Ohio AAPI Voter Outreach Visualization Tool

Darian Bhathena, Noah Faro, and Rishi Shah

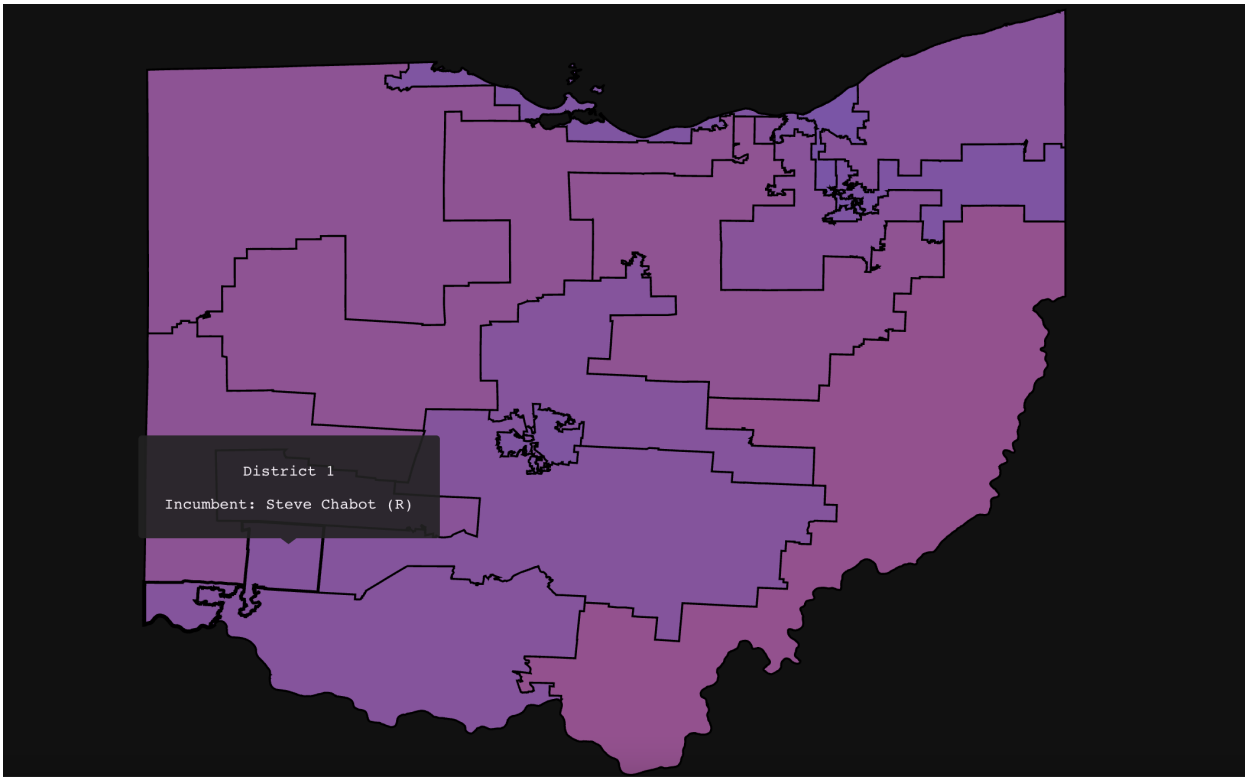

Fig. 1. Displaying US House districts in the visualization tool.

Abstract— In the last decade, significantly more work has done to integrate data-based decision making into political campaigns, particularly in the US. This is resulted in hyper-focused targeting of voters online, particularly via social media and a boooming digital advertising/data marketplace industry that is collecting an immense amount of information on the general public. However, both election results and recent work supports the relatively small effects of advertising compared to building a robust local turnout and engagement operation. This project creates a visualization tool to enable grassroots voter outreach to Asian American and Pacific Islander (AAPI) voters in the state of Ohio.

Index Terms—Visualization, voter outreach, asian american pacific islander

1 Introduction

With the growing importance of grassroots outreach in political campaigns, it is more important than ever that organizations organizations have comprehensive and expressive visualization tools to to target and customize their efforts to be as effective as possible. In order to create these tools however, an intimate knowledge of data visualization and data science has to combined with the expertise and requirements of local grassroots leaders in order to be as effective as possible.

For our final project in 6.859, Interactive Data Visualization, our team worked closely with a not-for-profit grassroots organization in Ohio and a political science researcher at the University of California San Diego (UCSD) to create a specialized voter outreach visualization [3]. Since our partner organization is specifically focused on encouraging civic engagement in the Asian American and Pacific Islander (AAPI) community in the state of Ohio, we were able to appropriately scope our project for the purposes of the course, while contributing to the civic engagement of this local community.

In discussions and interviews with our partner organization, it was clear that a visualization tool would be helpful to enable better targeting of outreach efforts. As such, our primary goals for the visualization tool were to:

1. Provide political organizers with the ability to gain insight into the AAPI voters of specific localities in Ohio, based on prominent voting districts in the state
2. Allow for users to export specific subsets of public voter data
3. Showcase a specific workflow or storyline about AAPI voter engagement

2 Related Work

One aspect of important design decisions we made was the encoding of data on the central state map of the visualization tab. In deliberating this

- *Darian Bhathena: darianb@mit.edu*
- *Noah Faro: nfaro@mit.edu*
- *Rishi Shah: rshah36@mit.edu*

design decision, we took inspiration from the work of Andy Woodruff, a notable cartographer and data visualization expert. Specifically, we took inspiration from his 'value-by-alpha' methodology, where data is represented through the alpha-value (transparency) of the representation [9]. Because this methodology is so central to the first-impression of the visualization, we also adopted a similar color scheme to that of Woodruff's, in order to further emphasize the value-by-alpha data embedding.

It is worth noting that in the creation of this visualization, we were consistently reflecting on the ethical implications of the visualization tool. Much of this was inspired by in-class review of work by Michael Correll and Catherine D'Ignazio. Specifically, we wanted to be conscious of the implications of how we showcase data on the underrepresented minority population of AAPIs. Particularly notable from the related work on this topic was how we chose what preset options are provided to users to filter data by (in our case, the localities of US House District, State Senate District, etc) and what steps can be taken to prevent the misuse of the voter-level individual information that is exportable from our visualization tool [7, 8].

3 Methodology

3.1 Data Sourcing and Wrangling

Our primary source of data was the Ohio Secretary of State's Voter Files [1], which is a public repository of all registered voters in the state. This database includes most information that each voter includes in their registration such as mailing and residential addresses and birth date. Notably, the database also includes the voting records of each registered voter (indicating whether or not individuals voted, rather than who they voted for).

Our second significant data source was from Dr. Tom Wong, our UCSD political science partner [3]. In his research, Dr. Wong has created a metric to estimate the likelihood of a voter being a member of the AAPI community. Utilizing this valuable augmentation, we were able to filter the publicly available dataset in order to focus the visualization on only AAPI voters. Select data sets of supplementary information, such as total populations of different localities, was sourced from the US Census Bureau [6].

A variety of geolocation data sets were also utilized from multiple open sources, including the Ohio Department of Transportation and the organization Open Data Delaware [2, 5].

A large amount of pre-processing was conducted utilizing the Tableau Prep Builder software in order to join the metrics from the different data sources for each individual voter file. In addition, summary statistics were pre-calculated for the different localities displayed on the map, to facilitate efficient loading of the visualization.

3.2 Data Storage

With our web-based visualization, we found that directly loading raw data files into the part of the public directory loaded into users' browsers was inefficient and introducing latency into the loading of the visualization. As such, we chose to store all non-geospatial data on an external database hosted by the service Firebase. Using their *Realtime Database* service, we were able to improve rendering efficiency of the visualization and prevent the web-page crashes we saw in our A4 [4].

4 Results

4.1 Voter Engagement Score

Given that the voting records of each of the voters were part of the publicly accessible data set, we were able to calculate a Voter Engagement Score (VES) for each voter. The VES represents the proportion of eligible elections voters have cast a ballot in since the beginning of the year 2000 (since this is the earliest included election public voting record). It is important to note that eligibility in the election was defined as all elections in which a voter was 18 years of age; for example, if a voter had not registered since the 2016 general election, but had turned 18 in 2012, all the elections that she had failed to register for between 2012 and 2016 would count against her VES.

Furthermore, we calculated three more specific VES variations besides a voter's overall VES, to assess their engagement in specifically general, primary, and special elections.

4.2 Scrollytelling

The first step of our visualization tool as shown to users is a scrollytelling feature. This showcases a specific story line, about the partisan divide in the AAPI community of Ohio. This visualization component fulfills the third central goal of our project, to showcase a specific workflow about AAPI voter engagement.

We made a specific design choice to include this scrollytelling component as the landing page for the visualization, because it provides users with an introduction to the purpose of our visualization tool, an overview of the key concepts underpinning our visualization, as well as a interesting series of visualizations that may prompt users to want to investigate further.

4.3 Map

After viewing a sample story line showcased on the 'Story' page of the visualization, users can move over to the 'Map' page to get a more interactive visualization of the data. This component of the visualization tool fulfills our project's first two primary goals, to allow users more exploration of the data in order to allow for targeted voter outreach as well as for the exporting of specific the voter files.

To allow for sufficient flexibility in the interactive components of the map, we allowed the user to subdivide the map of Ohio by four different localities: US House District, Ohio Senate District, Ohio House District, and City. These four localities each represent the voting districts of prominent publicly elected officials, and ask such serve as food filters for the users of this visualization tool. Furthermore, we utilize zip code to display even more detailed summary statistics about the AAPI population, as zip code is the smallest level of geolocation that was realistic to be visualized. However, introducing zip codes into the visualization posed an interesting design challenge, since the boundaries of the main localities (voting districts) rarely lined up perfectly with its zip codes. As such, we made the decision to display the zip codes while also showing an outline of the larger locality, instead of binning the zip code, since that would have conveyed inaccurate information (conveys that the full population in the zip code resides in the larger locality, which may not necessarily be true).

As part of the development of this map component, we made the design decision of allowing users to export specific subsets of the data, in order to make the visualization tool as useful as possible. We do however recognize that this introduces concerns regarding data privacy. As such, we've restricted the columns of the exported data files to only those that are available from the public data set (rather than data that was received from our project partners). This small step ensures that the tool can still be used for legitimate voter engagement (if users have external access to similar data, they can correlate the filtered data subset from our visualization tool via voter UID) while preventing abuse of this functionality.

5 Discussion

In interviews with our partner organization, it was clear that the creation of this visualization tool enabled specific insights and practices for voter engagement. Specifically, perhaps the largest contribution is the ability for the organization to identify in which area they will conduct their next round of phone bank outreach calls.

These engagement operations, where trained callers contact a specific subset of voters, is primarily for the purpose of expanding the information included in an organization's voter file data repository. With this more detailed dataset, organizations are able to better target voters, advertise their initiatives more strategically, and recruit volunteers by engaging with these voters on issues they are passionate about.

6 Future Work

In future iterations of this project, we expect to expand the visualization to incorporate the person-level data, that is currently only interfaced

through the export functionality. This could take multiple different forms, such as the expansion of a list of individual voters when a specific locality is selected or the visualization of the residential addresses of each voter (which would require a paid subscription to a geocoding service able to encode residential addresses into geospatial coordinates).

Furthermore, we hope to build a functionality such that users can upload supplementary data collected through outreach into the system, to make the visualization more expansive. This import functionality could also open the possibility of analytics running in the background to highlight salient details about the imported data, or a dynamic scrollytelling that changes based upon the imported data.

ACKNOWLEDGMENTS

The authors wish to thank Arvind Satyanarayan, Katie Bacher, Doğa Doğan, and Ebenezer Sefah for their fantastic teaching and mentorship, Dr. Tom Wong of University of California San Diego for the data source, and Rising Tide Ohio, a not-for-profit grassroots political organization focused on civic engagement of AAPI Ohioans, for their guidance and cooperation.

REFERENCES

[1] County Voter Files Download Page.
[2] Data Download ODOT TIMS.
[3] Dr. tom k. wong, university of california san diego.
[4] Firebase realtime database.
[5] Open Data Delaware.
[6] U. C. Bureau. Census data.
[7] M. Correll. Ethical dimensions of visualization research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 1–13. ACM, May 2019. doi: 10.1145/3290605.3300418
[8] C. D'Ignazio. *Data Feminism*. Mar 2020.
[9] A. Woodruff. Value-by-alpha maps — andy woodruff.

A VISUALIZATION PAGE

The visualization can be accessed via our GitHub repository: `https://6859-sp21.github.io/final-project-ohio-voters/#/story`.