Introduction to Graphics with Stata

Alicia Doyle Lynch
Harvard MIT Data Center

Documents for Today

- Find class materials at: http://libraries.mit.edu/guides/subjects/data/training/workshops.html
  - Several datasets
  - Presentation slides
  - Handouts
  - Exercises
- Let's go over how to save these files together

Organization

- Please feel free to ask questions at any point if they are relevant to the current topic (or if you are lost!)
- There will be a Q&A after class for more specific, personalized questions
- Collaboration with your neighbors is encouraged
- If you are using a laptop, you will need to adjust paths accordingly

Organization

- Make comments in your Do-file rather than on hand-outs
  - Save on flash drive or email to yourself
- Stata commands will always appear in red
- “Var” simply refers to “variable” (e.g., var1, var2, var3, varname)
- “pathname” should be replace with the path specific to your computer and folders

Assumptions and Disclaimers

- This is an **INTRODUCTION** to graphing in Stata
- Assumes basic knowledge of Stata
- Not appropriate for people already well familiar with graphing in Stata
- If you are catching on before the rest of the class, experiment with command features described in help files

Assumptions and Disclaimers

- I’m going to give you an overview of Stata’s capabilities
- I won’t be able to cover every graphing capability you’ll ever need!
- Take these skills – build on them and find what works for you

Useful Stata Graphing Resources

- 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” by Michael N. Mitchell
- Stata 11 users guide, “Graphics”

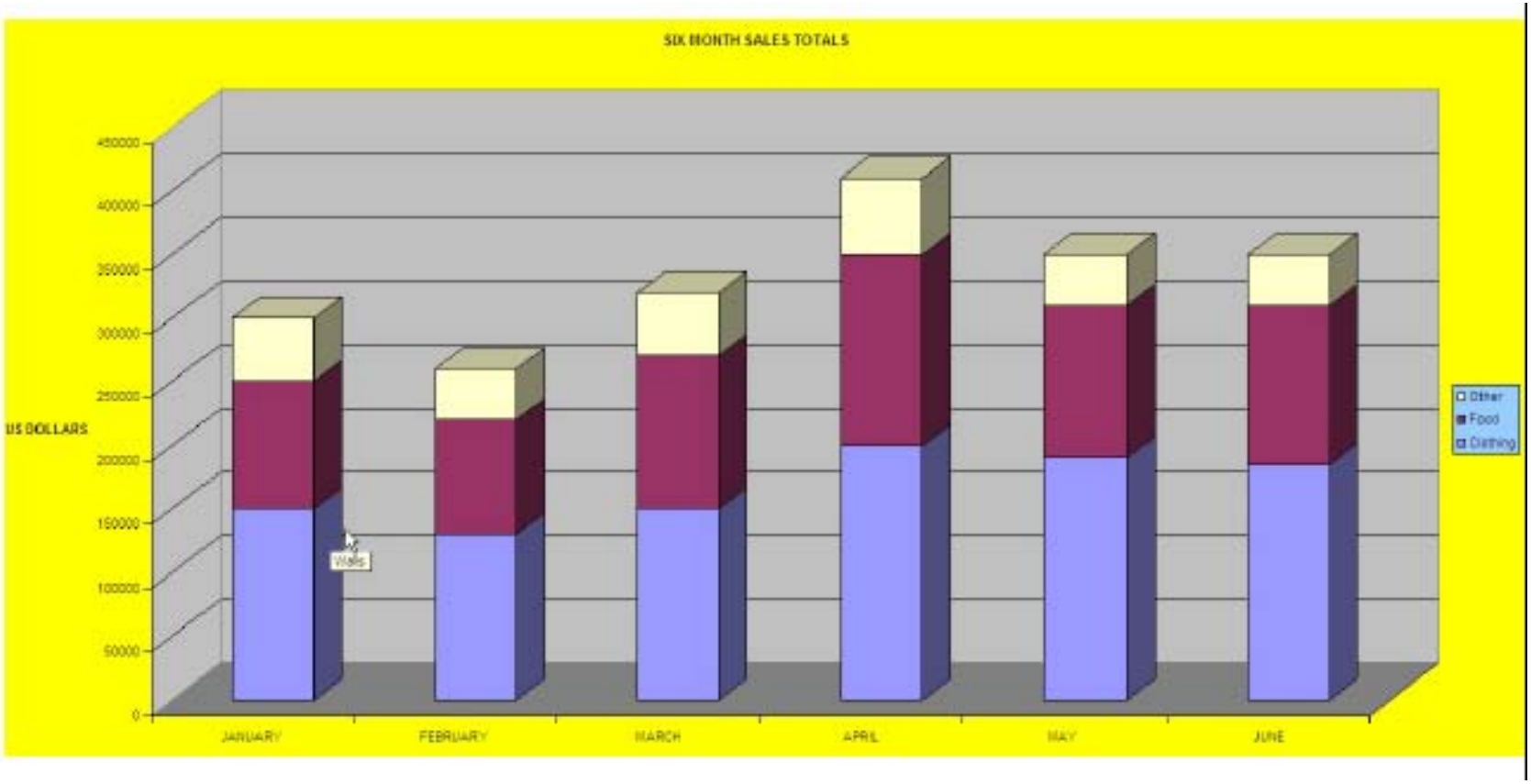
Why do we use graphs?

- You have a major point that is emphasized or easier to understand when displayed graphically
- Graphs are excellent means of communicating quantitative information
- More memorable than simply presenting numbers
- Easier for lay audience to interpret

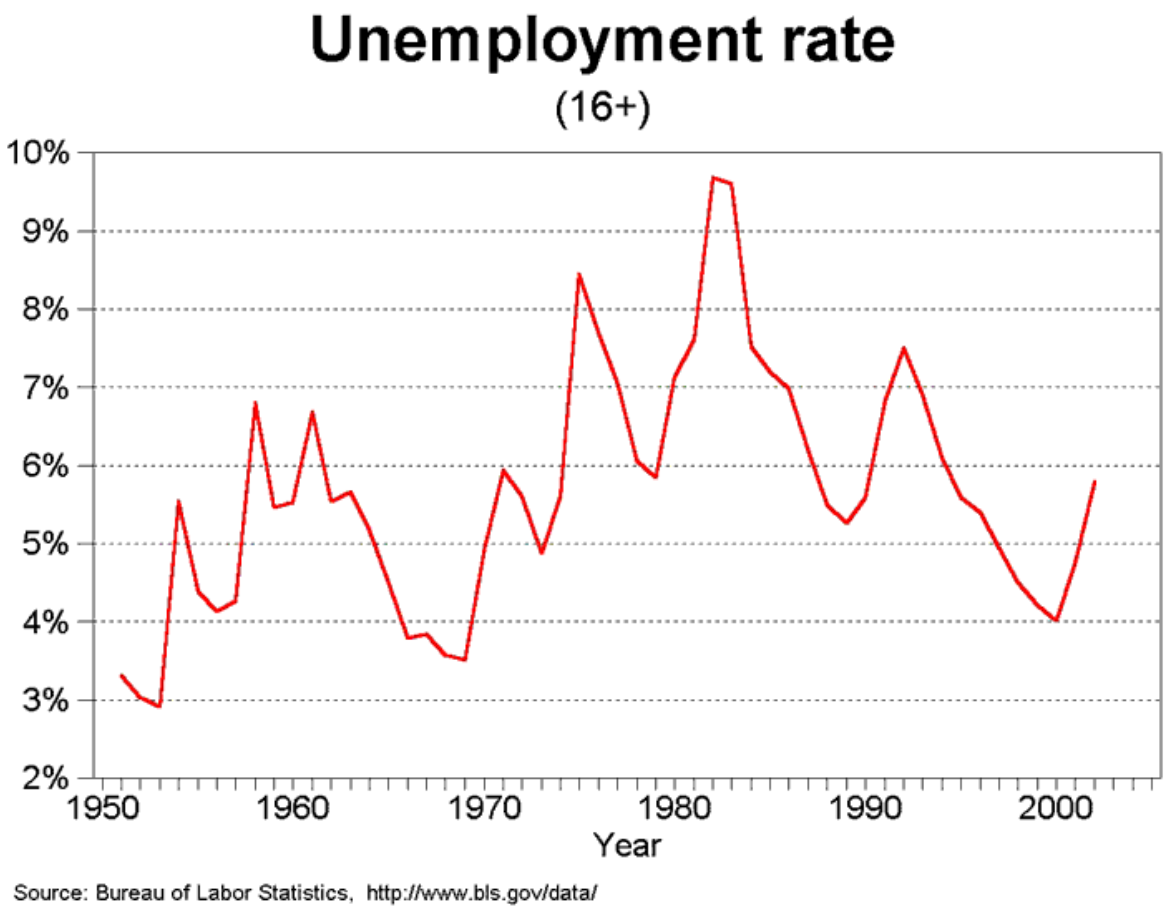
Graphing Strategies

- Keep it simple
- Labels, labels, labels!!
- Avoid cluttered graphs
- Every part of the graph should be meaningful
- Avoid:
  - Shading
  - Distracting colors
  - Decoration

Terrible Graphs



Less Terrible



Better Graph

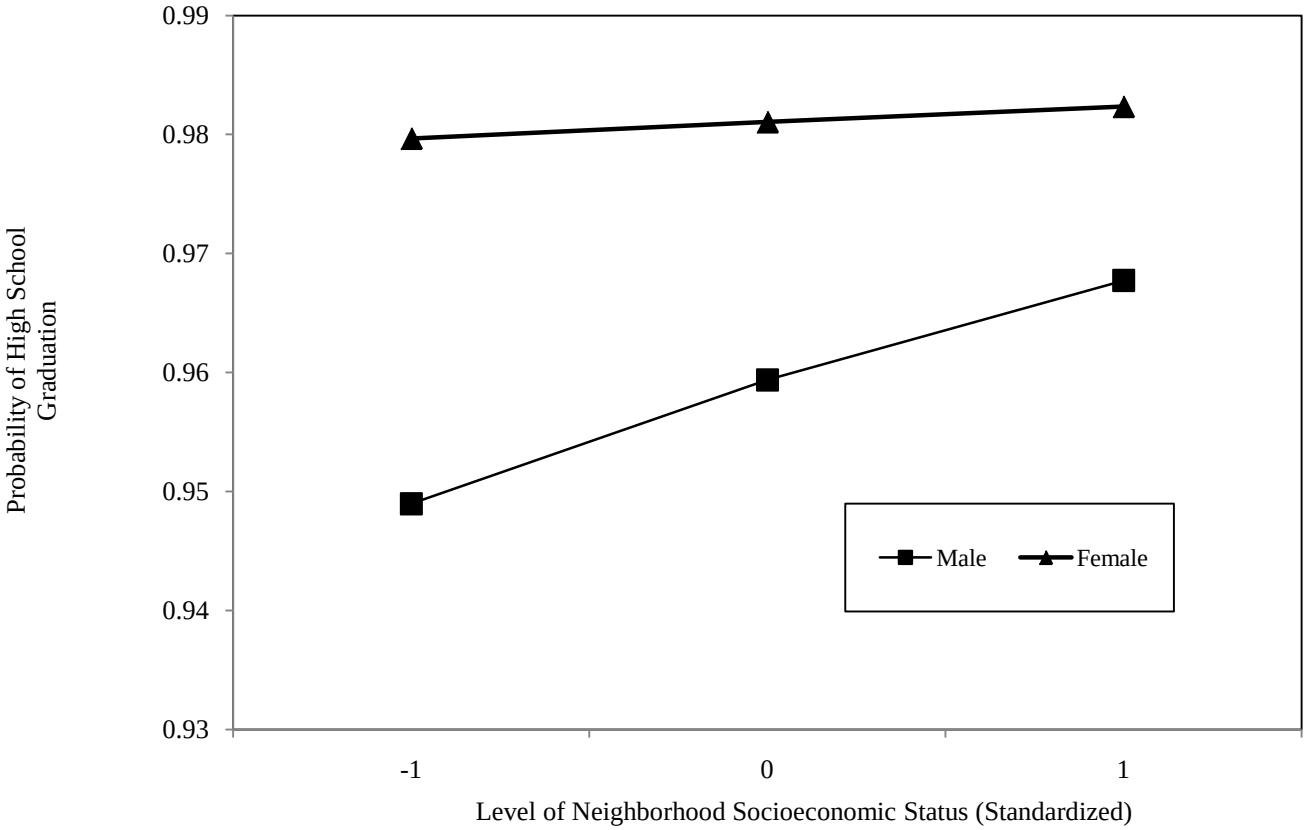


Figure 1. Two-way interaction of gender by the standardized measure of neighborhood socioeconomic status on probability of graduating from high school.

Opening Files in Stata

- When I open Stata, it tells me it's using the directory:
  - afs/athena.mit.edu/a/d/adlynch
- But, my files are located in:
  - afs/athena.mit.edu/a/d/adlynch/Graphing
- I'm going to tell Stata where it should look for my files:
  - cd "~/Graphing"

Basic Graphing

- Always know what you're working with before you get started
  - Recognize scale of data
  - If you're using multiple variables – how do their scales align?
- Before any graphing procedure review variables with codebook, sum, tab, etc.

- HELPFUL STATA HINT: If you want your command to go on multiple lines use " ///" at end of each line

Basic Graphing: Single Continuous Variables

Example: Histograms

- Stata assumes you're working with continuous data
- Very simple syntax:
  - hist varname
- Put a comma after your varname and start adding options
  - bin(#) : change the number of bars that the graph displays
  - normal : overlay normal curve
  - addlabels : add actual values to bars

Our First Dataset

- Time Magazine Public School Poll
- Based on survey of 1,000 adults in U.S.
- Conducted in August 2010
- Questions regarding feelings about parental involvement, teachers union, current potential for reform

Basic Graphing: Single Continuous Variables

Example: Histograms

- Change the numeric depiction of your data
- Add these options after the comma
  - Choose one: density fraction frequency percent
    - hist varname, percent

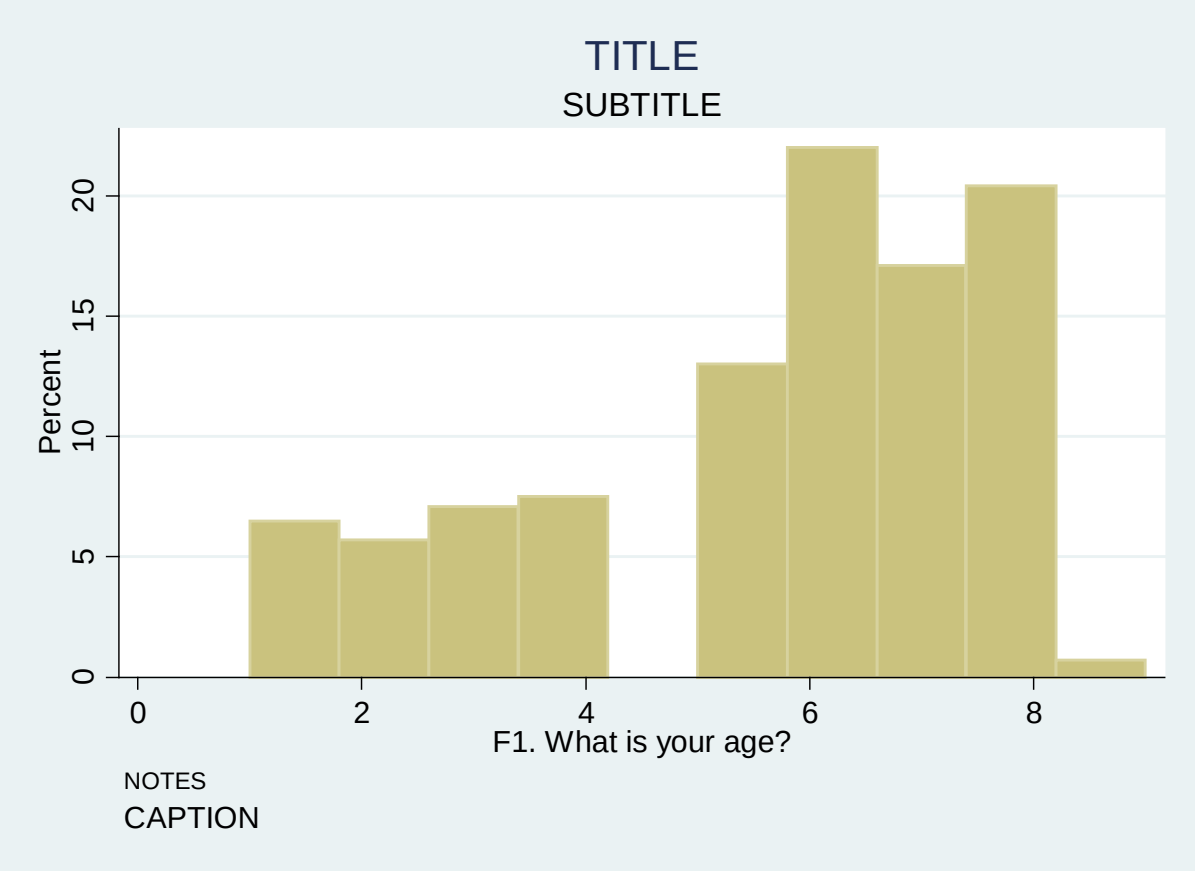
Basic Graphing: Single Continuous Variables

Example: Histograms

- Be sure to properly describe your histogram:
  - title(*insert name of graph*)
  - subtitle(*insert subtitle of graph*)
  - note(*insert note to appear at bottom of graph*)
  - caption(*insert caption to appear below notes*)

Basic Graphing: Single Continuous Variables

hist F1, bin(10) percent title(TITLE) subtitle(SUBTITLE) caption(CAPTION) note(NOTES)



Basic Graphing: Single Continuous Variables

Example: Histograms

- Axis title options (default is variable label):
  - xtitle(*insert x axis name*)
  - ytitle(*insert y axis name*)
- Don't want axis titles?
  - xtitle("")
  - ytitle("")

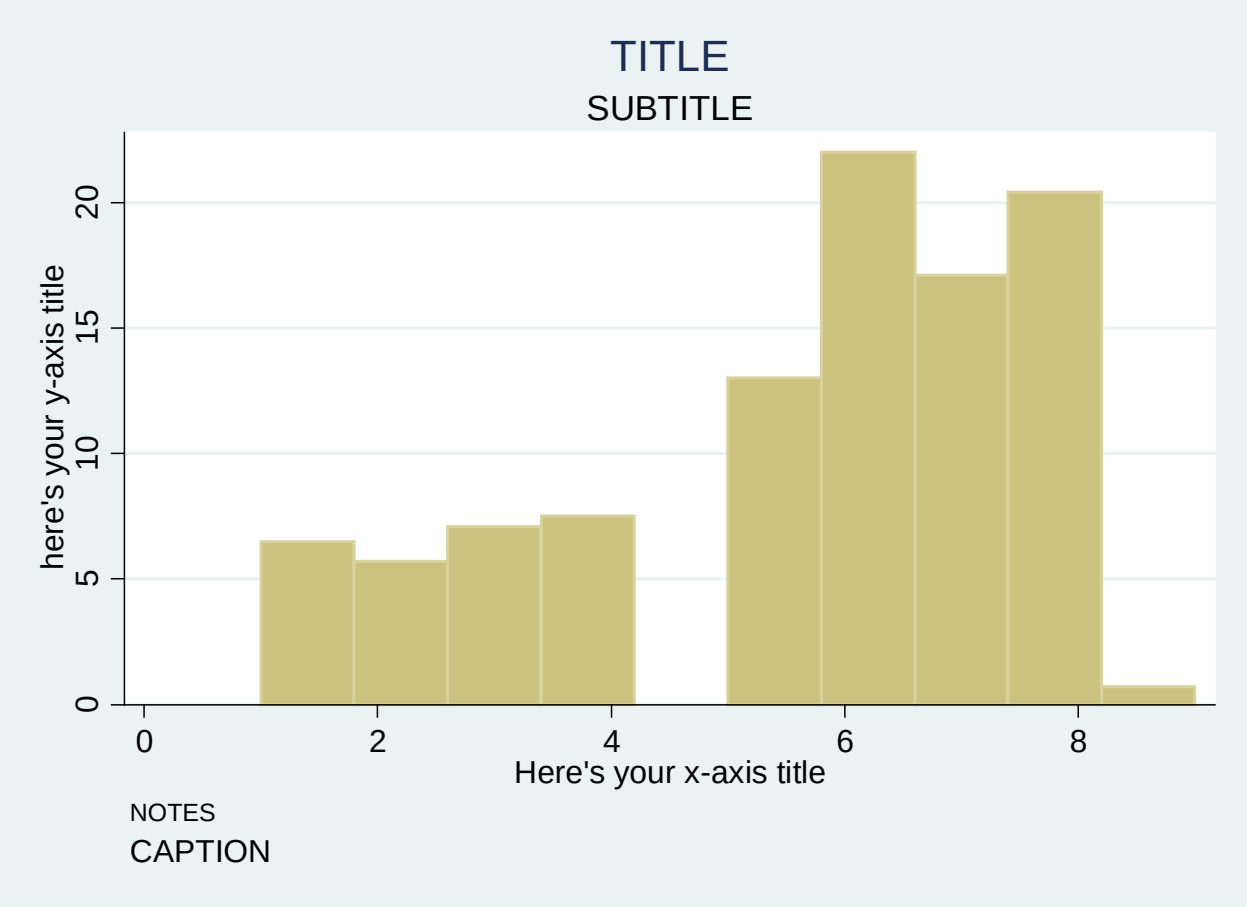
Basic Graphing: Single Continuous Variables

Example: Histograms

- Add labels to X or Y axis:
  - xlabel(*insert x axis label)*
  - ylabel(*insert y axis label)*
- Tell Stata how to scale each axis
  - xlabel(*start#(increment)end#)*
  - xlabel(0(5)100)
    - This would label x-axis from 0-100 in increments of 5

Basic Graphing: Single Continuous Variables

hist F1, bin(10) percent title(TITLE) subtitle(SUBTITLE) caption(CAPTION) ///
note(NOTES) xtitle(Here's your x-axis title) ytitle(here's your y-axis title)



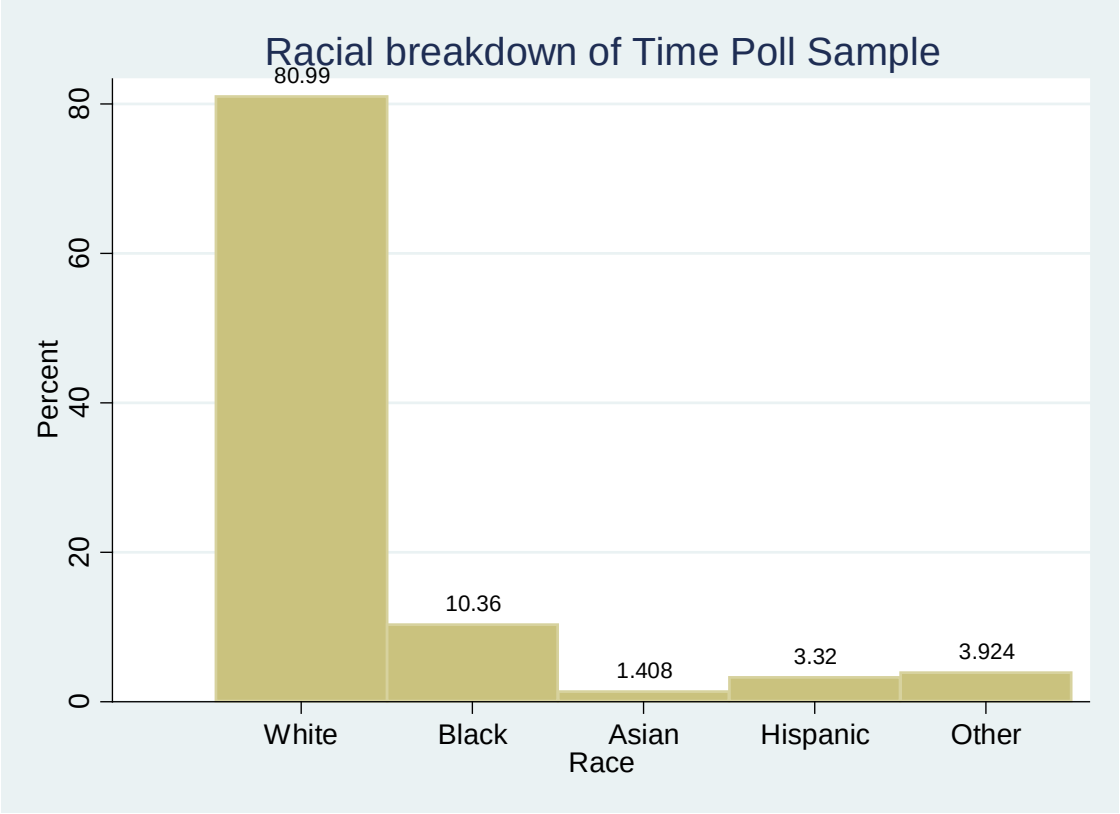
Basic Graphing: Single Categorical Variables

Example: Histograms

- What if my variable is not continuous?
  - Simply specify "discrete" with options
- Stata will produce one bar for each level (i.e. category) of variable
- Use xlabel command to insert names of individual categories
  - …, xlabel(1 "White" 2 "Black" 3 "Asian" 4 "Hispanic" 5 "Other")

Basic Graphing: Single Categorical Variables

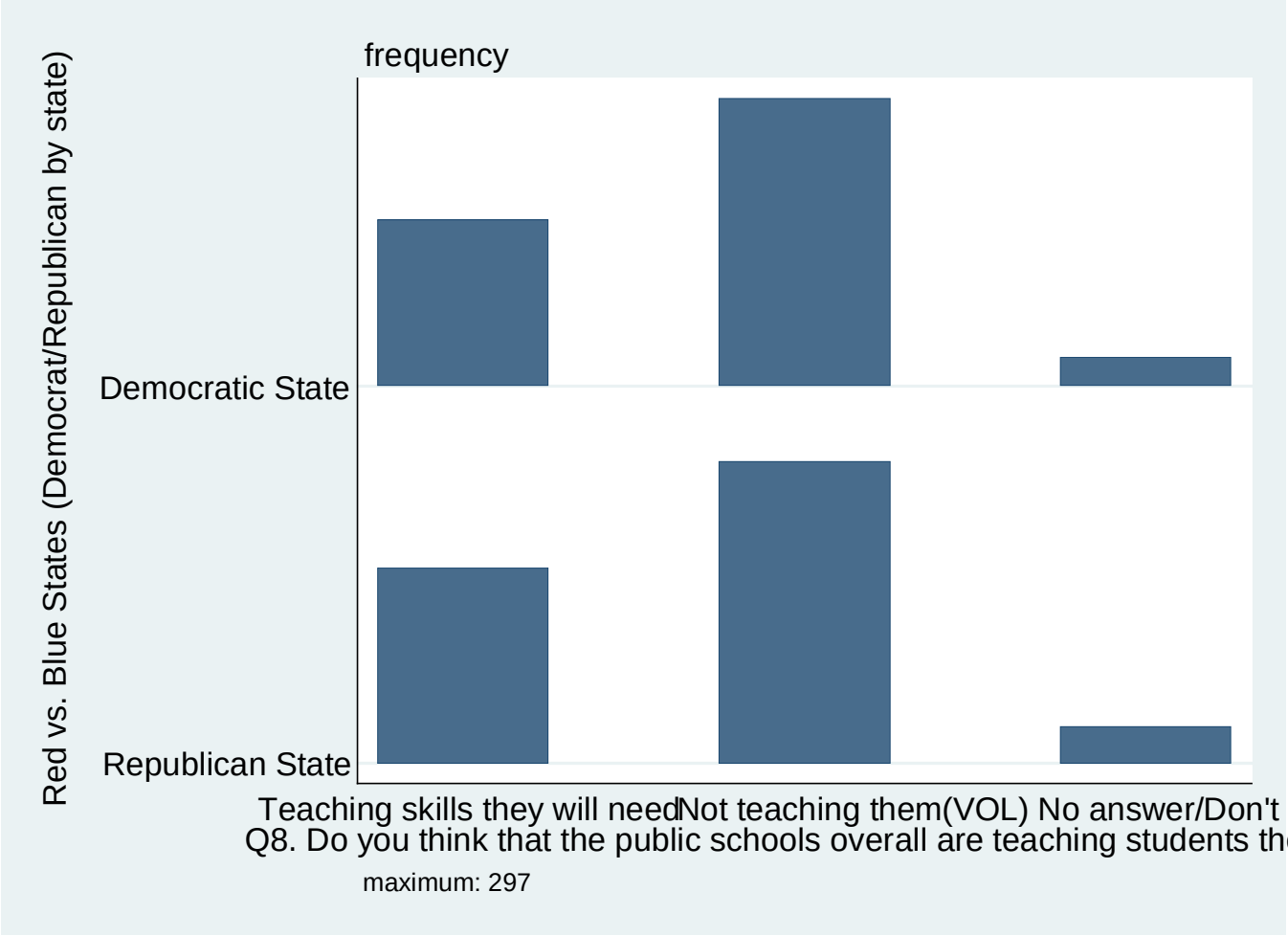
```
hist F4, title(Racial breakdown of Time Poll Sample) xtitle(Race) ///
ytitle(Percent) xlabel(1 "White" 2 "Black" 3 "Asian" 4 "Hispanic“ ///
 5 "Other") discrete percent addlabels
```



*Note my use of the “ ///” to allow the command to continue on multiple lines

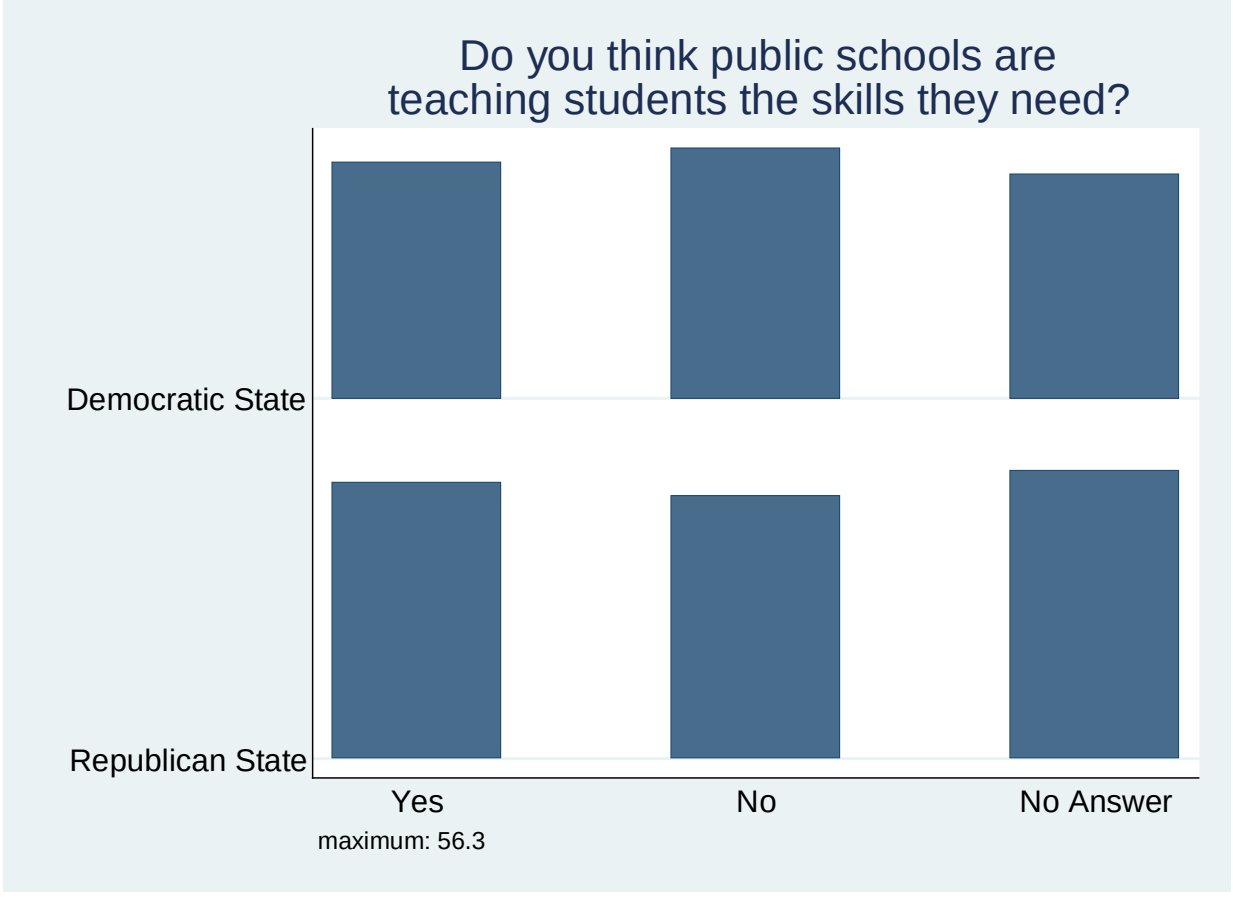
Comparing Responses Across Categorical Variables

tabplot rvb Q8



Comparing Responses Across Categorical Variables

```
tabplot rvb Q8, percent(Q8) title("Do you think public schools are" ///
 "teaching students the skills they need?") subtitle ("") xtitle("") ytitle("") ///
 xlabel(1 "Yes" 2 "No" 3"No Answer")
```



Exercise 1: Histograms and Tab Plots

The Twoway Family

- Next Dataset:
  - National Neighborhood Crime Study (NNCS)
  - N=9,593 census tracts in 2000
  - Explore sources of variation in crime for communities in the United States
    - Tract-level data: crime, social disorganization, disadvantage, socioeconomic inequality
    - City-level data: labor market, socioeconomic inequality, population change

The Twoway Family

- twoway is basic Stata command for all twoway graphs
- Use twoway anytime you want to make comparisons among variables
- Can be used to combine graphs (i.e., overlay one graph with another
  - e.g., insert line of best fit over a scatter plot

The Twoway Family

- Most basic:
  - tw scatter T_PERCAP T_VIOLNT
  - tw dropline T_PERCAP T_VIOLNT
  - tw lfitci T_PERCAP T_VIOLNT

Twoways and the By Statement

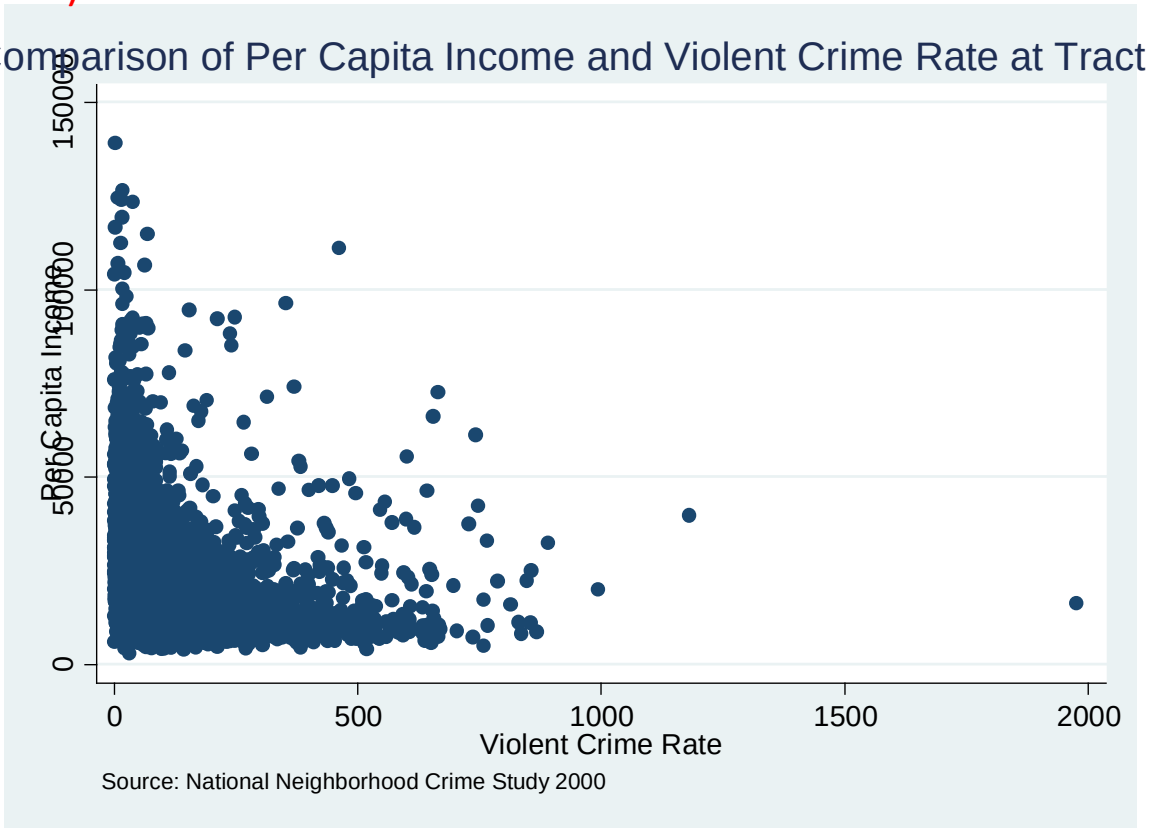
twoway scatter T_PERCAP T_VIOLNT, by(DICEMP)

Twoway Title Options

- Same title options as with histogram
  - title(*insert name of graph*)
  - subtitle(*insert subtitle of graph*)
  - note(*insert note to appear at bottom of graph*)
  - caption(*insert caption to appear below notes*)

Twoway Title Options

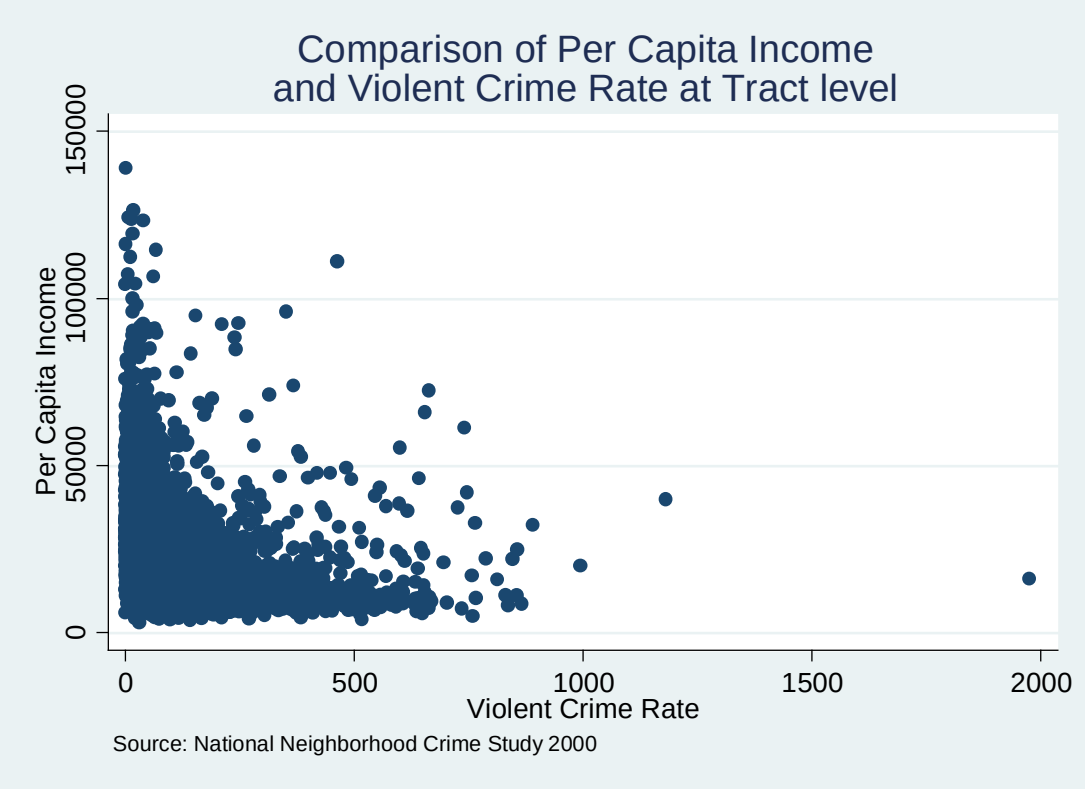
twoway scatter T_PERCAP T_VIOLNT, title(Comparison of Per Capita Income and Violent Crime Rate at Tract level) ///
xtitle(Violent Crime Rate) ytitle(Per Capita Income) note(Source: National Neighborhood Crime Study 2000)



Let's fix that graph title – it is too cramped….

Twoway Title Options

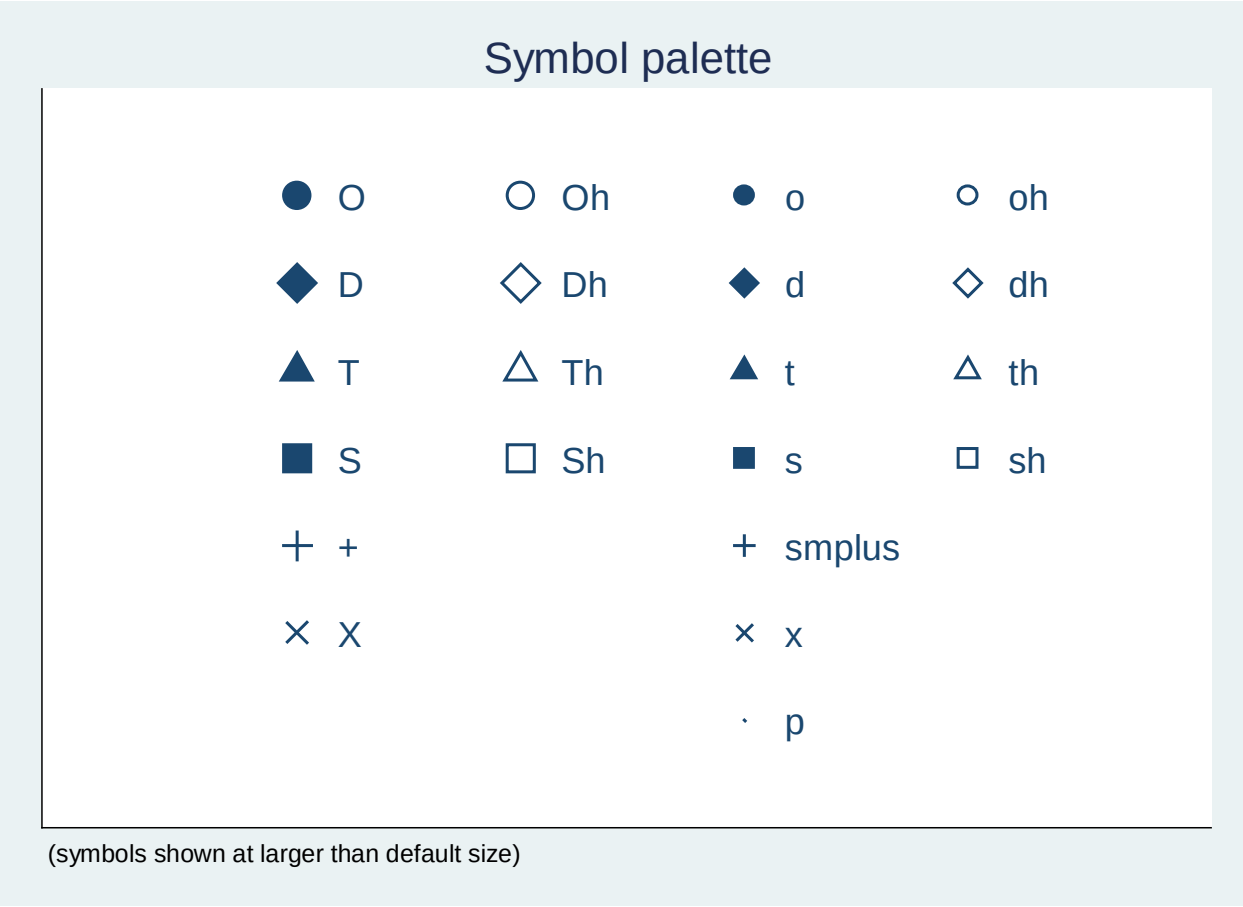
```
twoway scatter T_PERCAP T_VIOLNT, title("Comparison of Per Capita Income" ///
"and Violent Crime Rate at Tract level") ///
xtitle(Violent Crime Rate) ytitle(Per Capita Income) ///
note(Source: National Neighborhood Crime Study 2000)
```



*Note how we got our title to go onto two lines

Twoway Symbol Options

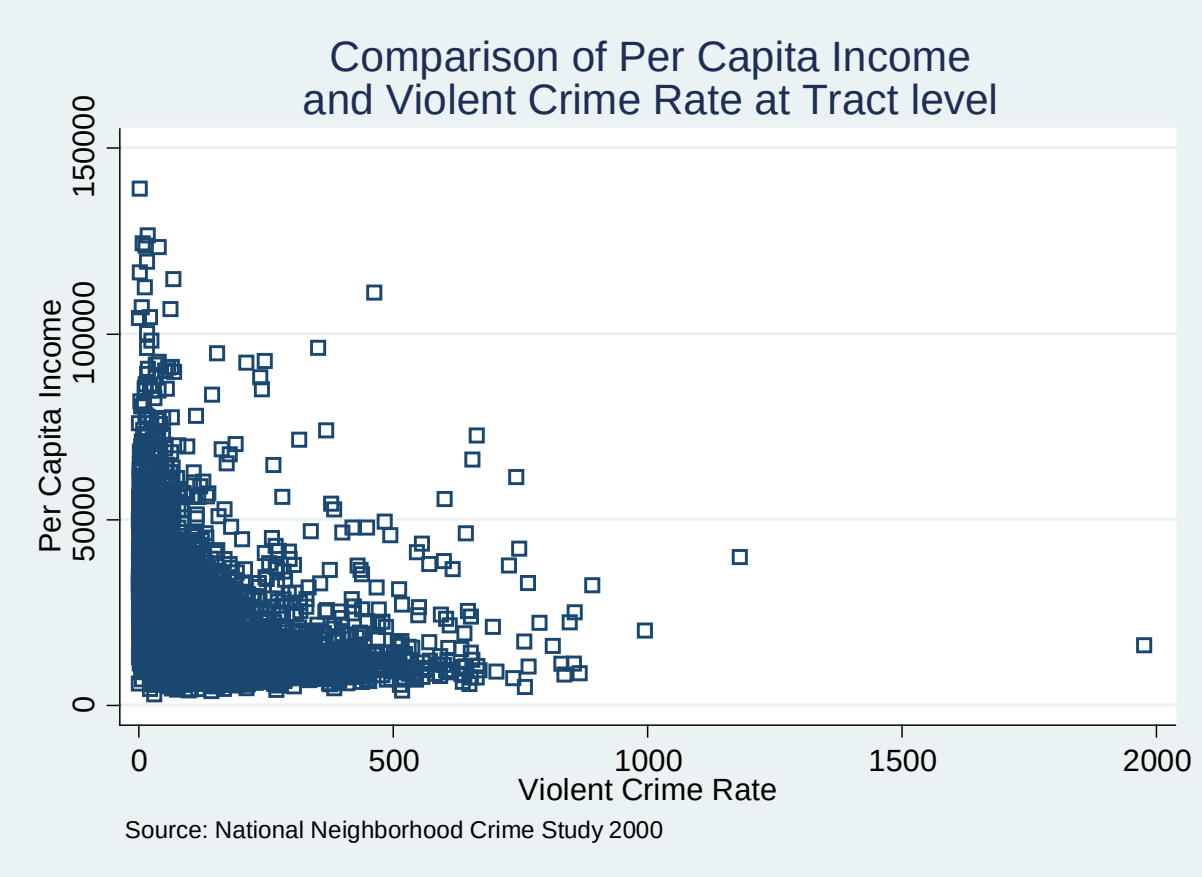
- To call this chart up in Stata, type: palette symbolpalette
- Use msymbol() in graph options to change symbol



Twoway Symbol Options

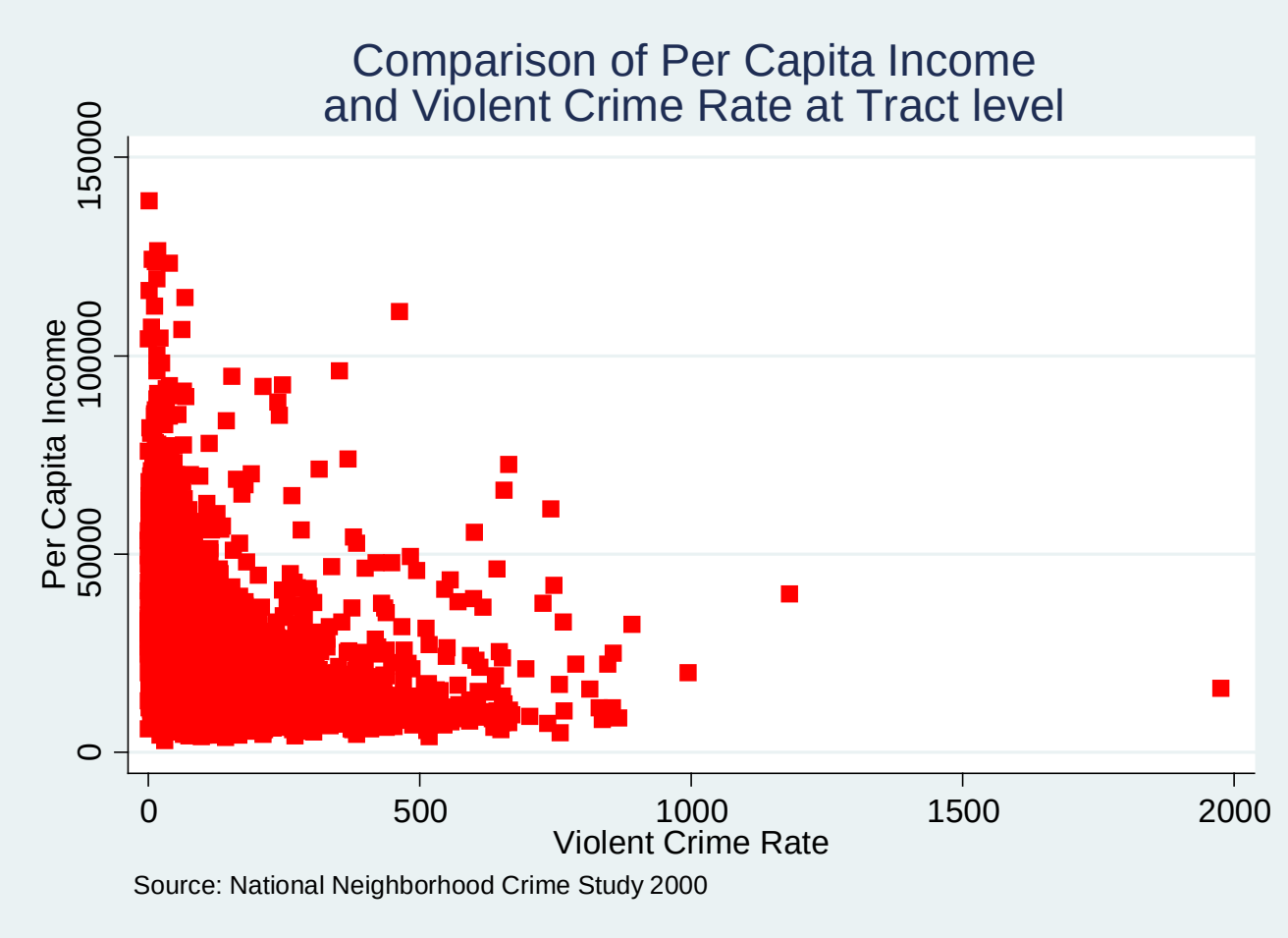
```
twoway scatter T_PERCAP T_VIOLNT, title("Comparison of Per Capita Income" ///
"and Violent Crime Rate at Tract level") ///
xtitle(Violent Crime Rate) ytitle(Per Capita Income) ///
note(Source: National Neighborhood Crime Study 2000) ///
msymbol(Sh)
```

Here's my msymbol() option



Twoway Symbol Options

Add "mcolor(*insert color*)" option to change color of symbol. Here, I just added "mcolor(red)" to the graph options.

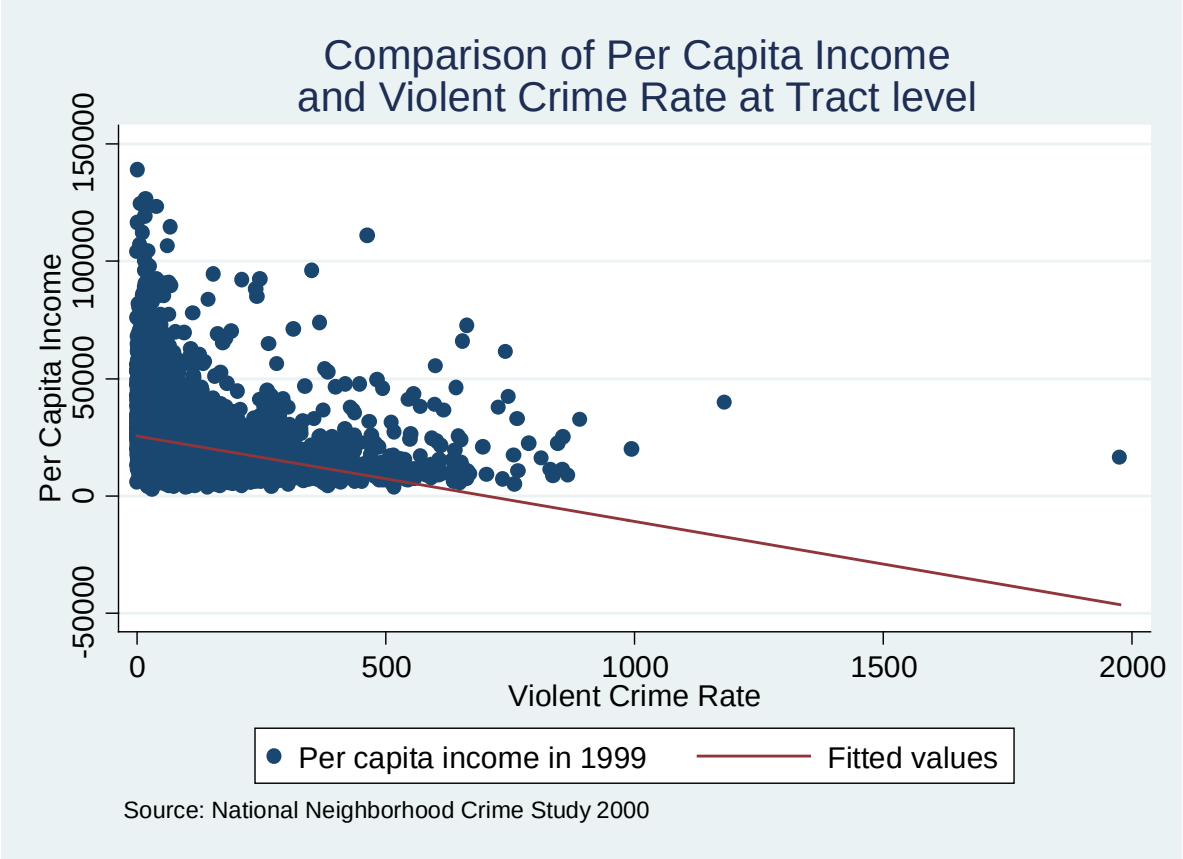


Overlaying Twoway Graphs

- Very simple to combine multiple graphs…just put each graph command in parentheses
  - twoway (scatter var1 var2) (lfit var1 var2)
- Add individual options to each graph within the parentheses
- Add overall graph options as usual following the comma
  - twoway (scatter var1 var2) (lfit var1 var2), options

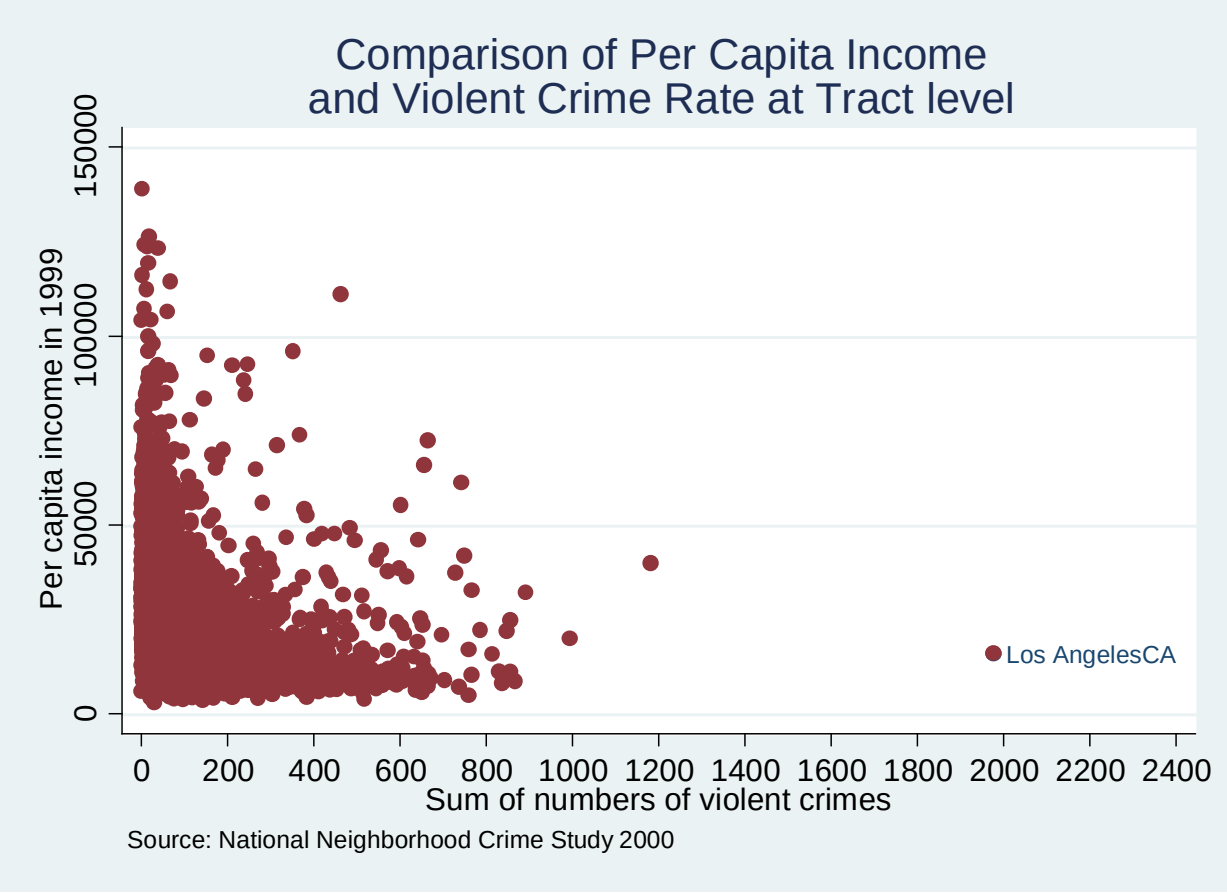
Overlaying Twoway Graphs

```
twoway (scatter T_PERCAP T_VIOLNT) (lfit T_PERCAP T_VIOLNT), title("Comparison of ///
Per Capita Income" "and Violent Crime Rate at Tract level“) ///
xtitle(Violent Crime Rate) ytitle(Per Capita Income) note(Source: National ///
Neighborhood Crime Study 2000)
```



Overlaying Twoway Graphs

```
twoway (scatter T_PERCAP T_VIOLNT if T_VIOLNT==1976, mlabel(CITY)) (scatter T_PERCAP T_VIOLNT), ///
title("Comparison of Per Capita Income" "and Violent Crime Rate at Tract level") xlabel(0(200)2400) ///
note(Source: National Neighborhood Crime Study 2000) legend(off)
```



Exercise 2: The TwoWay Family

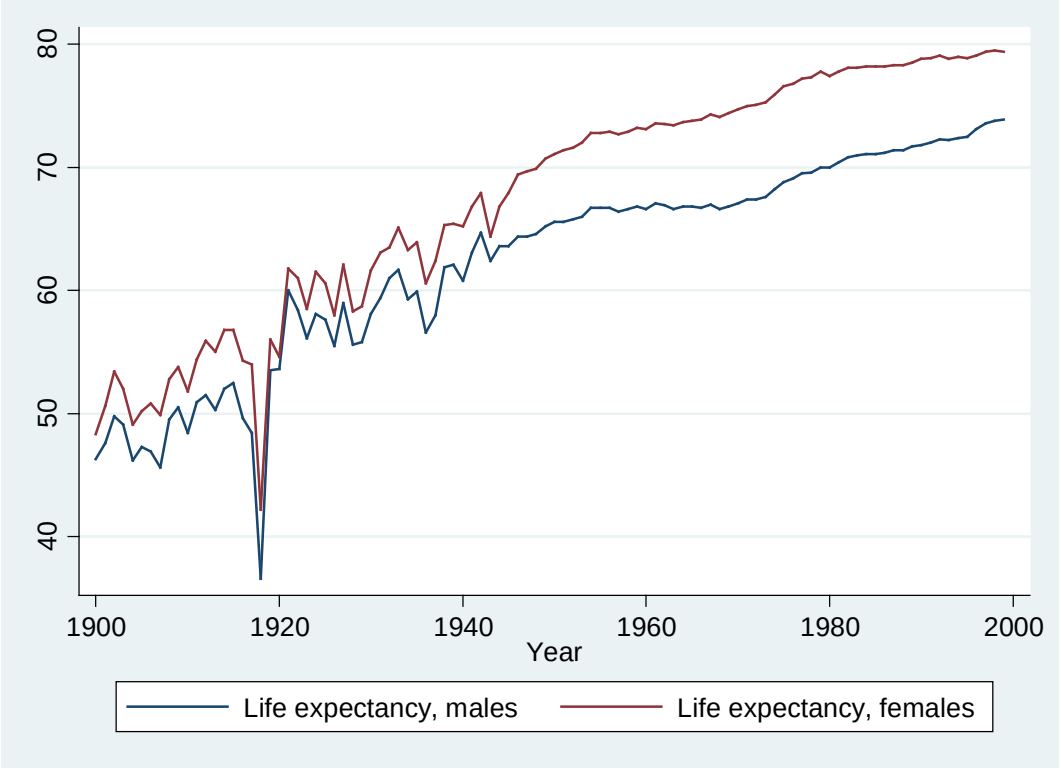
Line Graphs

- Line graphs helpful for a variety of data
  - Especially any type of time series data
- We'll use data on US life expectancy from 1900-1999
  - webuse uslifeexp, clear
- ok

Line Graphs

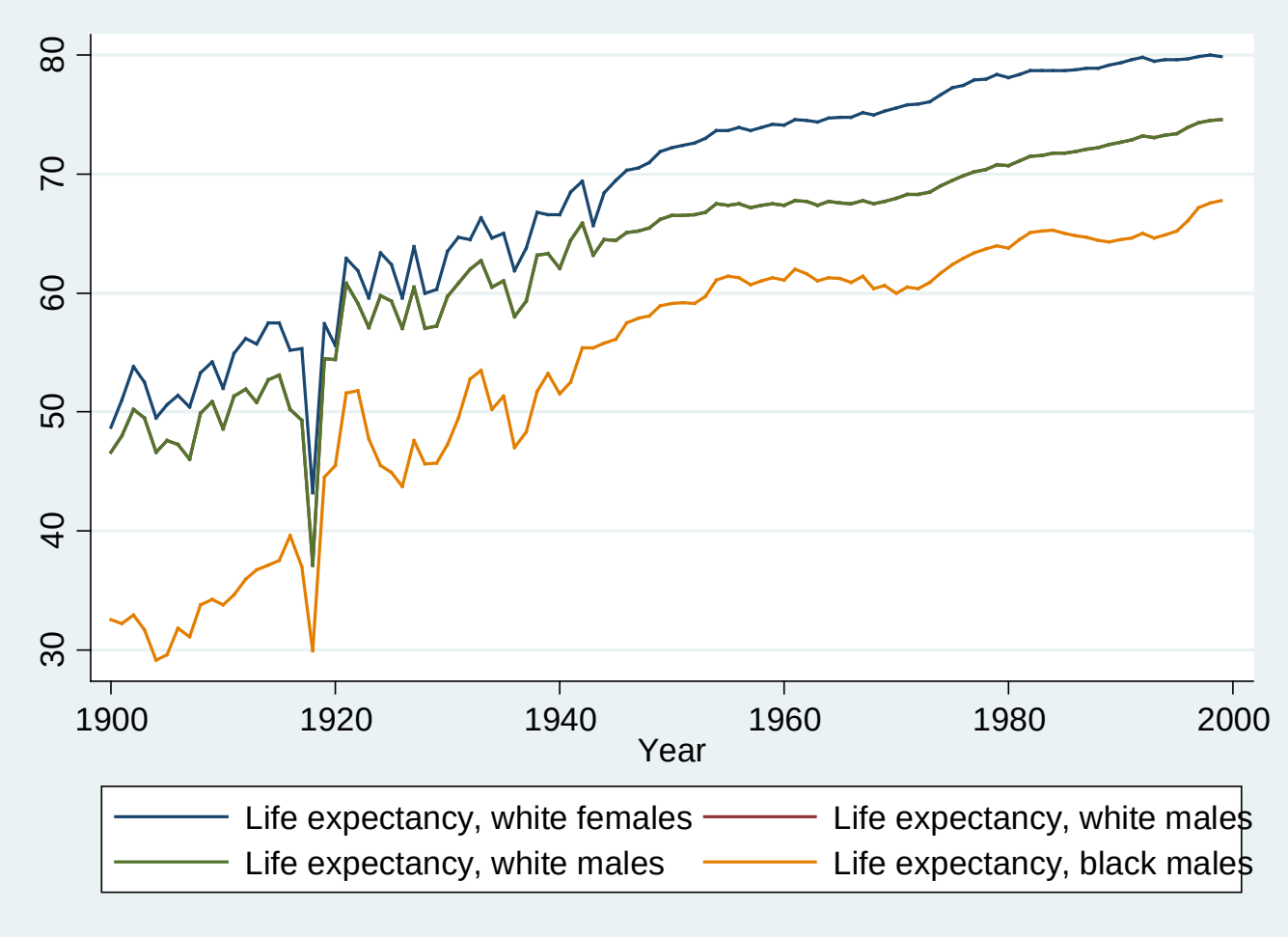
line le_wm le_bm year

*Simple line graph of men and women overtime



Line Graphs

line le_wfemale le_wmale le_wm le_bm year

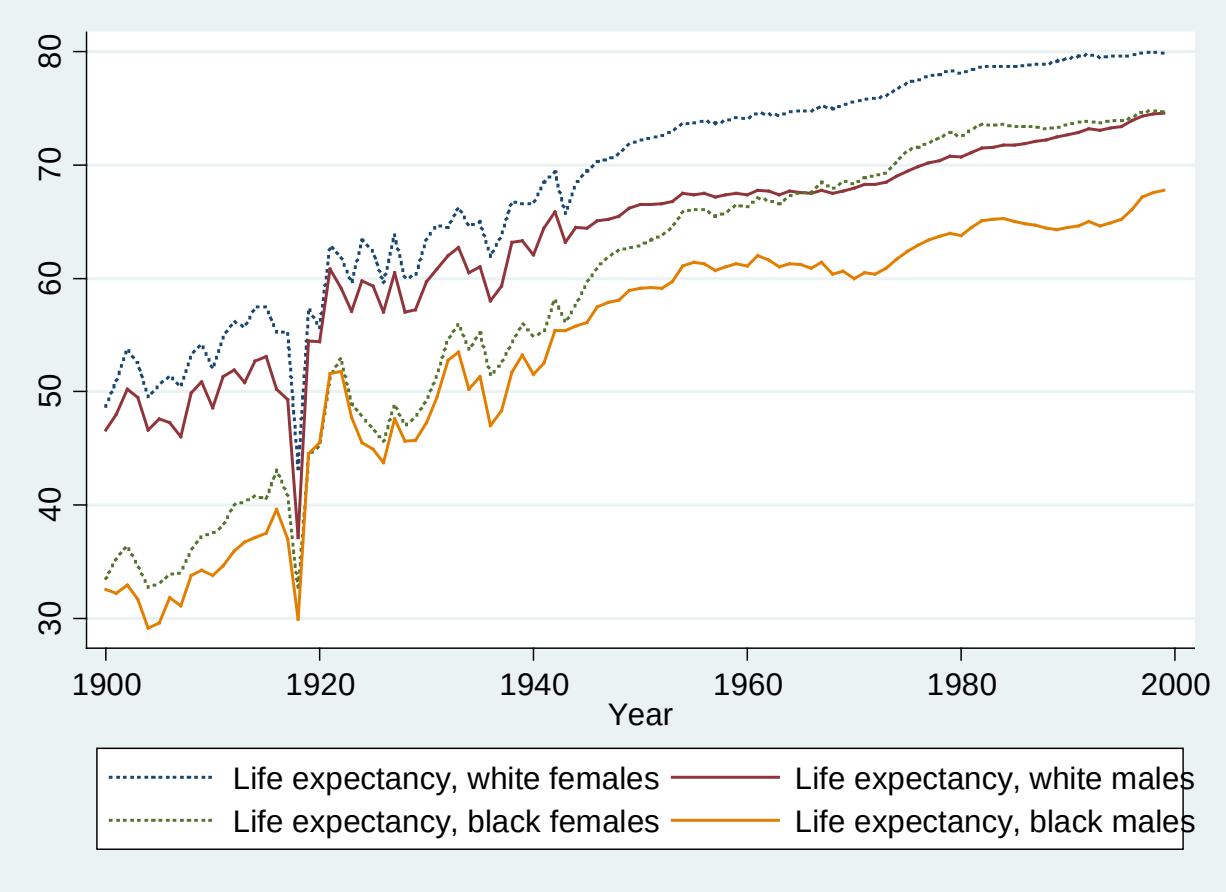


Line Graphs: Adding Options

- As usual…just keep adding options after the comma!
- Same rules apply for titles that we've already seen for histograms and the twoway graphs
- Let's review how we can play with the appearance of our lines
- Full listing of options type "help line_options"

Line Graphs: Changing Options

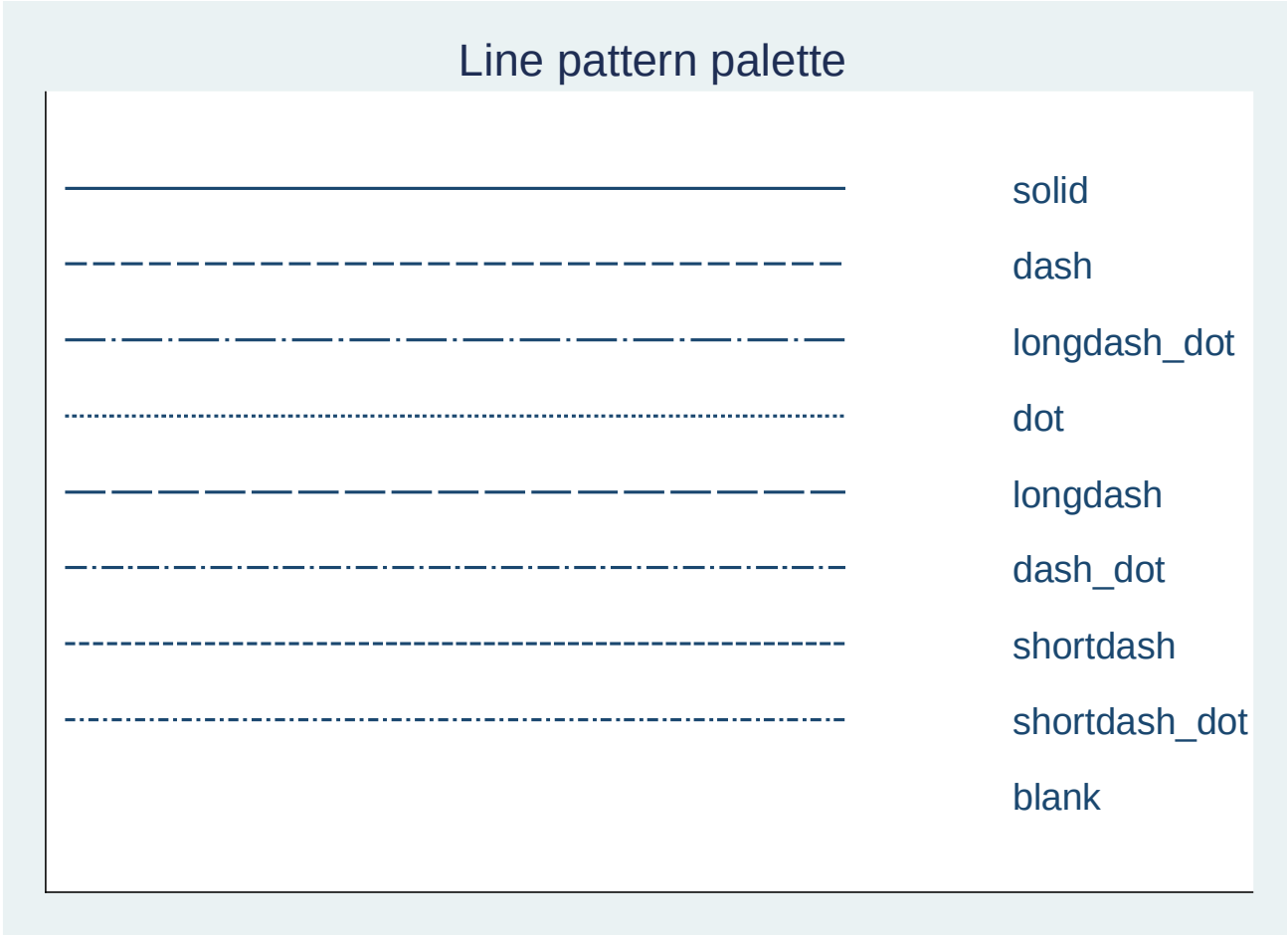
line le_wfemale le_wmale le_bf le_bm year, lpattern(dot solid dot solid)



"lpattern()" command allows me to change pattern from solid to dotted

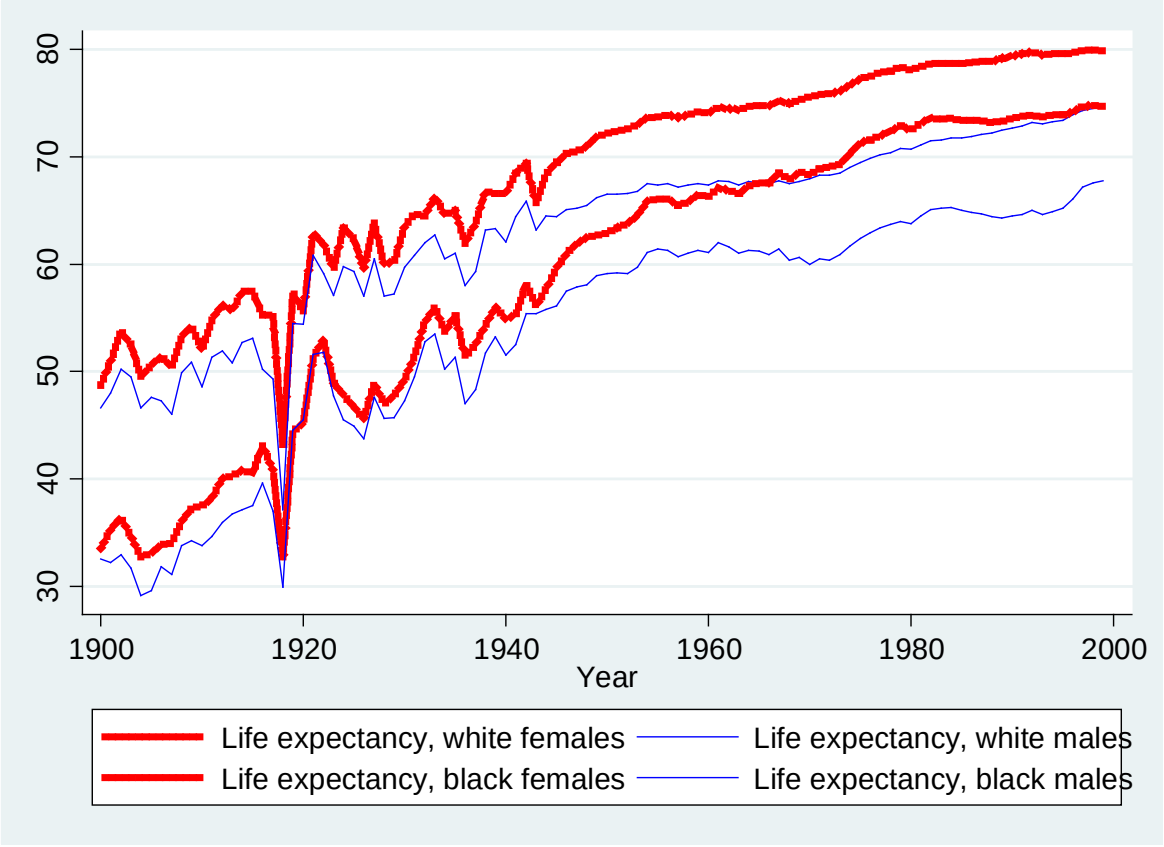
Stata Graphing Lines

To call this up in Stata, type: palette linepalette



Line Graphs: Changing Options

```
line le_wfemale le_wmale le_bf le_bm year, lpattern(dot solid dot solid) ///
lcolor(red blue red blue) lwidth(thick thin thick thin)
```



Now I've used several different options to change line pattern, color and width

Profile Plots

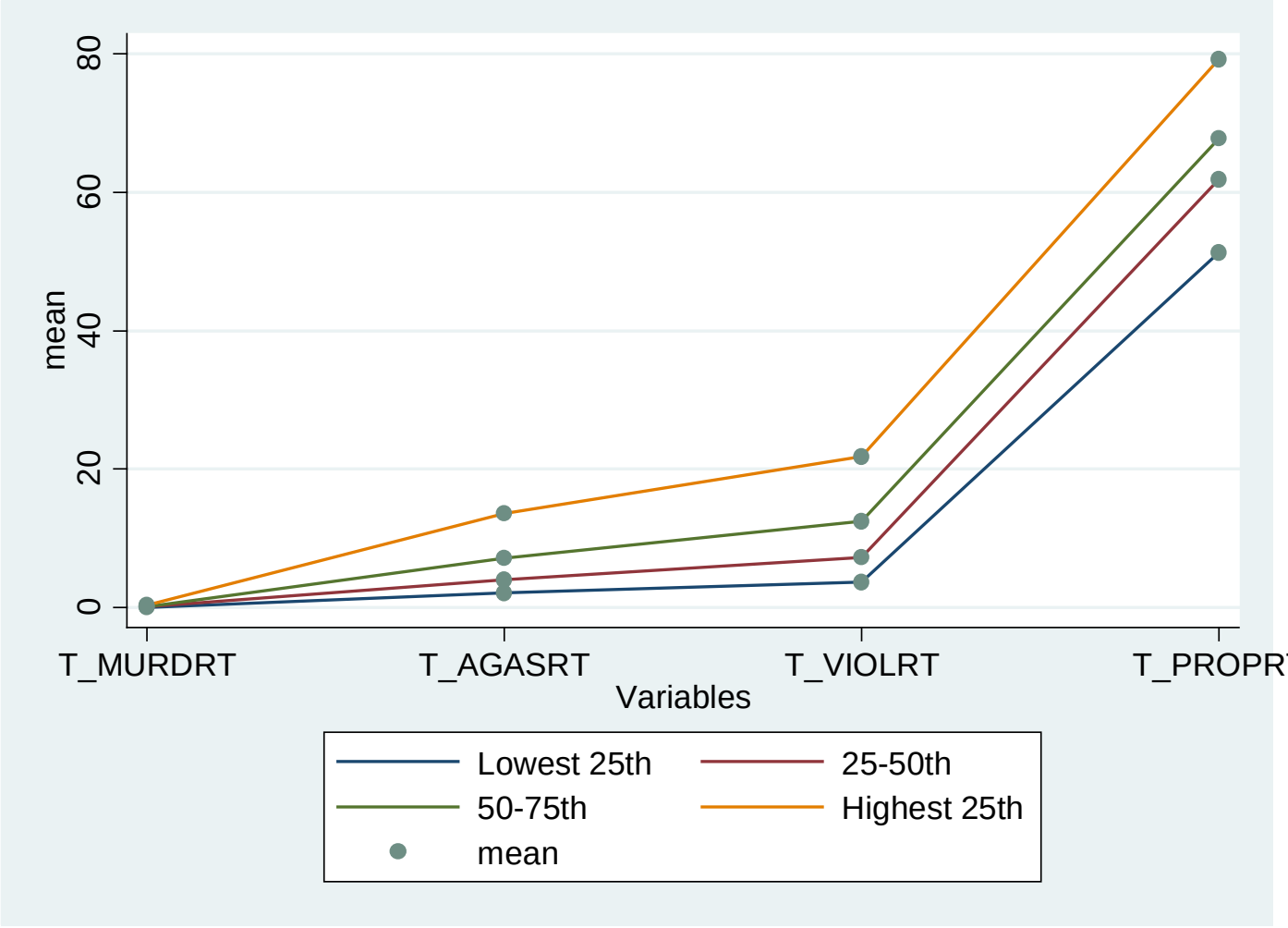
- Great way for comparing outcomes on continuous variables across different levels of categorical variables
- Example: math, science and reading scores (continuous variables) across different curriculum programs
- Profile plots is a Stata add-on (not in base package)
  - findit profileplot

Profile Plot

- Let's go back to the National Crime Survey and look at crime rates across different levels of unemployment at the tract level
- First, create categorical variable separating unemployment rates into quartiles
  - *pay attention to what happens with missing data
- Label new variable

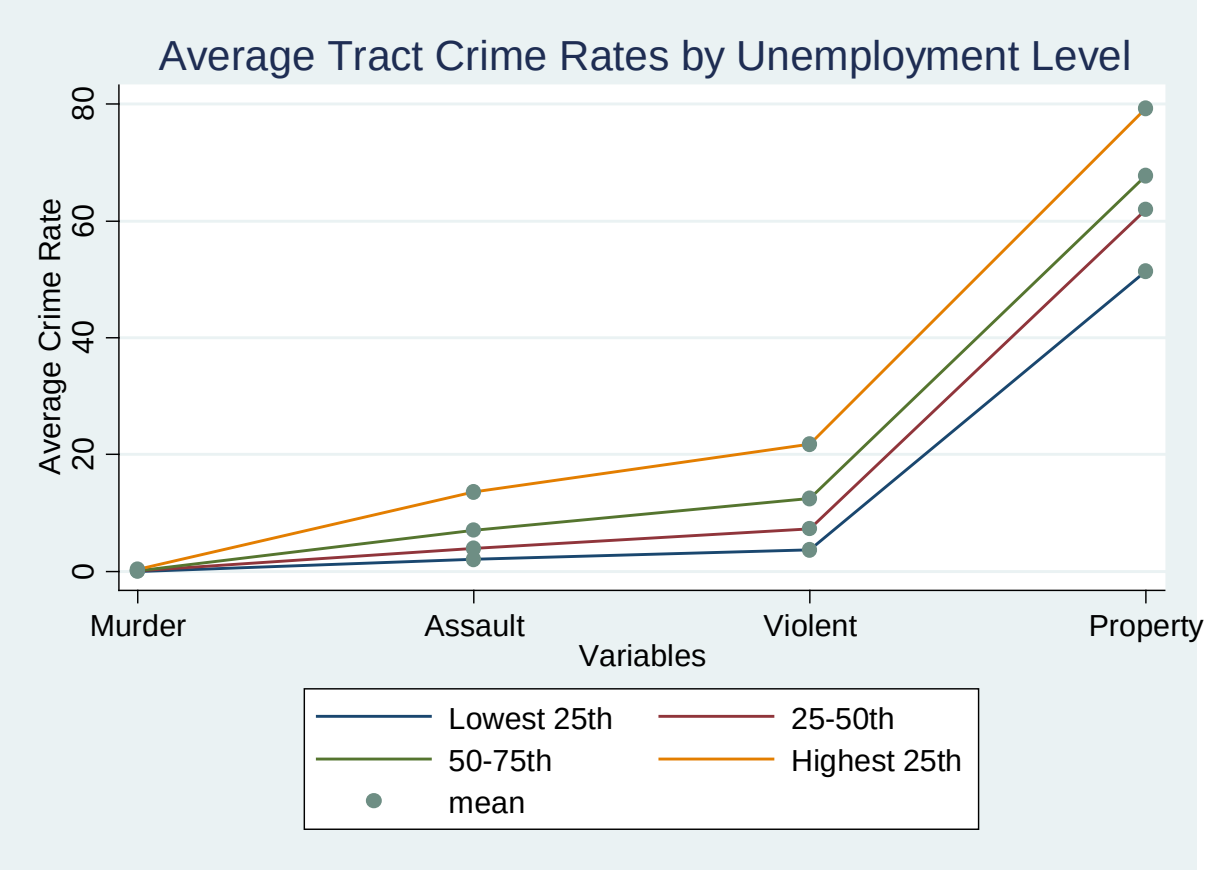
Profile Plots

profileplot T_MURDRT T_AGASRT T_VIOLRT T_PROPRT, by(unempquart)



Profile Plots

profileplot T_MURDRT T_AGASRT T_VIOLRT T_PROPRT, by(unempquart) xlabel(1 "Murder" 2 "Assault" 3 "Violent" 4 "Property") ///
ytitle(Average Crime Rate) title("Average Tract Crime Rates by Unemployment Level") xtitle("")



Exporting Graphs

- From Stata, right click on image and select "save as" or try syntax:
  - cd "~/Graphing"
  - graph export myfig.esp, replace
- In Microsoft Word: **insert > picture > from file**
  - Or, right click on graph in Stata and copy and paste into Word

Other Services Available

- MIT's membership in HMDC provided by schools and departments at MIT
- Institute for Quantitative Social Science
  - www.iq.harvard.edu
- Research Computing
  - www.iq.harvard.edu/research_computing
- Computer labs
  - www.iq.harvard.edu/facilities
- Training
  - www.iq.harvard.edu/training
- Data repository
  - http://libraries.mit.edu/get/hmdc

Thank you!

Thank you for participating in HMDC's Introduction to Stata Workshop.

We offer additional statistical workshops in Stata, SAS and R throughout the semester:

Introduction to R:
Monday December 6th: 1-4pm

*Note: This workshop is currently wait listed but will be offered again over IAP

Introduction to SAS:
Monday November 15th: 1-4pm

Sign up at:

http://libraries.mit.edu/guides/subjects/data/training/workshops.html

Thank you!

Can't make it to the workshops at MIT? MIT users are also welcome to attend these same workshops at Harvard. Sign up anytime by emailing: dataclass@help.hmdc.harvard.edu

Graphics in Stata:
Fri, Nov. 19th: 9 am to Noon

Introduction to R:
Fri, Dec. 3rd: 9 am to Noon

Introduction to SAS:
Fri, Nov. 5th: 9 am to Noon

http://support.hmdc.harvard.edu/kb-20/statistical_support