

Emissions and Energy Resource Relationships
By: Ashika Mulagada, Joanna Hoang, Sarah Schibel
Domain: Environment

Domain

Electric power is a dominant force in greenhouse gas emissions. In 2016, the electricity sector was considered the second largest contributor to greenhouse gas emissions in the US, contributing to 28.4% of greenhouse gas emissions. This sector includes the generation, transmission, and distribution of electricity. The emissions from electricity is mainly made up of carbon dioxide emissions, but also includes methane, nitrous oxide, and less than 1% of sulfur hexafluoride. This is largely due to coal combustion in power plants. In the US, CO₂ emissions due to electricity have decreased since 2005 by 28%. This is due to a slower growth in the demand for electricity and changes to electricity generation. Since 2005 there has been an increase in non-carbon generated electricity, and the use of natural gas plants has increased. Other trends in emissions are due to changes in economy, weather, and more. Currently, 28% of greenhouse gas emissions in the US comes from electric power generation which includes the combustion of coal or natural gas for electricity.

The United States Environmental Protection Agency is an independent government agency for environment protection. The EPA considers many different factors such as sampling methods, budgets, and regulatory requirements when collecting new data. The EPA's quality assurance system ensures that the data found is based on scientific principles, resources are justified, and activities are properly evaluated to reduce waste. The main purpose of the EPA's greenhouse gas data is to inform the public and policy makers about current emission rates.

URL of dataset:

<https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid>

Dataset Description

There are four different items of data that have been gathered for this specific dataset. These items include; Subregion Output Emission Rates, Subregion Resource Mix, State Emissions and Output Emission Rates. All four items correlate with one another upon the topic. Between both Subregion Output Emission Rates and State Emissions Rates, they both contain one piece of the same information which is the 'Total Output Emission Rates' which contains different environmental elements that affect the total output of emission rates. The difference between the two is that the 'Subregion Output Emission Rates' also has a category of 'Non-Baseload Output Emission Rates'. The other two items in the dataset, 'Subregion Resource Mix' and 'State Resource Mix', both contain the same information as well. With both items containing information on 'Generation Resource Mix', which is the different resources used throughout the different regions and states, and the 'Net Generation'.

The population of the data set is exclusive to only The United States and all 50 states as well as the subregions of The U.S.. The data set does not appear to be biased as it completely covers all regions of the U.S..

The data is fairly clean as it was gathered by the EPA which must meet government standards and requirements. All the columns and rows are properly labeled as well as have subdivisions amongst the different pieces of information. The only item not completely clean about the data set is a small note located at the bottom of the data set which states that "Percentages may not sum to 100 due to rounding." which could be an issue. We can clean up this part by rounding down instead of up to get a more precise amount.

Exploratory Data Analysis

This dataset was fairly large, containing 4 spreadsheets for each of the 4 topics listed in the above dataset description. Due to the wide array of variables, we focused on 2 variables to explore relationships in the dataset. These two variables were "Coal," (measured in percentage out of total resources), taken from the State Resource Mix section, and "CO2" (measured with units lb/MWh), taken from the State Output Emissions Rate page. The relationship between Coal and CO2 was chosen based on that coal is known to be a source of energy with negative environmental impacts, and our team wanted to see if official EPA statistics would back up that knowledge.

Descriptive statistics:

mean(state\$CO2) = 990.5 lb/MWh

mean(state\$Coal) = 30.07255%

Call:

lm(formula = CO2 ~ Coal, data = state)

Residuals:

Min	1Q	Median	3Q	Max
-440.99	-109.05	-7.58	118.59	776.86

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	497.891	45.565	10.93	9.78e-15 ***
Coal	16.381	1.138	14.39	< 2e-16 ***

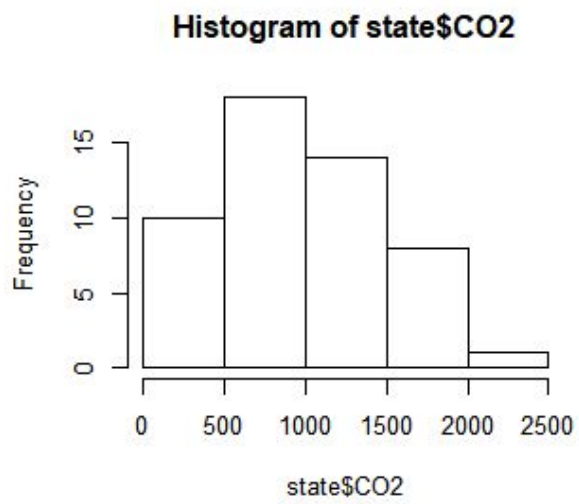
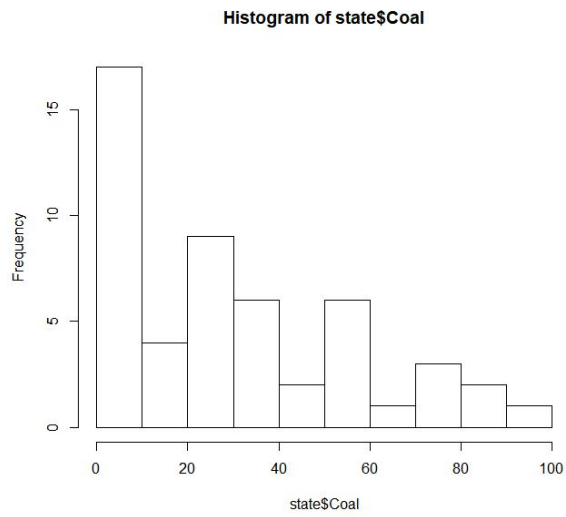
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

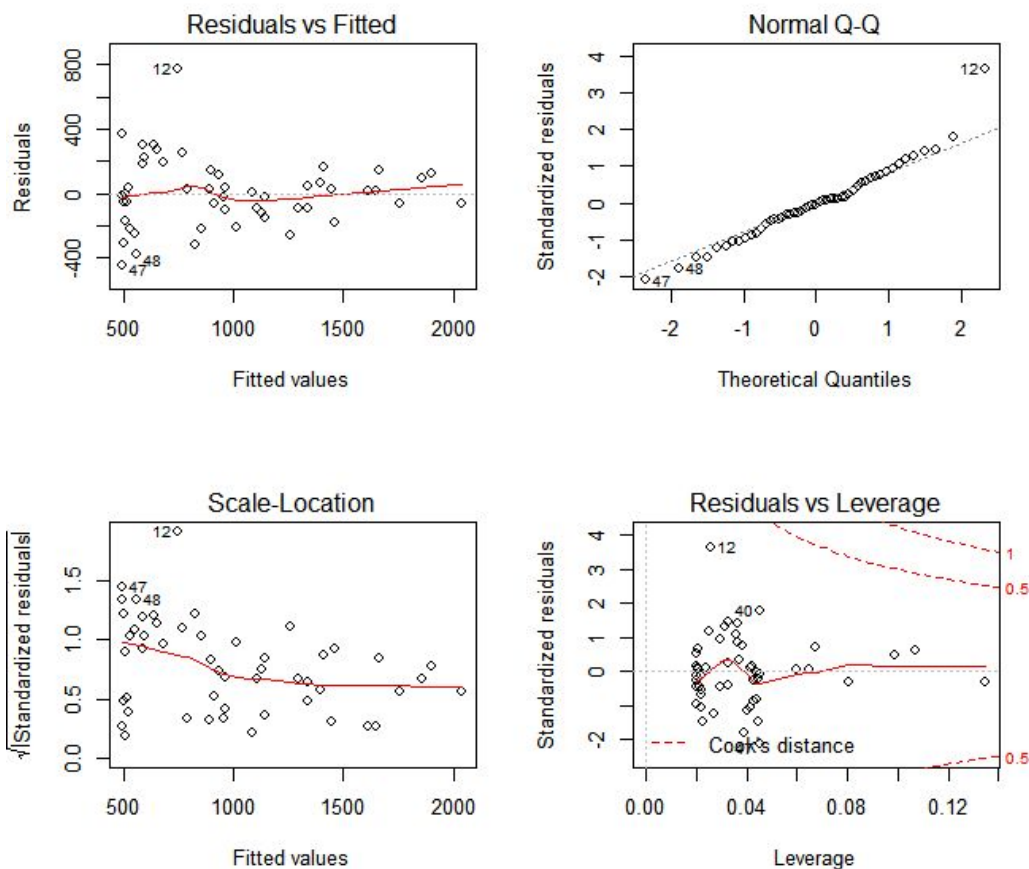
Residual standard error: 214.8 on 49 degrees of freedom

Multiple R-squared: 0.8087, Adjusted R-squared: 0.8048

F-statistic: 207.1 on 1 and 49 DF, p-value: < 2.2e-16

Histograms:





Covariation:

```
> cov(state$CO2, state$Coal) = 11668.89
```

Project Goals and Questions

For this project, our team aims to explore relationships between emissions and various energy sources to find out which areas of the country have the biggest issues with emissions outputs, and how that might be related to what energy sources are being used. Some questions we plan on exploring include:

Which energy resources are most correlated with harmful emissions?

What's the distribution of energy resources across the nation?

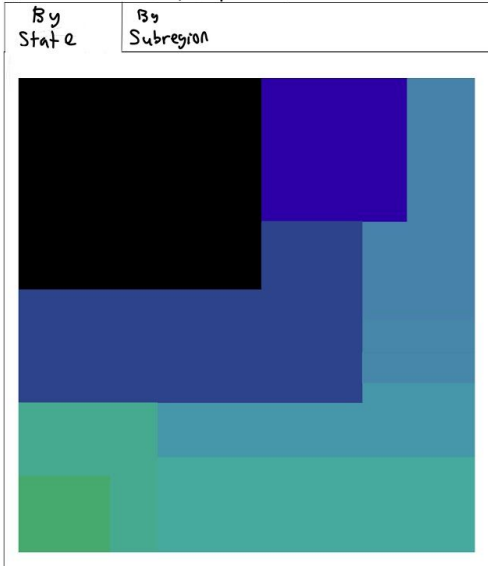
Are certain energy resources more popular in certain regions than others?

Do certain areas of the nation have more emissions outputs, and is that related to their distribution of energy resources?

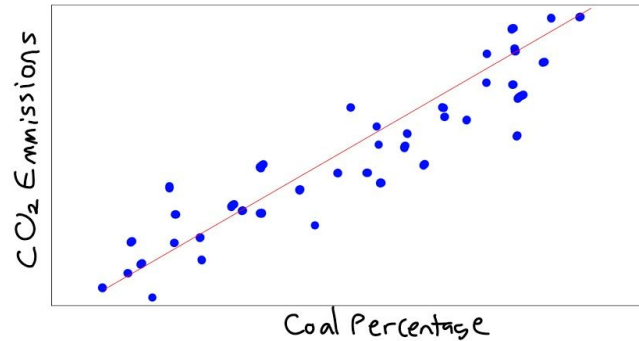
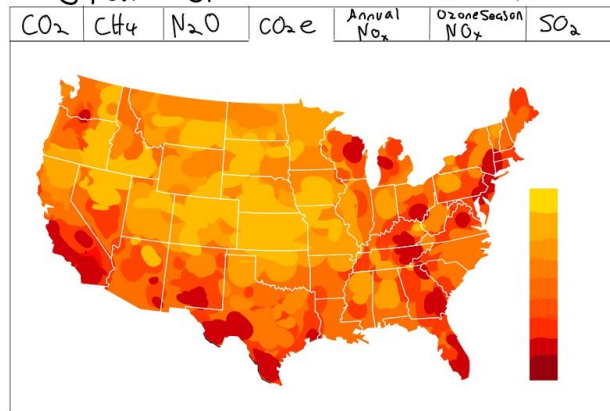
Dashboard Mockup

Project Info

Resource Distribution



State Emissions Outputs



Resources

<https://www.eia.gov/todayinenergy/detail.php?id=37392>

https://www.eia.gov/energyexplained/index.php?page=electricity_environment

<https://www.c2es.org/content/regulating-power-sector-carbon-emissions/>

<https://www.sciencedirect.com/science/article/pii/S2542435118305622>