CMPS 161 Program 2 Writeup

User Documentation:

You will receive a directory containing an html file (“driver.html”) and several subdirectories (“/csv”, “/js”, “/json”).

If you do not already have this directory, you can retrieve it at:

https://people.ucsc.edu/~ydchoe/CMPS%20161/prog2/ElectionVisualization/

Due to a security protocol most web browsers implement, running “driver.html” on its own will cause a Cross Origin Resource Sharing (CORS) error. This is a result of D3.js (Data Driven Documents) using asynchronous requests in order to parse through JSON and CSV files. To work around this, you must run the program on a web server. This can be done by running a simple Python HTTP server within the directory. Open the terminal and cd into the directory and type either "$python -m SimpleHTTPServer" or "$python3 -m http.server". Then using a web browser of your choice, type in “http://127.0.0.1:8000/driver.html”. If you’re unable host a local web server you can also visit “https://people.ucsc.edu/~ydchoe/CMPS%20161/prog2/ElectionVisualization/driver.html” to find already hosted version of this program. Please note that internet connection is required in order to use the D3.js library.
Counties in blue represent Democrat strongholds, while counties in red show Republican majority. Solid colors represent a Yes on Proposition 6, while opaque colors represent a No. Hovering your cursor over a particular county will color it green, give you the county’s name, and identify its relevant data on the parallel coordinate plot in fuschia.

**Technical Write-Up:**

**Hypothesis:**

The assignment requested for an “in-depth analysis” on the Californian 2018 midterm elections. I hypothesized that Republicans were more likely to vote No on Proposition 6: Repeal of the Road Repair and Accountability Act (a fuel tax) while Democrats were more likely to vote Yes. As for an in-depth analysis, I also hypothesised that counties with a younger population, a higher population density, a higher use of public transportation, and a higher average income would vote No on Proposition 6. Younger citizens, such as college students, may not have access to personal vehicles and therefore rely more on public transportation. Places with a higher population density are more likely to have better infrastructure for public transit. People living in counties with evaluated public transportation usage would probably care less about a fuel tax as the need for a car goes down. Lastly, counties with a lower median income would most likely vote Yes to repeal the fuel tax in order to cut down spending.

**Results:**

(\textbf{Image to the Left: Red represent Republican counties and Blue represent Democratic counties. Solid colors represent a majority Yes vote, while opaque colors represent a majority No vote.})

It appears that my hypothesis about Democrats and Republicans was correct. The map to the left clearly shows that Proposition 6 was a partisan issue. This makes sense considering the ideologies of each party. Democrats often tend to vote for more taxes while Republicans tend to vote for less taxes.

(\textbf{Image to the Right: Visualizes data between Public Transportation Percentage (percentage of people who use public transit compared to other forms of transportation), Median Age, and Percentage of people who voted Yes for Proposition 6 in each county using a parallel coordinate plot})

My theory that counties with a younger population would be more likely to vote No on Proposition 6 was not supported by the data I gathered. It appears that there is no correlation between age and the Proposition 6 Ballot result. There also doesn’t seem to be a correlation between age and public transportation usage.
However, there seems to be some truth with my hypothesis about population density, public transportation, and median household income. My parallel coordinate plot suggests that there is some inverse correlation between voting Yes for Proposition 6 and the Median Household Income for each county. As a county’s median income increases, the likelihood of it voting Yes decreases. There also appears to some inverse correlation between Public Transportation Percentage (percentage of people who use public transit compared to other forms of transportation) and voting Yes for Proposition 6, although not as obvious as the previous example. Most counties seem to have a Public Transportation Percentage of 3.5% or less. These areas seem more eager to repeal the fuel tax. However, counties with more public commuters seem less eager to repeal the fuel tax. Lastly, there is somewhat of a linear correlation between Population Density and Public Transportation Percentage. Therefore it is probably safe to say that there is also some inverse correlation between Population Density and voting Yes for Proposition 6.

The two counties the best represented my hypothesis were San Francisco County and Lassen County. Their datasets are polar opposites of each other (except for age).
Program Approach:

I decided to use HTML, JavaScript, and D3.js (Data Driven Documents) to represent my data. JavaScript provided most of the data structure needed for the program, while D3.js provided most of the interactive data visualization, such as the county map and parallel coordinate plot. This method was the most efficient because much of the source code used in my program was readily available elsewhere, such as the geoJSON needed to create the map of California. Much of the data came in the form of a Comma Separated Value (CSV) file, which D3.js can handle with ease.

Sources for the Data:

Much of the data structure for the program was just combining different datasets from various sources and storing it into the geoJSON. Party majority for each county and the 2018 ballot results came from the California Secretary of State website (www.sos.ca.gov). Data such as median age, population density, public transportation, and median income came from the United States Census Bureau at (www.factfinder.census.gov).

Visualization Choices:

Using a color coded county map was probably the best way of showing the partisan divide in California. We can clearly see that the coastal counties tend to be more liberal while the inner counties tend to be more conservative. Using opacity to see the ballot result further strengthens the idea that the Two Party System is more alive than ever in America, even in a blue state such California. One major drawback with this method is that it is difficult to tell which counties are swing counties. We could have used opacity to instead show how much more control a party has over a county percentage wise. However, it would then be more difficult to show the ballot result in effective manner.

Using a parallel coordinate plot allows for a simple visualization of a multivariate dataset. It is easy to see correlations between an axis and its neighbors. However, it does come with disadvantages. The biggest of which is that it is difficult to see patterns with a neighbor of a neighbor of an axis. Therefore, many reorderings of the axes are needed to find relevant correlations. My code currently does not allow users to dynamically move the axes freely on the webpage. This must be done in code. Another issue with parallel coordinates is that as the dataset grows the messier it looks.