An Analysis of Energy Production Data

By: Andrew Kang & Caitlin Cavanaugh

Key Questions:

* Which states have produced the most and the least CO2 emissions in lbs per MW hour at the start year (2008) and end year (2018)?
* How do trends differ between **CO2 Emissions (lbs per MW Hour)** and **Total CO2 Emissions (thousand metric tons)**?
* How have energy costs changed from start year (2008) to end year (2018)? Have they increased, decreased, or stayed consistent?
* What are some factors that might make energy costs more expensive for different states and regions?

Data sourced from https://www.eia.gov/opendata/browser/electricity/state-electricity-profiles

The data source for this project is from the US Energy Information Administration (EIA). The EIA is a part of the US Department of Energy and is dedicated to the collection and synthesis of information regarding the production, usage, and cost of energy within the United States. Using this data we found a set of states at certain statistical boundaries (max, min, median, upper boundary) regarding three key statistics:

* Average Price (Cents per KW Hour)
* CO2 Emissions (lbs per MW Hour)
* Total CO2 Emissions (thousand metric tons)

From there we further analyzed the selected states by:

* Production capability by energy source in 2008 and 2018
* The proportion of each energy source 2008 and 2018
* The change in the above data points from 2008 to 2018

Utilizing a t-test we found that between 2008 and 2018 there was likely a statistically significant change in CO2 Emissions (lbs per MW Hour), but there likely wasn’t a statistically significant change in Average Price (Cents per KW Hour). In addition in the states that we selected to further analyze we found that they all, barring Vermont, decreased usage of fossil fuels. Vermont is an exception, shutting down a nuclear power plant instead of decreasing fossil fuel usage.

Data

Variables in original data source - State Rankings for Key Statistics

* State
* Year
* Average Price (Cents per KW Hour)
* CO2 Emissions (lbs per MW Hour)
* Total CO2 Emissions (thousand metric tons)

Variables in original data source - Generating Capacity

* Production Capability (MW)
* State
* Year
* Energy Source
  + Source of energy by various types and subtypes
    - The Production Capability (MW) of the “All” energy source is an aggregate of all other energy sources for that state for that year
    - Energy sources with hyphens (-) are subtypes of a broader energy source - eg Natural Gas - CC vs Natural Gas. The unhyphenated energy sources include all subtypes
* Producer Type
  + Dropped, not used in this analysis.

Cleanup:

**State Rankings for Key Statistics**

Merged 2008 and 2018 tables into one ungrouped. Used for boxplot and histographic analysis.

Merged 2008 and 2018 tables into one on the ‘State’ column. Renamed and reordered columns by relevant year.

Added the below columns by subtracting the 2018 values from the 2008 columns for our key statistics:

* Change in Average Price (Cents per KW Hour)
* Change in CO2 Emissions(lbs per MW Hour)
* Change in Total CO2 Emissions (thousand metric tons)

Added the below columns by dividing the 2018 values from the 2008  key statistics:

* Percent Change in Average Price (Cents per KW Hour)
* Percent Change in CO2 Emissions(lbs per MW Hour)
* Percent Change in Total CO2 Emissions (thousand metric tons)

**Generating Capacity**

Removed all rows with hyphens to leave the major energy source categories for 2008 and 2018. Removed all rows with the energy source “All” to leave only data in major categories for 2008 and 2018.

Added Proportion of Production column to each year by using the transform function on each year’s Production Capability column grouped by state .

Added the below columns by subtracting respective 2018 columns from 2008 columns from each other

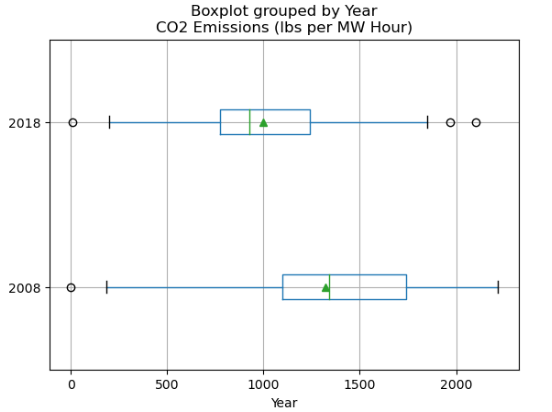
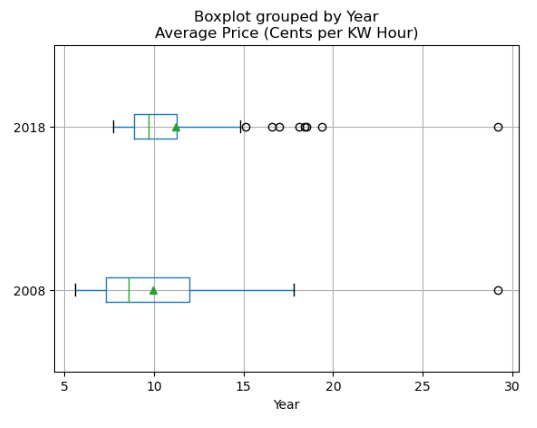
* Change in Production Capability (MW)
* Percent Change in Proportion

Analysis & Graphs

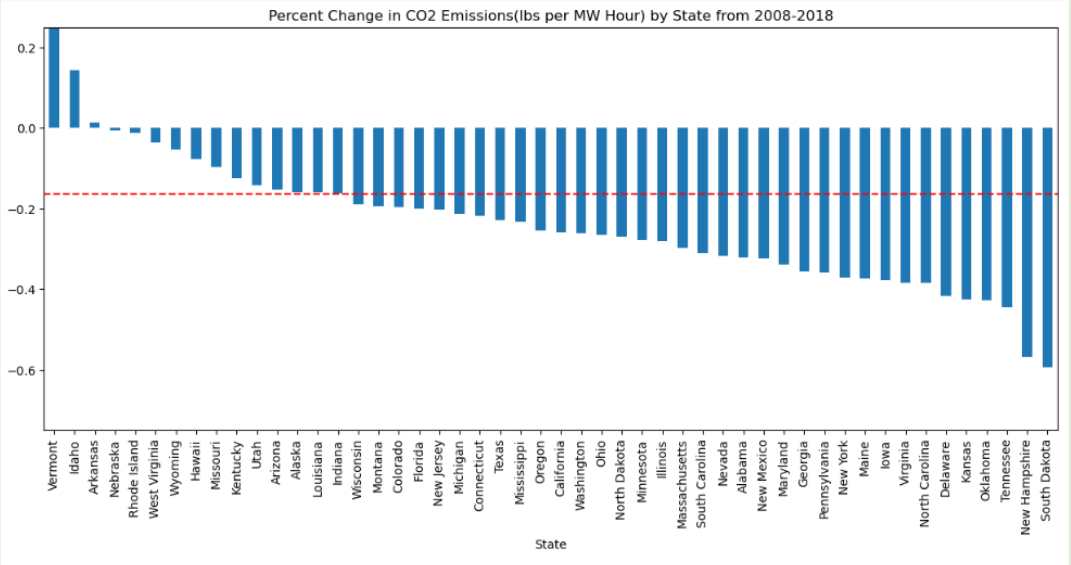
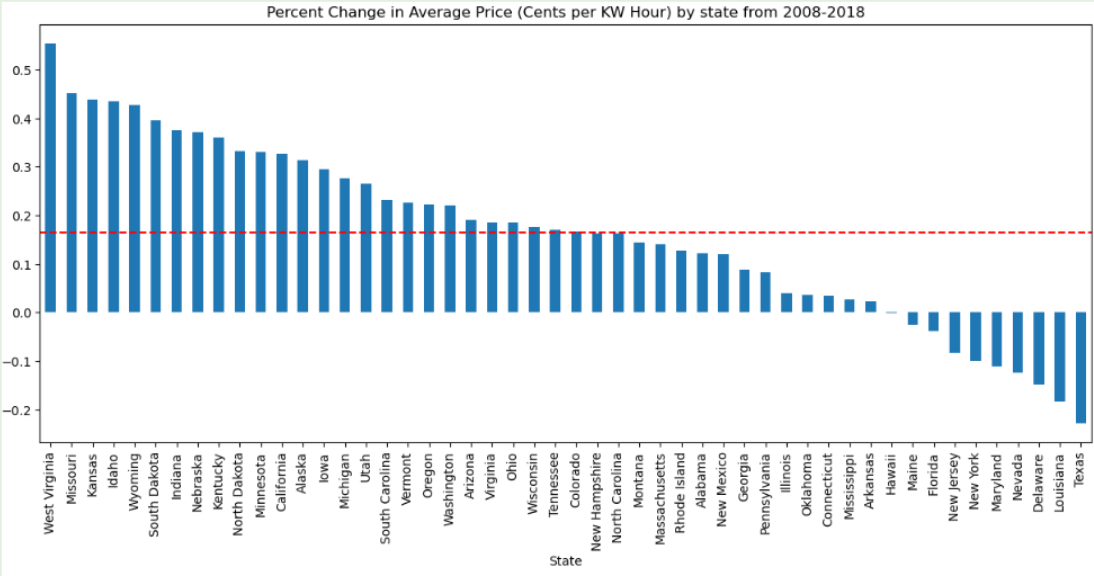
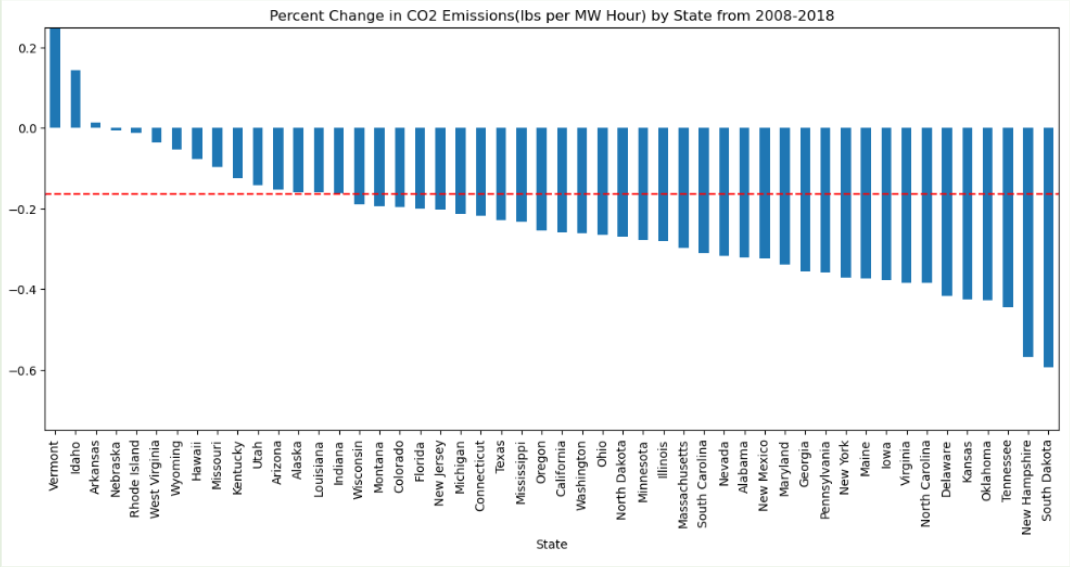
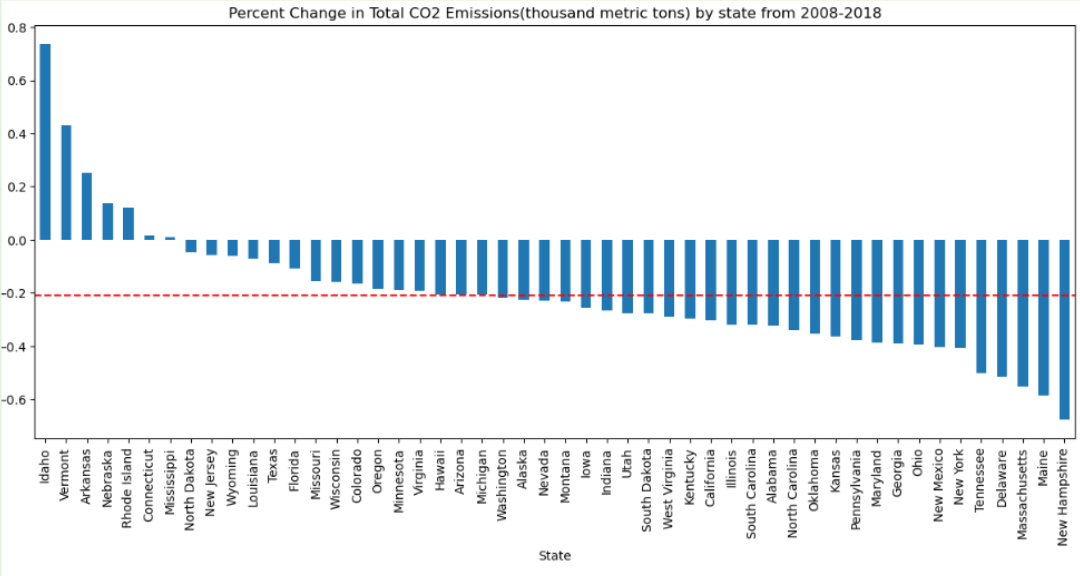
Initial analysis was done on the State Rankings for Key Statistics dataset via t-test to see if there was a statistically significant change in values between the means in 2008 and 2018. Results are as follows:

* CO2 Emissions (lbs per MW Hour):
  + p-value of .002 indicating it is likely that there was a statistically significant change between 2008 and 2018
* Average Price (Cents per KW Hour):
  + p-value of .13 indicating it is unlikely that there was a statistically significant change between 2008 and 2018
* Total CO2 Emissions (thousand metric tons):
  + p-value of .15 indicating it is unlikely that there was a statistically significant change between 2008 and 2018

The boxplots produced for Average Price and CO2 Emissions support the conclusion of the t-test, as we see a notable movement in the mean and median in the CO2 Emissions in lbs per MW Hour, but the Average and median price largely stay the same between the periods.

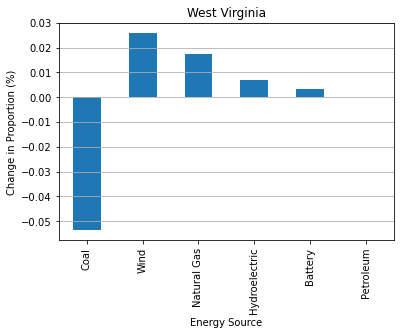
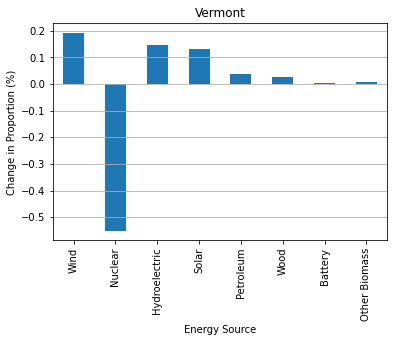
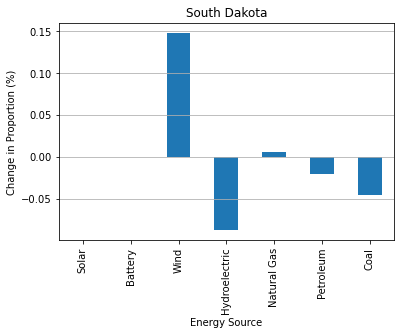
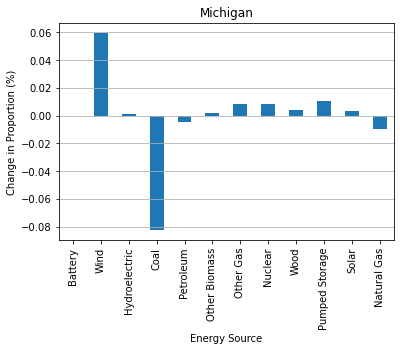
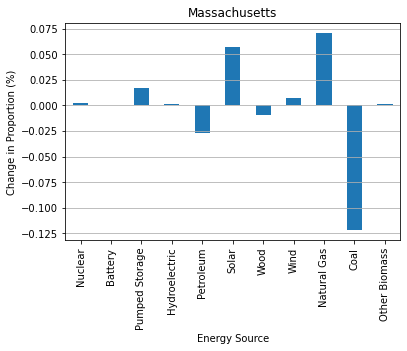
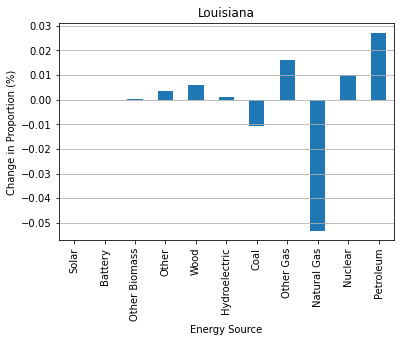
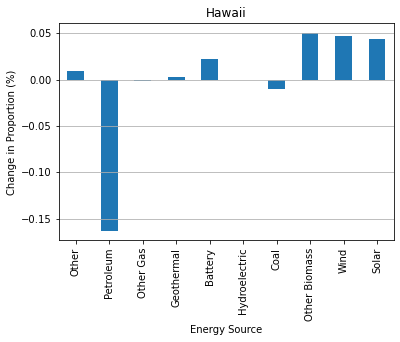
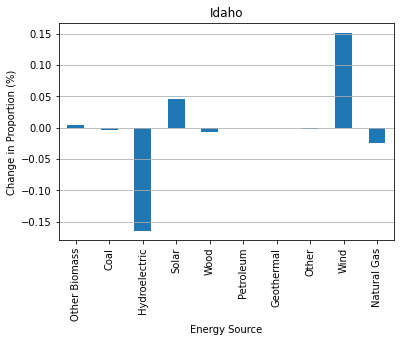


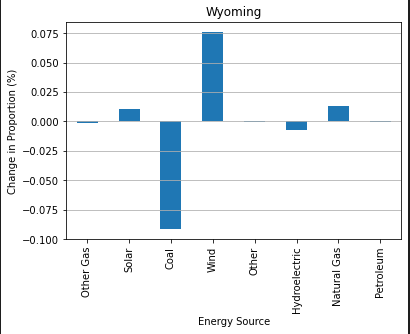
The following bar charts by percent change support the box plots, but do not seem to align with the conclusion of the t-test, as on average the change in values sits between 16% and  20% in absolute values.

States are then selected from the above from the above analysis based for the following reasons:

* Hawaii: The state with the most expensive electricity in 2008 and 2018 to the point of being a possible outlier (29.20 & 29.18 respectively)
* Louisiana: The state with the least expensive electricity in 2018 (7.71)
* Vermont: The state with the lowest lbs of CO2 emissions per MW Hour of energy produced in 2008 and 2018 (2 & 10 respectively)
* Wyoming: The state with the highest lbs of CO2 emissions per MW Hour of energy produced in 2008 and 2018 (2216 & 2100 respectively)
* Michigan: The state closest to the mean of all states in 2018 in average price per KW hour (11.40)
* Massachusetts: highest price at the upper bound of energy cost in 2008 (16.23)
* West Virginia: Greatest percent change in average price (+55.4%)
* South Dakota: Greatest percent change in CO2 Emissions per MW Hour (-59.4%)
  + Vermont excepted due to being an outlier and already being on the list
* Idaho: Greatest percent change in total CO2 Emissions (+73.6%)

And then analyzed for percent change in energy source and plotted onto bar charts:



Conclusion

* Which states have produced the most and the least CO2 emissions in lbs per MW hour at the start year (2008) and end year (2018)?
  + Vermont with the least at both time points.
  + Wyoming with the highest at both time points
* How do trends differ between **CO2 Emissions (lbs per MW Hour)** and **Total CO2 Emissions (thousand metric tons)**?
  + There are some states that seem to be related between these two statistics. For example New Hampshire and Idaho are similarly positioned on their respective percent change graphs.
  + However some states are not as strongly related. South Dakota with the largest percent decrease in CO2 emissions (lbs per MW hour) is close to the mean in the percent change in total CO2 emissions.
* How have energy costs changed from start year (2008) to end year (2018)? Have they increased, decreased, or stayed consistent?
  + The t-test of average price between the two time points indicates the change is not statistically significant with a p-value of p-value of .13, though we still find the average increase of 16% notable
* What are some factors that might make energy costs more expensive for different states and regions?
  + Inconclusive. Findings are cursory and require additional investigation, but initial examples seem to show maintaining fossil fuel usage to be related to lower energy costs
    - Hawaii dropped it’s usage of petroleum but essentially had no change in energy price
    - Massachusetts dropped its usage of coal and increased usage of natural gas and solar energy. Prices increased.
    - Louisiana dropped natural gas usage for other gas and petroleum usage and has the least expensive energy in 2018

Limitations and Further Exploration

* The sample size of states examined for their percent change was very small at only nine states.
* The criteria used to select the above states for further examination was not consistent