

# Final Project Data Memo

## Data Visualization (STAT 320)

Apoorva Shivaram

2/20/2021

## Contents

### Data Source

*Identify and describe your data source or sources – citations would be appropriate. Citations will typically include links to the data source or describe your plan for obtaining the data.*

I will be working with the immigration legislation dataset that was posted on Canvas. Citation:

Filindra, Alexandra and Shanna Pearson-Merkowitz, (2020) Database of State-Level Immigration Legislation, 1990-2015. Chicago: University of Illinois

### Why this Data

*The reason for your choice of data. Is it interesting to you? Challenging? Is this part of a larger research project? State any questions of interest that might be motivating your choice of data. Include any other motivations or background that you believe to be relevant.*

I think immigration legislation is extremely interesting, relevant, and has an impact on so many people's lives. As an international student, I see how (positive or negative) immigration legislation can make such a big difference and I'm curious to explore past trends in this dataset. I'm also hoping to expand my data science portfolio, so this will be a neat addition to my skill set. Taking a peek at the data, it looks like I might have to do a bit of cleaning/wrangling which is always an exciting prospect.

### Potential Data Issues

*It is important that you think about possible issues and determine if you have a plan to handle them - this might entail budgeting time to get help with this.*

*Identify any potential issues you may face in preparing or analyzing your data, including any merging, cleaning, or processing that may be necessary. Consider anything that may take extra time that is unique to your dataset.*

I will have to inspect the data to make sure it has been coded correctly. There might be a fair bit of data cleaning/wrangling to be done with this dataset, so I'll have to allocate a bit of time for that. I'll have to convert some or all of the numeric variables into factors with labels for each levels so that the visualizations are more meaningful and straightforward.

The data set is well documented with a code book and the questionnaire, so identifying the variables/values shouldn't be a problem. There's also no missing-ness, so that's a good sign!

## Final Project Format

Ideally, I'd like to build a shiny app with a combination of static and interactive graphics.

Some ideas for questions to visualize:

1. `year_introduced` and `enacted`: Trends in whether the bill was enacted across time?
2. `bill_type` and `enacted`: is a certain type of bill enacted more often than others?
3. `state` and `enacted`: do bills introduced in certain states have a higher probability of being enacted than others?
4. What group of people do the bills generally target? how has that varied across time?
5. What sector does the bill engage? How has that changed over time?