

```

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, over(country, sort(1) descending label(labsize(*0.5)))
107 graph hbar (mean) gdppc (median) gdppc if gdppc>40000, ///
108     over(country, sort(1) descending label(labsize(*1))) ///
109     legend(label(1 "GDPpc (mean)" label(2 "GDPpc (median)"))
110
111 help graph bar
112
113 ***** Box plots
114
115 * Need to recode polity2
116 recode polity2 (-10/-6=1 "Autocracy") ///
117     (-5/6=2 "Anocracy") ///
118     (7/10=3 "Democracy") ///
119     (else=.), ///
120     gen(regime) label(polity_rec)
121
122 tab regime /* Frequency*/
123 tab regime, nolabel /* See numeric values*/
124 tab country regime /* Cross tabulations */
125 tab country regime, row /* Adding percent per row */
126
127 help tab
128 * http://dss.princeton.edu/training/StataTutorial.pdf
129
130 graph hbox gdppc
131 graph hbox gdppc if gdppc<40000
132 graph box gdppc, over(regime) yline(4517.94) marker(1,mlabel(country))
133
134 help graph box
135
136 ***** Scatterplots
137 * scatter y x
138
139 scatter import export
140
141 twoway scatter import export || scatter import export if export>1000000, mlabel(country)
142 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 if country=="United States" | country=="United Kingdom", bin(10) by(country)
169 twoway hist gdppc if country=="United States", bin(10) || ///
170 hist gdppc if country=="United Kingdom", bin(10) ///
171 fcolor(none) lcolor(black) legend(label(1 "USA") label(2 "UK"))
172
173 help hist
174
175 ***** Setup panel data
176 * See http://dss.princeton.edu/training/Panell01.pdf
177
178 xtset country year /*Gives an error, 'country' is string*/
179 encode country, gen(country1) /*Assign numeric value to strings*/
180 xtset country1 year /*No error, 'country1' is coded variable*/
181
182 xtline gdppc
183 xtline gdppc if gdppc>39000, overlay
184
185 help xtline
186
187 ***** Combining graphs
188
189 graph drop _all /*Drop graphics saved in memory*/
190 hist gdppc if country=="United States", name(gdppc, replace)
191 line unemp year if country=="United States", name(unemp, replace)
192 graph combine gdppc unemp, col(1)
193
194 help graph combine
195
196 ***** Scatterplots with linear fit and confidence intervals
197
198 use "http://dss.princeton.edu/training/students.dta", clear
199 twoway (lfitci sat age) ///
200 (scatter sat age, mlabel(lastname)), title("SAT scores by age") ytitle("Sat")
201
202 * Changing position
203 generate position=3
204 replace position=6 if lastname=="DOE01"
205 replace position=6 if lastname=="DOE10"
206 replace position=12 if lastname=="DOE14"
207 replace position=12 if lastname=="DOE29"

```

```

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
231 bysort gender1 major1: gen gendermajor = _N
232 spineplot gender1 major1, text(gendermajor)
233 spineplot gender1 major1, percent bar1(bcolor(yellow)) bar2(bcolor(green)) text(gendermajor)
234
235 * See the graphs here:
236 * http://www.princeton.edu/~otorres/mosaic1.pdf
237 * http://www.princeton.edu/~otorres/mosaic2.pdf
238
239 ***** Using catplot, see (pages 53-55):
240
241 * http://dss.princeton.edu/training/StataTutorial.pdf#page=53
242 * http://dss.princeton.edu/training/StataTutorial.pdf#page=54
243 * http://dss.princeton.edu/training/StataTutorial.pdf#page=55
244
245 * Chernoff faces
246 * Few cases, each face is a row case.
247
248 use "http://www.princeton.edu/~otorres/chernoff.dta", clear
249 ssc install chernoff /*User-written command, need to install*/
250 chernoff, hdark(gdppc) bdens(trade) nose(unemp) mcurv(polity2) order(gdppc) ilabel(country)
251
252 * See the graph here:
253 * http://www.princeton.edu/~otorres/chernoff.pdf
254
255 * For more complex graph examples (code included) see here
256 * http://statistics.ats.ucla.edu/stat/stata/library/GraphExamples/default.htm
257
258 * Do not forget to close the log
259
260 log close

```