

```

1 *=====
2 * Intro to data visualization      *
3 * Oscar Torres-Reyna             *
4 * DSS Princeton University       *
5 * http://dss.princeton.edu/      *
6 *=====
7
8 * NOTE, commands should be type either in the command window, see here page 5 here
9 * http://dss.princeton.edu/training/StataTutorial.pdf#page=5
10 * on in a do-file, see here page 9:
11 * http://dss.princeton.edu/training/StataTutorial.pdf#page=9
12
13 * Stata has a color-coded system, see here page 13
14 * http://dss.princeton.edu/training/StataTutorial.pdf#page=13
15
16 ***** Setting working directory
17 *NOTE: If using Mac go to File -> Change Working Directory, and select the folder
18
19 cd "H:"
20
21 ***** Creating a log file.
22 * Anything you type will be saved in the log file. You can open it with any word processor
23
24 log using mylog.log, replace
25
26 * Data transfer other than Excel see
27 * http://dss.princeton.edu/training/StatTransfer.pdf
28
29 * Opening a Stata data file
30
31 use "http://www.princeton.edu/~otorres/wdipol.dta", clear
32
33 * See the data
34
35 browse
36
37 ***** Getting to know your data
38
39 describe
40 summarize
41
42 ***** Line graphs
43
44 line unemp unempf unempm year if country=="United States"
45 summarize unemp unempf unempm
46 replace unemp=. if unemp==0
47 replace unempf=. if unempf==0
48 replace unempm=. if unempm==0
49 summarize unemp unempf unempm
50 line unemp unempf unempm year if country=="United States"
51
52 twoway line unemp unempf unempm year if country=="United States", ///
53     title("Unemployment rate in the US, 1980-2012") ///
54     legend(label(1 "Total") label(2 "Females") label(3 "Males")) ///
55     lpattern(solid dash dot) ///
56     ytitle("Percentage")
57
58 twoway connected unemp unempf unempm year if country=="United States", ///
59     title("Unemployment rate in the US, 1980-2012") ///
60     legend(label(1 "Total") label(2 "Females") label(3 "Males")) ///
61     msymbol(circle diamond square) ///
62     ytitle("Percentage")
63
64 twoway connected unemp year if country=="United States" | ///
65     country=="United Kingdom" | ///
66     country=="Australia" | ///
67     country=="Qatar", ///
68     by(country, title("Unemployment")) ///
69     msymbol(circle_hollow)
70

```

```

71 twoway (connected unemp year if country=="United States", msymbol(diamond_hollow)) ///
72     (connected unemp year if country=="United Kingdom", msymbol(triangle_hollow)) ///
73     (connected unemp year if country=="Australia", msymbol(square_hollow)) ///
74     (connected unemp year if country=="Qatar", ///
75     title("Unemployment") ///
76     msymbol(circle_hollow) ///
77     legend(label(1 "USA") label(2 "UK") label(3 "Australia") label(4 "Qatar")))
78
79 twoway connected gdppc year if gdppc>40000, by(country) msymbol(diamond)
80
81
82 bysort year: egen gdppc_mean=mean(gdppc)
83 bysort year: egen gdppc_median=median(gdppc)
84
85 twoway connected gdppc gdppc_mean year if country=="United States" | ///
86     country=="United Kingdom" | ///
87     country=="Australia" | ///
88     country=="Qatar", ///
89     by(country, title("GDP pc (PPP, 2005=100)")) ///
90     legend(label(1 "GDP-PC") label(2 "Mean GDP-PC")) ///
91     msymbol(circle_hollow)
92
93 help twoway line
94 help twoway connected
95
96 ***** Graph markers
97 palette symbolpalette
98 palette linepalette
99 palette color green
100 help palette
101
102 ***** Bar graphs
103
104 graph hbar (mean) gdppc /*Mean is the default*/
105 graph hbar (mean) gdppc, over(country, sort(1) descending)
106 graph hbar (mean) gdppc (median) gdppc if gdppc>40000, ///
107     over(country, sort(1) descending label(labsize(*1))) ///
108     legend(label(1 "GDPpc (mean)") label(2 "GDPpc (median)"))
109
110 help graph bar
111
112 ***** Box plots
113
114 * Need to recode polity2
115 recode polity2 (-10/-6=1 "Autocracy") ///
116     (-5/6=2 "Anocracy") ///
117     (7/10=3 "Democracy") ///
118     (else=.), ///
119     gen(regime) label(polity_rec)
120
121 tab regime /* Frequency*/
122 tab regime, nolabel /* See numeric values*/
123 tab country regime /* Cross tabulations */
124 tab country regime, row /* Adding percent per row */
125
126 help tab
127 * http://dss.princeton.edu/training/StataTutorial.pdf
128
129 graph hbox gdppc
130 graph box gdppc if gdppc<40000
131 graph box gdppc, over(regime) yline(4517.94) marker(1,mlabel(country))
132
133 help graph box
134
135 ***** Scatterplots
136 * scatter y x
137
138 scatter import export
139
140 twoway scatter import export || scatter import export if export>1000000, mlabel(country)
141 legend(off)

```

```

140
141 twoway (scatter import export, ytitle("Imports") xtitle("Exports")) ///
142 (scatter import export if export>1000000, mlabel(country) legend(off)) ///
143 (lfit import export, note("Constant values, 2005, millions US$"))
144
145 bysort year: egen gdppc_mean=mean(gdppc)
146
147 twoway (scatter gdppc year, jitter(13)) ///
148 (connected gdppc_mean year, msymbol(diamond)) , xlabel(1980(1)2012, angle(90))
149
150 help twoway scatter
151
152 ***** Scatterplot matrix
153
154 graph matrix gdppc unemp unempf unempm export import trade polity2, maxis(ylabel(none)
155 xlabel(none))
156
157 graph matrix gdppc unemp unempf unempm export import trade polity2, half maxis(ylabel(none)
158 xlabel(none))
159
160 help graph matrix
161
162 ***** Histograms
163
164 hist gdppc /* Shows density*/
165 hist gdppc, frequency /*Shows frequency*/
166 hist gdppc, kdensity /* Combo histogram and density plot */
167 hist gdppc, kdensity normal /* Adding a normal curve */
168 hist gdppc, kdensity normal bin(20)
169 hist gdppc if country=="United States" | country=="United Kingdom", bin(10) by(country)
170 twoway hist gdppc if country=="United States", bin(10) || ///
171 hist gdppc if country=="United Kingdom", bin(10) ///
172 fcolor(none) lcolor(black) legend(label(1 "USA") label(2 "UK"))
173
174 help hist
175
176 ***** Setup panel data
177 * See http://dss.princeton.edu/training/Panell01.pdf
178
179 xtset country year /*Gives an error, 'country' is string*/
180 encode country, gen(country1) /*Assign numeric value to strings*/
181 xtset country1 year /*No error, 'country1' is coded variable*/
182
183 xtline gdppc
184 xtline gdppc if gdppc>39000, overlay
185
186 help xtline
187
188 ***** Combining graphs
189
190 graph drop _all /*Drop graphics saved in memory*/
191 hist gdppc if country=="United States", name(gdppc, replace)
192 line unemp year if country=="United States", name(unemp, replace)
193 graph combine gdppc unemp, col(1)
194
195 help graph combine
196
197 ***** Scatterplots with linear fit and confidence intervals
198
199 use "http://dss.princeton.edu/training/students.dta", clear
200 twoway (lfitci sat age) ///
201 (scatter sat age, mlabel(lastname)mlabv(position) jitter(21)), title("SAT scores by
202 age") ytitle("Sat")
203
204 * Without confidence intervals
205 twoway (lfit sat age) ///
206 (scatter sat age, mlabel(lastname)mlabv(position) jitter(21)), title("SAT scores by
207 age") ytitle("Sat")
208
209 help twoway lfit
210 help twoway lfitci
211
212 ***** Plotting categorical variables
213
214 ***** Mosaic plots (a.k.a spineplots)
215 * May need to install it, type:
216 ssc install spineplots
217
218 use "http://dss.princeton.edu/training/students.dta", clear
219 encode gender, gen(gender1) /* Assign numeric values to categories in string format */
220 encode major, gen(major1)
221 spineplot gender1 major1
222 bysort gender1 major1: gen gendermajor = _N
223 spineplot gender1 major1, text(gendermajor)
224 spineplot gender1 major1, percent bar1(bcolor(yellow)) bar2(bcolor(green)) text(gendermajor)
225
226 * See the graphs here:
227 * http://www.princeton.edu/~otorres/mosaic1.pdf
228 * http://www.princeton.edu/~otorres/mosaic2.pdf
229
230 ***** Using catplot, see (pages 53-55):
231
232 * http://dss.princeton.edu/training/StataTutorial.pdf#page=53
233 * http://dss.princeton.edu/training/StataTutorial.pdf#page=54
234 * http://dss.princeton.edu/training/StataTutorial.pdf#page=55
235
236 * Chernoff faces
237 * Few cases, each face is a row case.
238
239 use "http://www.princeton.edu/~otorres/chernoff.dta", clear
240 ssc install chernoff /*User-written command, need to install*/
241 chernoff, hdark(gdppc) bdens(trade) nose(unemp) mcurv(polity2) order(gdppc) ilabel(country)
242
243 * See the graph here:
244 * http://www.princeton.edu/~otorres/chernoff.pdf
245
246 * For more complex graph examples (code included) see here
247 * http://statistics.ats.ucla.edu/stat/stata/library/GraphExamples/default.htm
248
249 * Do not forget to close the log
250
251 log close
252
253
254
255
256
257
258

```

```

208 twoway (lfitci sat age) ///
209 (scatter sat age, mlabel(lastname)mlabv(position) jitter(21)), title("SAT scores by
210 age") ytitle("Sat")
211
212 * Without confidence intervals
213 twoway (lfit sat age) ///
214 (scatter sat age, mlabel(lastname)mlabv(position) jitter(21)), title("SAT scores by
215 age") ytitle("Sat")
216
217 help twoway lfit
218 help twoway lfitci
219
220 ***** Plotting categorical variables
221
222 ***** Mosaic plots (a.k.a spineplots)
223 * May need to install it, type:
224 ssc install spineplots
225
226 use "http://dss.princeton.edu/training/students.dta", clear
227 encode gender, gen(gender1) /* Assign numeric values to categories in string format */
228 encode major, gen(major1)
229 spineplot gender1 major1
230 bysort gender1 major1: gen gendermajor = _N
231 spineplot gender1 major1, text(gendermajor)
232 spineplot gender1 major1, percent bar1(bcolor(yellow)) bar2(bcolor(green)) text(gendermajor)
233
234 * See the graphs here:
235 * http://www.princeton.edu/~otorres/mosaic1.pdf
236 * http://www.princeton.edu/~otorres/mosaic2.pdf
237
238 ***** Using catplot, see (pages 53-55):
239
240 * http://dss.princeton.edu/training/StataTutorial.pdf#page=53
241 * http://dss.princeton.edu/training/StataTutorial.pdf#page=54
242 * http://dss.princeton.edu/training/StataTutorial.pdf#page=55
243
244 * Chernoff faces
245 * Few cases, each face is a row case.
246
247 use "http://www.princeton.edu/~otorres/chernoff.dta", clear
248 ssc install chernoff /*User-written command, need to install*/
249 chernoff, hdark(gdppc) bdens(trade) nose(unemp) mcurv(polity2) order(gdppc) ilabel(country)
250
251 * See the graph here:
252 * http://www.princeton.edu/~otorres/chernoff.pdf
253
254 * For more complex graph examples (code included) see here
255 * http://statistics.ats.ucla.edu/stat/stata/library/GraphExamples/default.htm
256
257 * Do not forget to close the log
258
259 log close
260
261
262
263
264
265
266

```