Final Project Proposal - S&DS 625 Statistical Case Studies

Megan Ayers

1 November 2021

For my final project I would like to explore the following question: Does living in an American county that is frequently affected by extreme weather events increase the likelihood that residents believe in climate change (or in the necessity of acting to mitigate it)? To investigate this kind of relationship between climate change beliefs and extreme weather, I plan to use data collected and prepared by Dr. Jennifer Marlon (YSE) and her team at the Yale Program on Climate Change Communication. This data was sourced from a "large national survey dataset (n > 25,000)" (webpage here), collected as part of the Climate Change in the American Mind project. It contains county-level estimates of climate change opinions (i.e. 74% of adults in New Haven County believe that global warming is happening) with a somewhat wide range in question specificity. These estimates were produced using multilevel regression and poststratification to address issues of sparsity in the raw survey data. As a sidenote, I am meeting with Dr. Marlon on Wednesday Nov. 3 to discuss potential practical work next semester and will ask her if she has additional data at the individual level and/or over time that she would be willing to share with me to potentially work into this project.

In addition to this climate change opinion dataset, I would like to collect extreme weather event data from NOAA (user interface for the database here) as well as information about county demographics such as age, race, and religion distributions from the Census and an election returns dataset from the Harvard Dataverse (here) to better understand political affiliations. I expect the election and Census data preparation process to be fairly cut-and-dried, and I have had practice working with both in the past on the county and congressional district levels. The NOAA data may require some more decisions to be made, such as if I want to look at a subset of the most extreme, enduring/repetitive weather events, and it also has a bit of an additional challenge of sometimes including weather "zones" instead of just counties. But because NOAA provides detailed information about the number of deaths, injuries, property damage, dates, and locations of these events, it seems like I will have the tools I need to make these modeling decisions as well as to merge all the data sources by county.

Once all this data is collected and combined, I am planning to do some initial data exploration to gut check some of my prior expectations, for example that democrat-leaning counties have higher estimates of belief in climate change compared to republican counties, and that there are NOAA records for hurricane events in the southeast and wildfire/drought events in states like California and Oregon, etc. Then I will model how these different factors predict county climate change belief estimates, probably considering multiple individual models where the focus for each will be on different regions and the weather events more specific to that region.