

Problem Set 5

Will Miller

March 3, 2025

1 Static Webpage Scrape

I chose to scrape data from the NOAA Storm Events Database. This database records all reported storms for every county in the United States starting in the early 1950s. One idea I had for the project in this class was to see if I could find evidence for repeated hail storms in the OKC metro leading to higher insurance costs. I am not exactly sure how I would go about quantifying this because most severe weather damage pass-through studies I have found focused primarily on large, infrequent storms with damage that is (relatively) easy to observe and quantify. Hail storms are so common and so ubiquitous - with most storms only doing very localized damage - that I think this project might not be doable with the data I have available to me.

Nonetheless, it was still quite interesting to see the data on hail storms. I gathered data on every hail storm in Cleveland County from January 1st, 2010 through the last reported hail storm in the county (November 3rd, 2024 in Stella, Oklahoma) using the `rvest` package in R. The CSS selector was very easy to locate for this webpage because it had me request the specific data I wanted before it would display a webpage and then produced some tidy tabular data with a clear CSS selector (identified using `SelectorGadget`). I looked to see if they had an API but it looks like they are in between systems with a new API coming soon. The NOAA storm database provides information on the date and time of the reported storm, the magnitude of the storm (in this case the diameter of the hailstones), the estimated damages, and the reported injuries and fatalities.

2 API Data

For this section of the problem set, I chose to get data from the FRED API on global Gini Indexes over time. This was inspired by a talk Dr. Burge gave at the Economics Club last Thursday on the economics of taxing the rich. He had some slides about rising income inequality in the US and I wanted to see if that was true globally as well.

I used the `fredr` package in R to gather this data using my API key that we set up in class. I started by getting data for just the US which was very easy. My next goal was to gather data for every country series they had (158 in total). It was clear that I couldn't select all of these manually, so used the text search function in the `fredr` package to get a list of the series name that corresponded to every country and then I defined a function that would iterate through that list and ping the FRED API individually for each series. That function returned a list, which I then converted to a panel dataframe before doing some minor cleaning.

I am not totally sure that I'll use this data for research purposes in the future, but I have recently become quite interested in the `plotly` package in R for making animated and interactive visualizations. I am currently working on creating an animated world map that shows the progression of each country's gini index over time. Ideally, it would be animated in a way that the end user could both watch the progression over time automatically and select specific years to interact with using some sort of slider at the bottom. I hope to have this graph ready to turn in for problem set 6.