

Computational History (Spring 2020)

This syllabus comes from <https://lincolnmullen.com/courses/data.2020/>. Only the online version of this syllabus is authoritative, and it may be updated as necessary.

Course: HIST 697-001 (https://historyarthistory.gmu.edu/courses/hist697/course%5c_sections/46187). Spring 2020.
Department of History and Art History, George Mason University. 3 credits. Meets Mondays, 7:20–10:00pm in Music Theater Building 1008.

Instructor: Lincoln Mullen (<https://lincolnmullen.com>) <lmullen@gmu.edu>. Office: Research Hall 484. Office hours: By appointment. [Book an appointment.](#) ([/page/office-hours/](#))

Course description

In this course you will learn to use computational methods to create historical interpretations. You will work with historical data, which includes finding, gathering, manipulating, analyzing, visualizing, and arguing from datasets, with special attention to geospatial, textual, and network data. These methods will be taught primarily using the [R programming language](https://www.r-project.org/) (<https://www.r-project.org/>). While data analysis 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 then apply the methods to write a research paper using a dataset from your own historical field.

Learning goals

After taking this course, you will be able to

- gather historical data from print and manuscript sources; use existing historical data sets; clean, tidy, and manipulate data; perform exploratory data analysis; 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 the field of digital history and the discipline of history.
- conceive of and execute a short research project in computational history.

Essential information

Most required readings are available online or through the GMU libraries. These are the main books that we will be using.

- Hadley Wickham and Garrett Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (<http://r4ds.had.co.nz/>) (O'Reilly, 2017). ISBN: 978-1491910399.
- Kieran Healy, *Data Visualization: A Practical Introduction* (<http://socviz.co/>) (Princeton University Press, 2018). ISBN: 978-0691181622.

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, which should be treated the same as other assigned readings.

All communication for this course will happen in [our Slack group](https://mason-dh-grad.slack.com/signup) (<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) (<https://get.slack.help/hc/en-us/articles/218080037-Getting-started-for-new-users>) if you need help. The Slack group is your primary place to ask for help. Please ask for help in the public channels rather than private

messages. You are almost certainly not the only person to have your question, and asking and answering questions publicly benefits everyone. When you ask a question, help me help you by including the code that you are asking about and any error messages that are relevant.

You are always welcome to talk with me during office hours. While you can drop in, I strongly encourage you to [book an appointment](#) ([/page/office-hours/](#)). If the scheduled times don't work for you, please contact me and suggest a few other times that would work for you.

Bring a computer to each class meeting. For the most part, we will be using an RStudio Server instance hosted by [RRCHNM](#) (<https://rrchnm.org>), which you can log in to using a web browser. But you should also install some key software on your computer. See the list under the heading for the first week. I will assume that you have 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: R has very good support for Windows, and 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.

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 field 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 could try to use the final assignment as a test run for work you might want to do in one of your own research projects, such as a conference presentation, article, or dissertation.

Assignments

Assignments should be [submitted via this form](#) (<https://forms.gle/A5hxmPgPBYJ9J1GMA>). For each assignment you will be expected to turn in the URL to a web page you have created using [RMarkdown](#) (<https://rmarkdown.rstudio.com>). You are responsible for finding some place to host the documents. A personal website, perhaps hosted at [Reclaim Hosting](#) (<https://reclaimhosting.com>) or [GitHub pages](#) (<https://pages.github.com>), will do.

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, though you do not need to submit them and I will not check them. 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 (25%). 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 be graded by completion.

Analysis assignments (4 × 10% = 40%). You will do four 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](#) (<http://rmarkdown.rstudio.com/>) document 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 interpretation using 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 [Distill](#) (<https://rstudio.github.io/distill/>) format for RMarkdown.

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 (<https://doi.org/10.1080/00031305.2017.1375986>) that I can re-run on my own computer. 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. For models, see the articles published in for Current Research in Digital History (<https://crdh.rchnm.org>). Due Monday, May 11 at 5pm.

Schedule

Week 1 (January 27): Introduction to computational history

Assignment:

- Find one example of a digital history project that uses visualization or data analysis. Be prepared with the URL and a three-minute answer to these questions: What is interesting or insightful about this project? What did this project do that you would like to learn how to do for your own research?

Readings:

- Frederick W. Gibbs, “New Forms of History: Critiquing Data and Its Representations (<http://tah.oah.org/february-2016/new-forms-of-history-critiquing-data-and-its-representations/>),” *The American Historian* (February 2016).
- Taylor Arnold and Lauren Tilton, “New Data: The Role of Statistics in DH (<https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/a2a6a192-f04a-4082-afaa-97c76a75b21c#ch24>),” in *Debates in DH 2019*, ed. Matthew K. Gold and Lauren F. Klein (University of Minnesota Press, 2019).
- “Introduction (<http://dh-r.lincolnmullen.com/introduction.html>),” in *Computational Historical Thinking*.

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

- Join the class Slack group (<https://mason-dh-grad.slack.com/signup>).
- 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 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.
- Install Homebrew (<https://brew.sh>) (only if you use macOS).

These are mostly optional, but it would be helpful to have them:

- Install Visual Studio Code (<https://code.visualstudio.com>), a general-purpose text editor for developers.
- Install Cyberduck (<https://cyberduck.io>), an FTP client.
- Install Git (<https://help.github.com/articles/set-up-git/>) (more detailed guide (<https://happygitwithr.com/install-git.html>)).

Week 2 (February 3): Data from history and historians

Assignment:

- Find at least three primary source data tables, datasets, or corpora from your field of historical research. These could include sources that are in print or manuscript, as well as datasets that have already been created. 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.

Readings:

- Roger Finke and Rodney Stark, *The Churching of America, 1776-2005: Winners and Losers in Our Religious Economy* (<http://www.jstor.org/stable/j.ctt5hhwnc>) (Rutgers University Press, 2005), ch. 1.
- Chad Gaffield, “Words, Words, Words: How the Digital Humanities Are Integrating Diverse Research Fields to Study People,” *Annual Review of Statistics and Its Application* 5, no. 1 (2018): 119–39, <https://doi.org/10.1146/annurev-statistics-031017-100547> (<https://doi.org/10.1146/annurev-statistics-031017-100547>).
- Abraham Gibson and Cindy Ermus, “The History of Science and the Science of History: Computational Methods, Algorithms, and the Future of the Field,” *Isis* 110, no. 3 (2019): 555–66, <https://doi.org/10.1086/705543> (<https://doi.org/10.1086/705543>).
- Jessica Marie Johnson, “Markup Bodies: Black [Life] Studies and Slavery [Death] Studies at the Digital Crossroads,” *Social Text* 36, no. 4 (2018): 57–79, <https://doi.org/10.1215/01642472-7145658> (<https://doi.org/10.1215/01642472-7145658>).
- Laurie F. Maffly-Kipp, “*If It’s South Dakota You Must Be Episcopalian: Lies, Truth-Telling, and the Mapping of U.S. Religion* (<http://www.jstor.org/stable/4146694>)” *Church History* 71, no. 1 (2002): 132–42.
- Shari Rabin, “‘Let us Endeavor to Count Them Up’: The Nineteenth-Century Origins of American Jewish Demography,” *American Jewish History* 101, no. 4 (2017): 419–440, <https://doi.org/10.1353/ajh.2017.0060> (<https://doi.org/10.1353/ajh.2017.0060>).

Browse:

- Robert K. Nelson et al., *Atlas of the Historical Geography of the United States* (<https://dsl.richmond.edu/historicalatlas/>) (Digital Scholarship Lab, University of Richmond).
- Herman Carl Weber, *Presbyterian Statistics through One Hundred Years, 1826-1926* (<http://catalog.hathitrust.org/Record/007109885>) (Philadelphia: Presbyterian Church in the U.S.A., 1927), part II.
- Jasmine Weber, “How W.E.B. Du Bois Meticulously Visualized 20th-Century Black America” (<https://hyperallergic.com/476334/how-w-e-b-du-bois-meticulously-visualized-20th-century-black-america/>), 5 February 2019.

Week 3 (February 10): Basics of R

Assignment:

- [Getting familiar with R worksheet](https://dh-r.lincolnmullen.com/worksheets.html) (<https://dh-r.lincolnmullen.com/worksheets.html>).
- Either use a primary source dataset that you found last week or, as a backup, the *Minutes* (<http://catalog.hathitrust.org/Record/006771726>) of the Methodist Episcopal Church from after 1851. Create a well-structured spreadsheet and transcribe at least 25 rows of the data. Upload a CSV file to Slack before class. Be prepared to describe in class how you decided on the structure of your data, and how you identified what the variables were. Use the Broman and Woo article as a guide.

Readings:

- Wickham and Grolemund, *R for Data Science*, ch. 1, 4, 6, 8, 27.
- Karl W. Broman and Kara H. Woo, “Data Organization in Spreadsheets,” *American Statistician* 72, no. 1 (2018): 2–10, <https://doi.org/10.1080/00031305.2017.1375989> (<https://doi.org/10.1080/00031305.2017.1375989>).
- “[Getting Started](https://dh-r.lincolnmullen.com/getting-started.html) (<https://dh-r.lincolnmullen.com/getting-started.html>)” and “[An R Primer](http://dh-r.lincolnmullen.com/primer.html) (<http://dh-r.lincolnmullen.com/primer.html>)” in *Computational Historical Thinking*.
- [RMarkdown documentation](https://rmarkdown.rstudio.com/) (<https://rmarkdown.rstudio.com/>).

Week 4 (February 17): Data manipulation

Assignment:

- [Data structures worksheet](https://dh-r.lincolnmullen.com/worksheets.html) (<https://dh-r.lincolnmullen.com/worksheets.html>).
- [Functions worksheet](https://dh-r.lincolnmullen.com/worksheets.html) (<https://dh-r.lincolnmullen.com/worksheets.html>).

Readings:

- Wickham and Grolemund, *R for Data Science*, ch. 5, 12, 13.
- Documentation for the [tidyverse](https://www.tidyverse.org) (<https://www.tidyverse.org>).
- Documentation for [databases in R](https://db.rstudio.com) (<https://db.rstudio.com>).

Week 5 (February 24): Data visualization

Assignment:

- [Data manipulation worksheet](https://dh-r.lincolnmullen.com/worksheets.html) (<https://dh-r.lincolnmullen.com/worksheets.html>).

Readings:

- Healy, *Data Visualization*, ch. 1, 3, 4.
- Wickham and Grolemund, *R for Data Science*, ch. 3, 28.
- Kieran Healy and James Moody, “[Data Visualization in Sociology](http://kieranhealy.org/files/papers/data-visualization.pdf)” (<http://kieranhealy.org/files/papers/data-visualization.pdf>) *Annual Review of Sociology*, 40:105–128.
- Lauren F. Klein, “The Image of Absence: Archival Silence, Data Visualization, and James Hemings,” *American Literature* 85, no. 4 (December 1, 2013): 661–88, <https://doi.org/10.1215/00029831-2367310> (<https://doi.org/10.1215/00029831-2367310>).
- John Theibault, “Visualizations and Historical Arguments,” in *Writing History in the Digital Age*, ed. Kristen Nawrotzki and Jack Dougherty (University of Michigan Press, 2013), <https://doi.org/10.3998/dh.12230987.0001.001> (<https://doi.org/10.3998/dh.12230987.0001.001>).

Week 6 (March 2): Exploratory data analysis

Assignment:

- [Data visualization worksheet](https://dh-r.lincolnmullen.com/worksheets.html) (<https://dh-r.lincolnmullen.com/worksheets.html>).

Readings:

- Wickham and Grolemund, *R for Data Science*, ch. 7, 17–21, 30.
- Healy, *Data Visualization*, ch. 5, 8.
- Roger Peng, *Exploratory Data Analysis with R* (<https://bookdown.org/rdpeng/exdata/>) (Leanpub, 2016), ch. 1, 4–6.

Spring break (March 9)

Week 7 (March 16): Maps

Assignment:

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

Readings:

- Healy, *Data Visualization*, ch. 7.
- Documentation for the [sf package](https://r-spatial.github.io/sf/) (<https://r-spatial.github.io/sf/>).
- Documentation for the [leaflet package](https://rstudio.github.io/leaflet/) (<https://rstudio.github.io/leaflet/>).
- Richard White, “[What is Spatial History?](http://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29)” (<http://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29>), *Spatial History Project* (Stanford University, 2010).
- Cameron Blevins, “Space, Nation, and the Triumph of Region: A View of the World from Houston,” *Journal of American History* 101, no. 1 (2014): 122–47, <https://doi.org/10.1093/jahist/jau184>

(<https://doi.org/10.1093/jahist/jau184>).

- Todd Presner and David Shepard, “Mapping the Geospatial Turn,” in *A New Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Wiley, 2016), 201–212. [GMU library](http://mutex.gmu.edu/login?url=http://www.gmu.eblib.com/EBLWeb/patron/?target=patron&extendedid=P_4093339_0) (http://mutex.gmu.edu/login?url=http://www.gmu.eblib.com/EBLWeb/patron/?target=patron&extendedid=P_4093339_0)

Browse:

- Robert K. Nelson and Edward L. Ayers, eds., *American Panorama: An Atlas of United States History* (<https://dsl.richmond.edu/panorama/>) (Digital Scholarship Lab, University of Richmond).
- [Gallery](http://web.stanford.edu/group/spatialhistory/cgi-bin/site/gallery.php) (<http://web.stanford.edu/group/spatialhistory/cgi-bin/site/gallery.php>) from the *Spatial History Project*, Stanford University.

Week 8 (March 23): Networks

Assignment:

- [Mapping assignment](https://github.com/ClioGMU/clio2-mapping) (<https://github.com/ClioGMU/clio2-mapping>).

Readings:

- Documentation for the *ggraph* package (<https://ggraph.data-imaginist.com>).
- Mark E. J. Newman, *Networks: An Introduction* (<http://math.sjtu.edu.cn/faculty/xiaodong/course/Networks%20An%20introduction.pdf>) (Oxford University Press, 2010), ch 1, 3, 4, 6, 7.
- Matthew Lincoln, “Social Network Centralization Dynamics in Print Production in the Low Countries, 1550–1750,” *International Journal for Digital Art History* 2 (2016): 134–157, <https://doi.org/10.11588/dah.2016.2.25337> (<https://doi.org/10.11588/dah.2016.2.25337>).
- “[AHR Forum: Mapping the Republic of Letters](https://academic-oup-com.mutex.gmu.edu/ahr/search-results?f_TocHeadingTitle=AHR%20Forum:%20Mapping%20the%20Republic%20of%20Letters) (https://academic-oup-com.mutex.gmu.edu/ahr/search-results?f_TocHeadingTitle=AHR%20Forum:%20Mapping%20the%20Republic%20of%20Letters),” *American Historical Review* 122, no. 2 (2017): 399–463.

Browse:

- Analysis repository for *civil procedure codes* (<https://github.com/lmullen/civil-procedure-codes>).

Week 9 (March 30): Texts

Assignment:

- [Network assignment](https://github.com/ClioGMU/clio2-networks) (<https://github.com/ClioGMU/clio2-networks>).

Readings:

- Kasper, Welbers, Wouter van Atteveldt, and Kenneth Benoit, “Text analysis in R,” *Communications Methods and Measures* 11, no. 4: 245–265, <https://doi.org/10.1080/19312458.2017.1387238> (<https://doi.org/10.1080/19312458.2017.1387238>).
- Taylor Arnold, Nicolas Ballier, Paula Lissón, and Lauren Tilton, “Beyond Lexical Frequencies: Using R for Text Analysis in the Digital Humanities,” *Language Resources and Evaluation* 53, no. 4 (2019): 707–733, <https://doi.org/10.1007/s10579-019-09456-6> (<https://doi.org/10.1007/s10579-019-09456-6>).
- Tim Hitchcock and William J. Turkel, “The *Old Bailey Proceedings, 1674–1913*: Text Mining for Evidence of Court Behavior,” *Law and History Review* 34, no. 4 (2016): 929–955, <https://doi.org/10.1017/S0738248016000304> (<https://doi.org/10.1017/S0738248016000304>).
- Joshua Catalano, “Digitally Analyzing the Uneven Ground: Language Borrowing Among Indian Treaties,” *Current*

Research in Digital History 1 (2018): <https://doi.org/10.31835/crdh.2018.02> (<https://doi.org/10.31835/crdh.2018.02>).

- Ryan Cordell, “Reprinting, Circulation, and the Network Author in Antebellum Newspapers,” *American Literary History* 27, no. 3 (2015): 417–445, <https://doi.org/10.1093/alh/ajv028> (<https://doi.org/10.1093/alh/ajv028>).
- David A. Smith, Ryan Cordell, and Abby Mullen, “Computational Methods for Uncovering Reprinted Texts in Antebellum Newspapers,” *American Literary History* 27, no. 3 (2015): E1–E15, <https://doi.org/10.1093/alh/ajv029> (<https://doi.org/10.1093/alh/ajv029>).
- Documentation for [quanteda package](https://quanteda.io/) (<https://quanteda.io/>).
- Documentation for [cleanNLP documentation](https://statsmaths.github.io/cleanNLP/) (<https://statsmaths.github.io/cleanNLP/>).
- Wickham and Grolemund, *R for Data Science*, 14.

Week 10 (April 6): Word embeddings

Assignment:

- Send me a one-page proposal for the final paper that you are going to write. You may organize the proposal however you like, but it should address the following questions: What historical questions are you asking? How will this paper advanced historical understanding in your field? What dataset and other sources will you be using? Which methods and visualizations will you use to answer your questions? What additional computational skills will you need to learn?

Readings:

- Ben Schmidt, “[Vector Space Models for the Digital Humanities](http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html) (<http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html>)” (October 25, 2015).
- Ben Schmidt, “[Rejecting the Gender Binary: A Vector-Space Operation](http://bookworm.benschmidt.org/posts/2015-10-30-rejecting-the-gender-binary.html) (<http://bookworm.benschmidt.org/posts/2015-10-30-rejecting-the-gender-binary.html>)” (October 30, 2015).
- Ryan Heuser, “[Word Vectors in the Eighteenth Century](http://ryanheuser.org/word-vectors/) (<http://ryanheuser.org/word-vectors/>). ”
- Matthew K. Gold and Lauren F. Klein et al., “[Forum: Text Analysis at Scale](http://dhdebates.gc.cuny.edu/debates/text/93) (<http://dhdebates.gc.cuny.edu/debates/text/93>),” in *Debates in the Digital Humanities 2016* (University of Minnesota Press, 2016), 525–568.
- Jo Guldi, “Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora,” *Journal of Cultural Analytics* (2018): <https://doi.org/10.22148/16.030> (<https://doi.org/10.22148/16.030>).

Week 11 (April 13): Clustering (unsupervised classification)

Assignment:

- [Text analysis assignment](https://github.com/ClioGMU/clio2-texts) (<https://github.com/ClioGMU/clio2-texts>).

Readings:

- Roger Peng, *Exploratory Data Analysis with R* (<https://bookdown.org/rdpeng/exdata/>) (Leanpub, 2016), ch. 12.
- Robert K. Nelson, *Mining the Dispatch* (<http://dsl.richmond.edu/dispatch/>) (Digital Scholarship Lab, University of Richmond).
- Benjamin Schmidt, “Stable Random Projection: Lightweight, General-Purpose Dimensionality Reduction for Digitized Libraries,” *Journal of Cultural Analytics* (2018): <https://doi.org/10.22148/16.025> (<https://doi.org/10.22148/16.025>).
- Skim Gareth James, et al., *An Introduction to Statistical Learning: With Applications in R* (Springer, 2013), ch. 10. [GMU library](https://link.springer.com.mutex.gmu.edu/chapter/10.1007/978-1-4614-7138-7_2) (https://link.springer.com.mutex.gmu.edu/chapter/10.1007/978-1-4614-7138-7_2)

Week 12 (April 20): Prediction (supervised classification)

Assignment:

- In the Slack group, post a one-paragraph summary of the work you’ve done since last week on your final paper. What

challenges are you facing, and what progress have you made? Feel free to post visualizations or other work in progress.

Readings:

- Wickham and Grolemund, *R for Data Science*, ch. 23–24.
- Matthew L. Jockers and Ted Underwood, “Text-Mining the Humanities” in *A New Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Wiley, 2016), 291–306. [GMU library](#) (http://mutex.gmu.edu/login?url=http://www.gmu.eblib.com/EBLWeb/patron/?target=patron&extendedid=P_4093339_0)
- Rebecca Sutton Koeser, “Trusting Others to ‘Do the Math’ (<https://doi.org/10.1080/03080188.2016.1165454>)” *Interdisciplinary Science Reviews* 40, no. 4 (2015): 376–392, <https://doi.org/10.1080/03080188.2016.1165454> (<https://doi.org/10.1080/03080188.2016.1165454>).
- Benjamin Schmidt, “[Do Digital Humanists Need to Understand Algorithms?](#) (<http://dhdebates.gc.cuny.edu/debates/text/99>)” in *Debates in the Digital Humanities 2016*, ed. Matthew K. Gold and Lauren F. Klein (University of Minnesota Press, 2016).
- Skim 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 13 (April 27): Interactive visualizations with Shiny

Assignment:

- In the Slack group, post a one-paragraph summary of the work you’ve done since last week on your final paper. What challenges are you facing, and what progress have you made? Feel free to post visualizations or other work in progress.

Readings:

- Documentation and gallery for [Shiny](#) (<https://shiny.rstudio.com>).

Week 14 (May 4): Final project workshop

Assignment:

- Circulate a draft of your final project in Slack by Friday, May 1. Be prepared to present your work in class for approximately ten minutes. Read each person’s draft and come prepared to offer helpful comments on their work.

Read:

- Ben Marwick, Carl Boettiger, and Lincoln Mullen, “Packaging Data Analytical Work Reproducibly Using R (and Friends),” *American Statistician* 72, no. 1 (2018): 80–88, <https://doi.org/10.1080/00031305.2017.1375986> (<https://doi.org/10.1080/00031305.2017.1375986>).

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/) (<http://catalog.gmu.edu/>) for general policies, as well as the university [statement on diversity](http://ctfe.gmu.edu/professional-development/mason-diversity-statement/) (<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/) (<http://oai.gmu.edu/>) and the [honor code](http://oai.gmu.edu/understanding-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) (<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-2020/) (<http://registrar.gmu.edu/calendars/spring-2020/>).

This syllabus draws ideas and assignments from many people and syllabi, including Taylor Arnold, Andrew Goldstone, Jason Heppler, Ben Schmidt, and Lauren Tilton.