Author's Note - Sanchita S. Kamath

Learners becoming familiar with data visualization and the interpretation of produced graphics was the goal of this workshop. The main principle I tried to follow, while facilitating meaningful learning experiences around data visualization and interpretation methodologies for novice digital humanists working with data, was to seek an optimal balance between foundational knowledge about the process and theory behind tool usage as well allowing for hands-on practice with exploring the dataset.

My responsibilities began with manually extracting attributes relevant to visualization; this was a herculean task, because it would define how our research and workshop went further. Therefore, I looked through all the attributes from the datasheet available, and moved on to weeding irrelevant attributes to the social context we wished to explore as defined in the scope of our workshop. After that, I was able to classify relevant attributes into four categories: Living Conditions, Opinions, Upbringing and Attitudes. Then, I had to determine which of these attribute collections I could include which brought in the aspect of data contextualization. A GSS published report (*General Social Survey Has Created Social Media Archive, a New Source for Public Opinion Data | NORC at the University of Chicago*, n.d.), contextualized data by using current social media data they crawled; while we could not do this due to the limiting scope and time constraints of our workshop, I had the idea to think of a human life in context. Through the class readings on data and visualization in a humanist context, specifically on topic modeling, I was able to understand which approach to take (Callaway et al., 2020; 'Data Feminism', n.d.) when I was initiating discussions for the workshop. Initially, I had begun with correlating attributes within the dataset based on what I found to be interesting or what seemed probable; through the readings, I came by the process of manual cleaning and categorization.

This is where I felt the connection between Digital Humanities (DH) and raw data, and how both can be applied to generate insights. I began searching for DH principles, visualizations, and techniques I could apply to facilitate our workshop, and found several articles that were taking a certain individual variable and finding correlations within the data (Barringer et al., 2020; *Repeat Victimization and the Canadian Context: An Empirical Analysis of General Social Survey Data*, n.d.). I however, did not want to singularly contextualize the dataset; instead, I wanted to view

how affecting external influence is, on an individual. This meant that I wished to find the correlation between living conditions and their opinions, while having some attributes from the upbringing category, for me to be able to visualize how the past and present of an individual can affect their opinions. In lieu of this, I updated the categories of living conditions and opinions to include certain attributes from the other two, and discarded those categories for this workshop.

The aim of our workshop was to help participants understand that it is essential for them to divide the data into manageable sections, based on what they wish to investigate. I also took it as a responsibility to inform participants that the data, while representative of a large American population over the years, are self-reported, and may contain some biases.

After the manual cleaning, it was important to decide how this workshop would be delivered, in the sense that which tools would be used for visualization, which correlations to prioritize, and how to relay the knowledge without being over directive. Ergo, what I tried to do, while developing the content for the workshop was to not let participants execute isolated visualization techniques through several tools without any context; instead I attempted to ensure that students understand how they can unveil embedded socio-cultural narratives from the real world social survey data exploration. While the nature of delivering the workshop did not allow for real-time feedback and iteration, I tried to develop the script and flow of the workshop in such a way that it was easy to understand and follow for any novice learner.

I also modified the coding scheme of the dataset, to make it more understandable for the participants. The instruments I chose for visualization were RawGraphs, Palladio and Tableau. The aim was to help participants understand how to create and interpret graphs. The demonstration of one example which helped guide the participants through the tools in a fashion that did not overwhelm them but allow them to know where they could apply the data. The focus was to help participants understand which tool could be used for what purpose and how they could explore a dataset and tool to understand this purpose, so that they are able to discern what they wish to use for their further projects. While limiting the workshop's scope to non-programmatic browser-based instruments,

I conduct a small demonstration of how one can interpret the graphs generated by Tableau and also reference Tableau's enhanced capabilities in the explanation, should attendees wish extending their explorations.

I held a narrow focus on tools applicable specifically to social survey data analytics to maximize attendees' ability directly applying lessons to their own research. The reason for using RawGraphs and Palladio was, that they could be easily accessed on a browser, and could be used easily due to their simplistic and intuitive interface and the lack of prerequisite knowledge.

The narrative sequence was intentionally framed in a way that would allow users explore the social contexts associated with the data and how different tools can be leveraged to explore them. I designed this session's arc expressly not as an enumerated set of discrete software tutorials, but an integrated start-to-finish narrative guiding learners thoroughly from raw dataset through refined analysis-ready corpus to elucidating visualizations. This intentional sequencing flows in a repetitive pattern with the information transmission being uniform for all tools to prevent cognitively overloading learners with disjointed narratives as they incrementally tackle key steps themselves; since providing an opportunity for hands-on experience was a key goal of the workshop.

Finally, comparison between multiple tools was conducted in a systematic fashion, so that one could revisit the tools cognitively, and validate their beliefs about the usage of the tool. Through conducting this workshop, I was able to understand how to apply tools, purport information in a cognitively simplistic fashion, and what it truly means to handle humanistic data - complexities, correlations, biases (which hasn't been well addressed throughout the workshop, I admit) and hypotheses that more often than not, may not be supported by data.

Reproducibility

Should other instructors wish to reproduce this workshop, there are some factors they should consider:

1. Content Adaptation:

    a. Tailor the workshop content to suit the specific needs and backgrounds of your participants. Based on their learning experiences, the workshop level should be modified to cater to their skills.

    b. The research of the participants should be an important factor to consider. Instructors must ensure that the dataset used aligns with the interests or research focus of your audience.

2. Tool Selection:

a. While I used RawGraphs, Palladio, and Tableau, instructors should feel free to adapt the tool selection based on the familiarity and preferences of their participants.

b. Depending on the participants, tool type would change from browser-based to more technical.

3. Contextual Emphasis:

a. It is important that instructors emphasize the importance of understanding the social context of data throughout the workshop.

4. Hands-On Practice:

a. Instructors should provide ample opportunities for participants to engage in hands-on practice with the selected tools by demonstrating one or more examples to guide participants through the visualization process without overwhelming them.

This workshop was designed to provide learners with transferrable end-to-end foundations communicating stories designed through numbers. By taking care to focus the scope on nurturing mindsets and postures approaching data visualizations, I hoped to build learners' confidence asking ever-better questions through visualization they generated based on this need to produce data shaping policies governing people's lives. Their enhanced capacities put at society's service promise expanding who wields influence crafting the datasets of tomorrow more equitably by exposing voices and priorities long obscured.

Limitations and Shortcomings of my contributions to the workshop

While I mentioned how it is important to be cautious of underlying assumptions and biases rampant in self-reported data, I failed to explain what these biases could be due to a shorter span of time spent on exploring the data itself. Further, due to the time taken to actually sort the data, and lower computational resources in Personal Computers, I could not demonstrate how I actually cleaned the data, instead I handed over the pre-cleaned data which Inyoung further filtered through her workshop demonstrations.

Further, I could not create a general hypothesis that every tool could address due to the vastly different benefits of every tool and what its capabilities were. Instead, I left an overarching hypothesis for the interpretation section powered by Tableau., and drew conclusions based on generated visualizations.

Acknowledgements

Citations

Barringer, M. N., Sumerau, J. E., & Gay, D. A. (2020). Generational Variation in Young Adults' Attitudes toward Legal Abortion: Contextualizing the Role of Religion. *Social Currents*, *7*(3), 279–296. https://doi.org/10.1177/2329496520905020

Callaway, E., Turner, J., Stone, H., & Halstrom, A. (2020). The Push and Pull of Digital Humanities: Topic Modeling the What is digital humanities? Genre. *Digital Humanities Quarterly*, *014*(1).

Data Feminism. (n.d.). *Data Feminism*. Retrieved 16 December 2023, from https://data-feminism.mitpress.mit.edu/

General Social Survey Has Created Social Media Archive, a New Source for Public Opinion Data | NORC at the University of Chicago. (n.d.). Retrieved 15 December 2023, from https://www.norc.org/research/library/advisory--general-social-survey-has-created-social-media-archive.html

Repeat Victimization and the Canadian Context: An empirical analysis of General Social Survey Data. (n.d.). Retrieved 15 December 2023, from https://central.bac-lac.gc.ca/.item?id=MQ86092&op=pdf&app=Library&oclc_number=56752021