Slaying the Gerrymander:
Using Python to Safeguard Democracy

THE GERRY-MANDER, BY BOSTON CENTINEL, PUBLIC DOMAIN. ADAPTED BY SIGHTLINE INSTITUTE.
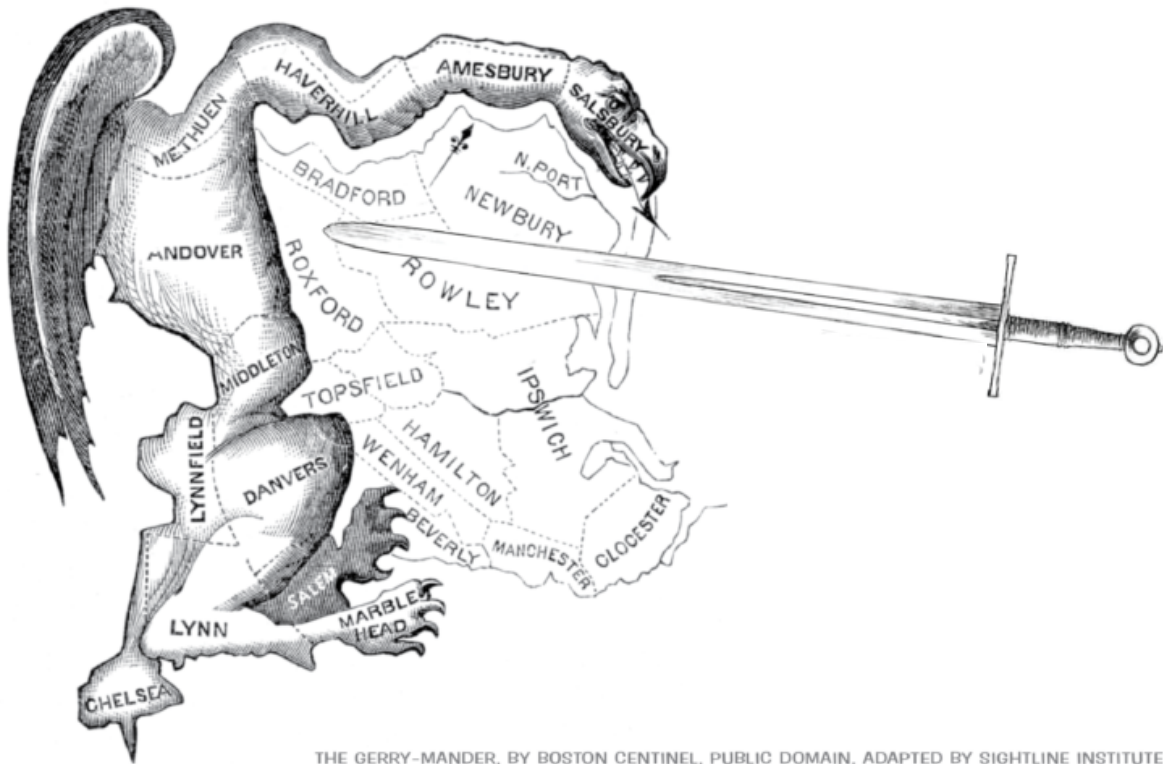
Navo Emmanuel, Charet Bolton, Morgan McCarthy, Jay Aswani

Table of Contents

INTRODUCTION

Abstract

Our project team developed a python program that takes a user's address as an input and returns a personalized call script based on their location. The Graphical User Interface (PySimple GUI) utilizes ChromeDriver and Selenium to find the user's representatives on an external website. Then, it returns a call script with their representatives name and phone number, as well as several measurements of gerrymandering specific to their district. These measurements were drawn from multiple sources including CSV files, web scraping, and API data. These were imported using pandas, sqlite3, BeautifulSoup, re, JSON, and requests. The program output also displays a map of Pennsylvania with gerrymandering measurements developed using Plotly and OGR.

Scope and Purpose

While we know gerrymandering is a problem in state elections, how do social scientists define if a district is gerrymandered? And more importantly, are the districts we live in gerrymandered? What steps can we take to fix this?

Gerrymandering cases have been argued in front of the Pennsylvania Supreme Court, which ruled that US Congressional district maps be redrawn before the 2018 election. However, there have been limited steps taken to examine gerrymandering at the Pennsylvania state legislature level. Critics argue that more must be done to fix gerrymandering in state elections; instead of the voters choosing representatives, the representatives currently get to choose their voters.

Our group aims to support the work of Fair Districts PA, which is a nonpartisan, statewide coalition of organizations and individuals working to create a process for redistricting that is transparent, impartial, and fair. To this end, we developed an application that empowers users interested in ending the undemocratic practice of gerrymandering. A user will input their address to receive information on:

- Their district number
- How gerrymandered the state is overall
- How gerrymandered the district is, based on state rank
- A choropleth map showing levels of gerrymandering by district
- Their local representative's name and contact information

- A call script to contact their representative and speak against gerrymandering

Our group project aims to build a Python program that will give residents of Pennsylvania more information about the state congressional districts they live in, and if they could be gerrymandered. Social scientists have identified several ways to identify if elections are as fair as they could be, including the efficiency gap and district population, which are what our program examines. By creating a tool that casts a light on this issue and giving Pennsylvania residents a representative to contact, we believe we can further improve the value of a vote in our democracy.

Quick Overview

1. Download and extract the zip file
2. Enter your name, address, and download destination folder (ideally the newly extracted folder which contains the CSV)
3. Retrieve the call script and map by going to the chosen download destination folder

SPECIFIC DATA SOURCES

Downloading CSVs

For our program, we downloaded a CSV from Kaggle that contained results from House Congressional elections in 2020. That dataset can be found here. As long as the CSV is in the same folder as the main program itself, the program will be able to access the CSV file. However, it is also recommended that the destination folder chosen be the same as the newly extracted folder.

Web Scraping

Representative Phone Number

We used the Pennsylvania House contact pages of each representative to get their primary phone number. If there were multiple numbers for the member's different offices, we used the first phone number indicated. Due to the recent election and office transitions, there were a number of representatives who did not yet have a contact page. We noted this in our output to the user. An example for Representative Summer Lee of the 34th District is shown below.

The template for the web scrape was found here. Also, we used this link for the example below:

Rep. Summer Lee
34th District / Allegheny County

Contact Rep. Summer Lee

Braddock Office

501 Braddock Ave.
Braddock, PA 15104-9998

P: (412) 273-3400

F: (412) 273-3434

Representative Name

To extract the names of the members by district, our program used the Pennsylvania House of Representatives Members of the House directory. Below is a snapshot of the page. Our program uses the page version with the members sorted by their district number to ensure the data is extracted and stored in order to improve accuracy. The web scrape extracts each member's full name and district number.

API

Our program uses the Census Bureau's API to find the population of each Pennsylvania State House of Representative District. We utilized values from the American Community 5-Year Survey published in 2014, as it was the first state representative election using 2010 Census Data. Therefore, these represented the population values that the 2011 Legislative

[Reapportionment Commission](#) would have used to develop the state house districts, and should therefore be equal.

Below is the API query returns the population for each district in Pennsylvania:

https://api.census.gov/data/2014/acs/acs5?get=B01003_001E&for=state%20legislative%20district%20(lower%20chamber):*&in=state:42

- "2014" is the year that the data was published
- "acs" represents the American Community Survey
- "acs5" indicates it is the 5-year American Community Survey
- "B01003_001E" stands for the Total Population
- The "&for=" indicates that these results are displayed for every lower chamber state legislative district
- The "&in=state:42" filters the results to Pennsylvania

Changing anything in the API query will alter the code and will negatively affect the results of the program.

USAGE INSTRUCTIONS

Necessary Packages and Libraries

Before running the program, please follow the directions at [this site](#) and do so **before** installing the following packages.

Also, please install the current versions of both gdal and pygdal (preferably with both being similar versions. If these are not installed, the choropleth map mentioned later will not generate. However, even if you are unable to install these libraries, the rest of the program will work.
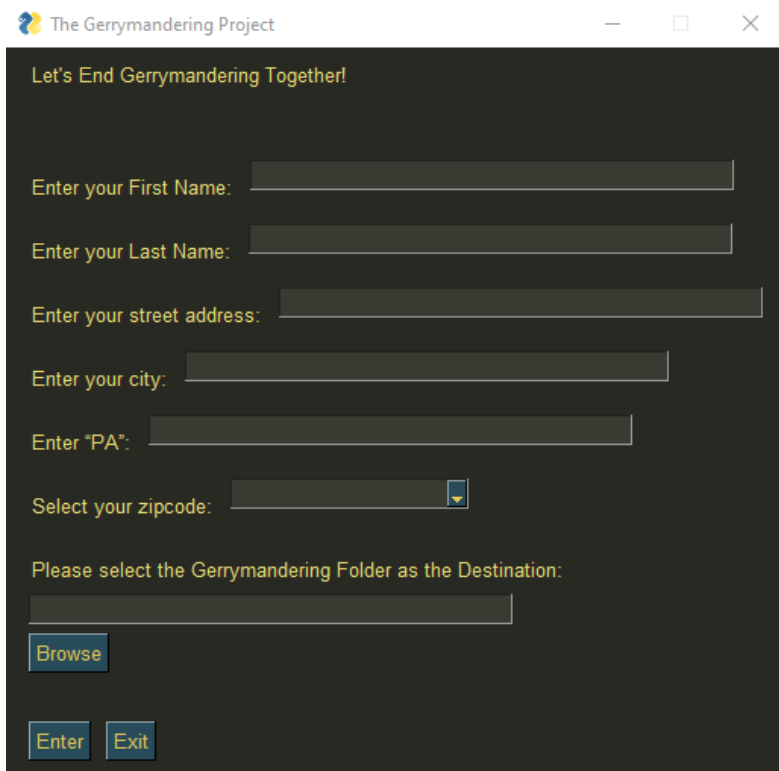
We have made sure to include several "!pip install" statements at the beginning of the program, but we would like to reiterate the packages and libraries you may need to install here:

- chromedriver_autoinstaller
- osgeo
- plotly-geo
- geopandas==0.3.0
- pyshp==1.2.10
- shapely==1.6.3
- selenium
- numpy
- pandas
- json
- requests

- io
- sys
- os
- plotly.offline
- plotly.figure
- plotly.express
- time
- re
- sqlite3
- PySimpleGUI
- bs4

Running the Program

NOTE: When errors appear, simply click the error button to continue. Also, please be patient as the program may take some time to load, we apologize for the delay.
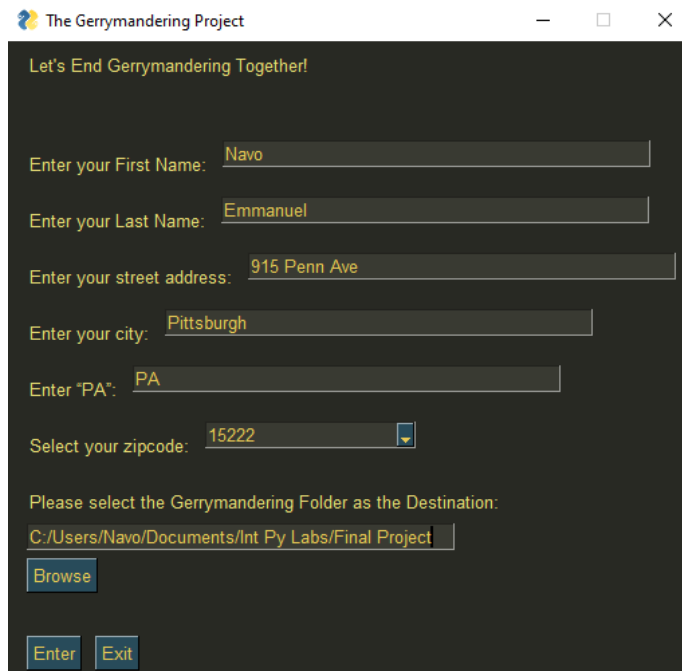
After downloading the zip file, your first step is to open the gerrymandering.py file in your IDE of choice. We recommend doing so in Spyder as that was the IDE we mainly used in building this program. Also, **please make sure that you have the latest version of Google Chrome installed.** Our program uses Google Chrome to web scrape necessary data.

Afterward, you simply run it. You may notice at the top of the program that there are several "!pip install …." commands. These are to ensure that you will automatically download any packages that you may be missing that will be required to run this program.

The program takes some time to start up, so please be patient. However, in a few moments, if you are keeping an eye on your taskbar, you will notice that a new window has appeared. This is the graphic user interface (GUI). Instead of inputting your commands at the console, this GUI will allow you to conveniently enter your information. Once it appears, please input your information. Keep in mind that this program is **only meant to work for Pennsylvania addresses** and that the program is sensitive to the inputs that you provide. **You must enter a valid Pennsylvania address for the program to work correctly.**

Once you have entered your name and address, please select a folder. You can do this by clicking the browse button. This will open up your file explorer so you can select a folder. It is recommended that you select the folder that the program is contained in as some users have reported facing errors when they choose a different folder.

Once you have selected a folder, please select the enter button. In a few moments, Google Chrome will automatically open and you will see the address you entered is automatically inputted into this website here. After your state legislator's information is pulled up, the program will automatically web scrape the page for your lower house representative's name. Google Chrome will then automatically close.

After this, the program will do some calculations in the background. It will use your address to identify your state house district number in Pennsylvania and use this information to identify the

population in your current district with an API. It will then calculate the average population of all districts in Pennsylvania and then compare how far away your district's population is from the average to calculate your district's level of gerrymandering. The farther away your district is from the mean, the more gerrymandered your district is.

Next, using a CSV containing House Congressional election results from 2020, the program will calculate how gerrymandered the state of Pennsylvania is on a national level.

Lastly, it will use your address to also web scrape for the contact information of your district's state legislator. Once all this is complete, if you check the destination folder you selected, you will notice the presence of this new file:

gerrymandering.py	12/17/2020 6:08 PM	PY File	17 KB
EndGerrymanderingCallScript	12/17/2020 6:07 PM	File	1 KB
12.17.20 GUI Integration.py	12/17/2020 6:02 PM	PY File	17 KB
electiondb.sqlite	12/17/2020 2:46 PM	SQLITE File	100 KB
cb_2014_42_sldl_500k.dbf	12/17/2020 2:19 PM	DBF File	32 KB
cb_2014_42_sldl_500k.shp	12/17/2020 2:19 PM	SHP File	1,430 KB
cb_2014_42_sldl_500k.cpg	12/17/2020 2:19 PM	CPG File	1 KB
house_candidate.csv	12/17/2020 2:19 PM	CSV File	65 KB
cb_2014_42_sldl_500k.shp.ea.iso.xml	12/17/2020 2:19 PM	XML File	19 KB
cb_2014_42_sldl_500k.shx	12/17/2020 2:19 PM	SHX File	2 KB
cb_2014_42_sldl_500k.prj	12/17/2020 2:19 PM	PRJ File	1 KB
cb_2014_42_sldl_500k.shp.iso.xml	12/17/2020 2:19 PM	XML File	31 KB
cb_2014_42_sldl_500k.shp.xml	12/17/2020 2:19 PM	XML File	21 KB

If you double click this file, it will ask you how you would like to open it, we suggest using Notepad:

Once selecting Notepad, you will see something like this, but the information will be slightly

How do you want to open this file?

Adobe Acrobat DC

Firefox

Internet Explorer

Notepad

Paint

RStudio

VLC media player

Windows Media Player

OK
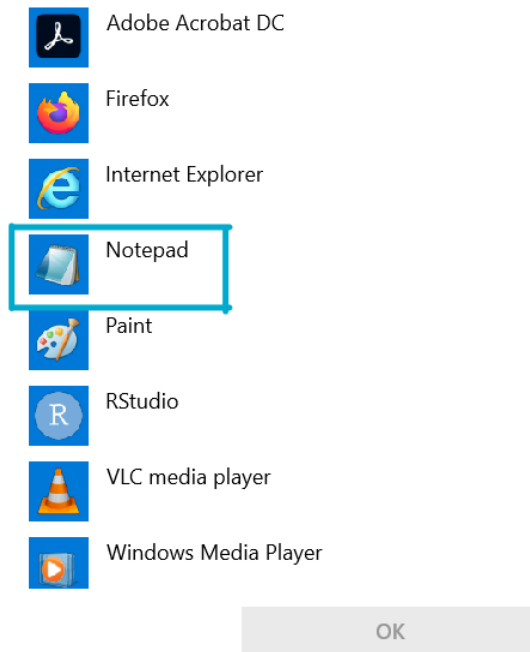
different, depending on what you initially entered:
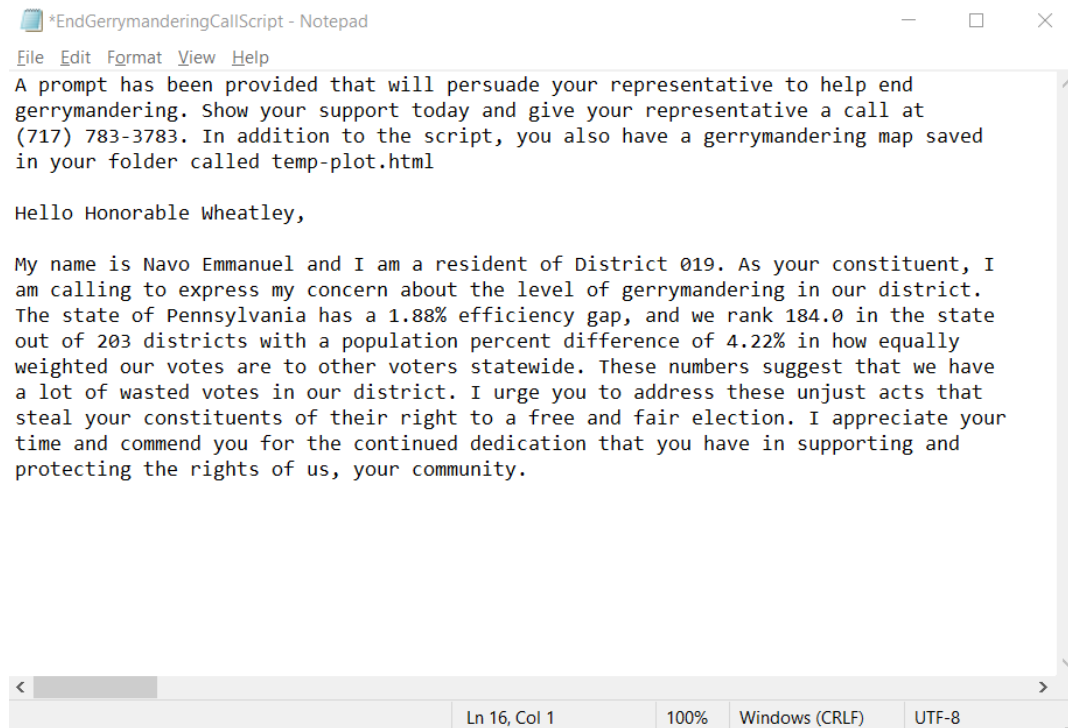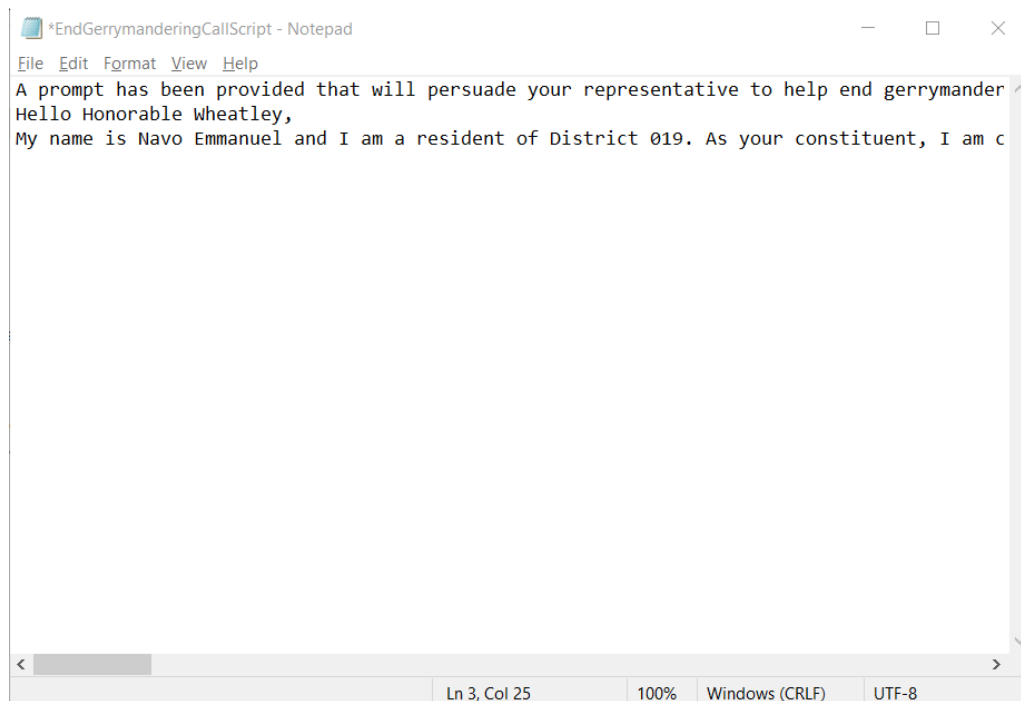
*EndGerrymanderingCallScript - Notepad

File Edit Format View Help

A prompt has been provided that will persuade your representative to help end
gerrymandering. Show your support today and give your representative a call at
(717) 783-3783. In addition to the script, you also have a gerrymandering map saved
in your folder called temp-plot.html

Hello Honorable Wheatley,

My name is Navo Emmanuel and I am a resident of District 019. As your constituent, I
am calling to express my concern about the level of gerrymandering in our district.
The state of Pennsylvania has a 1.88% efficiency gap, and we rank 184.0 in the state
out of 203 districts with a population percent difference of 4.22% in how equally
weighted our votes are to other voters statewide. These numbers suggest that we have
a lot of wasted votes in our district. I urge you to address these unjust acts that
steal your constituents of their right to a free and fair election. I appreciate your
time and commend you for the continued dedication that you have in supporting and
protecting the rights of us, your community.

Ln 16, Col 1   100%   Windows (CRLF)   UTF-8

For your own Notepad file, you will actually have three lines that stretch far into the Notepad that is outside of view. It will look more like this:

*EndGerrymanderingCallScript - Notepad

File Edit Format View Help

A prompt has been provided that will persuade your representative to help end gerrymander
Hello Honorable Wheatley,
My name is Navo Emmanuel and I am a resident of District 019. As your constituent, I am c

Ln 3, Col 25   100%   Windows (CRLF)   UTF-8

However, for the sake of this walkthrough, I have decided to manually enable the "Word Wrap" setting under "Format" for the clarity of this walkthrough.

In addition to the text file that provides you information on gerrymandering in the state of Pennsylvania overall and specifically in your district, it also contains your lower house state legislator's name and contact information along with a script you can use while calling them.

Lastly, you will see another file in your folder named temp-plot.html (this should also have popped as a new window). This file contains a choropleth map showing the level of gerrymandering in each district in your state, including yours. If the file doesn't download, this means that there was an issue with the ogr library. By hovering over different districts, you will see their respective information. It will look like this:
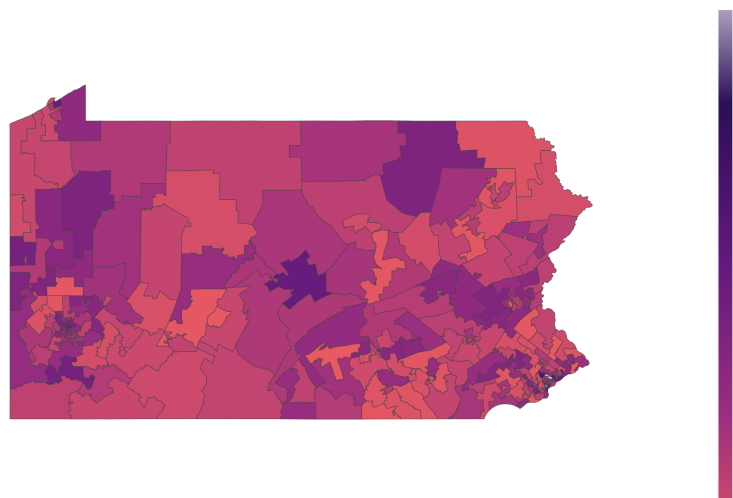
ERRORS AND CONCLUSION

Errors

While the Gerrymandering Project has quite a bit of functionality to it, there are certainly flaws to note. To run effectively, the program requires that you enter a valid Pennsylvania address, otherwise the program will not work effectively. The reason for this is two-fold: first, the open states website being accessed by Selenium has its own search function that requires that the address being entered is a valid address entry. If it isn't, then the GUI interface will restart and the user will have to input valid information and manually close out of the failed selenium window. Another area of concern is that if the user inputs a valid address as defined by open states, the program will still work so long as the congressional district that is scraped corresponds with one of the ones in our API data frame, but the output will be meaningless in this context. This is because the congressional district information in our database possesses Pennsylvania-based congressional districts.

To address the limitations of the program, we incorporated the following:

- The user is required to enter values for all textboxes
- The user is required to choose a file path
- The last character in the first name, last name, city, and state text boxes cannot be a number
- The user is required to enter a Pennsylvania zipcode
- If the user enters an invalid address as defined by open states, the program restarts
- If the user closes the selenium web browser, the program restarts

In the event that the previous conditions are met and the program still crashes, use a new kernel and re-run the program to get it back up and running.

Conclusion

Our program was developed to help residents of Pennsylvania enhance their knowledge of gerrymandering in state-level elections, and to give them tools to contact their representatives and advocate for equal, fair elections. With the 2020 Census completed, and subsequent redistricting of Pennsylvania districts approaching, our work is especially relevant. Our team is excited about the future applications of our program.

Please reach out to any of our team members if you have any additional questions:

Navo Emmanuel:nemmanue@andrew.cmu.edu
Charet Bolton: crbolton@andrew.cmu.edu
Morgan McCarthy: mamccart@andrew.cmu.edu
Jay Aswani: aaswani@andrew.cmu.edu