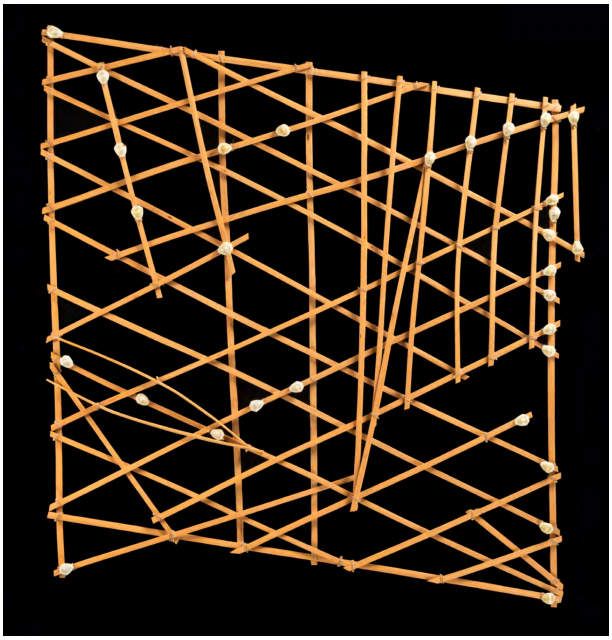
MO5161 Skills in Transnational History

Module Coordinator: Konrad M. Lawson

Fall, 2020



Marshallese Navigational Chart
Denver Museum of Nature & Science Catalog No.: A926.1

Repository for this handbook available at:
https://github.com/kmlawson/skills-in-transnational-history

Overview

Meetings:
16:00-18:00 on Thursdays, other times TBD.

Location
BWD 9 or on Microsoft Teams

Note: No meeting in Week 6, 10, or 11. Second meeting in Week 7 and 8 on Tuesdays 13:00-15:00.

Preparation: 12-18 hours per week

Week 0 - **Orientation: History in Practice - Balancing Skills and Critique in the Digital Humanities**
Week 1 - **Introduction to Historical GIS and First Look at QGIS**
Week 2 - **QGIS: Basics and Georeferencing Historical Maps**
Week 3 - **QGIS: Techniques for Exploration and Analysis of Spatial Data**
Week 4 - **Thinking With/Against Maps and Spatial Analysis**
Week 5 - **Text Analysis: Using the Command Line and Regular Expressions**
Week 6 - *Independent Learning Week*
26 Oct - First Assessment Due
Week 7.1 - **Introduction to R**
Week 7.2 - **Using R: Frequencies and Correlations**
Week 8.1 - **Using R: Sentiment Analysis and Topic Modeling**
Week 8.2 - **Networks: Introduction to Social Network Analysis**
Week 9 - **Mapping Texts and Networks; Mediums for Sharing Your Work**
Week 10 - *Work on Final Project* - Thursday Open Office Hours 16-18:00
Week 11 - *Work on Final Project* - Thursday Open Office Hours 16-18:00
11 Dec - Second Assessment Due

Key Info

Coordinator: Konrad M. Lawson **Email:** kml8@st-andrews.ac.uk
Office: St. Katharine's Lodge Room B3
Office Hours: By appointment via Microsoft Teams

Description

This digital humanities module is offered will introduce MLitt students in the School of History at the University of St Andrews, Scotland to historical GIS, the computational analysis and visualisation of texts, and social network analysis. Students will read and critique examples of humanities scholarship that apply these digital methodologies. The seminar will, throughout the module, engage with the critical debates on the advantages, limitations, challenges, and costs of employing these approaches in the humanities.

Assessment Summary

50% - **26 October** before midnight - First Project or Essay
50% - **11 December** before midnight - Second Project or Essay

Assessments

You have two equally weighted assessments. For each of these two you may choose to submit one of the three following possibilities. You are very welcome to do more than one of the same type of essay or project. You may use a similar body of sources in more than one project only if they are of a different type. If you do a map project and choose to do another map project, then the second map project should use another body of sources.

Option One: Map Project

This assessment is designed to give you a chance to demonstrate your ability to effectively use QGIS to prepare map layers, produce exported maps for incorporation into a historical project; be able to articulate what the maps can be useful for; and be able to describe the limitations and problems associated with your map and the data on which it depends. For Submission you must upload, as a zipped collection of files, to the MMS:

1. 1,500 word explanatory essay which provides some basic historical background to the inspiration for the map project, discusses the ways in which the map can be used for heuristic, illustrative, or analytical purposes, and the limitations and problems with map. In other words, discuss both what the map might reveal, but also what it may potentially conceal or distort. *Important*: At the end of the essay or embedded into the body of the essay should be the exported maps, with any post-processing required, such as you would want them to appear in a historical essay, with an appropriate caption. You will be marked in part on the presentation quality of the map and its ability to clearly and cleanly communicate what it sets out to show.

2. An appendix, at least 500 words must also be uploaded to the MMS which describes the technical process by which you assembled the QGIS project and its layers. What did you do to create each of the layers? Where exactly did you get the data for each of the layers and why did you choose that data over possible alternatives? What other technical steps did you carry out in order to produce the project?

3. If your files cannot, for some reason be uploaded directly to MMS as a zip file, you may instead upload a single page document with a link to a place where the QGIS files can be downloaded (publicly accessible or shared Google Drive or Dropbox file link, or other file sharing resource). If you choose to use this method, in order to ensure that no further modifications to the documents are made after submission, your document with the link must include the "checksum" or MD5 hash for the zipped collection of documents (you can do this via the command line or online at a site such as this one: http://onlinemd5.com/). Alternatively, you may submit a USB drive on the previous meeting of class with the files.

Scope: The number of layers of data, the number of visualisations produced, and the quantity of data to include depends entirely on what map or maps are being created, that is what you are trying to show. It may consist of three or up to a dozen layers and may include natural coastal or topographical data, other features, and thematic data layers visualizing things. You may generate a single exported map if it is rich and complex (but not to the detriment of its ability to communicate), requiring a very significant amount of time in preparation or a collection of different maps using different data or snapshots of data at a particular time if they can be used effectively for historical argumentation. You may combine maps with tables or graphs, or maps combined with presentation of spatial analysis. Think of your project as building a portfolio to pitch your work at a workshop, in a grant application, etc. showing a good balance of analysis, design skills, and judicious of choice in what to include and exclude. If you are concerned about finding the balance, feel free to consult with the tutor as your potential outputs become clearer.

Considerations for Assessment:

- Context: Did the project explanatory essay set the context sufficiently?
- Analysis: Did the project explanatory essay effectively explore the limitations and advantages of the maps produced or spatial analysis performed?
- Combination: Did the chosen combination of a) the number of maps produced b) the challenge of compiling the spatial data for visualisation c) the challenge of specific techniques employed in the creation of the visualisations d) the employment of any specific spatial analysis e) any alternative ways of visualising the data (tables, graphs, as appropriate) represent an effective and judicious balance for the presented project portfolio?
- Appendix: Did the appendix effectively describe the process of creating the visualisations and any spatial analysis performed and did it properly cite the sources for the data.
- Design: Did any visualisations or maps produced clearly communicate their contents and demonstrate a good sense of design, well-chosen scales, and appropriate density of visual information?

- Restraint: Did the project avoid superfluous visualisations that contribute little to the overall project and show good judgement in what to include and exclude?

Option Two: Text Analysis or Social Network Project (50%)

You may do a project which makes use of one or more of the new computational techniques learned in the second half of the course: text analysis and/or social network visualisation and analysis. You are free to use the combination of tools and techniques that you feel most appropriate and comfortable in using (R, Python, Cytoscape, Gephi, Orange, Voyant Tools, Palladio, Shiny, regular expressions and command line utilities etc.). You are also welcome to teach yourself and make use of other techniques and tools that you have come across if these are well documented and described in your project.

Your project should include from one to a dozen visualisations depending on the difficulty and complexity of the visualisations produced and what they can potentially be used for.

These may include any of the following: 1) a collection of graphs and tables identifying patterns in the data or the product of computational analysis of text or networks. 2) Visualisation of the text or social network data in combination with GIS in the form of a choropleth or other thematic map, visualisations of networks using Cytoscape, Gephie, Palladio, or other visualisation tool, maps which visualise networks or material extracted from a text or corpus of texts. 3) a list of regular expressions, including description of what they do, that were used to clean data or extract useful material from texts.

For Submission you must upload to the MMS:

1. 1,500 word explanatory essay which provides some basic historical background to the inspiration for the project and visualisations. It must also discuss the limitations, assumptions, and advantages of the material produced, including its heuristic, illustrative, and analytic goals, but also what they may conceal or distort. *Important*: Embedded in the body of the essay or at the end of the essay you should produce the exported visualisations.

2. At least 500 word appendix discussing the technical process you used to create the database and visualisation. What decisions did you make in coding the data, where did you get the data from exactly, and what steps did you take in producing the visualisations.

3. If your files cannot, for some reason be uploaded directly to MMS as a zip file, you may instead upload a single page document with a link to a place where the QGIS files can be downloaded (publicly accessible or shared Google Drive or Dropbox file link, or other file sharing resource). If you choose to use this method, in order to ensure that no further modifications to the documents are made after submission, your document with the link must include the "checksum" or MD5 hash for the zipped collection of documents (you can do this via the command line or online at a site such as this one: http://onlinemd5.com/). Alternatively, you may submit a USB drive on the previous meeting of class with the files.

Considerations for Assessment:

- Context: Did the project explanatory essay set the context sufficiently?
- Analysis: Did the project explanatory essay effectively explore the limitations and advantages of the visualisation produced or computational analysis performed?
- Combination: Did the chosen combination of a) the number of visualisations etc. produced b) a challenge in compiling the data for visualisation c) the challenge of specific techniques employed in the creation of the visualisations d) the employment of any specific forms of computational analysis of networks, text, or spatial data. Were they an appropriate combination of materials?
- Appendix: Did the appendix effectively describe the process of creating the visualisations and any analysis performed and did it properly cite the sources for the data.
- Design: Did any visualisations or maps produced clearly communicate their contents and demonstrate a good sense of design, well-chosen scales, and appropriate density of visual information?
- Restraint: Did the project avoid superfluous visualisations that contribute little to the overall project and show good judgement in what to include and exclude?
- Development: If the project used historical GIS and use the same or similar source base, then did the project show a substantive development beyond the first assessment, and make clear in the explanatory essay what that development consisted of.

Option Three: Critical Essay (50%)

You may, alternatively, for one or both of the assignments, write a 4,000 word essay in the genre of the critical humanities. For this alternative, you should engage effectively with the theoretical scholarship on critical digital humanities both generally and specifically for the area of focus for your essay and it must engage critically and effectively with several several pieces of scholarship that may be described as applications of the technologies studied in this module.

Alternatively you may target a particular project in the digital humanities or a published work which uses digital approaches and write a critique or evaluation of this work. Again, however, this should employ some of the critical scholarship that is relevant in your analysis.

Considerations for Assessment:

- If the essay evaluated a digital project or publication which made use of skills in the digital humanities such as computational methods etc., and was the evaluation fair, constructive, and show a good understanding of the project being critiqued.
- Did the essay engage effectively with an appropriate amount of theoretical literature for its topic?
- Did it have a clear argument which was well supported throughout the essay?

Option Four: Video Project (50%)

Finally, you may create a 20 minute video project. The output should be submitted as a video file. The contents of the essay may follow completely the guidelines for Option Three above, engaging critically with the digital humanities. Alternatively, it may generally follow the guidelines for Option One or Option Two above. The key is that it is to combine the voice of the creator as narrator offering explanation and analysis, combined with the visual element of slides, maps, graphs, etc. as appropriate. In addition, the student is asked to attach a "bibliography" of sources used in the form of "film credits" at the end of the video (this may be static on separate slides or scrolling movie style). The "credits" at the end are not counted in the 20 minute limit.

Considerations for Assessment:

- Did the project meet the relevant considerations for assessment from the option above that most closely matches the contents?
- Was the voice narration effectively delivered in terms of speed, modulation, clarity etc.
- How effective was the visual elements of the video?
- Did the video avoid overly heavy use of text.

Guidelines

Your submitted work should meet the following requirements:

Headers and Formatting

At the top of all your written work or on a cover page, please include:

- The date of submission
- The assignment you are submitting (e.g. Historiographical Essay, Empirical Research Essay, etc.)
- Your student number
- A title for your essay or project
- The total number of words including footnotes (use the word count feature of your word processor)

When formatting your assignments, please follow these guidelines:

- Add page numbers
- Use a minimum of 12 sized font
- Use a serif font (such as Times Roman, Georgia, Garamond), not a sans serif font (such as Arial, Helvetica, Verdana)

Footnotes and References

Please follow the St Andrews School of History Style Sheet, available on the school website. This document, sections 1-4, contains extremely valuable information on how to compose your essay, including how to format your footnotes and bibliography. Please note that failure to follow the style sheet can result in penalties on the final mark for the essay.

Online Submission and Late Work

Unless otherwise indicated, work will considered submitted by the date the document was submitted online on the MMS website for the course. If you are concerned that any given assignment was not correctly submitted to the MMS, you are free to email a copy of your submitted assignment to the module coordinator, if you like. In the event an assignment was not correctly uploaded to the MMS for some reason, but an emailed copy was sent in time, that date of submission will be used, but a copy will still need to be submitted to the MMS thereafter. This module follows the official School of History penalties for late work found in the PGT handbook.

Word Limits

Assessed work with word limits should be always submitted within those limits. Writing in a clear and concise manner, and being able to structure and execute an argument that may be shorter than you feel is required is a skill that is of great use in academic fields as well as the workplace beyond. Please do not go over the limit and force yourself to work within them as a practice that will be important for writing assignments in your future careers. Please see the PGT handbook.

Please Note: The appendix for the two project submissions has a minimum but no maximum.

Academic Misconduct and Plagiarism

Academic integrity is fundamental to the values promoted by the University. It is important that all students are judged on their ability, and that no student is allowed unfairly to take an advantage over others, to affect the security and integrity of the assessment process, or to diminish the reliability and quality of a St Andrews degree. Academic misconduct includes *inter alia* the presentation of material as one's own when it is not one's own; the presentation of material whose provenance is academically inappropriate; and academically inappropriate behaviour in an examination or class test. For more information:

https://www.st-andrews.ac.uk/students/rules/academicpractice/

If you are unsure about the correct presentation of academic material, you should approach your tutor. You can also contact CAPOD, which provides an extensive range of training on Academic Skills.

http://www.st-andrews.ac.uk/capod/

Advice and Support for Students

For advice and support on any issue, including academic, financial, international, personal or health matters, or if you are unsure of who to go to for help, please contact the Advice and Support Centre, 79 North Street, 01334 462020, theasc@st-and.ac.uk.

Semester Dates

The Semester Dates are available at:

https://www.st-andrews.ac.uk/semesterdates/

Marking Scale

We use a 20 point marking scale for your assignments. For more on interpreting the marking scale, see the School of History marking scale for taught Postgraduates. Please see the PGT handbook.

Good Academic Practice

http://www.st-andrews.ac.uk/media/teaching-and-learning/policies/gap.pdf
http://www.st-andrews.ac.uk/media/teaching-and-learning/documents/September 2014 Updated GAP Guide for Students.pdf

Absence from Classes

Attendance is a basic assessment requirement for credit award, and failure to attend classes or meetings with academic staff may result in your losing the right to be assessed in that module.

Please read the university policy on student absences:

https://www.st-andrews.ac.uk/students/rules/selfcertification/

Disability Support

If you require support for disability reasons, for example teaching and exam arrangements, please contact the Disability Team in Student Services who can provide support for a wide range of disabilities such as learning difficulties, visual and hearing impairments, mobility difficulties, Asperger's, mental health, long standing medical condition and much more.

Main Readings

SPATIAL Gregory, Ian N., Don DeBats, and Don Lafreniere, eds. *The Routledge Companion to Spatial History*. Routledge 2018

GREGORY Gregory, Ian N., and Paul S. Ell, eds. *Historical GIS: Technologies, Methodologies, and Scholarship*. Cambridge Studies in Historical Geography, no. 39. Cambridge: Cambridge University Press, 2008. Library ebook

LÜNEN Lünen, Alexander von, Charles Travis, eds. *History and GIS: Epistemologies, Considerations and Reflections*. Dordrecht ; New York: Springer, 2013. Library ebook

BODENHAMER Bodenhamer, David J., John Corrigan, Trevor M. Harris, eds. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press, 2010. Library ebook

GEDDES Gregory, Ian N., and Alistair Geddes. *Toward Spatial Humanities: Historical GIS and Spatial History*. Indiana University Press, 2014. Library ebook

KNOWLES Knowles, Anne Kelly, and Amy Hillier. *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. ESRI, Inc., 2008.

PAST Knowles, Anne Kelly *Past Time, Past Place* ESRI Press, 2002.

DEEP Bodenhamer, David. John Corrigan, Trevor M. Harris *Deep Maps and Spatial Narratives* Indiana University Press, 2015 Library ebook

RAU Rau, Susanne, Ekkehard Schönherr, eds. *Mapping Spatial Relations, Their Perceptions and Dynamics: The City Today and in the Past. Lecture Notes in Geoinformation and Cartography*. Cham ; New York: Springer, 2014. Library ebook

SCOTT Scott, John *Social Network Analysis* 4th Edition (3rd will also suffice) London: Sage, 2017.

SILGE Silge, Julia and David Robinson *Text Mining with R* O'Reilly (2017) free online version

JOCKERS Jockers, Matthew L. *Text Analysis with R for Students of Literature* Springer (2014) Library ebook

LUKE Luke, Douglas A. *A User's Guide to Network Analysis in R* Springer (2015) Library ebook

Managing Your Time During the Semester

You should expect to spend about 20-25 hours each week in preparation for this seminar module (40 credits), with the remaining third of your weekly study time dedicated to your core module (20 credits). I strongly suggest you set aside regular time every week, across at least three days of the week, that is especially allocated to this module in order to balance out the time you dedicate to it throughout the semester. Readings for the module will rarely involve more than half the time and you should never consider your preparation done for the week when the readings are complete. The majority of your time should be spent on your projects. You don't need extensive notes on the readings, focus less on the particulars of specific historical projects we may read about but instead on the methodologies, weaknesses and strengths regarding them, and bigger connections that connect to other readings and themes we have been discussing.

After you finish the readings for the week, use the remaining time you have set aside for source *research*, *experimentation*, *play*, and data *exploration*. This should eventually, but not immediately, be directed towards the broader end of becoming more familiar with the technologies introduced in the module, and most of all, helping you towards the development of the projects for your two assessments. Bring your ideas, as they develop, into the module discussion and share them with your fellow students as well. I cannot stress enough that this time is important. You will quickly find that, even when you have a firm project idea in mind, the sources settled, and a goal in sight, the majority of time in project development will go towards cleaning or preparing data, tinkering with software, googling obscure errors, watching YouTube tutorials or diving into bulletin boards about specific features or tasks you wish to perform, and, alas, it is *very common* to discover that a project is ultimately not viable with the original approach or data, and a switch in direction is needed. *A critical ingredient for success is, therefore: to allow time for productive failure and new beginnings.*

This is a computer screen intensive module. Though this may seem like an obvious point, it is worth reminding you here that it is not healthy to sit and stare at a computer screen for long periods. I suggest you try working in periods of 25 minutes or so with the help of a timer, forcing yourself to get up and take a break regularly.

Optional Pre-Semester Preparation

If you would like to get a head start on the module and get a flavor for the module, consider trying some of the following tasks. If you have challenges with software installation and such, then don't worry too much, we will try to address some of these issues during the semester.

1. Read the short book by Franco Moretti, *Atlas of the European Novel, 1800-1900* and the fun but very informative book by Mark Monmonier *How to Lie with Maps*
2. Attempt to install QGIS and Cytoscape on your laptop
3. Attempt some of the tutorials at Mapping and GIS for Historians - please note that these tutorials were developed for QGIS 2.x and some screenshots will not indicate the correct location of buttons, etc. in the 2018 Summer release of QGIS 3.x
4. Attempt to download and install R and the free R Studio Desktop OpenSource License. If you would like to get started with the world of R, consider a book such as *Learning R* by Richard Cotton, or any of the many online tutorials. *Note:* All the skills using R we will encounter in this module can also be done in the even more popular programming language of Python. If you think you might someday want to learn more programming, I'm happy to guide students towards resources that will allow them to do any of the module tasks in Python instead of R.
5. Browse and experiment with instructional materials at A Gentle Introduction to QGIS, Cytoscape Tutorials Wiki, and if you have your access to St Andrews e-resources, try Chapter 1 of *Text Analysis with R for Students of Literature* (library ebook)

DataCamp for Education

Before the module begins, you will be given an invitation email for free access to the courses available at datacamp.com. Your free access will be active from August for a period of six months and you may take the courses that are assigned through the online representation of this module on the platform. The platform is widely used for digital skills and data analyst training. If any of you are thinking of exploring careers in this area, I encourage you to use this (usually expensive) free account to learn a range of new digital skills.

MO5161 DataCamp Classroom 2020

DataCamp offers short interactive online courses (each of which can be completed in 3-6 hours of work), mostly showing you ways to employ the programming languages of R and Python for data science. It alternates between short film clips and interactive exercises in which you input code (I suggest using a desktop instead of tablet). During the semester, you will be required to either complete 4-6 short courses on DataCamp from a selection possible courses, or you may request an alternative research and writing task if you wish to avoid DataCamp altogether. While completion is required, your performance on these courses is not assessed with a formal mark.

All students are asked to complete: Introduction to R and Introduction to the Tidyverse. In addition, complete at least three additional mini-courses (* = suggested). There are a number of courses you will find in the "tracks" offered on DataCamp through our online class there. These include:

- Text Analysis: Intermediate R, *Introduction to Text Analysis in R, Sentiment Analysis in R: The Tidy Way*
- Strengthening R Foundations: String Manipulation in R with stringr, *Joining Data in R with dplyr*, Working with Data in the Tidyverse, Cleaning Data in R*, Communicating with Data in the Tidyverse, Working with Dates and Times in R, Importing Data in R (I, II), Data Visualization with ggplot2 (I, II)
- Data Science: Introduction to Data, Modeling with Data in the Tidyverse, Correlation and Regression, Fundamentals of Bayesian Data Analysis in R, Exploratory Multiple and Logistic Regression, Foundations of Inference, Experimental Design in R, Survival Analysis in R
- Spatial Analysis in R: Working with Geospatial Data in R, Spatial Analysis in R with sf and raster, Spatial Statistics in R
- Network Analysis: Network Science in R: A Tidy Approach, Network Analysis in R

Resources to Explore

As you do work on your projects you will find plenty of inspiration surveying the explosion of scholarship over the last decade or two. The following are some journals, hosts of projects, link hubs, and aggregators:

Digital Humanities Quarterly
International Journal of Humanities and Arts Computing (formerly *History and Computing*)
Digital Scholarship in the Humanities (formerly *Literary and Linguistic Computing*)
Language Resources and Evaluation (formerly *Computers and the Humanities*)
Historical Methods: A Journal of Quantitative and Interdisciplinary History
Journal of Cultural Analytics
Digital Literary Studies
Digital Studies / Le champ numerérique
Journal of Data Mining and Digital Humanities Digital Humanities Now
Spatial History Project at Stanford University
The Historical GIS Research Network
Historical GIS Lab Bibliography

Other journals that also have some relevant publications:

Social Science History
Historical Geography
Journal of Historical Geography
Social Science Computer Review
Transactions in GIS
Environment & Planning A/B/C/D/E
Environmental History
Annals of GIS

Striking Out on Your Own

During the semester, you may decide that you want to deepen your knowledge and skill set in one or more of the technologies that are introduced in the module. A great many of the humanities scholars who have adopted some of the digital methodologies, tools, and techniques that we will encounter in this module do not have much in the way of formal training in computer science or related fields. With the exception of some short training courses in GIS and one-off sessions at digital humanities workshops and unconferences, I have no formal training in any of the skills in this module. Self-learning with online resources, videos, and books are a common way to expand your abilities in these fields.

Programming Historian - This is by far the best resource online for tutorials covering a wide variety of digital humanities related tasks that are useful for historical research.
Mapping and GIS for Historians
GQIS Tutorials and Tips
ProfHacker
TAPoR 3 Research Tools for Studying Texts
GIS Lounge
Introduction to Programming for Humanists (Python)
Digital Humanities Slack Channel
The Spatial Community Slack Channel
Digital Humanities Answers
Stack Overflow

Learning GIS

Learning QGIS Anita Graser
Mastering QGIS Kurt Menke et al.
QGIS Map Design 2nd Edition Anita Graser & Gretchen N. Peterson
Discovery QGIS 3.x Kurt Menke
Historical GIS: Technologies, Methodologies, and Scholarship Gregory, Ian N., and Paul S. Ell, eds.
Introduction to GIS Victor Olaya
Geospatial Analysis: A Comprehensive Guide Michael J de Smith

An Introduction to Geographical Information Systems Ian Heywood, Sarah Cornelius
Thematic Cartography and Geovisualization Terry A. Slocum, Robert B McMaster et al
Spatial Data Analysis: An Introduction For GIS Users Christopher Lloyd
GIS: A Short Introduction Nadine Schuurman
An Introduction to R for Spatial Analysis and Mapping Chris Brunsdon, Lex Comber

Text Analysis

Text Mining with R Julie Silge, David Robinson
Text Mining in Practice with R Ted Kwartler
Text Analysis with R for Students of Literature Matthew L. Jockers
Applied Text Analysis with Python Benjamin Bengfort
Natural Language Processing with Python Steven Bird
Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text Taylor Arnold, Lauren Tilton
Mastering Regular Expressions Jeffrey E. F. Friedl
An Introduction to Statistical Learning with Applications in R Gareth James et al

Social Networks

A User's Guide to Network Analysis in R Douglas A. Luke
Mastering Gephi Network Visualization Ken Cherven
Social Network Analysis John Scott
Social Network Analysis: Methods and Applications Stanley Wasserman

General Texts

Data Visualization: A Practical Introduction Kieran Healy *Basic Statistics for Social Research* Augustine Kposowa, Mark D. Riddle, and Robert Hanneman
R for Data Science Hadley Wickham, Garret Grolemund
ggplot2: Elegant Graphics for Data Analysis Hadley Wickham
Learning R Richard Cotton
The Book of R Tilman M. Davies *Learning Base R* Lawrence Mark Leemis
Learning Python Mark Lutz
The Linux Command Line William E. Shotts

Week 0 - Orientation: History in Practice - Balancing Skills and Critique in the Digital Humanities

Our emphasis in this course is on a set of skills related to GIS and mapping, social networks, and tools for the analysis of texts which, to obtain even a moderate proficiency in, would each require several modules. These are skills that are often associated with a field that has come to be known as the *digital humanities*. This module should be seen as a kind of taster course that allows you to encounter a world of digital tools and methodologies that may be useful for working with large collections of sources. You may decide that most or all of these methods are not appropriate for your own research going forward, but some exposure to them will greater afford you the opportunity to understand them and, when appropriate engage or critique other historical scholarship that makes use of them.

We will have three primary goals in this semester:

1. In three broad parts, we will learn some of the basic principles and techniques of *historical GIS*, digital techniques for *text analysis and distant reading*, and some basic *visualisation of social networks*. In the two major assessments of the module you will have the opportunity to practice some of the techniques learned or, in the case of the final assessment, alternatively write a critical engagement with scholarship that makes use of these methodologies. This part of the module is arguably both the most fun and the most frustrating: you should prepare yourself for the sometimes considerable amount of time spent sorting out what can often be rather trivial technical obstacles along the way. One of the goals of this course is to introduce you to the problem solving mindset required to face these difficulties by effective use of online resources.

2. In the case of each set of technologies, we will read some historical scholarship that has made use of these approaches in order to better see how they are being employed in scholarship today. When reading these texts, it is not important for you to focus on the particulars of each historical case, but instead pay close attention to how the historian makes use digital methodologies in the support of their argument. We will try to highlight what added benefits there are in employing these methodologies, either alone, or in mixed method scholarship that integrates them with other historical approaches.

3. Finally, this module will, throughout, introduce you to some of the work of what has been called the *critical digital humanities*. We will examine some of the powerful critiques that have been made of GIS both among historians and within the world of geography, where it had its biggest impact on scholarship. We will similarly examine critical discourses on other digital methodologies and in our discussions of the essays applying them that we read. Ask yourself, how much does the argument depend on the technological approach? What assumptions are made in setting the stage for the scholarship, in preparing the date, and in using the technologies? What are the costs implied, and the benefits gained? To what degree does the choice of methodology influence the kinds of questions asked and outcomes obtained?

Preparation

Introductory Reading

Moretti, Franco. *Atlas of the European Novel, 1800-1900* Verso 1998, p1-73 (Ch 1: The Novel, the Nation-State)

Harley, J. B. "Deconstructing the Map" in Michael Dear and Steven Flusty eds. *The Spaces of Postmodernity: Readings in Human Geography*, 277-288

Suggestions for Play and Exploration

- Browse the table of contents of some recent issues in the journals listed above in the *Resources to Explore* section and take note of titles you may want to revisit for inspiration for your future projects.
- If you haven't done so already, consider signing up for Slack, and joining the Digital Humanities slack and Spatial Community slack channels, both good places to learn from others passively on the channels hosted or to ask questions.

Week 1 - Introduction to Historical GIS and First Look at QGIS

This week will explore the field that has become known as Historical GIS, but more broadly the analysis of geographic data within the humanities in general, as well as some of its basic approaches and promise. Beginning this week, and continuing through week four, you will all be asked to give short presentations and prepare summary handouts on some examples of scholarship making use of GIS. We will set aside time this week to discuss the first project assessment, and some potential topics and sources that you may consider using for the assessment.

Preparation

1. Examine two of the most famous maps cited in scholarship on mapping and GIS: Charles Joseph Minard's map of Napoleon's Russian Campaign, and the famous cholera map of John Snow (you can read more about this here: Map-making and Myth-making in Broad Street. Then take a look at Joseph Priestley's *A New Chart of History*. Why have the come to be regarded as excellent historical examples of visualisations that communicate an idea very clearly? Where might a critique of them begin?

2. If you haven't already done so, please try to **download and attempt to install QGIS**. If you have problems, please make note of what issues appear and bring them in.

General Tips Regarding Installation Problems: Many of the multi platform open source applications out there these days make use of Java. Sometimes the version of Java pre-installed on your laptop (or, indeed, not installed or active) will be the key problem for getting the application to function smoothly. Installing or updating the appropriate Java version can often fix these issues. Also, if you find yourself presented with unusual error messages, searching google with a large part or all of the error message in quotation marks will often lead you to bulletin boards where people have faced similar issues and tried different steps to resolve the issue (though *please* evaluate recommended instructions on random bulletin boards with great care and ensure your machine is backed up before trying anything too elaborate!). The website Stack Exchange often has useful discussions that address software issue. The PC Classrooms on campus should have QGIS 3.x installed and you may use them for your project work.

3. Using the Microsoft Teams Sign-up tab, **sign up to do one 5-10 minute non-assessed presentations** (not a minute more!) each on the "application" texts (one student per text) in the *Reading Selection* list below. These presentations will be given Week 2, Week 3, or Week 4. In addition, sign up for a second and third text which you will prepare a handout for, but not actually present on.

    - **Make a handout about each of the three texts** (only one of which you will present on) which is on a single or double-sided page. Submit the digital file for the handout as a PDF to the module Team Files tab for everyone's access before the class day that you will present on.
    - In the presentation comment on the way that GIS or mapping was used in the text. Comment on what data was used for the project and what kind of effort might have been involved in preparing it for GIS. Were there any innovative or useful ideas to draw from the text as potential inspiration for our projects? Was GIS used primarily for *heuristic*, *illustrative*, or *explorative* purposes, visualising sources or data in order to discover new questions and spot patterns to press the historical research forward? Was GIS employed using formal *analysis*, that is to say, did it make use of quantitative methodologies, spatial statistics, or other algorithmic techniques (if so, list what specific ones were used in the handout) towards the end of making concrete spatial claims?

Required Readings

- KNOWLES Ch 1 "GIS and History" 1-27 (see Files)
- GREGORY Ch 1 "GIS and its role in historical research: an introduction" 1-19
- GEDDES Fragment of Ch 4, only section "Environmental Management" 104-111

Reading Selection for Presentations on Historical GIS Scholarship:

For the coming weeks, choose one of the following chapters for the week of your presentation with a handout plus two for just a handout. Please skim 3-5 other chapters of interest over the next few weeks for inspiration for your map project.

See the preparation section above for presentation info.

GEDDES
Ch 1 "Railways and Agriculture in France and Great Britain" 4-30 Schwartz and Thevenin
Ch 2 "The Development, Persistence and Change of Racial Segregation in U.S. Urban Areas" 35-59 Andrew Beveridge
Ch 3 "Troubled Geographies: A Historical GIS of Religion, Society, and Conflict in Ireland since the Great Famine" 62-83 Ian Gregory
Ch 5 "The Politics of Territory in Song Dynasty China" 118-140 Ruth Mostern

KNOWLES
Ch 2 "Creating a GIS for the History of China" Peter K. Bol (see Files)
Ch 4 "Scaling the Dust Bowl" Geoff Cunfer (see Files)
Ch 6 "Mapping Husbandry in Concord: GIS as a Tool for Environmental History" Brian Donahue
Ch 8 "New Windows on the Peutinger Map of the Roman World" Talbert and Elliot

PAST
Ch 2 "Teaching the Salem Witch Trials" Benjamin Ray (see Files)
Ch 3 "Similarity and Difference in the Antebellum North and South" Aaron C. Sheehan-Dean (see Files)
Ch 6 "Redlining in Philadelphia" Amy Hillier (see Files)

BODENHAMER
Ch 7 "Mapping Text" May Yuan

DEEP
Ch 4 "Inscribing the Past: Depth as Narrative in Historical Spacetime" Ethington and Toyosawa
Ch 7 "Spatializing and Analyzing Digital Texts: Corpora, GIS, and Places" Gregory, Cooper, Hardie, and Rayson
Ch 8 "GIS as a Narrative Generation Platform" Yuan, McIntosh, and Delozier

SPATIAL
Ch 1 "Re-focus on Women in an Industrial Revolution…" Sherry Olsen
Ch 4 "Railroads and Population Distribution…" Alvarez-Palau and Martí-Henneberg
Ch 6 "Mapping the American Iron Industry" Anne Kelly Knowles
Ch 9 "Geographies of Welfare in Nineteenth Century England and Wales" Douglas H L Brown
Ch 13 "Kleindeutscland: the lower east side in new york city…1880s…" Kurt Schlichting
Ch 15 "A City of the White Race Occupies its Place…" Lutz et al
Ch 17 "The Post, The Railroad and the State" Gustavo Velasco
Ch 19 "Food, Farms, and Fish in Great Britain and France, 1860-1914" Robert M. Schwartz
Ch 20 "White Maps and Black Votes…" Don DeBats
Ch 21 "The Spatial History of State Power" Ruth Mostern
Ch 22 "Peasants and Politics - How GIS Offers new insights into the German Countryside" George Vascik
Ch 23 "Mapping Inequality 'Big Data' meets social history in the story of redlining" Connolly et al
Ch 25 "Urban Property in Nineteenth-Century Rio De Janeiro: Rent, neighborhoods, and networks" Zephyr Frank

RAU
Ch 3 "Peopling the Past: Interpreting Models for Pedestrian Movement in Ancient Civic-Ceremonial Centres" Morton et al
Ch 5 "Lyons, the Spatial Analysis of a City in the 17th and 18th Centuries. Locating and Crossing Data in a GIS Built from Written Sources" Gauthiez and Zeller
Ch 6 "Mapping Long-term Urban Space Structures: Barcelona as a Case Study" Bassols and Garriga

Dear, Michael J. *GeoHumanities: Art, History and Text at the Edge of Place* Routledge, 2011. Library ebook Ch 25 "Without Limits: Ancient History and GIS" Alexander von Lünen et al.
Ch 26 "History and GIS: Railways, Population Change…" Robert M. Schwartz et al.
Ch 30 "What do Humanists Want? What do Humanists Need? What Might Humanists Get?" Peter K. Bol

Suggestions for Play and Exploration

- Why All World Maps Are Wrong - watch this nicely composed Vox video on projections, and check out this XKCD comic on map projections.
- Play with the Mercator project and nation sizes here.
- Stanford Spatial History - Gallery of Projects - Visit this gallery of projects and examine a few of them for possible inspiration. Note that some of them merely have screenshots (many of them were once interactive with the plug-in Flash, which is no longer supported by its developers) and reflect on the problem of technological obsolescence on the web.
- Think about historical works that have impressed you or which you think are important for your area of interest. Are there aspects or sources which are or might benefit from a more spatially explicit approach? Do the works make use of maps or spatial visualization in an effective way? Could they have? If so, how might it have been done? What sources in your area of interest might lend themselves well to *heuristic* or *analytical* or *illustrative* use of maps and GIS?
- If your chosen readings for presentations include work that use analytic GIS techniques or spatial statistics, look up the method that was used and read more about it. Find out if/how the method can be done within QGIS, evaluate the strengths or weaknesses of the method. What kinds of assumptions does it make or other problems may arise?
- Look up one or more of the source bases used in some of the reading selections for presentations. Are any of them available online and easily accessible? Are sources like it or of the same type or class available in your area of interest and easily accessible? How might you get a hold of them for use in your own research?
- Start thinking about your first assessment and come up with half a dozen possible ideas to explore. Start chasing down information for it, or putting in inter-library loan requests if required.
- If you don't have it yet, apply for a SCONUL card in the library to help you get access to other large research libraries in the area. Consider getting a reader card at the National Library of Scotland and read up on the procedure for the use of the library, which is in Edinburgh.
- Look to the weeks ahead and identify techniques or skills that you think you might be especially interested in. Jump ahead and start learning in that area.

Week 2 - QGIS: Basics and Georeferencing Historical Maps

We will open today with a discussion of some basic GIS terminology, concepts, and the components which are required for using GIS software. We will discuss the readings and have presentations on some applications of GIS in historical scholarship. We will then use most of the remainder of the time to work on learning the basic functionality of QGIS and the georeferencing of historical maps, which is a common task for many history projects. We will set aside class time to further talk about the first assessment, the expectations and share thoughts on what you wish to do in this regard.

Some of the QGIS basics we will work on, time permitting include:

- Adding and manipulating layers in QGIS
- Adding a simple layer with some points from a CSV file
- Importing shapefiles
- Adding basemaps
- Georeferencing and georectifying a historical map
- Adding additional features to a layer using a georeferenced or other basemap
- Basic querying of spatial data on maps
- Basic choropleth and graduated point symbol maps

Preparation

1. Please come to class with your laptop, having installed QGIS: QGIS Download and Installation. In class we will together work through a few of the tutorials at Mapping and GIS for Historians
2. Watch some of these videos and try some of this yourself, some of it will be review material from last week: QGIS 3 Tutorials for MO5161
3. We will have some presentations today, see last week for more on how to prepare.
4. Some of the readings today point to the controversies over the use of GIS that created considerable tension within the field of Geography and has similarly sparked debates among historians. As these will emerge multiple times, consider noting down some of the main issues and different perspectives along with some of the major participants in these debates mentioned.
5. There are a number of technical concepts that arise this week. In fact, some of them show up in two or three of the readings so if you understood it clearly the first time, you can skim or merely review the understanding of the concept in the other readings. If they are not clear from the readings, consider looking into them a bit more: Data models, Vector and raster data, MAUP modifiable areal unit problem, Ecological Fallacy, Spatial autocorrelation, Generalization (in the context of GIS), Least cost path analysis, Epistemology and Ontology (in the context of GIS)
6. Browse the David Rumsey historical map collection and especially its offerings in the areas of your interest.

Required Reading

Victor Olaya *Introduction to GIS* 2018 pp1-36, pp87-98 - This is a wonderful short volume that covers a great deal of the basic terminology and concepts. Very affordable physical copy available for order online, or various free formats here)

Nadine Schuurman *GIS: A Short Introduction* Ch 2 "GIS, Human Geography, and the Intellectual Territory Between Them" pp21-52 (see Files)

DEEP Ch 3 "Genealogies of Emplacement" pp54-71

Further Reading

GREGORY Ch 4, 5, 6, 8

Monmonier, Mark. *How to Lie with Maps*. 3rd Edition. University of Chicago Press, 2018.

Schuurman *GIS: A Short Introduction* Ch 3 "The Devil is in the Data", Ch 4 "Bringing it All Together"

John Pickles ed. *Ground Truth: The Social Implications of Geographic Information Systems* 1995 esp. Preface, Ch 1, 3, 7

Hill, Linda L. *Georeferencing: The Geographic Associations of Information*. Digital Libraries and Electronic Publishing. Cambridge, Mass: MIT Press, 2006, Ch 5 Gazetteers and Gazetteer Services Library ebook

Fotheringham, A. Stewart et al. *The SAGE Handbook of Spatial Analysis*. (2009).
Ch 6 Spatial Autocorrelation
Ch 7 The modifiable area unit problem (MAUP)
Ch 23 Spatial Analaysis on a Network

Reference Reading

Note: QGIS is now at 3.x so some instructions may be slightly off when it comes to some features in various links below

QGIS 3 Tutorials for MO5161 - This is a collection of YouTube tutorials I created for this module. Over the next few weeks, please watch as many of these as you can, and try the exercises yourself in QGIS. Mapping and GIS for Historians - This was designed for QGIS 2 but most tutorials should still work with little change. A Gentle Introduction to QGIS - Documentation for QGIS 3.4

The following may be useful to you in getting acquainted with QGIS. See also the *Striking Out on Your Own* section above in this handbook:

Graser, Anita. *QGIS Map Design* 2nd Edition (Not in Library) - Really great collection of task oriented tutorials.
Menke, Kurt *Discovery QGIS 3.x* (Not in Library) - Lots of interactive tutorials in this work.

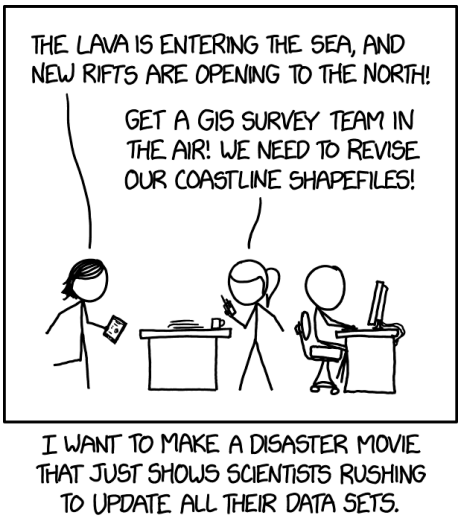
QGIS Training Manual
GQIS Tutorials and Tips

Suggestions for Play and Exploration

As you learn some of the basics of using QGIS, especially the georeferencing of historical maps, creating new vector layers from the information, and visualising historical data with choropleth maps and graduated symbols, spend some serious time this week looking at potential sources out there: historical maps, historical data (already digitised and georeferenced, or not), spatial claims that can be illustrated or potentially questioned with spatial analysis. Also use play time to practice your skills working with georeferencing maps of different kinds. Start thinking about the kinds of obstacles and challenges you may face among your project ideas.

Week 3 - QGIS: Techniques for Exploration and Analysis of Spatial Data

Our goal in this second hands-on QGIS session is complete a review of any techniques we didn't find time for in the previous week, and then continue with further skills from the list below. We will continue with the presentations that were signed up for in week one, but the readings for this week, which are generally in the area of critical GIS, will be mostly rolled into our discussion in week four.



XKCD Comic 2029 "Disaster Movie"

Some QGIS techniques for today, as time permits:

- Creating a heat map
- Creating and using buffers
- Clipping, difference, and intersect
- Joins (Spatial, Table, Summary), Counting Points in Polygons, and Field Calculation
- Counting points in polygons
- Distance Matrix and Nearest Neighbor Analysis
- Depicting routes
- Creating a time series in QGIS + alternative online tools
- Some other geoprocessing techniques
- Exporting your map and post-processing
- GeoJSON and its uses

Preparation

1. QGIS 3 Tutorials for MO5161 - Please continue doing tutorials on our website
2. See the module Moodle for an updated list of YouTube videos to watch on the some of the above techniques, along with some reference websites related to the QGIS techniques we will discuss. Please watch as many of the videos as you can, but you may skip those for techniques you are don't think you will ever need. There are some exercises there for you to try in preparation for our meeting.
3. If you are giving a 3-5 minute presentation, please prepare it and if you haven't distributed it yet, submit and bring copies of a third chapter as instructed above.
4. This week you should should use almost all of your time practising QGIS, in play and experimentation and directly working on your project. Make progress this week to allow for changed directions, frustrations, or distractions to come.

Required Reading

Joanna Drucker "Humanities Approaches to Graphical Display"

BODENHAMER Ch 3 "Geographic Information Science and Spatial Analysis for the Humanities" pp31-73

KNOWLES Ch 5 "'A Map is Just a Bad Graph': Why Spatial Statistics are Important in Historical GIS" pp123-150 (see Files)

Further Reading

Joanna Drucker "GIS Analysis and Critical Issues"

Albrecht, Jochen *Key Concepts & Techniques in GIS* Sage, 2007.

Lloyd, Christopher *Spatial Data Analysis: An Introduction For GIS Users* Oxford University Press, 2010

Smith, Michael J. *Geospatial Analysis: A Comprehensive Guide*. Drumlin Security Ltd, 2018.

Reference Reading and Links

QGIS Documentation: Vector Overlay - Clip, Difference, Intersection, Union, etc.
QGIS Documentation: Buffers
QGIS Training Manual: Vector Analysis
QGIS Training Manual: Network Analysis
QGIS Training Manual: Spatial Statistics
QGIS Documentation: Interpolation
QGIS Tutorial: Spatial Joins
QGIS Tutorial: Table Joins
QGIS Tutorial: Points in Polygon
QGIS Tutorial: Heat Maps - Note: this has changed in QGIS 3.x
QGIS Tutorial: Nearest Neighbor Analysis
QGIS Tutorial: Interpolation

Introduction to QGIS Time Manager
Anita Graser - Index of Posts on QGIS Time Manager

Brief Introduction to GeoJSON
GeoJSON.io

Suggestions for Play and Exploration

- Determine what QGIS skills will be most important for your project and look for more instructional websites and youtube clips about these techniques. Practice on some data, even if you don't have your own project data ready.
- Use remaining time to explore potential sources and work on your project, experimenting and testing ideas.

Week 4 - Thinking With/Against Maps and Spatial Analysis

For several weeks now we have been learning some very basic skills in QGIS and mapping. The promise of using GIS software for mapping, and the more advanced geographical analysis that becomes possible with it should be apparent. This week we will check with all of you on the progress in your work on the map project. We will discuss what constitutes effective use of maps and some of the ways that your map project will be evaluated. We will also have any remaining short presentations on the readings on applications of historical GIS. All of our remaining time, however, will be dedicated to a discussion of the various critical and theoretical readings we have done about GIS as a tool and methodology.

Preparation

1. Any remaining presentations should be prepared for and your additional handout should be submitted and distributed.
2. Be prepared to share your work on your map project so far, if you are working on one, including the kinds of sources you are using or evaluating, what challenges you have had so far, and questions you might have. If you are working on a critical essay, be ready to share what direction you will be taking it.
3. Make substantive progress on your first project or essay.
4. Optional: Sometimes GIS is just overkill and creating a good map for illustration just needs some time spent tracing existing maps and adding your own features. For this, learning how to use Bezier curves is a great plus but requires that you have Vector drawing software. Commercial packages like Adobe Illustrator and Affinity Design among many other options work great but cost money. If you think you might want to trace maps, try learning how to use Bezier curves with the Bezier game. Install Inkscape or download free version of Gravit Designer (or use the online version) and try tracing a simple map. A few more videos on how to use Bezier curves: 1, 2, 3,4

Required Reading

Curry, Michael *Digital Places: Living with Geographic Information Technologies* Library ebook
Ch 1 "Reason and Language in Geographic Information Systems" 11-23
Ch 2 "On Space in Geographic Information Systems" 24-38

LÜNEN
Ch 11 "GIS and Research into Historical 'Spaces of Practice': Overcoming Epistemological Barriers" Sam Griffiths
Ch 12 "GIS and History: Epistemologies, Reflections, and Considerations" Charles Travis
Ch 14 "Tracking in a New Territory: Re-imaging GIS for History" Alexander von Lünen

Further Reading

Wilson, Matthew W. *New Lines: Critical GIS and the Trouble of the Map* 2018 (not in library)

Pickles, John. *A History of Spaces: Cartographic Reason, Mapping and the Geo-Coded World*. Routledge, 2012.

Massey, Doreen B. *For Space*. London; Thousand Oaks, Calif.: SAGE, 2005.

Cope, Meghan, and Sarah Elwood *Qualitative GIS: A Mixed Methods Approach* Sage, 2009

"Elements of Historical Knowledge About Urban Spaces: Reflections on the Requirements for a Dynamic Map" Ekkehard Schönherr in Rau, Susanne, Ekkehard Schönherr, eds. *Mapping Spatial Relations, Their Perceptions and Dynamics: The City Today and in the Past. Lecture Notes in Geoinformation and Cartography*. Cham ; New York: Springer, 2014. Library ebook

Jobst, Markus, et al. *Preservation in Digital Cartography: Archiving Aspects*. Berlin ; London: Springer, 2010. Library ebook

Reference Reading

Slocum, Terry A., Robert B. McMaster, Fritz C. Kessler, and Hugh H. Howard. *Thematic Cartography and Geovisualization*. 3 edition. Pearson, 2013.

Tufte, Edward R. *The Visual Display of Quantitative Information*. Graphics Press, 2001.

Suggestions for Play and Exploration

- See earlier weeks for some suggestions. However, the majority of time should now be devoted to the development of your map project.

Week 5 - Text Analysis: Using the Command Line, Regular Expressions

The main task of today is to work with regular expressions (regex). The art of regex is an important foundation skill for mining textual information in many different contexts. A regex is a collection of symbols and characters used to identify a pattern, allowing you to search large bodies of text to *match* a designated combination of characters, words, or phrases which might, for example, simultaneously exclude other characters or words. We can do this directly through a web browser on certain sites but some dialect of regular expressions is built into powerful word processing software, many command line utilities such as *grep* and almost every programming language. We will also explore some of the other possibilities of using the command line on your computer, and software for OCRing documents that can then be searched or cleaned up with regular expressions. The various tasks below asking you to install software are optional. They are useful if you want to learn and try working with utilities on the command line (and you don't have a Mac) or work with Tesseract, but this will not be required, especially if you don't plan on doing any OCRing of historical documents, or plan to use other commercial OCR software.

XKCD Comic 208 "Regular Expressions"

Preparation

1. Please install a good text editor that can support regular expressions. You can download the free Microsoft Visual Studio Code available for Mac, Windows, and Linux. There is a significant learning curve, but if you want to go full scale 1970s classic editing power, I am personally a daily user of (and this handbook was composed in) the powerful tool VIM (Mac OS X users see also MacVim and VimR). Its bitter arch-enemy (the two of them have been battling it out for supremacy in the geek community since 1976) is Emacs.

2. Optional: Install Tesseract. This can be done easily via HomeBrew on OS X, or the package manager if you will be using your VirtualBox as per instructions on the Tesseract github page. There is also a windows installation but I'm not familiar with its use.

3. In class we will work though the Interactive Tutorial for regular expressions, but give it a try before coming to class. Read two somewhat silly examples of using regular expressions here: Reformatting Confucius with Regular Expressions and Finding the Women of Heimskringla with Regular Expressions

4. If you get the hang of regular expressions before class, try Regex Crossword (This only works in Safari and Chrome for me, try a different browser if it doesn't work for you).

5. There is very little required reading today. Take advantage of this time to work on your mapping project, due after independent learning week.

6. By this week, please have completed at least the Introduction to R course on DataCamp.

Reading

Introduction to Bash

Interactive Tutorial - Again, please go through this tutorial. We'll practice more in class.

'Q i-jtb the Raven': Taking Dirty OCR Seriously Ryan Cordell

OCR Failures in 2016 Ben Schmidt

Further Reading

Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers Carolyn Strange et al. in *Digital Humanities Quarterly*

Understanding Regular Expressions - a tutorial using LibreOffice Writer

Cleaning OCR'd text with Regular Expressions - a tutorial using Python programming language

In the Beginning was the Command Line, 1999 by novelist Neal Stephenson (skim)

ProfHacker posts on the Command Line

Reference Reading

Jeffrey E. F. Friedl *Mastering Regular Expressions* 2nd Edition O'Reilly - very helpful and thorough overview of regular expressions Tony Stubblebine *Regular Expressions Pocket Reference* 2nd Edition O'Reilly

Week 6 - Independent Learning Week

Week six of the fall semester is an independent learning week at St Andrews. Independent Learning Week is a learning week, not a holiday; it is integrated into the curriculum. The majority of time this week should be dedicated to the completion of your map project and on catching up on any reading you may have fallen behind on. In addition, please read the following short pieces to prepare for the transition to the second part of the semester:

Joanna Drucker, "Humanistic Theory and Digital Scholarship" in Gold, Matthew K., eds. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. Library ebook

Where to Start with Text Mining

Literature Is not Data: Against Digital Humanities

Using Metadata to find Paul Revere

Putnam, Lara. "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast- The Transnational and the Text-Searchable." The American Historical Review 121, no. 2 (April 1, 2016): 377–402. https://doi.org/10.1093/ahr/121.2.377.

Week 7.1 - Introduction to R

This week we'll talk about R, which, together with the programming language Python are the two most commonly used programming languages for digital humanities, statistics, machine learning, and the whole world of data science. Although I personally enjoy Python more, and it is a more versatile programming language used in a wide range of tasks, my experience been that the welcoming environment of R Studio and the ease of adding and making use of libraries seems to be slightly better for new beginners. Anything in the module that can be done with R can be done with Python. Let me know if you want resources in that area and feel free to tackle the Python equivalents of our module on DataCamp instead of R.

Preparation:

1. Attempt to download and install R and the free R Studio Desktop OpenSource License.

2. If you haven't done anything except the Introduction to R on DataCamp, proceed with at least the Introduction to the Tidyverse course and consider making progress on other courses there.

Required Reading

JOCKERS
Ch 1 R Basics

SILGE
Ch 1 The Tidy Text Format

Further Reading

Cotton, Richard *Learning R* O'Reilly (2013) - if you are thinking of doing your project with R, I would suggest buying yourself a copy of this for reference
Wickham, Hadley et al *R for Data Science* O'Reilly (2017) - another book you might consider owning
Davies, Tilman M. *The Book of R: A First Course in Programming and Statistics*

Week 7.2 - Text Analysis: Frequencies and Correlations

This week we will focus on some of the basic tasks carried out in text analysis: breaking texts down into tokens, getting frequencies for these in a text, and exploring correlations. You will *not* be required to use R in your final project, but take this opportunity to acquaint yourself with some of the basics of R and some of the ways it can be used. While here we use it for text analysis, it can also be used for regular expression search, GIS, network analysis and visualization, and many other useful tasks.

Preparation:

1. The technical readings this week will reproduce some of what we try in class so don't worry if you can't make it through everything. We'll step through some of this within class. To take a break from R, download a raw text file from one of the books at Project Gutenberg and input it, or a piece of it, into Voyant Tools, which does some of the basic tasks that we are learning to do by other means.

Required Reading

Note: See below for the 8.1 meeting readings as well

Read One of the Following:

A) Ramsay, Stephen *Reading Machines: Toward an Algorithmic Criticism* University of Illinois Press (2011) Ch 1 An Algorithmic Criticism pp1-17 Ch 5 Patacomputing pp69-81 Library ebook

B) Underwood, Ted *Distant Horizons : Digital Evidence and Literary Change* Preface, Ch 1-2. Library ebook

Do Digital Humanists Need to Understand Algorithms? Ben Schmidt in *Debates in Digital Humanities*

The Foreign Language of 'Mad Men'

SILGE
Ch 3 Analyzing Word and Document Frequency: tf-idf
Ch 4 Relationships Between Words: N-grams and Correlations (up to p54 where network visualization begins)

If you struggle with SILGE, which uses a new convenient package and tidytext appraoch, but offers less in the way of explanation in the book, you can, alternatively, try: JOCKERS Ch 2-5

Further Reading

Textual Analysis introduction by David Hoover

Text Analysis, Data Mining, and Visualizations in Literary Scholarship

Alien Reading: Text Mining, Language Standardization, and the Humanities Jeffrey Binder in *Debates in Digital Humanities*

Mad Men anachronism hunting Making Downton more traditional

Weiss, Sholom M., Nitin Indurkhya, Tong Zhang, and Fred J. Damerau. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. New York: Springer, 2005.

Daniel Rosenberg, "Data before the Fact" from *"Raw Data" is an Oxymoron*, ed. Lisa Gitelman (MIT Press, 2013)

Rockwell, Geoffrey. "What Is Text Analysis, Really?" Literary and Linguistic Computing 18, no. 2 (June 1, 2003): 209–19. https://doi.org/10.1093/llc/18.2.209.

Text Data Preparation: a Practice in R using the Sheng Xuanhuai Collection - example of tackling the problem of segmenting Chinese text using jiebaR package and then quanteda for analysing the text.

Reference Reading

Google Ngrams Culturomics Hathi Trust Digital Library

Week 8.1 - Text Analysis: Topic Modelling and Sentiment Analysis; Introduction to Alternative Tools

This meeting will continue the work we did in the last session. It is simply not possible to do more than a taster of either topic modelling or sentiment analysis but it opens the way to students who wish to explore more on their own, but also be aware of some of the debates that have emerged around the use of these techniques. We'll also look briefly at Orange as an example of a more interactive visual interfaced application to analyse text.

Preparation

1. You will not have had much time to prepare given this is our second meeting this week. Do you best to look through the online resources below and at least skim some of the technical reading. We'll try to do some of the technical tasks in class together.
2. Come to class with some ideas for your final project. As you think about what to work on, browse the work in the sessions to come. Especially start thinking about interesting primary sources that would be useful to approach with some of the methods we have been looking at unless you are going to be writing a critical essay.

Suggestions for Play and Exploration

If you are interested in using Sentiment Analysis in your final project, then go through Julia Silge's course on DataCamp: Sentiment Analysis in R: The Tidy Way. If you are interested in exploring Topic Modeling more, consider the DataCamp course Topic Modeling in R with Pavel Oleinikov.

If you are interested in alternatives to R (or Python) for text analysis, take a look at Orange which is a data analysis software package. If you are interested in the tool or looking for ways of analysing text without using R, take a look at some of its documentation or their Youtube tutorials

Required Reading

Probalistic Topic Models David M. Blei *Communications of the ACM* (2012)

Words Alone: Dismantling Topic Models in the Humanities Benjamin M. Schmidt

Language is Biased. What Should Engineers Do? Ben Schmidt

Odd Numbers: Algorithms Alone Can't Meaningfully Hold Other Algorithms Accountable Frank Pasquale

Either SILGE Ch 2 Sentiment Analysis with Tidy Data and Ch 6 Topic Modelling OR JOCKERS Ch 11 Clustering and Ch 12 Topic Modeling

Further Reading

The Hermeneutics of Data and Historical Writing

Discovery and Justification are Different: Notes on Science-ing the Humanities

When you have a MALLET, everything looks like a nail

Vector Space Models in the Digital Humanities and Rejecting the gender binary: a vector-space operation - Ben Schmidt

The Ground Truth of DH Text Mining Tanya E. Clement

Problems with the Syuzhet Package Annie Swafford

Mining the *Dispatch* project

Getting Started with Topic Modeling and MALLET - this tutorial runs MALLET directly from command line

Text Mining: An Annotated Bibliography

Bibliography on Topic Modeling by David Mimno

Week 8.2 - Networks: Introduction to Social Network Analysis and Cytoscape

At one level the study of historical networks and relationships between people is at the heart of the sub-field of "prosopography" which we will examine more closely today. The heart of prosopography work has often been the work of developing a good database, which unfortunately is not something we will have time for here. We'll examine some examples, however, of prosopography projects and think about how social network analysis can be put to good use in historical work.

Preparation

1. Please visit and look over the following websites:

Prosopography Research
Domesday - Prosopography of Anglo-Saxon England.
Prosopography of the Byzantine World
China Biographical Database Project

2. Please attempt to install Cytoscape on your laptop if you can and look over some of these Cytoscape Tutorials

Suggestions for Play and Exploration

If you are interested in exploring more quantitative approaches to Network Analysis, using R, then you can explore the DataCamp courses for this in R (Network Analysis in R, Network Science in R: A Tidy Approach). I would also suggest you take a look at Barabási's textbook *Network Science* to tackle the methodological foundations.

Reading

"Demystifying Networks" Scott Weingart

SCOTT 4th edition (see Files) Ch 1 What is Social Network Analysis
Ch 5 Terminology for Network Analysis
Ch 6 Popularity, Mediation and Exclusion

(If you have 3rd edition of Scott then read instead: Ch 1 Networks and Relations, 4 Lines, Neighborhoods and Densities, 5 Centrality, Peripherality and Centralization instead)

Further Reading

Consider reading more of SCOTT

Robert A. Hanneman and Mark Riddle *Introduction to social network methods* Online textbook

Miriam Posner's Glossary of Social Network Analysis

Peter Bol "GIS, Prosopography, and History"

Folsom, Ed. "Database as Genre: The Epic Transformation of the Archives" *PMLA* Vol. 122 No 5 Oct 2007, 1571-1579.

Lev Manovich Database as a Symbolic Form

The following SAGE handbook has sometimes very technical chapters but is especially valuable for anyone interested in using network analysis in their final project given its rich bibliographies for each sub-topic. Can be a great starting place for exploring the range of scholarship on networks of particular kinds and applied to particular themes:

Scott, John, and Peter J. Carrington, eds. *The SAGE Handbook of Social Network Analysis*. (2011)
Ch 2 Social Network Analysis: An Introduction pp11-25
Ch 4 Network Theory pp40-54
Ch 23 A Brief Introduction to Analysing Social Network Data pp331-339
Ch 24 Concepts and Measures for Basic Network Analysis pp340-369
Ch 27 Qualitative Approaches pp404-416
Ch 28 Analyzing Affiliation Networks pp417-433
Ch 29 Positions and Roles pp434-446
Ch 8 Personal Communities: The World According to Me pp101-115

Ch 13 Corporate Elites and Intercorporate Networks pp180-195
Ch 14 Political Dimensions of Corporate Connections pp196-209
Ch 15 Policy Networks pp210-222
Ch 16 Social Movements and Collective Action pp223-235
Ch 17 Crime and Social Network Analysis 236-255
Ch 19 Scientific and Scholarly Networks 271-285
Ch 20 Cultural Networks 286-300

Also useful for anyone wanting to explore more advanced topics with SNA, the following book is a classic core text in the field and covers many of the details of the quantitative foundations of the field:

Wasserman, Stanley, and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

Examples of Research on Networks

Lindner, Ulrike. "Transnational Movements between Colonial Empires: Migrant Workers from the British Cape Colony in the German Diamond Town of Lüderitzbucht." European Review of History: Revue Europeenne D'histoire 16, no. 5 (2009): 679–95.

David S. Lux and Harold J. Cook, 'Closed Circles or Open Networks? Communicating at a distance during the scientific revolution', History of Science 36 (1998): 179-211.

James Secord, 'Knowledge in Transit', *Isis* 95 (2004): 654-672

Rodogno, Davide, Bernhard Struck, and Jakob Vogel, eds. *Shaping the Transnational Sphere: Experts, Networks and Issues from the 1840s to the 1930s*. New York: Berghahn Books, 2014. (Introduction)

Jasanoff, Maya. *Liberty's Exiles: How the Loss of America Made the British Empire*. London: HarperPress, 2011.

John Law, 'On the methods of long-distance control: vessels, navigation and the Portuguese route to India', in John Law (ed.), Power, Action and Belief. A New Sociology of Knowledge? (London, 1986), pp. 234-263. Not in library, but available to download here

David Livingstone, *Putting Science in its Place: Geographies of Scientific Knowledge* (Chicago, 2003), Chapter 1 'A Geography of Science?' and Chapter 4 'Circulation: Movements of Science'.

Bruno Latour, 'The Powers of Association', in John Law (ed.), *Power, Action and Belief*. A New Sociology of Knowledge? (London, 1986), pp. 264-280.

Pierre-Yves Saunier, 'Circulations, connexions et espaces transnationaux', *Genèses* 57 (2004), 110-126

Historical Network Research - Bibliography

If you are interested in Social Network Analysis and want to explore this more:

Reading

LUKE Ch 2-3 pp11-41

Wetherell, Charles. "Historical Social Network Analysis." *International Review of Social History* 43, no. Supplement S6 (December 1998): 125–44. doi:10.1017/S0020859000115123.

Lemercier, Claire. "Formal Network Methods in History: Why and How?," December 7, 2011.

If Everything is a Network, Nothing is a Network Mushon Zer-Aviv

Further Reading

Consider reading on in LUKE to find more about network visualization design and some of the possibilities for more advanced analysis.

Erickson, Bonnie H. "Social Networks and History: A Review Essay." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 30, no. 3 (January 1, 1997): 149–57. doi:10.1080/01615449709601182.

More from Scott Weingart Series on Demystifying Networks:
DN2 Degree, DN3 Power Law Rant, DN4 Co-Citation Analysis, DN5 Communities, PageRank, and Sampling Caveats, DN7 Doing Co-Citation Analysis, DN8 When Networks are Inappropriate, DN9 Bimodal Networks

Week 9 - Mapping Texts and Networks; Mediums for Sharing Your Work

This final meeting will discuss some ways that maps, texts, and networks can be combined and visualised together. We'll talk about the problem of preservation in the digital realm, and discuss again some of the critical discourses around the digital humanities.

Preparation

1. Focus your efforts on developing ideas for the final project and ask questions in class if you want help with ideas or direction
2. Some tools to explore:
    - Palladio
    - Recogito
    - InfraNodus or Textexture
3. Take a look at Cytoscape.js, a way to put your network graphs online and make them interactive.
4. Prepare for a discussion about the problem of the maintenance and longevity of formats, outputs and software in the digital world. Think about the example of Gephi. Look at the history of commits to the open source software Gephi: Gephi Github Commit History, the Gephi Release History and the Github Insights for the project. What observations can we make about the history and the health of the software? What concerns might we have about its future? Is it as easy to gauge the health of Cytoscape from their GitHub repositories? What about their Release Notes and their Roadmap?

Reading

CORRIGAN
Ch 4 "Inscribing the Past"
Ch 7 "Spatializing and Analyzing Digital Texts"

Tom Elliot and Sean Gillies "Digital Geography and the Classics" *Digital Humanities Quarterly*

Schmidt, Benjamin M. "Theory First." *Journal of Digital Humanities*, March 9, 2012.

Gibbs, Fred. "Critical Discourse in Digital Humanities." *Journal of Digital Humanities*, March 9, 2012.

Where is Cultural Criticism in the Digital Humanities Alan Liu

Technology Is Taking over English Departments: The False Promise of the Digital Humanities Adam Kirsch

Difficult Thinking About the Digital Humanities Mark Sample

Further Reading

From Hermeneutics to Data to Networks: Data Extraction and Network Visualization of Historical Sources - SNA tutorial using Palladio

"Digital History" Can Never Be New Scott Weingart

Isn't it Obvious? Lincoln Mullen

Humane Computation Stephen Ramsay in *Debates in Digital Humanities*

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for Literary History*. Verso, 2007, p34-64 (Ch 3: Maps) Library ebook

Cherven, Ken. *Mastering Gephi Network Visualization*. Packt Publishing Ltd, 2015, selections (not in library)

Posts by Amanda Visconti on Using Gephi for Network Visualization

Reference Reading

Bunsdon, Chris and Lex Comber *An Introduction to R for Spatial Analysis and Mapping*. Second edition. Thousand Oaks, CA: SAGE Publications Ltd, 2018.

Hanneman, Robert A., Augustine J. Kposowa, and Mark D. Riddle. *Basic Statistics for Social Research*. 1 edition. San Francisco, CA: John Wiley & Sons, 2012.

Week 10 and Week 11

There are no formal meetings during these two weeks, but I'll be available at the regular class time for an open office hour. This is to allow you more time to focus in on the approaches from the previous weeks of most interest to you. Make use now of the further reading, examples of scholarship, and "strike out on your own" section to further develop your understanding and proficiency in the direction you want to go. I cannot stress enough how important experimentation, fruitful failures, and making good use of digital research skills in helping you to deepen your skills.

Attributions

This handbook takes ideas for readings and tasks from many of the works listed above in the readings and further readings. It also used material found in the blogs, syllabi and modules of other scholars out there including:

- the QGIS videos of Eric Brelsford
- the QGIS videos Lex Berman and the Harvard CGA
- many links and ideas from the members of the Spatial Community and the Digital Humanities Slack groups
- the various courses taught by David Mimno
- Courses and pages of David J. Birnbaum
- Seminar by Matt Wilkens
- Courses and links of Mark Sample
- Courses and blog entries of Lincoln Mullen
- Courses and writing of Ryan Cordell
- Writings of William Turkel
- Tutorials of Programming Historian
- Various postings by authors on ProfHacker
- Useful ideas and suggestions from Antonis Hadjikyriacou
- This is far from a complete list…