Jennifer Wegmann-Gabb May 9, 2019

Final Project

For my final project I wanted to create a programming project that would help me in my political work. I would like to share a bit of personal background as it reflects on my short-term goal for this project and long-term ideas moving forward.

Upon finishing my MA in Art History in May, I returned to Illinois to figure out my next step. While figuring that out, I reignited my passion for political activism and am now very active in politics (for pay and as a volunteer). My political background is primarily in field operations (holding events, contacting voters, organizing volunteers, getting voters out to the polls) which is still decidedly non-technical in its basics. There is some movement towards developing digital and data tools for national campaigns, by political consultants, or the national party committees (primarily on the Democrats’ side as the Republicans heavily rely on consultants), but it is slow and has not trickle down to local elections or activists. Most recently I worked as a field organizer for a congressional campaign and saw several opportunities for tools to benefit ground operations. In the middle of campaign season is not the best time to develop new programs, and I had not yet started to learn programming. After admission to the iSchool, I decided to focus on programming and data management, as opposed to my original intent of archives and special collections, in order to gain the technical skills to make better use of data and technology in strategizing field operations.

My concept for this project is to take the overwhelming election results data file from the Illinois State Board of Elections and parse the information by precinct to help paint a more specific picture about voter engagement and turnout. I am focusing on my county (McHenry) because it is a manageable dataset, I am familiar with the area, and I am very involved with the local party and this will help our voter engagement efforts in 2020. I plan on continuing to work on this project by developing visualizations and, hopefully, a web-based user-interface for campaigns and local activists to access the data.

I had big plans for this project in my proposal which quickly became tempered as my efforts to read-in and modify the csv files were met with failure and a panic attack manifesting in my final check-in assignment. It wasn’t until one of my classmates lamented her programming project in our class chat and mentioned that she used PANDAS to import her csv files that I turned the corner. I read-in the csv files quickly, but then had to spend a lot of time reading the PANDAS documentation and a variety of blog entries to learn how to manipulate the files. I found that many of the concepts we covered in class helped me to develop my code and understand how to use PANDAS. I felt very rewarded as I worked through many of the code elements we studied.

Once I got the csv files read-in and manipulated as I wanted, I focused on trying to read out the information from a specific precinct to a csv file. I began by just printing the information. I actually achieved this fairly quickly. I also read out the csv fairly quickly too. My trouble began as I transformed this code into a function. The file name process caused me the most trouble. I wanted to pull the information for the file name from the data frame but kept getting errors. I even tried following a suggestion from Stack Overflow to create a column called ‘filename’ and use that but kept getting errors. Finally, I decided not to spend any more time on this particular problem and just create three different functions for my three election files. Not an elegant solution, but it works. I will figure out how to achieve the file name protocol eventually. I also wanted to figure out how to have the files load into a subdirectory. I started to go down that path and tried a couple of options, but none worked, and I still had work to do, so I put that aside for the future.

Once I hard coded in the election year and pulled the precinct name from the function parameter I had my file naming protocol working. Now I needed to figure out a way to loop over the files and pull out each precinct. I created a text file with just the precinct names and used this as the file the for loop goes over and where the function pulls its parameter from. I was so excited when it worked. Honestly, it was kind of a transcendent moment when all my csv files popped up. I created my other two for loops for the other election result files and ran them. Et voilà, the csv files appeared.

I thought I was done. Yay! Except. No. I found that for precincts that had more than 9 precincts the single digit precincts pulled more than just their data (ex. Algonquin 2 pulled its data, but also that of Algonquin 20- 29). At first, I thought it was just because I use ‘contains’ instead of ‘match,’ but when I changed it I still had that problem for the 2016 and 2018 files. So back to the internet I went. I had difficulty finding anything about my specific problem. I’m sure it’s out there, but I just don’t know how to best phrase my question to get the answer I need. I soon realized that Regular Expressions could work to force my query to match. The only problem, I’m not really sure how to best use them. I watched the video from last semester, but I still had questions because I wasn’t sure how to force a match. I also thought about using if-else but as I was writing one I took a closer look at my csv files. A-ha! I thought the files were all formatted the same, but they were not. In the 2018 file, the precinct names with only one number had an extra space and the 2016 file used only three letters for each precinct followed by an extra space for precincts with one number. At first, I tried to figure out how the alter the column in the data frame, but I was having problems changing it. I tried OpenRefine but couldn’t find a way to alter everything easily. I’m sure there is an easier way, but I just do not know enough programming and at a certain point, cleaning by hand was just faster because I could use a program I know well. It was also late, so I decided to clean the data by hand in Excel with ‘find and replace’. Presto, with the cleaned column the for loops worked as intended and I had 636 precinct data files.

After getting over my panic and really focusing on this project, I had a lot of fun and felt rewarded when the code worked. I feel more confident about diving into new python packages to do more with my coding. I am also much more comfortable uploading a project to GitHub which is good, because the political data jobs often ask for your GitHub repository, and now I have something to share. My GitHub is ‘travelingmedievalist’ and this project is at: <https://github.com/travelingmedievalist/McHenry_Election_Data>