Introduction to Graphics With Stata

Adam Okulicz-Kozaryn
Harvard-MIT Data Center (HMDC)

dataquest@help.hmdc.harvard.edu

Organization

- Please do interrupt and ask questions if questions are relevant to the currrent topic or if you are lost
- For further questions there will be a Q & A after the class
- Collaboration with your neighbours is encouraged
- Slides/Exercises assume you use lab computer; If you have laptop adjust (e.g. paths) accordingly
- If you are ahead of time:
  - help others
  - experiment with commands
  - read help files

Organization cont'd

- Make comments in the code file (we will download it), not on your handouts – you are going to use code/commands, not the handouts
- Save commented code file on flash drive or email to yourself

Outline

some theory (Tufte)

stata graph basics

descriptive graphs

postestimation graphs

bonus

good practices and next step

assumptions and disclaimers

- this is **introduction to graphics with stata**
- assumes basic knowledge of stata
- not appropriate for people not familiar with stata

download do-file, slides and exercises

- find class materials `http://stathelp.iq.harvard.edu/stata_gph`
- right-click, save link as, and put on c:\ drive,
  go to c:\ and right-click, select **win-zip** and **extract to here**
- there are presentation slides, exercises, and the do-file

Outline

why use graphs ?

- graphs are an excellent way to communicate quantitative information
- people will remember graphs better than just numbers
- graphs will enliven your argument

when use graphs ?

- display of many numbers, e.g. time series
- presentations/talks
- remember: keep it simple

be simple

- everything should be made as simple as possible, but not one bit simpler
- avoid data padding – present only data needed for a specific purpose
- avoid clutter – put in a single graph only the data that are highly related and must be compared
- put data into appendix if it is not very relevant but may be useful

avoid visual clutter

- all parts of graph should be meaningful
- good practices:
  - do not use shades
  - do not use fancy colors
  - do not use any decoration

chartjunk

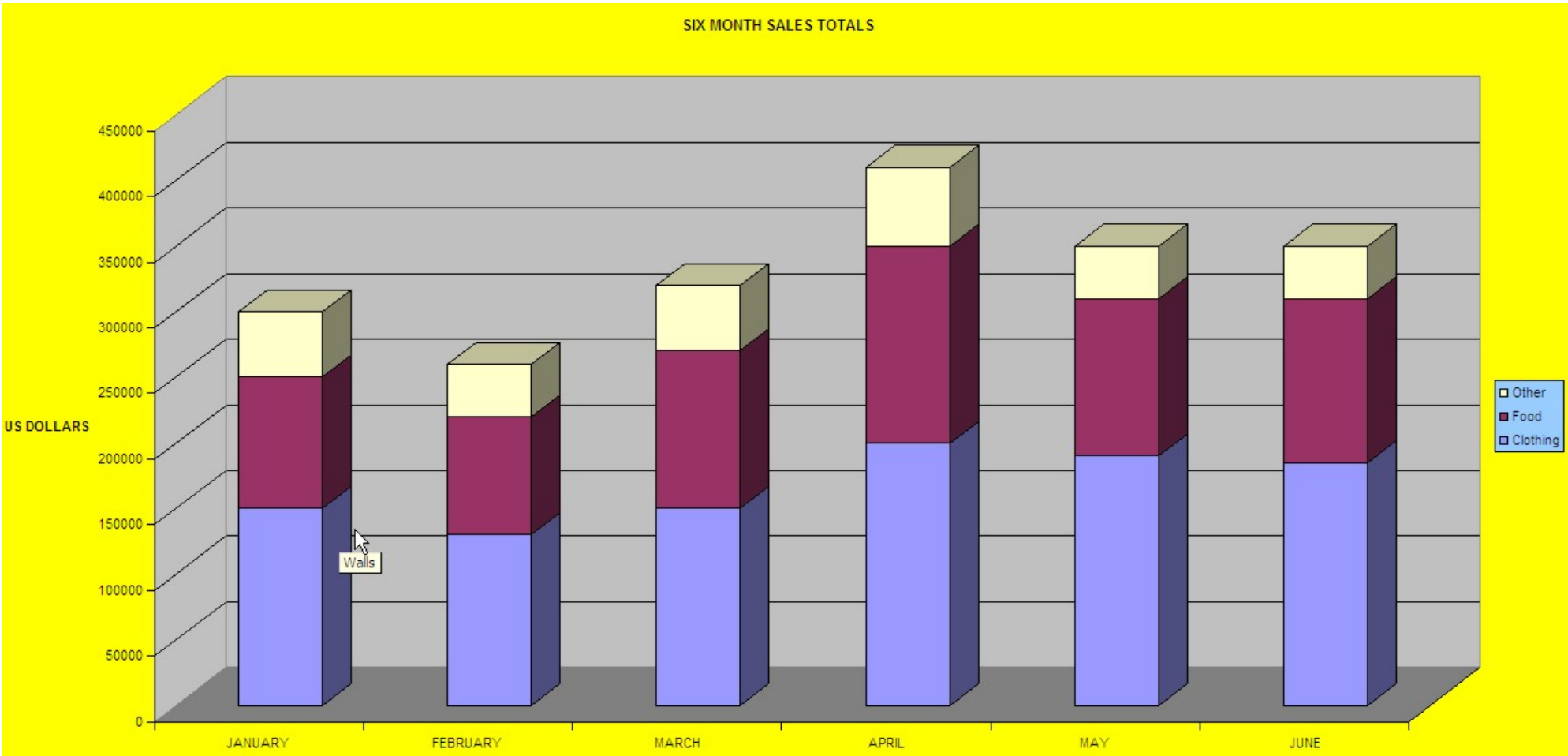
◇ *The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its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

Edward Tufte "The Visual Display of Quantitative Information"

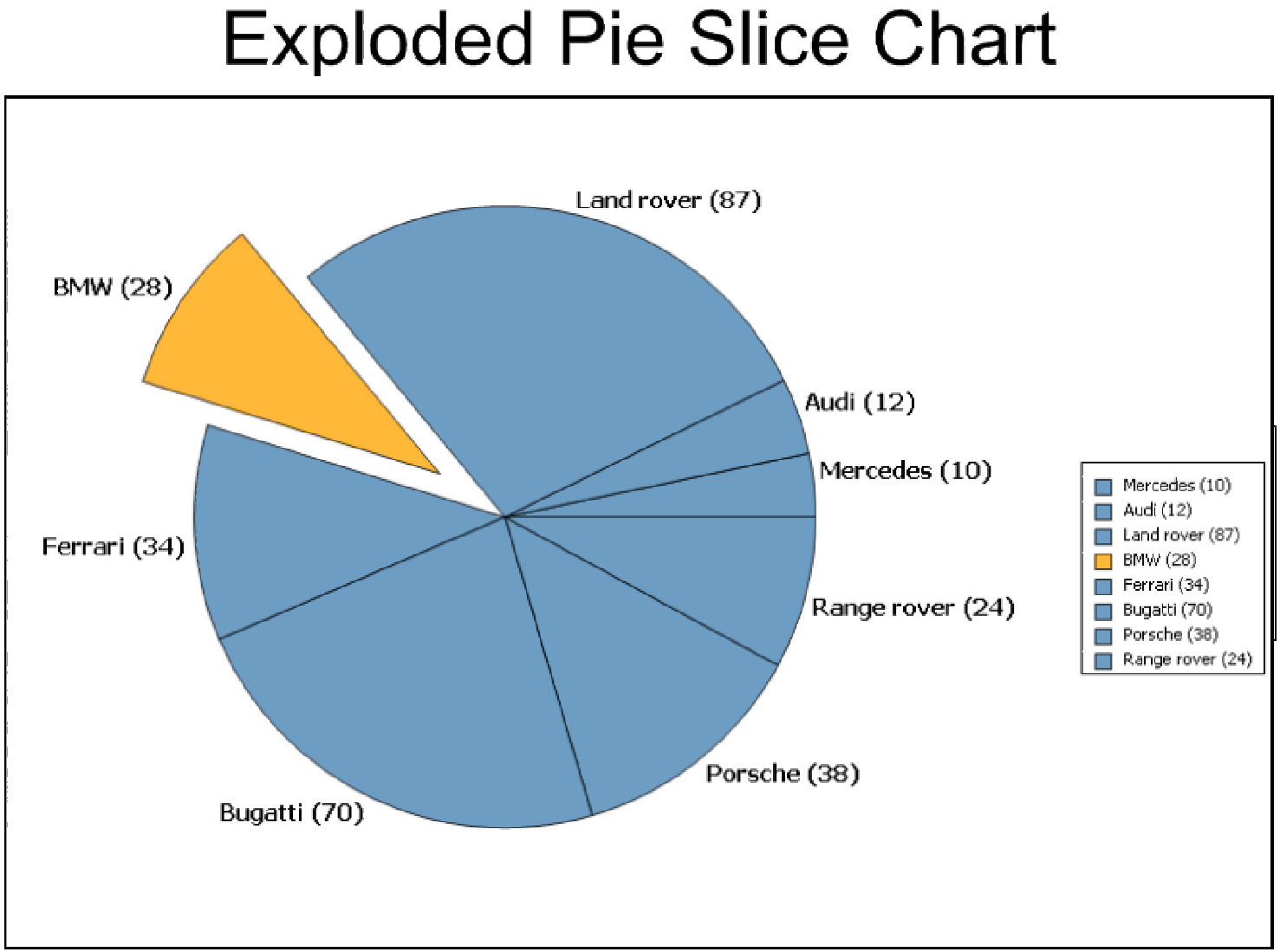
◇ examples follow:

- chartjunk or "business graphs"
- good fancy graphs

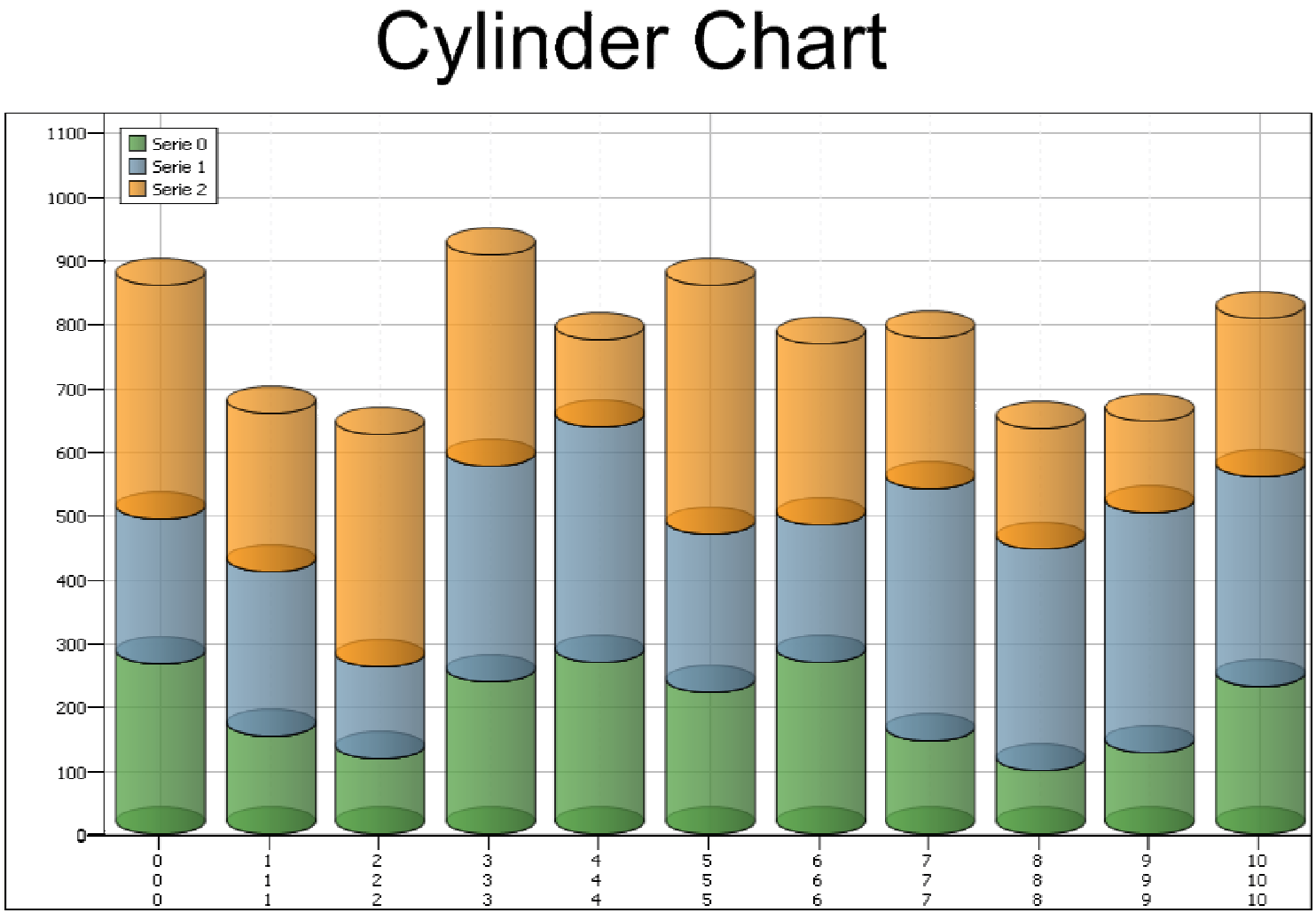
chartjunk


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

chartjunk

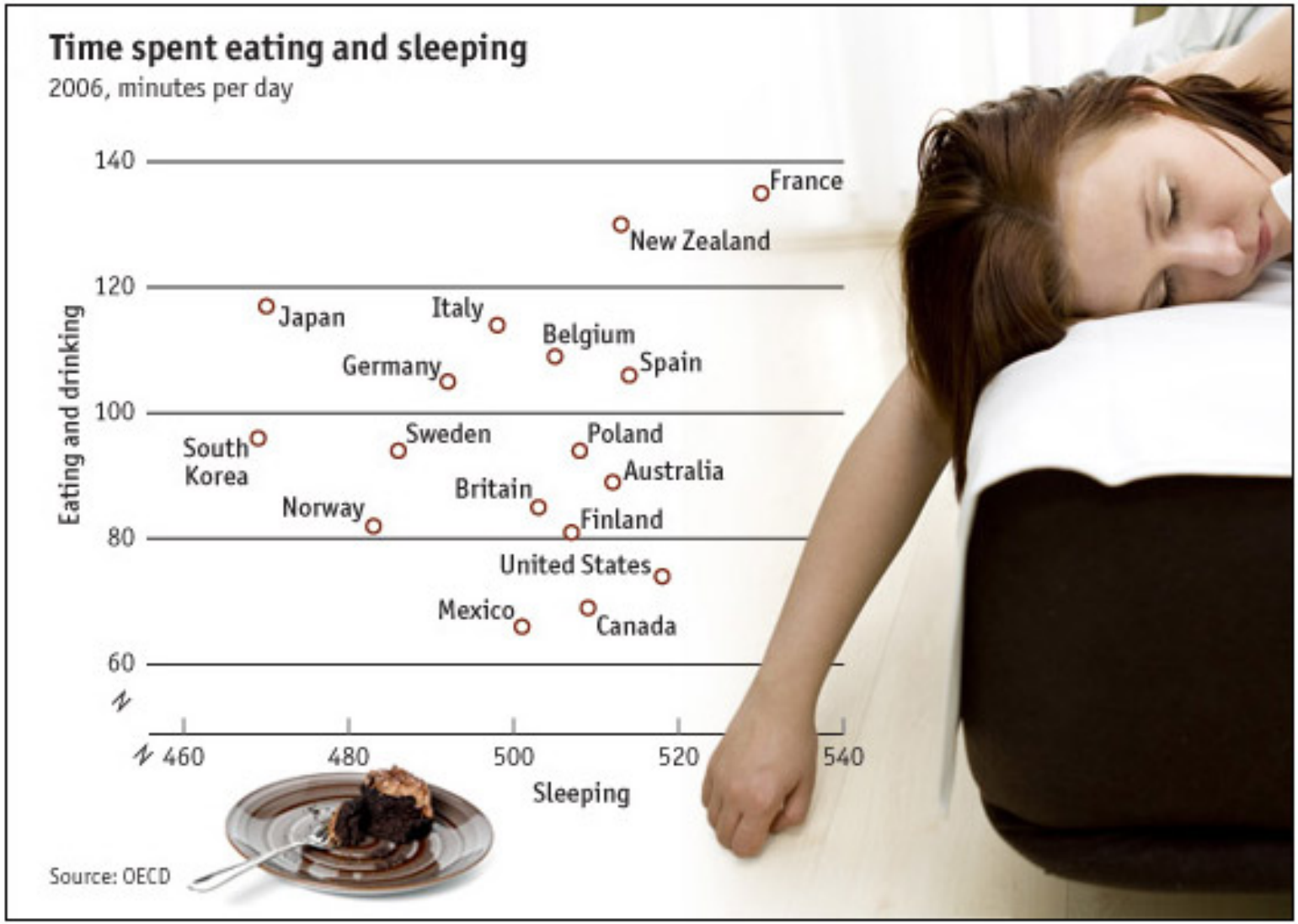

Exploded Pie Slice Chart
Land rover (87)
BMW (28)
Audi (12)
Mercedes (10)
Ferrari (34)
Range rover (24)
Porsche (38)
Bugatti (70)
Mercedes (10)
Audi (12)
Land rover (87)
BMW (28)
Ferrari (34)
Bugatti (70)
Porsche (38)
Range rover (24)

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

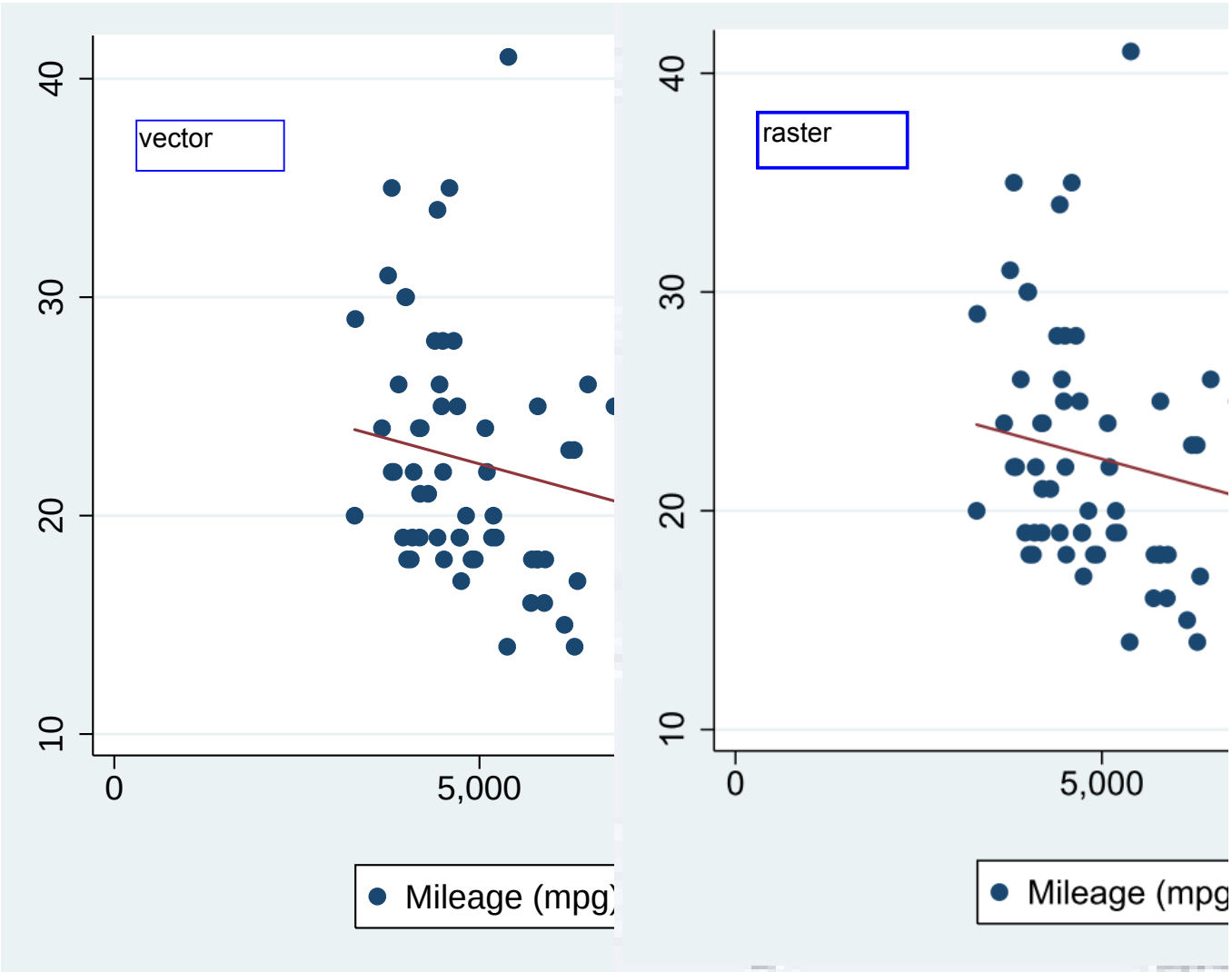
not chartjunk (the economist)

not chartjunk (the economist)

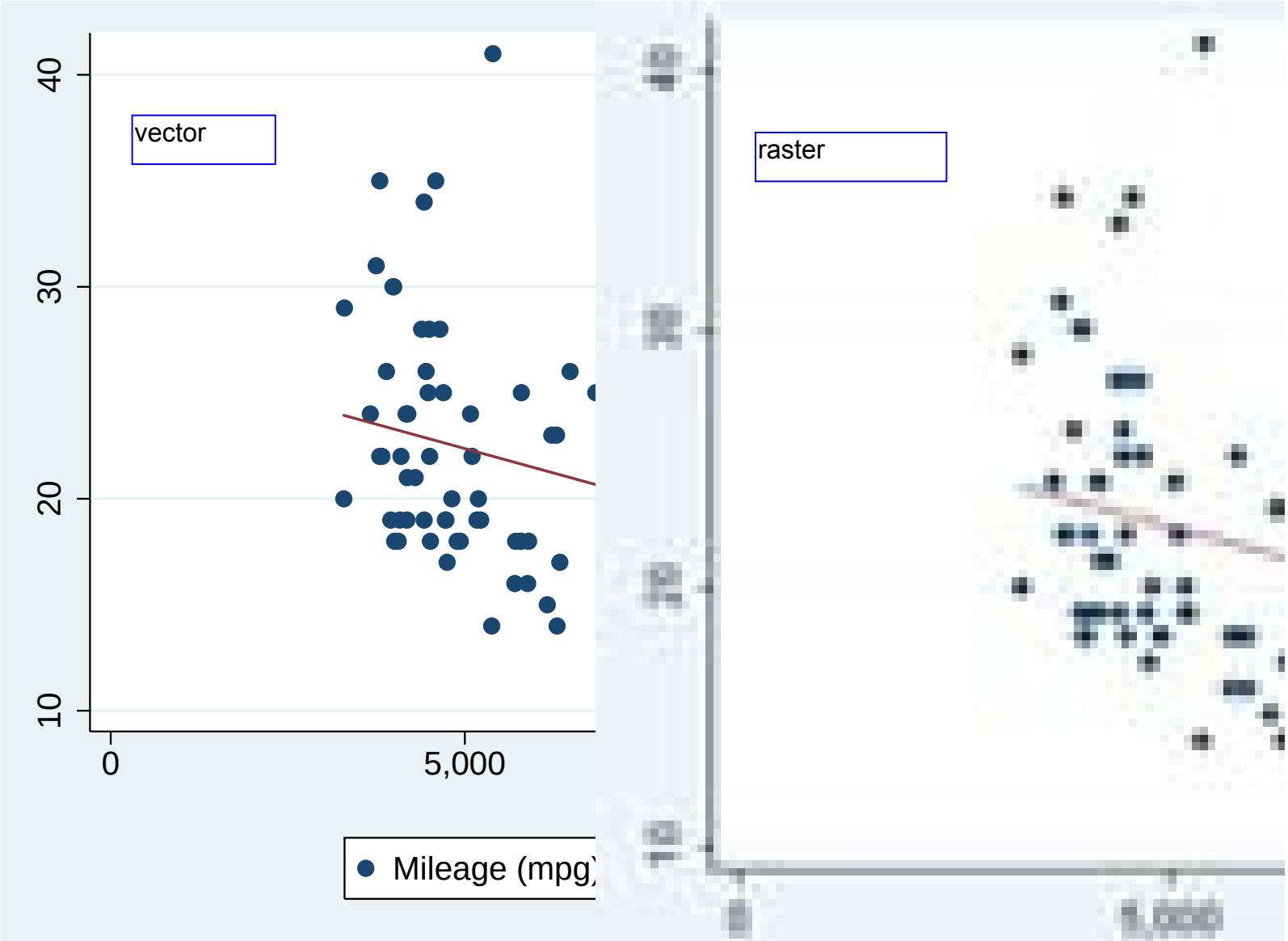


Outline

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 (stata 10+)
- hi-tech: use PS, EPS
- windows: use EMF

formats, exporting, importing

- formats depend on OS `help graph export`
- to export either **right click** >**save as** or
  `gr export myfig.eps, replace`
- MS word: **insert** > **picture** > **from file** or
  drag it over into MS word
  or **right click** in stata > **copy** and **paste** in MS word
- LaTeX `includegraphics{myfig.eps}`
- do-file

strategy

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

  extra: a short tutorial from stata help file (good but boring)

  help graph intro

galleries

- useful galleries
  - `http://www.ats.ucla.edu/stat/Stata/library/GraphExamples/default.htm`
  - `http://www.stata.com/support/faqs/graphics/gph/statagraphs.html`
  - a visual guide to stata graphics, available in the lab
- fancy galleries
  - `http://www.survey-design.com.au/Usergraphs.html`
  - `http://www.ats.ucla.edu/stat/stata/faq/graph/atsplot.htm`
  - `http://www.ats.ucla.edu/stat/stata/faq/graph/njcplot.htm`

GUI

- GUI is a good idea with graphs: **graphics** >
- experiment
- figure out your code for a particular task
- save it in a do-file
- next time, just change variables in the do-file

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** and **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

useful options

- graphs have lots of code; use:
  ///
  #delimit; and #delimit cr or
- usually you want graphs to be black and white and more compact than the default, but there are many more options extra: help region_options
- extra: here is the gallery of symbols
  http://www.ats.ucla.edu/stat/stata/faq/showmark.htm
  or:
  palette linepalette
  palette symbolpalette
  palette color

graph options

```
tw (scatter mpg price, ///
msymbol(smcircle_hollow)), scheme(s2mono) ///
graphregion(margin(tiny)) ///
plotregion(margin(tiny))


#delimit;
twoway (scatter mpg price,
msymbol(smcircle_hollow)),scheme(s2mono)
graphregion(margin(tiny))
plotregion(margin(tiny)) ;
#delimit cr
```

do-file

graph combine

- again, to overlay graphs use twoway
- to put graphs next to each other gr combine
- do-file

graph editor [extra]

- starting with stata 10 you can edit graphs
- let's produce a simple graph and edit it
- sysuse auto, clear
  tw (scatter mpg price) (lfitci mpg price)
- extra: help graph_editor
- exercise 1

Outline

why description

- most people have urge to run as complicated models as possible as quickly as possible – this is wrong
- without descriptive statistics
  - you may not know what models to run
  - you may not know of interesting hypotheses to test
  - you are likely to make mistakes:
    (1) coding of data (e.g. income quartiles, not $);
    (2) available variables (e.g. no ability, but education);
    (3) relationships in data (e.g. quadratic)

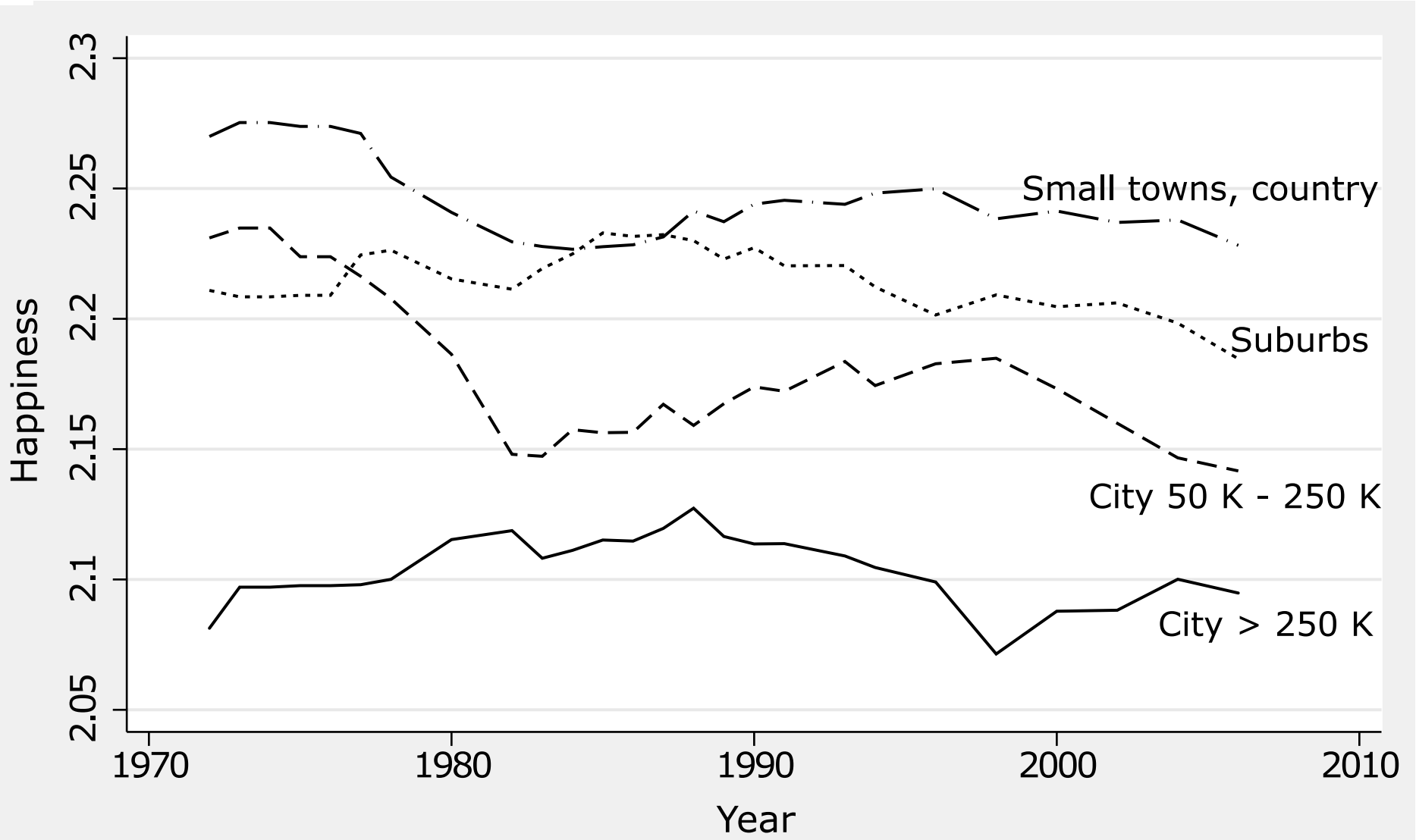
howto graph it ?

- again, look at the galleries
- graphing possibilities depend on level of measurement of vars and on what you want to do
  - continuous
  - distribution
  - categorical, continuous, summary statistics

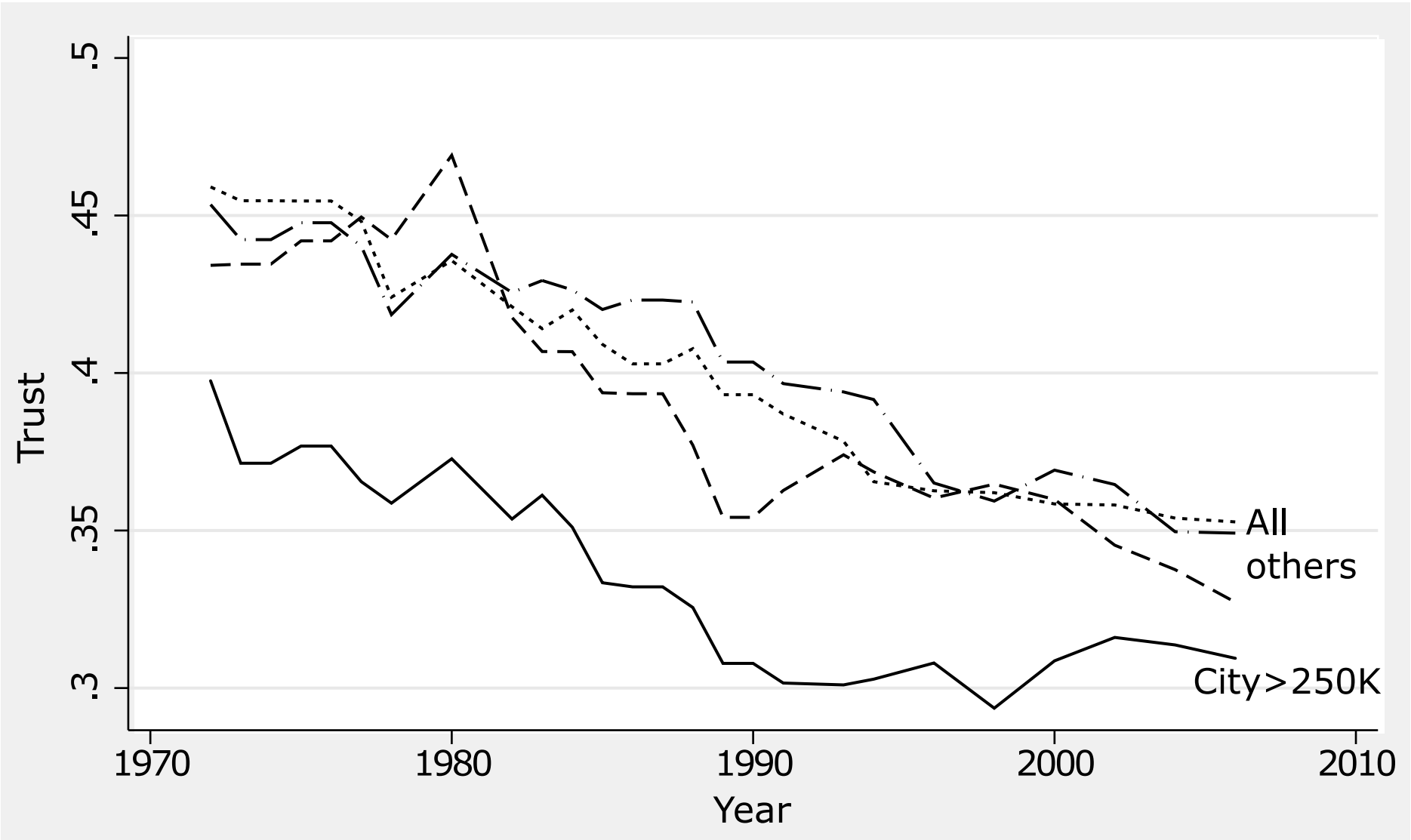
continuous vars

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

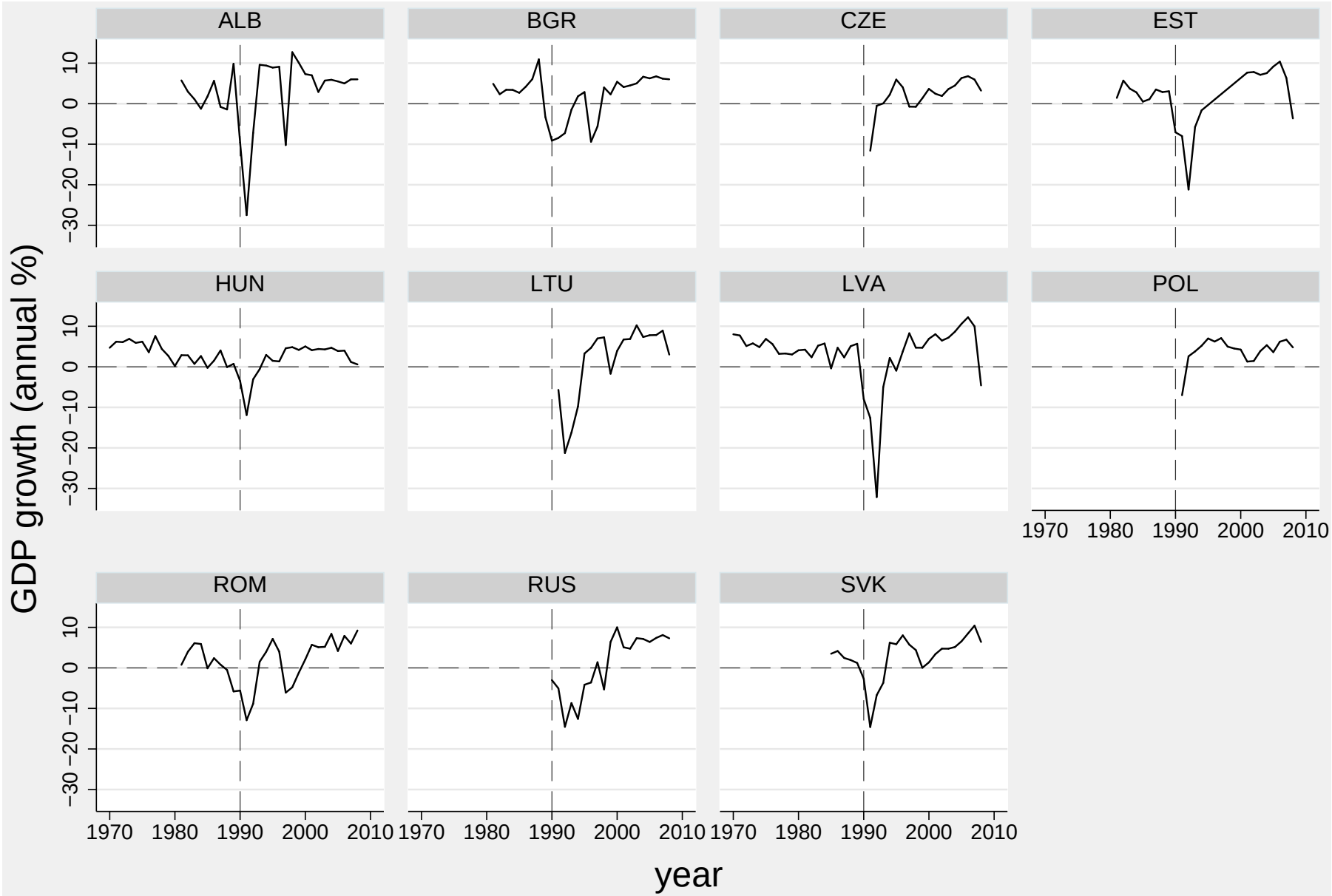
line graph example 1



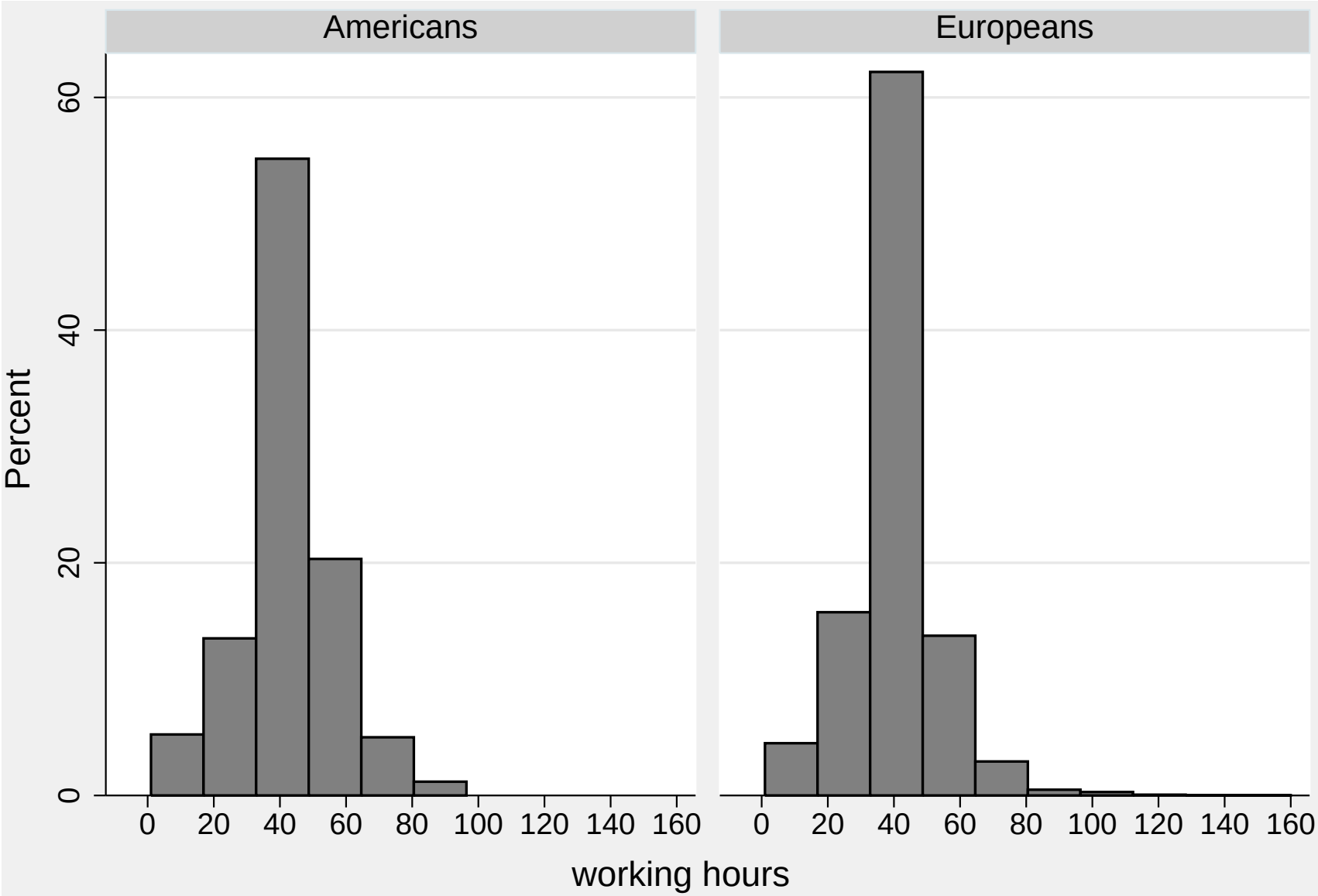
line graph example 2



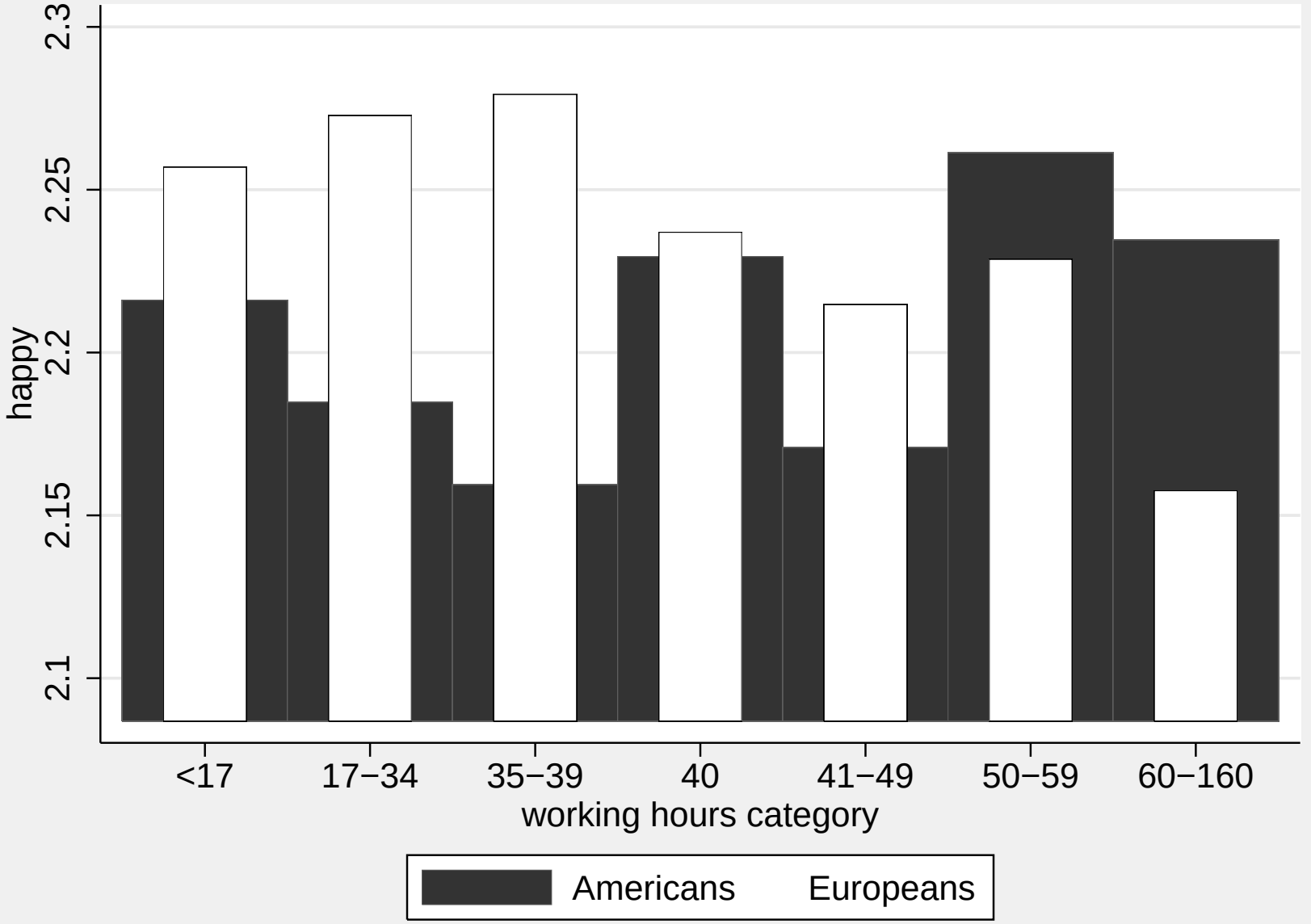
line graph example 3 (do-file)



distribution graph example (do-file)



categorical, continuous, summary statistics (do-file)



advanced descriptive graphs

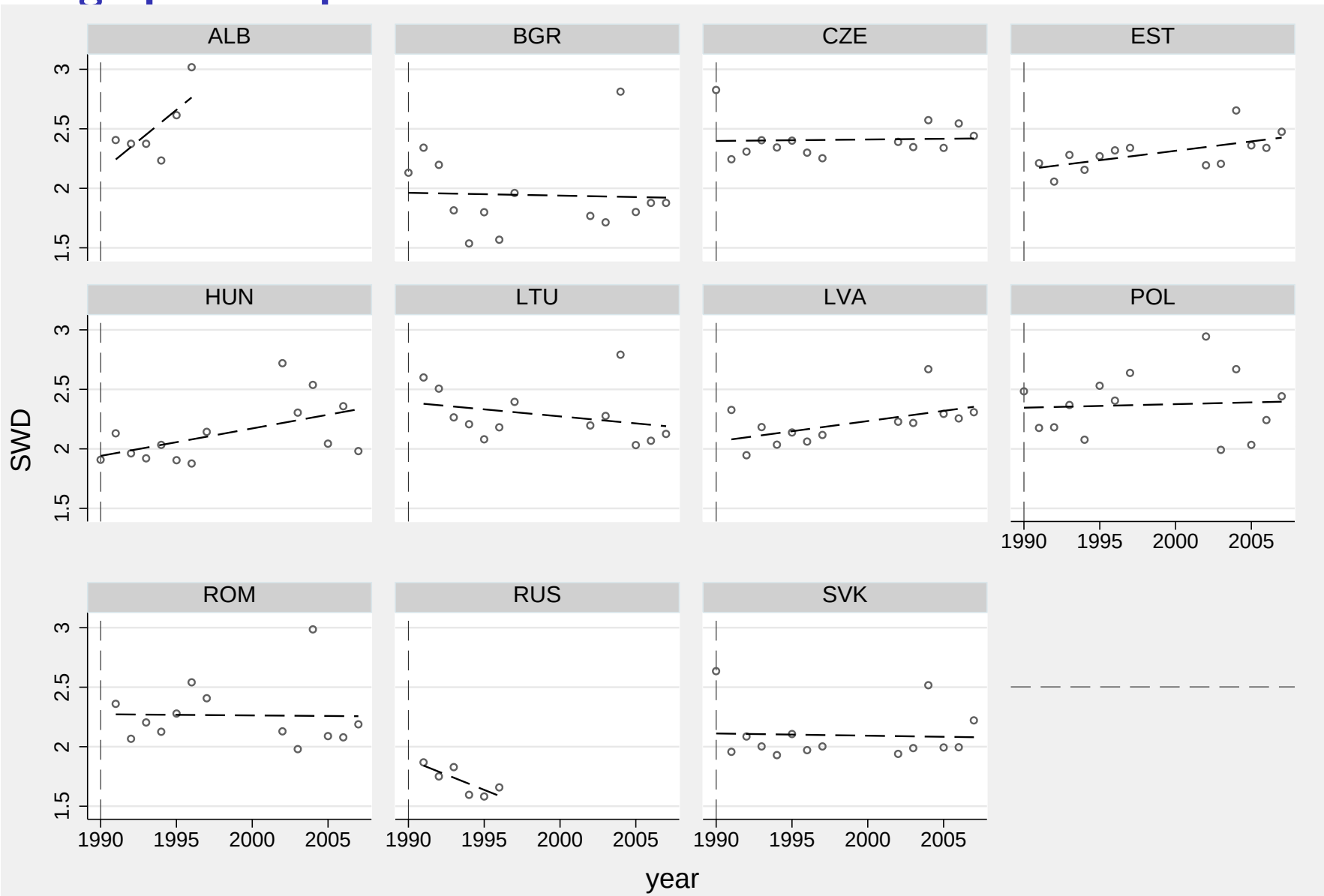
- present usual numeric statistic (e.g. frequency table) in graphical format
- be careful not to overdo with graphs
- EXTRA: do-file
- exercise 2

Outline

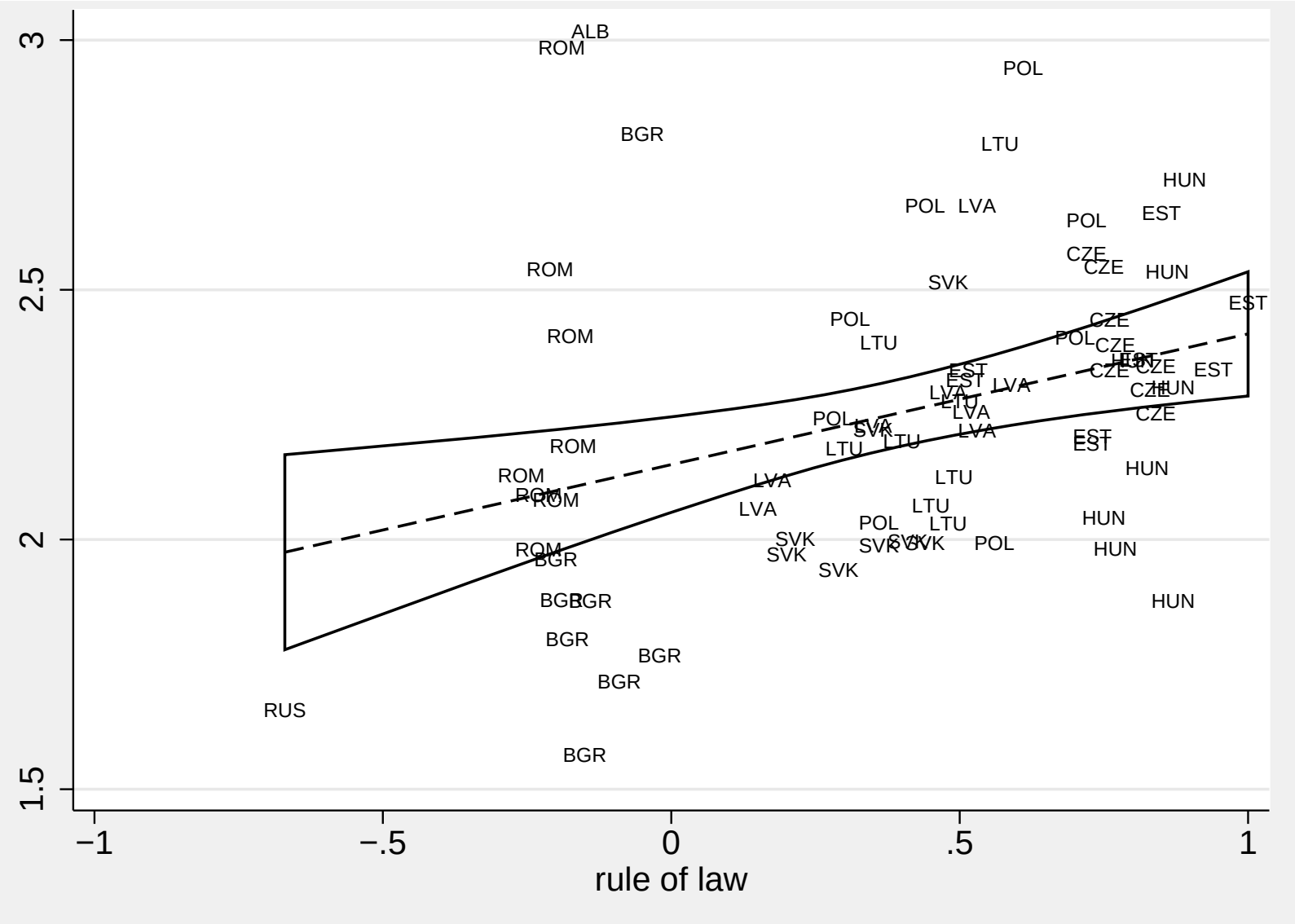
why graphing postestimation results

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

fit graph example 1



fit graph example 2



graphs 1

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

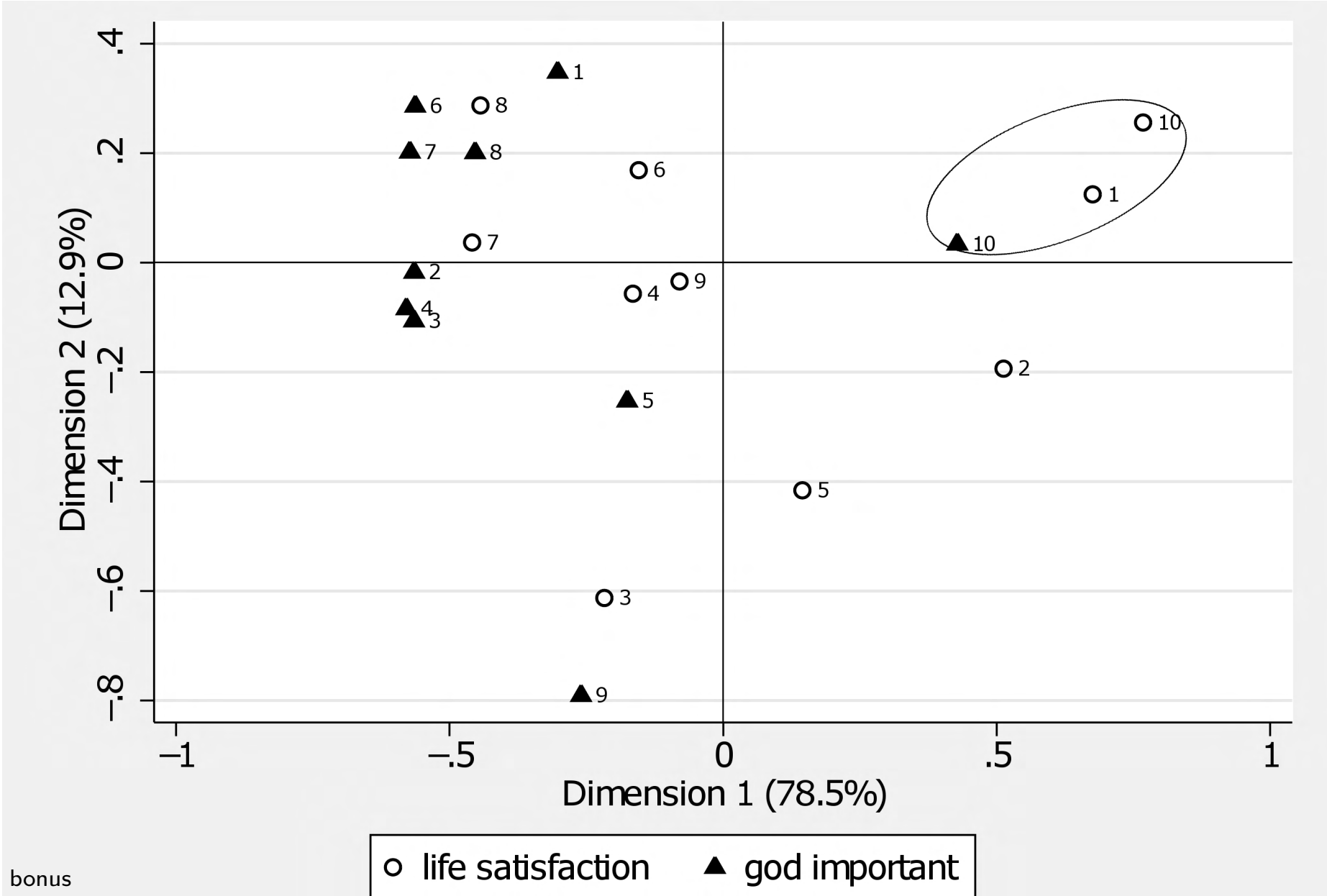
graphs 2

- 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
- exercise 3

Outline

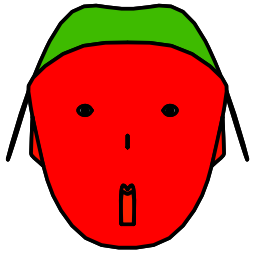
ca biplot graph example

(do-file)

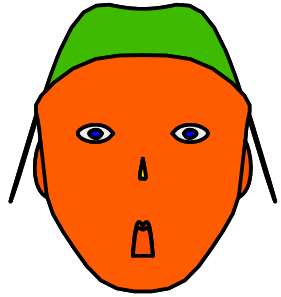


chernoff faces (do-file)

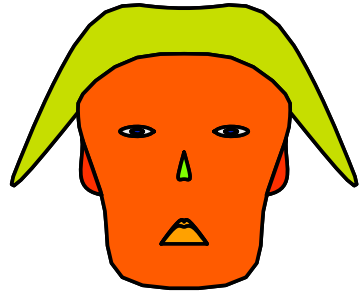
1947



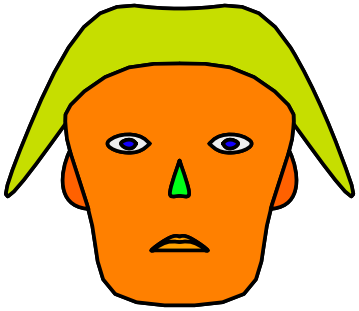
1948



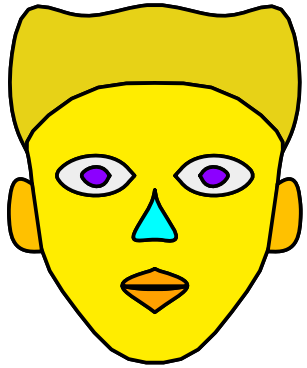
1949



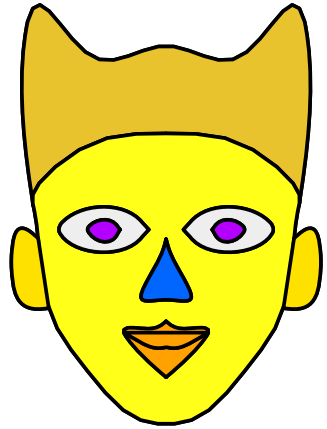
1950



1951



1952



1953

1954

1955

maps in stata

- maps are almost always good to describe data: everything takes place somewhere
- stata can produce maps, but the are clunky
- `http://www.stata.com/support/faqs/graphics/tmap.html`
- `http://huebler.blogspot.com/2005/11/creating-maps-with-stata.html`

maps in R

- R can produce maps, but the are still quite clunky (less clunky than Stata)
- `http://geography.uoregon.edu/geogr/topics/maps.htm`
- `http://www.omegahat.org/GoogleEarth/CityTemperatures/`
- `http://cran.r-project.org/web/packages/RgoogleMaps/vignettes/RgoogleMaps-intro.pdf`
- `http://cran.r-project.org/web/packages/RgoogleMaps/`

maps in GIS software

- For crating maps best use GIS software such as free GeoDa (geodacenter.asu.edu) or ArcGIS, see our workshops
- http://www.iq.harvard.edu/geographic_information_system_gis_software_2009_2010
- http://libraries.mit.edu/gis
- http://gis.harvard.edu

advanced graphs

- do-file
- fancy galleries
  - `http://www.survey-design.com.au/Usergraphs.html`
  - `http://www.ats.ucla.edu/stat/stata/faq/graph/atsplot.htm`
  - `http://www.ats.ucla.edu/stat/stata/faq/graph/njcplot.htm`

Outline

good practices

- use graphs often
- never use chart junk
- do not use graphs if they take up more space than text or numbers
- avoid graph padding and within-graph data padding
- as always, keep code in do-file and reuse it for future projects
- share code with others – make your project open-source – show how you came up with the results

next step: more links

- http://www.ats.ucla.edu/stat/stata/topics/graphics.htm
- http://www.ats.ucla.edu/stat/Stata/library/GraphExamples/default.htm
- http://www.stata.com/support/faqs/graphics/gph/statagraphs.html

next step: other graphs

- there are more graphs than that – explore GUI, for instance we did not discuss survival analysis graphics
- use web search

next step: when stata is not enough

- if you need fancy/complicated graphs **and** like programming use R
- `http://addictedtor.free.fr/graphiques/thumbs.php`
- `http://research.stowers-institute.org/efg/R/`
- R can produce maps too, but if you like GUI use GeoDa (free) or ArcGIS (not free)

Thank You !

- Please fill evaluations AND give us some comments/feedback – we do care for these classes and want to make them better
- Come to other classes we offer and tell your friends about our classes `http://www.iq.harvard.edu/statistical_software_2009_2010`

A Word From Our Sponsor !

- Institute for Quantitative Social Science `http://iq.harvard.edu`
- Data Collection, Management

  `http://www.iq.harvard.edu/data_collection_management_analysis`
- Research Computing Environment

  `http://www.iq.harvard.edu/research_computing`
- Computer Labs (Software, Books) `http://www.iq.harvard.edu/facilities`
- Training `http://www.iq.harvard.edu/training`