CTW Project Reflection

Casey Alvarado and Jay Woo

Assessment evidence and interpretation

Our original project proposal stated that our MVP was making a static map image, coloring the states according to which ones had the most instances of sexual assault. Our stretch goal was an interactive map with d3. However, after doing a couple of brief mockups with various Python packages (iPython notebook), we realized that the map didn't really fit with our storytelling purposes.

So, we decided to look at other datasets to see if we could find any notable correlations. In our iPython notebook data_cleaning_exploration.ipynb, we noticed a fairly high, positive trend between the percentage of Roman Catholics per state and the number of accused priests in each state. The correlation was 0.46 for the whole dataset, but for states with populations larger than 3,000,000, the correlation went up to 0.59. We decided to focus on this scatterplot and structured our story around it. In that notebook, we also tried out a bunch of different methods of cleaning the data, and as a result, we learned a lot about how to use pandas effectively.

In the end, we ended up making a semi-interactive story using D3 (found here) that goes through a bunch of visualizations that we made to describe the data. Our main focus was to ensure that the visualizations were presented well, and we would like to be assessed on not only the technical aspect of the project but also whether our story was told effectively.

Changing the world

Though our visualization could use a lot of extra polishing, we think our project could change the world. There is currently a lot of discussion about this issue. For instance, The New York Times has written at least 16 articles on this topic since the beginning of 2016. However, there only seem to be articles about one-time occurrences. By visualizing all of the data from 1950 to 2002, we think that our approach is different from what most people have already done.

We do think that our project could have been improved in a lot of ways, mainly in the way that we cleaned the data. Many of the entries in the dataset were misspelled or did not follow consistent naming conventions, and as a result, we decided not to look at certain features of the data (i.e., the "Source" column). If we had worked a bit more on the data exploration side, I think we could have found more compelling pieces of information to add to our visualization.

Learning goals

We learned a lot of what we set out to learn at the beginning of the project. After doing our final iteration, we feel like we are much more comfortable using tools like D3 and Pandas to analyze our dataset. Though we had a difficult time integrating all of the individual pieces, we ultimately got everything to work as intended. We also learned a lot about designing good visualizations and telling powerful narratives. Although we don't feel like we've mastered these skills, we feel more confident in our ability to communicate our findings.