users a high level of flexibility to explore the data on both the macro- and micro-levels is NameVoyager, which uses data from the US Government to display the popularity of the top 1,000 baby names by gender since the 1880s, primarily so that prospective parents can make informed name choices [33]. Scrolling over an area/mountain graph (Fig. 7) shows the top 1,000 names of boys (blue) and girls (pink), with the thickness of the area representing the number of children per million with a particular name in a given year. For example, the names John and Mary have always been popular names. On the micro-level, users can click on a specific name or enter a name at the top of the display to generate a separate graph for a particular name.

As shown in Fig. 8, the name Elizabeth was more popular in the past than the present, though it experienced a resurgence in the 1980s. Scrolling over the graph reveals the ranking (from 1 to 1,000) of the name Elizabeth in any given era. Users can graph any names they wish, so long as the names appeared in the top 1,000 in a given year. In this way, individual users can continually tailor NameVoyager to their own goals.

Interactive maps are also a rich source of data that users can manipulate to explore data at various depths. For example, Fig. 9 shows the “USA National Gas Temperature Map,” which
displays gasoline prices across the US [34]. On the macro-level, users can view the entire country to see how prices compare nationally by county and state, or they can zoom in to the micro-level to compare prices, as shown in Fig. 10, which displays the Chicago metropolitan area. Additional zooming enables users to map gas prices, almost literally, in their own neighborhoods. The rhetorical impact of this display would seem to be highly volatile because its kairos—or timing in a given situation—would be marginal when oil prices are stable and consumers drive little, but its kairos would intensify when oil prices spike upward during summer vacations, holidays, and international crises. For internet displays like this one, some “kairotic” moments may be unpredictable.

A more complex and sophisticated example of an interactive map can be found in the New York Times 2004 Election Guide of the US, which voters can use to assess information about the candidates, the issues, and the voting patterns and preferences of the electorate [35]. Users can select from numerous political contests (President, Senate, House, Governors) and then click on states to reveal additional data about a given contest. This interactive map supplies users with a rich pool of data that they can fish in as deeply and broadly as they like. Although its kairotic moment has passed for voters, the interactive map narrates compelling political stories that users can now read retrospectively and perhaps educate themselves as the next national election draws near.

Of course, interactivity does not assure clarity, and no amount of interactivity can compensate for a poorly conceived design. Usability research in human–computer interaction can help us understand how readers respond to various aspects of interactive design—color, zooming, and the like—by identifying optimal display techniques. Research in this field is underpinned by a cognitive-perceptual approach and appears in journals like Human Factors and in Card, Mackinlay, and Schneiderman’s anthology Readings in Information Visualization [14]. Mirel critiques this kind of “object-oriented” research as too narrowly focused and she advocates, as a complement, usability studies that are situation-based and that account for the fuller rhetorical picture of a given user’s experience [36]. Short of such holistic studies, user feedback can be looped into the design process. For example, internet visitors are counted, monitored, and often profiled, and they can contact designers with their input so that displays can be revised.

**Animating Small Multiples** Map animations can reveal both space and time, enabling users to visualize time geographically in what Tufte calls “small multiples,” a series of miniature displays with the same plot frame and data variables [1, pp. 170–175]. However, instead of envisioning a handful of small multiples in a print display that is confined to a single visual field, animations can visualize dozens or even hundreds of small multiples that stream together to create a continuous display, revealing a much richer macro-level view of the data, as conceptualized in Fig. 11. Clarity is enhanced by the proximity...
Fig. 11. Tufte’s concept of “small multiples” [1, pp. 170–175], transferred from print to digital animation.

Fig. 12. Three of the small multiples in an online animated display of the Annual Population Growth in the Washington, D.C./Baltimore area from 1792 to 1992 [37]. Reprinted with permission of the U.S. Geological Survey and of Keith Clarke, Geography Department, University of California, Santa Barbara.

and speed with which users can process multiple displays in rapid succession.

For example, a website of the US Geological Survey and the University of California at Santa Barbara contains an animated map that shows the population growth of the Washington DC/Baltimore region over two centuries, with the density of the population data points progressively increasing as the story unfolds each year from 1792 to 1992 [37]. Along the way, users can stop the animation to get a micro-level view of the population data in any given year. Fig. 12 shows images for three of the years—1932, 1962, and 1992—that are displayed in the animation. In just seconds or minutes, digital animations like these can tell highly accessible, macro-level narratives.

Animated data displays demonstrate what Blair calls the “evocative power” of the visual, with many images appearing in rapid succession to create arguments [38, p. 51]. An animation with considerable “evocative power” appears on the Center for Remote Sensing of Ice Sheets (CReSIS) website, which displays animated maps visualizing the effects of rising sea waters around the globe [39]. By visually narrating the incremental impact of rising waters on low-lying areas—such as Florida, the Netherlands, and Bangladesh—the maps create a stunningly powerful argument that global warming poses extreme danger to coastal communities. Users can select regional maps, which enlarge the images considerably, to focus on anywhere in the world they wish. Fig. 13 shows an image from the animated map for southeastern United States at six meters, which would undoubtedly alarm anyone living in southern Florida or along the Atlantic or Gulf coast—or those with friends, relatives, or financial, political, or environmental interests in these areas. The micro- and macro-level clarity of sea waters streaming onto the mainland will certainly impel users to think seriously about the future.

The power of the animated narrative in multimodal form can be experienced in a dramatic map of Iraq that visualizes the daily distribution of casualties of coalition forces since the start of the war, plotting circular red bursts accompanied by audio gunfire to locate the place and magnitude [40]. Piecemeal war stories are thus threaded together into a coherent and compelling narrative that graphically and audibly represents the grim data. By vividly visualizing the heavy toll on human life, the animated map generates pathos appeals that arouse the reader’s emotions about the war and thereby, as Dragga and Voss advocate, “humanizing” the data [41].
Impact on Conventions Although the interactive nature of digital display makes the user’s experience with data visualization highly individual, this medium also has social consequences by proliferating existing genres like bar charts, lines graphs, and the like—conventions that lay readers come to understand and expect and that reinforce existing visual literacy standards. Because digital design is such a potent medium, it provides an avenue for inventing new forms of display, which because they appear online gives them the possibility, if not the prospect, of becoming conventions. Digital data design also sometimes develops its own graphing communities where conventions emerge and are shared among specialized groups of users [3].

Digital data design has the ability to reinvent displays that have appeared in the past but which were unsustainable because they were too complex or too labor-intensive to reproduce. Just like images and texts that have been revived and widely disseminated through digital technology, data displays can also reappear electronically. Mosaics, for example, are a type of area chart that shows several variables in the same display through rectilinear shapes that are proportionate to the data [13], [42]. Mosaics appeared as early as the Statistical Atlas of the United States Based on the Results of the Ninth Census 1870 [24]. The image on the top in Fig. 14 shows a page of mosaics from this statistical atlas that visualize the population of each state, segmented into several groups. The image on the bottom in Fig. 14 shows an enlargement of the mosaic for Kentucky, with labels explaining each category of population.

Clever in design and rich in information, mosaics failed to become a popular genre, probably because they were too time-consuming to create and possibly because they were too demanding interpretively for readers unaccustomed to reading displays of any kind and drawn initially to simpler forms of representing data. The production problem can be solved by an innovative piece of software—Manet: Missions Are Now Equally Treated—which enables designers to create a variety of plots, including mosaics [43]. Fig. 15 shows a mosaic created in Manet that displays hypothetical data about students in a university, categorized by college (Business, Engineering, or LAS), citizenship (US or international), sex, and rank (graduate or undergraduate), with graduate students in each category shown in red and undergraduates in gray. Because Manet is interactive, readers can select and reconfigure the variables to visualize a variety of plots [42], [43] (for additional examples of mosaics, see [13, pp. 20–23, 49], [44, pp. 382–385], [45, pp. 396–398]). To the uninitiated, learning to read mosaics can still be daunting in the 21st century, and thus clarity is far from assured. However, by actively engaging users in visualizing data, Manet diminishes the interpretive problem of learning a novel display system. User participation itself fosters clarity, as users learn by doing.

Many other graphical inventions that have languished in obscurity wait to be rediscovered digitally. The 19th century in Europe and America...
(the golden era of statistical graphics) is a particularly fertile place to recover imaginative designs. Epic charts that correlate numerous variables on the same plot frame, wind roses that display weather and disease, [16, pp. 46, 113, 138-143], and bi-polar population distribution charts [25, pp. 219–221]—these are among the many forms that could be revived interactively in a digital age if designers had the curiosity and the technical skills to rediscover them [16]. Today their sustainability has a greater chance because technology makes them easier to reproduce and because readers participate in shaping their interactive successors.

Collaboration/Self-Expression Sometimes online interactive data displays not only ask readers to play an active role in shaping the display but also invite them to collaborate in the design process. The website “Information Aesthetics” provides a space where designers from around the world display innovative and experimental data designs in a wide variety of media—from paper and electronic to sculpture and clothes—and where readers post critiques, fostering a continuous conversation about data visualization [46]. In many of the examples on the Information Aesthetics website, data design functions as much as a vehicle for self-expression as for understanding, analysis, or decision-making. The rhetoric of this design is clearly expressive and even poetic, and functional usability is subordinate to the designer’s desire to create a novel and aesthetically engaging artifact that elicits a subjective response from the user. The rhetoric of this design is highly expressive and even poetic, and clarity here is subsumed into an aesthetic experience.

CONCLUSION
Far from being simple and straightforward, clarity in data design is a multifaceted and sometimes ambiguous and elusive concept. Perceptual principles, cognitive models, aesthetics, the socializing influence of learning and experience, the exigencies of the rhetorical situation, as well as other rhetorical factors such as pathos elements that generate reader interest—all contribute something important to our understanding of clarity. As readers’ encounters with data displays shift from print to screen, clarity is again redefined, partly because readers are often allowed control over how data are visualized. Interactive displays also afford readers access to complex data sets without having to process all of the data in the same display simultaneously, thus reducing their interpretive stress.

Over the next 50 years, technology will likely continue to fuel the revolution in data design, perhaps even surpassing the discoveries and innovations of the 19th century. The innovations in digital data design will afford readers even greater flexibility to visualize data, further democratizing visual access to information and intensifying the rhetoric of participation. As a result, readers will likely have a much larger array of internet-based conventional forms to interpret, as well as encounter many novel forms, all of which will compete for their attention. As these developments further redefine and expand visual literacy, the rhetoric of clarity will likely become even more fluid and contingent.

ACKNOWLEDGMENT
I wish to thank two colleagues from the Iowa State University Department of Statistics: Professor Heike Hoffman for teaching me to use the Manet software, and both she and Professor Dianne Cook for enabling me to envision new ways of displaying data. I also wish to thank Marty Teply, Iowa State University, for technical support and Professor John Kostelnick, Haskell Indian Nations University, for directing me to websites with animated maps. Finally, I appreciate the feedback.
from two anonymous reviewers and the editor that helped me focus and revise this article.

REFERENCES


Charles Kostelnick is a Professor of English at Iowa State University, Ames, IA, where he has taught an undergraduate and a graduate course in visual communication. He was the co-author (with Michael Hassett) of Shaping Information: The Visual Rhetoric of Conventions and co-author (with David Roberts) of Designing Visual Language: Strategies for Professional Communicators.
Day 4:
Ideologically Laden Data —
Propaganda and Polemics

Morning 4:
Becoming Acquainted with the Mores of Data Collection and Visualization
9:00am to 12pm (3 hours)

- Activity: Creating Visualizations from Census Data

Afternoon 4:
Discussion and Project Time
1:30pm to 4:00pm (2.5 hours)

Two activity options
Students will do prepared activities on the day’s theme, or they can work on their own datasets / projects
- Activity: Students will create two visualizations of the same data: one propagandized and one ethical. These visualizations can be created using provided sample data or their own treacherous data.
- Activity: Students will present either their propagandized or ethical visualization to the rest of the class, and fellow students will identify whether the visualization can be trusted based on what they have learned in the class so far.
DAY 4 SLIDES

[add slides here]
Counting Race

Problematising our Categories to create more Ethical Visualizations
☐ WHAT
☐ IS
☑ YOUR
☐ RACE?

The Census and Our Flawed Efforts to Classify Americans

Kenneth Prewitt
Three-fifths compromise (1787)
Americanus: reddish, choleric, and erect; hair black, straight, thick; wide nostrils, scanty beard; obstinate, merry, free; paints himself with fine red lines; regulated by custom.

Asiaticus: sallow, melancholy, stiff; hair black; dark eyes; severe, haughty, avaricious; covered with loose garments, ruled by opinion.

Africanus: black, phlegmatic, relaxed; hair black, frizzled; skin silky; nose flat; lips tumid; women without shame, they lactate profusely; crafty, indolent, negligent; anoints himself with grease; governed by caprice.

Europeanus: white, sanguine, muscular; hair long, flowing; eyes blue; gentle, acute, inventive; covers himself with close vestments; governed by laws.¹

Carl Linnaeus, application of Systema Naturae to humans, 18th century
Josiah Nott, *Types of Mankind*, 1854
**EXISTING FACTS OF HUMAN ASCENT**

Racism, Public Schooling, and the Entrenchment of White Supremacy

A CRITICAL RACE ETHNOGRAPHY

Sabina E. Vaught
DAY 4 READINGS

Gini Coefficient - Income Inequality

A popular measure of inequality is the Gini coefficient, which ranges from 0 (perfect equality) to 1 (perfect inequality), but is typically in the range of 0.3-0.5 for per capita expenditures. The Gini is derived from the Lorenz curve, which sorts the population from poorest to richest, and shows the cumulative proportion of the population on the horizontal axis and the cumulative proportion of income on the vertical axis. While the Gini coefficient has many desirable properties – mean independence, population size independence, symmetry, and Pigou-Dalton Transfer sensitivity – it cannot easily be decomposed to show the sources of inequality.

The Gini coefficient is a relatively reliable means of quantifying income inequality. Consider the criteria that make a good measure of income inequality, namely:

- **Mean independence.** This means that if all incomes were doubled, the measure would not change. The Gini satisfies this.

- **Population size independence.** If the population were to change, the measure of inequality should not change, ceteris paribus. The Gini satisfies this too.

- **Symmetry.** If you and I swap incomes, there should be no change in the measure of inequality. The Gini satisfies this.

- **Pigou-Dalton Transfer sensitivity.** Under this criterion, the transfer of income from rich to poor reduces measured inequality. The Gini satisfies this too.
**Limitation:** with the Gini coefficient, inequality may not be broken down by population groups or income sources or in other dimensions.
[add readings here]
Chapter 3

MAP GENERALIZATION:
LITTLE WHITE LIES AND LOTS OF THEM

A good map tells a multitude of little white lies; it suppresses truth to help the user see what needs to be seen. Reality is three-dimensional, rich in detail, and far too factual to allow a complete yet uncluttered two-dimensional graphic scale model. Indeed, a map that did not generalize would be useless. But the value of a map depends on how well its generalized geometry and generalized content reflect a chosen aspect of reality.

Geometry

Clarity demands geometric generalization because map symbols usually occupy proportionately more space on the map than the features they represent occupy on the ground. For instance, a line 1/50 inch wide representing a road on a 1:100,000-scale map is the graphic equivalent of a corridor 167 feet wide. If a road’s actual right-of-way was only 40 feet wide, say, a 1/50-inch-wide line symbol would claim excess territory at scales smaller than 1:24,000. At 1:100,000, this road symbol would crowd out sidewalks, houses, lesser roads, and other features. And at still smaller scales more important features might eliminate the road itself. These more important features could include national, state, or county boundaries, which have no width whatever on the ground.

Point, line, and area symbols require different kinds of generalization. For instance, cartographers recognize the five fundamental processes of geometric line generalization described in figure 3.1. First, of course, is the selection of complete features for the map. Selection is a positive term that implies the suppression, or nonselection, of most features. Ideally the map author approaches selection with goals to be satisfied by...
Chapter Three / 26

FIGURE 3.1. Elementary geometric operations in the generalization of line features.

a well-chosen subset of all possible features that might be mapped and by map symbols chosen to distinguish unlike features and provide a sense of graphic hierarchy. Features selected to support the specific theme for the map usually require more prominent symbols than background features, chosen to give a geographic frame of reference. Selecting background details that are effective in relating new informa-
tion on the map to the viewer’s geographic savvy and existing “mental map” often requires more insight and attention than selecting the map’s main features. In the holistic process of planning a map, feature selection is the prime link between generalization and overall design.

The four remaining generalization processes in figure 3.1 alter the appearance and spatial position of linear map features represented by a series of points stored in the computer as a list of two-dimensional (X, Y) coordinates. Although the growing use of computers to generalize maps led to the isolation of these four generalization operations, traditional cartographers perform essentially the same operations by hand but with less structure, less formal awareness, and less consistency. Simplification, which reduces detail and angularity by eliminating points from the list, is particularly useful if excessive detail was “captured” in developing a cartographic data file, or if data developed for display at one scale are to be displayed at a smaller scale. Displacement avoids graphic interference by shifting apart features that otherwise would overlap or coalesce. A substantial reduction in scale, say, from 1:25,000 to 1:1,000,000, usually results in an incomprehensibly congested collection of map symbols that calls for eliminating some features and displacing others. Smoothing, which also diminishes detail and angularity, might displace some points and add others to the list. A prime objective of smoothing is to avoid a series of abruptly joined straight line segments. Enhancement adds detail to give map symbols a more realistic appearance. Lines representing streams, for instance, might be given typical meander loops, whereas shorelines might be made to look more coastlike. Enhanced map symbols are more readily interpreted as well as more aesthetic.

Point features and map labels require a somewhat different set of generalization operators. Figure 3.2 illustrates that, as with linear features, selection and displacement avoid graphic interference when too many close symbols might overlap or coalesce. When displacement moves a label ambiguously far from the feature it names, graphic association with a tie line or a numeric code might be needed to link the label with its symbol. Abbreviation is another strategy for generalizing labels on congested small-scale maps. Aggregation is useful where many equivalent features might overwhelm the map if ac-
FIGURE 3.2. Elementary geometric operations in the generalization of point features and map labels.

corded separate symbols. In assigning a single symbol to several point features, as when one dot represents twenty reported tornadoes, aggregation usually requires the symbol either to portray the "center of mass" of the individual symbols it replaces or to reflect the largest of several discrete clusters.

Where scale reduction is severe, as from 1:100,000 to 1:20,000,000, area conversion is useful for shifting the map viewer’s attention from individual occurrences of equivalent features to zones of relative concentration. For example, instead of showing individual tornadoes, the map might define a belt in which tornadoes are comparatively common. In highlighting zones of concentration or higher density, area conversion replaces all point symbols with one or more area symbols. Several density levels, perhaps labeled "severe," "moderate," and "rare," might provide a richer, less generalized geographic pattern.

Area features, as figure 3.3 demonstrates, require the largest set of generalization operators because area boundaries are
subject to aggregation and point conversion and all five elements of line generalization as well as to several operators unique to areas. Selection is particularly important when area features must share the map with numerous linear and point features. A standardized minimum mapping size can direct the selection of area features and promote consistency among
the numerous sheets of a map series. For example, 1:24,000-scale topographic maps exclude woodlands smaller than one acre unless important as landmarks or shelterbelts. Soil scientists use a less precise but equally pragmatic size threshold—the head of a pencil—to eliminate tiny, insignificant areas on soils maps.

Aggregation might override selection when a patch otherwise too small to include is either combined with one or more small, similar areas nearby or merged into a larger neighbor. On soils maps and land-use maps, which assign all land to some category, aggregation of two close but separated area features might require the dissolution or segmentation of the intervening area. A land-use map might, for example, show transportation land only for railroad yards, highway interchanges, and service areas where the right-of-way satisfies a minimum-width threshold. Simplification, displacement, smoothing, and enhancement are needed not only to refine the level of detail and to avoid graphic interference between area boundaries and other line symbols, but also to reconstruct boundaries disrupted by aggregation and segmentation.

Generalization often accommodates a substantial reduction in scale by converting area features to linear or point features. Line conversion is common on small-scale reference maps that represent all but the widest rivers with a single readily recognized line symbol of uniform width. Highway maps also help the map user by focusing not on width of right-of-way but on connectivity and orientation. In treating more compact area features as point locations, point conversion highlights large, sprawling cities such as London and Los Angeles on small-scale atlas maps and focuses the traveler’s attention on highway interchanges on intermediate-scale road maps. Linear and point conversion are often necessary because an area symbol at scale would be too tiny or too thin for reliable and efficient visual identification.

Comparing two or more maps showing the same area at substantially different scales is a good way to appreciate the need for geometric generalization. Consider, for instance, the two maps in figure 3.4. The rectangles represent the same area extracted from maps published at scales of 1:24,000 and 1:250,000; enlargement of the small-scale excerpt to roughly the same size as its more detailed counterpart reveals the need for
considerable generalization at 1:250,000. The substantially fewer features shown at 1:250,000 demonstrate how feature selection helps the mapmaker avoid clutter. Note that the smaller-scale map omits most of the streets, all labels in this area, all individual buildings, and the island in the middle of the river. The railroad and highway that cross the river are smoother and farther apart, allowing space for the bridge symbols added at 1:250,000. Because the 1:24,000-scale map in a sense portrays the same area in a space over a hundred times larger, it can show many more features in much greater detail.

How precisely are symbols positioned on maps? The U.S. Office of Management and Budget addresses this concern with the National Map Accuracy Standards, honored by the U.S. Geological Survey and other federal mapping agencies. To receive the endorsement "This map complies with the Nation-
al Map Accuracy Standards," a map at a scale of 1:20,000 or smaller must be checked for symbols that deviate from their correct positions by more than 1/50 inch. This tolerance reflects the limitations of surveying and mapping equipment and human hand-eye coordination. Yet only 90 percent of the points tested must meet the tolerance, and the 10 percent that don’t can deviate substantially from their correct positions. Whether a failing point deviates from its true position by 2/50 inch or 20/50 inch doesn’t matter—if 90 percent of the points checked meet the tolerance, the map sheet passes.

The National Map Accuracy Standards tolerate geometric generalization. Checkers test only “well-defined points” that are readily identified on the ground or on aerial photographs, easily plotted on a map, and conveniently checked for horizontal accuracy; these include survey markers, roads and railway intersections, corners of large buildings, and centers of small buildings. Guidelines encourage checkers to ignore features that might have been displaced to avoid overlap or to provide a minimum clearance between symbols exaggerated in size to ensure visibility. In areas where features are clustered, maps tend to be less accurate than in more open areas. Thus Pennsylvania villages, with comparatively narrow streets and no front yards, would yield less accurate maps than, say, Colorado villages, with wide streets, spacious front yards, and big lots. But as long as 90 percent of a sample of well-defined points not needing displacement meet the tolerance, the map sheet passes.

Maps that meet the standards show only planimetric distance, that is, distance measured in a plane. As figure 3.5 shows, a planimetric map compresses the three-dimensional land surface onto a two-dimensional sheet by projecting each point perpendicularly onto a horizontal plane. For two points at different elevations, the map distance between their “planimetrically accurate” positions underestimates both overland distance across the land surface and straight-line distance in three dimensions. Yet this portrayal of planimetric distance is a geometric generalization essential for large-scale flat maps.

The user should be wary, though, of the caveat “approximately positioned” or the warning “This map may not meet the National Map Accuracy Standards.” In most cases such maps have been compiled from unrectified aerial photographs,
FIGURE 3.5. Planimetric map generalizes distance by the perpendicular projection of all positions onto a horizontal plane.

FIGURE 3.6. A vertical aerial photograph (and any map with symbols traced directly from an air photo) is a perspective view with points displaced radially from their "accurate" planimetric positions.

on which horizontal error tends to be particularly great for rugged, hilly areas. Figure 3.6 shows the difference between the air photo’s perspective view of the terrain and the planimetric map’s representation of distances in a horizontal plane. Because lines of sight converge through the camera’s lens, the air photo displaces most points on the land surface from their planimetric positions. Note that displacement is radially outward from the center of the photo, that displacement is greater for points well above the horizontal plane than for lower points, and that displacement tends to be greater near the edges than near the center. Cartographers call this effect...
"radial displacement due to relief," or simply relief displacement. An exception is the orthophoto, an air-photo image electronically stretched to remove relief displacement. An orthophotomap, produced from orthophotos, is a planimetrically accurate photo-image map.

For some maps, though, geometric accuracy is less important than linkages, adjacency, and relative position. Among the more effective highly generalized maps are the linear cartograms portraying subway and rapid transit systems. As in figure 3.7, scale is relatively large for the inner city, where the routes converge and connect; stops in the central business district might be only four or five blocks apart, and a larger scale is needed here to accommodate more route lines and station names. In contrast, toward the fringes of the city, where stations are perhaps a mile or more apart, scale can be smaller because mapped features are less dense. Contrasting colors usually differentiate the various lines; the Washington, D.C., Metro system, in fact, calls its routes the Blue Line, the
Red Line, and so forth, to enhance the effectiveness of its map. By sacrificing geometric accuracy, these schematic maps are particularly efficient in addressing the subway rider’s basic questions: Where am I on the system? Where is my destination? Do I need to change trains? If so, where and to what line? In which direction do I need to go? What is the name of the station at the end of the line? How many stops do I ride before I get off? Function dictates form, and a map more “accurate” in the usual sense would not work as well.

Content
As geometric generalization seeks graphic clarity by avoiding overlapping symbols, content generalization promotes clarity of purpose or meaning by filtering out details irrelevant to the map’s function or theme. Content generalization has only two essential elements, selection and classification. Selection, which serves geometric generalization by suppressing some information, promotes content generalization by choosing only relevant features. Classification, in contrast, makes the map helpfully informative as well as usable by recognizing similarities among the features chosen so that a single type of symbol can represent a group of similar features. Although all map features are in some sense unique, usually each feature cannot have a unique symbol. Even though some maps approach uniqueness by naming individual streets or numbering lots, these maps also use very few types of line symbols, to emphasize similarities among roads and property boundaries as groups. Indeed, the graphic vocabulary of most maps is limited to a small set of standardized, contrasting symbols.

Occasionally the “template effect” of standardized symbols will misinform the map user by grouping functionally different features. Standard symbols, designed for ready, unambiguous recognition and proportioned for a particular scale, are common in cartography and promote efficiency in both map production and map use. Traditional cartographers use plastic drawing templates to trace in ink the outlines of highway shields and other symbols not easily rendered freehand. Drafters can cut area and point symbols from printed sheets and stick them onto the map and can apply dashed, dotted, or parallel lines from rolls of specially printed flexible tape. Elec-
Electronic publishing systems allow the mapmaker not only to choose from a menu of point, line, and area symbols provided with the software but also to design and store new forms, readily duplicated and added where needed. Consistent symbols also benefit users of the U.S. Geological Survey’s series of thousands of large-scale topographic map sheets, all sharing a single graphic vocabulary. On highway maps, the key (or “legend”) usually presents the complete set of symbols so that while examining the map, at least, the reader encounters no surprises. Difficulties arise, though, when a standard symbol must represent functionally dissimilar elements. Although a small typeset annotation next to the feature sometimes flags an important exception, for instance, a section of highway “under construction,” mapmakers frequently omit useful warnings.

Generalized highway interchanges are a prime example of how information obscured by the template effect can mislead or inconvenience a trusting map user. The left panel of figure 3.8 is a detailed view of the interchange near Rochester, New York, between highways 104 and 590, as portrayed at 1:9,600 on a state transportation department map. Note that a motorist traveling from the east (that is, from the right) on N.Y. 104 cannot easily turn north (toward the top of the map) onto N.Y. 590. The upper right portion of the left-hand map shows that the necessary connecting lanes from N.Y. 104 were started but not completed. In contrast, the right panel shows how various commercial map publishers portray this interchange on their small-scale statewide highway maps. Two diamond-shaped interchange symbols suggest separate and equivalent connections with the eastward and westward portions of N.Y. 104. Yet the large-scale map clearly indicates that a driver expecting an easy connection from N.Y. 104 westbound onto N.Y. 590 northbound must travel to the next exit west or south and then double back. Until the road builders complete their planned connecting lanes, such discrepancies between reality and art will frustrate motorists who assume all little diamonds represent full interchanges.

Effective classification and selection often depend on a mixture of informed intuition and a good working definition. This is particularly true for geologic maps and soils maps, commonly prepared by several field scientists working in widely
separated places. A detailed description is necessary if two people mapping areas a hundred miles apart must identify and draw boundaries for different parts of the same feature. These descriptions should also address the mapping category's internal homogeneity and the sharpness of its "contacts" with neighboring units. In soils mapping, for instance, small patches of soil B might lie within an area labeled as soil A. This practice is accepted because these enclaves of soil B are too small to be shown separately, and because the soil scientist cannot be aware of all such enclaves. Soil mapping, after all, is slow, tedious work that requires taking samples below the surface with a drill or auger and occasionally digging a pit to examine the soil's vertical profile. Map accuracy thus depends upon the field scientist's understanding of the effects of terrain and geology (if known) on soil development as well as on expertise in selecting sample points and intuition in plotting boundaries.
That crisp, definitive lines on soils maps mark inherently fuzzy boundaries is unfortunate. More appalling, though, is the uncritical use in computerized geographic information systems of soil boundaries plotted on “unrectified” aerial photos subject to the relief-displacement error described in figure 3.6. Like quoting a public figure out of context, extracting soils data from a photomap invites misinterpretation. When placed in a database with more precise information, these data readily acquire a false aura of accuracy.

Computers generally play a positive role in map analysis and map display, the GIGO effect (garbage in, garbage out) notwithstanding. Particularly promising is the ability of computers to generalize the geometry and content of maps so that one or two geographic databases might support a broad range of display scales. Large-scale maps presenting a detailed portrayal of a small area could exploit the richness of the data, whereas computer-generalized smaller-scale displays could present a smaller selection of available features, suitably displaced to avoid graphic interference. Both the content and scale of the map can be tailored to the particular needs of individual users.

Computer-generalized maps of land use and land cover illustrate how a single database can yield radically different cartographic pictures of a landscape. The three maps in figure 3.9 show a rectangular region of approximately 700 mi² (1,800 km²) that includes the city of Harrisburg, Pennsylvania, above and slightly to the right of center. A computer program generalized these maps from a large, more detailed database that represents much smaller patches of land and describes land cover with a more refined set of categories. The generalization program used different sets of weights or priorities to produce the three patterns in figure 3.9. The map at the upper left differs from the other two maps because the computer was told to emphasize urban and built-up land. This map makes some small built-up areas more visible by reducing the size of area symbols representing other land covers. In contrast, the map at the upper right reflects a high visual preference for agricultural land. A more complex set of criteria guided generalization for the display at the lower left: forest land is dominant overall, but urban land dominates agricultural land. In addition, for this lower map the computer dissolved water
FIGURE 3.9. Land-use and land-cover maps generalized by computer from more detailed data according to three different sets of display priorities. Areas, which were discontinuous because of variations in the width of the river. These differences in emphasis might meet the respective needs and biases of demographers, agronomists, and foresters.

Generalized maps almost always reflect judgments about the relative importance of mappable features and details. The systematic bias demonstrated by these generalized land-cover maps is not exclusive to computer-generated maps; manual cartographers have similar goals and biases, however vaguely defined and unevenly applied. Through the consistent application of explicit specifications, the computer offers the possibility of a better map. Yet whether the map's title or description reveals these biases is an important clue to the integrity of the mapmaker or publisher. Automated mapping allows ex-
perimentation with different sets of priorities. Hence computer generalization should make the cartographer more aware of choices, values, and biases. But just because a useful and appropriate tool is available does not mean the mapmaker will use it. Indeed, laziness and lack of curiosity all too often are the most important source of bias.

The choropleth map (introduced as the right-hand elements of figs. 2.13 and 2.14) is perhaps the prime example of this bias by default. Choropleth maps portray geographic patterns for regions composed of areal units such as states, counties, and voting precincts. Usually two to six graytone symbols, on a scale from light to dark, represent two to six nonoverlapping categories for an intensity index such as population density or the percentage of the adult population voting in the last election. The breaks between these categories can markedly affect the mapped pattern, and the cautious map author tests the effects of different sets of class breaks. Mapping software can unwittingly encourage laziness by presenting a map based upon a "default" classification scheme that might, for instance, divide the range of data values into five equal intervals. As a marketing strategy, the software developer uses such default specifications to make the product more attractive by helping the first-time or prospective user experience success. Too commonly, though, the naive or noncritical user accepts this arbitrary display as the standard solution, not merely as a starting point, and ignores the invitation of the program’s pull-down menus to explore other approaches to data classification.

Different sets of categories can lead to radically different interpretations. The two maps in figure 3.10, for example, offer very different impressions of the spatial pattern of homes in the northeastern United States still lacking telephones in 1960. Both maps have three classes, portrayed with a graded sequence of graytone area symbols that imply "low," "medium," and "high" rates of phonelessness. Both sets of categories use round-number breaks, which mapmakers for some mysterious reason tend to favor. The map at the left shows a single state, Virginia, in its high, most deficient class, and a single state, Connecticut, in its low, most well-connected class. The casual viewer might attribute these extremes to Virginia’s higher proportion of disadvantaged blacks and to Connecticut's af-
Figures 3.10 and 3.11 illustrate the effect of different class breaks on choropleth maps. The map on the left, with class breaks set at 10%, 20%, and 30%, and above 30%, shows a concentration of states with high percentages of occupied housing units lacking a telephone in the northeastern United States. The map on the right, with class breaks at 15%, 25%, and above 25%, presents a different distribution, with a higher percentage of states clustered in the northeastern region.

Fluent suburbs and regard the remaining states as homogeneously "average." In contrast, the map at the right portrays a more balanced distribution of states among the three groups and suggests a different interpretation. Both states in the high category have substantial dispersed rural populations, and all four in the low category are highly urban and industrialized.
Moreover, a smaller middle group suggests less overall homogeneity.

Machiavellian bias can easily manipulate the message of a choropleth map. Figure 3.11, for example, presents two cartographic treatments with substantially different political interpretations. The map on the left uses rounded breaks at 10 percent and 15 percent, forcing most states into its high, poorly connected category and suggesting a Northeast with generally poor communications. Perhaps the government is ineffective in regulating a gouging telecommunications industry or in eradicating poverty. Its counterpart on the right uses rounded breaks at 20 percent and 30 percent to paint a rosier picture, with only one state in the high group and eight in the low, well-served category. Perhaps government regulation is effective, industry benign, and poverty rare.

The four maps in figures 3.10 and 3.11 hold two lessons for the skeptical map reader. First, a single choropleth map presents only one of many possible views of a geographic variable. And second, the white lies of map generalization might also mask the real lies of the political propagandist.

Intuition and Ethics in Map Generalization

Small-scale generalized maps often are authored views of a landscape or a set of spatial data. Like the author of any scholarly work or artistic creation based on reality, the conscientious map author not only examines a variety of sources but relies on extensive experience with the information or region portrayed. Intuition and induction guide the choice of features, graphic hierarchy, and abstraction of detail. The map is as it is because the map author "knows" how it should look. This knowledge, of course, might be faulty, or the resulting graphic interpretation might differ significantly from that of another competent observer. As is often the case, two views might both be valid.
Day 5:
Defining Position Statements —
Taming Treacherous Data

Morning 5:
Framing your own visualization practise through position statements

9:00am to 12:00pm (3 hours)

- **Reading and discussion:** Kirk St. Amant. 2016. "Introduction to the special issue: Cultural considerations for communication design: integrating ideas of culture, communication, and context into user experience design". *Communication Design Quarterly Review*. 4, no. 1: 6-22.
- **Activity:** Writing 500-1000 word position statements on your own visualization practise, based on socio-cultural considerations of their subject and audience.
DAY 5 SLIDES

[add slides here]
Chart Suggestions—A Thought-Starter

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?

Relationship

Composition

Distribution

Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart

Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories

One Variable per Item

Among Items

Over Time

Comparison

What would you like to show?
DAY 5 READINGS

[add readings here]
Introduction to the special issue
Cultural Considerations for Communication Design: Integrating Ideas of Culture, Communication, and Context into User Experience Design

Kirk St. Amant
East Carolina University
kirk.stamant@gmail.com

Culture and the Context of Communication Design

Culture can be difficult to define, yet it is central to almost everything humans do. Culture shapes how individuals view the world – what they consider right and wrong or appropriate and inappropriate – and often provides the lens through which they perceive communication and create messages (Sardi & Flammia, 2011; Varner & Beamer, 2015). As such, culture can be one of the most important aspects communication designers need to consider when developing materials for an audience – any audience. When extended to broader intercultural or international contexts, the need to understand how culture affects expectations and perceptions becomes even more acute. For this reason, the more communication designers know about researching, considering, and addressing cultural communication expectations, the more effectively they can develop materials that meet the information seeking and usage needs of a greater global audience.

But where to begin?

For starters, none of us is born with a culture. Rather, we start our lives as a blank slate, and our culture is something we acquire over time via living as a part of it, observing actions that take place in it, and learning from these observations (Berry, Poortinga,
Breugelmans, Chasiotis, & Sam, 2011). This process of learning our own culture is known as *enculturation*, and it is based on exposure over time. That is, the more we see a particular behavior that the members of our culture reward, the more likely we are to adopt the perspective that behavior is correct and seek to emulate it. The more often we see the members of our culture being chastised for a given behavior, the more likely we are to view it as negative and to avoid that behavior. This connection between exposure and patterns of emulation or avoidance undergirds everything from how individuals dress to how they move to how they design to how they communicate.

So why is all of this important to communication designers? Because it means we cannot assume what is an acceptable way to communicate (e.g., to draft texts, craft visuals, or design interfaces) in one culture is universal and works with all others. Superficially, this idea—that no cultural universal exists for communication—seems self obvious. At a deeper level, however, it has important implications. And it is these deeper-seated concepts that affects how communication designers need to approach the process of developing informational and instructional materials for audiences from other cultures.

**The Complex Dynamics of Culture and Communication**

When it comes to creating messages or materials for individuals from other cultures, communication designers need to consider a range of factors. Some of these factors can appear self evident. Others can be more nuanced. In many cases, they are interconnected, and a failure to address one or more of them effectively can affect how the members of a given culture respond to and use a particular item – be it a document or an interface. For these reasons, it is important for communication designers to familiarize themselves with the various elements they need to consider when crafting different informational or instructional items for individuals from another culture.
Language and Rhetoric

To begin, culture and language are often interconnected. This connection, however, involves more than just language itself. Rather, it is often a matter of how the members of a given culture use a language in particular ways to convey information. In essence, just because individuals can say something a given way in a particular language does not mean they should say it that way or that the members of the related culture consider it expected or appropriate to convey ideas in that manner. The notion of rhetoric – or how one structures or presents ideas in given language – is therefore often as important as the vocabulary and syntax of the language itself (see, for example, Campbell, 1998 and St.Amant, 2006).

These rhetorical expectations – or how to say something appropriately, effectively, or credibly (i.e., worthy of attention and consideration) – can affect everything from how to construct a sentence to what kind(s) of information should appear in certain types of documents to if a particular genre for conveying information is used at all (Driskill, 1996; Tebeaux, 1999; Woolever, 2001). What makes these cultural rhetorical expectations particularly challenging is they are a construction of the culture using a given language and not necessarily an aspect of the language itself. This factor means cultural groups that speak the same language can have different expectations of what are considered expected, acceptable, or credible approaches to convey ideas (Driskill, 1996; St.Amant, 1999). Moreover, individuals often use the rhetorical expectations of their native language and culture to assess the credibility and the effectiveness of messages in other language and constructed by the members of different cultures (Uijjn, 1996).

Image Use and Visual Design

Further complicating this situation is the fact that such communication expectations generally transcend language. Different cultures, for example, also often have different expectations relating to visual communication. In some cases, these differences can involve what an item or object should look like in order for it to be recognizable (Atchison, 1994; Kostelnick, 1995;
Gillette, 1999). (What, for example, should a mailbox look like so potential users correctly identify a “send mail” icon?) In other cases, these differences can involve expectations of what constitutes an “appropriate” or a “credible” visual depiction of an object or a person (St. Amant, 2005). These ideas of recognizeability and credibility, moreover, are both related to the notion of exposure over time (Atchison, 1994). That is, the more one sees a particular item and is told, “This item is X.” (e.g., this item is a “mailbox”), the more the individual will expect the related item to appear that way. Variations from these exposure-based norms can, in turn, affect if individuals from other cultures can recognize the items depicted in a visual. Similarly, they can affect if individuals from other cultures consider a particular visual a credible or an appropriate one that merits consideration (or if the overall document or interface in which that visual appears is also a credible one that merits consideration or use).

Closely related is the practice of using visuals to convey abstract ideas. If, for example, one wishes to use an image as a metaphor to represent a particular concept, what should that visual be? (To convey a particular product is associated with education and learning, what should the related product logo depict?) In making this selection, communication designers need to be aware that the same item can represent different qualities or traits (i.e. have different metaphoric connotations) depending on the culture of the related audience (Horton, 1993 & 1994). Similarly, how should visuals be used to convey ideas more indirectly? When, for example, is it appropriate to have an advertising visual depict two competing products in order to imply that one is better than the other? It depends on the culture (and related cultural preferences) of the given audience (Kamath, 2000).

These variations in visual communication expectations, moreover, extend beyond the use of specific images or the design of certain graphics. Rather, they often also apply to the construction of overall visual elements such as interfaces. Consider the design of an organizational website. Where should the menu bar (or bars) be located to facilitate use? How many links should appear in them? And how many images should one include on such sites (and where should these images be located)? Such factors are not
universal. They instead depend upon the expectations of the user and, in many cases, are connected to the culture of the user and what expectations members of that culture have in relation to the overall design of different communication materials (Yunker, 2003; Sun, 2012).

**Usability and the User Experience**

Perspectives on what constitutes a usable design add another layer of complexity to intercultural and international communication contexts. In essence, just because a given technology or communication product or device exists in multiple cultures does not mean that item will be used in the same way(s) – or at all – across cultures (Sun, 2012; Getto & St.Amant, 2014). Many social media, for example, can be accessed and are considered acceptable in a range of nations and cultures, but research indicates these media are not always used as often, in the same ways, or at all depending on the culture of the user (St.Amant, 2015, April). Similarly, other research notes that the idea of creating communication materials for one specific culture and then trying to adapt them for others is not always an effective approach. In fact, the design of an item is generally so closely connected to the norms and expectations of the culture that created it, users in other cultural contexts often need to re-configure the item to use it effectively (see, for example, van Reijswoud & de Jager, 2011). For these reasons, a number of individuals have begun to advocate individuals begin the design and development process with different culture in mind from the start and then create materials in a way that best address varying cultural expectations of and conditions of uses (Langmia, 2011; van Reijswoud & de Jager, 2011; Sun, 2012; Getto & St.Amant, 2014).

These ideas of design and use, moreover, are further complicated by the fact that no culture is a monolith. Rather, there are different groups within a given culture, and each group brings with it different expectations and needs that reflect different attitudes, lifestyles, and situations or settings (Yu, Chan, & Ireland, 2007; Getto & St.Amant, 2014). For these reasons, communication designers studying culture and communication have begun to advocate approaches – such as the use of personas – to better understand the different populations that can exist within a greater
cultural group (see, for example, Getto & St. Amant, 2014). While such a degree of granularity can seem daunting, the rewards for these investments can be quite high. (This situation is particularly the case in relation to online media and interface design as international Internet access continues to expand to more nations and cultures.)

Understanding and addressing such cultural communication factors, however, are only part of the overall puzzle communication designers need to understand in order to engage in effective design for other cultures. The other major factor to address is the context in which the related materials will be used. The central issue here is that different contexts require the communication designer to account for and address different factors in order to create materials a given cultural audience will find usable. While these contexts are often manifold, one way to approach this overall situation is in terms of three meta-categories that examine the settings in which information is shared with different groups.

The Contexts in Which Cross-Cultural Communication Occurs

One of the greatest challenges in addressing aspects of culture and communication has to do with the context in which information will be used or will be exchanged. To address this factor from a communication design perspective, let’s think of three different contexts for interaction. The idea is that by better understanding such contexts, communication designers can make more informed and more effective choices about how to approach a particular situation that involves designing materials for users from other cultures.

**Context 1: Cross-Cultural and Intercultural Communication**

For the purposes of discussing these contexts, let’s define culture as a world view. That is, culture is a framework the members of a group use to identify what is important to the group and how to assign value based on that common notion of importance. In this way, our culture identifies what the members of our group/our culture communicate about (i.e., what the group values) and how
they communicate it (i.e., what the members of the group consider a credible way to interact in relation to that which is valued).

When groups with two different world views, or different frameworks for considering and valuing the world around them, interact, information moves from one culture to another and thus from one value system—and system for communicating and assessing the conveying ideas—to another. This context for interaction where information moves from one value system/world view/set of expectations for conveying ideas to another could be considered cross-cultural communication (i.e., ideas move across cultures). The process of conveying ideas and information back and forth across cultures – and different systems for conveying and evaluating the presentation of ideas and information – would thus be intercultural communication, for it is a context in which the members of two different cultural groups interact.

The central idea in this situation is the factors affecting communication practices are connected to identifying, understanding, and addressing cultural expectations vs. other items or aspects (e.g., different legal systems) that affect communication practices. This perspective also means that such cross-cultural or intercultural exchanges are not connected to geopolitical boundaries, but can take place within the boarders of the same nation (e.g., Canada has large Anglo-Canadian and French Canadian cultures, both of which exist in the same country and communicate within the contexts of that same nation). Thus, when communication designers study intercultural communication contexts, the focus of their research would be on identifying the different world views of the cultures interacting and then identifying how those differences affect communication patterns and expectations. The idea is by identifying, understanding, and addressing such cultural factors (i.e., differences), one can design more effective – and, ideally, more usable – communication products for users from that cultural group.

**Context 2: International Communication**

In other cases, communicating across cultures involves interacting across different nations. In these cases, communication is not only intercultural (i.e., involving more than one cultural group), it is also
international. This distinction has to do with geopolitics and economics. When different cultures interact in the same nation, they are often doing so under a common political and legal system (e.g., a set of national laws) that governs exchanges. As a result, the communication designer only needs to be familiar with and account for one set of national laws when creating information for these different cultural groups. Moreover, as the legal system under which the parties interact also generally stipulates the nature of a range of behaviors (e.g., how to settle a business dispute, what constitutes legitimate business practices, how to file a grievance or challenge a decision), the communication designer only needs to be familiar with and consider the nuances of this one system when creating materials for the different cultural groups interacting within it.

When communication becomes international, or involves more than one nation state, the legalities one needs to consider becomes more complex. In such cases, what might be permissible or legal behavior in one nation might be actionable or illegal in another (St. Amant, 2008). For example, the laws that govern what information can or must be shared with others, how, where, and when can also vary markedly from nation to nation (Markel, 2006). (Consider how the different national approaches to personal information and data disclosure affect how US companies can interact with individuals in EU member nations.) Similarly, privacy laws that regulate if and when the government can monitor communications (and hold individuals responsible for what they say) can vary from one nation to another – and such factors can affect how individuals in different nations use the same medium to communicate (e.g., uses of social media in the US vs. in the People’s Republic of China) (see St. Amant, 2008, 2015 March, & 2010 April). These different national laws can also affect how much control organizations in one nation might have over their information or materials once those items have been shared with individuals in another nation (Herrington, 2013). (Consider, for example, how different national statutes on copyright have influenced the ways in which organizations in one nation share proprietary information with partners in another country.)
In a similar way, the various political and economic systems within a nation can affect communication – and other – practices in a range of ways. The economic system of a nation, and the related political system that governs the nation, can affect what resources/funds are available and how they are used. Consider, for example, infrastructure. The amount of funds available to a particular national government and how that government decides to allocate those funds can affect how developed, widespread, and reliable different kinds of infrastructure are in a given nation or a particular region of that nation (Tawileh, 2011; van Reijswoud & de Jager, 2011; St.Amant, 2015, March). When two cultures are interacting under one governmental system, aspects such as these can be easier to identify and account for. When they expand to include different nations and political and economic systems, then assumptions about what is “expected” in terms of telecommunications or transportation infrastructure – as well as the mean income of prospective clients or consumers – needs to be re-thought, researched, and accounted for to make sure individuals in other systems can access and use materials (ideally, as intended) (Tawileh, 2011; van Reijswoud & de Jager, 2011; St.Amant, 2015, March). Accordingly, the more nations involved in an international interaction, the more varying national/geopolitical factors need to be identified and accounted for in relation to design and developing materials that can work in different national contexts.

**Context 3: Global Communication**

Within this framework, the idea of global communication represents the highest level of complexity. In such cases, an organization is attempting to release a product to or share information with as many markets in as many nations as possible (or feasible). Thus, the complexities of attempting to address not only cultural differences and national differences expands exponentially as communication designers try to develop materials that will work with a wide range of cultural groups distributed across an array of nations. In such cases, the question often becomes what exactly does a global market or a global audience mean? Is an organization truly attempting to share information with all cultures and nations everywhere, or is the idea to connect to only a select group of individuals across a broad range of nations (e.g., the middle
class/individuals who can afford a particular kind of product). And even when the desire is to connect to a narrow group within a greater global context, the number of cultural and national issues that need to be considered creates high levels of complexity that the communication designer must address.

While this range of contexts, when combined with cultural communication factors, can seem daunting, the situation is not impossible. The key to communicating effectively in or designing effectively for these different contexts involves one central factor: knowledge. The more the communication designer knows about the cultures for which he or she is creating materials and the contexts in which those items will be used, the more effectively he or she can develop products that meet user expectations and needs.

Considerations in Culture and Communication Design

Addressing these various and complex factors often comes down to two central concepts: information and approaches. Information has to do with what one knows about the culture for which she or he is designing materials and the context in which the related audience will use those materials or engage in interactions involving or relating to them. The more one knows about the cultural expectations and preferences of the related audience, the environment in which the audience will make use of that item, and how, when, and where the item will be used, the better the communication designer can create materials that meet those needs and address the dynamics of the related context. (In essence, knowledge of audience is power to design effectively for it.)

Gaining such knowledge, however, can be difficult, for cultures can change in rapid and unexpected ways over time. As a result, the approaches communication designers use to learn about the expectations and needs of users from different national and cultures and the contexts in which they interact need to be multifaceted and extend beyond a simple literature search for prior work published on cultural communication patterns. Rather, the approaches communication designers use must involve observing the members of a given culture in order to learn about perceptions.
and patterns of use. Similarly, communication designers need to consider ways to test different designs with the members of a given cultural audience to assess the usability of those designs and revise and adapt them to better meet audience needs and expectations when possible. Thus, resources that provide models of such approaches for acquiring this information and discuss the uses of different resources that contain certain kinds of information can help communication designers better navigate the various contexts in which culture can affect communication and design expectations. The entries in this special issue represent initial examples of the approaches communication designers can employ and the resources they can use to enhance the information they have on factors of culture, context, and communication.

The Objective of This Issue

The purpose of this special issue is not to provide communication designers with a wide range of strategies and solutions for engaging in effective cross-cultural communication in different contexts. Nor should it be seen as a definitive resource on such topics. (Those objectives would require a series of texts – if not an entire library – to address effectively.) Rather, the objective of this special issue is to provide communication designers with an overview of ideas to consider, approaches to try, and resources to use when developing materials for users from other cultures. To this end, the entries in this issue should be viewed as the start of a discussion related to investigating how individuals think about communication practices in different intercultural, international, and global contexts. The editor of this issue therefore encourages readers to consider how the ideas, information, and approaches examined in these articles might be applied, modified, or built upon to extend our understanding of designing for and engaging with users from other cultures.

As noted, communication technologies often reflect the expectations and practices of the cultures in which they emerged. They can also shift the ways in which individuals interact via such technologies and lead to changes in cultural communication patterns. These ideas are central to the first entry in this issue – Xiaobo Wang and Baotong Gu’s “The Communication Design of
WeChat: Ideological as Well as Technical Aspects of Social Media.” In this article, the authors examine how WeChat – a social media technology developed in the People’s Republic of China (PRC) – both reflects and is changing communication practices in that nation. To do so, Wang and Gu analyze how different features of WeChat allows citizens of the PRC to communicate in ways that circumvent certain governmental restrictions and engage in more open dialogue around sensitive political issues. By analyzing two cases where individuals employed WeChat to criticize and discuss different government actions, Wang and Gu reveal how the technologies cultures use to communicate can challenge conventional aspects of cultural discourse and of political control in national and international contexts. In so doing, the authors also present an approach for studying uses of communication technologies within the contexts of other cultures.

It is one thing to study the technologies created by and used by the members of a different cultural group. It is another to design materials, such as interfaces, for different cultural audiences. In such cases, success involves more than understanding audience expectations related to language and culture. Rather, communication designers also need to account for the various economic, political, and technological factors that can affect if, when, and how the members of another culture use certain materials. The idea of understanding such cultural contexts is the focus of Uttaran Dutta and Swayang Das’ article “The Digital Divide at the Margins: Co-designing Information Solutions to Address the Needs of Indigenous Populations of Rural India.” In the article, the authors present the results of a case study in which a research team developed interfaces for sharing information with individuals living in rural India. Early on in the process, it became apparent that new approaches were needed to better understand the intended users – individuals who had to overcome a range of challenges involving language, literacy, and familiarity with computers. By overviewing the approaches used to gather information on this population, Dutta and Das provide examples of the research communication designers can and should do to learn more about different cultural groups and contexts of use. In so doing, the authors also reveal how integrating users into the
design, testing, and research processes can result in materials that better reflect specific international contexts of technology use.

It is one thing to design materials for the members of a different culture; it is another to collaborate with individuals from other cultures on international design projects. In such cases, both parties need to understand the other’s culture to facilitate communication and collaboration to benefit all involved. Developing such understanding involves studying the ongoing relationship between collaborators and identifying areas in which miscommunication might occur and relationships might break down. Rudy McDaniel and Lanlan Kuang’s article “Cross-cultural Cinematic Communication: Learning from the Information Design Process for a Sino-American Film Competition” examines the complexities of such cross-cultural collaborations. In their entry, McDaniel and Kuang review a case in which a team comprised of individuals in the US and the PRC collaborated to co-host an international film festival spanning two nations. During the planning process, a number of cultural and political factors emerged that could have created problems between the collaborating groups. By examining the steps taken to address such issues and to open channels of communication, McDaniel and Kuang provide an effective example of approaches for learning about cultural dynamics while engaging in collaborations. Through this examination, the authors offer strategies for researching cultural practices and preferences while working with the members of another culture. In this way, McDaniel and Kuang build upon Dutta and Das’ prior entry by expanding how one can partner cross culturally to engage in more effective international design practices.

As discussed, a first step in studying cultural communication practices involves identifying the contextual factors that can affect interactions. These factors can be cultural and linguistic, or they can be legal, political, and economic in nature. In either case, access to effective informational resources on such topics can facilitate effective designs or collaborations across cultures. The idea of resources is central the issue’s final entry: Hilary Sarat-St. Peter’s “Designing with HDR Data: What the Human Development Report Can Tell Us about International Users.” In this concluding article,
Sarat-St. Peter discusses the Human Development Report (HDR), an annual publication of the United Nations Development Program. The report – which is a collection of data on different political, economic, and other factors for specific nations – provides relatively current information on a range of factors that could affect communication and design practices in different nations. As such, the HDR can serve as a valuable resource communication designers can consult to learn more about factors affecting individuals in different nations and regions. Using this resource effectively, however, requires an understanding of the data it contains as well as approaches to applying such information in different design contexts. Sarat-St. Peter, in turn, provides suggestions for using HDR data to guide communication design practices. She also notes how resources like the HDR can be used in classroom contexts to familiarize the next generation of communication designers with approaches for creating materials for other cultural groups.

By themselves, each entry in this issue provides important concepts and insights that can guide a range of activities related to developing materials for and sharing information with individuals from other cultures. When read as a volume, the combined ideas and approaches covered in this issue represent an important introduction to the resources, methods, and strategies communication designers can use to work effectively in different global contexts. In this way, the overall issue can serve as a foundation upon which communication designers can build and expand the ideas and approaches described here to a wider range of practices across different settings. Doing can help further our understanding of culture and communication design. It can also enhance our understanding of design, communication, and usability approaches and practices across a range of contexts associated with culture and communication.

References


Extra Materials

Optional readings
Notes