```
clear
set more off
capture log close
cd "/Users/anayatruss-williams/Desktop/pa research"
log using "Truss_Coding_Sample.log", replace

/*Anaya Truss-Williams' [Brief] Coding Sample

Excerpt of Stata code from my urban economics research project focusing on Philly

For this project, I found the effect of liquor store openings on property values of homes sold around a newly opened store. Pennsylvania—an ABC state—has a state monopology on the retail sale of liquor!

I use data from the PA Liquor Control Board (location data going to back in the day) and City of Philadelphia (recent housing transaction data). This was real world data and it was messy. An issue to be mindful of: uninhabitable properties. This is accounted for with the exclusion of outliers.

Also, some homes sold are in close proximity to more than one store. I assume that they aren't and only focus on whether there is any store within specific distance thresholds (within 0.5 mile, 0.5–1 mi, 1–2 mi) of the property.

Sections:
i. Summary stats
ii. Dealing with zie outliers
iii. Regression  */

use "pa research sept 14/PhillyHouses_sept.dta", replace

browse
des

label var number_of_bedrooms "Number of Bedrooms"
label var number_of_bathrooms "Number of Bathrooms"
label var total_livable_area "Square Footages"
label var phalf "Within 0.5 mi"
label var pmile "Within 0.5–1 mi"
label var ptwomiles "Within 1–2 mi"
label var pfourmiles "Within 2–4 mi"

*** Summary stats ***
sum treated
sum sale_price total_livable_area age number_of_bedrooms
```

```
number_of_bathrooms

*For my spatial analysis in Stata, I use the geonear and geodist packages. Geonear was initally used to find nearest neighbors. In the spring, I had some issues using the package so I transition to using geodist. I calculated the distance between each home and my four stores of interest (the stores opening in philly during my time period in zip codes that didn't have a store already')

*Create a single var for the various distance thresholds
gen dist = 0
replace dist = 1 if phalf == 1
replace dist = 2 if pmile == 1
replace dist = 3 if ptwomiles == 1
replace dist = 4 if pfourmiles == 1

sort dist
by dist: sum sale_price

*** Dealing with zie outliers ***
sum sale_price, detail

*making a histogram to see the frequency distribution
hist sale_price

*Defining the outliers I use for the regression:
gen outlier = 0
replace outlier = 1 if sale_price > 1450000
replace outlier = 1 if sale_price < 55000

*The outlier var is a dummy (1 = outlier, 0 = else.) for observations with sale prices below $55,000 and above $1.45M. This way, I can easily exclude sales below the 25th percentile and above the 99 percentile

*Based on my search of single family properties on Zillow and similar sites, these properties below $55,000 are likely inhabitable, whether because it is a parking space, storage or house that is in ruins

/*Here is the better way to define outliers that I did not previously use:
outside 1.5 times the interquartile range above the upper quartile and below the lower quartile (Q1 - 1.5 * IQR or Q3 + 1.5 * IQR)
IQR = Q3 - Q1 = 75th% - 25th% = 240,000 - 55,000 = 185,000
Max outlier/upper quart = Q3 + 1.5*IQR = 240000 + 1.5(185) = 240000 + 277500 = 517500

```

```
gen outlier2 = 0
replace outlier2 = 1 if sale_price > 517500
replace outlier2 = 1 if sale_price < 55000

This is a better way to define outliers since it uses the IQR
range. The DID results are relatively close (compared to when
including outliers) whether I used the first or second method */

*making a histogram that excludes outliers
hist sale_price if outlier == 0

*** Running Regression ***

*making a macro for the indep vars
local x "total_livable_area age number_of_bedrooms
number_of_bathrooms"

*regression on sale price
reg sale_price `x'

*regressing the natural log of the sales price on the independents
vars for observations that don't have outlying sale prices, with
fixed effects
reg logprice `x' phalf#postone i.year i.zip_code if outlier == 0

*same regression but only using observations not within the second
definition I used for outliers
reg logprice `x' phalf#postone i.year i.zip_code if outlier2 == 0

save "Philly_sample.dta"

log close

```