ECON 2330: Notes on making graphs and tables¹

Juan Pablo Uribe

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.

Note that the document has a lot of links to relevant resources. Every word in blue has a relevant link.

I. Graphs

Good figures are essential in applied economics research. They are usually essential in conveying the validity of the research design, help explaining the data, and summarize the key results.

In this note, I provide some general best practices to create figures with an emphasis on how to produce them in sata. Most of the content is based on the work and suggestions of **Jonathan Schwabish** (Schwabish, 2014, 2021), and Edward Tufte. Another book that is word reading is Calling Bullshit.

1. 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*. Thoroughly describe the evidence. Provide a detailed title, indicate the authors and sponsors, document the data sources, show completely measurement scales, point out relevant issues.
- 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

 $^{^1{\}rm Notes}$ prepared by Juan Pablo Uribe (juan_uribe@brown.edu) for the class ECON 2330 at Brown University. Last updated March 11, 2022

content. That's the only way to improve your presentation.

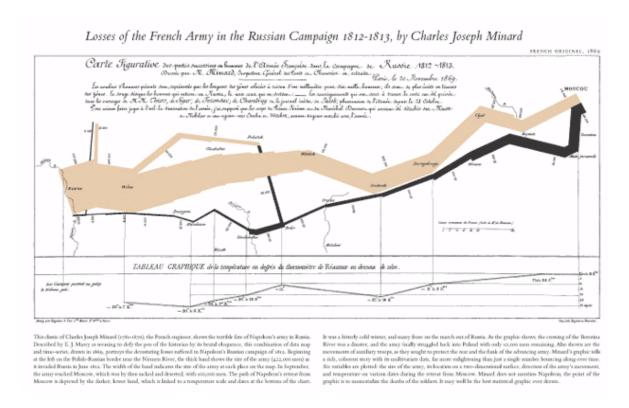


Figure 1: Example of a good use of the principles of analytical design

Note: based on Tufte (2006)

2. Graphical Excellence

- Well designed presentation of interesting data—a matter of the 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. Some general recommendations

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

Schwabish (2014)

4. Usefull Resources for graphs

- Asjad Naqvi has website has created and incredible amount of resources to create figures in stata.
- World Bank-DIME created a website with resources to create better graphs in stata and R.

5. 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 scheme 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. Also the following GitHub repository by Asjad Naqvi has more schemes. You can download them using

```
ssc install schemepack, replace
```

For example the "white_viridis" 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 "white_viridis" by default.

```
scheme white_viridis, perm
```

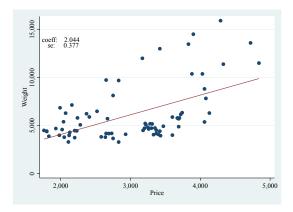
More Generally, you can customize your own scheme. For that, you can copy the scheme that is closer to what you want, save it with a different name, and do the adjustments

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 circles are solid. To see how to add particular features to your own scheme:

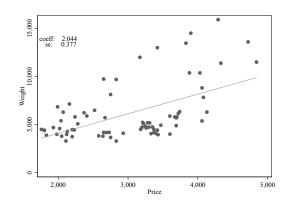
help scheme entries

If you want to explore different options to edit your graph in a systematic way, see Ben Jann's slides and Asjad Naqvi guide. This guide is very comprehensive and it walks you step by step on how to adjust your schemes.

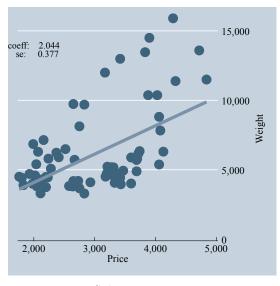
i) Examples with different schemes



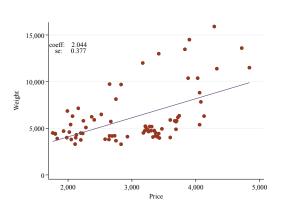
 $a. \ Scheme: \ s2color$



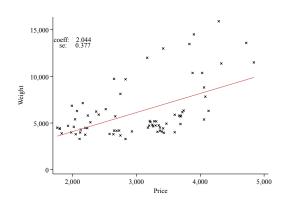
b. Scheme: s1mono



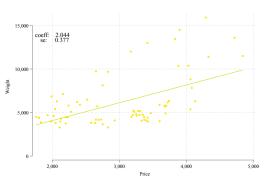
 $c. \ Scheme: \ economist$



d. Scheme: JPU



 $e. \ Scheme: \ labor 2$



 $f. \ Scheme: \ white_viridis$

```
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{ //
       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');
       gr export "${output}/figure1'x'.pdf", replace font(times)
}
```

6. Add information to the graphs

i) Add Scalars

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

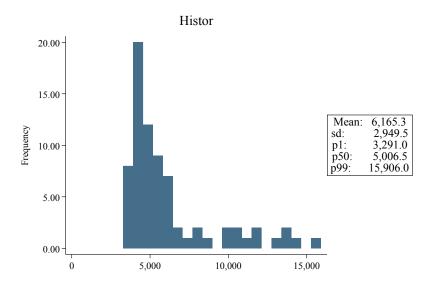


Figure 3: Histogram

```
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 ;
```

ii) 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.

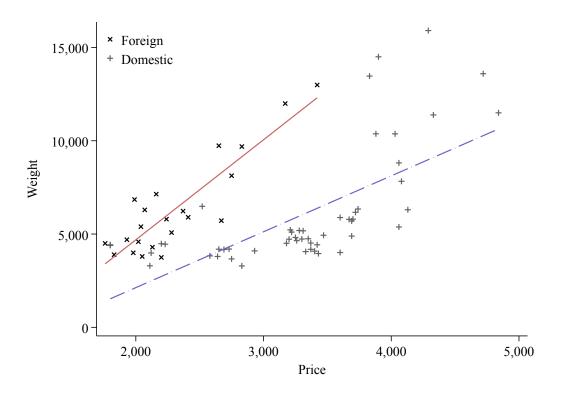


Figure 4: Add Leyend

iii) 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 too small. In the graphs below, I show how to add symbols and adjust the size and symmetry of the graph.

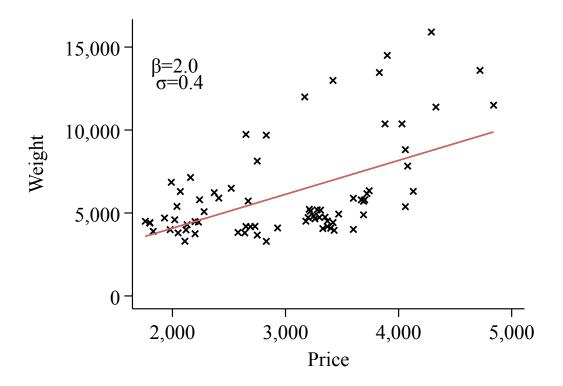


Figure 5: Math Symbols and bigger fonts

```
**** 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)

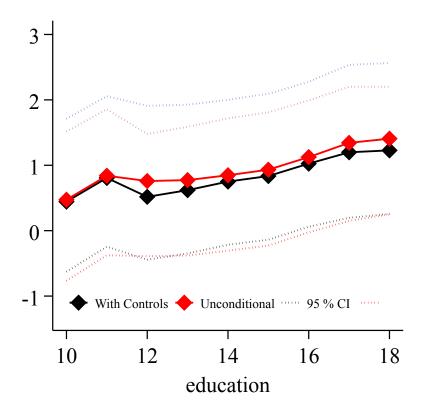
local beta: display %3.1fc b
```

```
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)
```

iv) Plot a set of coefficients

Figure 6: log wages vs hourly wages



```
tempfile lwage lwage_controled
eststo : reg lwage i.educ_fixed
parmest , lab saving('"'lwage'"', replace)
eststo : reg lwage i.educ_fixed age age2 female white married tenure uncov daded momed,
parmest , lab saving('"'lwage_controled'"', replace)
}
# d cr ;
* - Graph
preserve
         clear
       append using $'$"'lwage'"'
       replace var="lwage" if var==""
        append using ""'lwage_controled'",
       replace var="lwage_controled" if var==""
       gen education=subinstr(parm,".educ_fixed","",.)
        destring education, replace force
         # d;
         tw (line min95 max95 education if var=="lwage_controled", lp(dot dot) lc(black))
           (line min95 max95 education if var=="lwage", lp(dot dot) lc(red))
          (con estimate education if var=="lwage_controled" , color(black))
          (con estimate education if var=="lwage" , color(red) ) ,
          legend(label(1 "95 % CI") label(3 "") label(5 "With Controls") label(6 "Unconditional")
          ring(0) row(1) pos(6) size(*0.6)
               region(color(none)) order(5 6 1 3 ) )
           scale(*1.5)
           xsize(5) ysize(5)
       # d cr;
```

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 used them but they seem useful. You can give it a try.

1. Useful Resources

- Ten Guidelines from (Schwabish, 2020)
- World Bank-DIME Tables
- Asjad Naqvi

2. 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

i) 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

ii) Example summary statistics

Table 2: Summary Statistics

(1)					
mean	sd				
6146.0	2912.4				
21.29	5.866				
3.406	0.990				
69					
	mean 6146.0 21.29 3.406				

```
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 (4.753)	25.29 (6.310)	21.29 (5.866)
Repair Record 1978	3.021 (0.838)	4.286 (0.717)	3.406 (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 ;
```

3. Regression outputs

i) 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 \mathbb{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

ii) Adding Fixed Effects Indicators

Table 5: Including FE

	(1)	(2)	(3)
	$Q1_1$	$Q1_2$	$Q1_3$
Weight (lbs.)	2.044***	2.408**	2.431*
	(0.3897)	(0.8011)	(1.0768)
Repair Record 1978			573.405
			(462.2821)
Constant	-6.707	260.709	-1083.997
	(1032.3939)	(1409.9157)	(4445.5871)
Trunk FE	No	Yes	Yes
Other	No	No	Yes
Observations	74	74	69
Adjusted \mathbb{R}^2	0.280	0.290	0.288
F	27.506	•	

Standard errors in parentheses others iclude mpg rep78

^{*} p < 0.05, ** p < 0.01, *** p < 0.001

iii) Create panel A and B from different set of estimates

Sometimes stacking two sets of results, one on top of the other, is very effective to show how your coefficients change when you include a new set of controls like in this example. It can also be useful to show different identification approaches, for example, panel A can have OLS estimates and panel B IV) estimates.

You can automate this process by using the append option and modifying the header and footer for latex in your do file.

	A. Wi	thout Occu	pation Dun	nmies		
	(1)	(2)	(3)	(4)	(5)	(6)
computer	0.172***					0.120***
	[0.0058]					[0.0067]
pencil		0.125***				0.031***
		[0.0061]				[0.0082]
telefon			0.137***			0.042***
			[0.0057]			[0.0082]
calc				0.130***		0.046***
				[0.0057]		[0.0069]
hammer					-0.088***	-0.035***
					[0.0061]	[0.0063]
Constant	1.701***	1.650***	1.667***	1.666***	1.726***	1.751***
	[0.0199]	[0.0196]	[0.0197]	[0.0198]	[0.0211]	[0.0211]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.333	0.320	0.324	0.323	0.312	0.343
	B. Ac	lding Occup	pation Dum	nmies		
computer	0.0827***					0.0682***
	[0.00721]					[0.00736]
pencil		0.0490***				0.00745
		[0.00680]				[0.00830]
telefon			0.0721^{***}			0.0484***
			[0.00682]			[0.00846]
calc				0.0539***		0.0223**
				[0.00631]		[0.00698]
hammer					-0.0206*	-0.0206*
					[0.00806]	[0.00804]
Constant	2.084***	2.068***	2.071***	2.080***	2.094***	2.081***
	[0.0287]	[0.0289]	[0.0288]	[0.0288]	[0.0292]	[0.0291]
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20042	20042	20042	20042	20042	20042
Adjusted \mathbb{R}^2	0.407	0.405	0.407	0.405	0.403	0.410

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

```
global var computer pencil telefon calc hammer
foreach var of global var{
eststo: reg lnw lwage ed exp exp2 female mar , robust
eststo: reg lnw ${var} ed exp exp2 female mar , robust
global var computer pencil telefon calc hammer
foreach var of global var{
  areg lnw 'var' ed exp exp2 female mar, abs(occ) robust
                    matrix V_alt = [vecdiag(e(V)),1]
                    matrix V=diag(V_alt)
                    matrix b=[e(b),1]
                    matrix colnames b= ': colnames e(b)' occ
                    matrix colnames V= ': colnames e(V)' occ
                    erepost b=b V=V ,rename
                    eststo
areg lnw ${var} ed exp exp2 female mar, abs(occ) robust
                    matrix V_alt = [vecdiag(e(V)),1]
                    matrix V=diag(V_alt)
                    matrix b=[e(b),1]
                    matrix colnames b= ': colnames e(b)' occ
                    matrix colnames V= ': colnames e(V)' occ
                    erepost b=b V=V ,rename
eststo
# d;
                    esttab est1 est2 est3 est4 est5 est6 using "Tables/table_ex2.tex", replace
                    b(\%9.3fc) se(\%9.4fc) t(\%9.4fc)
                    t br se par ar2 label nogaps drop() compress nomtitle noobs
                    indicate( "Controls=ed exp exp2 female mar")
                    order( "computer" "pencil" "telefon" "calc" "hammer")
                    prehead(\begin{table}[H]\centering
                    \begin{tabular}{1*{@span}{c}}
                    \hline\hline
                    \mbox{\colored} c) \ A. \ \mbox{\colored} \ \m
                    postfoot(\hline);
                    # d cr;
```

```
esttab est7 est8 est9 est10 est11 est12 using "Tables/table_ex2.tex", append
indicate("Occupation dummies=occ" "Controls=ed exp exp2 female mar")
t br se par ar2 label nogaps drop() compress nomtitle nonumbers
order( "computer" "pencil" "telefon" "calc" "hammer" )
prehead( \multicolumn{@span}{c}{B. Adding Occupation Dummies} \\ )
postfoot(\hline\hline
\multicolumn{@span}{m{1\textwidth}}{\begin{singlespace} \footnotesize Robust Standard Errors in
\label{T:SBpro}
\end{tabular}\end{table})
;
# d cr;
```

iv) 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											
weight	2.044***	2.408**	2.431*	2.044***	2.408**	2.431*	2.044***	2.408**	2.431*	2.044***	2.408**	2.431*
	(0.3897)	(0.8011)	(1.0768)	(0.3897)	(0.8011)	(1.0768)	(0.3897)	(0.8011)	(1.0768)	(0.3897)	(0.8011)	(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	260.709	-1083.997	-6.707	260.709	-1083.997	-6.707	260.709	-1083.997	-6.707	260.709	-1083.997
	(1032.3939)	(1409.9157)	(4445.5871)	(1032.3939)	(1409.9157)	(4445.5871)	(1032.3939)	(1409.9157)	(4445.5871)	(1032.3939)	(1409.9157)	(4445.5871)
Trunk FE	No	Yes	Yes									
N	74	74	69	74	74	69	74	74	69	74	74	69
adj. \mathbb{R}^2	0.280	0.290	0.288	0.280	0.290	0.288	0.280	0.290	0.288	0.280	0.290	0.288
F	27.506			27.506			27.506			27.506		

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}{1*{@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. 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 table.

Another alternative is to use *landscape* option in latex if the table is too wide or *longtable* if it is too long

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 from 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 Patrick Ball Plain text in your workflow

2. Resources for Graphs and 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 Management 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). Another resource is mathcha
- 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 on many different topics

III.CONTENTS

1	Grap	Graphs											
	1.1	Fundamental principles of analytical design											
	1.2	Graphical Excellence											
	1.3	Some	general recommendations	3									
	1.4	Usefull Resources for graphs											
	1.5	Schemes in Stata											
		1.5.1	Examples with different schemes	5									
	1.6												
		1.6.1	Add Scalars	6									
		1.6.2	Leyends	8									
		1.6.3	Math Symbols	9									
		1.6.4	Plot a set of coefficients	10									
2	Export Tables												
	2.1	Useful Resources											
	2.2	2 Basic summary statistics											
		2.2.1	Example t-test	13									
		2.2.2	Example summary statistics	14									
	2.3	Regression outputs											
		2.3.1	Basic	16									
		2.3.2	Adding Fixed Effects Indicators	17									
		2.3.3	Create panel A and B from different set of estimates	18									
		2.3.4	Big Tables and Table Adjustments	21									
3	Usef	ul Addi	itional Resources:	22									
	3.1	Guide	for "Good Practices" to write codes	22									
	3.2	Resour	rces for Graphs and Tables	23									
	3.3	Guides	s to learn R	23									
	3.4	Resour	rces for Stata	23									
	3.5	Miscellaneous											
Re	feren	ces		25									

III.References

- 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.
- Schwabish, J. A. (2021). Better data visualizations: A guide for scholars, researchers, and wonks. Columbia University Press. Retrieved from link
- Tufte, E. R. (2006). Beautiful evidence (Vol. 1). Graphics Press Cheshire, CT.