Computer Science w/ Applications II Project Proposal

Project members: Christopher Summers, Yang Hou, Huanye Liu

For our project we hope to analyze the effect of unemployment and demographics on the election results of 2016. Specifically, there were two narratives in the aftermath of the 2016 election that we wish to analyze. The first was that white, male voters swung the election towards Donald Trump due to gender or racial animosity. The second was that voters who had manufacturing jobs created the swing due to economic anxiety over the decline of U.S. manufacturing. These factors are connected; many manufacturing jobs were predominantly white and male. Thus in order to determine the relative influence of each factor we will need to control for the other variables.

The use of job type in our analysis presents the difficulty that voter occupation is not something typically collected in voting surveys. So instead we intend to do our analysis on the county/district level. The data for county voting results in the past election can be found either from a [MetroCosm dataset](#), which can be found on [Kaggle](#), or from the [US Election Atlas](#), which has been cited in publications such as The New York Times and The Economist. The US Election Atlas data will require us to build a crawler to scrape the data from the website.

Data on the demographics of these counties also comes from a [Kaggle dataset](#) on primary results. And finally we will be using the [Equal Employment Opportunity data](#) from the United States Census, which gives data on the number of workers in different occupations and their demographics, split up by various county regions. Unemployment data will come from the Bureau of Labor Statistics [data on unemployment by occupation](#).

Finally, a further variable to consider if there is time is the relative strength of unions and how that may further affect the correlations between manufacturing jobs, unemployment, and voting. While we are currently unsure of how to determine data about how strong unions are, it would likely be based on what laws have been passed in various states such as those restricting collective bargaining rights or various "right-to-work" laws.

Modules

This project will require 3 modules:

1. **Retrieving Data**: several of the datasets are simply available for download, but the US Election Atlas will require a website crawler to be built to retrieve the election results from each county. The US Census data also has an API that we will use to retrieve data on the given counties.

2. **Data Cleaning/Munging**: We will build a module that checks the data we retrieve for missing data and potentially combine datasets if we can to fill the holes. Furthermore, the county variables are not consistent across different datasets; the EEO county definitions combine multiple smaller counties into one, which will require us to create a module that retrieves the EEO list of counties and reformats the other datasets to use those combined county variables. Thankfully, the BLS and the EEO use the same codes for identifying occupations, though some reformatting may still be necessary.

3. **Regression Module**: This module will perform the regressions on the data that we get. Specifically it will determine whether voting results for each county/combined county group are best predicted by the demographics of that county, by the occupations of that county, or by the unemployment by occupation of those counties.

Timeline:

Our raw timeline is listed below:

Week 5: determine all the deliverables for different modules. In other words, in what ways and with which variables or files will different modules communicate with each others.

Week 6: complete 80% for each module, attending the project check-in with any questions arised. Determine if we have time for labor union module.

Week 7: most likely on the weekend, finish first draft of all modules.

Week 8: try to connect all modules together, and fix any bugs arises,attend project check-in.

Week 9: do any necessary improvement to the project.

Week 10: present and submit the project.