

Empirical Analysis of Energy Markets - U6616
Empirical Exercise I - Data Exploration

Prof. Ignacia Mercadal

This problem set is due on **October 2**. I strongly recommend to start working on the homework early. I would suggest to plan on completing the first two parts by September 20. You can work in pairs and submit a common solution. Please submit the homework as an R markdown file (if there are data files, they put all the files in a zip file). The code must run without errors. To make this easier, set the working directory at the beginning so it can be easily changed by someone else running the code.

The purpose of this dataset is to analyze how the capacity of different fuels used to generate electricity in the United States has changed over time. We will also look at how emissions of different pollutants have changed and think about potential links between the two.

1. Start at the [EIA Electric Power Annual](#).
 - (a) First download Tables 4.2.A and 4.2.B., which contain capacity by source over time.
 - (b) Load the files to R and describe each variable. Make sure to use the appropriate class for each variable.
 - (c) What is capacity? Make sure you understand the difference between capacity and generation.
 - (d) Plot total capacity over time. How has this changed?
 - (e) Plot the share of capacity from the different sources over time. How has this changed?
2. Now we will obtain data on emissions over time. For this, follow [this link](#) to the EIA Monthly Energy Review. Section 11 contains data on many emissions sources, but just download the data set on carbon emissions from energy consumption in the electric power sector (11.6).
 - (a) Load the files to R and describe each variable. Make sure to use the appropriate class for each variable.
 - (b) Plot total emissions over time. How have they changed?

- (c) Plot the share of emissions from the different sources over time. How has this changed? How do you relate this to what you observed in the previous question?
3. Since the previous dataset only had capacity for 10 years, the next step is to download data from [EIA Form 860](#) to build capacity for a period as long as the one for emissions.
- (a) Follow the steps covered during recitation to build the total capacity of source for each year.
 - (b) Plot total capacity over time. How has this changed?
 - (c) Plot the share of capacity from the different sources over time. How has this changed?
 - (d) Add total carbon emissions to the graph. Is there any correlation?
4. Suppose you are interested in the causal effect of coal capacity on carbon emissions and you want to use the dataset you just build to estimate it.
- Run a regression of carbon emissions on coal capacity and present your results using the stargazer package. Use robust standard errors.
 - Does the previous regression capture the causal effect of coal capacity on carbon emissions? Why? If there is bias, do you expected to go in a particular direction? Explain.
 - Are there any controls you could add to correct this problem? Explain. You don't have to run more regressions, just to discuss how you would do it assuming you had access to typically available data.