Nathan Bush

Data Analytics Term Project

Data and Topic

**Research Questions**

1. Have wild salmon populations in the Snake River declined since dams were built in the 1970s?
2. Can the increased number of visitors to National Parks and Outdoor recreation areas be linked to outdoor oriented social media posts?
3. **Is the amount of renewable energy a state uses correlated with its political views?**
4. Does proximity to a substantive body of water (large lake or river) impact housing prices?

Initially, these were the four research questions I came up with when contemplating a topic for this assignment. I was initially very interested in the first question: whether or not damming the snake river has impacted fish populations. But the data that I did find was very basic. Just a count of fish populations from 1940 – 2018 and I figured that might be a bit too easy to work with given a large part of this assignment seems to be cleaning and organizing the data we are working with. So, I settled on a comparison of the amount of renewable energy a state uses and its political leanings.

**Why it Matters**

It is important to understand whether or not political views effect the amount of renewable energy states are using for many reasons. First, it effects policy. If we find that conservative leaning states tend to use less renewable energy compared to their liberal leaning peers, then it is important to incentive these states to a greater degree when making a policy intended to increase the amount of renewable energy production. Second, it could indicate that states are not acting in their best interests if the amount of renewable energy produced in their state is not in line with its potential. For example, a state like Utah, with an abundance of sunshine, rivers, and wind, should be taking advantage of these opportunities to produce renewable energy in comparison to states like South Dakota with a relative lack of the same resources. If the conservative values of a state get in the way of opportunities like this, this could be considered irrational and not in the state’s objective best interest. Lastly, as climate change continues to increasingly effect energy infrastructure and the earth’s natural resources, it is important to understand the dynamics behind energy systems and how politics plays a role in these systems

**Data**

I have three datasets that I will be using. Two .csv files I have found in my research and one table that I will be web scraping.

My first dataset is data on energy usage statistics on a wide variety of metrics in each state. This is taken from the US Energy Information Administration. The kay variables here are the percentages of petroleum, natural gas, coal, and renewable energy as a percentage of the total energy use in the state, the capacity of each state to use renewable energy, and the emissions of each state. This dataset contains information from 2019-2022.

[*https://www.eia.gov/state/compare/?sid=AL#?selected=US-AL-AK-AZ-AR-CA-CO-CT-DE-DC-FL-GA-HI-ID-IL-IN-IA-KS-KY-LA-ME-MD-MA-MI-MN-MS-MO-MT-NE-NV-NH-NJ-NM-NY-NC-ND-OH-OK-OR-PA-RI-SC-SD-TN-TX-UT-VT-VA-WA-WV-WI-WY*](https://www.eia.gov/state/compare/?sid=AL#?selected=US-AL-AK-AZ-AR-CA-CO-CT-DE-DC-FL-GA-HI-ID-IL-IN-IA-KS-KY-LA-ME-MD-MA-MI-MN-MS-MO-MT-NE-NV-NH-NJ-NM-NY-NC-ND-OH-OK-OR-PA-RI-SC-SD-TN-TX-UT-VT-VA-WA-WV-WI-WY)

My second dataset is county voting data on the 2020 presidential election taken from the MIT Election and Data Science lab and the Harvard Dataverse. The variables here are State, County, Candidate, # of votes in each county, total votes. This dataset contains results from 2000-2020.

[*https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ*](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ)

In addition, I will be webscraping a table from Ballotpedia in order to compare which party a given states governor is from with who they voted for in the presidential election. My hypothesis is that a state who voted for a republican in the 2020 election and has a republican governor is more conservative than a state that has a democratic governor but voted republican in the 2020 election. I will be able to test whether or not there is any difference in energy usage based on these observed differences by adding in the governor variable.

[*https://ballotpedia.org/Partisan\_composition\_of\_governors*](https://ballotpedia.org/Partisan_composition_of_governors)

These datasets can be joined by state and ultimately compared across states.