1. The whole semester, we used real survey data and R to practice how economists work with data:

- Making graphs and descriptive stats to understand patterns in data
- Running regression models to see how one variable is related to others
- Estimating binary-choice models (logit/probit for yes/no outcomes)
- Trying machine learning methods like random forests, SVM, and LASSO

So by the time of the final project, we'd practiced:

- Cleaning and subsetting large datasets
- Computing means, standard deviations, shares, and correlations
- Building regression models with polynomials, interactions, and fixed effects
- Doing joint hypothesis tests and using robust / clustered standard errors

---

2. What the final project was supposed to do

The final project was meant to be a mini empirical paper on a topic of our choice, using real microdata and the econometric tools from class.

Assignment expectations:

- Use real micro data, not pre-summarized tables.
- Pick an economics question that has been studied in the academic literature.

Important parts from the paper:

1. **Data and descriptive statistics**
    - Describe the dataset: where it comes from, what years, what population.
    - Explain any restrictions (e.g., only certain regions, ages, or types of plants).
    - Show means, standard deviations, min/max, etc. for all key variables.
    - Add simple graphs (time series, scatterplots, histograms) and talk about what they show.
2. **Simple regressions**
    - Run basic linear regressions, clearly interpret coefficients.
3. **More complex regressions/robustness checks**
    - Add things like:
        - Nonlinear terms (squares, cubes, logs)
        - Fixed effects (e.g., by date or region)
        - Possible interactions
    - Do joint hypothesis tests (e.g., whether wind and wind² are jointly zero).
    - Check robustness: does the main result change if you change the sample or specification?

3. Project Gist on wind & fossil fuel generation

- **My project's question**: in ERCOT (Texas grid), when wind power output increases, how much does fossil-fuel generation (gas + coal) fall, after controlling for electricity demand and time patterns?
  In econometrics terms, I was trying to estimate the "offset":
- For each extra 1 MWh of wind, how many MWh of fossil generation (especially coal and gas) are displaced?
    - ➢ Built an hourly dataset for 2018–2020 from ERCOT's Native Load and Internal Generation by Fuel files.

- ➢ Calculated summary statistics, shares, correlations, and plotted patterns of wind vs demand and fossil fuels.

- ➢ Ran a series of regressions (static, dynamic, and with date fixed effects) of fossil generation on wind and load (plus time dummies), and then separate regressions for gas and coal.

Rubric of Econometrics Final Projects

1. Does the paper show basic statistics for all of the relevant variables? Does it explain what these mean?
2. Does the paper show simple correlations or conditional means, and demonstrate how to incorporate these basic statistics with the question being considered?
3. What regressions are being done – does the paper demonstrate ability to execute basic linear regressions? (These alone hardly make a C paper) Does the paper explain the meaning of these simple regressions, sufficient to demonstrate that the author understands the concepts?
4. More regressions? Things like interactions, squared or higher-order terms? Again, explain? Variety of specifications, debating considerations like possible endogeneity. These get to B.
5. Advanced statistics and regressions – does the paper demonstrate advanced techniques beyond those carefully covered in class? Is student able to take some advanced concepts, maybe just mentioned in class, learn them by herself?
6. Alt (not required but nice): did students find and assemble their own data, or did they use one of the ones from class?