Teaching Graduate Students to Code

Lincoln Mullen
Assistant Professor
History and Art History
George Mason University
lincoln@lincolnmullen.com

Sara Collini
PhD student, George Mason Univ.
scolli16@gmu.edu

Peter Carr Jones
PhD student, George Mason Univ.
pjones15@gmu.edu

Jannelle Legg
PhD student, George Mason Univ.
jannellelegg@gmail.com

Anne Ladyem McDivitt
PhD student, George Mason Univ.
amcdivit@gmu.edu

George D. Oberle III
PhD student, George Mason Univ.
goberle@gmu.edu

Amanda Regan
PhD student, George Mason Univ.
aregan2@gmu.edu

For a PDF of this poster, links to the syllabus, to a book titled *Digital History Methods in R* drafted during the course, and to the student projects, which include citations and open-source code:

http://lincolnmullen.com/projects/coding-poster/

A Description of the "Programming for Historians" Course

"Programming for Historians" (Clio 3) is an elective course for PhD students in History and Art History at George Mason University. The primary aim is to teach interested students to program in order to take advantage of digital history methods in their dissertations. Students in the course learned two languages, mainly the R language for data analysis but also JavaScript and the D3.js library for online visualizations.

Conclusion: Programming methods—including mapping, text mining, quantitative methods, and network analysis—proved applicable to a wide array of topics. Most of the students saw the course as a chance to pick up needed digital methods before the prospectus stage and to use them on sources intended for their dissertations. Though there is a large technical literature on programming and a growing literature on digital history theory and practice, only a few resources such as *The Programming Historian 2* and *The Historian's Macroscope* bridge the gap between technical knowledge and its application by historians.

The remainder of this poster features brief descriptions and selected visualizations from the projects created by the graduate students in the course.

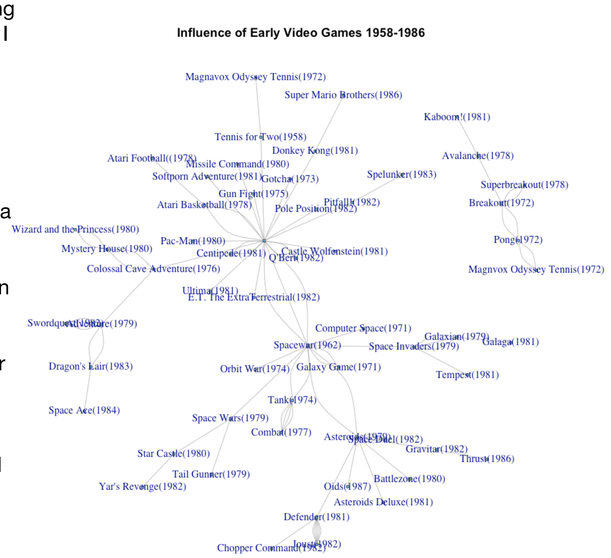
Mapping the Church Mission to Deaf-Mutes, 1873–1879 / Jannelle Legg

The Episcopal Church Mission to Deaf-Mutes sent missionaries across the United States. Using its records, I created an interactive map demonstrating the prevalence of the mission before the founding of other Deaf organizations.

Project URL: http://jannellelegg.com/portfolio/FinalProject/

Networking the Early Video Game Industry / Anne Ladyem McDivitt

This network graph shows of a selection of video games connected by their known developers. It demonstrates the connections that took place from the late 1950s into the mid-1980s. Going into this project, I had a clear opinion that many of the early video games up to the middle of the 1980s were derivative, especially from *Space War!*. After doing network analysis in R, I concluded that there were also a significant number of original games in the period. I came to this revised conclusion because I had to expand the data set to include any game with known developers, rather than just games that were known to have been derived from a another game. Programmatic analysis thus allowed me to make new and better use of sources I might otherwise have passed over.

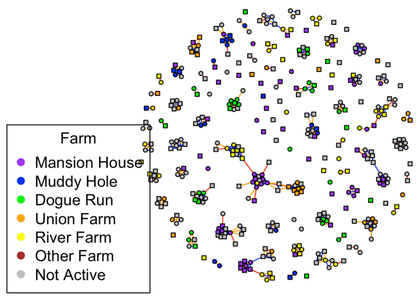


Project URL: http://anneladyem.org/?page_id=503

Network Analysis of the Enslaved Community at Mount Vernon, 1786–1799 / Sara Collini

This project explores the kinship relations of George Washington's Mount Vernon enslaved community from 1786–1799 using the igraph package in R. The Mount Vernon Historic Preservation and Collections department is currently building a database of slaves from every available primary source, such as farm reports and slave censuses. This project explores the kinship relations from the two most complete censuses, 1786 and 1799. In this network graph, each vertex is a slave, and each edge is a kinship relationship. The vertices are colored according to each of the five farms comprising Washington's estate. The Mount Vernon enslaved community follows the pattern of other eighteenth-century Chesapeake plantations as described in Philip D. Morgan, *Slave Counterpoint*: the number of slave kinship ties increased and grow more stable in the late eighteenth-century. Two-parent families increase, and the number of single persons decrease.

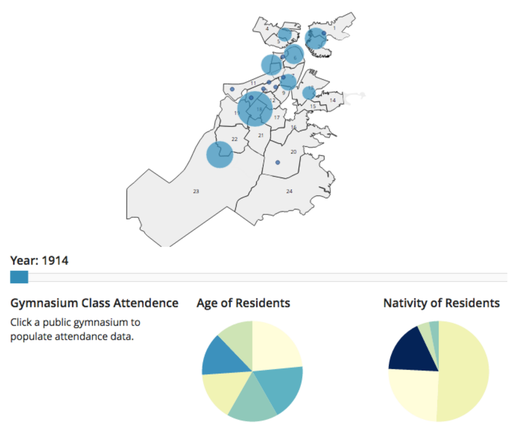
Slave Family Structures on Mount Vernon Farms in 1786



Project URL: http://www.rpubs.com/scollini/final

Mapping Gymnasiums in Boston, 1914–1925 / Amanda Regan

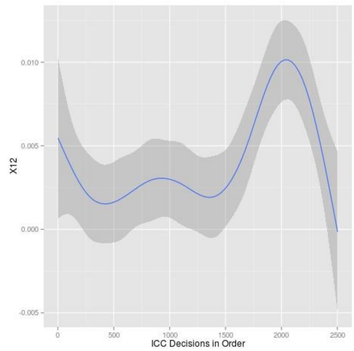
In early twentieth century, Boston's gymnasiums were designed to improve morality and hygiene, foster community, and encourage exercise for the city's residents. This visualization shows the location of each municipal and private gymnasium in Boston between 1914 and 1925 as well as demographic and class information for Boston's wards. The visualization was created using D3.js, combining data from reports of the Boston Parks Department, city directories, census records from the Integrated Public Use Microdata Series, and a map of Boston's wards digitized with QGIS from an original in the Boston Public Library. This maps showed me how municipal gymnasiums were usually placed in immigrant neighborhoods while privately owned and membership-based gymnasiums were located in middle-class neighborhoods. This finding contradicts the discourse about gymnasiums from those who supported these institutions. The map suggests that the aims of the public and private institutions were quite different. The public gymnasiums were thought to help clean up the streets and promote proper hygiene, while private gymnasiums were a social gathering place for men. The visualization along with my related secondary and primary research leaves me with new questions about the role of the gymnasium in the body that will be explored in a future project.



Project URL: http://www.amanda-regan.com/Clio3-Final

Mining the ICC: Macroanalysis of the Decisions of the Indian Claims Commission / Peter Carr Jones

The Indian Claims Commission was a legal body that adjudicated hundreds of claims that Indian Tribes had against the United States for past wrongs. It produced 43 volumes of decisions over more than 30 years. My project analyzes how the Commission used historical knowledge to arrive at legal decisions by using text mining to explore a large corpus of legal documents computationally. To mine the decisions, I first scraped them from Oklahoma State's digital repository with wget. Next, I used a Makefile and Google's tesseract software to convert PDFs into machine-readable text. I then used the rmallet package in R, created stoplists for the names of tribes and places, and visualized the resulting topics. A cluster dendrogram helped me see which topics were similar to others. I initially set out to examine what themes were similar or dissimilar in decisions where tribes were compensated for historical wrongs. While topic modeling did not prove which themes were most supported by history, it did provide new questions to ask. For example, the natural resources cases—timber, oil, and gold topics—were clustered together even though the cases involved different nations, geography, and terminology. I am now exploring why these cases are clustered using traditional close reading. I was also able to compare and contrast my topics to the prevailing historiography. I noted that my "accounting" topic peaked as these cases were in vogue during the latter part of the Indian Claims Commission lifetime.



The "Accounting" Topic

Project URL: http://petercarrjones.com/projects/mining-the-icc/

A Network Analysis of Selected Organizations in the Early American Republic / George D. Oberle III

This project served as a prototype to determine if visualizations of shared relationships between the members of different organizations could be useful in exploring specific questions for my dissertation, "Institutionalizing the Information Revolution: Debates over the National University in the Early American Republic." Learned societies have been studied in isolation from each other; therefore it is difficult to understand the relationships between different types of institutions. Network analysis assists with this challenge. The main sources used in this project were the membership rolls of the associations: in this case, 2,024 names from eight different organizations that supported the founding of a national university. This network graph—created using the igraph package in R—shows organizations as vertices and shared members as edges. The American Philosophical Society (APS) and the American Academy of Arts and Sciences (AAAS) were clearly connected. In addition, an interesting and important connection exists between the AAAS and the Massachusetts Medical Society (MMS). Further research may demonstrate a strong relationship between the burgeoning medical profession and significant connections to scholars of natural history and chemistry. Network analysis offers new ways to deal with the paucity of direct evidence linking people across different groups. Adding new computational methods into Clio's toolbox may offer historians new insights using previously unused sources.

Organizations Supporting the Founding of a National University

Project URL: http://georgeoberle.org/earlyrepublic/final-project-clio-3/