Project 1

CS 3300

Evan Zhao (etz3), Justin Kuang (zk65), Anya Chopra (ac948)

Hosted Link:
https://ap-ib-offerings-visualization.herokuapp.com/index.html

Our overall visualization seeks to map the percentage of public high schools in each state that offer AP and/or IB programs, and also show correlation between said percentages and political leanings (voting percentages in the 2016 presidential election).

Our specific data consisted of several different sources that were integrated together.. Data on the number of high schools that offer APs came directly from College Board (https://secure-media.collegeboard.org/digitalServices/pdf/research/2017/School-Report-Exams-2017.xls)- we deleted non-essential information, and also converted the file to .csv for easier parsing. Similarly, our IB data came straight from the IB website (http://www.ibo.org/programmes/find-an-ib-school), but there was no dataset immediately provided. We manually searched for the number of public high schools in each state that offered the IB diploma, and entered such data into the AP information .csv file as a separate column. Finally, we gathered information on the total number of high schools in each state from the National Center for Education Statistics (https://nces.ed.gov/programs/digest/d17/tables/xls/tabn216.80.xls). This specific dataset contained a lot of information that we didn't require, so we stripped it down to a single column (total number of secondary schools in each state), and also converted to .csv for easier parsing.

For our political plot, we gathered data from Kaggle (https://www.kaggle.com/stevepalley/2016uspresidentialvotebycounty/data). We cleaned the data to only use percentages for each state, even though we were presented with an overwhelming amount by county.

On the mapping side, we used the same USA .json map as the one provided in lecture, and we relied on the topojson library to graph the map. Since the USA .json file didn't have any direct correlation (fields) that mapped paths to states, we found a third party dataset (https://github.com/GovLab/opencorporatesd3/blob/master/us-state-names.tsv) that relates state identification numbers and the state's location on the map. By

integrating these two sources, we figured out how to correctly link data (in our case, the percentages) to their respective states.

The overall variables we used are the state identification number, the state name, the number of AP programs, the number of IB programs, the total number of AP and IB programs and the average state income.

The integration of these data sources was primarily mathematical- we calculated a percentage of each state's high schools that offered AP/IB schools using the simple formula (# of high schools that offer AP) divided by (# of total high schools in the state). On the maps, the darker colored states have the higher percentage schools that offer said programs. To create the maps, we used linear scales that transformed the direct percentage of schools to an opacity percentage from 0.00 to 1.00.

We decided to extend the analysis of education in each state and attempt to search for correlations in the recent 2016 elections. With a scatterplot, we directly graphed the percentages of schools in each state that offer AP curriculum against the percentage of people in each state who voted for Donald Trump. Note that Independent Parties were cleaned out from the dataset. Using D3 methods, I made states that eventually voted for Trump red, and states that eventually voted for Clinton blue. To show the overall trend of the graph, the least squares regression was used to compute the slope of the regression line -- I simply used d3.line() to graph against this regression line. A legend was added onto this graph using D3 libraries.

We expected to see a correlation in our mappings towards coastal states (that are anecdotally wealthier)- this correlation was mostly confirmed. Interestingly enough, IB programs only seem to be emphasized by certain states. On the political side, we expected to see a correlation possibly between higher AP offering and Democratic leaning, but we found only a very weak correlation between the two -- it seems like the quality of education as measured by AP curriculum in each state is relatively apolitical.