Winter 2023

CAPP 30122 Project: The Civic Side

Katherine Dumais (kdumais) &
Francesca Vescia (fvescia)

Project Abstract

We built a civic life dashboard exploring several measures of civic engagement across Chicago. Specifically, we paired mayoral campaign contributions data scraped from the Illinois Board of Elections website with voter turnout and 311 non-emergency service calls data from the city of Chicago to investigate how different zip codes engage in political processes and access civic services.

Our dashboard consists of maps that display metrics specified by the user and tables with more details for each zip code. Map metrics include various measures of civic engagement and service utilization, such as average campaign contribution and number of 311 calls, as well as average housing prices as a proxy for income. The maps are displayed side-by-side, so users can visualize relationships between different metrics of interest. Below the maps, tables display top five 311 complaints, total 311 calls, and campaign contribution statistics for the city of Chicago (default) or a specified Chicago zip code. Lastly, for users curious about hyperlocal politics, a final table maps wards and precincts to Chicago zip codes.

Software Structure

```
30122-project-the-civic-side
      civic-side/
            campaigns
                  contributions
                  saved_pages
            the-polis
                  datasets
            dashboard.py
            graphs.py
            data_prep_funcs.py
            prep_data.py
            __init__.py
            __main__.py
      poetry.lock
      pyproject.toml
      .gitignore
      README.md
```

Our data, and most of the code that scrapes, cleans, and processes it live in two sub modules: **campaigns** and **the_polis**. **campaigns** houses campaign finance data scraped from the Illinois Board of Elections website, while **the_polis** houses data on socioeconomics and civic life – Zillow housing price data, which we use as a proxy for income; voter turnout data; ward-and-precinct-to-zip-code data; and data on 311 calls – for each of Chicago's zip codes. **data_prep_funcs.py**, run by a command in **prep_data.py** (in our primary module, **civic_side**), merges processed data from the submodules into one file that underlies our dashboard maps (generated in **dashboard.py** with graphs from **graphs.py**).

`campaigns` Submodule

As detailed below, Francesca saved static snapshots of each campaign's contributions web page to the saved_pages folder and then scraped those HTML files. get_contributions() and save_contributions() from crawler.py scrape the HTML and save the resulting data to the contributions folder. clean(), merge_candidates(), and process_contributions() from cleanup.py clean the raw data, merge clean data from multiple candidates into a single data frame, and drop contributions received outside the specified window. Finally, contribution_stats() in stats.py generates summary statistics from the processed data. utils.py stores parameters, such as start and end dates for the 2019 election cycle, that are passed to the campaigns functions when they are called in prep_data_funcs.py.

`the-polis` Submodule

In terms of process, for the Polis section of the work, including the Zillow data, 311 data, voting data, and ward to zip code data, the original datasets are kept within the datasets folder. There is a .py file for the cleaning of each of the datasets which has an overarching function that produces individual clean versions of the original datasets in the civic_side folder. Using prep_all_data() from data_prep_funcs.py, we run all of the clean_data functions, including the campaign data cleaning functions, and merge the datasets together. We then use this merged dataset to produce our visuals within the dashboard.py file.

Interacting with the Application

Users interact with our application via the command line and our browser-based dashboard. The user begins by running **poetry install** in the top-level directory, 30122-project-the-civic-side, to activate the virtual environment. Next, the user runs **poetry run python3 civic_side/prep_data.py** to scrape, clean, process, and merge the data underlying the dashboard. Finally, the user runs **poetry run python3 civic_side/dashboard.py** and follows the resulting link from the command line terminal to the browser-based dashboard, where they can interact with our data.

Within our dashboard, voters can view choropleth maps side by side, detailing a variety of indicators across zip codes including voting turnout rates, campaign donation rates, and income. There are two drop downs that allow users to customize their view of these data and to see correlations between indicators. Below this, there is a table with contribution numbers, 311 call information, and voting districts so, with a drop down, a user can learn more about a zip code of their choice, or about the city of Chicago as a whole.

Team Member Responsibilities

Francesca scraped and processed the campaign contributions data, while Katherine gathered, processed, and merged data from Zillow and the City of Chicago. (This work is demonstrated within the campaigns and polis submodules of our application respectively.) prep_data.py, in the main module, is primarily Katherine's code and is used to bring the datasets together. We wireframed the dashboard together and collaborated on initial experiments with Dash and Plotly, then Katherine developed the dashboard while Francesca implemented the project's final file structure and drafted this paper, which we then finalized together.

Reflections on Process and Product

We set out to build a “Civic Life Dashboard” exploring civic engagement and civic service utilization across Chicago, and we accomplished that goal. We used each of the data sources we identified in our project proposal: campaign contributions data from the Illinois Board of Elections (BOE), Zillow housing price data as a proxy for income, data on 311 non-emergency calls as a measure of service utilization, and voter turnout data.

However, the exact focus of our analyses evolved over the course of the project. Initially, we planned to analyze aldermanic campaign contributions in addition to mayoral ones, and to explore to the extent to which candidates were funded by donors within vs. beyond their jurisdictions. Unfortunately, the BOE website uses ASP.net, which made saving static snapshots of each campaign’s contributions web page by hand and then scraping those HTML files the most effective way to access campaign finance data. It was not feasible to repeat this process for every aldermanic candidate, so we narrowed our focus to the mayoral race. Conversely, voter turnout data, which we identified as optional in our initial proposal, became a key part of our dashboard. We also added a city-wide element to our project, aggregating our zip-level data to give dashboard users the option to compare zip codes not just to each other, but also to the city as a whole.

Challenges we faced over the course of our project involved scraping the ASP.net-based BOE website (as described above); determining how best to map Chicago wards and precincts to Chicago zip codes; and navigating post-2019-election redistricting. We had hoped that that precincts, which are the smallest civic voting unit in Chicago, would be unique to zip codes, and therefore we could get statistically unique data for voting rates, and therefore could add more robust analysis about voters vs. non-voters in terms of engagement. This turned out to not always be the case. We chose to double count voters, one in each zip code, understanding that voting is likely uniform in a four to six block radius and that a group of 300 voters within an on-average 6000 person zip code would weight the calculated average and bias our estimate, but not enough to lead to results that this would drastically change the rate of our result. The redistricting after the 2019 election also made it so, while we were able to get a FOIA request complete, shapefiles were the only way to map to zipcodes for the historic ward and precinct info.

Final conclusions were that for areas for higher housing prices, there were higher numbers of campaign contributions. This is negatively correlated with residential 311 calls, emphasizing that those who financially contribute to the government, through taxes and donations, have better access to government services. This also comes up in the types of complaints within areas: airplane noise and graffiti had some preponderance on the North Side, while potholes and street lights were more common problems on the South Side. However, surprisingly, the largest number of 311 calls were from zip codes with high concentrations of critical infrastructure, more specifically 60666 where the airport is and 60612 where RUSH and UIC Hospitals are, demonstrating the most calls are in relation to getting people access to critical city services.

Future Directions

With more time and resources, we would love to dive deeper into the campaign finance data and explore hyperlocal (i.e., ward-level) elections. These analyses look at number and amount of donations

by zip code, but we expect the additional details available on the BOE website would reveal additional insights. For example, one limitation of this work is that it does not distinguish between different types of campaign contributions. By parsing the donation sources and descriptions available on the BOE website, we could learn more about within- vs. beyond-ward contributions and potentially exclude anomalous data points from our analyses. For example, Francesca observed one donation in the data that was described as a refund. Future analyses could parse donation descriptions to check for other potentially anomalous contributions that could be adding noise to our summary statistics. (These anomalies did not appear common, but it would be valuable to check for them more systematically.)

Richer analyses would also move beyond average voter statistics to look at how voter disenfranchisement at the local level connects to civic engagement and civic service utilization, and would broaden the definition of civic services to include more services that affect community members' well-being, like quality of and access to public transportation, hospitals, schools, and other infrastructure critical to civic life. Finally, we would redo our analyses with data from the 2023 election, which was still underway as we were working on this project, to see if any of the patterns visible in the 2019 data have changed over the past four years.

Notable Discoveries
Highlights from what we learned along the way

- If you look at Woodlawn (60637) and Hyde Park (60615), there is a notably larger number of 311 calls (about 15,890 vs. 9,492) in Woodlawn, and a substantially different average campaign donation ($1,168.71 vs. $588.55), demonstrating the need to ask for services and the correlation between campaign funding and automatic service provision.

- While there are no campaign contributions from the airport zip code which made it not a concrete comparison point, the largest number of 311 complaints were from the zip code within the airport, with airport noise being one of the top 311 complaints. There is also a large concentration of information-only 311 calls centered around major hospitals (specifically in 60612). This suggests many non-emergency calls relate to accessing critical city infrastructure.

- Kanye West paid $73,540 to the Board of Elections to pay off candidate Amara Enyia's debt from her previous campaign.[1]

- Lori Lightfoot received three $0.01 donations. We initially expected a bug in our code, but we confirmed the contributions are listed on her campaign contributions web page. We suspect the donor – all $0.01 donations were from the same person – was testing the donation portal, since they proceeded to make several additional, larger donations over the following months.

[1] https://www.bbc.com/news/newsbeat-45949201