

graphics

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Friday 12th October, 2018 12:43

outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

postestimation graphs [qm2]

bonus

time to check on grades

- ◇ see predicted letter grade on sakai

why? graphics and data management?

- ◇ i was emphasizing importance of understanding your data
- ◇ graphics is the best way to visualize/understand data
- ◇ data are numbers, usually many and in a matrix
 - graphics is a great tool to allow humans to comprehend those many numbers
 - if you look at numbers you will be slower in understanding them than when looking at a picture
- ◇ pictures are not less “scientific” than numbers!
- ◇ actually the very best journals mostly do graphs!!!
 - nature, science, pnas, psy sci, etc

outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

postestimation graphs [qm2]

bonus

references/links

- ◇ Tufte (multiple) <http://www.edwardtufte.com/tufte/>
- ◇ Kosslyn “Clear and to The Point”
<http://www.amazon.com/Clear-Point-Psychological-Principles-Presentation/dp/0195320697>
- ◇ <https://www.stata.com/bookstore/visual-guide-to-stata-graphics/>

be simple: avoid clutter in general

- ◇ everything should be made as simple as possible,
 - but not one bit simpler
- ◇ avoid padding: present only data needed for a specific purpose
- ◇ avoid clutter: eg single graph must only present the data that are highly related and must be compared
- ◇ put data into appendix if it is not very relevant but may be useful
 - people looking for extra information will find it
 - people interested in the main story will not get distracted

avoid visual clutter

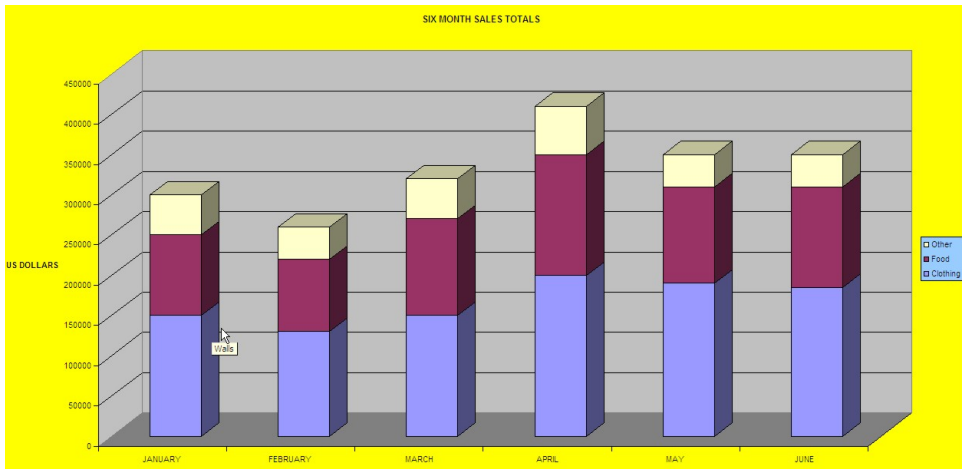
- ◇ all parts and attr of graph must be meaningful
- ◇ including:
 - shades
 - colors
 - decoration

chartjunk

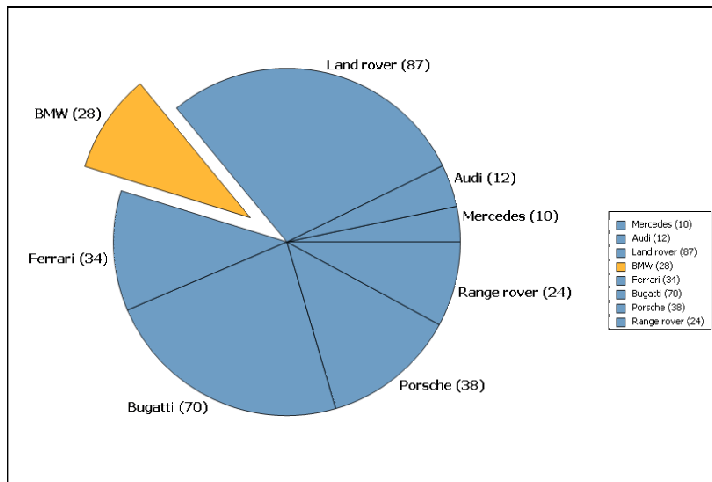
- ◇ *The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, it is all non-data-ink or redundant data-ink, and it is often chartjunk.*

Edward Tufte "The Visual Display of Quantitative Information"

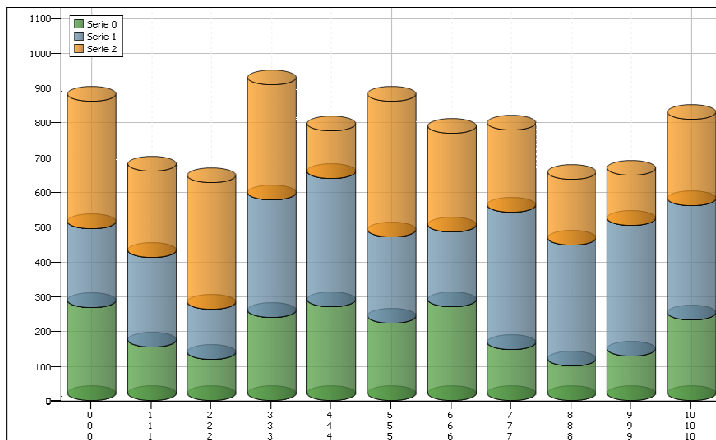
- ◇ examples follow:
 - chartjunk or "business graphs"
 - good fancy graphs



Exploded Pie Slice Chart



Cylinder Chart



not chartiunk (the economist)

Extra leisure time enjoyed by men compared with women
2006*, minutes per day



Source: OECD

*Or latest available

not chartjunk (the economist)

Time spent eating and sleeping

2006, minutes per day



Source: OECD



outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

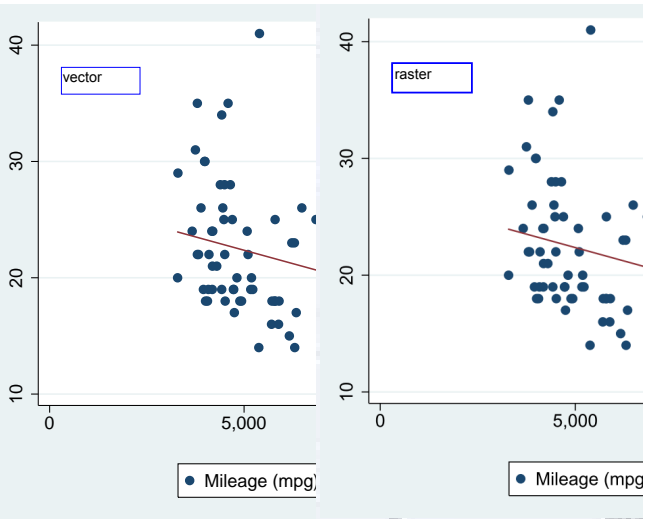
postestimation graphs [qm2]

bonus

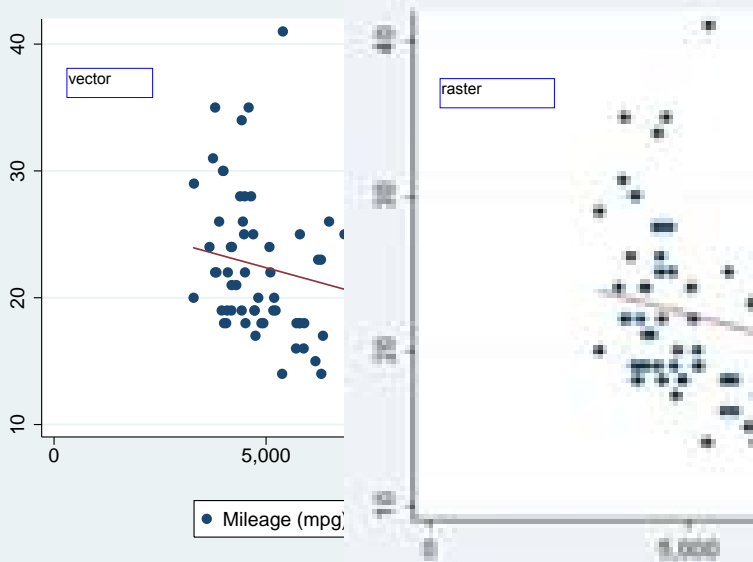
approach

- ◇ we will see lots of examples and lots of code
- ◇ much of it will be irrelevent to you
- ◇ try to pick graphs/chunks of code that you may use for your own research

vector vs raster , 100% zoom



vector vs raster, 500% zoom



stata graph formats

- ◇ raster: PNG, TIF
- ◇ vector: **PS**, **EPS**, GPH, PDF, WMF, **EMF**
- ◇ stata's GPH format good for editing
- ◇ hi-tech/linux: use PS, EPS, PDF
- ◇ low-tech/windows: use EMF

strategy

- ◇ complicated code – many graph options;
easier to figure out options either with
 - examples/galleries or
 - GUI (yes! its ok)
- ◇ learn graph options (for power/bored users only)

help graph intro

galleries

◇ useful galleries

- ▶ <http://www.stata.com/support/faqs/graphics/gph/statagraphs.html>
- ▶ <http://www.stata.com/bookstore/visual-guide-to-stata-graphics>

twoway

- ◇ `tw` is a generic stata command for graphics
- ◇ it can produce almost any graph in stata and overlays of graphs
- ◇ `tw` combines simple graphs, e.g. `scatter`, `line`

twoway GUI

- ◇ graphics-twoway; 'create' and 'accept' when done
- ◇ then overlay another graph with 'create' 'accept'
- ◇ then change options using 'if/in', 'Y axis' etc
- ◇ finally hit 'submit' and experiment further
- ◇ once happy with the result save code in the do-file
- ◇ GUI

graph combine

- ◇ again, to overlay graphs use `twoway`
- ◇ to put (different) graphs next to each other `gr combine`
 - (note: many stata graphs do sugraphs!)
- ◇ `dofile`

outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

postestimation graphs [qm2]

bonus

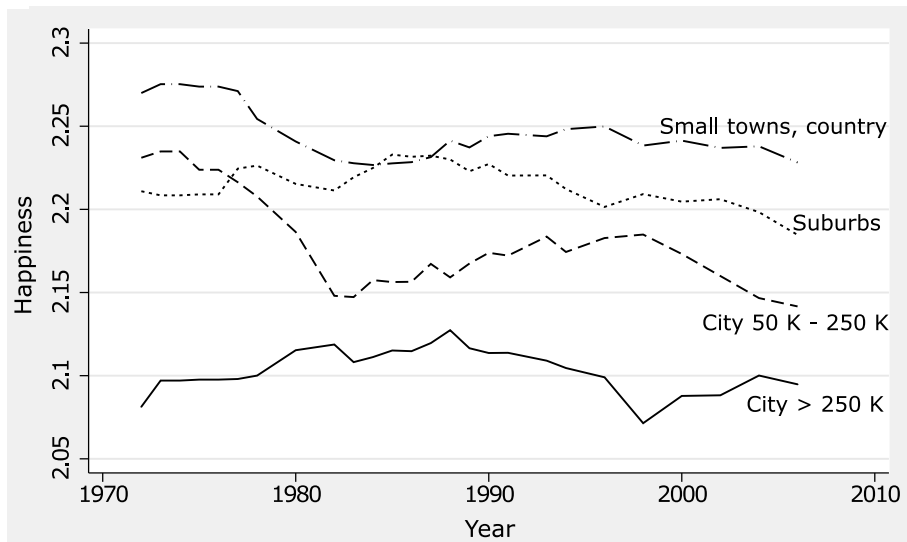
howto graph it?

- ◇ again, look at the galleries to get inspired!
- ◇ but also consider this:
 - continuous
 - distribution
 - categorical, continuous, summary statistics
- ◇ and we will break the discussion in this way

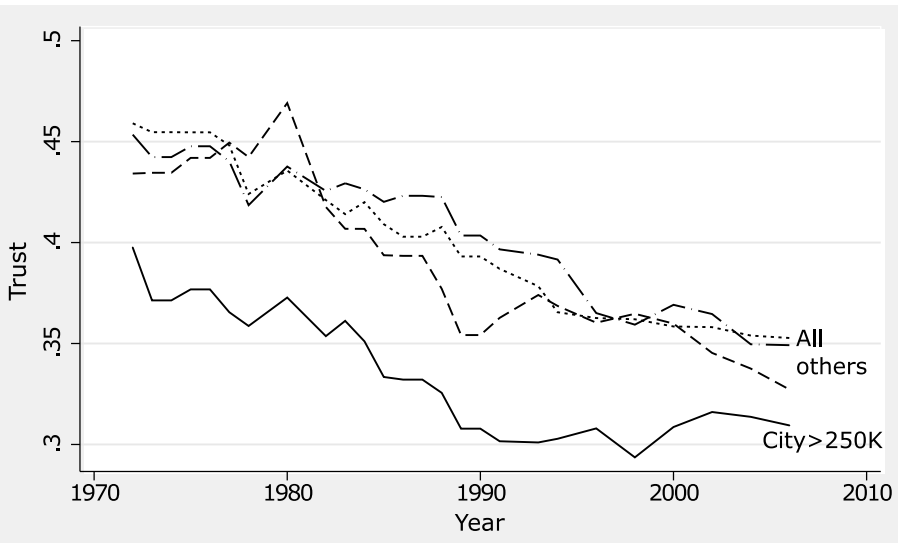
continuous vars

- ◇ continuous vars: scatterplots `scatter`
- ◇ time series: lineplots `line`
- ◇ combine a bunch of scatter plots using `graph matrix`
- ◇ show range `tw area` `tw rcap`

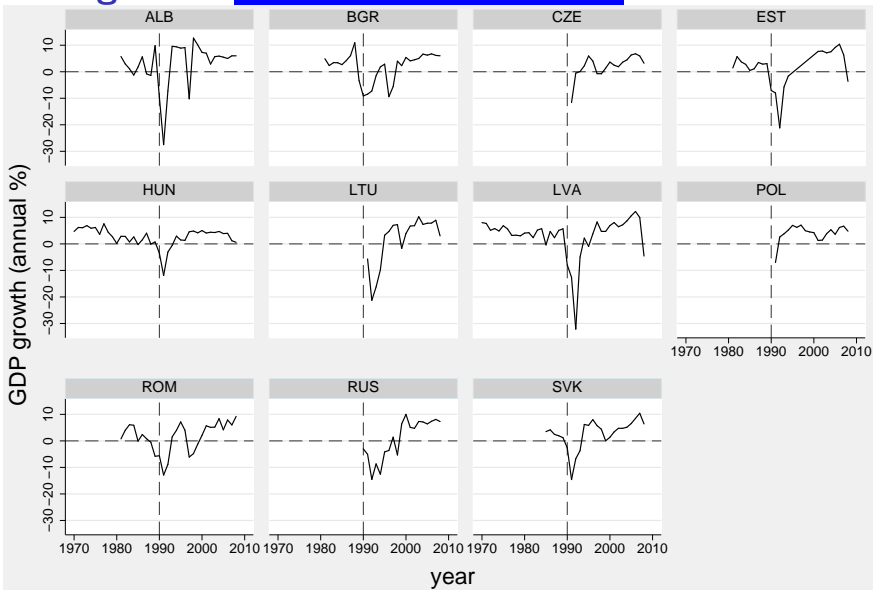
line ex 1



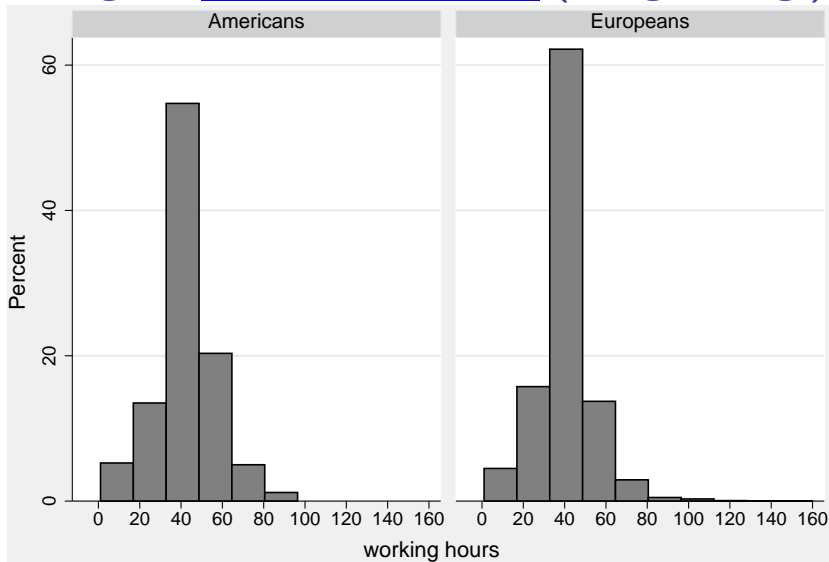
line ex 2



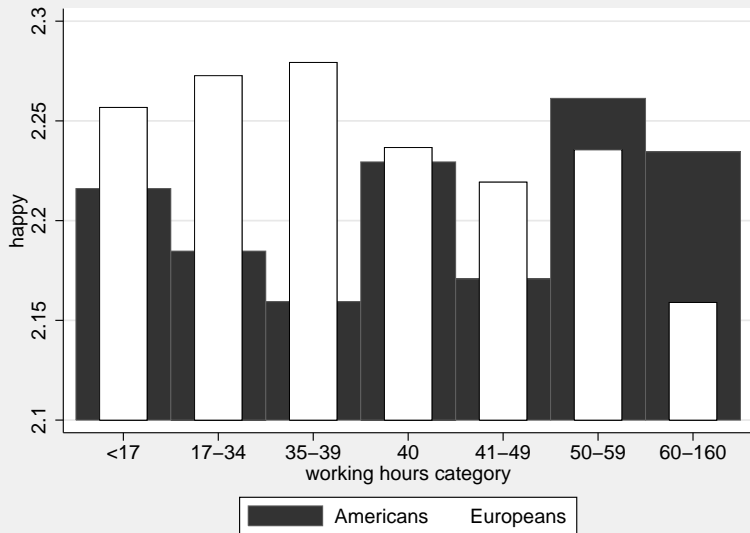
line gr ex 3 dofile: continous vars



dist gr ex **dofile: distribution** (sub-gr of 1 gr)



dofile: categorical, continuous, summary statistics



outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

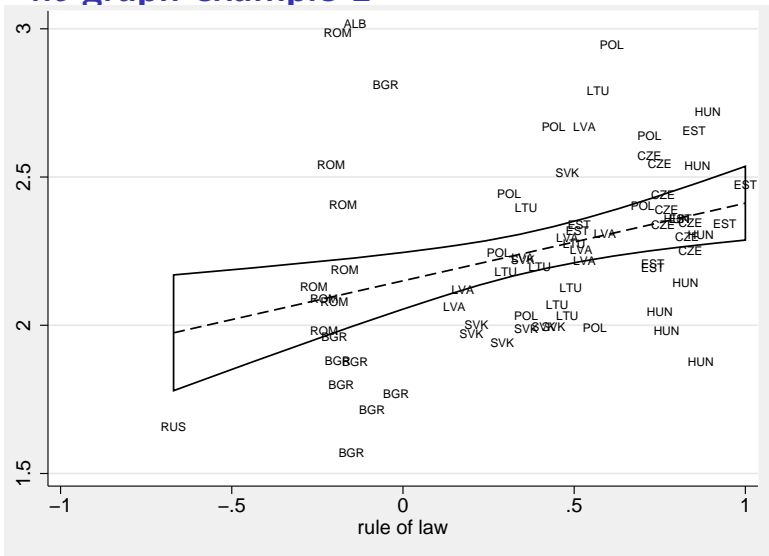
postestimation graphs [qm2]

bonus

why graphing postestimation results

- ◇ graphs provide more information than numbers
- ◇ graphs show the relationship better than numbers
- ◇ again, graphs are especially good for presentations

fit graph example 2



graphs 1

- ◇ `regplot` plots predicted and actual values also for nonlinear models
- ◇ `ecplot` plots estimates and confidence intervals
- ◇ `do-file`

graphs 2

- ◇ `postgr3` is very useful for categorical dependent variable models (multinomial, binary, ordinal)
- ◇ if your dependent variable is categorical , graph results because estimates are not constant
- ◇ `do-file`

good practices

- ◇ do not use graphs if they take up more space than text or numbers
- ◇ avoid graph padding and within-graph data padding
- ◇ always display measures of uncertainty, typically 95%CI
 - eg `marginsplot` `ciplot`

outline

graphics: some theory (Tufte)

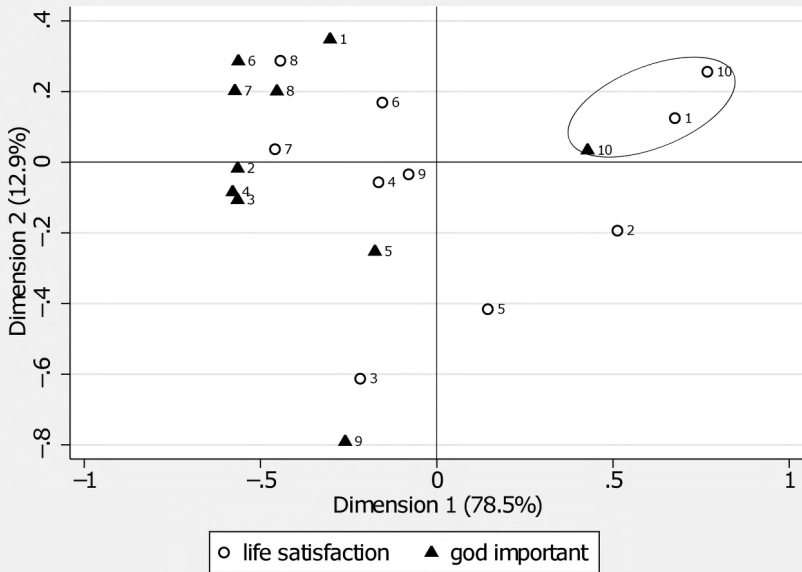
stata graph basics

descriptive graphs

postestimation graphs [qm2]

bonus

ca biplot graph example (do-file)



coordinates in symmetric normalization

ca

- ◇ ca is really nice !
- ◇ checkout
 - help ca
 - help mca

chernoff faces (dofile)

- ◇ <http://healthyalgorithms.com/2012/11/09/baby-faces-and-chernoff-faces/>
- ◇ <http://healthyalgorithms.com/2012/11/12/dataviz-in-python-chernoff-faces-with-matplotlib/>

maps: use python

- ◇ <http://sensitivecities.com/so-you-d-like-to-make-a-map-using-python-EN.html>
- ◇ <http://wrobstory.github.io/2013/04/python-maps-choropleth.html>
- ◇ <http://www.qgis.nl/2013/08/13/python-script-to-generate-series-of-maps/?lang=en>