Geographical Information Systems in the Digital Humanities

Ian Gregory
Alejandra Zubiria Perez
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 ....
Regional Map of Greater Victoria

Average Frequency
- **Regional Route**: 15–60 minute service with limited stops
- **Frequent Route**: 15 minute or better service, 7am-7pm, Mon-Fri
- **Local Route**: 20–120 minute service

Legend:
- **Direction of Travel**
- **Route Name**
- **Transit Exchange**
- **UVic**
- **Park & Ride Lot (no overnight parking)**
- **Major Stop**
DHSI Wi-Fi

Network name: DHSI
Passkey: dhsi2019
Psst: Some 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, Saltspring 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 Saltspring 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!
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 to 10:00</td>
<td>Welcome, Orientation, and Instructor Overview (MacLaurin A144)</td>
</tr>
<tr>
<td></td>
<td>Welcome to the Territory</td>
</tr>
<tr>
<td></td>
<td>Welcome to DHSI: Ray Siemens, Alyssa Arbuckle</td>
</tr>
<tr>
<td></td>
<td>Welcome from UVic: Jonathan Bengtson (University Librarian), Alexandra D’Arcy (Associate Dean Research, Humanities)</td>
</tr>
<tr>
<td>8:30 to 10:00</td>
<td>Classes in Session (click for details and locations)</td>
</tr>
<tr>
<td></td>
<td>1. [Foundations] Digitisation Fundamentals and their Application (Clearihue A103, Lab)</td>
</tr>
<tr>
<td></td>
<td>2. [Foundations] Introduction to Computation for Literary Criticism (Clearihue A102, Lab)</td>
</tr>
<tr>
<td></td>
<td>3. [Foundations] Making Choices About Your Data (Digital Scholarship Commons, McPherson Library A308, Classroom)</td>
</tr>
<tr>
<td></td>
<td>4. [Foundations] DH For Department Chairs and Deans (David Strong Building C124, Classroom)</td>
</tr>
<tr>
<td></td>
<td>5. [Foundations] Developing a Digital Project (With Omeka) (Clearihue A031, Lab)</td>
</tr>
<tr>
<td></td>
<td>8. [Foundations] Fundamentals of Programming/Coding for Human(s)ists (Clearihue A108, Lab)</td>
</tr>
<tr>
<td></td>
<td>9. Out-of-the-Box Text Analysis for the Digital Humanities (Human and Social Development A160, Lab)</td>
</tr>
<tr>
<td></td>
<td>10. Sound and Digital Humanities (Cornett A120, Classroom)</td>
</tr>
<tr>
<td></td>
<td>11. Critical Pedagogy and Digital Praxis in the Humanities (Clearihue D132, Classroom)</td>
</tr>
<tr>
<td></td>
<td>12. Digital Humanities for Japanese Culture: Resources and Methods (McPherson Library A003, Classroom)</td>
</tr>
<tr>
<td></td>
<td>13. Conceptualising and Creating a Digital Edition (McPherson Library 210, Classroom)</td>
</tr>
<tr>
<td></td>
<td>14. Retro Machines &amp; Media (McPherson Library 128, Classroom)</td>
</tr>
<tr>
<td></td>
<td>16. Geographical Information Systems in the Digital Humanities (Clearihue A105, Lab)</td>
</tr>
<tr>
<td></td>
<td>17. Introduction to IIIF: Sharing, Consuming, and Annotating the World’s Images (Cornett A121, Classroom)</td>
</tr>
<tr>
<td></td>
<td>18. Web APIs with Python (Human and Social Development A170, Lab)</td>
</tr>
<tr>
<td></td>
<td>19. Ethical Data Visualization: Taming Treacherous Data (Cornett A128, Classroom)</td>
</tr>
<tr>
<td></td>
<td>20. Linked Open Data and the Semantic Web (Cornett A132, Classroom)</td>
</tr>
<tr>
<td></td>
<td>21. Palpability and Wearable Computing (McPherson Library A025, Classroom)</td>
</tr>
<tr>
<td></td>
<td>22. The Frontend: Modern JavaScript &amp; CSS Development (Clearihue A030, Lab)</td>
</tr>
<tr>
<td></td>
<td>23. Modelling. Virtual. Realities. A Practical Introduction to Virtual (and Augmented) Reality (Human and Social Development A150, Lab)</td>
</tr>
<tr>
<td></td>
<td>25. Information Security for Digital Researchers (David Strong Building C114, Classroom)</td>
</tr>
<tr>
<td>10:15 to Noon</td>
<td>Lunch break / Unconference Coordination Session (MacLaurin A144)</td>
</tr>
<tr>
<td></td>
<td>(Grab a sandwich and come on down!)</td>
</tr>
<tr>
<td></td>
<td>Discussion topics, scheduling, and room assignments from among all DHSI rooms will be handled at this meeting.</td>
</tr>
<tr>
<td>1:30 to 4:00</td>
<td>Classes in Session</td>
</tr>
<tr>
<td>1:30 to 4:00</td>
<td>Institute Lecture: Jacqueline Wernimont (Dartmouth C): &quot;Sex and Numbers: Pleasure, Reproduction, and Digital Biopower&quot;</td>
</tr>
<tr>
<td></td>
<td>Chair: Anne Cong-Huyen (U Michigan) (MacLaurin A144)</td>
</tr>
<tr>
<td>4:10 to 5:00</td>
<td>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 &quot;in what ways do theories of biopower, critical gender and critical race studies, and media studies&quot; 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.</td>
</tr>
<tr>
<td>5:00 to 6:00</td>
<td>Opening Reception (University Club)</td>
</tr>
<tr>
<td>Tuesday, 4 June 2019</td>
<td>DHSI Conference and Colloquium Lightning Talk Session 1 (MacLaurin A144)</td>
</tr>
<tr>
<td>9:00 to Noon</td>
<td>Classes in Session</td>
</tr>
<tr>
<td>12:15 to 1:15</td>
<td>Lunch break / Unconference</td>
</tr>
<tr>
<td></td>
<td>&quot;Mystery&quot; Lunches</td>
</tr>
<tr>
<td>1:30 to 4:00</td>
<td>Classes in Session</td>
</tr>
</tbody>
</table>
Wednesday, 5 June 2019

9:00 to Noon
Classes in Session

12:15 to 1:15
"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

4:15 to 5:15
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"

Thursday, 6 June 2019

9:00 to Noon
Classes in Session

12:15 to 1:15
"Mystery" Lunches
[Instructor lunch meeting]

1:30 to 4:00
Classes in Session

4:15 to 5:15
DHSI Conference and Colloquium Lightning Talk Session 3 (MacLaurin A144)
Chair: Kim O'Donnell (Simon Fraser U)
- Colleen Kolba (U South Florida), "What Comics can Teach our Students about Multimodal Literacy"
- Trish Baer (ETCL; U Victoria), "Preserving Digital Legacies: Archived Websites and Digital Discoverability"
- Suchismita Dutta (U Miami), "The Importance of Archival Transcription for Genre Building"
- Jeffrey Lawler (California State U, Long Beach), "Twining our way through the Past: Video Game Authoring as History Pedagogy"
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!

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)

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 Savicicvic (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)
Keynote - Nathan P. Gibson (Ludwig Maximilians U, München): "Thinking in -JTR: Reorienting the Directional Assumptions of Global Digital Scholarship"

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"
- Ayo Osusanwo (U Ibadan), "Online Newspaper Construction of Agitation for the Sovereign State of Biafra in Nigeria"
- Joseph Jones (U British Columbia), "Testbed for an Approach to Distant Reading: Fictions That Represent Vietnam War Resisters in Canada"
- 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"
- Ryan Ikeda (UC Berkeley), "Disrupting Digital Literacy: Situating Electronic Literature Among Public Education Initiatives"
- Christopher Church, Katherine Hepworth (U Nevada, Reno), "We’re STEAMed! A call for balancing technical instruction and disciplinary content in the digital humanities"
- Chelsea Milbourne (Cal Poly, San Luis Obispo), "Finding the Right Fit between Technology and Class Content: Reflections on Including Web Development in a Digital Storytelling Course"

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 (HuCSo) - 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"
- Felix Bayode Oke, Stella N. Kpolugbo (Anchor U Lagos), "The Multimodal Technique as a Pedagogical Tool in Pelu Awofeso’s White Lagos: A Definitive and Visual Guide to the Eyo Festival"
- 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"
## 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 116, 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)

#### 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)
- **62. DIY Digital Editions: Workflow + Philosophy [9 June; AM]** (Clearihue D132, 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 A121, Classroom)
- **66. DHSI Knits: History of Textiles and Technology [9 June; PM]** (Fine Arts 109, Classroom)
- **67. Book History Pedagogy Using Scalar [9 June; PM]** (Cornett A229, 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 Figgstad 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. Processing Humanities Multimedia (Human and Social Development A150, Lab)
- 46. Digital Humanities Pedagogy: Integration in the Curriculum (Cornett A121, Classroom)
- 47. Accessibility & Digital Environments (Priestly Law Library 265, Classroom)
- 48. Agile Project Management (Cornett A132, Classroom/Lab)
- 49. XPath for Processing XML and Managing Projects (Clearihue A105, Lab)
- 50. Endings: How to End (and Archive) your Digital Project (Priestly Law Library 192, Classroom)
- 51. Text Processing - Techniques & Traditions (Clearihue D132, Classroom)
- 52. Introduction to Humanities Data Analysis & Visualization in R (Human and Social Development A160, Lab)
- 53. Introduction to Network Analysis in the Digital Humanities (Clearihue D132, Classroom)

### 12:15 to 1:15
- Lunch break / Unconference Coordination Session (MacLaurin A144)
- "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), "OurDHI anti-colonial: Questions and challenges in dismantling colonial influences in digital humanities pedagogy"
- Julia King (U Bergen), "Developing Network Visualizations of Syon Abbey's Books, 1415-1539"
- Luis Meneses (ETCL; U Victoria), "Identifying Changes in the Political Environment in Ecuador"
- Alicia Brown (Texas Christian U), "Digital Cartography of the Ancient World"
- Laura Horak (Carleton U), "Building the Transgender Media Portal"
- Andrew Boyles Peterson (Michigan State U), "Last Mile Tracking: Implications of Rental Scooter Surveillance"

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  Presentation: An Introduction Jupyter Notebooks for Researchers  (MacLaurin A144)
This presentation introduces Jupyter Notebooks for researchers, via a partnership between Compute Canada and the Pacific Institute for the Mathematical Sciences (PIMS) including a large number of Canadian institutions. Read more here. Presenting is James Colliander, PIMS Director and team.

1:30 to 4:00  Classes in Session

4:15 to 5:15  DHSI Conference and Colloquium Lightning Talk Session 5  (MacLaurin A144)
Chair: Lindsey Seatter (U Victoria)
- Calvin 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"
- Kent Emerson (U Wisconsin-Madison), "Digital Mappa and the George Moses Horton Project"

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  "Mystery" Lunches

1:30 to 4:00  Classes in Session

4:10 to 5:00  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 Akademie (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.

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

**Contact info:**

institut@uvic.ca  P: 250-472-5401  F: 250-472-5681
The course offers an introduction to the theory and practice of using Geographical Information Systems (GIS) to research the past. It will be primarily based on using the ArcGIS software package, the use of Google Earth to disseminate humanities data will also be explored. The course will be relevant to historians, historical geographers, demographers, and others with an interest in the geographies of the past. Quantitative and qualitative approaches will both be explored. We would welcome attendees bringing their own data so that we can explore how to get it into GIS form and what can then be done with it.

Organisation:

1. Introduction to the Course and Introduction to GIS in the Humanities.
   In the session we will introduce ourselves and provide an overview of the course content. A reading list will also be provided. Subsequent to this there will be a lecture that will look at: what GIS is, why and how it should be used for humanities research, and give case study examples.

2. Introduction to GIS in the humanities – practical.
   This session introduces ArcGIS and its major modules: ArcMap and its extensions, ArcCatalogue, ArcToolbox and the Help system. The practical will be based on place names from Samuel Taylor Coleridge’s 1802 tour of the English Lake District.

3. Cartography in ArcGIS.
   Although ArcGIS stresses map patterns, producing high-quality maps that communicate the patterns within your data is a skilled activity. This lesson has three aims: the first is to understand mapping as a form of exploratory analysis, the second is to introduce the basics of cartography, and the third is to show how ArcMap allows high-quality maps to be produced. Topics will include: choice of class intervals, shading schemes (including the ColorBrewer website), legends, north arrows, titles and captions, scale bars. The lesson will combine a short lecture and a practical exercise which will finish with the student exporting a map to a bitmap file and reading it into PowerPoint or Word. The practical will be based on demographic data for England and Wales from the 1900s.
4. Working with tabular data.

GIS stresses the spatial patterns in data but most of the data that it uses is in tabular form. Data in humanities research such as census data, vital registration data, and economic statistics are also frequently found in tables. This practical will introduce the basic concepts of using tabular data within ArcGIS including querying tabular data and joining tables.

5. Data integration through overlay and buffering

The ability to use locational information to integrate data from different sources is central to the utility of GIS in humanities research. This session focuses on the different overlay and buffering techniques that can be used to integrate map-based data.

6. Places, coordinates and point data

Much of the data that will be used by a researcher in humanities GIS are available as place names that can be represented as points. In the practical exercise we will demonstrate how to convert a list of places for which we have coordinates into a Shapefile and then to explore that patterns that it contains. This is the easiest way of creating GIS data and can be done from many sources including raw data, gazetteer-based data, GPS data and even reading paper maps. The exercise will introduce the Geonames gazetteer of Britain and extracting data from it to create Shapefiles. These will be used to geo-reference one of Wordsworth’s poems.

7. Geo-referencing historical data

This session will start with a lecture on the fundamental aspects of how spatial data are measured and recorded. It will then explore issues associated with map projections and coordinate systems and how these are handled within ArcGIS. We will then move to a practical that takes a historical map of the Lake District and explores how this can be geo-referenced.

8. Geo-visualization using Google Earth

In this we will explore the use of Google Earth and how humanities data can be incorporated into it. This allows 3D-visualizations to be created and placed on the internet quickly and easily.

9. Work on own projects

Students are encouraged to bring any data or information that they want to use within GIS with them. In this session students will be given help to work with this information to turn it into a GIS database and then to begin exploring it using the techniques introduced in the course. If students do not have their own data a range of material will be made available for them to use instead.
Reading list

1. Introductions to GIS


2. GIS in the Humanities

a. Overviews


b. Case studies

Papers in the edited volumes above.

Papers in:

- *Social Science History*, vol. 34(2), 2010 has a special section on “Railways and political economy in Britain, France, and the United States, 1840-1950” in which all of the papers are GIS based.
- *International Journal of Humanities and Arts Computing*, vol. 3(1-2), 2009. Not a special issue but largely devoted to HGIS.
- *Special edition of Social Science Computer Review*, vol. 27(3), 2009


3. Cartography and Visualisation

a. General cartography


b. Class intervals


c. Cartograms


4. GIS and change over time

a. In GIS and geography


b. Areal interpolation


5. Gazetteers


6. Internet resources

**Historical GIS Research Network:** [http://www.hgis.org.uk](http://www.hgis.org.uk)

**a. Academic organisations:**

European Social Science History Conference: [http://www.iisg.nl/esshc/](http://www.iisg.nl/esshc/)
Association of American Geographers: [http://www.aag.org](http://www.aag.org)

**b. Commercial companies, etc**

Old-Maps.co.uk: [http://www.old-maps.co.uk/](http://www.old-maps.co.uk/)

**c. Open source GIS software**

For an overview: See: [http://www.freegis.org](http://www.freegis.org) or [http://opensourcegis.org](http://opensourcegis.org)
Quantum GIS (QGIS): [http://www.qgis.org](http://www.qgis.org)
MapWindow GIS: [http://www.mapwindow.org](http://www.mapwindow.org)
GRASS: [http://grass.fbk.eu](http://grass.fbk.eu)
**Neatline** – developed at the University of Virginia for mapping spatio-temporal data: [http://neatline.org/](http://neatline.org/)

**d. Gazetteers, georeferencing tools and data sources**

**Gazetteers:**

Natural Earth Data: [http://www.naturalearthdata.com/](http://www.naturalearthdata.com/)
Open StreetMap: [http://www.openstreetmap.org/](http://www.openstreetmap.org/)
ESRI World Gazetteer: [https://gis.ucla.edu/geodata/dataset/world_gazetteer](https://gis.ucla.edu/geodata/dataset/world_gazetteer)
English Place-Name (DEEP) Data: [https://digitisation.jiscinvolve.org/wp/2017/05/24/digitisation-of-english-placenames-deep-project-data/](https://digitisation.jiscinvolve.org/wp/2017/05/24/digitisation-of-english-placenames-deep-project-data/)
Ordnance Survey gazetteer data: [https://www.ordnancesurvey.co.uk/opendatadownload/products.html](https://www.ordnancesurvey.co.uk/opendatadownload/products.html)

**Geo-referencing street addresses:**


**Geoparsing:**

Recogito: [https://recogito.pelagios.org/](https://recogito.pelagios.org/)
Shapefiles:
ArcGIS Online has a lot of data especially for North America however most of this is not in Shapefile format. To access it use the Add data from ArcGIS Online option under Add data in ArcMap
British data: Edina (http://edina.ac.uk/census), Mimas (http://mimas.ac.uk/expertise/geospatial-data) and the UK Data Service (http://ukdataservice.ac.uk) all have data for academic registered users. The Ordnance Survey also has a range of free data available from: https://www.ordnancesurvey.co.uk/opendatadownload/products.html
US census and boundary data: http://www.census.gov/geo/maps-data
A number of websites offer a range of free shapefiles. These include:
http://www.diva-gis.org/gdata
http://www.statsilk.com/maps/download-free-shapefile-maps
http://www.geofabrik.de/data/shapefiles.html

Scans of historical maps:
Old maps online: http://www.oldmapsonline.org. The British Library is making a wide range of its collection available through this, see: http://www.bl.uk/maps/georefabout.html
David Rumsey historical map collection: http://www.davidrumsey.com
National Library of Scotland’s Georeferenced map collection: http://maps.nls.uk/geo/explore
Glossary


**Arc**: See *line*.

**ArcCatalogue**: The module that *ArcGIS* uses as a file manager.

**ArcGIS**: A commonly used GIS package from Environmental Systems Research Institute (ESRI). It updated and merged two of ESRI’s earlier products: *ArcInfo* and *ArcView*.

**ArcInfo**: Was the market leading GIS software package when GIS computing was workstation-based. Is now available incorporated in *ArcGIS* and can be used if you have the full ArcInfo license.

**ArcMap**: One of *ArcGIS*’s three major modules. It is the one in which most work takes place.

**ArcToolbox**: A module within *ArcGIS* that consists of a set of macros and other tools.

**ArcView**: Was a commonly used desktop GIS software package produced by Environmental Systems Research Institute (ESRI) but is now the basic license for *ArcGIS*.

**Areal interpolation**: The process by which data from one set of source *polygons* are re-districted onto a set of overlapping but non-hierarchical target polygons.

**Areas**: See *polygons*.

**Attribute data**: Data that relate to a specific, precisely defined *location*. The data are often statistical but may be text, images or multimedia. These are linked in the GIS to *spatial data* that define the location.

**Attribute querying**: A *query* that extracts features from a *layer* based on the value of its attribute data: for example, ‘select polygons with an unemployment rate greater than 15%’ would be an attribute query.

**Backcloth**: A *raster* image, usually a map or aerial photograph, used to provide contextual information behind other *layers* of data.

**Buffering**: A buffer is a polygon that encloses the area within a set distance of a spatial features. *Points*, *lines*, and *polygons* can all have buffers placed around them. For example, if a user is interested in all areas within 1km of a church, a buffer would be placed around all the points representing churches. This would create a new layer consisting of polygons representing those areas within 1km of a church.

**Centroid**: A point at the geometric centre of a *polygon*. This can be used to represent a *polygon* as a point.

**Choropleth maps**: Maps of quantitative data that show patterns by using different colours or different shading for polygons classed in some way. For example, a map of polygon-based unemployment rates (expressed as percentages) might sub-divide rates into 0-5, 5-10, 10-15 and 15-20 and shade the polygons accordingly.

**Coordinate pair**: An x and y coordinate used to represent a location in two-dimensional space, for example (6.523,4.910).

**Coverage**: See *layer*. 

---

*GIS in the Digital Humanities*  
*Introduction*  
Ian Gregory
Data capture: The process by which data are taken from the real-world (primary source), or from a secondary source such as a paper map, and entered into GIS software. From primary data this is usually through the use of Global Positioning Systems or remote sensing. For data taken from secondary sources this is usually done by digitising or scanning.

Database Management Systems: Software systems specifically designed to store attribute data.

DBMS: See Database Management Systems.

DEM: See Digital Terrain Model.

DGPS: See Differential GPS.

Differential GPS: A way of collecting Global Positioning Systems data with increased accuracy. It involves using a fixed base station at a known position to help find the location of a roving receiver.

Digital Elevation Model: See Digital Terrain Model.

Digital Terrain Model: A data model that attempts to provide a three-dimensional representation of a continuous surface. Often used to represent relief (height).

Digitiser: In GIS terminology this refers specifically to a digitising table or digitising tablet.

Digitising: In GIS this has a more precise meaning than in other disciplines. It usually refers to extracting coordinates from secondary sources such as maps to create vector data.

Digitising table: A flat table with a fine mesh of wires under the surface used to allow accurate digitising of paper maps.

Digitising tablet: Similar to a digitising table only smaller.

Dissolve: An operation in which adjacent polygons are merged if the values within a specified column of their attribute data are the same. An example might be merging polygons representing fields based on the value of their crop type. This would a new layer containing crop type that would only have boundaries between fields of different crop types.

Drape: Involves laying features over a digital terrain model to provide information on features that lie on the terrain. The terrain model provides the shape of the terrain. Draped features may then include a satellite image of the terrain to show land-use, and vector data to show features such as roads.

DTM: See Digital Terrain Model.

Edge-matching: See rubber-sheeting.

Error: In the context of GIS this means the difference between the real world and its digital representation.

Error propagation: As layers of data are integrated through overlays the error present on the output layer will become the cumulative total of the error present on all the input layers.

Gazetteer: A database table that provides coordinate information for each named place. They often can also be used to standardise the spellings of place names or to locate place names within an administrative hierarchy.

Geographical Data Analysis (GDA): A way of analysing data that explicitly incorporates information about location as well about attribute. This term may be used almost interchangeably with spatial analysis.

Geographical Information Science: Methods of exploring and analysing spatially referenced data that take account of the benefits and limitations of such data.

Geographical Information System: A computer system that combines database management system functionality with information about location. In this way it is able to capture, manage, integrate, manipulate, analyse and display data that are spatially referenced to the earth’s surface.
Geo-referencing: The process of proving a layer of data with a real-world coordinate system such as the British National Grid or latitude and longitude.

GIF: A commonly used bitmap file format.

GIS: See Geographical Information System.

GIS data: Data stored in a GIS are represented in two ways: attribute data says what the feature is, and spatial data says where it is using points, lines, polygons, or pixels.

GISc: See Geographical Information Science.

Global Positioning Systems (GPS): A system based on satellites that allows a user with a receiver to determine precise coordinates for their location on the earth’s surface. These are a primary source of spatial data.

GPS: See Global Positioning Systems.

Graphic primitive: The basic representations of spatial features used in GIS. These are usually points, lines, polygons or pixels.

Head-up digitising: The process by which vector data are extracted from raster scans using a cursor on-screen.

Idrisi: A raster based GIS software package produced by Clark Labs, Clark University

Interpolation: A method of reallocating attribute data from one spatial representation to another. A simple example is to reallocate data from sample points to polygons using Thiessen polygons. Kriging is a more complex example that allocates data from sample points to a surface.

Isolines: A line joining points of equal value. The most common example is the contour line on a map. Isobars showing lines of equal pressure on weather maps are another example.

JPEG: A commonly used bitmap file format.

Kriging: A form of statistical modelling that interpolates data from a known set of sample points to a continuous surface.

Latitude: The angle of a location on the earth’s surface from the equator expressed in degrees north or south. The Arctic Circle, for example, is at approximately latitude 66° North.

Layer: The GIS data model represents the world by sub-dividing features on the earth’s surface according to a specific theme. Each layer is usually geo-referenced. Examples of layers for a study area might include: roads, railways, urban areas, coal mines, etc. A layer usually consists of both spatial and attribute data.

Line: A spatial feature that is given a precise location that can be described by a series of coordinate pairs. In theory a line has length but no width.

Location: The position of a feature on the earth’s surface. In GIS this is usually explicitly defined in terms of precise coordinates.

Longitude: The angle of a location on the earth’s surface usually expressed in degrees east or west of the Greenwich Meridian. New York, for example, is at approximately 74° West.

Map algebra: A form of overlay used with raster data. In it the values for pixels on the output layer is calculated by performing a mathematical operation on the pixels from the input layers. The calculation may be arithmetic (addition, subtraction, multiplication, etc.) Boolean (and, or, not, etc.), or a variety of other types.

MapInfo: A commonly used desktop GIS software package produced by the MapInfo Corporation.

MAUP: See Modifiable Areal Unit Problem.

Metadata: Data that describe a dataset to allow others to find and evaluate it.
Modifiable Areal Unit Problem (MAUP): Where data are published using totals for arbitrary areas such as administrative units, the patterns that they show may be simply the effect of the administrative units rather than genuine patterns among the underlying population.

MPEG: A video file format that can be used to publish animations.

Network: A topological GIS data structure that uses a series of lines to describe, for example a transport or river network.

Network analysis: Usually used to analyse flows along a network. An example is finding the shortest path between two locations on a road network perhaps taking into account the differing speeds and fuel costs of different types of roads.

Node: The start or end point of a line segment. As such a node is often the point at which lines intersect.

Non-spatial data: See attribute data.

Overlay: A formal geometric intersection between two or more layers of data. A layer produced by an overlay will contain the merged spatial data and attribute data from both of the input layers.

Pixels: The small units that sub-divide space to make up a raster surface. They are usually small grid squares.

Points: Spatial features that are given a precise location that can be described by a single coordinate pair. In theory a point has neither length nor width.

Polygons: Spatial features that are areas or zones enclosed by precisely defined boundaries. The boundaries of a polygon are formed from one or more lines.

Polyline: A term for a line used by some GIS packages.

Primary source: In GIS terms this usually means a digital data source that is derived directly from the real world such as through Global Positioning Systems or remote sensing.

Projection system: A method by which features on a curved earth are translated to be represented on a flat map sheet. This involves converting from longitude and latitude to x and y coordinates.

QGIS: See Quantum GIS.

Quantum GIS: An open source GIS software package.

Querying: The process by which data are retrieved from a database.

Raster data model: A way of representing the earth’s surface by sub-dividing it into small pixels, usually square cells. Each pixel has values attached to it providing attribute data about the pixel.

Raster-to-vector conversion: The process by which vector features (points, lines and polygons) are automatically extracted from raster data. This usually requires a large amount of user input and is often error prone.

RDBMS: See Relational Database Management Systems.

Reference points: A small number of points used to geo-reference a layer, often the four corners of the layer. Once a layer has been digitised we know the coordinates of the reference points in inches from the bottom left-hand corner of the digitising table or digitising tablet. We also know their locations in real-world units from the map. This allows us to convert the entire layer’s coordinates from digitiser inches to real-world coordinates.

Relational Database Management Systems: Software systems that store data in such a way that tables can be joined together by linking on a common item of data.

Relational join: The way by which two or more tables from a Relational Database Management System can be joined together based on one or more common items.
Remote sensing: The process by which satellite images are created by scanning the earth’s surface using sensors on satellites.

**RMS Error:** See Root Mean Square Error.

**Root Mean Square Error (RMS):** A measure of the average error across a map. It is used in digitising to give an approximate measure of the difference between the real-world coordinates and the registration points on the digital layer.

**Rubber-sheeting:** The process by which a layer is distorted to allow it to be seamlessly joined to an adjacent layer. Often this has to be done when layers created from adjacent map sheets are joined together. It is a process that inevitably introduces some error.

**Satellite images:** Raster models of the earth’s surface produced from sensors on satellites.

**Scanning:** The process by which raster data is captured from paper maps.

**Secondary data:** In GIS terminology this means data that have been taken from a source other than the real-world. A map is the most common secondary data source within GIS, a gazetteer would be another example.

**Segments:** See lines.

**Sliver polygons:** Small polygons formed as a result of overlaying two or more layers of vector data. These are formed due to small differences in the way that identical lines have been digitised.

**Space:** In a GIS context this means position on the earth’s surface. Its meaning is very similar to location.

**Spans:** A raster based GIS software package produced by PCI-Geomatics

**Spatial analysis:** A way of analysing data that explicitly incorporates information about location as well about attribute. This term may be used almost interchangeably with geographical data analysis.

**Spatial data:** Data that define a location. These are in the form of graphic primitives that are usually either points, lines, polygons or pixels.

**Spatial querying:** A query that extracts features from a layer based on their location; for example, clicking on a point and listing its attribute data is a spatial query.

**SQL:** See Structured Query Language.

**Structured Query Language (SQL):** A language used by many Relational Database Management Systems to manipulate their data.

**Surfaces:** A surface is a way of modelling space that attempts to treat it as continuous rather than subdividing it into discrete features such as polygons. Surfaces are usually modelled either as raster data or digital terrain models.

**Temporal data:** Data that explicitly refer to time.

**Tessellation:** A sub-division of space into discrete elements. Raster surfaces sub-divide space into regular tessellations such as pixels. Polygons are examples of irregular tessellations.

**Theme:** See layer.

**Thiessen polygons:** A method of allocating space to the nearest point. The input layer will contain a set of points. The output layer, containing the Thiessen polygons, will contain polygons whose boundaries are lines of equal distance between two points.

**Tic points:** See reference points.

**TIN:** See Triangular Irregular Network.
**Topology**: The description of how spatial features are connected to each other. The important point is that topology is only about connection, not about other features such as angles or distance. The London Tube map is a good example of a topological map.

**Travelling Salesman Problem**: A form of network analysis that attempts to find the shortest or cheapest route between a number of locations on a network.

**Triangular Irregular Network**: A data structure that produces a continuous surface from point data. Often used to create a digital terrain model.

**Vector data model**: Divides space into discrete features, usually points, lines or polygons.

**Vector-to-raster conversion**: The process by which vector data are converted to rasters. This is usually automated.

**Voronoi diagrams**: See Thiessen polygons.

**Zones**: See polygons.
What is GIS?

Ian Gregory
Department of History
Lancaster University
Structure of talk

• 1. Define GIS and related terms
  – Geographic Information
  – GIS as:
    • a type of software
    • a set of tools
    • an approach to analysis

• 2. Data in GIS
  – Spatial and attribute
  – Geo-referencing data
  – Raster and vector
  – Layers of data

• 3. Querying a GIS database

• 4. Integrating data with GIS
Geographical Information (GI)

- Information that refers to a location on the Earth’s surface
  - Has both a spatial and a thematic component
    - Census data
    - Hospitals admissions data
    - Relief data (e.g. from contours)
    - Information on transport networks
    - A text about a specific place (e.g. S.T. Coleridge’s tour of the Lake District)
    - A collection of photographs or paintings of buildings
  - Locational component can be an explicit (e.g. a co-ordinate or a precisely defined administrative unit) or vaguer (e.g. “The area around London” or “In Gaelic speaking areas”)

Geographical Information Systems (GIS)

1. GIS: A type of software
   - A computer system that allows us to handle information about the location of features or phenomena on the Earth’s surface
   - Has the functionality of a conventional DBMS PLUS functionality to handle the spatial component of the data (manipulating, mapping, analysing).
   - GIS as a DBMS that allows us to explicitly handle the spatial
   - Common examples:
     - ArcGIS
     - MapInfo
     - QGIS
Geographical Information Systems (2)

2. GIS: A tool-kit

- Manipulate spatially:
  - Calculate distances and adjacencies
  - Change projections and scales
  - Integrate disparate sources

- Analyse spatially:
  - Quantitative analysis
  - Exploratory spatial data analysis
  - Qualitative analysis

- Visualise data:
  - Maps!
  - Tables, graphs, etc.
  - Animations
  - Virtual landscapes
Geographical Information Systems (3)

3. Approach:

- Explore the database:
  - In conventional ways
  - AND geographically
- Allows us to think about the implications of location
- Allows us to think holistically
- Should not be restricted by vendor-provided functionality
- Should be used imaginatively taking into account:
  - the advantages and limitations of geographical information
  - the traditions of your discipline
Types of data

• Two types of data are stored for each item in the database
• 1. Attribute data:
  – Says *what* a feature is
    • Eg. statistics, text, images, sound, etc.
• 2. Spatial data:
  – Says *where* the feature is
  – Co-ordinate based
  – Vector data – discrete features:
    • Points
    • Lines
    • Polygons (zones or areas)
  – Raster data:
    • A continuous surface
Geo-referencing data

- Capturing data
  - Scanning: all of map converted into raster data
  - Digitising: individual features selected from map as points, lines or polygons

- Geo-referencing
  - Initial scanning digitising gives co-ordinates in inches from bottom left corner of digitiser/scanner
  - Real-world co-ordinates are found for four registration points on the captured data
  - These are used to convert the entire map onto a real-world co-ordinate system

Source: ESRI (1997)
Layers

- Data on different themes are stored in separate “layers”
- As each layer is geo-referenced layers from different sources can easily be integrated using location
- This can be used to build up complex models of the real world from widely disparate sources
Raster data: Hastings

Scale: 1:100,000
Grid cell size: 50 m.
Minimum altitude: 0 m.
Maximum altitude: 174 m.

© Ordnance Survey
Example: Vector data
Querying GIS data

- **Attribute query**
  - Select features using attribute data (e.g. using SQL)
  - Results can be mapped or presented in conventional database form
  - Can be used to produce maps of subsets of the data or choropleth maps

- **Spatial query**
  - Clicking on features on the map to find out their attribute values
  - Which features on one layer intersect with features from another?

- Used in combination these are a powerful way of exploring spatial patterns in your data
Attribute query: Lung disease in the 1860s

Spatial data: Registration Districts, 1/1/1870

Attribute data: Mortality rate per 1,000 from lung disease among men aged 45-64

Source: Registrar General’s Decennial Supplement, 1871

Query: Select areas where mortality rate > 58.0
Spatial query: Lung disease in the 1860s

District: Alston with Garrigill
County: Cumberland
M_rate: 68.4
Par1911: Select parish = ‘PORTSMOUTH’ (Attribute)
Stations: Select stations that intersect with Par1911 (Spatial)
Mapping through attribute query

Deaths from lung disease among men aged 45-64, 1861 to 1870

Mortality rate per 1,000 from lung disease among men aged 45 to 64

- Less than 27.8
- 27.8 to 38.2
- 38.3 to 56.0
- Greater than 58.0

Source: Registrar General's Decennial Supplement, 1871
Conclusions

• Advantages of GIS
  – Exploring both geographical and thematic components of data together
  – Stresses geographical aspects of a research question
  – Allows integration of data from disparate sources
  – Allows analysis of data to explicitly incorporate location
  – Allows a wide variety of forms of visualisation

• Limitations of GIS
  – Data are expensive
  – Learning curve on GIS software can be long
  – Shows spatial patterns and relationships but does not explain them
  – Origins in the Earth sciences and computer science. Solutions may not be appropriate for humanities research
From GIS…
To GIS in the Humanities
Components of data

• 1. Attribute (theme)
  – Says what the data is
    • Statistical, textual, image, etc.

• 2. Temporal
  – Says when the data existed/are relevant for

• 3. Spatial
  – Say where the data refer to
    • Can be precisely defined “at grid reference (x,y)”
    • Can be vague “in the west of Ireland”
  – Difficult to handle
    • On paper
    • In a database
  – Neglected
Advantages of GIS

• 1. Structures a database

• 2. Data integration

• 3. Data visualisation

• 4. Spatial analysis
1. Streets of Mourning

- The Reveille Website by Lancaster Military Heritage Group
  - Alphabetical list of over 1,000 “Men of Lancaster” killed in World War One.

http://www.lancasterwarmemorials.org.uk/
Casualties by street

913 fatalities had at least one address (85.5% of total fatalities)
848 of these had an address that could be mapped (80.3%)
14 fatalities had three addresses (1.3%)
122 fatalities had two addresses (11.6%)
Casualties by School

442 fatalities had at least one school (41.9%) – could be improved

Note: Ripley has no location. See notes field for problems with casualties. The 1913 Lancaster Directory was used to help locate schools. All of the boys and mixed schools on that list have been found. It is assumed the “National School” or “Boys National School” or “Nashy” refers to the National School on St. Leonards Gate (http://theripleyassociation.wordpress.com/gallery/)
Media interest...

Lancashire Evening
Post 12/2/14

BBC Online
20/5/14

Lancaster Guardian
29/5/14
2. Data Integration

History of the book

Source: MacDonald B and Black F (2000) “Using GIS for spatial and temporal analyses in print culture studies” Social Science History, 24, pp. 505-536
3. Data Visualisation

Infant mortality, 1900s

Choropleths

Cartograms
The dots show where and when ‘Men of Lancaster’ died as the war progressed.

Each new dot represents the first Lancaster death at a cemetery or memorial.
4. Spatial analysis
Infant mortality and the Core-Periphery Divide

1850s

1900s
Changes in Infant Mortality 1850s-1900s

Overall

Proportional to national rate

National rate dropped from 153.25 to 127.59 – an improvement of 16.74%

Shading indicates population deciles
Lake District literature GIS:
Coleridge & Gray
Smoothed surface places mentionned

Gray

Coleridge
Comparing Coleridge and Gray

All mentions

Visits

Green: Only in Gray
Yellow: Evenly in both
Red: Only in Coleridge
Conclusions

• GIS allows improved handling of spatially-referenced data
• Improved understanding of space allows:
  – Structuring
  – Integration
  – Visualisation
  – Analysis
• It does not answer questions
  – Identifies patterns
  – Challenges the researcher to explain them
Case Study:
Using GIS in English Literature:
Mapping and analysing the early literature of the Lake District

This research was done in collaboration with David Cooper, Dept. of English and Creative Writing, Lancaster University

From History to Literature
"...this is an essay on literary history: literature, the old territory (more or less),...within that old territory, a new object of study; instead of concrete, individual works, a trio of artificial constructs – graphs, maps and trees – in which the reality of the text undergoes a process of deliberate reduction and abstraction. 'Distant reading', I have once called this type of approach;" (Moretti, 2005: p. 1)

| Graphs (p. 16) | Maps (pp. 55) | Trees (p. 73) |

Literary Mapping of the Lakes:
A Pilot for a Humanities GIS

- British Academy funded pilot project with David Cooper and Sally Bushell
- Two tours of the Lake District
  - Thomas Gray, 1769 (9,000 words)
  - Proto-Picturesque
  - ST Coleridge, 1802 (10,000 words)
- Romantic
- Aims:
  - Can we create a GIS of text?
  - What can it offer to literary research?
- Method:
  - Texts typed up by hand
  - Places tagged manually
  - Conversion
  - Analysis
On Sunday Augt. 1st - half after 12 I had a Shirt, cravat, 2 pair of stockings, a little paper Amp; half a dozen Pens, a German Book (Voss’s Poems) Amp; a little Tea Amp; supper, with my Night Cap, packed up in my natty green oil-skin, neatly squared, and put into my <format format_type="I">net</format> knapsack & the knapsack on my back && the Besom stick in my hand, which for want of a better, and in spite of <person>Mrs C.</person> & <person>Mary</person>, who both raised their voices against it, especially as I left the Besom scattered on the Kitchen floor; off I sallied - over the Bridge & thro' the Hop-Field, thro' the <pl_name visited="Y">Prospect Bridge</pl_name> at <pl_name visited="Y">Portinscale</pl_name>, so on by the tall Birch that grows out of the center of the huge Oak, along into <pl_name visited="Y">Newlands</pl_name> - <pl_name visited="Y">Newlands</pl_name> is indeed a lovely Place the houses...
Coleridge & Gray in a GIS

Smoothed surface of Gray’s places

- All mentions
- Visits

Smoothed surface of Coleridges’s places

- All mentions
- Visits

Class intervals are 10 equal intervals of the all mentions/UrbanGrid = 10 km
Comparing Coleridge and Gray

All mentions

Visits

Mapping Emotional Response

Gray

Coleridge

Physical Characteristics of Tours
Physical Characteristics of Tours

Altitude of mentions

Population density

Gray

Logged Population

Height

Visited

Didn’t visit/Unclear

Close Reading with Google Earth

To Flickr...
Incorporating photos from Flickr

Conclusions

- Pilot work with more to do but:
  - We can use texts within a GIS
  - We can geo-reference them in a semi-automated way to ask “what were they saying about where?”
  - Allows us to integrate a vast amount of disparate data sources:
    - Historical texts, modern images, historical maps, modern topographical data, etc.
  - Allows us to analyse them in traditional detailed ways and more social science-based volume approaches

Case Study:
Using GIS in Quantitative History:
Population change and transport in Wales, 1841-1911

Railways and Population
• Census data:
  – Total population at parish level 1841-1911
  – 1,240 parishes in Wales
  – Boundary changes don’t allow direct comparisons
• Railway data:
  – Locations of lines and stations
  – Each has an opening date and a closing date
• Question:
  – What was the relationship between the growth of the rail network and population change at parish level?

Welsh parishes in 1911
Interpolating parish-level data for England & Wales

- Target units: 1911 parishes
- 1911 census published 1901 data so these did not need to be interpolated:
  - 1881 was published in 1891 and 1861 in 1871 so these are done together.
- 1851 and 1841 data also published together and taken from a different source.
- EM algorithm using 32 ancillary classes based on pop. density

The History of Rail Lines in Britain

The Railways of Great Britain: A Historical Atlas
Ian Allan: Shepperton
Digitised by: Prof. J. Martí Henneberg and his staff at the University of Lleida (Spain)

The Growth of the Network
Rail growth and Population

Growth of the Network by Parish
Before 1880s

<table>
<thead>
<tr>
<th>Period</th>
<th>Already had railway</th>
<th>Gained railway</th>
<th>No railway</th>
</tr>
</thead>
<tbody>
<tr>
<td>1851 and before</td>
<td>147</td>
<td>116</td>
<td>927</td>
</tr>
<tr>
<td>1850s</td>
<td>305</td>
<td>252</td>
<td>603</td>
</tr>
<tr>
<td>1860s</td>
<td>557</td>
<td>29</td>
<td>694</td>
</tr>
<tr>
<td>1870s</td>
<td>686</td>
<td>20</td>
<td>620</td>
</tr>
<tr>
<td>1880s</td>
<td>636</td>
<td>31</td>
<td>600</td>
</tr>
<tr>
<td>1890s</td>
<td>637</td>
<td>27</td>
<td>575</td>
</tr>
<tr>
<td>1900s</td>
<td>664</td>
<td>3</td>
<td>577</td>
</tr>
<tr>
<td>After 1911</td>
<td>667</td>
<td>27</td>
<td>573</td>
</tr>
</tbody>
</table>

N=1240
### The Railway and Population

<table>
<thead>
<tr>
<th>% Pop. gaining railway</th>
<th>Cum. % pop with railway</th>
</tr>
</thead>
<tbody>
<tr>
<td>1851 and before</td>
<td>30.8</td>
</tr>
<tr>
<td>1850s</td>
<td>10.5</td>
</tr>
<tr>
<td>1860s</td>
<td>20.8</td>
</tr>
<tr>
<td>1870s</td>
<td>1.6</td>
</tr>
<tr>
<td>1880s</td>
<td>1.7</td>
</tr>
<tr>
<td>1890s</td>
<td>1.4</td>
</tr>
<tr>
<td>1900s</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Populations are by parish as measured by previous census (1850s is 1851) expressed as % of total pop. of Wales

### The Impact of the Arrival of a Railway

Each graph charts median population growth for parishes that gained their first railway in the decade.

### Population growth before, during and after the arrival of a railway

Median parish population growth (%) relative to Welsh average.

<table>
<thead>
<tr>
<th></th>
<th>Before 1850s</th>
<th>1850s</th>
<th>1860s</th>
<th>1870s</th>
<th>1880s</th>
<th>1890s</th>
<th>1900s</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>-3.11</td>
<td>-0.30</td>
<td>0.90</td>
<td>3.88</td>
<td>-0.73</td>
<td>-3.43</td>
<td>-2.36</td>
<td></td>
</tr>
<tr>
<td>Before decade</td>
<td>9.87</td>
<td>6.21</td>
<td>2.05</td>
<td>2.20</td>
<td>1.79</td>
<td>-0.19</td>
<td>-2.53</td>
<td></td>
</tr>
<tr>
<td>After</td>
<td>10.14</td>
<td>3.25</td>
<td>0.23</td>
<td>0.73</td>
<td>4.40</td>
<td>2.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>147</td>
<td>158</td>
<td>252</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>27</td>
<td>576</td>
</tr>
</tbody>
</table>

Example: Parishes that gained a railway in the 1860s grew by 2.55% above the Welsh average in that decade. In the decades prior to that they had grown at 0.3% below average and in the decades afterwards they grew at 0.23% above average.

“Never” includes parishes that did not gain a railway until after the 1900s.
Conclusions

1. Integration:
   - Over time:
     • All parish-level population interpolated on 1911 boundaries
     • Allows inter-censal change to be calculated
   - Different themes:
     • Railway lines and parishes
     • Allows the number of parishes with a railway to be calculated along with their total population

2. Visualisation:
   - Maps of the data in various ways
   - Basic animations to show change over time
   - Tables of results
   - Time series graphs to show change over time

3. Analysis:
   - Allows the data to be manipulated to set up further analyses

Exercise 2: Introduction to Geographical Information Systems for Historians: Beginning ArcGIS

The Lake Poets

Background

Geographic Information Systems (GIS) technology makes available a wide range of tools that can aid in the visualization and analysis of humanities information in ways that are not otherwise possible. ArcGIS is one of the leading GIS software packages. This exercise will introduce you to the basic functionality that ArcGIS offers. We will return to many of these in more detail later in the course. The primary goals in this session are to develop an understanding of the basic navigation of ArcGIS and to begin to build up an appreciation of the power of this technology.

Data Sources

This exercise is primarily based on two layers of data that give the locations and some basic attribute information taken from the work of two of the Lake Poets of the early nineteenth century. First we have places mentioned in Samuel Taylor Coleridge’s 1802 ‘circumcision’ of the Lake District in which he climbed Sca Fell, the highest mountain in England. Secondly, we have some parts of William Wordsworth’s *Directions and Information for the Tourist*. There are also three “background” layers giving the outline of the coast of England (and part of the Scottish border), the major Lakes, and the towns.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STC_1802</td>
<td>A point shapefile containing places mentioned in Coleridge’s tour. Attribute data includes the date in which he mentions the place (cal_date), the day of the week (DoW), the place name as he spelt it (pl_name), whether he is actually at the place or talking about it from a distance (visited), a standardised version of the name (st_name) and which day of his tour it was (day_no).</td>
</tr>
<tr>
<td>WW_Guide</td>
<td>A point shapefile containing places mentioned the Wordsworth text. Attribute data includes the section in which the mention occurs (sub_section), the place name (pl_name), and a standardised spelling of the place name (st_name).</td>
</tr>
<tr>
<td>Coast</td>
<td>A line shapefile containing the outline of the north of England.</td>
</tr>
<tr>
<td>Inland water bodies</td>
<td>A polygon shapefile that gives the the major lakes.</td>
</tr>
<tr>
<td>Urban areas</td>
<td>A polygon shapefile containing the major towns.</td>
</tr>
</tbody>
</table>
Summarized Guidelines for Completing Exercise

The following outline describes the general process that you will follow to complete the exercise. Specific steps will be provided later in the step-by-step instructions component of the exercise.

- ArcCatalog – this is ArcGIS’s file manager. Its functionality and use are similar to Windows Explorer. It is needed primarily because a Shapefile consists of several files (*.shp, *.dbf, *.sbn, *.sbx, etc.). As a result renaming, deleting, copying or moving these files should always be done in ArcGIS and never in Windows Explorer.
  - Explore data organization
  - Explore menus
  - Explore buttons

- ArcMap – the main interface into ArcGIS
  - Basic map navigation and manipulation
  - Add data to a data frame
  - Change the symbology on maps
  - Spatial and attribute queries
Step-by-Step Instructions

First you should use Windows Explorer to copy the data for the practical to the folder that you want to work on it in. All the data should be in a the folder <My folder>\GIS2, which corresponds to, for example, C:\users\..\Session 2\GIS2.

In the following steps you will explore ArcGIS, beginning with ArcCatalog.

1. Open ArcCatalog so that you can explore the exercise data.

   - Start ArcCatalog by choosing All Programs>ArcGIS>ArcCatalog 10 from the Start menu. The window that opens for you will look slightly different to the below however the basics are the same: on the left the Catalogue Tree shows where we are on the computer’s folder system, the main window shows the files within that folder.

   - One unusual feature of ArcGIS is the concept of a connection to a folder. This is effectively a shortcut to the folder that we are working in. They are useful but also take a little of getting used to.
You need to find your `<My folder>/GIS2` folder. To find it, click on the `Connect to folder` button shown below.

This brings up a list of folders. Navigate through it to find your `GIS2` folder. Click on this so that it is highlighted then click `OK`.

![Connect to Folder dialog box](image.png)
You will see that a shortcut to the folder appears as one of the *Folder Connections* in the *Catalogue Tree*. The main window now has a list of all of the ArcGIS compatible files that are in this folder. You will see that the different type of files all have their own icon. There are different ones for respectively: Excel spreadsheets, line shapefiles, polygon shapefiles, ArcGIS map documents, point shapefiles, etc.

Click on the *STC_1802* data source in the catalog tree (on the left of the screen) to make this the target data source. You may need to press the small “plus” symbol to open the *GIS2* folder first.
When you select a data source in the Catalog Tree, the highlighted data source is the active data source. Any subsequent actions that you take such as reviewing metadata, viewing properties, and so forth will be based upon that data source.

- Click the Content tab to view the format information or a thumbnail of the STC_1802 data source.
A Thumbnail image of the shapefile can be created and will replace the generic format thumbnail provided by ArcCatalog by default.

- Click the **Preview** tab to view the data in the **STC_1802** data source. Initially it will show you the spatial data – a rather meaningless selection of points.

- Choose **Table** from the **Preview** pick list at the bottom of the screen. The screen will update to show you the attribute data for the **STC_1802** data source.
The Preview tab is very useful when you want to quickly view the geography and attributes of a data source without actually adding the data source to a map.

- Click the Description tab to view the metadata for the STC_1802 data source.

   ArcGIS provides a series of tools for editing metadata. These can be filled in to ensure that you remember what the data are, where you got them from, and what you have done to them. Without metadata a dataset will be of very limited use to another user.

2. In the next few steps you will explore the basic ArcCatalog Menus.

- Go back to the Contents tab. We will now explore the basics of the ArcCatalog menu.

  - Click on each of the menu items and review the available options. You will further explore these options in later exercises. For now the basics are:
File: Allows you to create new folders, shapefiles, etc.; to create new connections to folders, and to remove these connections.

Edit: Copy and paste files and folders.

View: Things such as refreshing the current view.

Go: Navigate around the folder structure.

Geoprocessing: Access various ArcToolbox tools, ArcToolbox itself, and various other ways of geoprocessing the data.

Customize: Provides access to additional functionality particularly through new toolbars and extensions.

Windows: Change the window that you are looking at.

Help: The ArcGIS help system is extensive and helpful.
3. Explore the basic buttons of ArcCatalog.

- There are various shortcuts to functionality here. The buttons shown are respectively: Go up one folder level; Create a connection to folder; Delete a connection to a folder; Copy; Paste; Delete; Four buttons that change the view of the data; Launch ArcMap (the globe button); Catalogue Tree window; the search window (which is a very useful way of finding the ArcTool that you want); Launch ArcToolbox; etc.

- Click on the ArcMap button to launch ArcMap
4. In the following steps you will explore ArcGIS ArcMap.

You just opened ArcMap by launching it from ArcCatalog. Alternatively, you can open ArcMap by choosing All Programs>ArcGIS>ArcMap 10 from the Start menu.

If you have a dialog box in the middle of your screen asking you to choose between several options: a new map, a template or an existing map, do the following:

- Under Existing maps select Browse for more… and navigate to <My folder>\GIS2\Lakes_lit.mxd
- Click OK to open the map.

If this dialogue box is not here then do:

- Choose File > Open then choose the <My folder>\GIS2\Lake_lit.mxd document.

The resulting window should look like:

If no map appears and there are red exclamation marks on the table of contents on the left hand side this means that ArcGIS has lost track of which folders the shapefiles lie in. This can happen if shapefiles and map documents are copied around, particularly across servers and between memory sticks. To correct this right click on the icon for one of the shapefiles on the table of contents. The select: Data > Repair data source… and find the
shapefile in the Data sourse window. When you have found it click Add. If you have any problems with this then ask!

5. Basic map navigation.

- Locate the ArcMap Tools toolbar which has the following buttons.

- Experiment with the Zoom In and Zoom Out, Pan, Full Extent, and Back/Forward Arrow buttons. Holding the mouse steady over a button will give a little bit of text saying what the button does. The Globe symbol (Full extent) is particularly useful.

   When using the magnifying glass tools, place the magnifying glass – your mouse pointer after clicking on the magnifying glass tools – near the location you want to zoom. Click and hold down the Left Mouse Button, and then move your cursor until you have drawn a box around the area of interest. Finally, release the left mouse button when satisfied with the selected area.

- Zoom in as shown below
Choose **Bookmarks> Create**. A Spatial Bookmark dialog box appears.

Name the bookmark *Inner Lakes*, and then click **OK**.

Click the Full Extent button on the Tools toolbar to view the entire map.

Go to the **Bookmarks** menu – the *Inner Lakes* bookmark that you just created is now here. Click on this.

*Bookmarks are useful when you find yourself focusing on a particular area frequently. It is a short cut to quickly orient the map.*

Click on the black **Select Elements** arrow on the Tools bar to return the cursor to showing the Select Elements arrow.

Click the check box to the left of the **Inland water bodies** and **Urban Areas** icons in the Table of Contents. These turn these layers on (and back off again)

Uncheck the **STC1802** check-box on the table of contents and turn on **WW_Guide** instead. This shows the places from Wordsworth in green.
Layers draw from the bottom first to the top last. This means that STC_1802 is on top as it was the last to draw. As it is turned off then WW_Guide appears on top. If you turn STC_1802 on you will see some of the Wordsworth places disappear as they lie underneath Coleridge’s points.

- At the top of the table of contents there are a number of buttons that provide slightly different functionality for the table of contents. Make sure that the left-hand button List By Drawing Order is pressed.

- Make sure the Inland water bodies layer is turned on (the check box by its name is ticked)

- Click on (highlight) the Inland water bodies layer and hold down the Left Mouse Button.

- Drag the Inland water bodies icon to the top of the list of layers on the Table of Contents and release. Note how the map changes. As the lakes move up, some of the points from first WW_Guide and STC_1802 disappear under them.

  Drag-n-Drop is a method for quickly changing the hierarchy of Layers in a data frame. Note, however, this functionality is not available if the List by Source, List by visibility, or List by selection options have been selected.

- Drag Inland water bodies back down to the bottom of the table of contents.

- Right-click (highlight) on STC_1802 layer to view the layer context menu.

- Click on Properties at the bottom of the layer context menu to open the Layer Properties window.
Click through the various property tabs to review the properties of the Layer. Then click **Cancel** to close the **Properties**.

*As you will learn in future exercises, a great deal of functionality is accessible through the layer properties dialog window.*

Right-click on (highlight) the STC_1802 layer to view the layer context menu.
- Click on **Open Attribute Table** to open the table for the STC_1802 layer.
- Review the attributes of the layer.
- You can move the window that contains the attribute table around by dragging and dropping the top bar of the window. When you do this you will see a number of blue arrows appear on the background of the main window. Move the mouse over the right-hand most of the arrows. You will see that the attribute table window now becomes part of the main window.
- Clicking on the pin symbol in the top-right of this window (circled above) will close it and hide it behind a tab as shown below.
- Clicking on the tab will re-open it. Clicking on the cross next to the pin will close the attribute table window altogether.

- Right-click on the STC_1802 layer to view the layer context menu, and then click on Remove to remove the STC_1802 layer from the data frame.

  Removing a Layer or Data Frame from the Table of Contents does not delete the data.

- Click on the Add data button at the top of the window - a black cross on a yellow square.

- Select the STC_1802 shapefile and click Add. STC_1802 will re-appear on the top of the map but will be represented using different symbols.

- Right click on the STC_1802 icon on the table of contents and choose Properties. Choose the Symbology tab.
Click on the Symbol button in the middle of the window. Change the symbol to “Circle 1”, colour it in red, and change the size to 6.

Click on OK to apply this and then either OK or Apply (Apply does the same thing as OK without closing the window).
Click on the arrow next to the Add data button and then click on Add Basemap...

Choose Terrain with Labels and press Add (this may take a minute – if the globe in the bottom right-hand corner is turning it is still working)

If you now do File > Save (or click on the Save button) this will save Lakes_lit.mxd map document to reflect the changes that you have made. This includes, for example, which layers are included, which order they are drawn in, whether they are turned on or off, what symbols and shading is used, how zoomed in we are, etc. It does not make any changes to the shapefiles themselves.
➢ Do File > Exit to leave ArcMap

➢ Now re-open ArcMap and re-open Lakes_lit.mxd. The changes you have made should appear.
6. Changing shading and symbols on a map.

- Set your map up as shown below (ie similar to when you started but zoomed in more closely).

- Right click on `STC_1802` on the Table of Contents. Then select `Properties...` from the context menu and choose the `Symbology` tab. Then click on `Categories – Unique values` under `Show`: Change the `Value Field` to `visited`. Click on `Add All Values` as shown below.
This means that there are three values of Visited: and “Y”, “N” and “U”. These mean that “Yes” Coleridge was actually visiting the place when he mentions it, “No” he was not there, he was talking about it from somewhere else, or that it is “Unclear” whether he was there or not. There is also one point that has no value for this field.

- Change the Color Ramp to change the basic shading of the different point symbols.

- Double click on the symbols to change the symbols, colours, size and angles. Experiment with these. You can also click under Label to change these from “Y”, “U”, “N” to “Yes”, “Unclear” and “No”.

- When you are happy press Apply or OK.
You will notice that the legend on the Table of Contents also changes to reflect the changes that you have made.

7. Querying the data.

GIS allows you to query a dataset in two different ways: Spatial querying where you usually click in features on the map to see what their attributes are, and Attribute querying where you select using the attribute data as you would in Excel or Access, but the results are returned spatially.

The easiest form of spatial querying is done using the Identify button on the Tools menu. Click on this and the Identify window appears. Change Identify from to STC_1802, and click on any point. Attribute information will be returned to you as below. You will note that you can move this window around and pin it in the same way as you can with the attribute table window.
- Can you find St. Bees Head? (It is on the peninsular in the west). What was the date when Coleridge mentioned this in his notebook?

- An attribute query is best done by looking at a dataset’s attribute data. We will do this on **STC_1802**. Right click on its icon on the table of contents and chose *Open attribute table* from the context menu. Now click on the **Table Options** at the top-left of this window to bring up the menu shown below. Choose *Select by attributes...* (You can also do this from the **Selection** menu at the top of the screen).
➢ Double click on “cal_date” the single click on the = button then press Get Unique Values. Scroll down the list to find ‘3/8/1802’ and double-click on it. This should give you the window shown below.

![Select by Attributes Window](image)

➢ Click on Apply and the Close to get rid of this window.

➢ You will see that Coleridge only mentioned five places on his account of this day. They are highlighted in blue on both the map window and on the attribute table. If you press the Show selected records button at the bottom of the Table window (circled below) then only these records will be shown.
You will see that there appears to only be three on the map window. There are two possible explanations for this: either the other points are off of the current display or two or more points are on top of each other. One way to check this is to use the Zoom to selected button at the top of the Table window. This is also available under the Selection menu or on the context menu on the STC_1802 icon under Selection). All three options are shown below.
From this it should be clear that all five mentions are very close together and two must lie under the other three.

Now clear the selection by using *Clear selected features*. This is available close to any of the three options above and is also on the *Tools* menu.

Close the *table* window by clicking on the black X in the top-left corner.

8. **ArcToolbox.**

We now want to see how many points lie on top of each other to avoid the problem above. This can be done using a macro on ArcToolbox.

To start ArcToolbox click the red toolbox button at the top of the screen. The tools will appear in the right of the screen.

Open the tools using the + symbol to find *Spatial Statistics Tools > Utilities > Collect Events* as shown below. Double click on this to open it.
Double click on this tool to open it.

Drag the **STC_1802** icon into the **Input Incident Features** box. It may also help to have **Show help** turned on. Now press the browse button (circled above) and select where you want to save the new layer that you are about to create. Call this **STC_count**.
When the window looks like the above (the file path may vary – it should be GIS2) then press okay. The progress of the macro is given in the bottom left of the screen. This can be a little slow to respond. When it has finished a blue window will appear in the bottom-left.

You will see that a new shapefile has been created and has been added to the window.
9. Toolbars, extensions and back to ArcCatalog.

- ArcGIS has a large number of different menus called Toolbars that hold much of its functionality. You have already been using the Tools toolbar to zoom in and out.

- Go to Customize > Toolbars > Tools and click on this. You should see the Tools toolbar disappear. Do this again so that it re-appears.

- Explore the range of other toolbars on offer.

- Some of the functionality of ArcGIS is only available through extensions. If the extension is not turned on the functionality will not be available. Go to Customize > Toolbars > Geostatistical Analyst. This gives some functionality for exploring your data. If you try to use any of the tools on the Geostatistical Analyst toolbar you will see that they are all greyed out and unavailable. This is because the extension is not turned on.

- To turn it on go to Customize > Extensions… and check them all. Now click Close. This is shown below. This will provide additional functionality if you have the appropriate licenses.
Ensure that ArcCatalog is not open. If it is close it.

You can open a cut-down version of ArcCatalog from within ArcMap. Click on the Catalog Window button, two to the left of the ArcToolbox button. This opens a cut down version of ArcCatalog on the left of the screen. You can delete, rename, move, etc. shapefiles from here.

Further work:

Review what you have learned. Make sure you are familiar with the basic functions of ArcCatalogue, ArcMap, ArcToolbox, and the Help system. Explore these basics using other layers or trying to do other things.
Recap

This practical teaches you the basics of ArcGIS without going into any details – these will follow in the next sessions. From doing this exercise you should begin to understand the following:

- The difference between ArcCatalog and ArcMap
- The basic menus and toolbars in ArcCatalog and ArcMap
- The difference between a shapefile and a map document
- How to open a shapefile’s attribute data and change its shading using the context menus on the Table of Contents (sections 5 and 6)
- How to create a folder connection (section 1)
- How to fix the red ‘!’ problem when a map document has lost the locations of its shapefiles (section 4)
- How to add and remove data from a map document and to save the map document (section 5)
- How to add a base map to a map document to provide background context (section 5)
- How to do simple querying of attribute data (section 7)
- How to use an ArcToolbox tool (section 8)
- How to turn on extensions that you might have a license for (section 9)
Cartography: Visualising data using maps

First Principle of Design

Keep the map simple and avoid graphic clutter.

Good cartography is founded on simplicity of design and complexity of data.

Basics of map design

- The 10 seconds rule:
  - A reader should be able to understand the basic message of the map within 10 seconds of looking at it.
- The map should be a stand-alone product
  - Its message should be clear without the reader having to refer to the accompanying text.
Map Elements

Each map should have:

- A title
- A legend
- A scale bar
- Visual balance
- A clear choice of shading or symbols
  - Hierarchy
  - Easily understandable symbols
- Justifiable class intervals
  - Where appropriate

Choice of shadings

Colours should have a clear hierarchy from smallest (light shading) to highest (dark shading)

Colours in a computer are shaded using CMYK or RGB combinations

CMYK: Cyan, Magenta, Yellow, Black in ranges of 0 to 100
RGB: Red, Green and Blue in ranges of 0 to 255

See: [http://www.colorbrewer.org](http://www.colorbrewer.org) for a good selection of shadings

Different CMYK combinations

- Cyan=0
  - Magenta=0
  - Yellow=0
  - Black=0
- Cyan=100
  - Magenta=0
  - Yellow=0
  - Black=0
- Cyan=50
  - Magenta=0
  - Yellow=0
  - Black=0
- Cyan=0
  - Magenta=100
  - Yellow=0
  - Black=0
- Cyan=0
  - Magenta=50
  - Yellow=100
  - Black=0
- Cyan=100
  - Magenta=0
  - Yellow=100
  - Black=0
- Cyan=50
  - Magenta=50
  - Yellow=0
  - Black=0
Different CMYK combinations – colorbrewer.org

Use of symbols and colour

- Try to make the map as understandable as possible without the reader having to refer to the legend
- Symbols:
  - Make these as intuitive as possible
  - Large sizes and/or heavy shading mean important
- Colours:
  - Certain colours have certain meanings:
    - Heavy shading – important, light shading – less important
    - Blue – water
    - Red – danger/bad
    - Green – good, rural/vegetated

Choice of class intervals

Interval/ratio data need to be split into a number of classes defined by class breaks

The choice of class breaks will heavily influence the pattern that the map shows

Number of classes: With grey-scale 4 or 5 is the maximum, in colour a few more are possible
Choice of class intervals

Equal interval:
- 0-20, 20-40, 40-60, etc.

Quantiles:
- Equal numbers of observations are put in each class.

Mean and standard deviation
- Natural breaks:
  - Minimizes the variances within classes.

Arithmetic progression:
- \( 1, 4, 10, 19, \ldots (3^n, 3^n+3) \)

Geometric progression:
- \( 1, 4, 13, 40, \ldots (3^n, 3^n*3) \)

Choose your own. BUT choices should never be arbitrary.

Four maps of the same data

Population density in Warwickshire, 1951

- a. Nested means
- b. Quantiles
- c. Arithmetic progression
- d. Geometric progression

Conclusions

Keep the map simple and avoid graphic clutter.

Good cartography is founded on simplicity of design and complexity of data.
Exercise 3: Cartography in ArcGIS

Mapping population data

Background

GIS is well known for its ability to produce high quality maps. ArcGIS provides useful tools that allow you to do this. It is important to note, however, that being able to use the tools does not make you a good cartographer. Cartography is the art of presenting your map to a reader in a way that makes an honest and comprehensible picture of the data. This requires good cartographic skills and effective choices of such things as: numbers of classes, class intervals, choice of shading and symbols, presence of a title, legend and scale bar, and overall balance of the map.

This exercise will introduce you to the basics of cartography. You will use ArcGIS to produce a map of population density in England & Wales from the 1911 census. Population density is a difficult variable to map because its has a skewed distribution, in other words there are many low density polygons (in the countryside) but the cities have very high densities even though there are only a few of them and they have very small areas. This makes choice of class intervals in particular very important.

Data Sources

GIS is particularly well-suited to quantitative data that contain a large amount of spatial information. One particularly strong source of this type of information is the census and vital registration data (births, marriages and deaths) which were published in much the same way, namely statistics such as the total population, number of unemployed people, number of deaths aged under one, etc. are given for a precisely defined administrative unit. In England & Wales until the First World War the Registration District was the main unit used. Once clear difference between census and vital registration data is that the census is a count of the people resident on a particular night and is taken once every ten years. Vital registration data counts the number of events, such as births, over a period of time, usually a decade between two censuses. This can cause problems, for example if you want to calculate the number of births (vital registration data) per 1,000 women aged 15-49 (census data) over a decade should you use the number of women at the start of the decade, the end of the decade, or the average of the two? (the answer is that there is no perfect solution).

This exercise introduces data from the 1900s. For attribute data we have some 1911 census data and some 1900s data on infant mortality (deaths amongst babies before their first birthday). These are both in Excel and dBase format respectively. Dbase is easily created from Excel using Save As... For spatial data we have the boundaries of Registration Districts in England & Wales in 1911, plus two sets of county boundaries: the registration county boundaries and the...
administrative county boundaries. These are significantly different because registration districts did not aggregate to form any existing type of county so a new type of county, the registration county was formed. The exercise is primarily concerned with data on population density from the 1911 census. This is stored in the item Pop_den on the layer RD1911.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD1911</td>
<td>Shapefile containing Registration Districts in 1911. Also has population density as an attribute.</td>
</tr>
<tr>
<td>Reg_county</td>
<td>Shapefile containing the Registration Counties of England &amp; Wales.</td>
</tr>
<tr>
<td>Admin_county</td>
<td>Shapefile containing Administrative Counties of England &amp; Wales.</td>
</tr>
<tr>
<td>1900s IMRs</td>
<td>Excel attribute table from the Registrar General’s Decennial Supplement of 1911 that gives the numbers of births and number of infant deaths (aged under 1) through the 1900s.</td>
</tr>
<tr>
<td>1911 Pop</td>
<td>dBase attribute table from the 1911 census giving the total population and total numbers of men and women taken from the 1911 census.</td>
</tr>
</tbody>
</table>

**Summarized Guidelines for Completing Exercise**

The exercise will introduce you to:

- Manipulating numbers of classes, class intervals, and shading schemes.
- The difference between the Data View and the Layout View. The data view is useful for data manipulation, exploration and analysis. The layout view is where you produce good quality maps.
- Adding titles, legends, north arrows, and text to the map
- Saving the map to a bitmap format so that it can be read into other packages such as Microsoft PowerPoint or Word.
- Advanced topic (if you have time): Data Frames. In the data view these allow you to have more than one view open. In the layer view they allow you to insert more than one map, such as a main map and a detailed insert, on a single page.

**Step-by-Step Instructions**

First you should use Windows Explorer to copy the data for the practical to the folder that you want to work on it in. All the data should be in a the folder `<My folder>\GIS3`.  

*GIS in the Digital Humanities*  
*Session 3: Cartography in ArcGIS*
1. Setting up the map

- Open ArcMap and start with a new empty map window.
- Use the Add data button to add <My folder>\GIS\RD1911 to the map window.
- Use File – Save As... to save this as a new map document. Call if Pop Density.mxd. This name will appear in the top left-hand corner of your ArcMap window on the blue header bar.
- Right click on RD1911 on the Table of Contents and select Properties... from the context menu. Click on the Symbology tab.
- Under Show select Quantities – Graduated Colors. Put Pop_den in the Fields – Value window. Click on Apply. This will produce a basic map as shown below.

- On this map the class intervals and colour scheme have been selected by default. There are 5 classes using Natural Breaks (Jenks) as shown on the right-hand side of the Layer Properties window. These need to be changed to something that produces a better map.
- Click on the Classify... button the right-hand side of the Layer Properties window.
This window gives you a graph of the frequency distribution of the dataset you are interested in. The current class intervals are shown as the blue lines on the graph and are expressed numerically on the right-hand side. The top right-hand corner gives some summary statistics about the dataset. As you can see, of the count of 736 polygons, the graph shows that nearly 600 are very close to the minimum value of 5.72. The maximum value is 49,616.9 but the Natural Breaks option leaves very few values in the higher classes.

Experiment using different types of class intervals by changing the option under Method and different numbers of Classes. To refresh the map you much click OK and then Apply.

Finish with a map that has 5 classes selected using Geometrical Intervals as shown below.
We now need to choose better colours. The easiest way to do this is simply by changing the *Color Ramp*. Try several of these. You need to press *Apply* to refresh the map. Try some sequential colours (where they go from light to dark) and also some diverging colours (for example from green to yellow to red).

You can reverse the colour by clicking on *Symbol* (above the colour boxes) and then clicking *Flip Symbols* as shown above.

The default colour schemes offered by ArcGIS are not very good. Use [ColorBrewer.org](http://colorbrewer.org) to select one that you prefer. In the example below we are using the sequential orange scheme in the top right, shown by the arrow below. Five classes are being used. Clicking on the *CMYK* button (circled) gives the CMYK combinations needed to reproduce this in ArcMap.
Back in ArcGIS, from the Symbology tab of the Layer Properties menu, double-click on the box that shows the colour of the lightest shade shown below.

Under Fill color select More Colors... This will bring up the menu below. If this is not set to CMYK then change it using the pull-down menu in the top right.
Change the values of C, M, Y and K to those given by Colorbrewer. In this case 0, 6, 12, 0.

Click OK and OK to apply these.

Repeat this for the remaining four classes to update all of the CMYK combinations.

Click Apply on the Layer Properties window to update the map in the new colours.

We now want to change the border for all of the polygons (i.e. the registration district boundaries) as the current lines are too thick and too dark. Click on Symbol again and now click on Properties for All Symbols...

Change the Outline Color to a pale shade of gray (e.g. gray 20%) and the Outline width to 0.2. Click OK and Apply to apply these changes.

Finally, click OK to close the Layer Properties window.

Under the main menu bar do File > Save to save these choices to the map window. These classes and colours are now saved.
2. Using the Layout View

- So far, we have been working in ArcMap’s Data View. To produce high quality maps we need the Layout View. To switch to this go to the main menu and use View > Layout View. The map window should change. If your map is not neatly in the middle of the page use the Full Extent button on the tools menu. This should leave it nicely centred in the middle of the window.

- To zoom in and out on the page in the layout view you must use the Layout toolbar (shown below). If this is not visible find it under Customize > Toolbars > Layout. If you use the Tools toolbar (which you have been using on the Data View) this will change the actual position of the map on the page. The Layout toolbar simply changes which part of it you are looking at. To put this another way, zooming using the Layout toolbar changes which part of the map you are looking at, zooming using the Tools toolbar actually moves the map on the page.

![Layout toolbar](image)

- When creating a map it makes sense to save the map window at regular intervals using File > Save.

- To add a title to the map use Insert > Title from the main menu. You can drag the title to where you want it to appear on the page.

- Double-clicking on the title opens the Properties menu. You can change the text here. Change it to “Population density, 1911”. You can also include XML-style tags to modify the formatting, for information on this press the About Formatting Text button.

- Under Change symbol... you can change the text size and font. Change it to 18pt Arial Bold. Click OK to close the window.
We now want to place a North Arrow on the map. Use Insert > North Arrow... to do this. Select the style you want and press OK. Drag it to the place you want. Again, you can change many of its properties by double-clicking on it.

Next we want to put a scale bar on the map. Use Insert > Scale Bar..., select the style you want, click OK, and drag it to the right position.

- You will need to make the scale bar a sensible length and to change the units from meters to kilometres.
- First zoom in on the scale bar using the zoom button from the Layout menu. Changing the length is done by clicking on the scale bar and dragging the blue box that appears at the end.
- Changing the properties is then done by double clicking on the scale bar. In the example below I have change the Division Units from meters to kilometres, changed the Number of divisions to 2, and changed the Label position to Below right.
- Use Zoom to whole page from the Layout menu when you have finished.
Now we want to put a legend on the map. The easiest way is use Insert > Legend and then simply accept all of the defaults on the wizard. Place the legend where you think is appropriate.

- Now double-click the legend, select the Items tab and press Style. (ensure that RD1911 is highlighted under Legend Items). Here you can experiment with different styles. Choose one that you think is appropriate. You may also want to change the title under the Legend tab.

As you can see there are far too many numbers after the decimal places. Rather than 5.725617 it would be better to have 5.7 or 5.73.

- Changing this is done on the Layer Properties window rather than on the legend itself. Right click on RD1911 on the Table of Contents and select Properties... Now under the Symbology tab click on the Label heading and select Format labels... (see below).
- Change the Rounding to 1 or 2 decimal places as you feel is appropriate. Click OK and OK. You will see that this changes both the table of contents and the legend as shown below.

- You should now have a map that looks like the one below (although feel free to make your own design choices).
Session 3: Cartography in ArcGIS

Ian Gregory
The final thing we want to do is remove the black box around the margin of the map that ArcMap has added by default.

- Right-click anywhere on the background of the map to bring up a context menu. Now choose Properties... and select the Frame tab.
- Change the border colour to No Colour and press OK.

Your final map should look like the one below.

![Map with black box removed](image)

- Save the map window.
- Change from layout view to the data view using View > Data view. Can you see the difference? Change back to the Layout view.

### 3. Reading your map into Microsoft PowerPoint.

You can print your map using File > Print (or the printer button) but it is often more useful to embed the map into another document such as Microsoft PowerPoint or Word. To do this you must first export it from ArcMap as a bitmap file such as a JPEG of BMP. These can then be imported into the new program. In this example we will export the file as a JPEG and import it into PowerPoint.

- Go to File > Export map. You can use this window to change the name of the file you are creating, the type of file e.g., JPEG, BMP, PDF, PostScript (EPS), etc. Sometimes you can also change the resolution and other settings. The higher the resolution, better the quality of the image but the bigger the file size. Click Save when you have set the Type to JPEG and selected the appropriate file name, folder and other settings.
Session 3: Cartography in ArcGIS

Ian Gregory
Now open Microsoft PowerPoint and press the New slide button under the Home tab to create a new slide.

Now go to the Insert tab and press the Picture button. And find the file you exported. Click OK. The picture should appear in the slide.

Right click on the picture to find the various options associated with changing the size and properties of the image.

Drag and drop the picture to the appropriate part of the slide.

When you are happy save the PowerPoint presentation and Exit PowerPoint.
4. Advanced option (if you have time): Using data frames

A data frame allows you to add another window to the map in which you can add a new layer or a different view of the same layer. In this example we will add a data frame that has a close-up of Liverpool and Manchester in the north-west of England (or feel free to use London or another part of the country).

- Do File > Save As... and save this new map window as **Pop density2.mxd**.
- Go to Insert > Data frame from the main menu.
- Drag and drop the new window as shown below. You will notice that the new map also appears on the table of contents as **New Data Frame** and that this is in **bold**. This means that the new data frame is active and the other (called **Layers**) is not.

- Use the Add Data button to add **RD1911** a second time. As shown below, it will appear in the active data frame.
It is easier to manipulate this using the Data View. Changing to this (using View > Data view) will give an unshaded map as shown below. Again, you can see that New Data Frame is in bold on the Table of Contents showing that this is the active data frame.

To change the name of a data frame double click on it on the Table of Contents and type in a new name on the General tab as shown below. Change the Name to Detailed map. Click OK.
To swap to the original data frame (Layers) you need to right-click on its icon on the table of contents and choose Activate (see below). This will switch you to the original map.

Switch back by doing the same thing on the Detailed Map data frame. The active data frame is always in bold.

Use the Tools toolbar to zoom in on part of the map such as the area around Liverpool and Manchester (shown below) on the Detailed Map data frame. Change the class intervals to Geometrical Intervals on Pop_den.
Now change back to the *Layout view*. As you can see, the new map has been added to the new data frame giving us a more detailed view of this area.

If you add legend, scale bars, etc, they will be added for the active data frame.
5. Going further (if you have time)

So far we have looked at creating a choropleth map of polygon data. The same principles can be used to create other types of map. Here we want to create a point map of Wordsworth and Coleridge using the shapefiles from the GIS2 exercise. We also want to include background information such as Lakes.

- Create a new blank map document called `Lakes_map.mxd`.

- Use the `Add Data` button to add all of the shapefiles from the GIS2 exercise (Coast, Inland Water Bodies, STC_1802, Urban Areas and WW_Guide.)

```
5. Going further (if you have time)

So far we have looked at creating a choropleth map of polygon data. The same principles can be used to create other types of map. Here we want to create a point map of Wordsworth and Coleridge using the shapefiles from the GIS2 exercise. We also want to include background information such as Lakes.

- Create a new blank map document called `Lakes_map.mxd`.

- Use the `Add Data` button to add all of the shapefiles from the GIS2 exercise (Coast, Inland Water Bodies, STC_1802, Urban Areas and WW_Guide.)

```

- Use these to create a map of Wordsworth’s locations in a different shade to Coleridge’s. **Hint:** The different ways of shading layers are found under the `Symbology` tab of the `Layer Properties` (see above). This can be by `Features`, `Categories`, `Quantities`, etc.

- **Either:** Use the `Categories – Unique Values` option under the `Symbology` tab to shade places Coleridge visited (`Value Field Visited = “Y”`) from those that he didn’t visit or it is unclear.

- **Or:** Use the `Collect Events` tool in ArcToolbox (from exercise 2) on the `WW_Guide` and `STC_1802` layers in turn and create a proportional circle map (`Quantities – Proportional Circles` under the `Symbology` tab)

- Add a legend, scale bar, north arrow, etc. and make the map look good.
Recap

In this lesson you should have learned:

- How to divide maps into classes and give each class the appropriate symbols/shading (using the Symbology tab under Properties which is found by right clicking the layer’s icon on the Table of Contents).
- How to design your own shading schemes using CMYK combinations.
- How to change between the Data and Layout views (under the View) menu.
- How to add legends, titles, north arrows, scale bars, etc. (under the Insert menu when in Layout view).
- How to export the map to a graphics file such as a Jpeg (under File – Export map).
- How to add another Data Frame (section 4) – if required.
Types of data in ArcGIS tables

- In an attribute table each column must be allocated to a specific type of data:
  - **Text:**
    - A text field that will normally be limited to a maximum of 255 characters.
    - You may specify the maximum number of characters that you want the field to have.
    - Values will sort alphabetically so numbers will sort: 1, 10, 11, ..., 19, 2, 20...
  - **Integer:**
    - A whole number (eg. 1, 2, 3, 134 but not 1.5, 2.78, or 134.5).
    - Can be long or short depending on how big the number can be.
    - Trying to put a "number" like "2.4" into an integer field will give an error.
  - **Float & double:**
    - Real numbers (ie decimals) such as 1.00, 1.24, 134.9, etc...
  - **Date:**
    - Self explanatory

Joining tables

### One-to-one

<table>
<thead>
<tr>
<th>ID</th>
<th>County</th>
<th>Pop</th>
<th>ID</th>
<th>Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Lancashire</td>
<td>100</td>
<td>A</td>
<td>Lab</td>
</tr>
<tr>
<td>B</td>
<td>Lancashire</td>
<td>110</td>
<td>B</td>
<td>Con</td>
</tr>
<tr>
<td>C</td>
<td>Cumbria</td>
<td>120</td>
<td>C</td>
<td>Lab</td>
</tr>
</tbody>
</table>

### One-to-many

<table>
<thead>
<tr>
<th>ID</th>
<th>County</th>
<th>Pop</th>
<th>ID</th>
<th>County</th>
<th>Pop</th>
<th>Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Lancashire</td>
<td>100</td>
<td>A</td>
<td>Lancashire</td>
<td>12,345</td>
<td>Lab</td>
</tr>
<tr>
<td>B</td>
<td>Lancashire</td>
<td>110</td>
<td>B</td>
<td>Cumbria</td>
<td>6,789</td>
<td>Con</td>
</tr>
<tr>
<td>C</td>
<td>Cumbria</td>
<td>120</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Types of join

- **Both sides must match**
  - ID County Pop  | ID Party
  - A Lancashire 100 | A Lab
  - B Lancashire 110 | B Con

- **Include all rows from left**
  - ID County Pop  | ID Party
  - A Lancashire 100 | A Lab
  - B Lancashire 110 | B Con
  - C Cumbria 120   | C Lab
Exercise 4: Working with tabular data
Exploring infant mortality in the 1900s

Background
Although people tend to think about GIS as being primarily concerned with mapping. It is better thought of as a type of database that uniquely is able to explore spatial and attribute data together. This exercise will focus on attribute data in tabular form. It will examine a series of data sources from the census and vital registration data. In the first part of the exercise you will explore basic selection using attribute data. In the next part you will explore the many tabular data management and viewing tools for individual tables that are found in ArcMap. In the final part of the exercise you will learn to associate tables in ArcMap by joining tabular datasets to spatial datasets.

Data Sources
These are again the same as the ones we started with, the layers from the 1900s demography map document. In this practical, however, we shall manipulate these in a new map document.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD1911</td>
<td>Shapefile containing Registration Districts in 1911. Also has population density (pop_den) as an attribute.</td>
</tr>
<tr>
<td>Reg_county</td>
<td>Shapefile containing the Registration Counties of England &amp; Wales.</td>
</tr>
<tr>
<td>Admin_county</td>
<td>Shapefile containing Administrative Counties of England &amp; Wales.</td>
</tr>
<tr>
<td>1900s IMRs.xlsx</td>
<td>Excel attribute table from the Registrar General’s Decennial Supplement of 1911. These are listed by registration district (reg_dist) and registration county (GIS_Cnty) and gives the numbers of births (Births) and infant deaths (aged under 1) (I_death) through the 1900s.</td>
</tr>
<tr>
<td>1911 Pop.dbf</td>
<td>dBase attribute table from the 1911 census. It is listed by registration district (reg_dist) and registration county (GIS_Cnty) and gives the year, total population (tot_pop) and total numbers of men and women (tot_male and tot_fem).</td>
</tr>
</tbody>
</table>
Summarized Guidelines for Completing Exercise

The exercise will introduce you to:

- Basic querying of attribute data
- How to manipulate tables of attribute data to import data from Excel, calculate new values, sort data, and find basic statistics about it.
- How to join tables with each other.
- Summarising data.

Step-by-Step Instructions

First you should use Windows Explorer to copy the data for the practical to the folder that you want to work on it in. All the data should be in a the folder `<My folder>\GIS4`.

1. Using queries

- Start ArcMap with a new Blank Document
- Add the Shapefile `Rd1911` from `<My folder>\GIS4` using the Add data button
- Right click on the `Rd1911` icon on the table of contents and select Open attribute table.
- Click on the Table Options button in the top left-hand corner of the Table window and select Select by attributes… (see below).

- Double click on “Reg_Dist”, single click on = and then click Get unique values
Scroll down the list of districts and double click on “Stepney”

The text box should now say “REG_DIST”='STEPNEY’. Click the Apply button and then Close.

Click on the Show selected records button at the bottom of the Table window.

The two selected records are shown in the attributes window (as above)

In addition their polygons are highlighted with a blue outline on the Map window. Under the context menu for RD1911 chose Selection > Zoom to selected features. This will zoom in on the district of...
Stepney which is in central London. You can see that Stepney district is in two parts separated by a narrow corridor or, to put it another way, there are two polygons called Stepney which happen to belong to the same district.

- Chose the Select Features button on the Tools toolbar (circled above) and click on the district that the narrow corridor belongs to. This selects it. What is the name of this polygon? What is its population density?

- Press the Clear selected features button, the one below of the Select Features button on the Tools toolbar shown above.

- Remove Rd1911 from the map by right clicking on its name on the table of contents and selecting Remove. This leaves you with a blank map window.
2. Exploring and manipulating tabular data

ArcMap can open attribute data in a variety of formats including Microsoft Excel, Microsoft Access, text, and Dbase (.dbf). Much of the functionality is only available in Dbase files. Until recently Excel could save spreadsheets as Dbase files but the newer versions of it do not allow this. There are, however, ways of working round this as we will see below. In this session we will open an Excel file containing infant mortality rates (IMRs) and save it as a Dbase file for further work.

- Start Microsoft Excel and open the file **1900s IMRs.xlsx**. You can see that this has four columns:
  - **Reg_dist**: The name of the Registration District
  - **Reg_Cnty**: The name of the county that the registration district is in
  - **Births**: The number of births during the 1900s
  - **I_death**: The number of infant deaths during the 1900s

- These are on a worksheet called **IMRs**. There are two other worksheets, **Sheet2** and **Sheet3** but these contain no data.

- Once you are happy with this close Excel.
In ArcGIS press the Add data button and double click on the **1900s IMRs.xlsx** (do not press the Add button at this stage). The three worksheets will appear in the menu with $ signs after them. Add IMRs$, the worksheet with data in it, either by double clicking on it or by pressing Add. This will appear on the table of contents.

Right-click on the IMRs$ icon and select Open from the context menu. The contents of the spreadsheet should appear.

Right click on the IMRs$ icon on the table of contents and choose Properties... and go to the Fields tab. This window allows you to explore the characteristics of the fields in the database, to change the order that they appear in (using the arrows at the top), and to hide fields. Click on GIS_Cnty and then uncheck the box next to it (see below, left) and click OK. You will see the item disappear from the table (below right). It has not been deleted, it is merely no longer being drawn.
Repeat this process but this time check the GIS_Cnty box to re-draw the field.

Right-click on the I_Death column heading and choose Sort Descending. This sorts the list so the places with the largest numbers of deaths among infants are at the top. Sort Ascending gives the opposite order. Which district had the most deaths? And which the fewest?

Right-click on the I_Death column heading and choose Statistics. This provides some basic statistics about the dataset including that there are 634 rows (count), a total of 1,186,621 infant deaths in England & Wales (sum), that the lowest number of deaths in a district was 30 (minimum) and the highest 26,679 (maximum). The average number of deaths was 1871.6 (mean). There is also a frequency distribution that shows that, like population density, the data were heavily skewed with a lot of districts having a small number of deaths and a few having a lot of deaths.

Click on the red X in the top-right of the window to close this.

We need another column that gives the infant mortality rate, not just the number of births and deaths. To do this we need to add a new column. ArcMap will only allow us to explore Excel files, not to change them. Instead we must first export it to a Dbase file and edit this.
Click on Table Options menu and choose Export... and then use the browse button to save a file called IMRs in your GIS4 folder. You must change the Save as type: option to dBase Table. ArcMap automatically adds the .dbf extension.

When you are prompted Do you want to add the new table to the current map? Click Yes.

This adds a new table to the table of contents called IMRs. Open this and you will see that it looks very similar to the original table except there is a new column called OID. One important difference is that if you click on the Table Options menu on the new IMRs table you will see that Add Field... is available. On the Excel version it is not. Note also that when you have two or more tables open, they appear in the same window, the tab at the bottom allows you to swap between them.
 Remove the Excel version (IMRs$) from the table of contents.

 Now select the Add field option from the Options menu of the IMRs Dbase table. We want to add a field IM_rate of Type Float (which means that it is a real number with decimal places rather than an integer).

 Click OK. A new field will be added and its values defaulted to 0.
Right-click on the heading of the IM_rate column and select the Field Calculator. Click Yes if it gives you a warning.

An infant mortality rate is calculated as the number of infant deaths per 1,000 births. To calculate this fill in the form such that IM_rate = \( \frac{1000 \times [I\_death]}{[Births]} \). Double-clicking on the fields in the top window brings them down to the bottom window.
Click OK.

We now have the new values as shown below. Note that this new field has been permanently added to the Dbase table.
Using what you have already learned find out: How many districts have an infant mortality rate of more than 150? What is their average infant mortality rate? Close the Attribute table window when you have finished.
3. Joining Tables

In the next part of the exercise you will explore the tools that are used to join two or more tables. This ability offers a great deal of potential to humanities research. Typically you may have a set of polygons or points representing administrative units or towns to which you want to join a range of tabular data from other sources. In the following steps you will join the table containing the infant mortality rates to the boundaries of 1911 Registration Districts.

- Add the Rd1911.shp shapefile to the map using the Add data button

- Right-click the Rd1911 icon on the table of contents and choose Open Attribute Table from the layer context menu to open the table. Do the same with the IMRs table.

- Notice that the REG_DIST field in the Rd1911 attribute table appears to have the same values as the REG_DIST field in the IMRs table. To verify that these values are the same, sort both fields by choosing Sort Ascending option from the Options context menu for each field. One thing to notice is that sometimes a Reg_Dist on RD1911 can have more than one polygon so there are 736 rows of data on the Rd1911 attributes but only 634 rows of data on IMRs. This was the case with Stepney as shown earlier.
**IMPORTANT! When you join two tables in a real project, remember the following rules:**

- The names of the fields that you are going to join on (the key fields) in two tables that you wish to join do **not** have to be the same.
- The format of the key fields in two tables that you wish to associate **must** be the same. In other words, a text field will not join to a Long Integer field.
- The lengths of the key fields in two tables that you wish to associate do **not** have to be the same.
- You do **not** have to have the same number of records in two tables to associate them. You only have to have at least some matching values.
- You can have multiple key fields in a table. However, **only one** key field can be used in any table join.

We will now join the **IMRs** table to the **Rd1911** layer attribute table.

- Right-click the **Rd1911** icon on the table of contents and choose **Joins and Relates > Join...** from the layer context menu.
- Complete the options in the **Join Data** dialog box as shown in the following example. This says that we will join **RD1911** to **IMRs** using the field **Reg_Dist** on both tables.
Click the OK button to process this join and to close the Join Data window.

You should now see that the RD1911 table also has the attributes from the IMRs table (NB: on the table below IM_rate is calculated by 100 rather than 1000).
Note that this has not permanently changed either the shapefile or the Dbase table, the tables merely appear to have been joined. If you save the Map Document the join will be saved.

- If you go to Properties on the RD1911 context menu you should be able to draw a map of 1900s infant mortality rates as shown below.

- You can remove a join by going to RD1911’s context menu and going to Joins and relates > Remove joins but do not do this yet.
4. Summarizing data

The IMRs table included a GIS_Cnty field which gives the name of the county that each registration district lies in. In this section we will calculate the average infant mortality rate for each county and output this to a new table.

- Open the attribute table of RD1911.
- Right-click on the column heading GIS_Cnty (a field from the joined IMRs table) and select Summarise...
- Use the + signs to open the lists so that you check the boxes for Sum of **Births** (shown below), Sum of **I_death**, and Average of **IM_Rate**. Then click on the Folders button (circled).

- Put **County_IMRs** as the *Name*, change *Save as type:* to *Dbase Table* and save it to the **GIS4** folder.
Now click Save and OK. When prompted you do want to add the new data to the map (click Yes).

Now open the **County_IMRs** table from the Table of Contents. You will see that it has five fields:

- GIS_Cnty: the name of the County
- Count_GIS_Cnty: the number of rows of data that makes up this county
- Sum_Births: the total number of births in the county
- Sum_I_Death: the total number of deaths in the county
- Average_IM_rate: the average IM_rate for districts in the country. (Note that this is not exactly the same as the county’s infant mortality rate as in this field weights each district evenly. To be correct, the county IMR should be calculated from the **sum_I_death** and **sum_births** fields as this will give more weight to districts with large numbers of births and deaths. You should also note that even this may be wrong because where a district consists of two or more polygons it will be counted more than once).
5. If you have time.

- Remove RD1911 and IMRs from the map.
- Add Reg_County to the map
- Join Reg_County to County_IMRs by joining from the field Reg_County to the field GIS_Cnty.
- Create a map of the average infant mortality rates. How does this differ from the registration district level map.

Although the practical has not explicitly told you how, it has given you enough information to allow you to permanently change data using a join. Once you have joined the data this is done by either:

- To include all data: Right click on the icon of the shapefile you have joined from on the Table of Contents and go to Data > Export to create a new shapefile.
- To create a single column that contains data from the joined table: Add a new field to the attribute data you are joining from before you do the join, then do the join, then use the Field Calculator to update the values in the new field based on the values in the joined table.
Recap

Much of the functionality for manipulating attribute data is found on the menus, buttons and column heading of the attribute tables themselves. This allows you to:

- Select data using attribute values
- Sort records in a table
- Add new fields
- Update field values using the Field Calculator
- Summarize data to create sub-totals

Importing data from Excel (and a range of other formats including Dbase, text and Access) is done using the *Add data* button. The imported table appears on the Table of Contents.

- ArcGIS will only allow you to update values in Dbase files, not Excel, text, etc. To convert to Dbase right-click on the table’s icon on the Table of Contents and use *Data > Export Data*.

Joining data is easy to do in ArcGIS by right clicking on the icon of the shapefile or table that you want to join from on the Table of Contents. This does not change any data and you can remove the join.
Introduction to buffering and overlay

Background

- Basic operations:
  - Buffer
  - Overlay
- Arguably the only functionality that is unique to GIS software
- Fundamental to GIS’s ability to integrate data
- Not really analysis — merely data manipulation
- Tests the quality of GIS data — error propagation

Buffering: The basics

- Input layer: can be point, line or polygon
- Output layer: Contains polygons that enclose all of the study region within a specified distance of the features on the input layer
- Examples:
  - Areas within 5 km of a hospital
  - Areas within 1km of a river
  - Areas within 10km of a lake
Examples of buffering

- Points
- Lines
- A polygon

Overlay: The basics

- Two or more input layers
  - Can be any combination of point, line or polygon
- Spatial data and attribute data are both combined to create an output layer
  - Feature type of the output layer will depend on the combinations of the input layers

Spatial data in overlay
Spatial and attribute data

Clip:
A "cookie-cutter" form of overlay
Input can be points, lines or polygons
Clip coverage must be polygon
Output coverage will be of the same feature type as the input
Features from the input coverage that lie within the clip coverage are preserved
Does not merge attributes

Overlay in ArcGIS

Intersect (points):
Input can be points, lines or polygons
Intersect coverage must be polygon
Only features from the input coverage that overlap with the intersect coverage are preserved
The output coverage contains the merged attributes of both coverages
Overlay in ArcGIS

Intersect (polygons):

Overlay in ArcGIS

Union:
- Input must be polygons
- Intersect coverage must be polygon features from both coverages are preserved.
- The output coverage contains the merged attributes of both coverages

Buffering and overlay combined
- How many farms lie within 2km of a major road
  - Input 1: Road layer (line) with type as attribute
  - Input 2: Farms (point)
Conclusions

- Overlay and buffering are tools that allow us to manipulate spatial data.
- They allow complex spatial queries of multiple layers of data.
- They are data manipulation tools that form an important component of the analytic toolbox.
- Overlay always makes error worse – output layers must be used with care.
Exercise 5: Data Integration

Writers, roads, towns and lakes

Background

One of the key advantages of GIS is its ability to integrate data from different layers based on where the features are found. This uses two main types of operation: overlay and buffering. In an overlay the spatial and attribute data from two or more layers are combined to form a single new layer. There are many different ways of doing this. A union combines all of the data on the input layers to create the output layer, an intersect only combines the data where the input layers intersect with each other, and a clip simply removes features from one input layer that are not enclosed by the second input layer. Buffering takes a single layer of data, which may be points, lines, or polygons and turns them into a polygon layer where the polygons enclose all of the areas within a set distance of the features on the input layers. ArcGIS gives two different ways of implementing overlays: as overlays using ArcToolbox or what it calls spatial joins. In this exercise we will use all of these operations.

The data provided are concerned with Lake District writing. There are point layers containing locations of the place-names mentioned in Thomas Gray’s 1769 proto-Picturesque tour, Samuel Taylor Coleridge’s 1802 tour and a selection of material from William Wordsworth’s Guide to the Lakes. There is also a point layer with the locations of various towns, a line layer showing the major turnpike roads, and polygon layers shows the counties. Various other shapefiles, described below, provide information on physical features.

We will use these data sets to explore the hypothesis that Thomas Gray travelled by turnpike road and concentrated on the areas around the towns. We also explore the theory that Wordsworth tended to concentrate on areas near to his home in Grasmere.
## Data Sources

The data are all in the Map Document **Lakes_overlay.mxd**.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray</td>
<td>A shapefile containing the places-names mentioned by Thomas Gray. <strong>Cal_date</strong> gives the date of the mention and <strong>DoW</strong> the Day of the Week. <strong>Pl_name</strong> is the placename as it appears in the text, <strong>Visited</strong> flags whether he is describing a place he is visiting (<strong>Y</strong>), a place he is describing from a distance (<strong>N</strong>) or whether it is unclear (<strong>U</strong>) which. <strong>St_name</strong> gives a standardised modern spelling of the place name. St_Name and the coordinates associated with it have been taken from the Ordnance Survey’s 1:50,000 gazetteer. <strong>Day_no</strong> gives the days of the mention in sequence from the first day (1) to the last (12).</td>
</tr>
<tr>
<td>STC_1802</td>
<td>A shapefile containing the places-names mentioned by Thomas Gray. The attributes are similar to <strong>Gray</strong>. <strong>Day_no</strong> goes from 1 to 9.</td>
</tr>
<tr>
<td>WW_Guide</td>
<td>A shapefile containing place-names mentioned in parts of Wordsworth’s <em>Guide to the Lakes</em>. Unlike the previous two texts which are descriptions of journeys, the <em>Guide to the Lakes</em> is a tourist guide. As such, it has fewer attributes. <strong>Pl_name</strong> and <strong>St_name</strong> are the original and standardised spellings of the place names respectively as for <strong>Gray</strong> and <strong>STC_1802</strong> while <strong>Sub_Secti</strong> gives thee sub-section of the text that the mention occurs in.</td>
</tr>
<tr>
<td>Lakes_towns</td>
<td>A shapefile containing the names and locations of the major towns in and around the Lake District. <strong>Town_name</strong> is the name of the town, <strong>Resident</strong> names any famous writers that lived there. The source of this shapefile, its selection criteria and accuracy are not given.</td>
</tr>
<tr>
<td>Main_Roads</td>
<td>A line shapefile that shows the major turnpike roads of the period. <strong>Route</strong> gives the start and end points of each segment (‘Br.’ stands for ‘Bridge’). Again, the source, accuracy and selection criteria are not specified.</td>
</tr>
<tr>
<td>Lake_Cnty</td>
<td>A polygon shapefile containing the three counties that made up the Lake District until reorganisation in the 1970s. <strong>County</strong> gives the county name.</td>
</tr>
<tr>
<td>Coast</td>
<td>A line shapefile that gives the coastline of England and Wales plus the Scottish border.</td>
</tr>
<tr>
<td>Lakes</td>
<td>A polygon shapefile containing the major Lakes in the Lake District. <strong>Name</strong> gives the name of the Lake.</td>
</tr>
</tbody>
</table>
Summarized Guidelines for Completing Exercise

The exercise will introduce you to:

- Overlaying different types of data using spatial joins.
- Overlaying different types of data using unions and intersects.
- Creating a subset of data using attribute data.
- Buffering data as a first stage in data integration operations.
- Combining overlaps and buffering to generate new information.
- Using a clip operation to remove features from one layer where they do not intersect with the other layer.
- Creating multi-ring buffers
- Adding areas to attribute tables using calculate geometry

We will also re-visit some functions used in earlier exercises. These include:

- Querying attribute data
- Summarising attribute data
- Adding fields to attribute tables
Step-by-Step Instructions

Please note that by now you should be beginning to understand the basic functionality of ArcGIS. While this practical, and the ones that follow, talk you through the detail of the operations that you need to perform to learn operations such as overlay, you should use your initiative about things such as turning layers on and off, zooming in and out, saving map documents, and so on.

All the data should be in a the folder <My folder>\GIS5.

- Open the map document Lakes_overlay.mxd.

1. Basic overlay operations: Adding county names to the layers

In this section we will find which county each point in the three writers’ shapefiles: Gray, STC_1802, and WW_Guide lies using three slightly different approaches: union, intersect and spatial join. Union and intersect are found in ArcToolbox. A union copies all features to the output layer, an intersect only copies those features from the input layer that overlap the intersect layer. Spatial joins use a different user interface and allow you to do overlays in various different ways. The main aim of this section is simply to explore the different ways that these three approaches to overlay work.

a. Using a spatial join to overlay STC_1802 and Lakes_Cnty

- Right click on the STC_1802 icon on the Table of Contents and go to Joins and relates > Join....

- The top drop-down menu on the dialogue box is usually Join attributes from a table (see above) which provides options for a conventional relational join. Change this to Join data from another layer based on spatial location to do a spatial join.
Fill in the remaining dialogue as shown below. This means that **STC_1802** will be spatially joined (overlaid) onto **Lake_Cnty** to create a new shapefile called **STC1802_Cnty** (use the *Browse* button to set this new layer name). This new shapefile will have the points from **STC_1802** with the attributes from **Lake_Cnty** added to them based on which polygon they fall inside.

![Join Data](image)

- Press **OK**
Once the tool has run a new shapefile called **STC1802_Cnty** will be added to the Table of Contents and the map. It will have the same locations as the original **STC_1802** but if you open its attribute table you will see that the attributes from **Lakes_Cnty** including **FID_2** which is just an internal ID number and, more usefully, **County**.

How many of places did Coleridge mention in each county?

**b. Intersecting Gray and Lake_Cnty**

Although not described as such by ArcGIS, the previous overlay was a union as all of the features from the input layer were copied onto the output layer.

With **STC_1802** some points from the input layer lay outside the three counties and thus the **County** field in **STC1801_Cnty** was blank for these points. If we didn’t want features that lie outside these counties to be included on the output layer we use an intersect instead. This adds the attributes from the intersect layer to the output layer but only copies over the features that lie within the intersect layer. We will explore this using **Gray**.

- Right-click on the **Gray** icon on the Table of Contents and use **Zoom to Layer**. You will see that Gray talked about some places that were a long way from the Lake District, including London, Warwick and the River Thames (represented by a point) in the south, and Auckland in the east (ensuring that the **Coast** layer is turned on may help with this).

- We want to remove the points that lie outside **Lake_Cnty** and add the county names to the output shapefile. For this we use the **Intersect** tool. Open this tool, it is in ArcToolbox under **Analysis Tools > Overlay**.

- Drag the icons for **Gray** and **Lake_Cnty** into the **Input Features** box (they will move to the lower box).

- Use the **Browse** button (circled below) to set the output layer name to **Gray_Cnty**.

- Ensure that **Join Attributes** is set to **All** and the other options are set as below.
Press OK. The tool will take a few minutes to run.

Once it has finished Gray_Cnty will be added to the map. Compare it and its attributes to the original Gray. Use Zoom to Layer on Gray_Cnty to help with this.

Now add county names to WW_Guide using whichever method you prefer to create WW_Cnty. You decide which method to use and whether you keep the points outside the counties or not. You can skip this stage if you are short of time.

2. Did Gray visit places near to towns?

In this exercise we will put a buffer around the towns to define what is ‘near’ to a town. We will then use an overlay operation to find out which points are ‘near’ to town and which are not. Before we can do this we need to do two things: define ‘near’ and extract the places that Gray actually visited from all of the places that he mentioned.

Ultimately defining ‘near’ is an arbitrary choice. As well as your common sense you should also think about the accuracy of the data. The Gray points were georeferenced using a gazetteer that specifies which 1km grid square the place name occurs in on Ordnance Survey 1:50,000 scale maps. This means that even if the place name is in the correct grid square it still may be inaccurate by ±500m in both x and y. For this reason alone it probably makes little sense to define ‘near’ as less than around 1km, below this any variations may simply be caused by data accuracy issues. A further question is whether the distance that represents near to a town is the same as near to a road, and so on. Come up with your definition(s) of ‘near’ in miles or kilometers (you can change them later if required).
To answer questions such as ‘did Gray mainly visit places near to towns?’ we first need to select the places that he visited. Next we need to create a buffer to define ‘near’ to towns. Finally, we overlay Gray’s locations and the towns buffer, and explore the results.

a. Extracting places that Gray visited

- Using the techniques you learnt in the previous exercise select all of the points in the Gray layer where visited is equal to “Y”. (The practical describes using Gray, however you can use Gray_Cnty if you prefer.)

- Right click on the Gray icon on the Table of Contents and select Data > Export Data

- Make sure that Export: is set to Selected features and use the Browse button (circled above) to save this to a shapefile called Gray_visited in your GIS5 folder.

- Press OK and when you are asked Do you want to add the exported data to the map as a layer? press Yes.

The new layer will appear on your table of contents.
Use the Clear Selected Features button to clear the selection on Gray.

Uncheck Gray and any other author data so that it is no longer drawn and change the symbology of the Gray_visited layer if required.

The map window should look similar to the above.

You now have a new layer, Gray_visited, that only contains the locations that Gray visited.
b. Putting buffers around the towns

Buffers are created using an ArcToolbox tool. This is found under Analysis Tools > Proximity > Buffer. It can also be reached under the Geoprocessing menu at the top of the screen.

- Open the Buffer tool.

- Set the Input Features to Lakes_towns, and use the Browse button to set the Output Feature Class to a shapefile called Towns_buff. It may make sense to also include the buffer size in the name, for example Towns_buff5m if you are using five miles for ‘near.’

- Set the Distance and linear unit to the distance that you have decided will represent ‘near’ to a town.

- Finally set the Dissolve Type to All. This is important as it means that overlapping buffers will be joined together rather than creating lots of separate overlapping polygons.

- Press OK.
The tool can take a few minutes. At the end it will add the new buffer layer to the map and the Table of Contents. It will look something like the above, however the details of the shapes of the buffer will vary depending on the buffer size that you selected.

If you look at the attribute table of the buffer shapefile you will see two things: first the shapefile consists of only one polygon (click on any of the buffers using the Identity tool if you want to confirm this). This is a multipart polygon which allows a single ‘polygon’ to cover more than one area. (If you want to turn this into the more usual singlepart polygons – there would be eight of these in the above – there is an ArcToolbox tool for this under Data Management Tools > Features > Multipart to singlepart but you don’t need to do this now). Second, there is only the bare minimum of attributes.

We want to add a new attribute to the shapefile for when we use it in further analyses. This will be called Nr_Town to flag that it encloses places near to towns. Using techniques from previous practicals add this field as a text field that is one character long. Use the Field Calculator to set this to ‘Y’ (don’t forget the quotes) for the one polygon as shown below.

c. Overlaying Gray_visited and the towns buffer to answer the question ‘Did Gray mainly mention places near to towns?’

Using whichever type of overlay you prefer, overlay Gray_visited and the buffer layer you created in the previous stage to create Gray_towns. This new shapefile should contain all of the points that Gray visited with attributes that include whether they are near a town or not.
How many of the places Gray visited were near a town?

3. Did Gray mainly visit places near to turnpike roads?

The shapefile Main_Roads holds the main roads through the Lake District. Placing a buffer around lines is similar to placing one around points. Repeat the process above to answer the question ‘did Gray mainly visit places near to turnpike roads?’ In answering this start with Gray_towns. By the end of the process, you should have a new shapefile that has attributes from both Nr_town (which you have already added) and Nr_road. Think about how far a buffer that defines ‘near’ to roads should be.

How many points did Gray visit that were not near either a road or a town?

4. Advanced if you have time: Did Wordsworth write mainly about places near his home?

One of the major problems in the use of buffer so far has been the binary definition of ‘near’ in which locations are either within an arbitrary threshold distance that says that they are near, or they are not. This is far from satisfactory. A way of avoiding this is to use multiple buffers to allow a more sophisticated understanding of whether things decline with distance. The Multiple Ring Buffer tool allows this to be implemented.

Wordsworth lived at Dove Cottage in Grasmere, did he favour places near Grasmere in his Guide to the Lakes more than those further away? In this section we will investigate this.

- You will need WW_Guide, Lakes_Cnty and Lakes_towns to start with. Turn these layers on and any you don’t want to see off.

- The first stage is to identify where Grasmere is and set up a shapefile that just has Grasmere in it. Using approaches used earlier, select Grasmere from the Lakes_towns shapefile (it is flagged by having Wordsworth in the Resident column) and export it to a new shapefile called Grasmere.

- If you look at the WW_Guide shapefile you will see that some of the points are Wordsworth describing routes to the Lake District. Arguably these should be removed and we should only include places within the three counties in and around the Lake District. Do this using a Clip (Analysis Tools > Extract). Clip WW_Guide with Lake_Cnty (the Clip Features) to create an output shapefile called WW_clip.

- Now we need to create the buffer. In ArcToolbox go to Analysis Tools > Proximity and open the Multiple Ring Buffer tool.

- The input feature is the Grasmere shapefile and you want to create a shapefile called Gmere_rings.

- You also need to decide on the size of the rings to create, in the example below a range from 2 miles to 20 miles has been used (on a first attempt you might use fewer rings than
this as it can be slow). Enter the distances by typing the numbers into the Distances box and pressing the + button.

- You also need to set the distance unit to miles (or kilometers if you prefer) – note that the default for British National Grid is meters.

- Leave the Field Name as Distance and the Dissolve Option as All.

- Press OK. This will take a few minutes.

The result should look like the below. If you look at the attribute table you will see that the distances are stored against each polygon.
It could be that later in the analysis we want to calculate the density of mentions (mentions per square kilometer or per square mile). To allow this to happen we might want to add a field that contains the area of each ring. To do this:

- Open the attribute table from **Gmere_rings** and add a new field called **Area** of type **Float**.
- Right-click on the heading of the **Area** field and click on **Calculate Geometry**.
Ensure that the Property is set to Area and change the Units to either Square Miles or Square Kilometers (your choice).

Press OK. The field will be updated to hold the areas of each polygon (buffer).

Now do an overlay (you choose which one) to find out how far each of the WW_clip points is from Grasmere. Once you have done it the attribute table of the shapefile should look similar to the one below. Note that any points that fall outside the largest buffer will be given a distance of 0 – you may want to change this to a large number (bigger than the largest distance).

Finally do a Summarize on distance to find out how many points lie in each buffer. If you output the results of this to a text file and read them into Excel you should be able to produce a graph like the one below.
Does this support the theory that Wordsworth paid more attention to locations near his home?

If you have time: Repeat the *summarize* operation but add the *Average* of *Area* to the results. In Excel use this area to calculate the density if place-name mentions in each buffer (rather than just the count. Density = PopCount/Area). How does this compare to the result above?

5. Further work (if you have time)

Set yourself some questions and see if you can answer them – if you don’t have time to actually carry the operations out at least think about how you might answer them. Questions might include:

- Did Coleridge mainly visit places near to his home?
- Did Wordsworth’s Guide to the Lakes concentrate on places near to roads?
- Did Coleridge visit places near to towns? Who had the higher percentage of visits near to towns – Gray or Coleridge?
- Which of the three writers concentrated the most on places near to lakes?
- The railway first arrived in the Lake District in 1847 at Windermere station which is just outside Bowness. This part of the Lake District is now visited extensively and Bowness is one of the largest towns in the Lakes. Based on Wordsworth’s Guide to the Lakes, which was written before the railways, does the area around Bowness seem to have been more or less popular than the area around Grasmere or the area around Keswick?
Recap

Buffering and overlay allow data to be integrated in a range of different ways but can lead to a bewildering variety of shapefiles many of which are just intermediate stages. Being careful with filenames can help.

- You can extract a subset of features to a new shapefile by selecting them and using Data > Export Data.

- Buffering is available from ArcToolbox. Don’t forget to set Dissolve Type to All and make sure that the units you are using are sensible (it is easy to create a 5 meter buffer when you mean 5 miles or 5 kilometers).

- There are a variety of types of overlay that can be implemented either with a Spatial Join (by right-clicking on the icon of the shapefile you want to join from in the Table of Contents), or using ArcToolbox.

- To add the area of a polygon to its attribute (or the x or y coordinates of its centroid) create the field you want to put these values in as a floating point number and use Calculate Geometry which is found by clicking on the column heading in the attribute table.
Map Projections

- Convert from Earth (a globe) to a flat piece of paper/computer screen
- All projections distort one or more of:
  - Shape
  - Area
  - Angles (bearings)
  - Distance
  - Distortions increase as you move away from the prime meridian

Examples:
- British National Grid
- Irish Grid
- Universal Transverse Mercator
- Robinson

Examples of projections

- Mercator:
  - Preserves shape
  - Distortion increases with distance from equator:
    - Greenland on map is same size as Africa
    - In reality, it is 1/13th the size
    - Poles appear to be the same length as the equator

- Gall-Peters or Peters:
  - Preserves area
  - Distortion increases with distance from 45°N & S

Do map projections matter?

“Kim’s missile is nuclear warning shot”, Sunday Times, 16th Dec 2012 p. 30

Mercator Projection

Polar Stereographic Projection
**British National Grid**

- Transverse Mercator projection with origin near Isle of Wight
  - Minimal distortions
- False origin south-west of the Scilly Isles
- Coordinates measured in meters from the false origin
  - All locations on mainland Britain are:
    - east and north of here
    - within 1000km
  - So. Any location can be expressed as a positive number of six figures or less
  - Eg:
    - Sca Fell (320656,506438)
    - Trafalgar Square (530013, 180419)

**Confusing factors**

- 100km grid squares
  - NY is 300km east and 500km north
- Six figure grid references
  - NY123456
  - Is 321320 east and 545600 north
  - Sca Fell (was 320656,506438) in NY206064

**Latitude and Longitude**

- Measure any location on the Earth's surface:
  - As an angle
    - Degrees, minutes and seconds
    - North/south of the equator (90° is the max)
    - East/west of the Greenwich Meridian (180° is the max)
    - Lancaster is 54°02' 54"N, 2°47' 58"W
  - Decimal degrees:
    - Computers don't work well with degrees, minutes and seconds
    - Divide minutes by 60, Divide seconds by 3600
    - Lancaster is 54.0496 DD in y, -2.7987 DD in x
- WGS1984 is the geoid used by the GPS satellites and Google Earth
Conclusion

• Once you have projected your data:
  – Calculate distances, areas, etc. (Calculate Geometry)
  – Integrate data
  – Change projections
    • On the map view
    • To create a new layer (the Project ArcTool)

• Further information
  – See the reading list
  – OR ArcGIS Help:
    • Essentials Library > What is GIS? > Georeferencing and Coordinate Systems
Exercise 6: Geo-referencing point data:

The use of gazetteers

Background

Place names are the most widely used way of expressing a location. They occur in a wide range of different types of sources, from quantitative sources such as census data, where each row of data is referred to by place name, to unstructured texts such as a description of a journey or an official report. As such, place names are extremely important to humanities geographies. One of the major challenges in building a GIS database is often to convert place names into the precisely defined, coordinate-based locations required by spatial data in GIS.

Sometimes place names refer to a specific and precisely defined administrative unit such as a municipality, district or county in census data. In these cases polygons are usually the most suitable way of representing the places that the names refer to. More often, however, the location a place name refers to is vaguer. In phrases such as “Scarlet fever was prevalent in Manchester and Salford” or “Helm Crag is a high mountain near Grasmere” (place names are shown in italics) the places that the text is referring to are not precisely delimited, indeed, Grasmere can refer to either a lake or the village next to it. In these cases a point is usually a much more satisfactory and convenient way of representing the location. The challenge, however, is to find coordinates for the place names.

Gazetteers can be particularly useful, if not crucial, for performing this task. A gazetteer can be thought of as a sort of ‘geographical directory’ that provides basic reference information about the geographical makeup of a country, a region, or even a continent. From a GIS perspective, a gazetteer must contain a minimum of place names and coordinates for those places. It may also contain additional information such as population size, areas, heights, feature types (town, mountain, river, etc.), and so on. Gazetteers are becoming increasingly available for free. In this exercise we will use the Geonames gazetteer (www.geonames.org), however, many other gazetteers are also available - the Course Overview introduces some of them.

Google Maps, Google Earth and other online mapping tools also have large gazetteers behind them as part of their search facilities. These are not available to download but can be used to locate places individually if they are not in one of the publically available gazetteers.
Data Sources

The aim of this exercise is to georeference two tables. The first is taken from Coleridge’s 1802 tour and already has coordinates. The second is taken from Wordsworth’s poem *To Joanna* and does not have coordinates. To add these we need to use the Geonames gazetteer. This will be joined to the table to find the coordinates. We then need to find places that fail to match and correct errors. For the full text of *To Joanna* see: [http://www.poemhunter.com/poem/to-joanna](http://www.poemhunter.com/poem/to-joanna).

The map document **Lakes_points.mxd** contains many of the background shapefiles.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB.txt</td>
<td>The UK data from Geonames</td>
</tr>
<tr>
<td>readme.txt</td>
<td>The readme file that accompanies GB.txt</td>
</tr>
<tr>
<td>GB.zip</td>
<td>The zipfile from Geonames from which the two files above have already been extracted</td>
</tr>
<tr>
<td>admin1CodesASCII.txt</td>
<td>A look-up table from Geonames that provides metadata for the admin1_code field.</td>
</tr>
<tr>
<td>To_Joanna.txt</td>
<td>A tab-delimited text file that contains the lines containing place names from Wordsworth’s poem <em>To Joanna</em>. It has three columns: <strong>line_no</strong> (the line number from the poem), <strong>line</strong> (the line), and <strong>pl_name</strong> (the place name).</td>
</tr>
<tr>
<td>Cole_other.xlsx</td>
<td>An Excel spreadsheet giving the places that Coleridge mentions in his 1802 tour. <strong>Pl_name</strong> the place name as it appears in his text. <strong>Visited</strong> whether he visited it (Y), talks about it from a distance (N), or if this is unclear (U). <strong>Date</strong> and <strong>cal_date</strong> the date in two different formats. <strong>St_name</strong> a standardised version of the name as taken from a gazetteer. <strong>Latitude</strong> and <strong>Longitude</strong> taken from a gazetteer.</td>
</tr>
<tr>
<td>Lakes_towns</td>
<td>A shapefile containing the names and locations of the major towns in and around the Lake District. <strong>Town_name</strong> is the name of the town, <strong>Resident</strong> names any famous writers that lived there. The source of this shapefile, its selection criteria and accuracy are not given.</td>
</tr>
<tr>
<td>Main_Roads</td>
<td>A line shapefile that shows the major turnpike roads of the period. <strong>Route</strong> gives the start and end points of each segment (‘Br.’ stands for ‘Bridge’). Again, the source, accuracy and selection criteria are not specified.</td>
</tr>
<tr>
<td>Lake_Cnty</td>
<td>A polygon shapefile containing the three counties that made up the Lake District until reorganisation in the 1970s. <strong>County</strong> gives the county name.</td>
</tr>
<tr>
<td>Lakes</td>
<td>A polygon shapefile containing the major Lakes in the Lake District. <strong>Name</strong> gives the name of the Lake.</td>
</tr>
</tbody>
</table>
Summarized Guidelines for Completing Exercise

First we will georeference the Cole_other table and convert it into a shapefile. This is straightforward and using the Display XY data option.

Next we georeference the To_Joanna table. This is more complicated. To do it we need to:

- Read To_Joanna and the gazetteer into ArcGIS.
- Join them together using a join.
- Update latitude and longitudes on To_Joanna where they match.
- Update spellings to cope with non-matches due to spelling variations.
- Add other coordinates manually.
- Convert the results into a shapefile.
- Use the Editor toolbar to tidy up the results.

The Geonames gazetteer

Geonames is on-line, freely available gazetteer, with which it is possible to view and download a variety of basic information about places (see www.geonames.org). The majority of place names in the gazetteer come from a few sources. These include the Geographic Names Information System of the US Geological Survey for the USA, www.geobase.ca for names in Canada, and the Board on Geographic Names of the US National Geospatial-Intelligence Agency for most names outside of North America. The database also contains a variety of other data from a variety of other sources. In addition, it is also an ‘open’ development, meaning that users may edit, correct and add new information to the existing database. At the time of writing this practical, the Geonames database is reported to contain over 8 million geographical names, from all around the Earth. Geographical coordinates for locations are given in latitude / longitude measurements in the WGS84 (World Geodetic System 1984) geographic coordinate system.
The main user interface to Geonames consists of a simple menu, as shown below:

This interface design is intended to aid users who wish only to look up information on a particular place; however, it is also possible to download all of the data for a particular country. In this practical, you will use the Geonames ‘country file’ containing entries for the UK, GB.txt. This has already been downloaded for you. Other country’s data is available by following the link to Free Gazetteer Data under Download.
Step-by-Step Instructions

1. Georeferencing a table that contains coordinates

The file Cole_other.xlsx is an Excel spreadsheet that contains the places mentioned in Coleridge’s text together with some other information and, crucially, the latitude and longitude that the place names refer to. (It doesn’t matter whether these files are Excel spreadsheets, database tables, or text files such as csv or tab delimited, when you have information that has coordinates, reading it into ArcGIS, georeferencing it and saving it as a shapefile is easy.)

a. Check the file in Excel

- Open Cole_other.xlsx in Excel and have a quick look at it.
  - You will notice that the file contains two versions of place name. The first, pl_name, is the spelling from the text, the second, st_name, (for standardised name) is the spelling that appears in the gazetteer that was used to georeferenced this. The file also contains latitude and longitude, and a range of other data.

- When you are happy, close the file in Excel.

b. Open the file in ArcGIS

- Start ArcGIS and open the Lakes_points.mxd map document. You will see that it contains a range of background data.
- Use the Add data button to add the Cole_other.xlsx spreadsheet to the Table of Contents.
- Open the table to make sure it has read in properly. It should look similar to the below:
c. Georeference the file to create an Events Layer

- Right-click on the Cole_other$ icon on the Table of Contents and click on Display XY data...
- Set X Field to Longitude and Y Field to Latitude
  - The way to remember that longitude is X and latitude is Y is that polar areas are at high latitudes and thus must be Y.
- Press the Edit... button and set the coordinate system to WGS1984. This is found under Geographic Coordinate Systems > World
  - Important: Options similar to WGS1984 are available under Projected Coordinate Systems > World however these give different results. Do not use them unless you are sure you want to. WGS1984 is Geographic notProjected.
- The menu should look as below. Press OK.
Two things should happen: **Cole_other$ Events** should appear on the Table of Contents and the points should appear in the map window. This should look as shown below
d. Export the Events Layer to a shapefile

It is important to note that an Events Layer is not a shapefile – it is simply an alternative way of looking at the table and has quite limited functionality associated with it. The final stage is to save the Events Layer as a shapefile and remove the Events Layer from the Table of Contents. To do this:

- Right-click on the **Cole_other$ Events** icon on the Table of Contents and go to *Data > Export Data*

![Image of Export Data dialog]

- Save this as **Cole_other** in your **GIS6** folder

![Image of saved data]

- Press *OK* and click *Yes* when asked whether you want to add the exported data to the map as a layer.

- Finally, remove both the **Cole_Other$ Events Layer** and the **Cole_other$ table** from the Table of Contents – these are no longer needed and can cause confusion.
Your map document should now look similar to the below. You will be able to treat the **Cole_other** shapefile like any other shapefile including querying it using the *Identify* button, opening its attribute table, selecting by attributes, changing symbology, and so on.
2. Georeferencing a table that does not contain coordinates

The file To_Joanna.txt is taken from Wordsworth’s poem of the same name. The file contains the lines that have a place name in them along with the line number and a column that just holds the place name.

a. Explore To_Joanna in Excel

- Open Excel.
- Go to File > Open.
- In the bottom-right hand corner of the menu change All Excel Files to Text Files (see below).

- Click on To_Joanna.txt.
- Press Open.

Now work through the wizard that Excel uses to open text files:

- On the first screen the button next to Delimited should be selected. Then press Next>.
- The Tab box under Delimiters should be ticked and Text qualifier should be set to . Press Next>.
- Press Finish.

The file (with a little bit of changing column widths) should look as below. If it does then close the file – you do not need to save it.
b. Read the file into ArcGIS

Use the Add Data button to add To_Joanna.txt to the Lakes_points.mxd map document that you were working with earlier in the practical. Open the table to check that it has read in properly.

As you know the file has no coordinates. These will be added by joining data from the Geonames gazetteer. To do this you need to add three new fields to the table: one for a standardised place name, and one each for latitude and longitude. You cannot edit a text file so:

- Right click on the To_Joanna.txt icon on the Table of Contents and go to Data > Export.
- Export it to a dBase Table (you will need to change this under Save as type) called To_Joanna. When prompted you do want to add the new table to the current map.
- Remove the original To_Joanna.txt from the Table of Contents to avoid confusion.

You now want to add the three fields.

- Open the To_Joanna table
- Under the menu in the top-left of the table (shown below) go to Add field...
• Name the field St_Name (for standardised name), it should be of Type “Text” and Length 30.
• Add two more fields: Latitude and Longitude which should be of type “Float”.

Finally, we assume that the initial values of St_Name should be the same as the values of Pl_Name:

• Use the Field Calculator to set make the values in St_Name equal the values of Pl_Name.

Your table should now look like the below.

  o Note that ArcGIS has defaulted Latitude and Longitude to zero. If these were not updated it would result in all of these places being georeferenced to a location in the Atlantic somewhere south-west of Nigeria (where the Greenwich meridian crosses the Equator).  
c. Explore the Geonames metadata

The gazetteer data from Geonames is difficult for the first-time user to understand. The contents of the readme.txt file, which was downloaded from the Geonames site, as part of the GB.zip file, is useful to understand both the content and structure of the gazetteer data file. Open the readme.txt file using Wordpad. To do so, you need to right-click on the file name in Windows Explorer, and then choose Wordpad from the menu, as shown below:

The file should then open in Wordpad, looking like this:

Read down quickly through the entire file. The important points are:

1. The table contains tab-delimited data
2. The field names in the main ‘geoname’ table are:

The “main ‘geoname’ table” referred to is in fact that gazetteer file for Britain, GB.txt, that has previously been downloaded and extracted from GB.zip. (This may not be clear to you yet, but it should become so shortly). Other points to note here are as follows:

- Each row refers to a field of information (see above). So for example ‘geonameid’ is the heading for the first field.
- In addition to the field heading, the information after the colon (:) provides some additional metadata about the content of the field. Thus, we are told that the entries in the ‘geonameid’ field are integer (numerical) identifiers.
- Counting down the rows, you will see that there are 19 separate fields in the gazetteer file. The most important ones are:
Three fields that contain place names (‘name’ is the basic one ‘ascii name’ and ‘alternatenames’ given variants on this)

- Coordinate information (‘latitude’ and ‘longitude’)

Finally, note also that the latitude / longitude information are recorded as decimal degrees (WGS84).

Leave this file open in the Wordpad window, as you now go on to look at the gazetteer file. Excel should open the GB.txt file in a spreadsheet format.

d. Explore the Geonames data file

Use Excel to open the file GB.txt in the same way as you opened To_Joanna.txt in the previous section. Remember that it is a text file and that it is delimited using tabs. Once it is opened it should look as below. You may want to widen some of the columns to get rid of the “###” effect.

There are two things to notice. First, the file does not contain column headings, you will need to add these as described below. Second, the file contains a lot of fields that you may not want – effectively everything after latitude, so from column G (“Feature Class” according to the metadata) onwards.

Although it is not essential, if you want to, it saves time to delete many of the unneeded fields. To do this highlight the fields by clicking on the heading for column G and dragging across to column Q. All of these columns should be highlighted. Now right-click on the column heading and click Delete on the menu that appears.

You now need to add a row containing column headings:
You will first need to insert a new top row into the Excel file. Do this by highlighting the top row, right clicking, and clicking *Insert*.

Either: Type the column names into cells A1, B1, etc. by hand. This is the safer option – see below under “Field names in ArcGIS”.

Or: Copy each of the column headings from *readme.txt* and paste them into the appropriate column as shown below. Be very careful not to copy over any spaces at the start or the end of the name.

- If you did not delete the data after the coordinates you will also note that the field before Country Code (which is a combination of feature code and feature class) is undocumented. Give it whatever name you want.

Field names in ArcGIS:

1. ArcGIS is not good at handling field names that contain certain characters, such as spaces. In the present case, this means that it would have difficulty in reading, for example, ‘field class’ (if you have not deleted it) as the title for one field. The easiest way to deal with this difficulty is to edit field names containing spaces, to remove the spaces – so, for example, ‘field class’ would be edited to ‘field_class’ or ‘fieldclass’. You also need to be careful that there are no spaces at the start or the end of the field...
name and that the field name does not start with a number or piece of punctuation. Spaces at the start and end can be a frustrating problem that is hard to spot.

2. ArcGIS does not like long field names. It may be worth truncating the names to less than around 10 characters. For example, ‘alternatenames’ might be better truncated to ‘altnames’.
   - Finally, save the file as a text file – a CSV file is often easier to find errors in than a tab delimited file, you could also use an Excel file if you prefer. Call the file **GB_Gaz**.
   - Close the file in Excel.

e. Read the Geonames data into ArcGIS and use it to update To_Joanna

   - Use the Add Data button to add the **GB_Gaz** file to the **Lakes_points** map document.
   - Open the file to ensure it has loaded correctly.
     - If you get an error it usually means that you have errors in your field names. Remove **GB_Gaz** from ArcGIS, open it in Excel or a text editor and see if you can correct the column names. Once you have done this, save the file, and close it in Excel before trying to add it to ArcGIS again.

   Once you are happy that the gazetteer has been read in correctly you need to join from **To_Joanna** to the **GB_Gaz** gazetteer file:

   - Right-click on the **To_Joanna** icon on the Table of Contents.
   - Go to Joins and Relates > Join…
   - You want to join from **St_Name** to the **GB_Gaz** field name. Press OK.

   If you have joined this correctly it should look as below. Note that <Null> means that the cell is empty. As you can see the join has not been entirely successful. How many rows have been successfully joined?

   ![Image](image.png)

   You want to set the values of **Latitude** and **Longitude** in the **To_Joanna** file to the values found in the gazetteer (they are currently 0).

   - Use the **Field Calculator** to do this. It will give warnings because many of the records have failed to join.
   - Now remove the join.
The **To_Joanna** table should now look as shown below.

![Table Image]

### 3. Updating values and correcting errors

This means that we still have a number of places for which we do not have coordinates. There are three possible strategies for finding these and updating the locations:

a) The place name is in the gazetteer but has a different spelling.

b) We can get the coordinates from a different source and update the table.

c) We move points on the map by editing a shapefile.

The remainder of the practical will explore these three approaches.

#### a. The place name is in the gazetteer but has a different spelling.

For a relational join to work the two fields must match exactly. This means even minor variations such as: different use of capital letters; hyphens against spaces; using or not using dots for abbreviations; or including or not including apostrophes will cause a join to fail to match. There may also be more significant variations in spelling. Where we find these variants, the **st_name** field is updated so that the join will match them.

- Re-open the gazetteer in Excel – it doesn’t matter which version of the gazetteer file you use.
- Highlight column B – the **name** column.
- Press **Ctrl-F** for **Find** or go to **Find & Select > Find** on the **Home** tab.
- Type “Kirkstone” into the **Find what** box (or try “Kirkston” or other variations – note that Excel **finds** are not case sensitive).
- It should find “Kirkstone Pass” as shown below.
Is “Kirkstone Pass” the “Kirkstone” that the poem is referring to? The answer to this is ultimately a judgement call based on your knowledge of the source. One reason for using `st_name` rather than simply updating the original spelling in `pl_name` is that it allows you (or other people) to see what you have done. Let us assume that we think that the coordinates for “Kirkstone Pass” do provide a realistic location for “Kirkstone” we now need to update the value of `st_name` for “Kirkstone” to be “Kirkstone Pass”:

- On the `To_Joanna` table select the “Kirkstone” row either by clicking on the left-hand end of the row, or by using `Select by Attributes`.
- Right-click on the `St_Name` column heading and use the `Field Calculator` to set the value to “Kirkstone Pass” (you need to include the double quotes and be careful of the capital letters).
  - Note: when you do this it is important that you have selected the row you are interested in. If you overlook doing this all values of `st_name` will be updated to “Kirkstone Pass”. A safer way of doing this is to use the `Editor` toolbar as described below.
If you were doing this for real, you would repeat this process for the other places that had not matched. Where the find does not work it is worth repeating it on the ‘alternatenames’ and ‘asciinames’ columns. For now, however, we will carry on to explore the other approaches.

The final stage, once you have updated all of the values of st_name you are going to do (in this case probably only ‘Kirkstone’) is to update the values of latitude and longitude as described above. This involves:

- Join To_Joanna to GB_Gaz from st_name to name.
- Use the Field Calculator to update the values of latitude and longitude.
  - Before you do this you might want to use Select by Attributes to select rows with a latitude of zero – ie. places that have not yet been allocated a latitude and longitude. If you do not do this, all values will be updated which will not cause a problem unless values of latitude and/or longitude have been updated in other ways such as those described below.
- Remove the join.

This should have successfully added coordinates to “Kirkstone”.

b. Getting coordinates from a different source.

Some place names will inevitably not be in your gazetteer and will have to be found elsewhere. Google Earth and Google Maps are useful sources of other coordinates. Google Maps is perhaps the easiest to get co-ordinates from.

“Glaramara” is not in the Geonames gazetteer so we will find its coordinates using Google Maps.

- In a web browser go to: http://maps.google.com.
- Type “Glaramara” into the search box in the top-left corner (do not use the auto complete options that appear). Press <Return>.
- You will be zoomed into a point as shown below – you may want to zoom out to check you are in approximately the right location.
• Right-click on the place-mark symbol and click on *What’s here?* You will see that coordinates in decimal degrees appear in the top-left corner.

We now need to get these coordinates into the *To Joanna* table. This can be done using the *Field Calculator* as described above, however, here we will introduce a different approach using the *Editor* toolbar in ArcGIS.

• Under *Customise > Toolbars* click on *Editor*

The Editor toolbar will appear on your screen.

• Under *Editor* select *Start Editing.*
Click on **To_Joanna** and press **OK**. You may get some warnings but you can ignore these.

You are now in a position to edit the table.

- Back in Google Maps, carefully highlight the latitude value and press **Ctrl-C** to copy it.
- In ArcGIS, click on the **Latitude** cell for “Glaramara” on the **To_Joanna** table.
- Use **Ctrl-V** to paste the value in. Make sure that the zero that was there gets deleted.
- Repeat this for **Longitude** (ensuring that you include the minus sign)
- Under the *Editor* menu on the *Editor* toolbar select *Save Edits*.

You could repeat this for the other values if you wanted, however, for now let us assume that you have found grid references for all of the places that you can.

- Under the *Editor* menu select *Stop Editing*

It should be noted that while you are in editing mode some functionality will not work therefore, when you have finished, make sure that you *Stop Editing*. Just closing the *Editor* toolbar by clicking on the cross in the top-right corner will **not** stop you editing.

c. **Moving points on the shapefile**

You should now have a table that looks like the above. Let us assume (for now) that this is as good as you think you can get it. We now want to turn this into a shapefile, check it is okay, and potentially edit the shapefile to correct any errors by moving points to new locations.
Use *Display XY data* to georeference these points as described above.
  o Make sure that longitude and latitude are the right way around and that you are using the Geographical Coordinate System for WGS1984.

Export the Events Layer to a shapefile called **To_Joanna**
  o Note that because you already have a Dbase file called **To_Joanna** you cannot just call the shapefile **To_Joanna.** 1 implies that we may need several attempts at this.

Add the exported data to the map when you are prompted.

Remove the **To_Joanna** Events Layer from the Table of Contents.

Turn the **Cole_Other** shapefile off so you are able to see the points.

You may also want to change the symbology on **To_Joanna** to make the points stand out more clearly.

We now want to check that our points are in sensible positions. If you right-click on the **To_Joanna** icon on the Table of Contents and go to *Zoom to Layer* the result should look as follows:

![Image of a map with a zoom highlight]()

There are two sets of problems:

1. Using the *Identify* button you should be able to see that the points at the bottom of the screen have latitudes and longitude of 0.
   o It is possible that the coordinates given at the bottom-right of the screen (circled above) are in meters on the British National Grid. If you want to change this (you don’t have to) you can do this by right-clicking on the background of the map window, selecting *Data Frame Properties* from the context menu, going to the *General* tab on the menu that appear and changing *Units > Display to Decimal Degrees.* This is shown below.
2. The second problem is “Fairfield” which is far too far west for the Fairfield in the Lake District (in fact it is in Northern Ireland). If you look in the gazetteer you will see that there are three Fairfields listed and that ArcGIS has simply joined to the first. One solution to this would be to delete the ‘wrong’ entries from the gazetteer and redo the join (and then re-create the shapefile as To_Joanna2). A second alternative would be to paste the correct coordinates from the gazetteer (or Google Maps or another source) into the To_Joanna table and then recreate the shapefile. Here, however, we shall use a third cruder option to demonstrate the use of the Editor toolbar with shapefiles. We will move Fairfield to approximately the correct location and delete the points at (0,0).

- Find the Editor toolbar, Start Editing, click on the To_Joanna1 icon on the Start Editing menu and press OK. You should now be in a position to edit To_Joanna1.

To delete the points at (0,0):

- Select all of the points at (0,0) either by using the Select by rectangle tool on the Tools toolbar or by using Select by Attributes.
- Either press the <Delete> (0,0)key or the black and blue cross (Delete selected) button that is available at the top of the table view (see below)
The (0,0) points should disappear from both the map window and the attribute table.

Save the edits using the Editor menu on the Editor toolbar.

Now we want to move the Fairfield point to approximately its correct position.

- Open Google Earth and find the Fairfield in Cumbria using the Search box in the top-left corner (Google Earth calls this “Fairfield Peak”). Zoom out so you get a clear idea of roughly where this is.
- Add a Basemap to your map window that shows the topography – the Imagery basemap does quite a good job of this.
- Press the black arrow button on the Editor toolbar (shown below)

- Now click on the point that represents Fairfield and drag and drop it to the Lake District.
- Zoom in on the Lake District and try to move the point to the correct location as identified in Google Earth.
- When you are happy that you have the point at the correct location, save your edits and stop editing.

One problem with this is that Fairfield’s latitude and longitude on the attribute table will be wrong (it is still in Ireland). You can correct this by:

- Select Fairfield in the attribute table.
- Right-click on the Latitude column heading and select Calculate Geometry.
• Ensure that you are setting the Y Coordinate of Point using the correct coordinate system (WGS1984) and decimal degrees. You also only want to calculate the selected records. Press OK.
• Repeat this for Longitude (which is X).

This sets these values using the actual coordinates the point is located at.

4. Further work

See if you can georeference the whole of To_Joanna using whichever methods are most appropriate. When you have created it make an attractive map of it.

Compare the locations of the places named in To_Joanna with the locations of the same place name from Cole_other. Are these always the same? What might account for the differences? Might these have implications for analysis?

Recap

If you have a spreadsheet that includes co-ordinates, turning it into a point shapefile is straightforward:

• Use the Add data button to add the file to the Table of Contents.
• Use Display XY data to turn the table into and Events Layer.
• Export the Events Layer to a shapefile
• Remove the Events Layer (and perhaps the original table) from the Table of Contents to avoid confusion.

The Editor toolbar can be used to add, move and delete features. Remember to save your edits and stop editing when you have finished.

If your table includes place-names but not co-ordinates, co-ordinates can be added by joining the table to a gazetteer such as GeoNames

• Add the table with place-names to ArcGIS using the Add data button and to stop editing when you have finished.
• Add the gazetteer to ArcGIS using the Add data button. Note that these can be text files or spreadsheets.
• Add a field called st_name to the table and calculate its initial values to be the same as the original place-names.
• Add fields for the latitude and longitude values to the table. Leave these blank for now.
• Join from st_name to the place name column in the gazetteer.

• If there are any rows that fail to join, investigate why. If it is due to spelling variations update the values of st_name so that it matches the spelling used by the gazetteer.

• Remove the join and try again with the updated values of st_name. Repeat this process of updating the values of st_name until you are happy that the results of the join match as many places in the gazetteer as possible.

• Now calculate the values of your latitude and longitude columns to set them to the joined values for the co-ordinates from the gazetteer.

• Remove the join.

• If there are any remaining rows that do not have x and y co-ordinates then:
  • Either, find co-ordinates for them from other sources such as Google Maps or Google Earth and update the latitude and longitude fields manually using the Editor toolbar.
  • Or, delete the rows.

• Convert the resulting table into a shapefile using the latitudes and longitudes you have just added using Display XY data and export the results to a shapefile.

• Use the Editor toolbar to tidy up the results if required.

Note that if you were doing this for real:
  • It might be a good idea to add one or more additional columns that give the source that each coordinate was taken from, eg. Geonames, Google Maps, etc.
  • You would need to think through the implications (if any) of any place names that you failed to find coordinates for. Are these simply going to be missed off? If so, is this going to affect any subsequent analysis?
Session 7: Geo-referencing Historical Maps

The English Lake District, 1815

Background

Frequently in humanities research you will have a historical map that you want to use in a GIS. This will often have been scanned or photographed and be available to you as a bitmap file such as a JPEG, GIFF or TIFF. The map needs to be imported in ArcGIS and geo-referenced, the process by which co-ordinates are given to the map. Geo-referencing a modern map is usually quite easy as the maps are good quality, are in good condition, will show a co-ordinate grid, and will have information about the projection included. Historical maps are more difficult. The quality of the original survey is usually unknown, they have often been folded for many years which will distort the paper and make it impossible to get completely flat for scanning or photographing, and they often have no information on coordinates or projections. This means that geo-referencing them is harder and there will always be issues associated with accuracy. Historians are, however, used to dealing with limited sources and maps such as these are no different.

This exercise takes a map of the English Lake District dated at 1815. It suffers from all of the problems listed above: it has been stored folded in a drawer and there is no information on the map on accuracy, projections or coordinates (although latitudes and longitudes are on the side). The information on date has been pencilled on. An additional problem is that the map has been scanned in two parts and these have been joined which may further limit its accuracy. We want to geo-reference the map and see how well we can integrate it with modern data. To do this we will have to get the co-ordinates of places from another source such as Google Earth. This raises further issues as the modern data are all on British National Grid while Google Earth gives latitudes and longitudes. Note that while we will do our best on this the results will inevitably not be perfect.
Data Sources

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lakes1815.jpg</td>
<td>A JPEG that holds the scanned map of the Lake District</td>
</tr>
<tr>
<td>Lake District.mxd</td>
<td>A map document that contains the following modern layers. These are all</td>
</tr>
<tr>
<td></td>
<td>projected using the British National Grid</td>
</tr>
<tr>
<td>Coast</td>
<td>The outline of the coast of the North of England</td>
</tr>
<tr>
<td>Inland_Water</td>
<td>The major lakes of the North-West of England. Their names are in the attribute Nam.</td>
</tr>
<tr>
<td>Lakes_towns</td>
<td>The major towns in and around the Lake District. Town_Name is the name of the town, resident names any famous writers that lived there.</td>
</tr>
<tr>
<td>Hindle.jpg</td>
<td>A JPEG of a map of the Lake District scanned from a book that you can use</td>
</tr>
<tr>
<td></td>
<td>to take the exercise further.</td>
</tr>
<tr>
<td>Europe1835.jpg</td>
<td>A JPEG of a scanned map of Europe that you can to take the exercise further.</td>
</tr>
</tbody>
</table>

Summarized Guidelines for Completing Exercise

In this exercise you have a scanned image of a historical map of the Lake District in JPEG format. You want to geo-reference this and convert to a TIFF. Within this you will need to:

- Read a Jpeg into ArcGIS as a raster layer
- Find the co-ordinates of features on the map using Google Earth
- Using the Geo-referencing toolbar.
- Create a link table that will convert the raw co-ordinates on the Jpeg into decimal degrees
- Convert the file to a TIFF
- Change its projection to British National Grid
Step-by-Step Instructions

First you should use Windows Explorer to copy the data for the practical to the folder that you want to work on it in. All the data should be in a the folder <My folder>\GIS7.

1. Explore the scan of the map using an imaging package

Open Lakes1815.jpg in a package such as Windows Photo Viewer and look around the map. How much information can you find about what it shows, when it was surveyed, and so on? Can you see the folds? Can you see where the two scans have been joined? (it runs east-west near Cockermouth and Penrith). Does this look like a good quality map? When you have finished close the map.

2. Read the scan into ArcMap

- Open ArcMap and open the Lake District.mxd map document. As you move the cursor over the map you will see the co-ordinates of the cursor change in the bottom-right corner of the map window. These are British National Grid co-ordinates expressed as meters east and north of a point just off of the south west corner of England.

- Use the Add Data button to add Lakes1815.jpg to the map. Important: do not double click on the Lakes1815 icon in the Add Data window, single click and press the Add button. If prompted, Yes you do want to build pyramids. If it warns you that there is an Unknown Spatial Reference on Lakes1815.jpg click OK.

- Although the icon for Lakes1815 appears on the Table of Contents as shown, nothing appears on the map.
Use **Zoom to Layer** on **Lakes1815** and it will appear but the other layers will disappear (note, if the quality of the map appears poor try zooming in on part of it). You will see that the co-ordinates for the map are very low compared to those for the other layers. This is because these are still in digitiser units (ArcMap wrongly assumes that it is meters because the projection is set to British National Grid). You will also see the (0,0) is at the top left of the map.
If you use the *Full Extent* button (on the *Tools* toolbar with the globe symbol) you will see that *Lakes1815* appears as a small square in the bottom left-hand corner.

Go back to zooming in on *Lakes1815*

3. Creating a spreadsheet of link points

Go to *Customize > Toolbars* and turn on the *Georeferencing* toolbar

Press the *View Link Table* button (circled above).
You can see that the link table requires an ID number (Link), the X and Y coordinates of the source location (XSource and YSource) and its X and Y coordinates in georeferenced form (XMap and YMap). Rather than add coordinates manually, it is usually easier to load them from a text file using the Load... button. First you need to set up the link table.

Create an Excel spreadsheet with six columns as follows:

Rename this worksheet Full data by right clicking on the tab at the bottom-left of the screen where is is currently called Sheet1 and selecting Rename from the context menu that appears.
- Start Google Earth. Find a location on Google Earth that corresponds to a location on the historical map. You may find it easiest to type the name of a place (e.g. Settle, from the south east corner of the map) into the Search box on Google Earth, this will take you straight there.
Session 7: Geo-referencing historical maps

Ian Gregory
Under **Tools > Options** ensure that on the **3D View** tab **Show Lat/Long** is set to “Decimal Degrees”.

Zoom in on the same place on **Lakes1815**.

On the spreadsheet fill in the co-ordinates from **Lakes1815** as **SourceX** and **SourceY**.
Fill in the latitude and longitude from Google Earth. Be as accurate as you can but remember Google Earth is far more accurate than the 1815 map so there is no point being too accurate on Google Earth.

Repeat this until you have 6 to 8 features. These should be well dispersed around the map. Suitable places may include: small towns and villages, coastal features that are still in existence today, conspicuous buildings such as castles, churches and lighthouses, junctions in rivers or obvious coastal features. Note that the location of most of these may have changed over time.

Save your spreadsheet.

This will form the basis of the Link Table but there are still some problems with the way you have structured your spreadsheet:

- A link table cannot have a location field as text
- ArcGIS reads coordinates in as (x,y) which is the oppose to latitude (y) and longitude (x)
- The table has column headings which will cause an error in a link table
- It is in Excel format, ArcGIS wants a text file
Move to a new worksheet on your spreadsheet and rename this to be ArcGIS data (Move to the Sheet2 tab at the bottom-left of the screen and rename this tab as ArcGIS data).

- The first cell should have the ID number of the first row of data from first worksheet (see below, if you do not know how to do this ask the instructor)

- The second cell should have the sourceX value from the first row of data

- The third cell should have the sourceY value from the first row of data

- The fourth cell should have the longitude values

- The fifth cell should hold the latitude values

The first row (Settle) should now look as shown:

Use Fill Down (under the Editing box on the Home tab) to complete the remaining rows.

Save your spreadsheet as an Excel file

Now use Save As to save the worksheet (with the data that we will use for the link table) as a text file (*.txt). This

The reason for doing is this way is that you have two versions of the same data: Full data is easy to enter and holds metadata such as the names of the locations so that if you want to go back and check you can. It also has headings and latitude and longitude in the normal way round for people to understand. ArcGIS data is linked to this but is a subset of the same information in a format that is suitable for ArcGIS. When you save this as a text file only this worksheet is saved (assuming that this is at the front when you save it). The important point, however, is that if you want to edit this you open the Excel version and make the changes to Full data. These will be automatically updated in the ArcGIS data version however when
you have finished making any updates you must save this twice – first as an Excel version and then, with ArcGIS data open, as a text file.

- Close the worksheet.

4. Geo-referencing the map

- Go back into ArcGIS

- The first problem is that this map document uses the British National Grid projection, the coordinates you got from Google Earth are now in decimal degrees. You need to re-project the map into the same projection as Google Earth.
  
  - Right click on the background of the map window and go to Data Frame Properties... on the context menu

  - Under the Co-ordinate system tab select WGS 1984 (this is the commonly used datum for GPS and other data). You will find this under Pre-defined > Geographic Coordinate Systems > World. Press OK and Yes to any warnings you are given.
You will see that your map has disappeared. Use *Zoom to Layer* on **Lakes1815** to bring it back.

- Now press the *View Link Table* button on the *Georeferencing* toolbar.
- Uncheck the *Auto Adjust* checkbox in the bottom-left of this window.
- Press *Load...* and open the text file you saved in the last stage.
- You should see some blue lines stretch from green crosses to red crosses. These are control points. The green ones are the source control points (from **SourceX** and **SourceY**). The red ones are the target control points (from **XMap** and **YMap**).
Zoom in on the area around the target control points. Are these in around the places you would expect them to be?

Check the Auto Adjust box. If you get a warning about your control points not being well distributed then ignore it and press OK. You will see that your map moves to the area defined by the target control points – if you can’t see it use Zoom to Layer to find it. You will also see that ArcGIS has calculated the Total RMS Error and the Residual for each link.

What is your Total RMS Error? How well does your map align with the coast, lakes and towns on the modern layers? Zoom in in detail to check this and turn layers on and off.
- You can delete a link by highlighting it and clicking on the *Remove* button in the top-right corner of the window (see below). Try this with the link with the highest *residual* value. What does it do to the *Total RMS Error*? Does it make the map line up modern features better?

- Experiment to make this as accurate as you can. Note that it is an old map and does not line up perfectly.
  - Note that deleting links may make the RMS error go down. This does not mean that the transformation is more accurate as RMS errors are more accurate (and often larger) where more locations are used.
  - If you make a mistake you can go back by going to *Reset transformation* under the Geo-referencing toolbar.
  - You can also turn *Auto Adjust* on and off.
  - You can reload your links from the text file at any time
  - If you want to change the text file make the change in the Excel version first, save it as Excel, then save it as a text file.
➢ To finalise the transformation go to *Rectify* under the *Georeferencing* toolbar. When filling this in multiply the *Cell Size* by 10 (ie delete one of the zeros after the decimal point). This lowers the quality of the image but dramatically reduces the file size. Change the *Resample type* as shown below, and save the file as a *TIFF* called **Lakes1815** in the appropriate folder. This is shown below however the user interface here is very confusing. The *Output location* should only be the name of the folder in which you are going to save the TIFF. The *filename* (Lakes1815) goes in the *Name* box. Press *Save*.

![Save As dialog box](image)

➢ Add the TIFF file to the map and remove the JPEG. You will see if you zoom in that the quality has been seriously degraded. This is just because of increasing the cell size. If you like, do it again with a smaller cell size. Be warned however that this can take a long time and create very big files if you set it too small.
5. Re-projecting the TIFF

➢ If you go to the Source tab under Properties for the **Lakes1815.tif** layer you will see that this is currently in WGS_1984 projection. All of our other data are in British National Grid.

➢ To change the projection on a raster layer go to ArcToolBox. The **Project Raster** tool is under Data Management Tools > Projections and transformations > Raster.

➢ Fill this in as shown above. You will find British National Grid under Projected Coordinate Systems > National Grids > Europe. The output raster should be called **Lakes_bng** (for British National Grid). Press **OK**.
➢ Now change the Data Frame’s projection back to British National Grid.

➢ The resulting map should appear as shown below:

6. Further work: If you have time

You also have JPG files called **Europe1835** and **Hindle**. Can you geo-reference one or both of these and save it/them as sensibly sized layer files?
Recap

Any image file may be added to ArcGIS but it will not have real-world locations.

An image file can be geo-referenced using real-world locations. A link table is used to match the co-ordinates of places on the image scan to real-world locations.

Real-world co-ordinates may be marked on the original map (for example using grid squares) in which case creating a link table is relatively easy. Hold the cursor over the relevant place on the scan (such as a grid intersection) to find the co-ordinates on the scan. Make a note of these and the real-world location they refer to and use these to create a link table.

If real-world locations are not marked on the map, Google Earth is a good place to find them using features that appear on the map and on Google Earth.

In ArcGIS a link table can only have five columns: ID, the co-ordinates on the scan and the real-world co-ordinates. These must be in a text file and can only be numeric. It is a good idea to have a more extensive version of this that gives more information on where each feature refers to. Excel allows you to create these two versions (a human-friendly version and the version for ArcGIS) in a seamless manner.

Geo-referencing is done using the Georeferencing toolbar. It is best to keep Auto Adjust off until you are sure you are ready.

Rather than use a link table you can just digitize links using the Georeferencing toolbar.

Once you have georeferenced the image you must export it. A number of formats are available such as georeferenced TIFFs. The file sizes for these can be large.

Once you have a georeferenced image these can be re-projected onto a new projection.
Google Earth file structure: Keyhole Markup Language

An example of KML

Header info
Define a place mark
Its name is Prospect Bridge
It has some other attribute info in html format

We can link to HTML pages from this

It is a point at these co-ords: (long, lat)
End of this placemark

In Google Earth
KML and KMZ

• KML is a single text file
• KMZ is a zip file that contains:
  – doc.kml: the KML
  – Images (usually in a folder called “Files”)
    • Any user defined marker symbols
    • Any scans of maps
• To explore a kmz file:
  – Change the extension to zip
  – Unzip it

KML/KMZ on the internet

• You may need to add a small settings file to the root folder of your website – ask your systems administrator
• Place the KML/KMZ file on the website
• Put a hyperlink to it from an HTML page
Exercise 8: Geo-visualisation with Google Earth

Literature in the English Lake District

Background

While ArcGIS is an excellent tool for GIS analysis it is not a good data viewer for a number of reasons: it is expensive, it is desktop-based, it does not come with any background imagery that provides context for your data, and it is difficult to use. Increasingly internet-based Virtual Globes are available that allow you to place spatial data on the internet and pan, zoom and query it in an easy manner. Examples of these include: Google Earth (http://earth.google.com), ESRI’s ArcExplorer (http://www.esri.com/software/arcexplorer), and Microsoft’s Virtual Earth (http://www.microsoft.com/virtualearth). Most of these combine a large amount of background satellite imagery with the ability to add your own data using a format called Keyhole Mark-up Language (KML), an XML implementation. Point, lines, polygons and attribute data can all be included in KML and background raster images can be created. Because KML is text-based, including raster images involves creating a KMZ file, effectively a zip file that includes the KML files and the images in Jpeg or similar format.

In this practical we take text files that include the names of places visited described in two early accounts of the Lake District: by Thomas Gray in 1769 and Samuel Taylor Coleridge in 1802. We will also include an 1815 map of the Lakes District as a Jpeg. We will read these into ArcGIS and then export them into KML/KMZ. We will then explore the resulting files in Google Earth.

Further reading on KML and Google Maps:


Udell S. (2008) Beginning Google Maps Mashups with Mapplets, KML and GeoRSS: From novice to professional. APRESS. – Information on how to create Google Maps displays within websites that can hold KML files (or other types of content).

See also: https://developers.google.com/kml/
Data Sources

<table>
<thead>
<tr>
<th>File Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray_Start.csv</td>
<td>A comma separated text file that includes the places that Gray started/finished his days at. It includes: the name of the place, whether he visited or not (which in this case is always ‘Y’), the day that he visited it, and the longitude and latitude of the place.</td>
</tr>
<tr>
<td>Cole_Start.csv</td>
<td>A comma separated text file that include the places that Coleridge started/finished his days at. It includes: the name of the place, whether he visited or not (which in this case is always ‘Y’), the day that he visited it (from day 1 to day 9), the date, and the longitude and latitude of the place.</td>
</tr>
<tr>
<td>Gray_all.csv</td>
<td>All of the places that Gray mentions (in a similar format to the above).</td>
</tr>
<tr>
<td>Cole_all.csv</td>
<td>All of the places that Coleridge mentions (in a similar format to the above).</td>
</tr>
<tr>
<td>Lakes1815_small.tif</td>
<td>A geo-referenced TIF of an 1815 map of the Lake District.</td>
</tr>
</tbody>
</table>

Summarized Guidelines for Completing Exercise

In this exercise you will take the text files, read them into ArcGIS, and then export them to KML. You will then open them in Google Earth to allow you to compare ArcGIS and Google Earth:

- Convert a textfile to a Shapefile
- Export a Shapefile to KML
- Export a map document to KMZ
- Explore KML and KMZ files in Google Earth
Step-by-Step Instructions

First you should use Windows Explorer to copy the data for the practical to the folder that you want to work on it in. All the data should be in a folder <My folder>\GIS8.

1. Convert a CSV file to a shapefile

- Open the file Cole_Start.csv in Excel and check its format. Then close it again.
- Start ArcMap in a new map window
- Add Cole_Start.csv to the map window.
- Right-click on its icon on the Table of Contents and select Display XY Data...

- Fill this in as shown below. The WGS 1984 projection is found under Edit... It is a Geographic Projection and then under World. Ensure the Warn me if the resulting layer will have restricted functionality box in unchecked. Click OK.
- The points should be displayed: Use the Identify button to check their attributes.
- Now click on the Cole_Start.csv Events icon on the Table of Contents. From its context menu select Data > Export Data...

![Image](image.png)

- Export the file to a Shapefile called Cole_Start. Add this to the map.

![Image](image.png)

- Now remove Cole_start.csv and Cole_start.csv Events from the Table of Contents.
2. Export the shapefile to KML

You now have a point Shapefile representing the places that Coleridge started and finished each day’s journey. We now want to create a KML file of this.

- Find the Layer to KML tool in ArcToolbox. It is under Conversion Tools > To KML. Drag the Cole_start icon into the Layer box, set the Output File to be Cole_start.kmz and set the Layer Output Scale to be 1. Press OK.

- Nothing obvious happens when the tool finishes but if you look in the appropriate folder in Windows Explorer you should see that Cole_start.kmz has been created.
3. Open the file in Google Earth and explore it.

- Open Google Earth. If a start-up Tips screen appears then close it.
- Use File > Open to open the Cole_start.kmz file.
  - If there are lots of icons that you don’t want you can remove them by unchecking the box next to Primary Database on the bottom half of the table of contents (circled below)

- Cole_Start should have appeared in the top half of the Table of Contents. Your points will also have appeared on the map but may not be obvious because they are too small and are green. To change the symbols of your points:
  - Press the triangle next to Cole_Start on the Table of Contents and then repeat with the next Cole_Start to open the folders as shown below:
Right-click on the lowest **Cole_Start** and go to Properties on the context menu

Click on the Style tab.

You should get a menu like the one below. Clicking on the point symbol at the end of **Name** (in the top-right of the menu) allows you to change the symbol, the **Label** options refer to the text that gives the place-names. Icon refers to the symbol. In both cases you can change the: **color**, **scale** (size), and **opacity** (how see-through this is. If this is 0% you will not be able to see it as it is completely transparent, 100% is completely solid). Change these to symbols/text format that you want.
- Experiment with:
  - Panning, zooming, tilting and rotating using the tools in the top-right corner. Can you create a view such as the one shown below of Scafell above Burnthwaite in Wasdale.
  - Clicking and double clicking on points
  - Opening the folders on the Places side-bar and clicking or double clicking on the places there.

- Right-click **Greta Hall, Keswick** on either the map of the Table of Contents and select *Directions from here* on. Then right-click on **Longmoor, Ennerdale** and click on *Directions To Here* on. Click on the magnifying glass symbol on the *Directions* tab.

- The directions that are returned are the modern directions (not the route Coleridge would have followed in 1802).
Click on the cross at the bottom right of the directions on the Table of Contents to get rid of this.
Right-click on either the symbol for **Greta Hall, Keswick** or its name on the sidebar. Select *Properties* from the context menu. This gives the properties for Greta Hall only. (Note that the HTML in the *Description* box may vary from what is show below).

- Change the symbol for Greta Hall to a red circle with a 1 in it using the button in the top-right.

- Select the *Style, Color* tab and change the colours of the label and icon.
- Make sure the *Opacity* is not set to 0% - if it is you will not be able to see the thing it is set for (100% is solid, lower than this and it starts to become see through)

- Then press *OK*.

- Repeat this for the other places to create a map as shown below.

- If you right-click on the **Cole_start** folder icon on the sidebar you can save the updated image as a new KML (or KMZ) file.

- Finally right click on the Cole_start icon and the sidebar and select *Delete*. This does the same as *Remove* in ArcGIS.
4. Exporting a map from ArcGIS to KMZ

- Add Gray_start to ArcMap and turn it into a shapefile as above.
- Make sure that Gray_start and Cole_start are shaded in different colours.
- Add lakes1815_small.tif to the map, you should now have three layers as shown below.

Save the map document as Lakes.mxd.

Now go to the Map to KML tool and fill it in as shown.

Open Lakes.kmz in Google Earth.

You will see that there are three folders in the Layers folder that has been added to the sidebar. You can turn on and off whole folders or individual features within them. In the example below all of Cole_start has been turned off. All off Gray_start has also been turned off except Brough, the place he started from.
5. Further work

- Experiment further with Google Earth. Try changing shadings and symbols. Try adding points, lines, and polygons yourself using the Add menu.
- What do you think that advantages and disadvantages of Google Earth are compared to ArcGIS?
Recap

Exporting data from ArcGIS shapefiles or raster images to KML or KMZ is straight-forward using ArcToolbox.

- If you just want to convert a single vector shapefile then create a KML file.
- If you want to convert a whole Map Document or any images such as a raster map then create a KMZ file.

Once in Google Earth it is easy to explore the data and to disseminate it over the internet. It is much better at spatial queries than attribute queries and has no analytic functionality.

Google Earth does provide limited functionality for editing KML files.

KML files are text files and can be edited using a text editor.

KMZ files also contain KML which can be accessed by uncompressing it first.
Extra Exercise: A quick introduction to QGIS

Background

QGIS is perhaps the most widely accepted free, open source GIS package. It is available for Windows, Mac and Linux.

Starting the practical

QGIS can be downloaded from: http://www.qgis.org. This practical is based on QGIS Wein (2.8.1).

The files for this practical are the same as was used in the first hands-on practical:

- Copy your GIS2 folder to a new folder called QGIS2

Please note: This is not a practical in which you follow clearly defined steps in order. It is intended to show the basics of QGIS and how it is similar to, and differs from, ArcGIS. Feel free to experiment and explore as you go along.

The basics of QGIS

- Open QGIS Desktop from the Start menu

The basic QGIS screen is shown above. Like ArcGIS, it has a Table of Contents to the left (with Browser written at the top), a map window that takes up most of the screen, menus at the top and a range of shortcut buttons on menus that can be moved around.
1. Setting up your project:

The QGIS equivalent of a map document (.mxd file) is called a Project and has a .qgs extension. To create a project:

- Go to the Project menu and select Save As... Call the project Lakes and save it in the QGIS2 folder.

Although the screen has not changed much (the project name 'Lakes' is now in the top left corner of the screen) you are now working with a project called Lakes.

To add shapefiles to the project:

- Under the Layer menu select Add Layer and then Add Vector Layer...
- Press the Browse button and navigate to the folder you want.
- Select the WW_Guide shapefile (It can help to change the type of file from All Files (*.*) to ESRC Shapefiles (*.shp) using the drop-down menu at the end of File Name.
- Press Open.
As with ArcGIS, the shapefile is added to the Table of Contents and appears in the map window.

A shorter way to do this is to use the Browser window at the top of the Table of Contents:

- Navigate through the file structure to find your files.
- As shown to the left, you can select multiple files by holding down the shift key and clicking on the file icon. Select Coast, Inland water bodies, STC_1802, and Urban Areas.
- Press the Add button at the top of the window (the green square with the white cross circled in the top left).

The files will be added to the Layers
If you want you can remove the *Browser* or other windows from the Table of Contents by clicking on the cross in the top-right (circled below).

To bring it back right-click on one of the dotted lines and check (for example) *Browser*.
2. Basic operations

As with ArcGIS, you can turn layers off and on, and change the drawing order using the Table of Contents. Experiment with this.

Panning and zooming is done using the tool bar shown above (the location of this may vary as it can be moved). It can also be done from the View menu as also shown above.
Various other features that appear on ArcGIS’s tools toolbar, such as Identify, are found on the toolbar shown above.

As with ArcGIS, a lot of functionality is found by right clicking on a layer’s icon on the Table of Contents as shown above.

Properties is again where a lot of functionality affecting a particular layer is found.

- Explore the Properties and experiment with changing symbols and colours (found under Style).

Unlike ArcGIS, one of the layers on the Table of Contents is always the active layer. This has its icon underlined. This means that in QGIS many of the functions only available in context menus in ArcGIS are available through the windows and toolbars at the top of the screen. In the example above, WW_Guide is the active layer so pressing the Open Attribute Table button (circled) will open its attribute table rather than any other layer’s.

To change which is the active layer, simply click on its icon.
3. Working with attribute data

Attribute tables are found by right-clicking on the layer’s icon on the Table of Contents and going to Open Attribute Table.

An example of a layer’s attribute table is shown above. As with ArcGIS, there are a number of buttons that allow, for example, rows to be selected according to a particular query.

To select one or more records:

- Click on the Select Feature using an Expression button (circled below)

- Click on the + symbol next to Fields and Values.
- Double-click on st_name. It should appear in double quotes in the dialogue window on the left of the menu
- Either click on the “=” button in the top-left of the menu or type “=” (no quotes) after “st_name”.
- Click on the all unique button next to Load values
- Double-click on ‘Ambleside’, it should appear in the text box in the text box in single quotes.
- The menu should look as shown below
- Click on Select
- Click Close.
- Now in the bottom-left of the attribute table change Show All Features to Show Selected Features.
Updating fields is similar to ArcGIS in that it involves the Field Calculator (circled below on the right). The dialogue is, however, more complicated. Before you do this it is best to turn editing mode on. This is done with the Toggle Editing Mode button (circled below on the left).

You can now edit the table and save your edits once you have finished.

- Try to add a new field called Test and set its values to 2 (or whatever you want). This differs somewhat from ArcGIS’s Field Calculator, as shown above. It allows you to save new values and create a field in the same operation (by checking the Create new field checkbox), update an existing field (Update existing field checkbox) and apply results to all records or only the selected ones (the Only update xx selected features checkbox).
- The example below shows how to do this
When you have finished, you need to turn edit mode of by pressing the *Toggle Editing Mode* button again (top left of the attribute table window). This will give you the option of saving or discarding your changes.
4. Creating a map

Unlike ArcGIS QGIS has no Layout view.

- You can add elements like scale bars and north arrows by going to Decorations under the View menu.

- Once you have done this, ensure that you check the Enable scale bar (or whatever element) checkbox.

North arrows, etc are done in a similar manner

You can export a finished map to file using Project – Save as image…
5. Analytic functionality

QGIS has a wide range of analytic functionality similar to many of the functions found in ArcToolbox. These are found under the Vector, Raster, and Processing – Toolbox menus. The Toolbox opens a series of tools to the right of the map window.

6. Working with Plugins: Open Layers

You can also install a wide variety of further functionality using the Plugins menu. In this example we will install and use the OpenLayers Plugin which provides functionality similar to ArcGIS’s Add Basemap.

Before you do this, zoom in onto, for example, the Lake District area.

a. Install the plugin

- Under the Plugins menu select Manage and Install Plugins.
- Click on the OpenLayers plugin and press Install Plugin (this will take a minute or two – once it is done the button will change to Uninstall plugin and Reinstall plugin)
- Click Close to close the menu
b. Using the plugin

- Now under Web you should find the *OpenLayers plugin* is available

- Select one of the options that you want (e.g., *Open Street Map*). It should then appear as a backcloth. It will also be on the Table of Contents.
Further Reading

(With thanks to Norma Serra, University of Victoria)

a. Manuals:


b. Books:


c. Tutorials:

Introducing GIS worksheets: http://linfiniti.com/dla

QGIS Tutorials: http://www.qgistutorials.com

QGIS basics for journalist: http://multimedia.journalism.berkeley.edu/tutorials/qgis-basics-journalists

d. Web sites:

Open Source Geospatial Foundation http://www.osgeo.org/content/foundation/about.html

Open Source GIS in Rural Townships: http://www.townshipgis.com/home
Other free GIS software

**gvSIG**: [http://www.gvsig.org/web](http://www.gvsig.org/web). A Spanish GIS software package that provides extensive GIS functionality.

**MapWindow GIS**: [http://www.mapwindow.org](http://www.mapwindow.org). A free open source GIS package. This is the package used by Owens *et al* in their *GIS Training Manual for Historians and Historical Social Scientists* ([http://www.geographicallyintegratedhistory.com](http://www.geographicallyintegratedhistory.com)).

**GeoDa and PySal**: [http://geodacenter.github.io/index.html](http://geodacenter.github.io/index.html). GeoDa is a relatively easy to use package that allows a range of spatial analysis functions to be carried out on shapefiles. PySal is an open source Python library for carrying out spatial analysis functions. These were developed by the Center for GeoSpatial Analysis and Computation at Arizona State University, and are now hosted at the Center for Spatial Data Science at the University of Chicago.

**Google Earth Pro**: [http://earth.google.com](http://earth.google.com). Not really a GIS but enhances the Google Earth functionality.