

Data and Visualization in Digital History

Course Description **Assignments** **Schedule** **Fine Print**

George Mason Univ., Dept. of History and Art History.

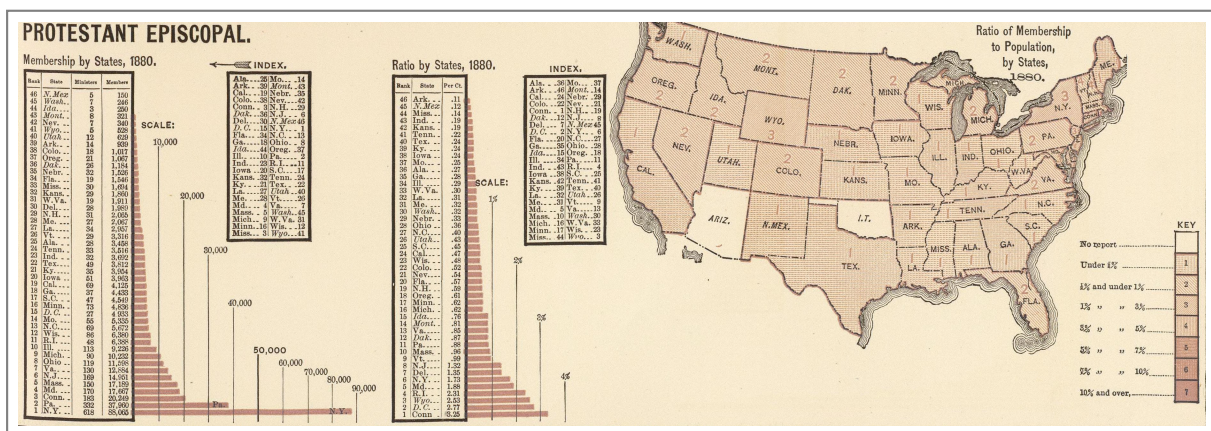
Spring 2016. [HIST 688-001](#).

Aquia Building 219. Wednesdays, 7:20 to 10:00 p.m.

[Lincoln Mullen](#), lmullen@gmu.edu, Research Hall 457.

Course Description

In this methods course you will be introduced to data manipulation and visualization for historians. You will learn to 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, using other command line tools as appropriate. 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. Therefore, in this course we will examine the historiographical tradition to see how historians have used data and visualization to understand the past. And we will work together to apply these methods to a linked series of datasets, some of which we will create ourselves, in the history of nineteenth-century American religion.



Detail of denominational statistics for the Protestant Episcopal Church in 1880 from Fletcher W. Hewes and Henry Gannet, *Scribner's Statistical Atlas of the United States Showing by Graphic Methods Their Present Condition and Their Political, Social and Industrial Development* (New York: Charles Scribner's Sons, 1883), [plate 59](#). Courtesy of the David Rumsey Map Collection.

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 provided by government or other research groups; create common visualizations; work with geospatial, textual, and network data; understand the basics of some machine learning techniques.
- write scripts using the R programming language and its extensive set of packages, as well as use command line programs.
- understand the place of data analysis and visualization within humanities computing, digital history, and the discipline of history.
- conceive of and execute a historical research project suitable for treatment in a dissertation chapter or journal article.
- take the course "Programming in History/New Media," a.k.a. [Clio 3](#), should you choose.

Books

You will need to secure copies of these books. All other readings are available on [the schedule](#).

- Taylor Arnold and Lauren Tilton, *Humanities Data in R* (Springer, 2015). [Free from Mason libraries](#).

- Shawn Graham, Ian Milligan, Scott Weingart, [*Exploring Big Historical Data: The Historian's Macroscopic*](#) (Imperial College Press, 2015).
- Daniel Kaplan, [*Data Computing: An Introduction to Wrangling and Visualization with R*](#) (Project Mosaic, 2015).
- Isabel Meirelles, [*Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*](#) (Rockport Publishers, 2013).
- Susan Schulten, [*Mapping the Nation: History and Cartography in Nineteenth-Century America*](#) (University of Chicago Press, 2012).

I may also post additional material to my [book in progress](#) on the subject of the course.

Software

We will use this software during the course. Plan on bringing a computer to each class meeting.

- An [RStudio Server instance](#) (kindly provided by RRCHNM) will give you access to R in your browser. You will be given a user name and password. You should, however, consider installing [R](#) and [RStudio Desktop](#) on your own computer.
- That same server will give you access to a Linux (CentOS, to be precise) command line which you can access through SSH. You will receive instruction on how to do this in class. Mac and Linux computers come with SSH installed; on Windows you will need to install [PuTTY](#).
- Much of your work for the course will go on [GitHub](#). Sign up for an account.
- [Slack](#) will replace e-mail and Blackboard for our course. You will receive an invitation to the [Mason Data and DH](#) team. You may wish to install [one of the apps](#). Here is an [introduction to Slack](#) from one of Kris Shaffer's courses.
- Get a real text editor and CSV reader. On Mac [Sublime Text](#) or [Atom](#) and on Windows [Sublime Text](#) or [Notepad++](#) seem to be popular choices. The one true text editor is [Vim](#), but it has a steep learning curve. Download [LibreOffice](#) because it does better than Excel at not messing with CSV files.

← [Back to home page](#). ⓘ [Licensed CC-BY](#).

Data and Visualization in Digital History

[Course Description](#) [Assignments](#) [Schedule](#) [Fine Print](#)

Assignments

N.B. In this course we will be working on U.S. religion in the long nineteenth-century. **You are more than welcome to work on your own research project.** Ideally your work in this course will advance your dissertation or other research project. Talk to me early in the course and we will figure out together what would work best for you.

Familiarize yourself with R (5%)

By the beginning of week 4 you must complete the [Try R](#) exercises at Code School. Students who want more of a challenge or who have come to class with some background in programming or data analysis are welcome to substitute another source of exercises. I recommend the first three chapters in Arnold and Tilton, *Humanities Data in R*, along with the exercises in the appendix. But the introductory chapters in Jockers, *Text Analysis with R for Students of Literature*, would also be a good choice if you prefer to work with textual data. Kaplan, *Data Computing*, will give you a different perspective focused on the “Hadleyverse” of R packages.

Weekly worksheets or assignments (50%)

After class each week you will be given an assignment. In the first several weeks of the course, these will be historical research assignments which might even get you inside of the library. The rest of the semester they will be data analysis worksheets. These are intended to practice what we have gone over in class to cement the ideas. They will also serve as a helpful reference sheet when you need to remember how to perform some kind of analysis. Some of the questions on the worksheet will be easy; most will be difficult; some you may

find nearly impossible. The aim is to *practice*, building on whatever level of skill you bring to the course. We will go over the worksheets in class the next week. Unless otherwise specified, you may work with up to three other people to solve the problems posed, and you are always welcome to ask for help in the Slack channel. But each student must turn in his or her own worksheet. If you attempt a problem and can't solve it, you should still turn in whatever work you did on it.

Here is how I will grade these worksheets. There will be no maximum level of points. Instead, I will give you points based on the truthfulness and skillfulness of your work. For instance, a correct answer to an obvious problem might garner you one point. A correct but sloppy answer to a more difficult problem might garner you two or three points, where a correct and elegant solution might garner you five or six. An incomplete answer to a very difficult problem which nevertheless reveals deep thought or historical insight might get you eight or nine points. I will total up your scores as we go, and grade them on a (very generous) curve at the end of the semester. Students who complete all the easy and moderate difficulty questions, attempt the very difficult questions, and ask me for help as needed will do just fine.

Visualization essays (45%)

You will be graded on three brief visualization essays. You must submit three, but you may submit as many as five, of which I will count the three best. This assignment is deliberately open ended. At a minimum these visualization should include the following: (1) explanatory text written for a scholarly but non-technical audience that frames the visualizations in both historical and historiographical terms and argues some worthwhile point; (2) as many well-crafted visualizations or data tables as are necessary to argue that point; (3) access to the code (though not necessarily in the essay itself) so that I can see how you got your results; and (4) citations. The most obvious way to accomplish this is in an knitr report (which we will discuss in class). But you may have better ideas about what you want to accomplish, and I will provide the direction and support you need to accomplish them. Students who wish to attempt some more ambitious project (which must still meet those requirements) are welcome to talk with me about combining two or three of the visualization essays into a single format. I will provide more formal rubrics for evaluation once we have made it further into the course.

Participation

Regular, informed participation is expected as a matter of course in a graduate class.

← [Back to home page.](#) ⓘ [Licensed CC-BY.](#)

Data and Visualization in Digital History

[Course Description](#) [Assignments](#) [Schedule](#) [Fine Print](#)

Schedule

N.B. Since this is a seminar where we will be working on a research topic together, I will probably modify and adapt the readings and assignments as the semester proceeds. Readings are to be completed prior to class. Assignments are to be completed before the start of the next class.

Week 1 (Jan. 20): The problem stated

Read:

- Robert Baird, *Religion in America; Or, An Account of the Origin, Relation to the State, and Present Condition of the Evangelical Churches in the United States. With Notices of the Unevangelical Denominations* (New York: Harper & Brothers, 1856).
- Fletcher W. Hewes and Henry Garnett, *Scribner's Statistical Atlas of the United States Showing by Graphic Methods Their Present Condition and Their Political, Social and Industrial Development* (New York: Charles Scribner's Sons, 1883), plates 58 to 61.
- Herman Carl Weber, *Presbyterian Statistics through One Hundred Years, 1826-1926* (Philadelphia: Presbyterian Church in the U.S.A., 1927), 438–539, 540–587.

Assignment:

- Find visualizations, data tables, data sets, or corpora about religion from the nineteenth century U.S. 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 in the class post.

Week 2 (Jan. 27): Data visualization from historical actors

Read:

- Schulten, *Mapping the Nation*
- Shari Rabin, "['Let Us Endeavor to Count Them Up:' Statistics and American Judaism in the Nineteenth Century](#)" (presented at the Association for Jewish Studies, Boston, MA, 2013).
- Rodney Stark, "The Reliability of Historical United States Census Data on Religion," *Sociology of Religion* 53, no. 1 (1992): 91–95, doi:10.2307/3711631.
- "[Colleges, Universities and Churches](#)" in Robert K. Nelson, Scott Nesbit, et al., *Atlas of the Historical Geography of the United States* (Digital Scholarship Lab).

Assignment:

- Find at least one instance of how historical U.S. religious data has been used by historians or sociologists. (Atlases are a sure bet, but be creative.) Post a scan or a photo of at least one visualization or data table and write a one paragraph critique.

Week 3 (Feb. 3): Data in social and digital history

Read:

- Graham, Milligan, Weingart, *Macroscope*, ch. 1–2.
- Roger Finke and Rodney Stark, [The Churching of America, 1776-2005: Winners and Losers in Our Religious Economy](#) (Rutgers University Press, 2005), at least chapters 1, 3, 5.
- Susan Hockey, "The History of Humanities Computing" in [A Companion to Digital Humanities](#), ed. Susan Schreibman, Ray Siemens, John Unsworth (Oxford: Blackwell, 2004).
- William G. Thomas II, "Computing and the Historical Imagination," in [A Companion to Digital Humanities](#), ed. Susan Schreibman, Ray Siemens, John Unsworth (Oxford: Blackwell, 2004).
- Laurie F. Maffly-Kipp, "If It's South Dakota You Must Be Episcopalian: Lies, Truth-Telling, and the Mapping of U.S. Religion." *Church History* 71, no. 1 (2002): 132–42.

- Laurie F. Maffly-Kipp, "Putting Religion on the Map." *The Journal of American History* 94, no. 2 (September 1, 2007): 522–29. doi:10.2307/25094964.
- Kieran Healy and James Moody, "[Data Visualization in Sociology](#)" *Annual Review of Sociology*, 40:105–128.

Assignment:

- Complete the [Try R](#) exercises at Code School.
- Worksheet

Week 4 (Feb. 10): Crash course in R

Read:

- Kaplan, *Data Computing*, chs. 1–4. If you want more of a challenge, try Arnold and Tilton, chs. 1–3 .
- Meirelles, *Design for Information*, ch. 1.

For reference:

- Hadley Wickham, [Advanced R](#) (Chapman & Hall, 2014).

Assignment:

- Worksheet

Week 5 (Feb. 17): Visualization with the grammar of graphics

Read:

- Kaplan, *Data Computing*, chs. 5, 6, 8.
- Meirelles, *Design for Information*, ch. 3.
- Graham, Milligan, Weingart, *Macroscopic*, ch 5.
- Johanna Drucker, "[Humanities Approaches to Graphical Display](#)" *Digital Humanities Quarterly* 5, no. 1 (2011).
- Fusion Charts, [Principles of Data Visualization](#). But see the data vizualization papers below.

For reference:

- Joanna Zhao and Jennifer Bryan, [R Graph Catalog](#)
- [ggplot2 documentation](#)
- Winston Chang, [companion website](#) to *R Graphics Cookbook*
- Hadley Wickham, [ggplot2: Elegant Graphics for Data Analysis](#) (Springer, 2009),

Data visualization research for reference:

- William S. Cleveland and Robert McGill. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79, no. 387 (September 1, 1984): 531–54. doi:10.1080/01621459.1984.10478080.
- C. G. Healey, K. S. Booth, and J. T. Enns, "[High-Speed Visual Estimation Using Preattentive Processing](#)," *ACM Transactions on Computer-Human Interaction* 3, no. 2 (1996): 108–35.
- Ben Shneiderman, "[The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations](#)," *Proc. Visual Languages* 96 (1996): 1–8.

Assignment:

- Worksheet

Week 6 (Feb. 24): Data manipulation

- Kaplan, *Data Computing*, chs. 1, 7, 9.
- Hadley Wickham, "[Tidy Data](#)" *The Journal of Statistical Software* 59, no. 10 (2014): 1–23.
- Hadley Wickham, "[The Split-Apply-Combine Strategy for Data Analysis](#)," *Journal of Statistical Software* 40, no. 1 (2011): 1–29.
- David Mimno, "[Data Carpentry](#)," August 19, 2014.

For reference:

- [The Quartz guide to bad data](#)

Assignment:

- Worksheet

- Begin transcribing historical data, chosen in consultation with me. Transcription should be substantially begun by next meeting, and substantially completed by the end of spring break.

Week 7 (Mar. 2): Data manipulation continued; gathering historical data

Read:

- Kaplan, *Data Computing* 10–12.

Assignment:

- Worksheet
- Complete transcription of historical data.

Spring break (Mar. 9)

Week 8 (Mar. 16): Exploratory data analysis and statistics

Read:

- Kaplan, *Data Computing*, ch. 14.
- Arnold and Tilton, *Humanities Data in R*, chs. 3–5.
- Peter Dalgaard, *Introductory Statistics with R* (Springer, 2008), chs. 4, 6, 3.

For reference:

- Roger D. Peng, *Exploratory Data Analysis with R* (Leanpub, 2015).
- Gareth James, *An Introduction to Statistical Learning with Applications in R* (Springer, 2013).

Assignment:

- Worksheet

Week 9 (Mar. 23): Mapping

Read:

- Meirelles, *Design for Information*, ch. 4.
- Arnold and Tilton, *Humanities Data in R*, ch. 7.
- Cameron Blevins, "Space, Nation, and the Triumph of Region: A View of the World from Houston," *Journal of American History* 101, no. 1 (June 1, 2014): 122–47, doi:10.1093/jahist/jau184.
- Richard White "[What is Spatial History?](#)"

For reference:

- [ggplot2 documentation](#)
- [Spatial Humanities Workshop](#)
- [leaflet documentation](#)
- Roger S. Bivand, Edzer Pebesma, and Virgilio Gómez-Rubio, [Applied Spatial Data Analysis with R](#) (Springer, 2013).

Assignment:

- Worksheet
- Find and, if necessary, transcribe a data set with a spatial component that has not been geocoded.

Week 10 (Mar. 30): Geocoding and georeferencing

Read:

- Meirelles, *Design for Information*, ch. 5.
- Todd Samuel Presner, David Shepard, and Yoh Kawano, *HyperCities: Thick Mapping in the Digital Humanities*, MetaLABprojects (Harvard University Press, 2014), 110–27.

Assignment:

- Worksheet

Week 11 (Apr. 6): Textual data

Read:

- Kaplan, *Data Computing*, ch. 16.
- Arnold and Tilton, *Humanities Data in R*, chs. 9 and 10.
- Meirelles, *Design for Information*, ch. 6.
- Graham, Milligan, Weingart, *Macroscopic*, chs. 3–4.

For reference:

- Matthew Jockers, [*Text Analysis with R for Students of Literature*](#) (Springer, 2014). You may also wish to consult Matthew Jockers, [*Macroanalysis*](#) (University of Illinois Press, 2013).

Assignment:

- Worksheet

Week 12 (Apr. 13): Textual data continued

Read:

- Jean-Baptiste Michel et al., "[Quantitative Analysis of Culture Using Millions of Digitized Books](#)," *Science* 331, no. 6014 (January 14, 2011): 176–82, doi:10.1126/science.1199644.
- Michael Whitmore, "[Text: A Massively Addressable Object](#)," December 31, 2010.

Assignment:

- Worksheet

Week 13 (Apr. 20): Network data

Read:

- Meirelles, *Design for Information*, ch. 2.
- Arnold and Tilton, *Humanities Data in R*, ch. 6.
- Graham, Milligan, Weingart, *Macroscopic*, chs. 6–7.

For reference:

- Eric D. Kolaczyk and Gábor Csárdi, [Statistical Analysis of Network Data with R](#)

(Springer, 2014).

Assignment:

- Worksheet

Week 14 (Apr. 27): Final projects

TBD

← [Back to home page.](#) ⓘ [Licensed CC-BY.](#)

Data and Visualization in Digital History

[Course Description](#) [Assignments](#) [Schedule](#) [Fine Print](#)

Fine Print

General policies

See the [George Mason University catalog](#) for general policies, as well as the university [statement on diversity](#). You are expected to know and follow George Mason's policies on [academic integrity](#) and the [honor code](#). Ask me if you have any questions about how these policies apply to this course. Note the dates for dropping and adding this course from the [academic calendar](#).

Accommodations

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 <http://ods.gmu.edu>. All academic accommodations must be arranged through that office.

Assignments

Do all the readings, work through all exercises, complete any assignments, and install any necessary software before the start of each class.

I may change due dates or assignments. I will always give you plenty of notice of changes, which will always be intended for your benefit. Students must satisfactorily complete all assignments (including participation assignments) in order to pass this course.

Attendance

Your attendance is expected at every meeting. If you must be absent, I request that you notify me in advance of the class meeting.

Communication

I am always glad to make an appointment to meet with you in person. I may send official course communications to your George Mason e-mail address, which you should check regularly. I will discuss grades only in person.

The online syllabus is the only authoritative version. Check it regularly.

Privacy

You may be required to make some of your work publicly available. If you wish, you may do so under a pseudonym to keep your identity private from the public, though I and the other students in the course will have access to your work.

Late work

I am sometimes willing to grant extensions 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 exams and final projects) will be accepted later than the last day that the class meets unless I have agreed otherwise.

License

This syllabus is copyrighted © 2016 Lincoln Mullen. This syllabus is licensed [CC-BY 4.0](#). You are free to use or modify this syllabus for any purpose, provided that you attribute it to the author, preferably at the course website.

Acknowledgments

I have drawn ideas or readings from the following syllabi:

- Jenny Bryan, [STAT 545: "Data wrangling, exploration, and analysis with R"](#)
- Shawn Graham, ["Crafting Digital History"](#)
- Kieran Healy, ["Data Visualization"](#)
- Ben Schmidt, ["Humanities Data Analysis"](#)

- Bill Turkel's [many digital history courses](#)
-

← [Back to home page.](#) ⓘ [Licensed CC-BY.](#)