graphics

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Thursday 3rd March, 2022 16:01

outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

postestimation graphs [qm2]

bonus

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 allows humans to comprehend those many numbers
  - if you look at numbers you will be slower in understanding
- pictures are not less "scientific" than numbers!
- actually the very best journals mostly do garphs!!!
  - nature, science, pnas, psy sci, etc
- dofile:checking (sometimes use tables/numbers)

outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

postestimation graphs [qm2]

bonus

make it interesting!

- The greatest value of a picture is when it forces us to notice what we never expected to see.

references/links

- Tufte (multiple) http://www.edwardtufte.com/tufte/
- Kosslyn "Clear and to The Point"
  http://www.amazon.com/
  Clear-Point-Psychological-Principles-Presentati
  dp/0195320697
- https://www.stata.com/bookstore/visual-guide-to-stata-graphics/
- new data sciency stuff:
  - https://chart.guide
  - https://datavizproject.com

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 distraced

avoid visual clutter

- all parts and attr of graph must be meaningful

  incl:

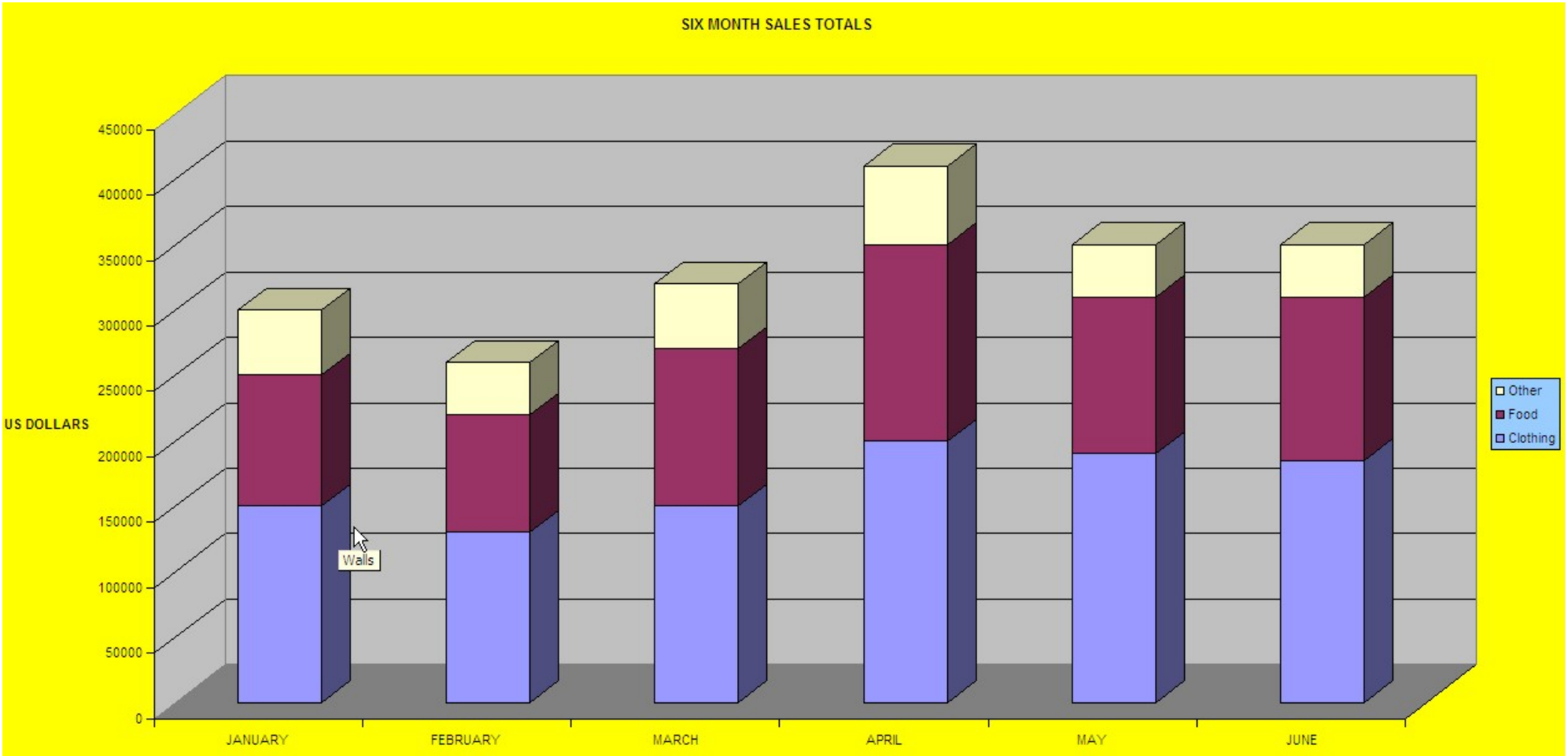
- shades
- colors
- decoration

chartjunk

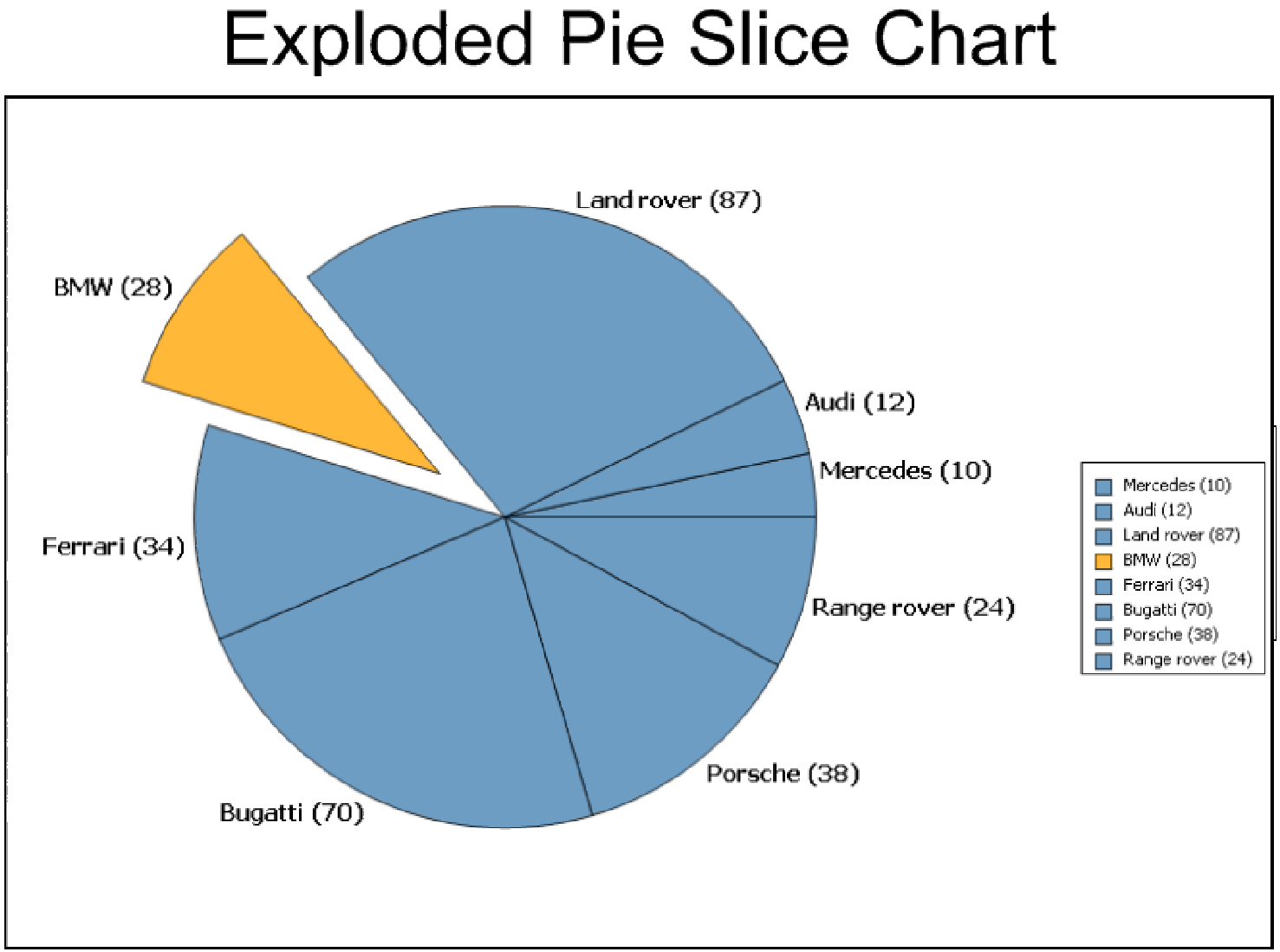
The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, 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

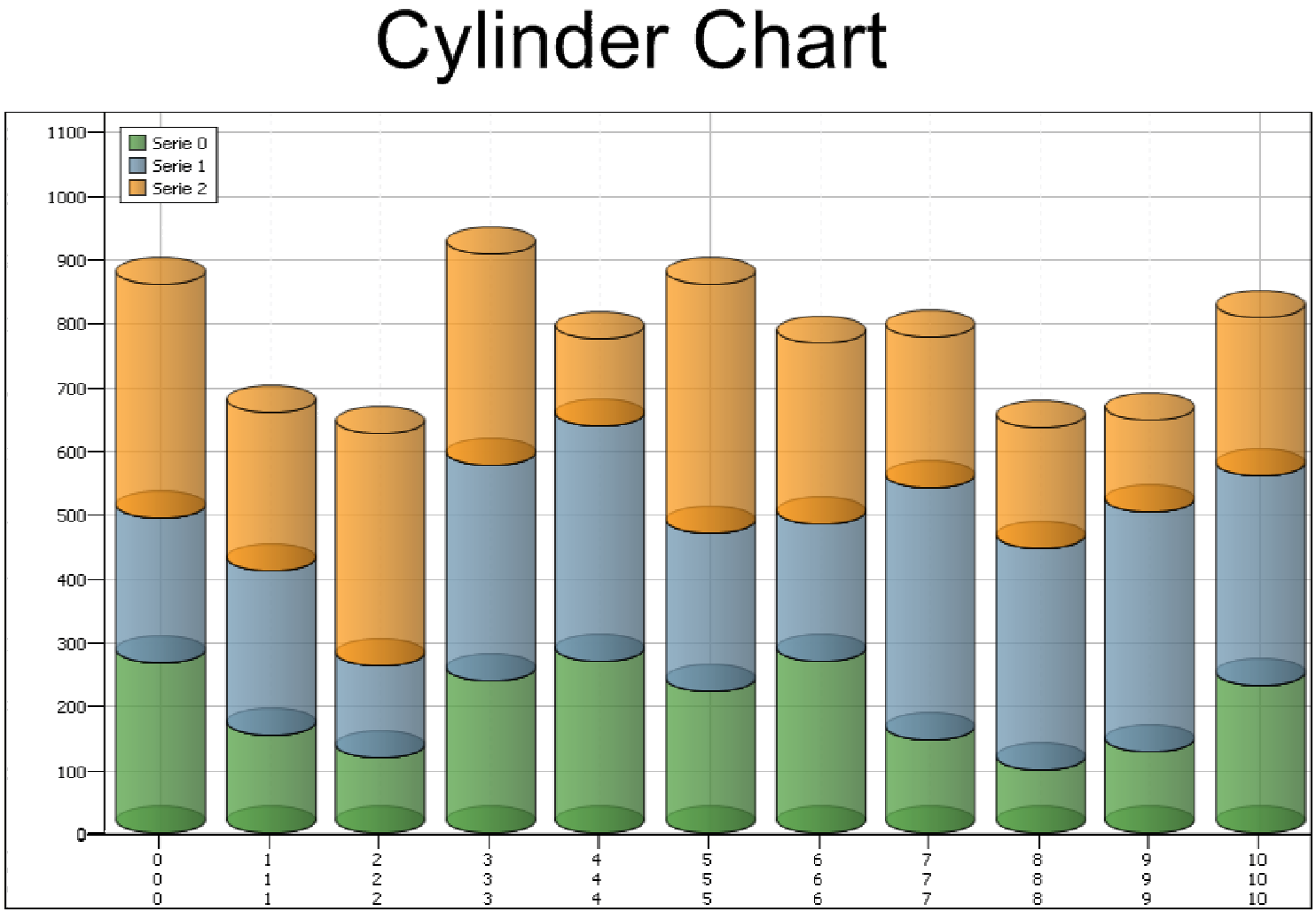
chartjunk


SIX MONTH SALES TOTALS
US DOLLARS
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
Walls
Other
Food
Clothing

chartjunk

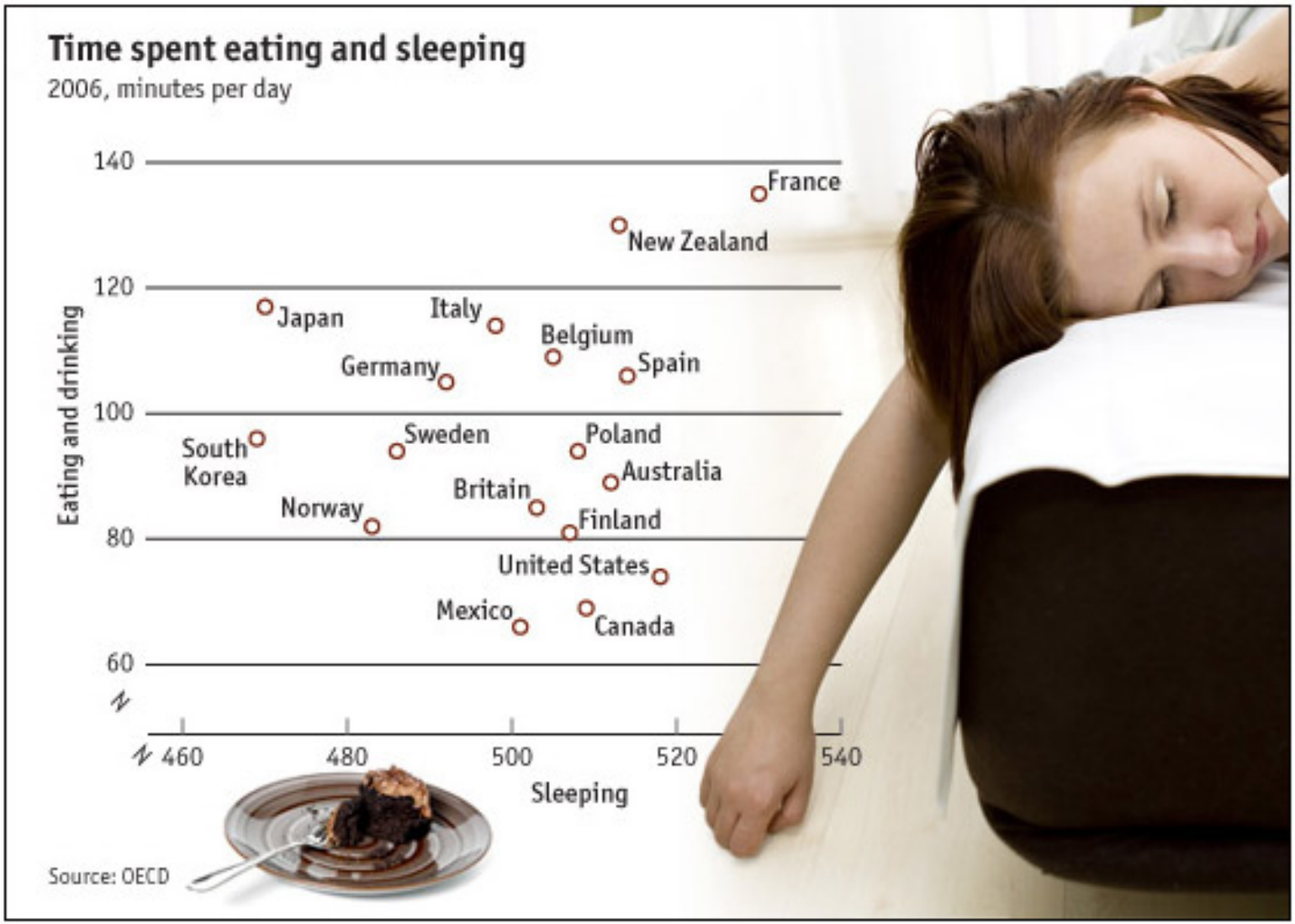

Exploded Pie Slice Chart
Land rover (87)
BMW (28)
Audi (12)
Mercedes (10)
Ferrari (34)
Range rover (24)
Porsche (38)
Bugatti (70)
Mercedes (10)
Audi (12)
Land rover (87)
BMW (28)
Ferrari (34)
Bugatti (70)
Porsche (38)
Range rover (24)

chartjunk


Cylinder Chart
Serie 0
Serie 1
Serie 2
1100
1000
900
800
700
600
500
400
300
200
100
0
0 0 0
1 1 1
2 2 2
3 3 3
4 4 4
5 5 5
6 6 6
7 7 7
8 8 8
9 9 9
10 10 10

not chartjunk (the economist)

not chartjunk (the economist)



outline

graphics: some theory (Tufte)

stata graph basics

descriptive graphs

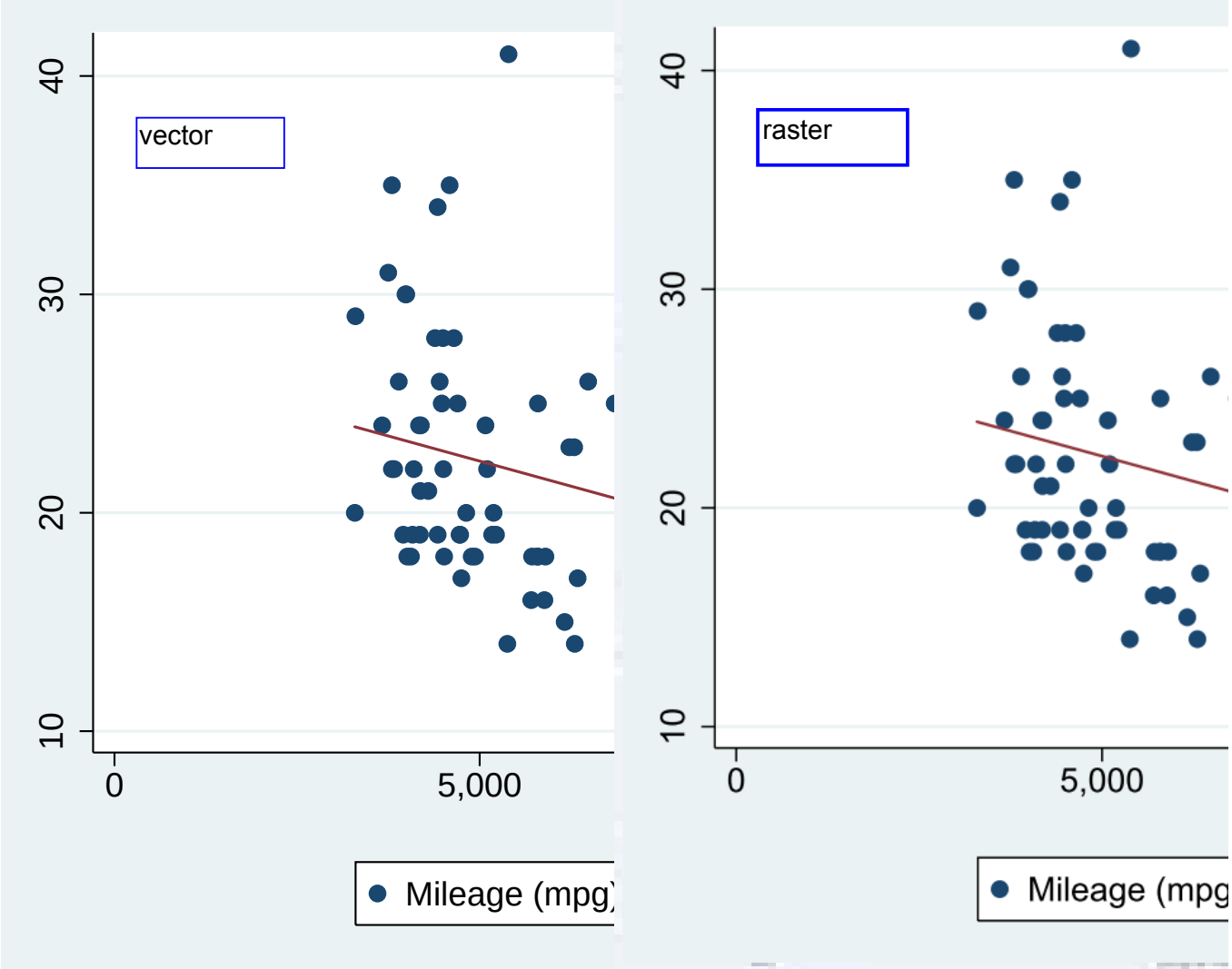
postestimation graphs [qm2]

bonus

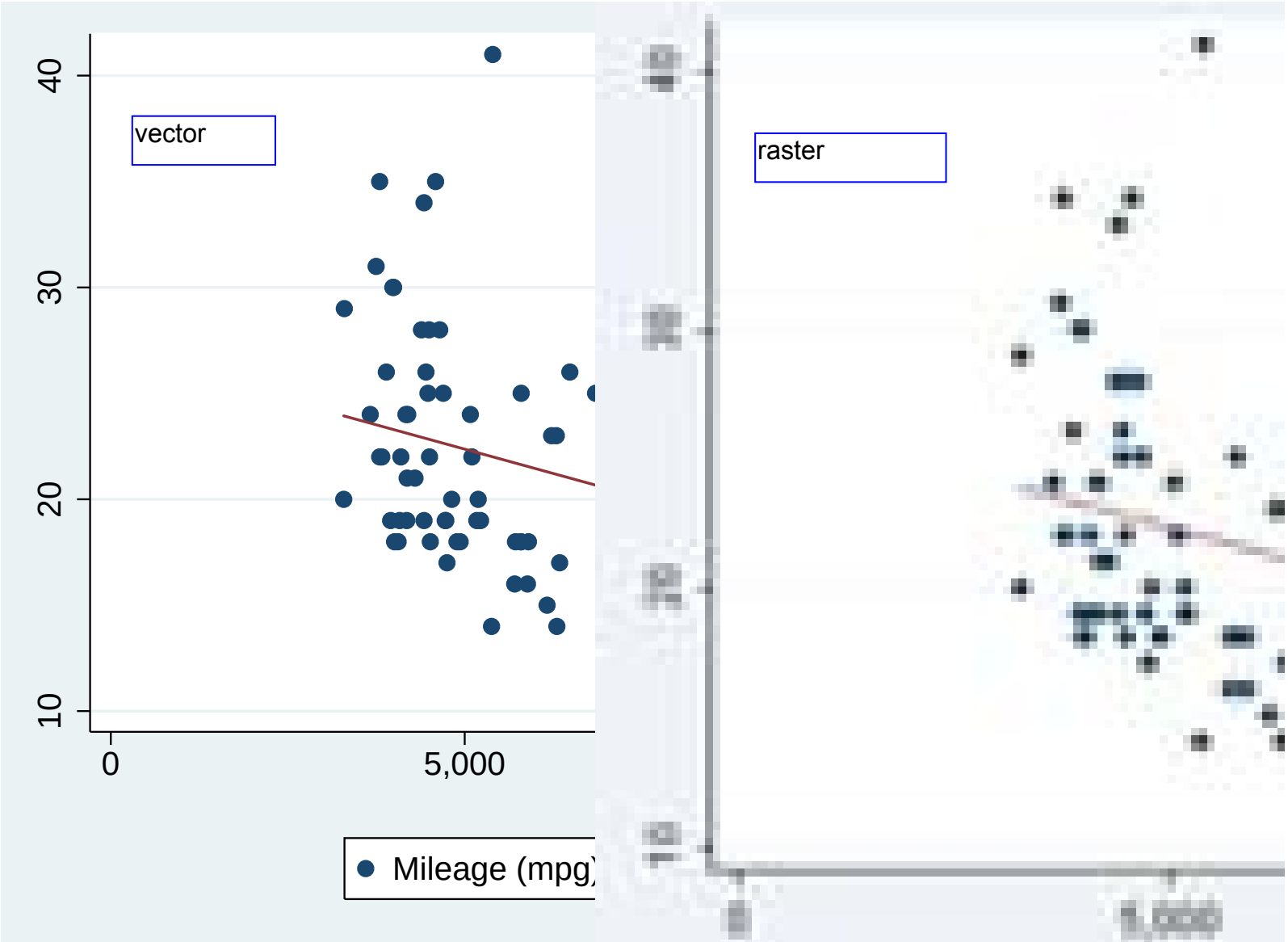
approach

- we will see lots of examples and lots of code
- much of it will be irreleveant 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)

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

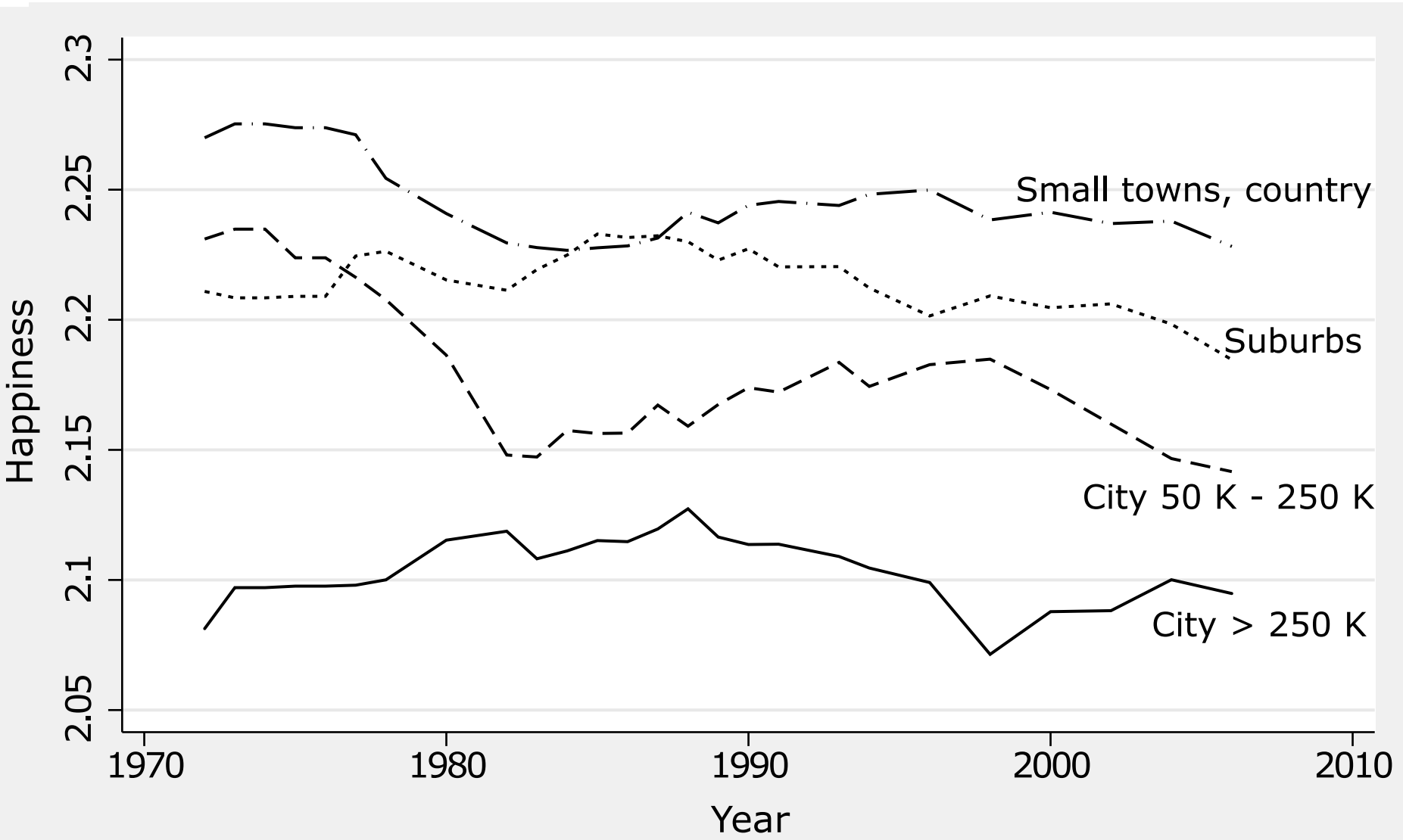
howto graph it?

- 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

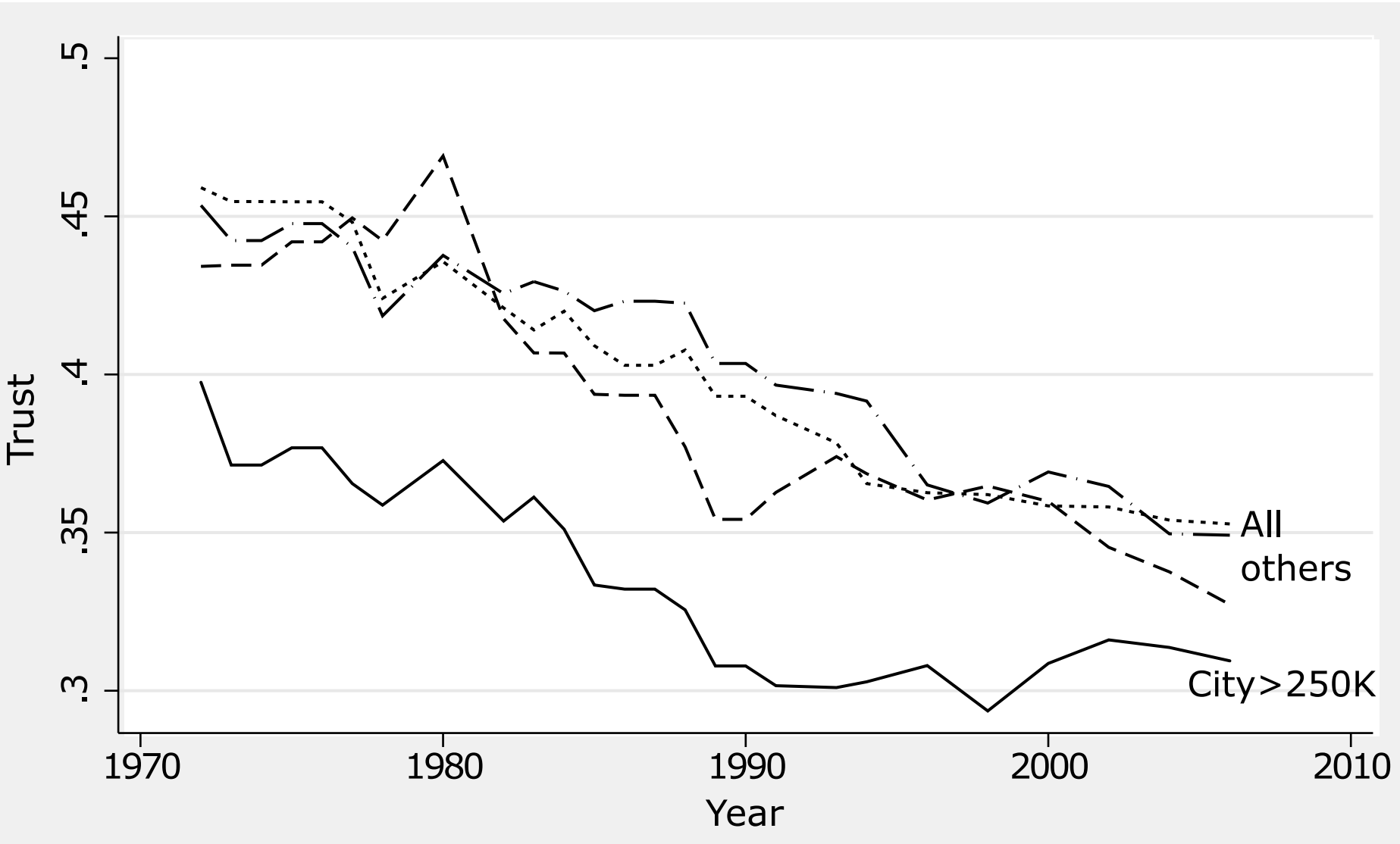
continuous vars

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

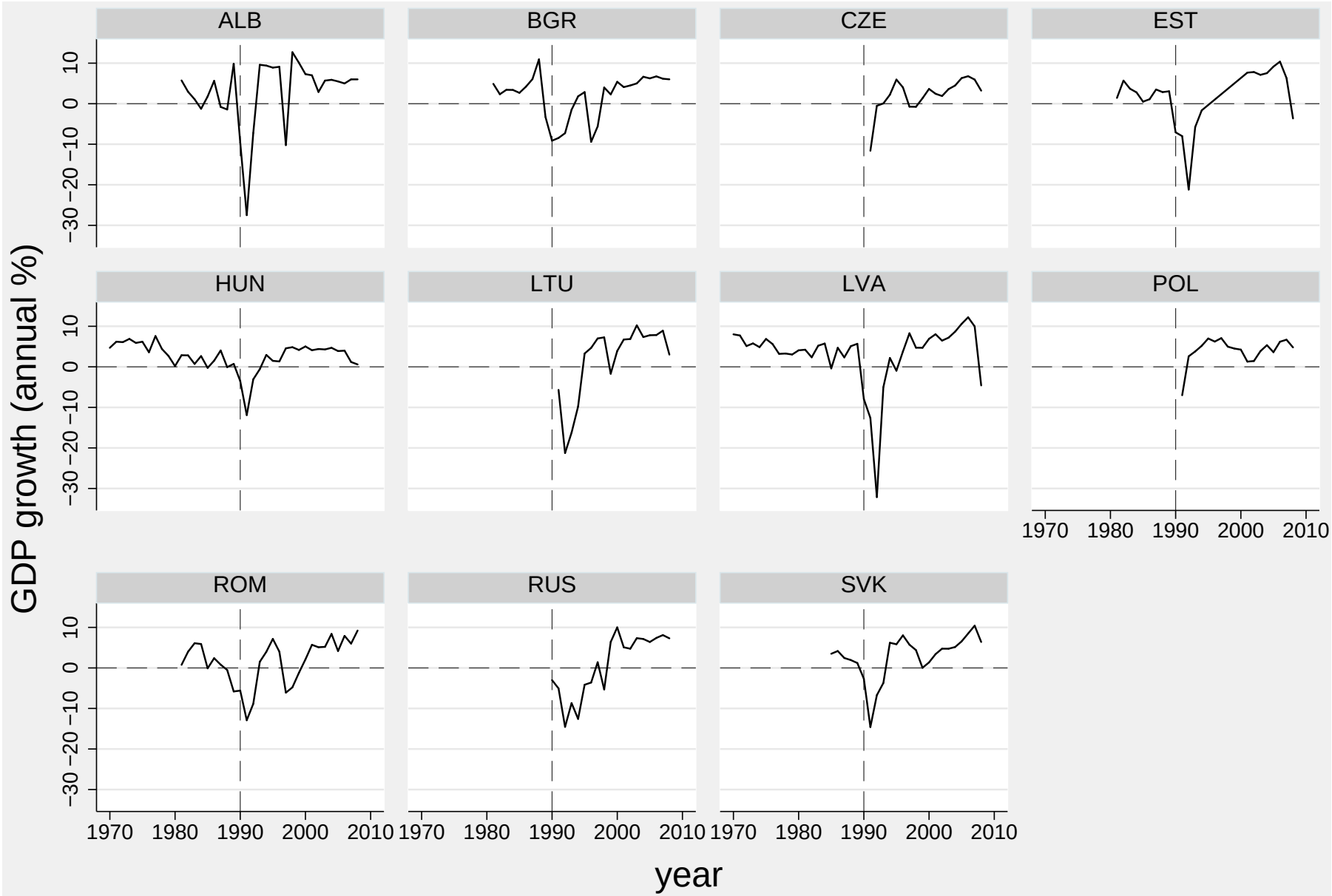
line ex 1


Happiness
2.05
2.1
2.15
2.2
2.25
2.3
1970
1980
1990
2000
2010
Year
Small towns, country
Suburbs
City 50 K - 250 K
City > 250 K

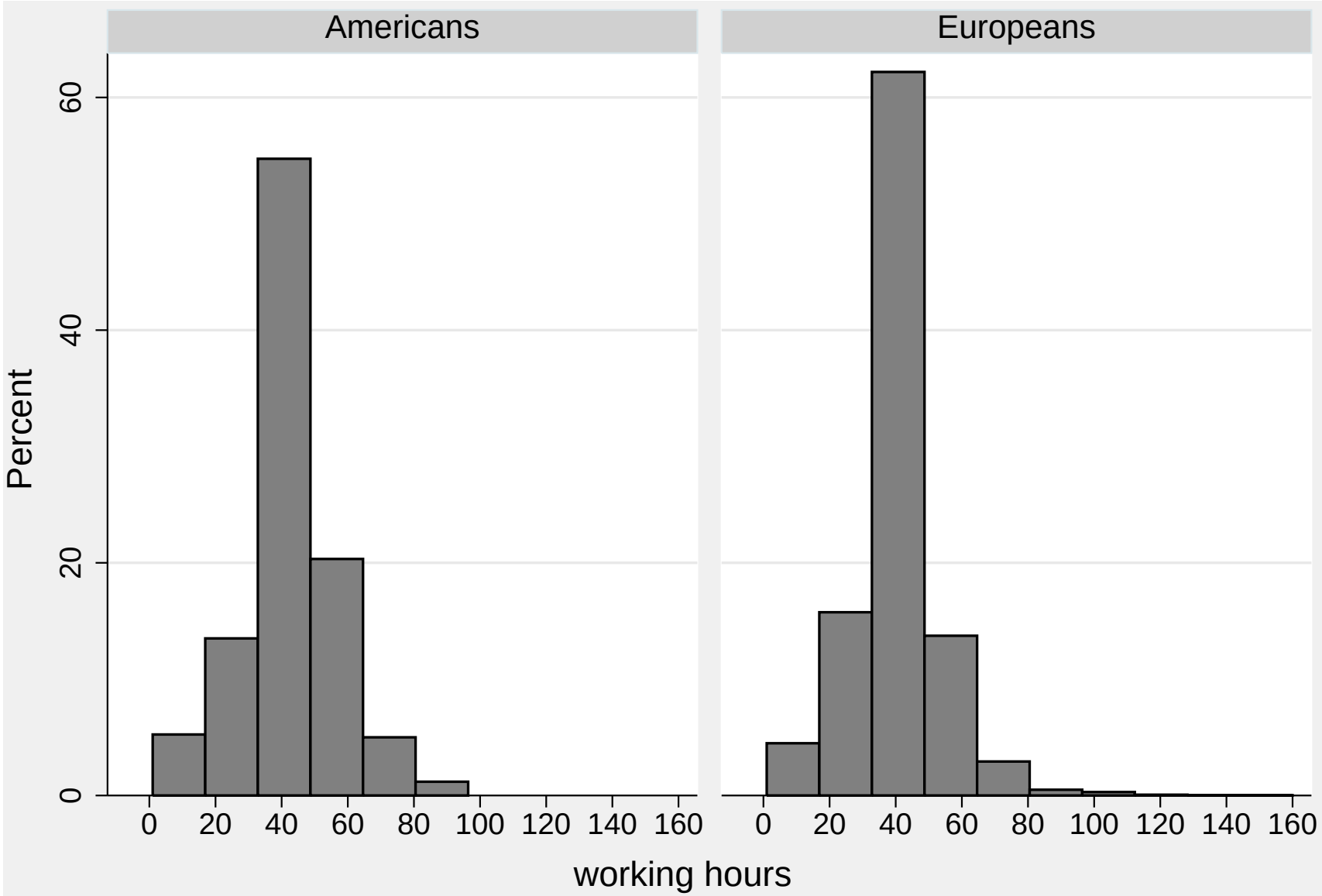
line ex 2


Trust
.5
.45
.4
.35
.3
1970
1980
1990
2000
2010
Year
All
others
City>250K

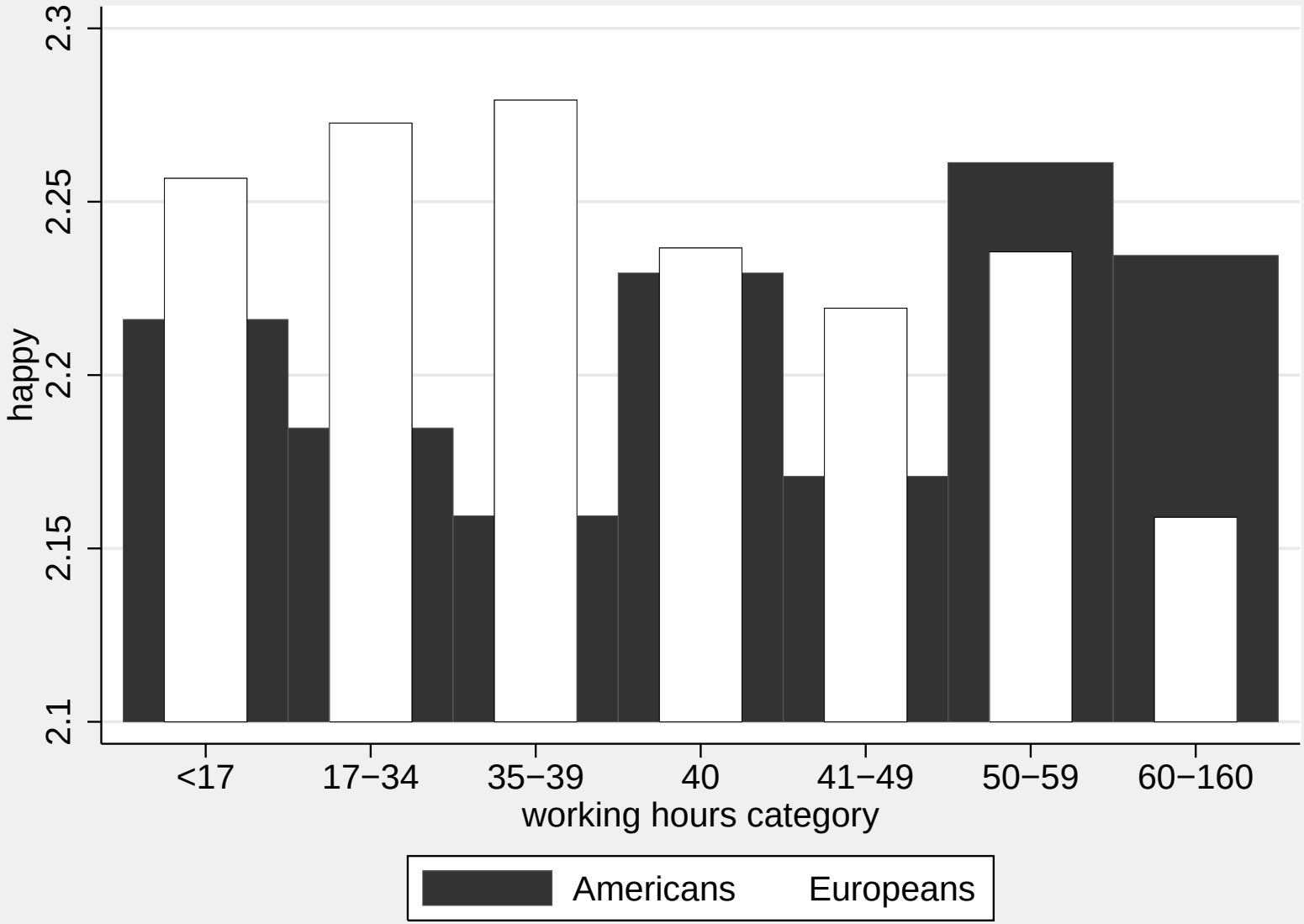
line gr ex 3 dofile: continous vars



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



dofile: categorical, continuous, summary statistics



radar

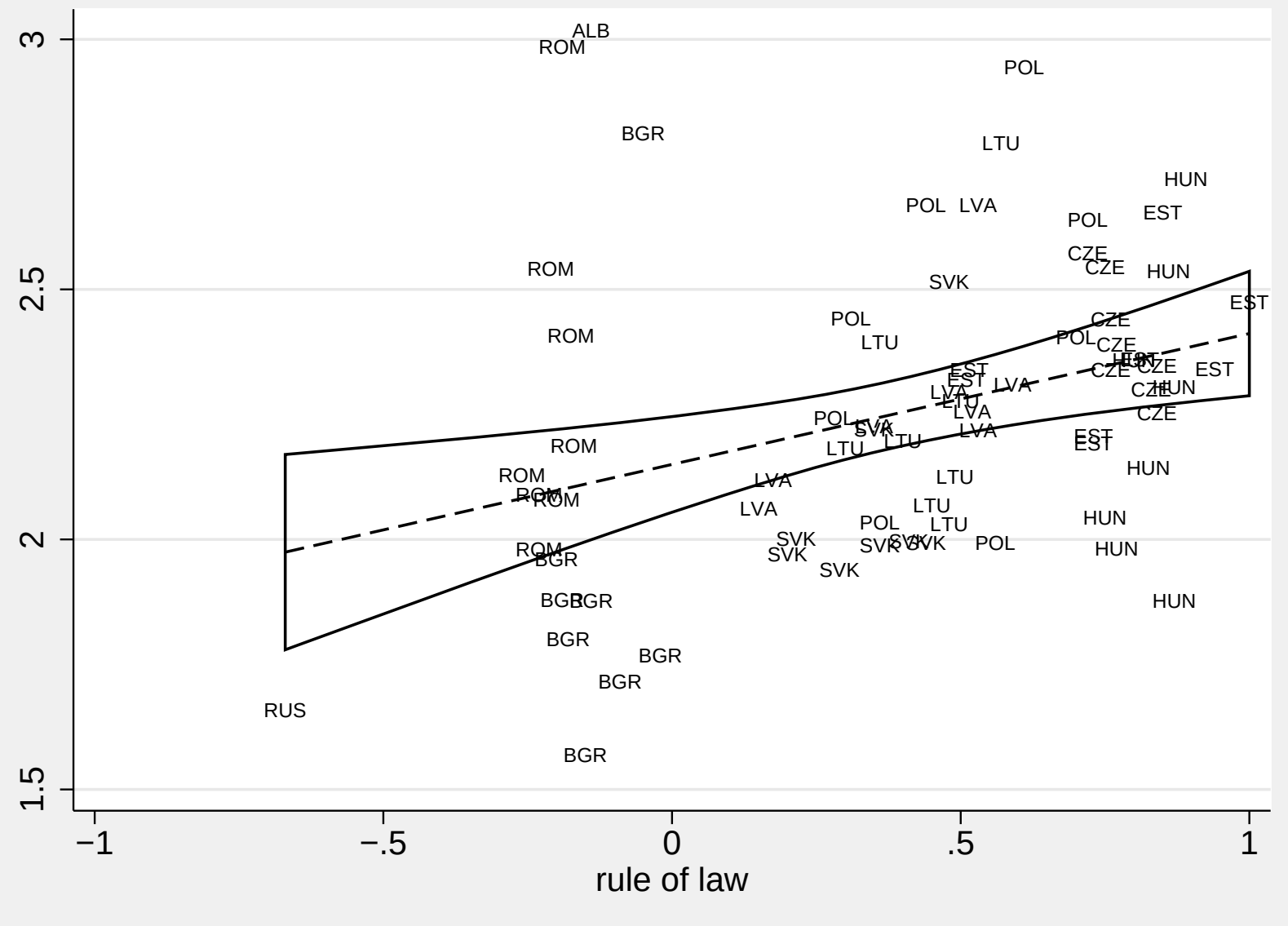
- very cool graph!
- https://www.google.com/search?q=stata+radar+plot&source=lmns&bih=681&biw=1200&rlz=1CAOTWH_enUS865&hl=en&sa=X&ved=2ahUKEwieueWt8ZDwAhWjfN8KHRx_DwEQ_AUoAHoECAEQAA

outline

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

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

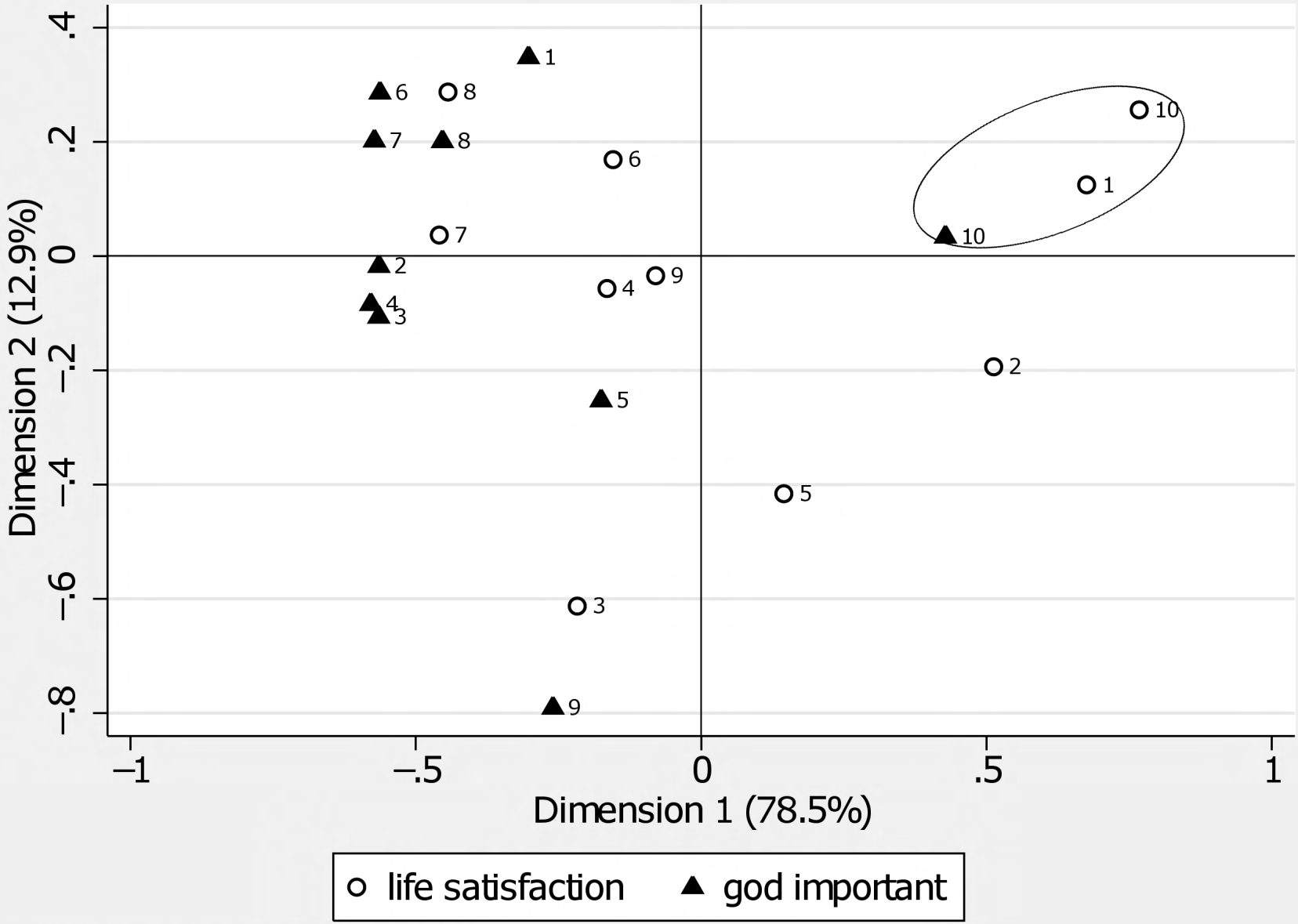
descriptive graphs

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-matplotli

maps: use python

- http://sensitivecities.com/so-youd-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