

# Introduction to Stata

## Lecture VII

Tomas R. Martinez

UC3M

September, 2019

- “Mankind invented a system to cope with the fact that we are so intrinsically lousy at manipulating numbers. It’s called the graph.”  
Charlie Munger
- All right, let’s do some graphs

- Stata can produce many kind of graphs
- Main command is **graph** combined with some chart type
- scatterplots, lines, bar charts, pie charts, densities...
- Each one has its own details
- We will focus in the general commands and the most common ones
- **USE AND ABUSE THE HELP FUNCTION!**

- Most are intuitive:
  - **Titles:** `title("yourtitle")`, `subtitle("subtitle")`, `note("A note here")`
  - **Legends:** `legend(suboption)`, `suboptions: label(1 var1)...`
  - **Axis:** `xtitle("name")` / `ytitle()`, `xlabel()/ylabel()`, `xscale()`, `yscale()...`
  - **Background:** `graphregion(color(white))...`
  - **Same graph for different groups:** `by(varname)`
  - **Saving:** saving saves the graph in `.gph` useful if you want to call `graph combine` and have multiple plots in figure
  - **Exporting figure:** `graph export namegraph.extension`
- You can also generate a “raw” graph and then play with the graph editor menu

# Twoway graphs

- Twoway produces a very intuitive two axis (x-y) plot
- Scatter, lfit, lfitci, qfit, line
- **Scatter:** Very useful to identify outliers in your data
- **lfit / qfit / lfitci:** Basically a linear/quadratic fit in your data
- **line:** plot a line
- When plotting a line your data must be sorted in the X variable!!!
- **Example:** cps05.dta with occupation information

# Plotting the distribution

- Sometimes you want to have a look at the distribution
- One way: **kdensity**
- Have a look at the option to smooth your distribution: **bwidth**
- Or you can plot histograms: it is key to select the right bin size!
- You can also manually create a CDF: `cumul varname, generate(cdfvar)`
- Then do a line plot: `twoway line cdfvar varname, sort`

- Summary graphs: **bar** and **pie**
- **Example:** bar (mean) incwage, over(educ)
- **Example:** pie (mean) incwage, over(educ)
- Others:
  - **rcap:** a plot with two y variables, useful create the “spikes” of confident intervals, max/min, etc.
  - **boxplot:** useful to examine the distribution over some group
  - **dot:** for me it is useless