Teachable Insights from Working with The Mary Eliza Project for Gastronomy Students Working with Data

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020751

Laura Kitchings
The National Coalition of Independent Scholars
North Grafton, Massachusettes, United States of America
0000-0002-1336-1038

This paper discusses the use of the Boston's City Archives dataset for a Data Analysis Course Completion Certificate from General Assembly. It documents insights gained while doing computational archival science using visualization tools designed for business use to be shared as part of a textual analysis workshop in October 2022 for graduate students in Boston University's Gastronomy Program. The workshop focuses less on computational tools, although students will be using Google Sheets, and focuses more on challenges and considerations when working on digital humanities projects.

Keywords—Data visualization tools, 19th Amendment, Dias in Data Analysis, Irish Immigration, Mapping Tools

I. Project Background

I began using Tableau Public for humanities research in 2020 when I had to transition my thesis to a digital project due to COVID-19 restrictions. My initial plan for Boston University's Gastronomy (Food Studies) Master of Liberal Arts final project was based on David Pearson's work on Book Provenance [1]. The goal was to incorporate my background in Cultural Heritage, most recently as an Archivist, with my food studies work. The proposed work involved examining multiple volumes of specific cookbooks in a variety of cultural heritage organizations for signs of use. Prior to 2020 the project had already been challenged by the fact that cookbooks accessioned by Museums and Special Collections are often pristine.

During the initial project, I developed an interest in how the cookbook creators used language to explain their food experiences, and often the prescriptive, aspirational worlds of the cookbook users. Through Twitter I discovered The Data-Sitters Club project exploring Baby-Sitters Club using computational text analysis tools, and based on their work, began using the concordancer AntConc [2].

I copied and pasted the word frequency results of specific cookbooks into Google sheets and determining how the same words were used in similar and different ways in the various cookbooks. At that time, I was unaware of computational tools that might have been available to me through Boston University and used Adobe Acrobat's Optical Character Recognition (OCR) capacities to create text files from PDF requiring significant time-consuming cleaning of the text documents. I have since found text files of specific cookbooks online that have been proofread already.

Through Internet searches, I found a few workshops on SQL and Python. While I did take some of the workshops, my focus needed to be on the cookbooks and not learning new computational tools and languages. I also took a one-day workshop though the for-profit education organization General Assembly focused on the data visualization tool Tableau Desktop, and I soon incorporated Tableau Public (the free, more-limited version) into my workflow. The eventual, inelegant workflow became the bulk of my thesis for the program [3]. I have since modified the always in-progress workflow and will be leading a planned virtual workshop in October 2022 for students in the Gastronomy Program's Cookbooks and History taught by Dr. Karen Metheny on creating text analysis visualizations to ask new questions of cookbook word usage.

After finishing the degree, I continued to use Tableau Public for wine certification classes, primarily the mapping tools, to document my wine experiences. In Winter 2021 I enrolled in the 10-week virtual Data Analytics workshop through General Assembly. The class focused on Data Analysis for Business with three required projects ideally using the same dataset. After finding myself uninterested in the presented business datasets, I saw Tweets related to the Archival Kismet conference by Dr Laura Prieto and Boston City Archives Archivist for Reference and Outreach Marta Crilly celebrating The Mary Eliza Project, which encourages exploration of the Boston City Archives releases of transcribed data from the 1920 Women's Voter Register Dataset. The data is notable because in 1920 the ratification of the 19th Amendment to the United States Constitution prohibited states from denying the right to vote to citizens of the United States based on sex. However potential

female voters could still be denied the right to vote for other reasons such as race [4][5].

The first release of transcribed data from the project intrigued me because it focused on Boston's Ward 6, which historically connects to the history of immigration in Boston [6]. While in the Gastronomy Program I worked on a project headed by Dr. Megan Elias on the history of Jewish food in Boston. Our team focused on the food history of the large Jewish community in Boston's North End between 1880-1930. Many members of that community immigrated from what I am currently referring to the "Former Russian Empire," e.g., Ukraine. I was curious if I could find traces of women from this immigrant community in the New Voter data.

The General Assembly instructor granted permission to use the data for the first project, based on using Microsoft Excel for data analysis. Our project rubric, which influenced my project, encouraged the use of advanced Excel visualizations and I decided to focus on Excel's mapping capabilities. Now, working with the dataset in light in light of Jacqueline Wernimont's work of the dangers of simply looking at people as numbers, I plan to change back to the original dataset which includes the women's names [7]. This will hopefully allow me to conduct research to their individual agency and determine possible social connections of the New Voters, not just their employment type which became a partial focus of the final General Assembly Project.

This paper includes data released after my final General Assembly project and includes the transcribed data accessible as of November 18, 2022. More data releases are expected and may change the direction of my research. So far, his has been a project of learning to work with data and is now moving into the stage of what Cole Nusbaumer Knaflic calls explanatory analysis [8].

As shown in Fig. 1, while new transcribed data is being regularly released from the project, there is now enough data to identify a few trends in the data, while identifying the audience for my insights.

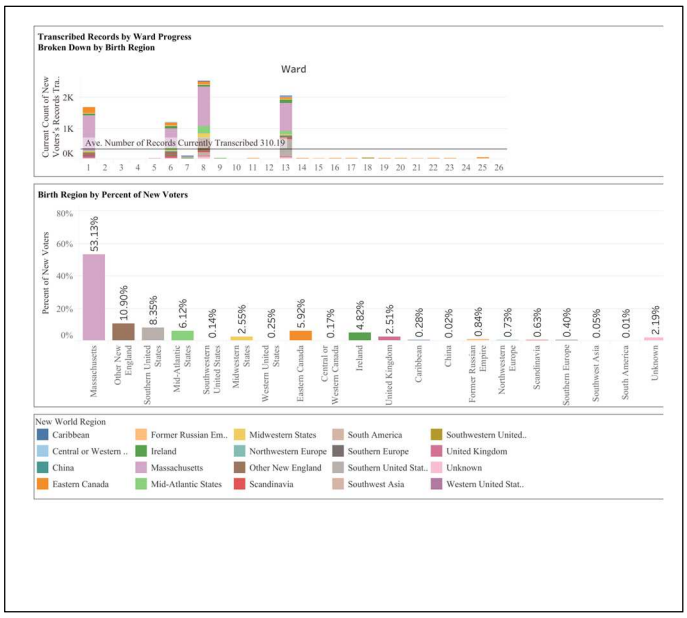

Fig. 1. New Voter Data broken down by transcribed ward and New World Region. New World Regions were created in Tableau to clarify the New Voters' birthplaces.

II. Initial Explorations of the Mary Eliza Project Data

A. Initial Findings

During my initial exploration of the data, I was surprised to find an equal balance between employed women and women who listed their employment using terms related to unpaid domestic labor. A common way to identify this in the registry was to use the term "At Home" which possibly suggests that the women were providing unpaid domestic labor for other family members. However, it is possible that many New Voters working "At Home" had this domestic labor provided by others. The insert in Fig. 2 shows this breakdown. While I assume that many of the employed women found their activism encouraged through workplace connections, I am now asking questions about the social connections of the women working At Home.

I was surprised to find that most of the New Voters were not born in Boston as shown in Fig 2. I am now assuming that the industrial revolution played a role in these women migrating to Boston. I hope to further explore the relationship between in the context of women moving to cities for employment reasons and their agency and activism in registering to vote

B. Mapping in Excel and Tableau Desktop

The initial exploration gave me false hope in the ease of creating map charts in Excel. I found it easy to explore the birth states and provinces of the North American born New Voters who had moved to Boston, and question if the New Voters born in the same state had pre-Boston social connections to each other. Fig. 3 shows the North American Birth States and Providences of the New Voters.

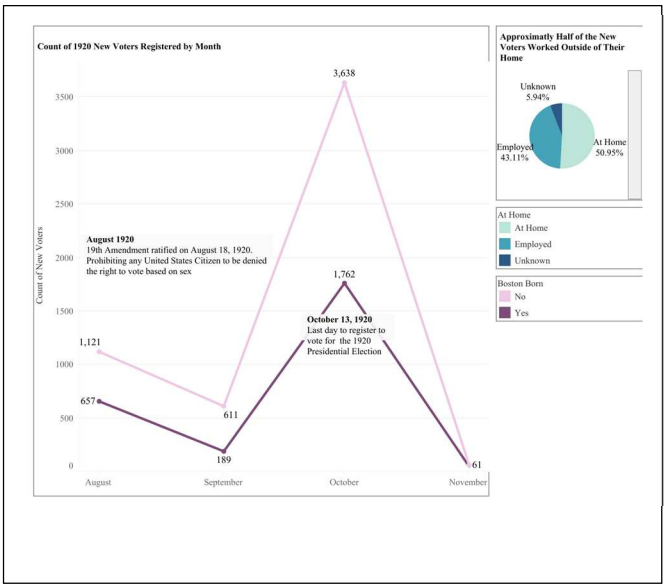

Fig. 2. New Voter Data broken down to consider the number of New Voters not born in Boston and the percentage of employed New Voters

I then began exploring the demographics of New Voters born outside of North America and became challenged by the mapping tools in Excel. Excel uses current official names of countries on their back end, not historic names. While the register gives no obvious indication of the religious identity of the New Voters, I did find New Voters from the Former Russian Empire and was curious about their demographics. The use country names by New Voters from the Former Russian Empire and Caribbean as shown in Fig 4. I created a new column in the spreadsheet to recorded what I assumed was the current official name of their recorded birth country. This became fascinating and time-consuming work. Often the New Voter had "Russia" recorded as their birth country but included additional information about their village or birthplace. I used resources from my earlier North End food research to determine the current country of birth as best as possible for each of these New Voters, and the explore the Russian Revolution and its impact on immigration demographics in the United States [9].

I am hoping to explore more about the New Voters from the Caribbean, curious if the demographics will change with future releases of transcribed data.

I had not considered the difficulties in mapping the birth places of New Voters who had "Ireland" recorded as their birth country. All of Ireland was under British Rule when these New Voters were born and the British Parliament divided the nation into Northern and Southern Ireland, and in 1922 the Southern section became independent from the United Kingdom [10]. The New Voters recorded as being born in Ireland often had their birth counties recorded, not states. This made it difficult to include Ireland and the United Kingdom on the same maps with North American born New Voters. I ended up mapping the Irish and United Kingdom born New Voters by county for future investigation as shown in Fig 5.

In my version of the spreadsheet, I used the created Current Country to separate out the New Voters into Ireland or the United Kingdom based on their recorded county. There has been significant scholarship on how their connections to labor movements in Ireland in and the United States led to political activism around voting for women who identified Ireland as their birth country [11].

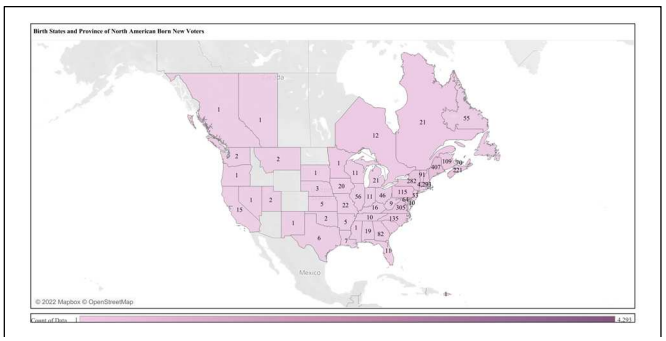

Fig. 3. New Voter Data broken down by North American State or Providence.

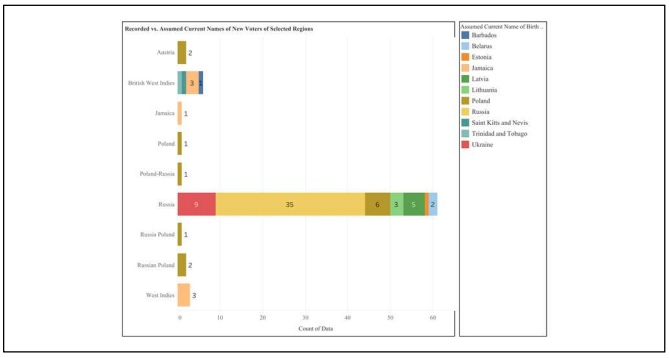

Fig. 4. New Voter Data showing how recorded birth countries can vary from current birth country names, reflecting world events.

In the second and third releases of data, I compared the demographics of the New Voters born in Ireland, the Former Russian Empire, and the Caribbean. I discovered that the women born in what is now defined as Ireland tended to be older than New Voters in other demographics as shown in Fig. 6. It turned the focus of my current data and secondary source exploration to considering how events occurring in rapidly changing parts of the world shaped the Right to Vote Movement in Boston. My current focus is if the women recorded as born in Ireland had connection to political movements in Ireland and exploring if large groups of New Voters immigrated from Ireland at the same point of time.

While these groups may be considered outliers in the larger data set secondary literature shows that members of these groups participated in the Right to Vote movement differently than the North American born activists, especially New Voters recorded as being born in Ireland. The transcribed data shows that 4.82% of the New Voters were born in what is currently Ireland, 2.51% born in the current United Kingdom (including what is now considered Northern Ireland). While only 0.28% of the New Voters were born in the Caribbean and 0.84% in the Former Russian Empire, I am curious if these percentages will remain the same as more transcribed records are released.

III. Insights for Gastronomy Workshop Students

This section is not intended to be infantizing but allow students to start their own digital humanities projects with awareness of challenges they may encounter. These insights may be obvious to those with data backgrounds; however, I am assuming no knowledge of data work for this workshop. Students in the workshop are attending the workshop as part of their work either a Certificate in Food Studies or a Master of Liberal Arts in Gastronomy. Students are generally working full-time while in the program and are not planning to attend doctorate programs, although in the past few years more students are continuing their research in an academic setting.

A. Document Moments of Challenge and Surprise

In *Data Analysis in Qualitative Research*, the authors use the term "Abduction" for moments of unexpected findings [12]. Organize your time to explore the data, not just look for things you expect to find. There will be challenging moments when you cannot yet understand what you are seeing in the data, but trust.

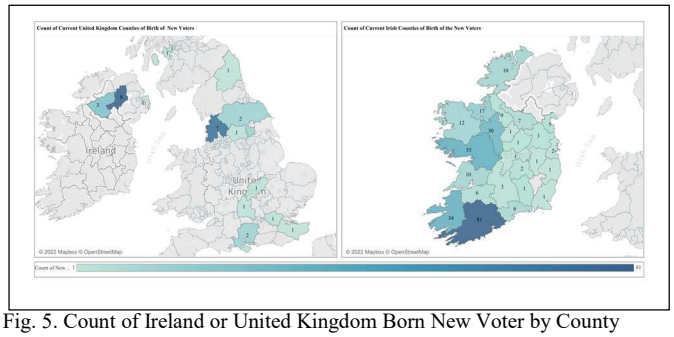

Fig. 5. Count of Ireland or United Kingdom Born New Voter by County

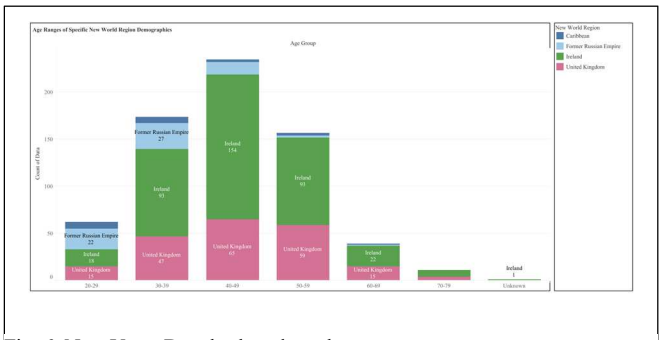

Fig. 6. New Voter Data broken down by age group (grouped in Tableau) or specific world regions

B. *Document Moments of Challenge and Surprise*

In *Data Analysis in Qualitative Research*, the authors use the term "Abduction" for moments of unexpected findings [12]. Organize your time to explore the data, not just look for things you expect to find. There will be challenging moments when you cannot yet understand what you are seeing in the data, but trust that with additional data or visualizations patterns will become clear.

C. *Start with Your Current Skill Set*

To create data humanities projects, it may be assumed that you need to learn new computational tools and programming languages. For the workshop we are using Google Sheets for our exploration and initial analysis. As a post from The Data-Sitters Club makes clear, start with known tools and expand your knowledge as works for your projects [13]

D. *Data Work is Never Neutral*

In the past few years data visualizations have been constant on cable news channels and it is necessary to remember that every time you clean the data or create a visualization you are changing the data in some way. As *Data Feminism* explains, data science projects can re-enforce existing hierarchies [14]. For this specific project it means that I must make sure I am not treating each demographic as a monolithic group and consider the individual agency of each New Voter. I also need to remember that many Boston women still could not register to vote in 1920 due to other laws and practices that served as gatekeeping for voter registration in Boston.

For students in the Gastronomy Program this can be likened to discussions in the field of Food Studies around how every food interaction deals with intersections of politics, heath assumptions, cultural traditions, and many other challenging issues.

E. *Prioritize Self-Care*

Working on a data humanities project can be frustrating and time-consuming, but the rewards may be worth it for your research. Find ways to celebrate small victories. Find ways to regularly take breaks from the data. This is challenging for Gastronomy Students who are often pressed for time. If possible, find others doing similar types of research work to discuss your process and frustrations for support.

F. *Importance of Backing Up Work*

While we will be using Google Sheets and Google Docs which save changes automatically, other tools do not save automatically. Also, have backups of important files. This may be obvious, but many people have stories of losing important work due to technical problems or human mistakes.

G. *Record Your Workflow*

Even if you begin by following an already created workflow, you will change the workflow. Sometimes, especially if you are changing it in increments, you may not realize you are adjusting how you are using computational tools in new ways or choosing to do one step before other steps. Review your workflow regularly and note changes.

ACKNOWLEDGMENT

This work would not have been possible without the support of Dr. Laura Prieto, Marta Crilly, and the transcribers of the Mary Eliza Project at Simmons University and the City of Boston Archives. At Boston University's Metropolitan College Dr. Megan J. Elias and Dr. Karen Metheny nourished my early attempts at data analysis. Former Gastronomy students Sara Steele, and Janaea Eads were teammates on our North End food research project. General Assembly instructor Candace Pereira-Roberts, Teaching Assistant Rachel Denton and the other students in my General Assembly Data Analysis Certification workshop gave useful feedback to the early stages of my New Voter Data Analysis. Sarah Keenan and Suzanne Kitchings provided significant unpaid domestic labor for me in the form of food procurement and preparation that allowed me to focus on my thesis and later Data Analysis work.

REFERENCES

[1] D. Pearson, Provenance research in book history: a handbook**.** Oak Knoll Press: New Castle, DE, 2019

[2] A. Lang, "DSC #4: AntConc Saves the Day." The Data-Sitters Club, 2020. https://datasittersclub.github.io/site/dsc4.html

[3] L. Kitchings, "Intersections of print cookbooks and information science: creating a workflow to visually explore historical printed cookbooks (1872–1920)," unpublished.

[4] Boston City Archives, The Women's Voter Register Dataset. updated August 2022, last accessed September 29, 2022. https://data.boston.gov/dataset/1920-women-s-voter-register

[5] Boston City Archives, The Mary Eliza Project: ward 8 voter records now avalible. updated August 2022, last accessed September 29, 2022. https://www.boston.gov/news/mary-eliza-project-ward-8-voter-records-now-available

[6] A. Goldfeld, The North End: A brief history of Boston's oldest neighborhood. Charleston, SC.: History Press, 2009, pp 143-152.

[7] J. Wernimont, Numbered lives; life and death in quantum media. Massachusetts Institue of Technology, 2018.

[8] C. Knaflic, Storytelling with data: a data visualization guide for business professionals. Wiley, Hoboken, NJ, 2015.

[9] History.com Editors, Russian Revolution. updated July 28 2022, last accessed September 30, 2022 https://www.history.com/topics/russia/russian-revolution

[10] D. Roos, "How the Troubles Began in Northern Ireland". updated November 12, 2021, last accessed September 30, 2022. https://www.history.com/news/the-troubles-northern-ireland

[11] E. Mahoney, host, "Ballot Blocked Episode 4: The Radicalism of Irish American Women," *Morning Edition*, National Park Service updated August 8, 2021, last accessed September 30, 2022. https://www.nps.gov/articles/000/ballot-blocked-episode-4.htm

[12] S. Timmermans and I. Tavory, Data Analysis in qualitatiove research: theorizing with adbuctive analysis. Chicago, IL; University of Chicago Press, 2022, p. 3.

[13] K. Bowers, Q. Dombrowski, and R. Risam. "DSC #12: The DSC and the New Coding Language." The Data-Sitters Club, 2021. https://datasittersclub.github.io/site/dsc12.html.

[14] C. D'Ignazio and L. Klein, Data feminism, The MIT Press, Cambridge, Massachusetts, 2020.

.