ECON 8000/9000 Empirical Energy Econ

Topic 07: Best Practice for Empirical Work

Christy Zhou

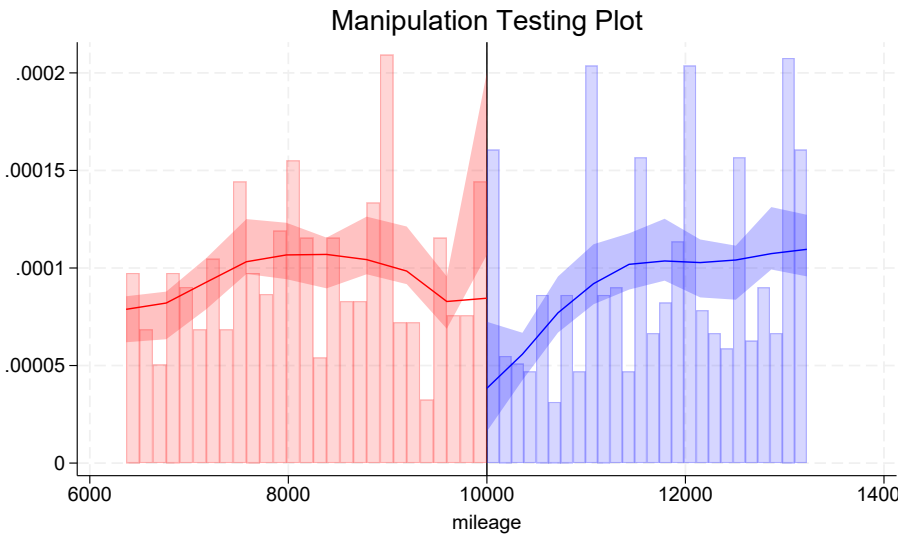
April 23, 2026

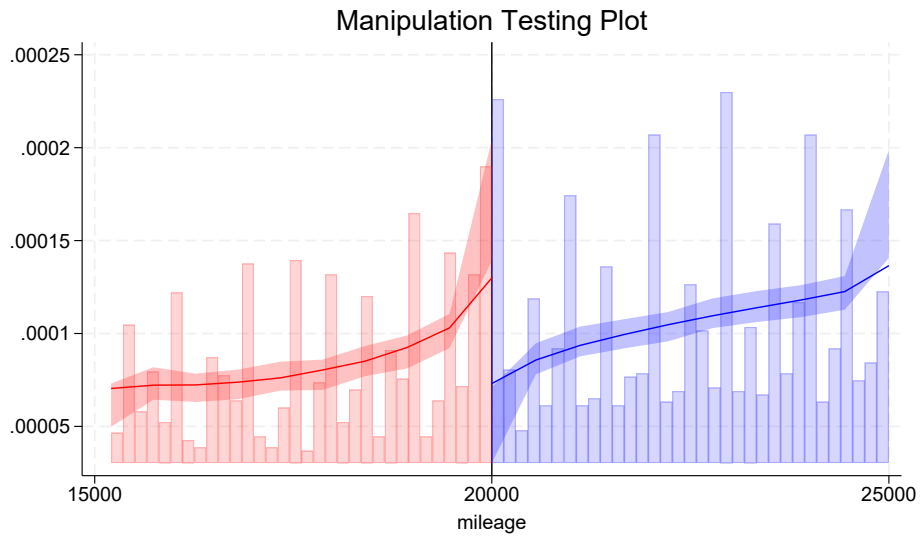
Outline

- PS3
- Student Presentations
- Best Practice for Empirical Project

PS3. Q1 Manipulation Test on Running Var Mileage

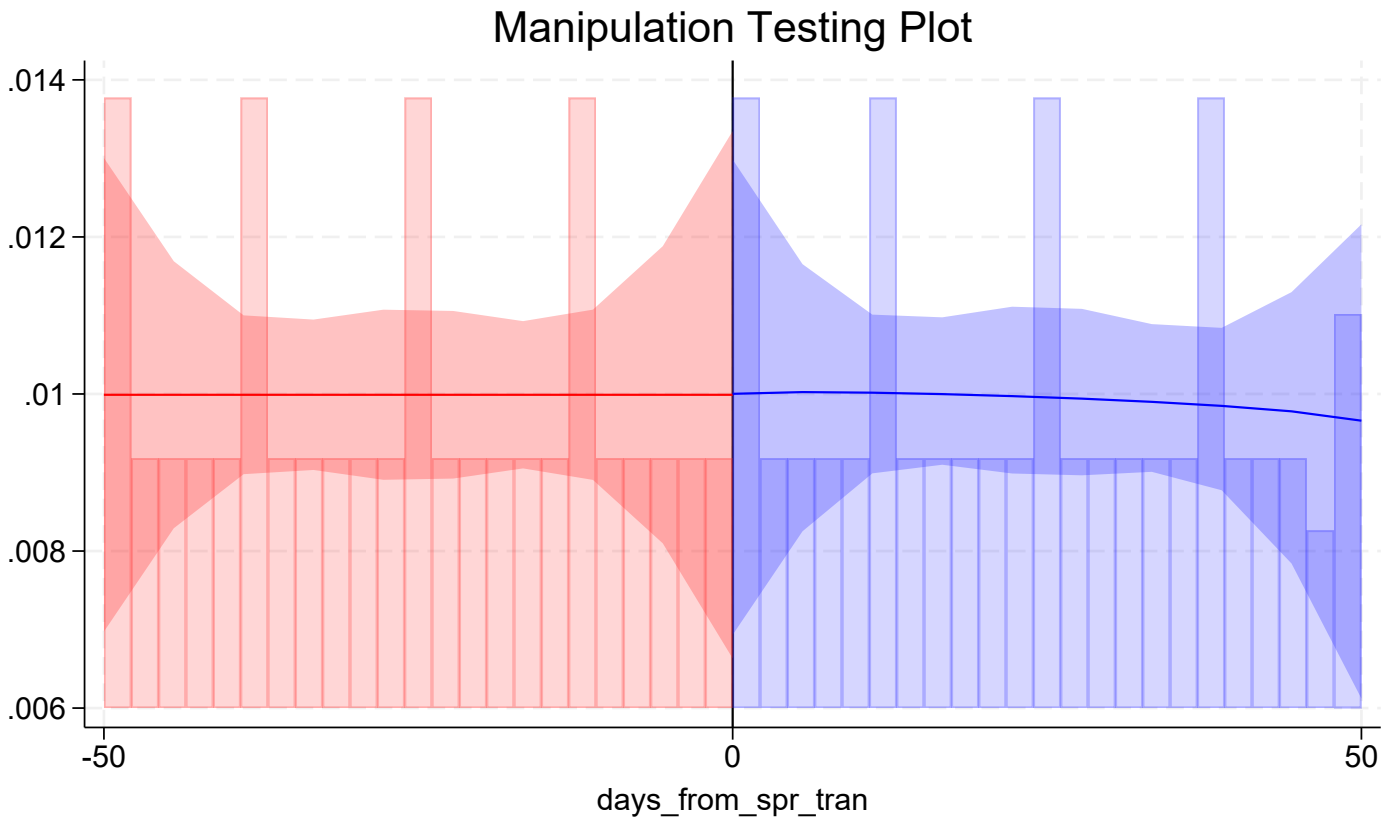
Non-overlapping CIs





PS3. Q2 Manipulation Test on Running Var Time t

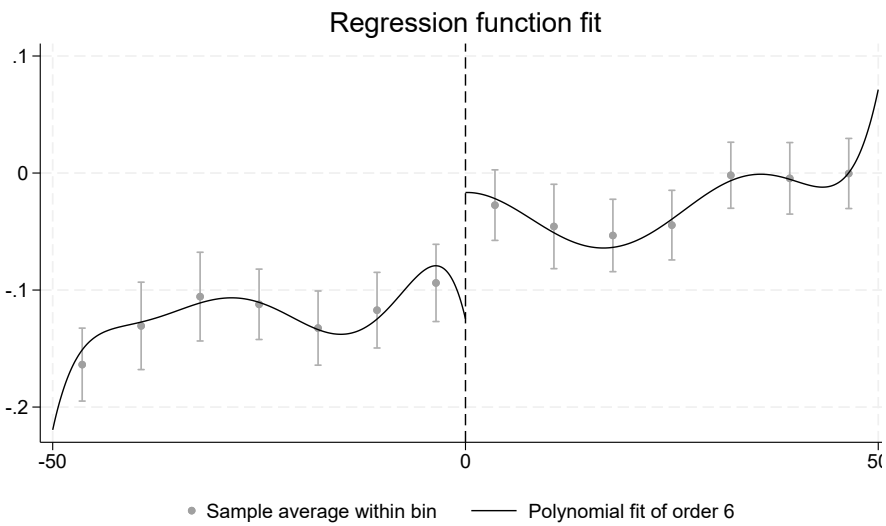
Overlapping CIs doesn't imply passing the test for RDiT



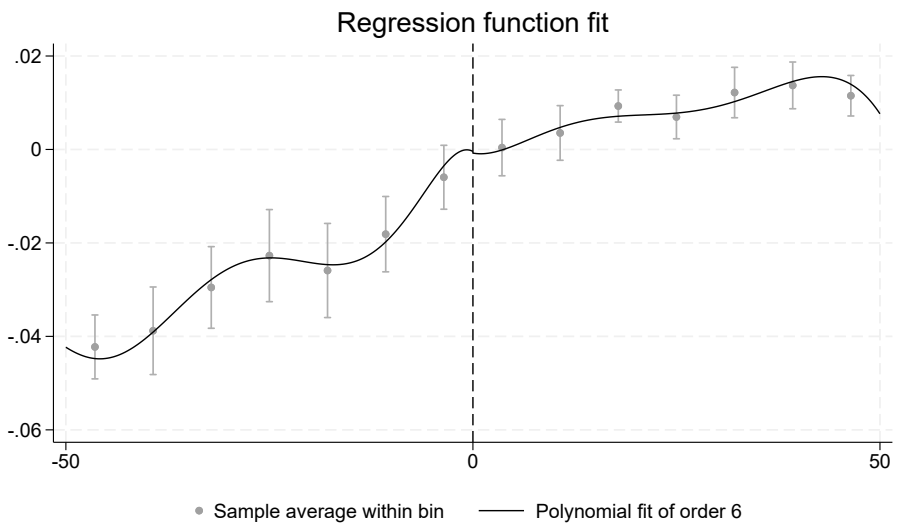
- Notice the uniform distribution in histogram (except weekly spike)

PS3. Q2 RD Plot

Two outcomes for spring DST: ln(Fatals) and ln(VMT)



Left: ln(Fatals)



Right: ln(VMT)

PS3. Q2 Part 10

Student Ideas for RD or RDiT

- "A possible RDiT would be to look at electric vehicles purchased following the Biden-Harris administration plan to build a national network of electric vehicle charging stations by 2030. Their might be a jump in purchases as people react to the easing of the concern that electric vehicles may not be able to travel long distances. It would be interesting to see if this announcement lead to an increase in the transition of combustion engines to electric vehicles."
- "I think that if you could find high-frequency data regarding wages and employment over a long period of time, you could study a minimum wage reform using an RDiT approach. The implementation date would serve as the cutoff and you could look at how employment varies around the policy implementation. This would be especially helpful in many international settings where the countries lack a pure control group and cannot use a DiD approach."
- "Does the switch to/from Daylight Saving Time affect residential electricity consumption patterns?"
- "What is the impact of renewable energy subsidy expiration dates on solar panel installation rates?"
- "An example could be the implementation of a renewable energy subsidy and seeing if it causes a jump in renewable energy building permits when the policy goes into effect"

PS3. Q2 Part 10

Student Ideas for RD or RDiT

- "For example, one could observe the effect that narcotic stimulants can have on local energy consumption. An RDIT could be used to analyze the change in energy demand in an area immediately following a known supply shock of methamphetamines (or other stimulants such as cocaine or crack) to see if residential utility consumption patterns change."

Outline

- PS3 ✓
- Student Presentations
- Best Practice for Empirical Project

Outline

- PS3 ✓
- Student Presentations ✓
- Best Practice for Empirical Project

Best Practices

Resource

Best Practices

- The GOAT of Best Practice: Gentzkow and Shapiro (2014) "Practitioner's Guide"
- Recent Update: Ristosvka (2019) "Coding for Economists"

Some Good Websites for Best Practices

- Julian Reif (link)
- Daniel Sullian (link)
- Grant McDermott (link)

My Recommendation

Top Principals

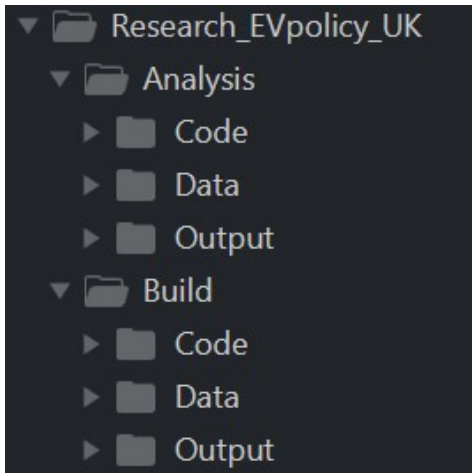
Priority

- A good folder structure
- Automate your code to avoid error-prone repetition
- Keep a good README for data cleaning

Additional comments on readability

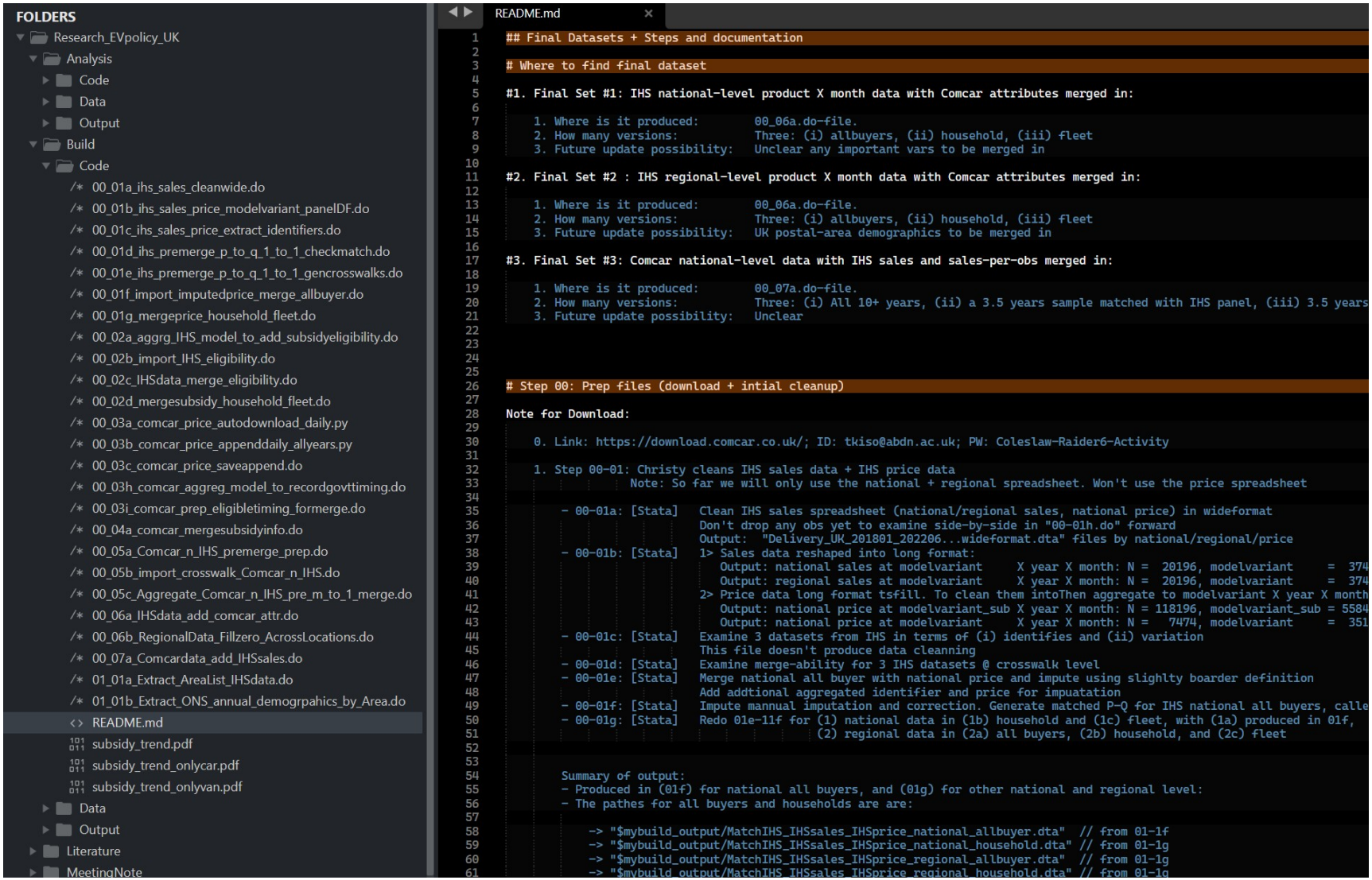
- Try to write your code in a way that it is readable even if you write no comments
- Try to name your variable in a way that you can understand (meaning & unit) even if you do not create a variable label (in Stata) or create a codebook (for other settings w/o the variable label feature). You should still maintain a codebook, though.

Folder Structure

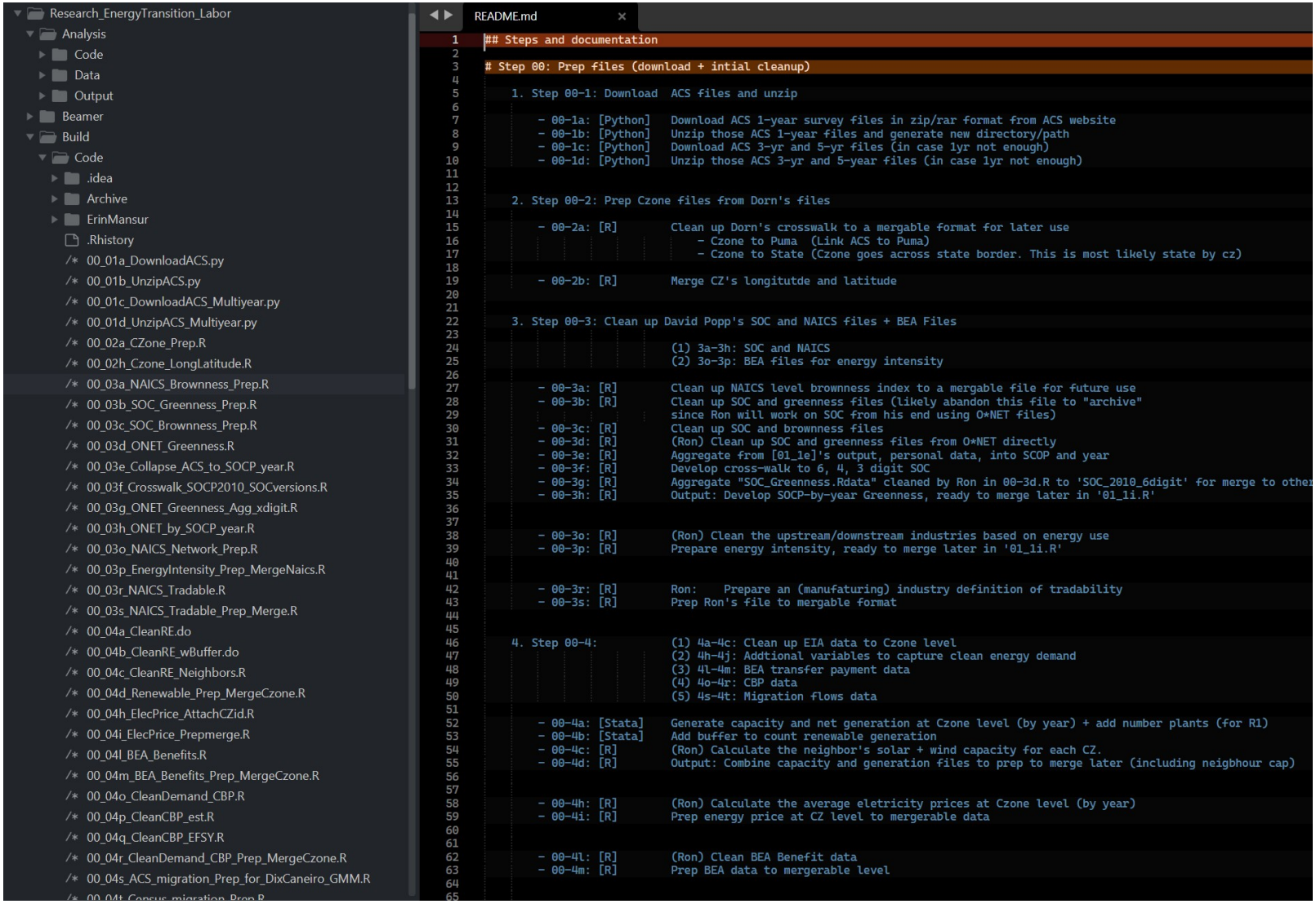


- You don't have to follow my preferred folder structure, but key principles:
- Write your data cleaning code and estimation code in different folders Designate a path for your final dataset(s) that you will directly use for estimation
- Designate a path to store all your outputs: visualization, estimation tables, etc.
- README is most important for data cleaning (Build folder). Some people also prefer to maintain one for analysis.

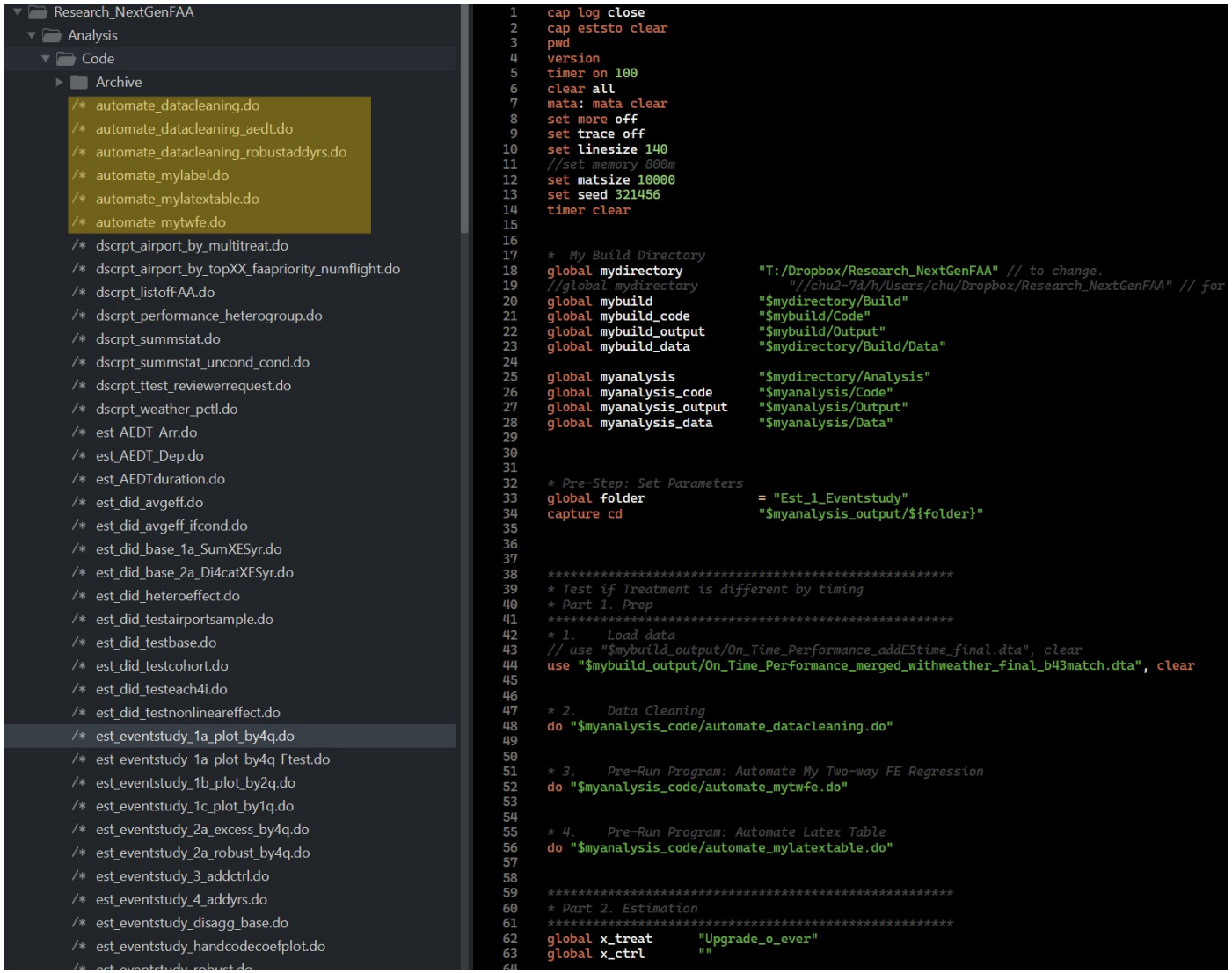
Maintain a README-file for Data Cleaning



Maintain a README-file for Data Cleaning



Use Automation to Reduce Error




```
cap log close
cap eststo clear
pwd
version
timer on 100
clear all
mata: mata clear
set more off
set trace off
set linesize 140
//set memory 800m
set matsize 10000
set seed 321456
timer clear


*  My Build Directory
global mydirectory          "T:/Dropbox/Research_NextGenFAA" // to change.
//global mydirectory            "//chu2-7d/h/Users/chu/Dropbox/Research_NextGenFAA" // for
global mybuild              "$mydirectory/Build"
global mybuild_code         "$mybuild/Code"
global mybuild_output       "$mybuild/Output"
global mybuild_data         "$mydirectory/Build/Data"

global myanalysis           "$mydirectory/Analysis"
global myanalysis_code      "$myanalysis/Code"
global myanalysis_output    "$myanalysis/Output"
global myanalysis_data      "$myanalysis/Data"



* Pre-Step: Set Parameters
global folder               = "Est_1_Eventstudy"
capture cd                  "$myanalysis_output/${folder}"



*****************************************************
* Test if Treatment is different by timing
* Part 1. Prep
*****************************************************
* 1.    Load data
// use "$mybuild_output/On_Time_Performance_addEStime_final.dta", clear
use "$mybuild_output/On_Time_Performance_merged_withweather_final_b43match.dta", clear


* 2.    Data Cleaning
do "$myanalysis_code/automate_datacleaning.do"


* 3.    Pre-Run Program: Automate My Two-way FE Regression
do "$myanalysis_code/automate_mytwfe.do"


* 4.    Pre-Run Program: Automate Latex Table
do "$myanalysis_code/automate_mylatextable.do"


*****************************************************
* Part 2. Estimation
*****************************************************
global x_treat      "Upgrade_o_ever"
global x_ctrl       ""
```

Other things

Length of scripts... this would be up to your preference

- I would avoid writing lengthy code whenever I can
- So that each script/file (data cleaning or estimation) usually serves one and only one purpose

The "prouduct" of the build folder includes all final datasets

- If there are more than 1 final dataset, then I would make it clear in README (Some ppl also maintain a Google Doc for collaborative work)
- If there is exactly one final dataset, then naturally the last step creates the final dataset