ECON2330: Notes on making graphs and tables[1]

In this document I introduce some basic examples to generate graphs and tables in stata and I provide some useful links to do applied research in Economics.

Contents

[1]Notes prepared by Juan Pablo Uribe (juan_uribe@brown.edu) for the class ECON 2330 at Brown University. Last updated November 10, 2021

I. Graphs

Presenting results usually involves "good" figures. Schwabish (2014), provides some useful discussions about creating effective visualizations.Edward Tufte has some books about "good visualizations".

1. Some general recommendations

1. Show the data
2. Reduce the clutter
3. Integrate the text and the graph

Schwabish (2014)

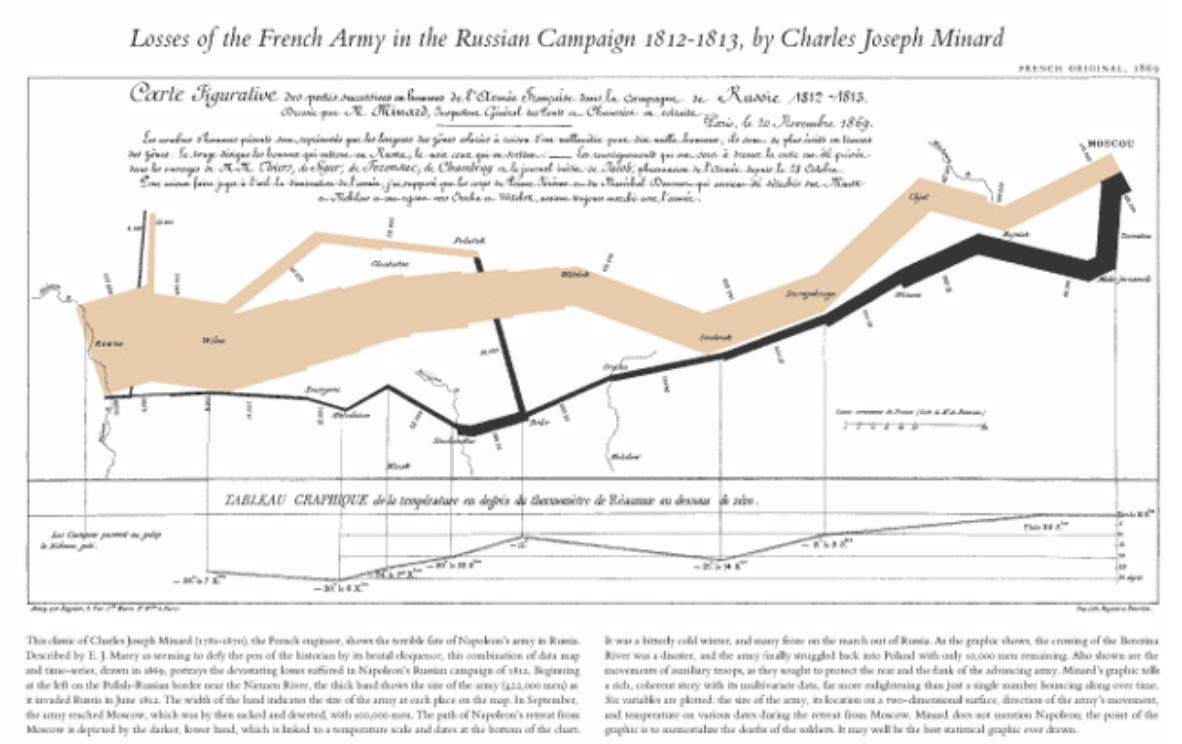
2. Graphical Excellence

- well designed presentations of interesting data-a matter of substance, of statistics and of design
- complex ideas communicated with clarity, precision and efficiency.
- gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- nearly always multivariate.
- requires telling the truth about the data.

3. Fundamental principles of analytical design

1. *Comparisons.* You want to answer the question: Compared to what?
2. *Causality, Mechanism, Structure, Explanation*
3. *Multivariate Analysis.* Show more than 1 variable.
4. *Integration of evidence.* Do not segregate the information by mode of production. Completely integrate words, numbers, images, diagrams.
5. *Documentation.*

6. *Content Counts Most of All.* Analytical presentations ultimately stand or fall depending on the quality , integrity of their content. Always try to get better content. That's the only way to improve your presentation.



This classic of Charles Joseph Minard (1781–1870), the French engineer, shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its brutal eloquence, this combination of data map and time-series, drawn in 1869, portrays the devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen River, the thick band shows the size of the army (422,000 men) as it invaded Russia in June 1812. The width of the band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally struggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. Minard does not mention Napoleon; the point of the graphic is to memorialize the deaths of the soldiers. It may well be the best statistical graphic ever drawn.

Figure 1: Example of a good use of the principles of analytical design

Note: based on Tufte (2006)

4. Schemes in Stata

If you use stata you will probably dislike the default scheme. In this document I provide some examples on how to create your own schemes in Stata. There are some default schemes (i.g., s1mono,s2color,economist) and I uploaded some schemes to canvas so you can use them. To use the schemes I provided,you need to copy them in your personal folder in stata. To find it, you can use "adopath". You can also download ready to use schemes. If you type "ssc d s" in stata, you will find some schemes. For example the "tufte" scheme I show below. To set a default scheme different than the one provided by stata use.In the example below, all your graphs will use the scheme "tufte" by default.

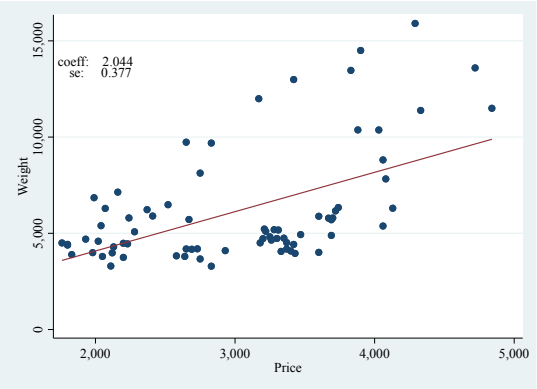
```
set scheme tufte, perm
```

More Generally, you can customize your own scheme.For that, you can copy the scheme that is closer to what you want, save it with a different name, and do the adjustments you want. For example, I use the scheme tufte and changed some features to create the scheme JPU. The main difference is that JPU is in Color and the lines and cirles are solid. To see how to add particular features to your own scheme use:

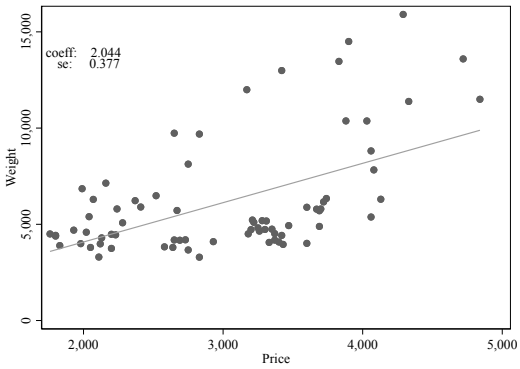
```
help scheme entries
```

If you want to explore the different options to edit your graphs in a systematic way see Ben Jann's slides and Asjad Naqvi

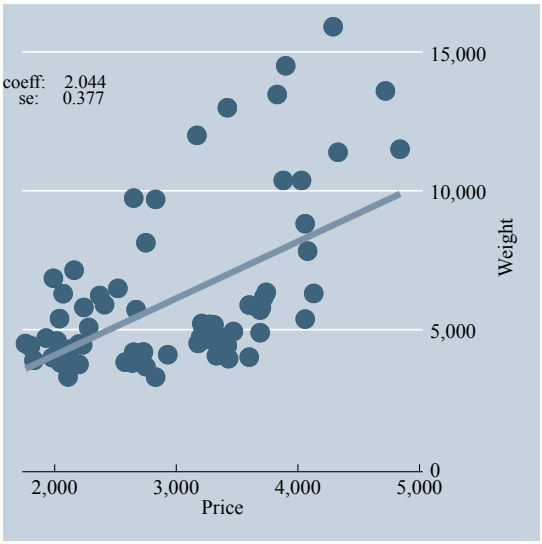
5. Examples with different schemes



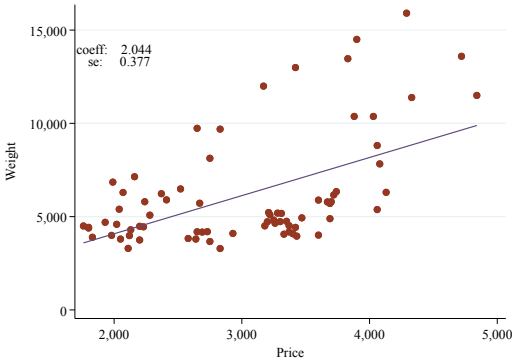
a. Scheme: s2color



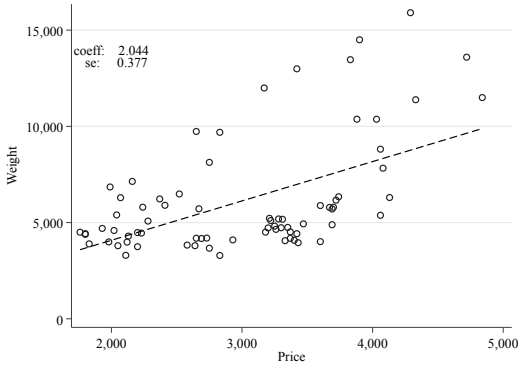
b. Scheme: s1mono



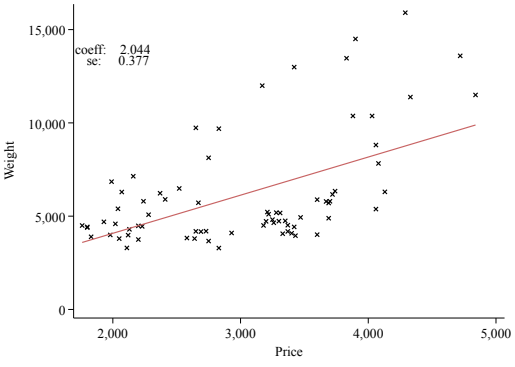
c. Scheme: economist



d. Scheme: JPU



e. Scheme: tufte



f. Scheme: labor2

```
sysuse auto, clear

eststo:  regress price weight



*Put some scalars in the graphs
matrix b=e(b)
scalar b=b[1,1]
matrix V=e(V)
scalar V=V[1,1]
scalar se=V^(1/2)

local beta: display %9.3fc b
local SE: display %9.3fc se
di "‘SE’"

foreach x in s2color labor2 labor tufte JPU s1mono labor economist{ //
        # d ;
        tw (scatter price weight ) (lfit price weight ) ,
        ytitle("Weight") xtitle("Price")
        leg(  label(2 "regression fit"))
         text(14000 2000  "coeff:‘beta’")
        text(13400  2000 "   se: ‘SE’")
        leg(off)
        scheme(‘x’);
        # d cr;
        gr export "${output}/figure1‘x’.pdf", replace font(times)
}
```

6. Add information to the graphs

6.1. Add Scalars

Note that I include The coefficient and se from the regression. Another example is to add the summary statistics to the Histogram.

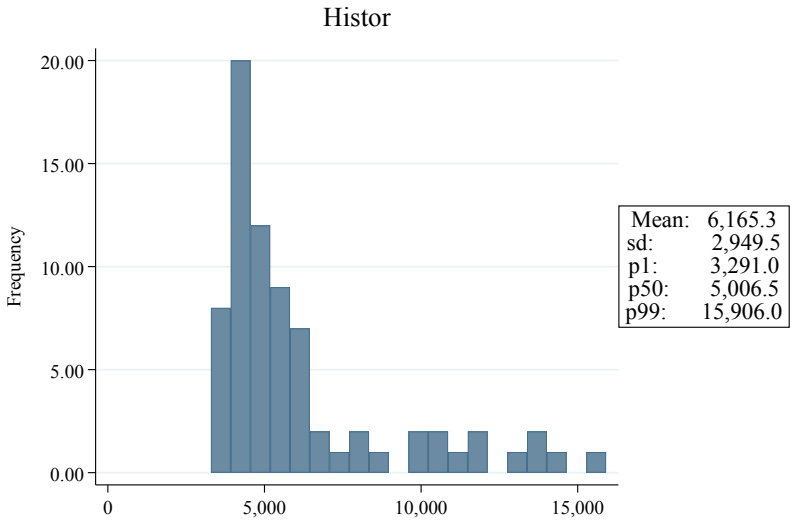


Figure 3: Histogram

```
sum price, d
local mean: display %9.1fc ‘r(mean)’
local sd: display %9.1fc ‘r(sd)’
local p1: display %9.1fc ‘r(p1)’
local p50: display %9.1fc ‘r(p50)’
local p99: display %9.1fc ‘r(p99)’
local max ‘r(max)’
*Histograma de la variable
# d ;
tw (histogram price, bin(20) freq ),   ylabel(,format(%9.2fc))
                   xtitle("")
                   note( "Mean: ‘mean’"
                         "sd:        ‘sd’"
                                "p1:       ‘p1’"
                                "p50:     ‘p50’"
                                "p99:     ‘p99’",
box size(*1.5) position(3)) xsize(6) title(Histor)
                   ;
        graph export "${output}/histogram.pdf" , replace;
# d cr ;
```

6.2. Leyends

The default in Stata is to put the leyend below the graph. However you should try to help the reader. In that sense, having the legend inside the graph in a corner could be

helpful. Also the labels should be meaningful. In the example below I show how to modify and move the legend.

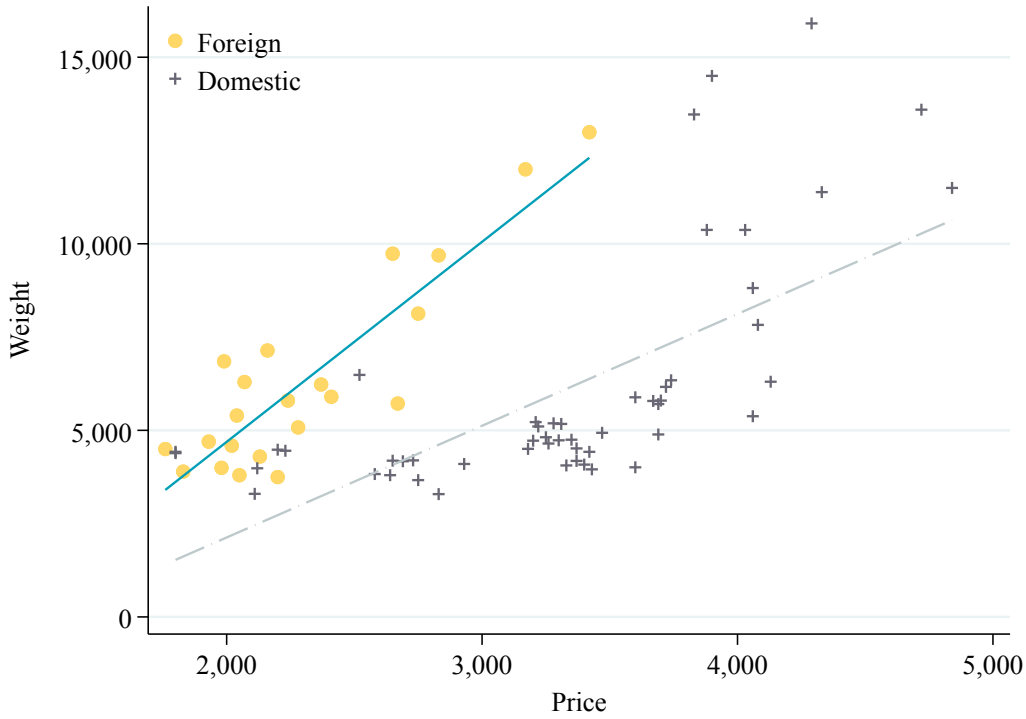


Figure 4: Add Leyend

```
**** Manipulate the legends
# d ;
tw (scatter price weight if foreign==1) (lfit price weight if foreign==1 )
(scatter price weight if foreign==0) (lfit price weight if foreign==0) ,
        ytitle("Weight") xtitle("Price")
        leg(order(1 3) label(1 "Foreign") label(3 "Domestic")
        pos(10) ring(0) col(1) bcolor(none)  )

        leg(on)

        ;
        # d cr;
        gr export "${output}/figure3.pdf", replace font(times)
```

6.3. Math Symbols

You can use math symbols to make it easier to compare the models you present with the graphs. Also, sometimes the font in the default stata graphs are to small. In the graphs below, I show how to add symbols and adjust the size and symmetry of the graph.

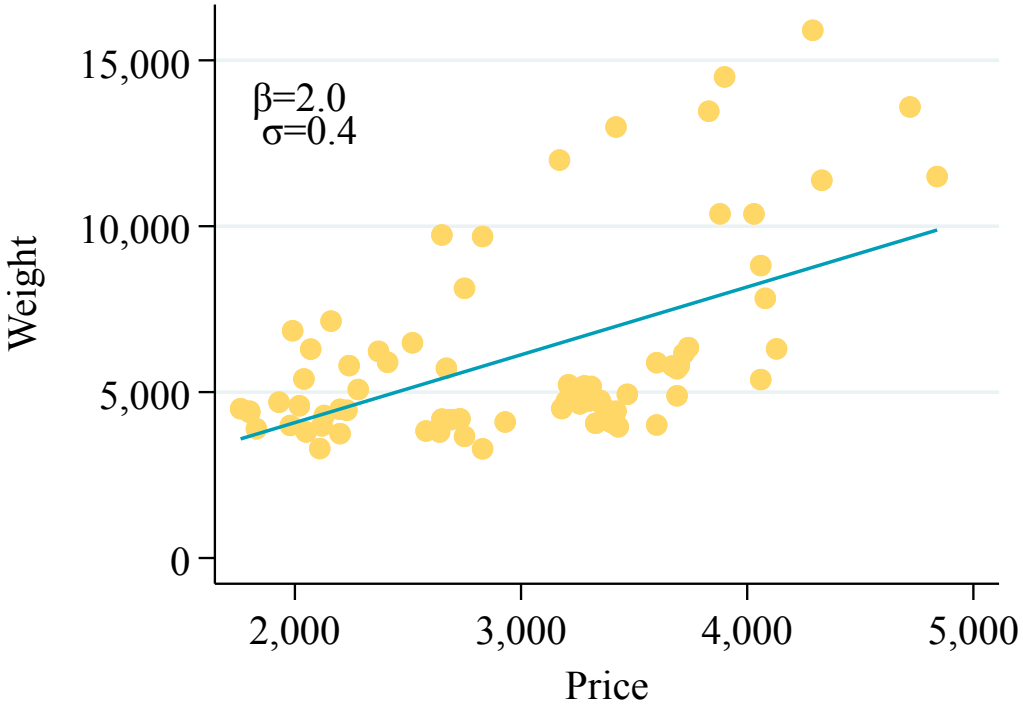


Figure 5: Math Symbols and bigger fonts

```
**** Math Symbols

eststo:  regress price weight

*Put some scalars in the graphs
matrix b=e(b)
scalar b=b[1,1]
matrix V=e(V)
scalar V=V[1,1]
scalar se=V^(1/2)


local beta: display %3.1fc b
```

```
local SE: display %3.1fc se
di ‘SE’



        # d ;
        tw (scatter price weight ) (lfit price weight ) ,
        ytitle("Weight") xtitle("Price")
        leg(  label(2 "regression fit"))
         text(14000 2020  "{&beta}=‘beta’")
        text(13000  2020 "  {&sigma}=‘SE’")
        leg(off)
        scheme(‘x’)
        scale(*1.5);
        # d cr;
        gr export "${output}/figure4.pdf", replace font(times)
```

II. Export Tables

There a lot of different ways to export tables from Stata. Some of the most popular are "*outreg2*", "*tabout*", "*xml_tab*", and many more. Each command has its advantages and which one you use is a personal choices. After using most of the previously mentioned, I converged to using *esttab/estout*, I find it extremely flexible and useful to create tables to use in a .tex compiler. If you save it in .rtf, or .csv you can have it in excel or word. Other methods are described by Julian Reif in his website. I have not use them but they seem useful. You can give it a try.

Ten Guidelines from (Schwabish, 2020)

1. Basic summary statistics

To create tables easy to export to a Tex table a very useful command is *estpost* see some of the examples below

1.1. Example t-test

Table 1: differences by origin

	(1)		
	Domestic	Foreign	Difference/se
Price	6072.423	6384.68	-312.26
			(754.45)
Mileage (mpg)	19.827	24.77	-4.95***
			(1.36)
Repair Record 1978	3.021	4.29	-1.26***
			(0.21)
Gear Ratio	2.807	3.51	-0.70***
			(0.08)
Observations	74		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
global variables price mpg rep78 gear_ratio
```

```
eststo clear
eststo, prefix(ttest): estpost ttest ${variables}, by(foreign)
 # d ;
        esttab ttest1 using "${output}/table_ttest.tex", replace
        cell("mu_1(fmt(3) label(Domestic)) mu_2(fmt(2) label(Foreign ))
        b(star fmt(2) label(Difference))" ". . se(par fmt(2))") label
        legend
          ;
 # d cr ;

 eststo, prefix(sumstat): estpost tabstat price mpg rep78, listwise
 statistics(mean sd) columns(statistics)
```

1.2. Example summary statistics

- Mean and sd for the whole sample

Table 2: Summary Statistics

	(1)	
	mean	sd
price	6146.0	2912.4
mpg	21.29	5.866
rep78	3.406	0.990
N	69	

```
eststo, prefix(sumstat): estpost tabstat price mpg rep78, listwise
statistics(mean sd) columns(statistics)

 # d ;
esttab sumstat2 using "${output}/table_sumstat.tex",
cells("mean(fmt(a3)) sd")
 replace
;
# d cr ;
```

- Mean and sd by subgroup

Table 3: Summary Statistics by foreign status

	(1)		
	Domestic	Foreign	Total
Price	6179.2	6070.1	6146.0
	(3189.0)	(2221.0)	(2912.4)
Mileage (mpg)	19.54	25.29	21.29
	(4.753)	(6.310)	(5.866)
Repair Record 1978	3.021	4.286	3.406
	(0.838)	(0.717)	(0.990)
Observations	69		

mean coefficients; sd in parentheses

```
 eststo, prefix(sumstat): estpost tabstat price mpg rep78  , by(foreign)
 listwise    statistics(mean sd) columns(statistics)
# d ;
esttab sumstat3  using "${output}/table_sumstat2.tex",
main(mean) aux(sd) nostar unstack  label
replace
;
# d cr ;
```

2. Regression outputs

2.1. Basic

Table 4: Basic

	(1)	(2)	(3)
	Est1	Est2	Est3
Weight (lbs.)	2.044***	2.266***	2.442***
	(0.3768)	(0.5111)	(0.6881)
Trunk space (cu. ft.)		-60.039	-99.367
		(92.8573)	(90.9304)
Mileage (mpg)			-63.210
			(84.2177)
Repair Record 1978			884.448**
			(325.6690)
Constant	-6.707	148.553	-1540.729
	(1174.4296)	(1203.4059)	(3635.3483)
Observations	74	74	69
Adjusted R^2	0.280	0.274	0.337
F	29.423	14.802	9.654
rmse	2,502.309	2,512.483	2,370.832

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
*- Basic case
 sysuse auto, clear
 eststo clear
 eststo ,prefix(Q1_):  regress price weight,r
 eststo ,prefix(Q1_): regress price weight trunk,r
  eststo ,prefix(Q1_): regress price weight trunk mpg rep78,r

# d ;
        esttab Q1_* using "${output}/table_reg.tex", replace
        b(%9.3fc) se(%9.4fc) scalars(F rmse) sfmt(%9.3fc)
        se  par  ar2  label nogaps  drop()  compress
        mtitles( "Est1 " "Est2" "Est3");
 # d cr ;
```

2.2. Adding Fixed Effects Indicators

Table 5: Including FE

	(1)	(2)	(3)
	Q1_1	Q1_2	Q1_3
Weight (lbs.)	2.044***	2.266***	2.431*
	(0.3897)	(0.6227)	(1.0768)
Constant	-6.707	148.553	-1083.997
	(1032.3939)	(947.5387)	(4445.5871)
Trunk FE	No	Yes	Yes
Other	No	No	Yes
Observations	74	74	69
Adjusted R^2	0.280	0.274	0.288
F	27.506	16.692	.

Standard errors in parentheses

others iclude mpg rep78

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
*- Indicate FE
eststo clear
 eststo ,prefix(Q1_):  regress price weight
 eststo ,prefix(Q1_): regress price weight trunk
 eststo ,prefix(Q1_): regress price weight i.trunk mpg rep78
# d ;
        esttab Q1_* using "${output}/table_regFE.tex", replace
        b(%9.3fc) se(%9.4fc) scalars(F) sfmt(%9.3fc)

        /***** This is the KEY  ******/
        indicate("Trunk FE=*trunk" "Other=mpg rep78")
        addnotes(others iclude mpg rep78)

        se  par  ar2  label nogaps modelwidth(7) drop()  compress
        mtitles( );
 # d cr ;
```

2.3. Big Tables and Table Adjustments

You have to add the package "booktabs,tabularx" "longtable" "adjustbox"

Table 6: Make it fit in the tex file

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	price	price	price	price	price	price	price	price	price	price	price	price
weight	2.044***	2.266***	2.431*	2.044***	2.266***	2.431*	2.044***	2.266***	2.431*	2.044***	2.266***	2.431*
	(0.3897)	(0.6227)	(1.0768)	(0.3897)	(0.6227)	(1.0768)	(0.3897)	(0.6227)	(1.0768)	(0.3897)	(0.6227)	(1.0768)
mpg			-35.351			-35.351			-35.351			-35.351
			(79.8520)			(79.8520)			(79.8520)			(79.8520)
rep78			573.405			573.405			573.405			573.405
			(462.2821)			(462.2821)			(462.2821)			(462.2821)
CONSTANT	-6.707	148.553	-1083.997	-6.707	148.553	-1083.997	-6.707	148.553	-1083.997	-6.707	148.553	-1083.997
	(1032.3939)	(947.5387)	(4445.5871)	(1032.3939)	(947.5387)	(4445.5871)	(1032.3939)	(947.5387)	(4445.5871)	(1032.3939)	(947.5387)	(4445.5871)
Trunk FE	No	Yes	Yes									
N	74	74	69	74	74	69	74	74	69	74	74	69
adj. R^2	0.280	0.274	0.288	0.280	0.274	0.288	0.280	0.274	0.288	0.280	0.274	0.288
F	27.506	16.692	.	27.506	16.692	.	27.506	16.692	.	27.506	16.692	.

Robust Standard Errors in Brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

```
 *- Adjust the size to fit in the table
# d ;
        esttab Q1_*  Q1_* Q1_* Q1_* using "${output}/table_regFE_adj.tex", replace
        b(%9.3fc) se(%9.4fc) scalars(F) sfmt(%9.3fc)
        se  par  ar2  nolabel nogaps modelwidth(7) drop()  compress
        substitute( "_cons" "CONSTANT")

        indicate("Trunk FE"=*trunk)
         prehead(\begin{table}[H]\centering
        \def\sym#1{\ifmmode^{#1}\else\(^{#1}\)\fi}
        \begin{adjustbox}{max width=\textwidth,max totalheight=\textheight}
        \begin{tabular}{l*{@span}{c}}
        \hline\hline)
        postfoot(\hline\hline
    \multicolumn{@span}{m{1.4\textwidth}}{\begin{singlespace}

    \footnotesize Robust Standard Errors in Brackets. @starlegend.  @note  \end{singlespace} }
        \end{tabular}\end{adjustbox}\end{table})        ;
 # d cr ;
```

Note that in the previous example you specify the beginning and end of the table directly into stata. Also note that in this example I change the name of the constant using the option "substitute". You can use this to automate changes to your tables.

III. Useful Additional Resources:

1. Guide for "Good Practices" to write codes

Write organized code is very important in applied work. In this section I link some useful resources

- Code and Data for the Social Sciences: A Practitioner's Guide (Gentzkow and Shapiro)
- Julian Reif includes some resources to other sources.
- Coding for Economists A Language-Agnostic Guide to Programming for Economists (Ljubica "LJ" Ristovska)
- Michael Stepner
- Jonathan Dingel (based on (Gentzkow and Shapiro) above and https://www.youtube.com/watch? Ball Plain text in your workflow

2. Resources for Graphs and Tables

- R and stata graphs
  - Stata with very usefull links
- Tables

3. Guides to learn R

- Official Website
- R for Data Science
- R FOR STATA USERS
- http://r-statistics.co

4. Resources for Stata

- Stata Cheat Sheet
- Data Managment guide

5. Miscellaneous

- Latex: Symbols, Help for Tables, +Overleaf + Wikibooks
- Writing equations in Latex: With Mathpix Snip you write the equations or take pictures and they translate it into a .tex code (or word).
- Theory Graphs with tikz: Chiu Yu Ko Guide,or some examples for economics
- Color Blind Friendly pallets
- Convert Documents
- Jonathan Dingel He has a lot of advice in many different topics
- Calling Bullshit

IIIReferences

Schwabish, J. A. (2014, February). An economist's guide to visualizing data. *Journal of Economic Perspectives*, *28*(1), 209-34. Retrieved from link

Schwabish, J. A. (2020). Ten guidelines for better tables. *Journal of Benefit-Cost Analysis*, *11*(2), 151–178.

Tufte, E. R. (2006). *Beautiful evidence* (Vol. 1). Graphics Press Cheshire, CT.