Clio 2: Computational History (Spring 2019)

This syllabus comes from https://lincolnmullen.com/courses/clio2.2019/. Only the online version of this syllabus is authoritative, and it may be updated as necessary.


Instructor: Lincoln Mullen (https://lincolnmullen.com) <lmullen@gmu.edu (mailto:lmullen@gmu.edu)>. Office: Research Hall 457. Office hours: Mondays, 11:00-12:00pm; Wednesdays, 10:30-11:45am.

Course description

In this course you will learn to use computational methods to create historical arguments. You will work with historical data, including finding, gathering, manipulating, analyzing, visualizing, and arguing from data, with special attention to geospatial, textual, and network data. These methods will be taught primarily through scripting in the R programming language (https://www.r-project.org/). While historical methods can be applied to many topics and time periods, they cannot be understood separate from how the discipline forms meaningful questions and interpretations, nor divorced from the particularities of the sources and histories of some specific topic. You will therefore work through a series of example problems using datasets from the history of the nineteenth-century United States, and you will apply these methods to a dataset in your own field of research.

Additional emphasis will be placed on publishing your scholarship on the web. While this is not a course in web development, you will learn the basics of publishing documents on the web, including familiarity with command line programs, getting files onto servers, basic web technologies such as HTML and CSS, Git and GitHub, and packages such as RMarkdown for publishing data analysis.

In other words, this class will teach you how to have something historically meaningful to say from data, and how to publish what you want to say on the web.

Learning goals

After taking this course, you will be able to

- perform exploratory data analysis; clean, tidy, and manipulate data; gather historical data from print and manuscript sources; use existing historical data sets; create common visualizations; work with geospatial, textual, and network data.
- write scripts using the R programming language and its extensive set of packages.
- understand the place of data analysis and visualization within humanities computing, digital history, and the discipline of history.
- conceive of and execute a short research project in computational history.
- publish your scholarship on the web.

Essential information

This is a graduate methods course in a field that moves reasonably quickly. The syllabus is likely to change over the course of the semester. In particular, I am likely to send you additional projects or visualizations to look at before class.
You are always welcome to talk with me during office hours. While you can drop in, I strongly encourage you to book an appointment ([office-hours](mailto:lmullen@gmu.edu)). If the scheduled times don’t work for you, email me ([lmullen@gmu.edu](mailto:lmullen@gmu.edu)) and suggest a few other times that would work for you.

All communication for this course will happen in our [Slack group](https://mason-dh-grad.slack.com/signup). Read this [getting started guide](https://get.slack.help/hc/en-us/articles/218080037-Getting-started-for-new-users) if you need help. This is your primary place to ask for help.

Bring a computer to each class meeting. See the list of software that you will need to install under the heading [for the first week](https://docs.microsoft.com/en-us/windows/wsl/about). This class is going to assume that you a computer with some kind of Unix-like operating system available. The easiest will be macOS or a Linux distribution. But if you use Windows, good news, you can install the [Windows Subsystem for Linux](https://docs.microsoft.com/en-us/windows/wsl/about), though after that you are mostly on your own to figure out the peculiarities of Windows.

You will need a basic web server and your own domain. If you do not already have a domain and web hosting, the personal shared hosting from [Reclaim Hosting](https://reclaimhosting.com/shared-hosting/) will be more than adequate.

One textbook is required in print (though it is partially available online).


All the other required readings are available online or through the GMU libraries, though they can also be purchased—sometimes in more complete editions—in print or e-books. These are the books we will use most frequently.


In general I have provided datasets and questions for you to work on for all the assignments except the final paper. But for any assignment, you may substitute a dataset from your own historical interests after checking with me. The forward-thinking graduate student will try to find such datasets early on in the semester so that you can use the intermediate assignments as preparation for your final assignment. If you can peer even farther into the future, you can try to use the final assignment as a test run of work you might want to do in one of your own research projects, such as an article or a dissertation.

**Assignments**

Assignments should be submitted via this form ([https://goo.gl/forms/J0a0sYVk4AANryBq2](https://goo.gl/forms/J0a0sYVk4AANryBq2)) unless otherwise instructed. For each assignment you will be expected to turn in two things. The first will be a web page with a public-facing presentation of your work. The other will be a GitHub repository with the source code that creates that web page.

Preparation and participation are expected as a matter of course in a graduate class. Complete all readings and submit all assignments before class. If the readings include sample code or questions at the end, work through them as part of doing the readings. Final grades will be calculated using the typical percentage-based grading scale (A = 93–100, A- = 90–92, B+ = 88–89, B = 83–87, B- = 80–82, … F = 0–59).
Worksheets and weekly assignments (20%). Many classes will have an assignment due before class begins. Some will require you to do library research; others will be practice data analysis worksheets. Some of the questions on the worksheets will be easy; most will be difficult; some you may find nearly impossible. The aim is to practice. We will go over the worksheets in class each week. If you attempt a problem and can’t solve it, you should still turn in whatever work you did on it. Students who complete all the easy and moderately difficult questions, attempt the very difficult questions, and ask for help as needed will do just fine. These assignments will graded by completion.

Analysis assignments (3 × 15% = 45%). You will do three analysis assignments, each demonstrating a specific skill in data analysis. For these assignments you will be given a historical dataset and asked some interpretative questions. You will prepare an RMarkdown document (http://rmarkdown.rstudio.com/) containing prose, code, and tables or visualizations to answer the historical questions and, as necessary, explain your methods. You will be given a starter GitHub repository that you can fork with the data and questions.

Research paper (35%). You will write a research paper suitable for a presentation at a disciplinary or digital humanities conference. This paper should advance a historical argumentation on the basis computational historical methods, though you can and should use more traditional historical methods as necessary. The body of the paper should be about 2,000 words in length. It should include notes in Chicago format like any other work of history. The paper should include embedded visualizations or tables as appropriate. Each table and figure must have a caption written in complete sentences. The paper should be attractively presented on your website using the Radix RMarkdown format (https://rstudio.github.io/radix/). Explain your methods as needed, but write in a way which would be understandable and compelling to any historian working in your field. The paper should be accompanied by a GitHub repository containing your data and code in a reproducible analysis. Ideally this paper could be presented at a conference, and it could serve as a trial for computational work you might do in a larger research project. As a model, see the most recent CFP (http://crdh.rrchnm.org/cfp/) for Current Research in Digital History. Due Monday, May 13 at 5pm.

Schedule

Week 1 (January 28): The web

Do your level best to get these set up before the first day of class:

- Get your own domain and web hosting, at Reclaim Hosting (https://reclaimhosting.com/shared-hosting/) if you don’t have it already.
- Install R (https://cran.rstudio.com/), a programming language for data analysis.
- Install RStudio (https://www.rstudio.com/products/rstudio/#Desktop), an environment for using R.
- Get a GitHub (https://github.com/) account and post it to the Slack group. (E.g., I am lmullen and this is my GitHub profile (https://github.com/lmullen).)
- Install Git (https://help.github.com/articles/set-up-git/) (more detailed guide (https://happygitwithr.com/install-git.html)).
- Install Homebrew (https://brew.sh) (only if you use macOS).

Read:

Arguing with Digital History working group, "Digital History and Argument," white paper, Roy Rosenzweig Center for History and New Media (November 13, 2017).

**Week 2 (February 4): Data from history and historians**

Assignment:

- Find at least three primary source data tables, datasets, or corpora from your field of historical research. Post full citations and URLs in the Slack group, along with a sentence or two explaining what you’ve found. Examine the links that other people post before class.

Read:


Browse:


**Week 3 (February 11): Basics of R**

Assignment:

- If you found a primary source dataset last week that is worth transcribing, then you can transcribe it. Otherwise, transcribe some of the *Minutes* (https://catalog.hathitrust.org/Record/006771726) of the Methodist Episcopal Church from after 1851. Whichever source you use, transcribe at least 25 rows of the data into a spreadsheet. Be prepared to describe in class how you decided on the structure of your
data, and how you identified what the variables were. (Hint: read the Broman and Woo article before you start this assignment.)

Read:

- Wickham and Grolemund, *R for Data Science*, ch. 1, 4, 6, 8.

Browse:

- Tutorial for RMarkdown (https://rmarkdown.rstudio.com/).

**Week 4 (February 18): Data manipulation**

Assignment:


Read:


**Week 5 (February 25): Data visualization**

Assignment:

- First data manipulation worksheet (https://dh-r.lincolnmullen.com/worksheets.html).

Read:

- Wickham and Grolemund, *R for Data Science*, ch. 3.
- Healy, *Data Visualization*, ch. 1, 3.

**Week 6 (March 4): More data manipulation and visualizations**

Assignment:


Read:

- Healy, *Data Visualization*, ch 4-5, 8.


Spring break (March 11)

Week 7 (March 18): Exploratory data analysis

Assignment:

- Second data manipulation worksheet ([https://dh-r.lincolnmullen.com/worksheets.html](https://dh-r.lincolnmullen.com/worksheets.html)).

Read:

- Wickham and Grolemund, R for Data Science, ch. 7, 19-21.
- Skim: Wickham and Grolemund, R for Data Science, ch. 27, 29.3-29.4, 30.

Week 8 (March 25): Mapping

Assignment:

- Exploratory data analysis assignment due ([https://github.com/ClioGMU/clio2-eda](https://github.com/ClioGMU/clio2-eda))

Read:

- Healy, Data Visualization, ch 7.

Browse:


Week 9 (April 1): Mapping

Read:


Browse:

- Ben Schmidt maps and visualizations gallery (http://benschmidt.org/maps-visualizations-gallery/)

Week 10 (April 8): Text analysis

Assignment:

- Mapping assignment due (https://github.com/ClioGMU/clio2-mapping)

Read:

- Graham, Milligan, Weingart, Macroscope, chs. 3-4.

Browse:

- quanteda documentation (https://quanteda.io/)
- cleanNLP documentation (https://statsmaths.github.io/cleanNLP/)
- Wickham and Grolemund, R for Data Science, 14.

Week 11 (April 15): Text analysis

Read:

- Robert K. Nelson, Mining the Dispatch (http://dsl.richmond.edu/dispatch/).
Week 12 (April 22): Network analysis

Assignment:
- Text analysis assignment due (https://github.com/ClioGMU/clio2-text)

Read:
- Graham, Milligan, Weingart, Macroscope, ch. 6-7.

Week 13 (April 29): Supervised classification

Read:
- Gareth James, et al., An Introduction to Statistical Learning: With Applications in R (Springer, 2013), ch. 1, 2, 4. GMU library (https://link-springer-com.mutex.gmu.edu/chapter/10.1007/978-1-4614-7138-7_2)

Week 14 (May 6): TBD

Topic and readings to be determined by the needs of student research papers.

Read:

Possible topic:
- Distant Viewing (https://www.distantviewing.org)

Fine print

This syllabus may be updated online as necessary. The online version of this syllabus is the only authoritative version.
Students must satisfactorily complete all assignments (including participation assignments) in order to pass this course. Your attendance is expected at every meeting. If you must be absent, I request that you notify me in advance of the class meeting. I am sometimes willing to grant extensions on assignments for cause, but you must request an extension before the assignment’s due date. For every day or part of a day that an assignment is late without an extension, I may reduce your grade. No work (other than final projects) will be accepted after the last day that the class meets. I will discuss grades only in person during office hours.

See the George Mason University catalog (http://catalog.gmu.edu/) for general policies, as well as the university statement on diversity (http://ctfe.gmu.edu/professional-development/mason-diversity-statement/). You are expected to know and follow George Mason's policies on academic integrity (http://oai.gmu.edu/) and the honor code (http://oai.gmu.edu/understanding-the-honor-code/). If you are a student with a disability and you need academic accommodations, please see me and contact the Office of Disability Services at 703-993-2474 or through their website (http://ods.gmu.edu). You are responsible for verifying your enrollment status. All academic accommodations must be arranged through that office. Please note the dates for dropping and adding courses from the GMU academic calendar (http://registrar.gmu.edu/calendars/spring-2019/).