STATA - Graphs

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

April 12, 2022

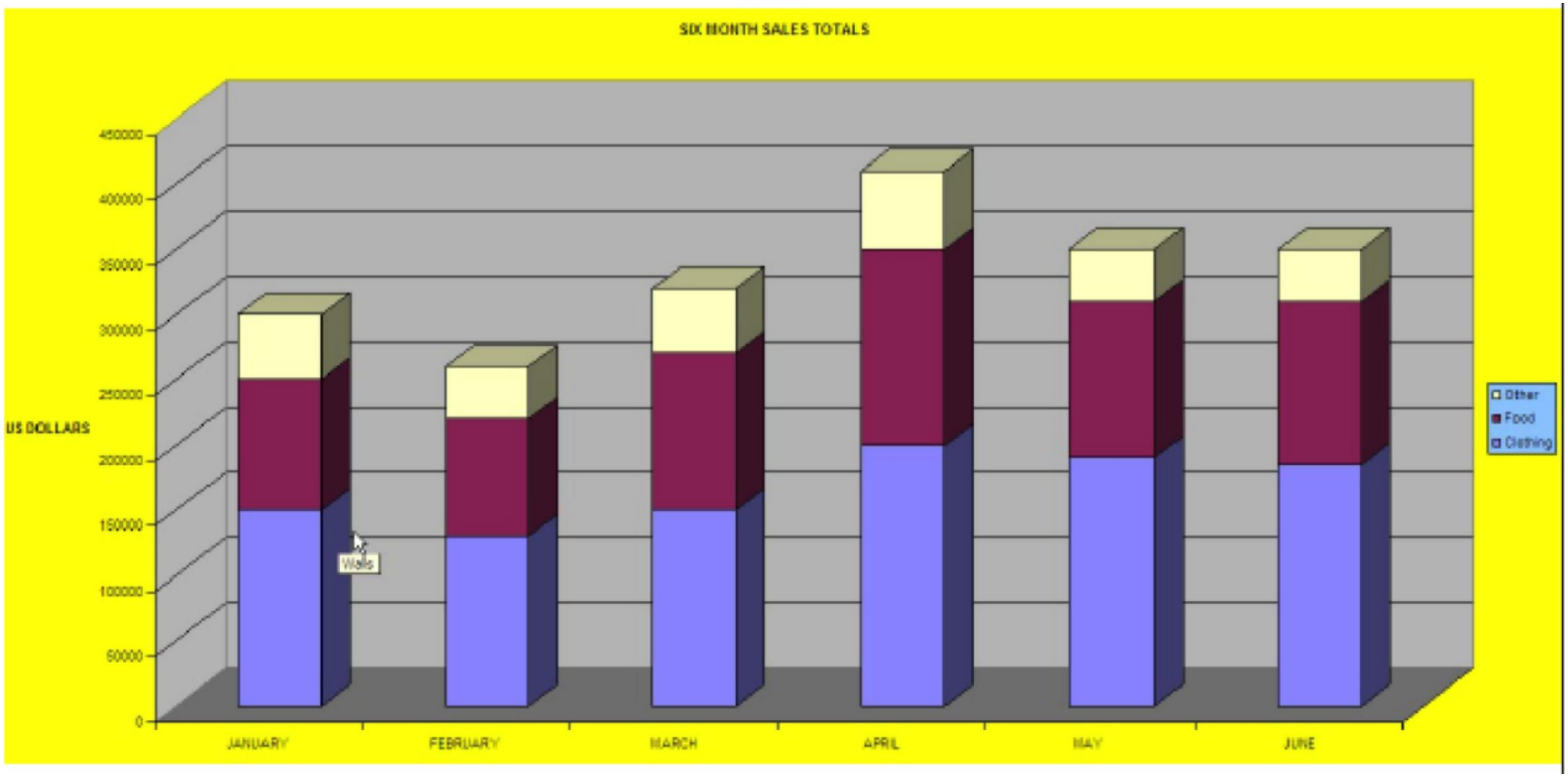
Graphing Strategies

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

Graphing Strategies

- 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.
- If you want your command to go on multiple lines use /// at end of each line

Terrible Graph



Much Better Graph

Source: Bureau of Labor Statistics, http://www.bls.gov/data/

Graphs

- Please see **C6_graphs_v2.do**

histogram: Create Histogram Graphs

Histograms

- It is useful if you want to explore **single continuous/discrete variables**
- Very simple syntax:
  - `historgam varname` or `hist varname`

STATA Command: histogram

- Syntax:

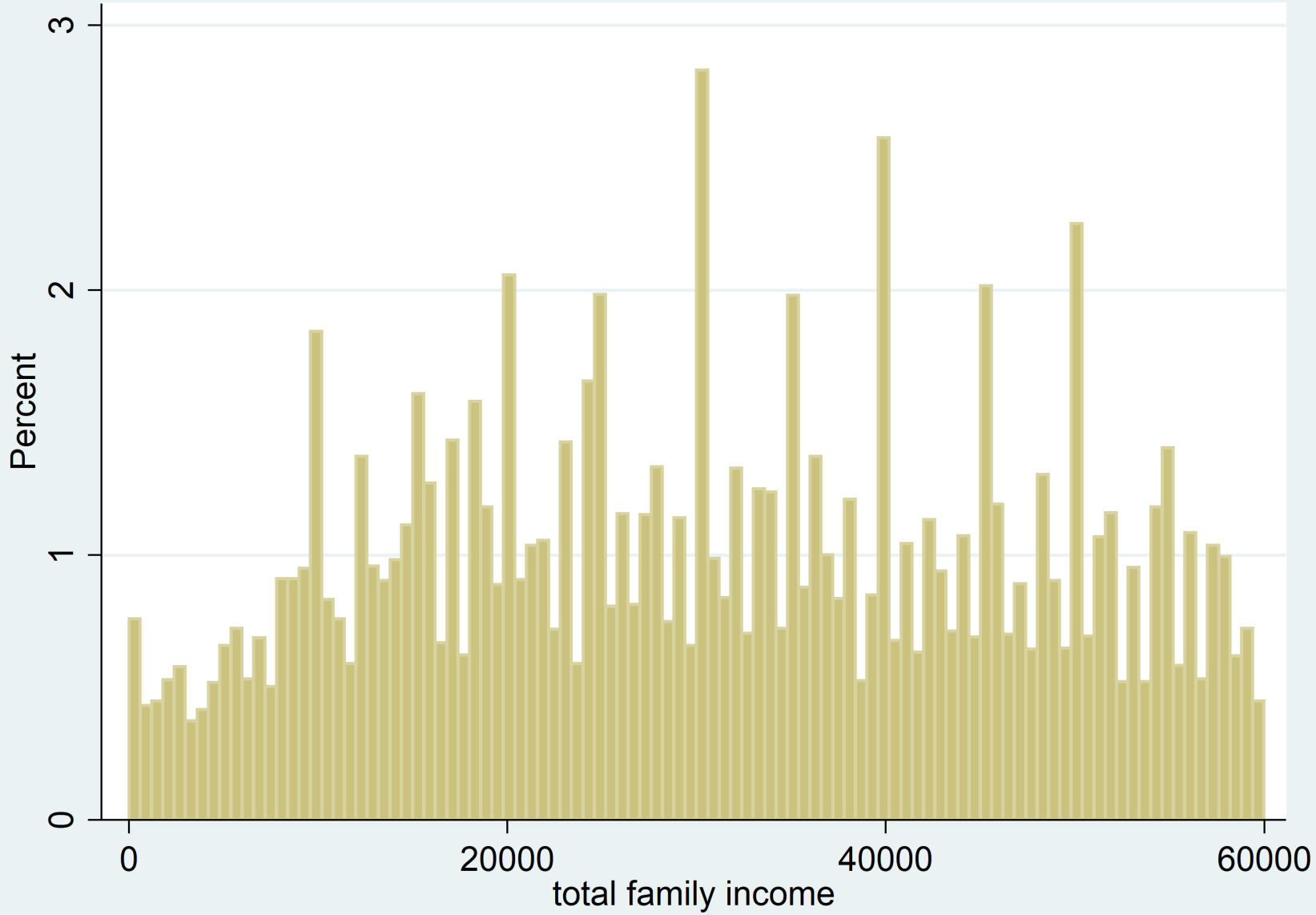
```
histogram varname [if] [in] [weight] [, [
    continuous_opts | discrete_opts] options]
```

- Example:

```
sum inctot,d
hist ftotval if ftotval< 60000 & ftotval>0, bin
    (100) percent
```

- Put a comma after your varname and start adding options
  - `bin(#)`: change the number of bars that the graph displays
  - `percent`: show percentage instead of density
    - To change the numeric depiction of your data add these options after the comma
    - Choose one: density, fraction, frequency, percent

Graph Example



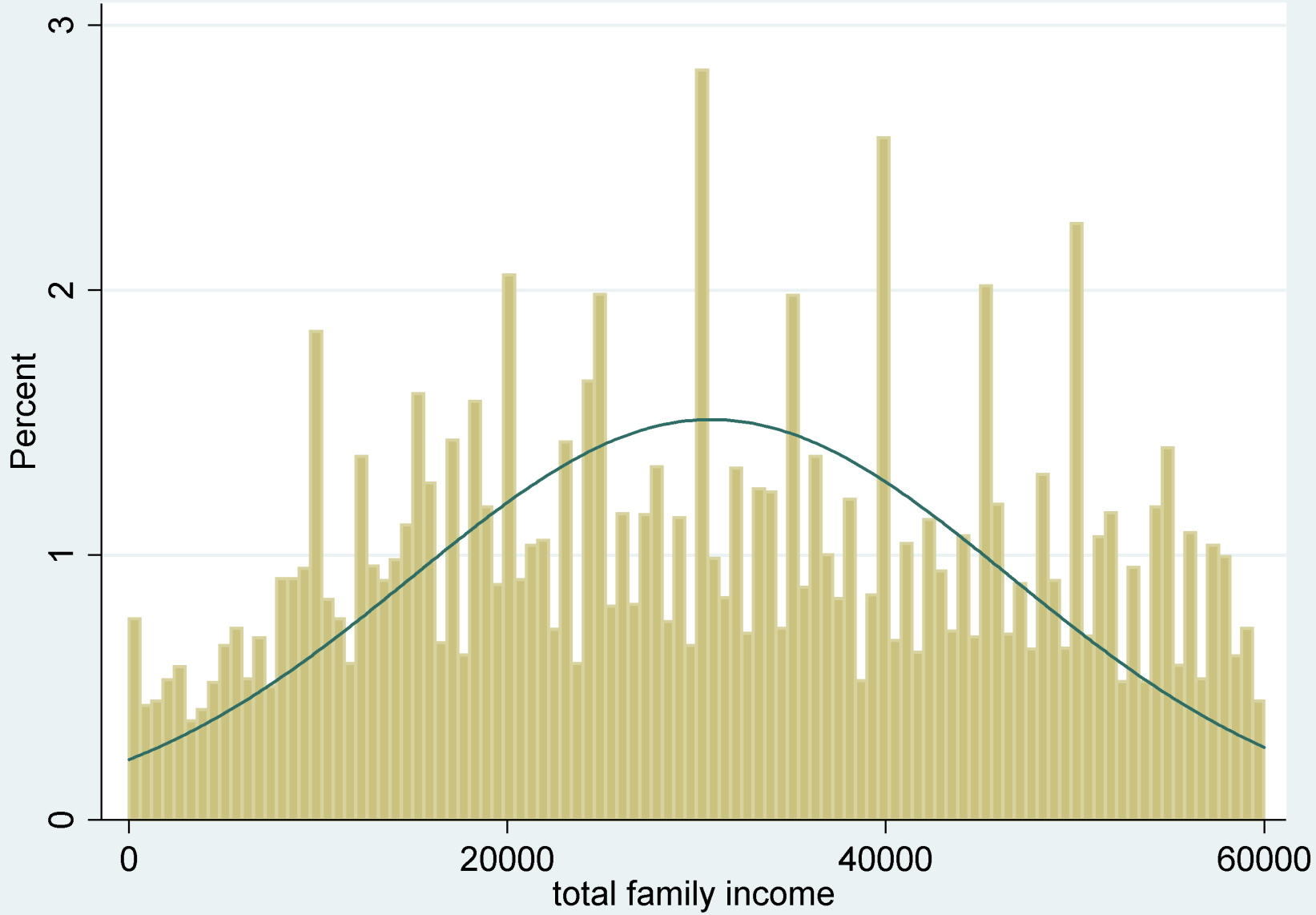
STATA Command: histogram

- Example:

```
1 hist ftotval if ftotval< 60000 & ftotval>0, bin
      (100) percent normal
```

- Put a comma after your varname and start adding options
  - `normal`: overlay normal curve

Graph Example



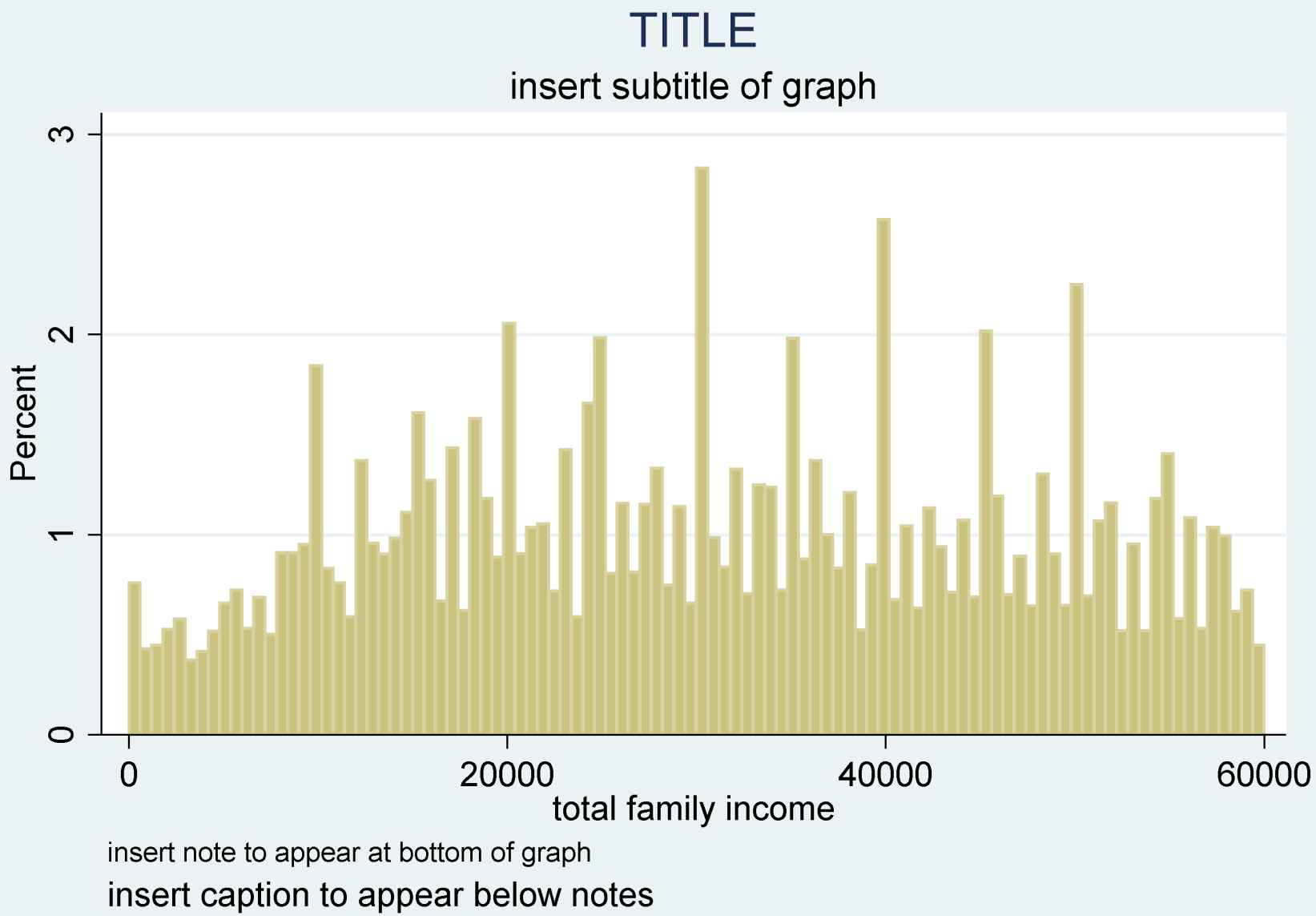
STATA Command: histogram

- Example:

```
1 hist ftotval if ftotval< 60000 & ftotval>0, bin
      (100) percent title(TITLE) subtitle(insert
      subtitle of graph) ///
2 note(insert note to appear at bottom of graph)
      caption(insert caption to appear below notes)
```

- 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)`

Graph Example

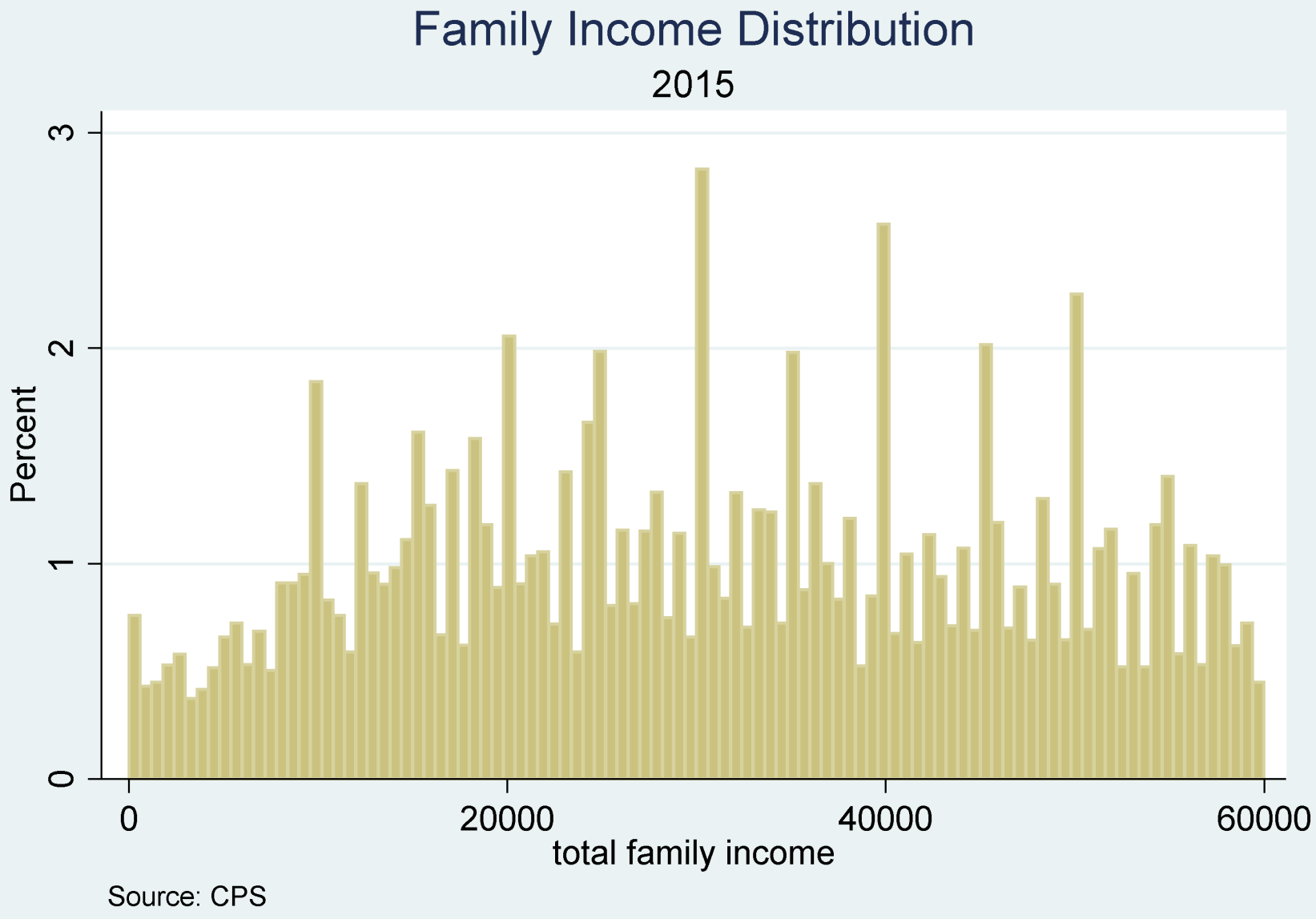


STATA Command: histogram

- Example:

```
1 hist ftotval if ftotval< 60000 & ftotval>0, bin
      (100) percent title(Family Income
      Distribution) subtitle(2015) note(Source: CPS
      )
```

Graph Example



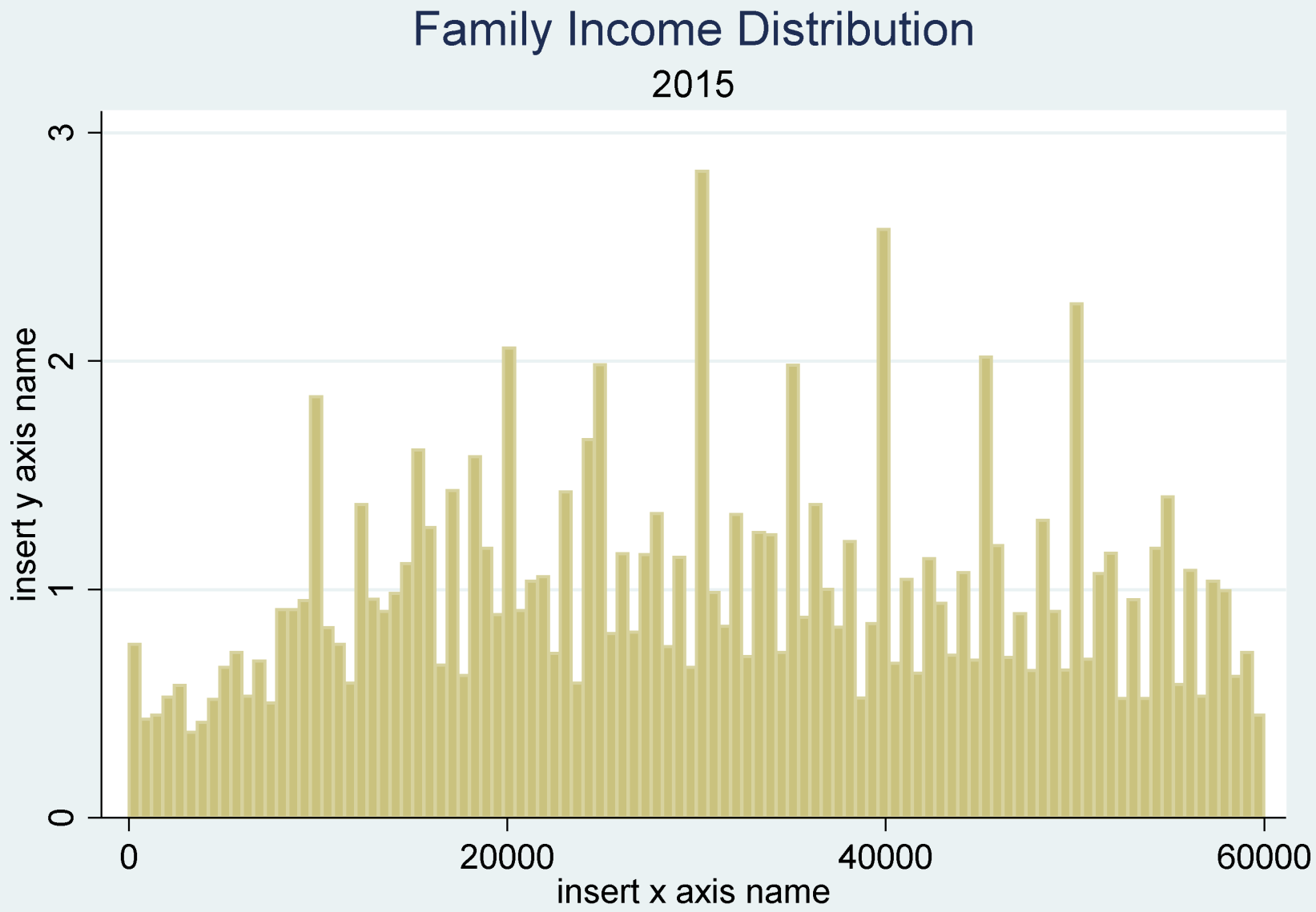
STATA Command: histogram

- Example:

```
1 hist ftotval if ftotval< 60000 & ftotval>0, bin
      (100) percent title(Family Income
      Distribution) subtitle(2015) ///
2 xtitle(insert x axis name) ytitle(insert y axis
      name)
```

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

Graph Example



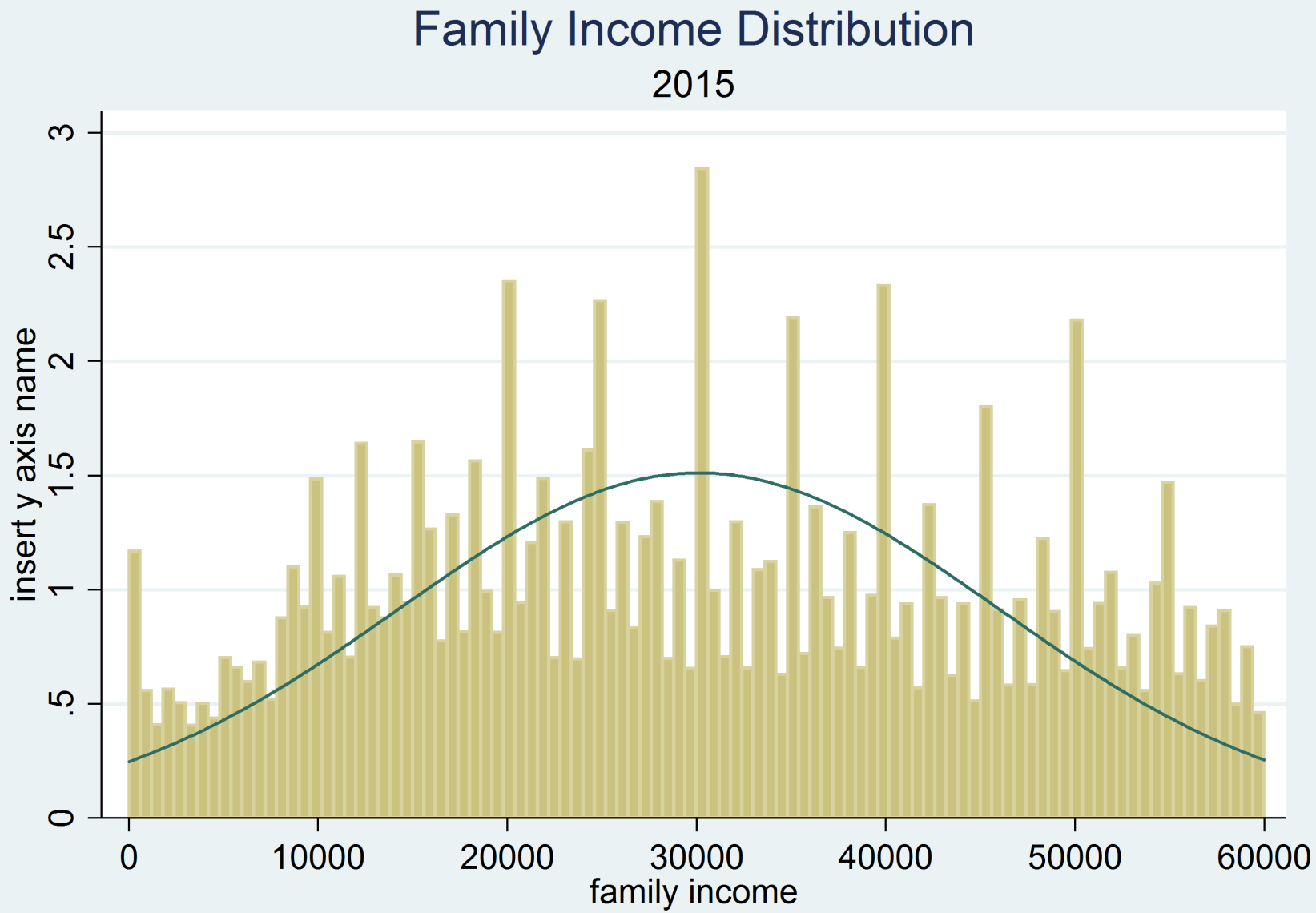
STATA Command: histogram

- Example:

```
1 hist ftotval if ftotval< 60000 & ftotval>0, bin
      (100) percent normal title(Family Income
      Distribution) subtitle(2015) ///
2 xtitle(family income) ytitle(insert y axis name)
      xlabel(0(10000)60000) ylabel(0(0.5)3)
```

- 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(star#(increment)end#)`
  - `xlabel(0(5)100)`
- This would label x-axis from 0-100 in increments of 5

Graph Example



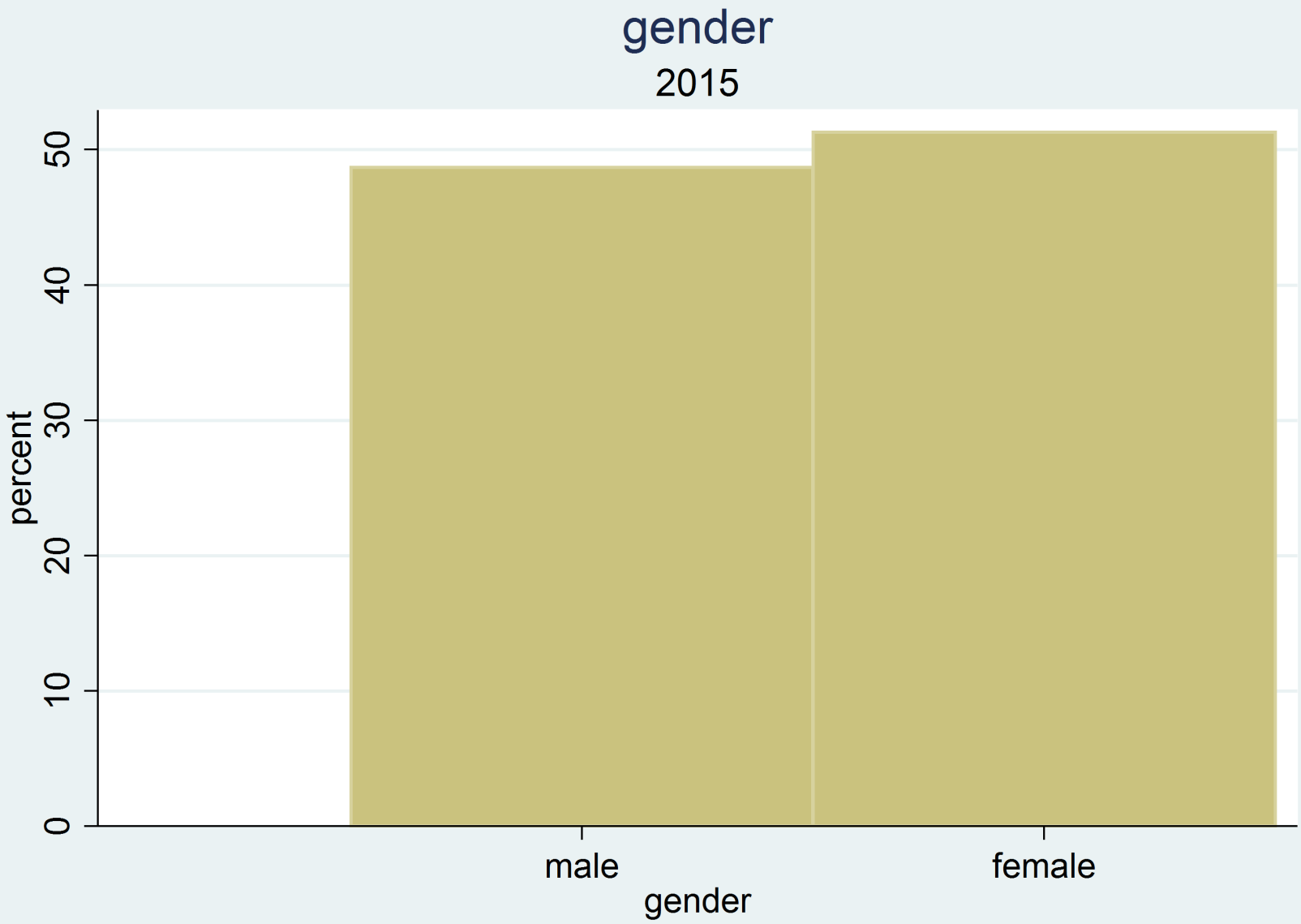
STATA Command: histogram

- Example:

```
1 hist sex, discrete percent title(gender) subtitle
    (2015) xtitle(gender) ytitle(percent) xlabel
    (1 "male" 2 "female")
```

- We can also use the `hist` command for bar graphs
  - Simple 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

Graph Example

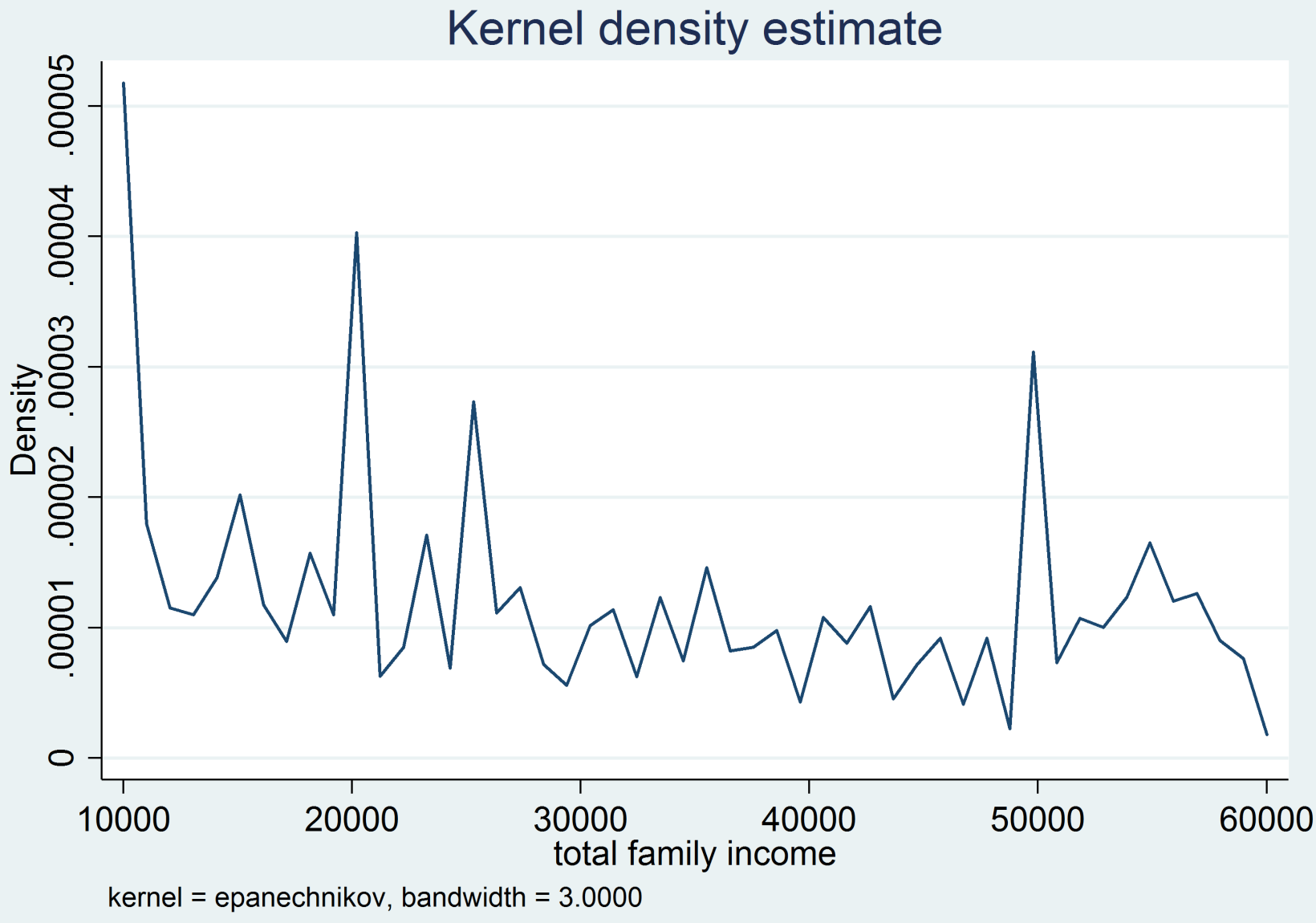


STATA Command: kdensity

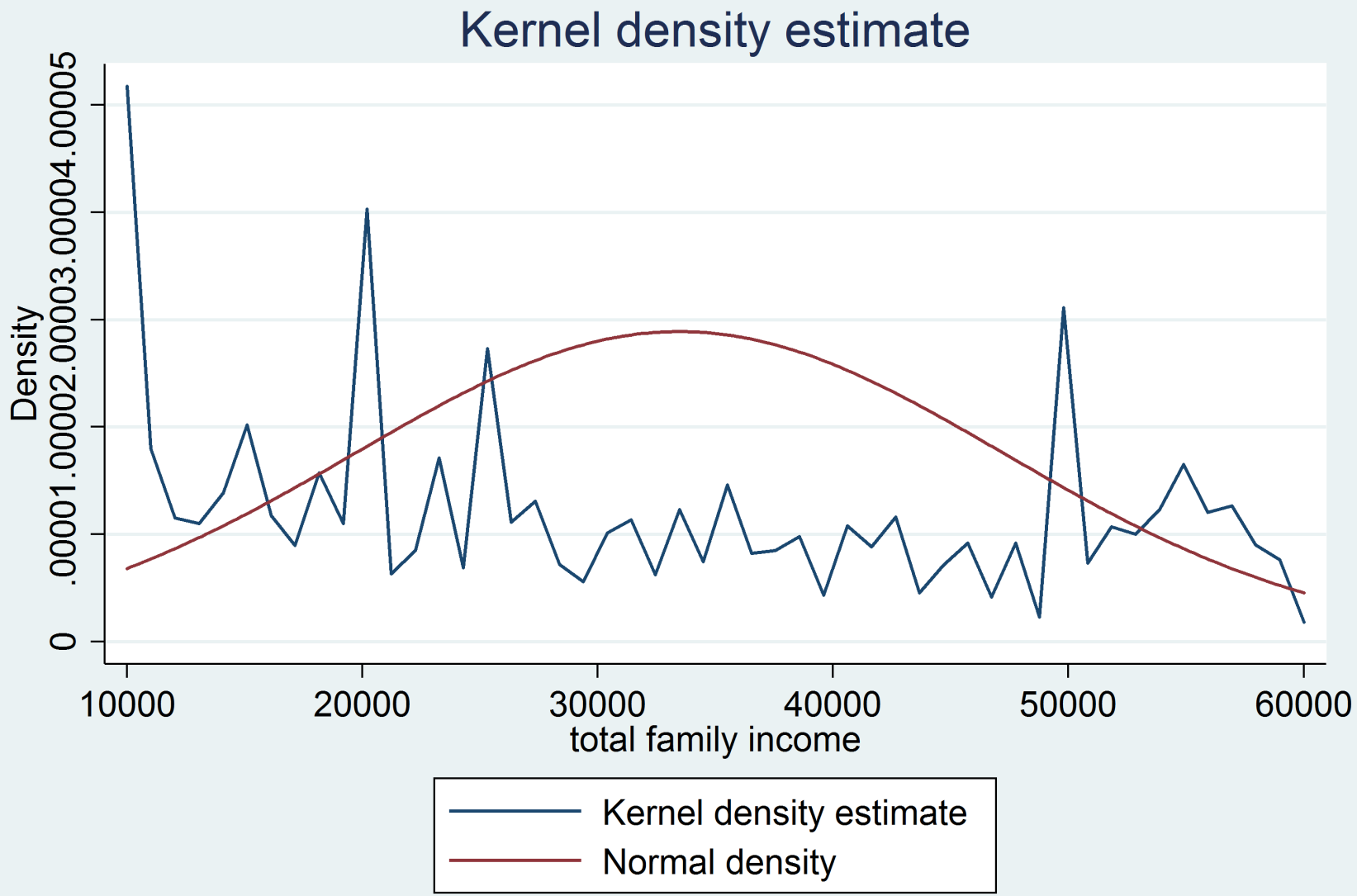
```
1 kdensity ftotval if ftotval< 60000 & ftotval>0,
      bwidth(3)
2 kdensity ftotval if ftotval< 60000 & ftotval>10000,
      bwidth(3) normal
```

- A more detailed view of the distribution of a variable may be obtained using a smooth histogram
- **kdensity**: smoothed histogram
  - option **bwidth()**: choose bandwidth (how smooth)
  - option **normal**: add normal density to the graph

Graph Example



Graph Example



graph twoway: Create X-Y Plots Showing Points or Lines

STATA Command: graph

- Stata has excellent graphic facilities, accessible through the **graph** command, see **help graph** for an overview

STATA Command: graph twoway

- The most common graphs in statistics are X-Y plots showing points or lines
- These are available in Stata through the **twoway** subcommand
- It as 42 sub-subcommands or plot types
- The most important of which are **scatter** and **line**

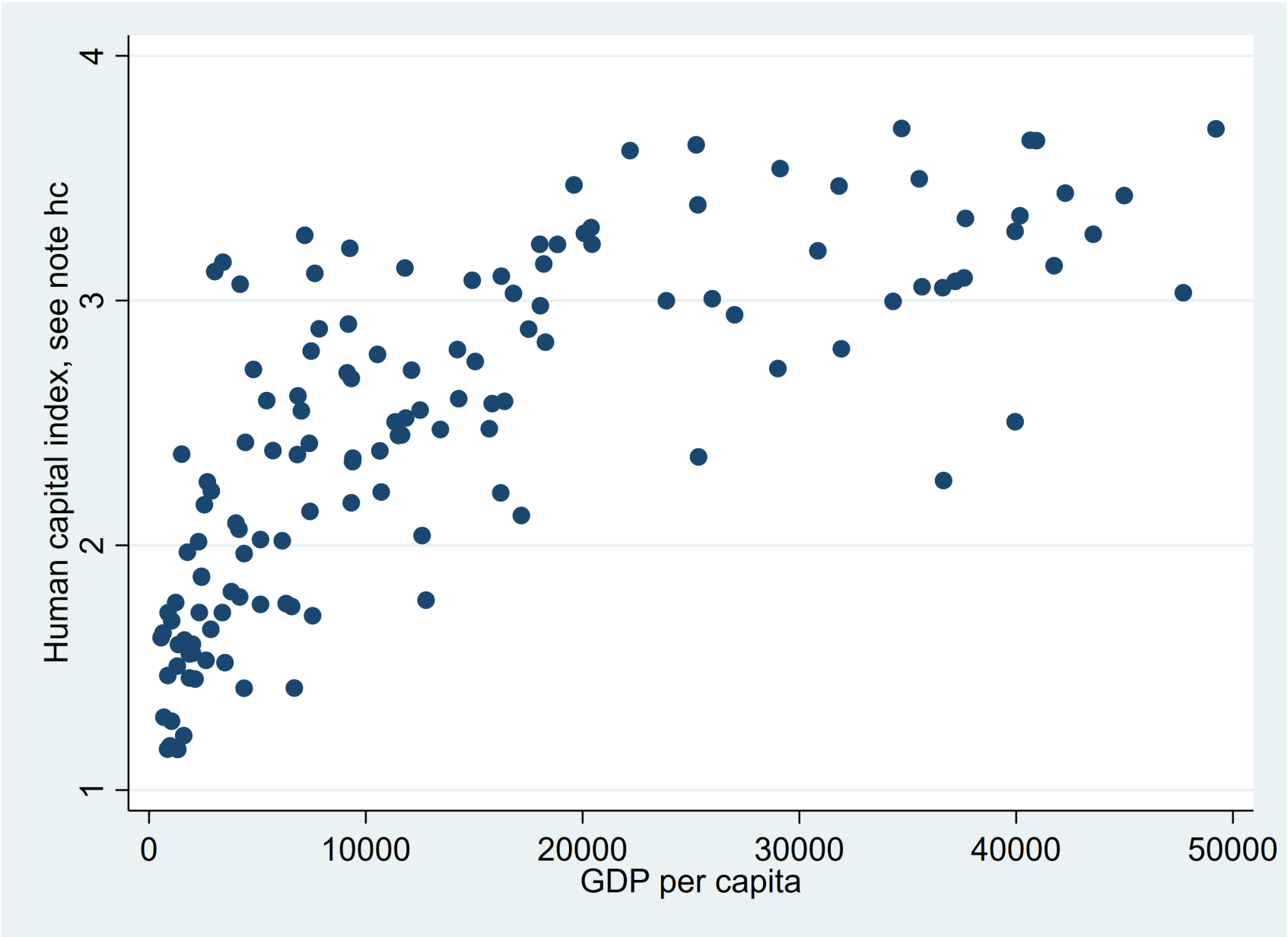
graph twoway scatter: Scatter Plots

STATA Command: graph twoway scatter

```
1 use $rawdata\pwt90.dta,replace
2
3 keep if year==2010
4 gen gdp_per = rgdpo/pop
5 label variable gdp_per "GDP per capita"
6 drop if gdp_per >50000
7
8 graph twoway scatter hc gdp_per
```

- **graph twoway scatter**: scatter plot
- Note that you specify y (change) first, then x (setting)

Graph Example

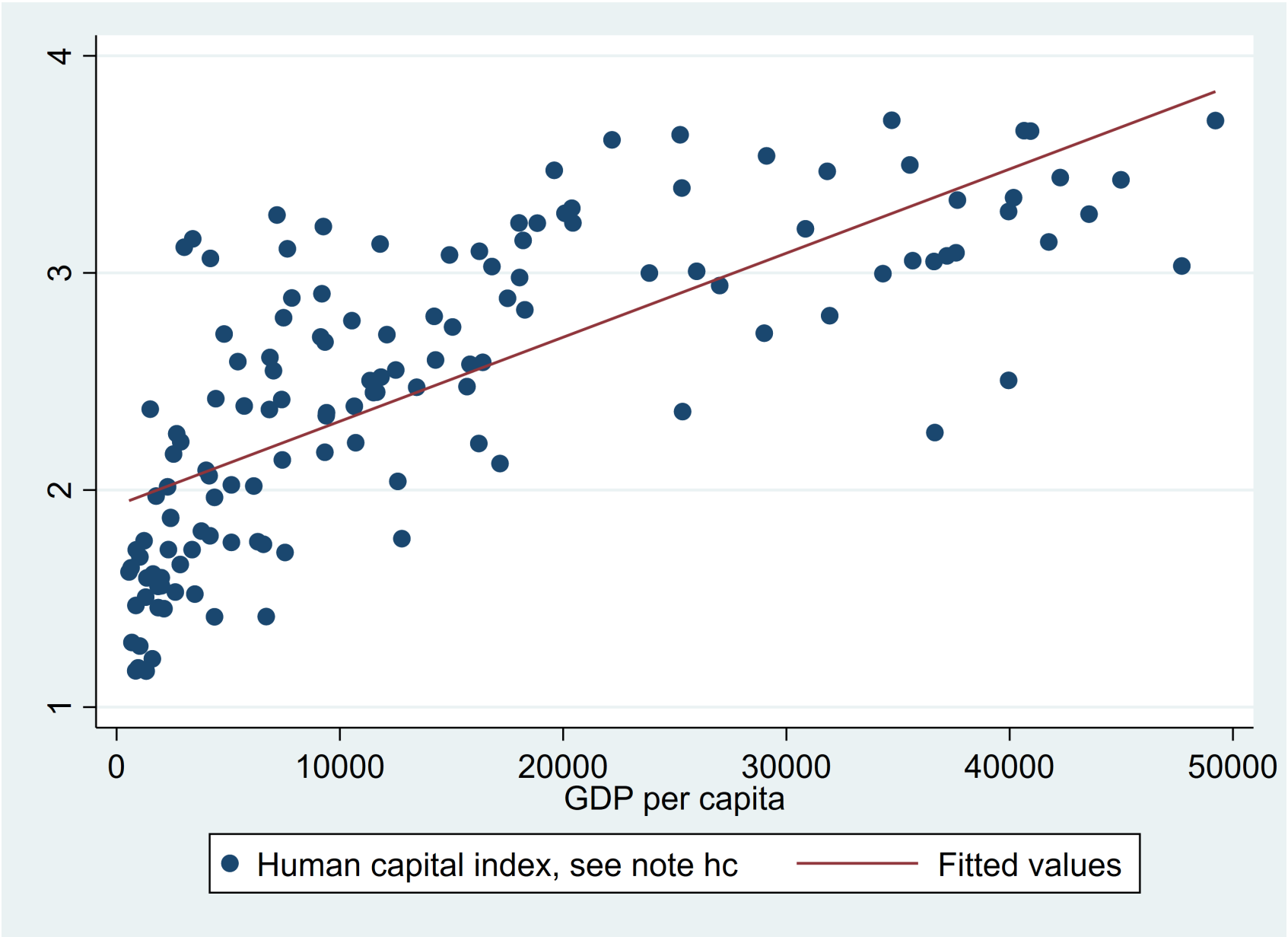


STATA Command: graph twoway scatter

```
1 graph twoway (scatter hc gdp_per) (lfit hc gdp_per)
```

- Suppose we want to show the fitted regression line as well
- Stata can do all that in one step using the **lfit** plot type
- There is also a **qfit** plot for quadratic fits

Graph Example

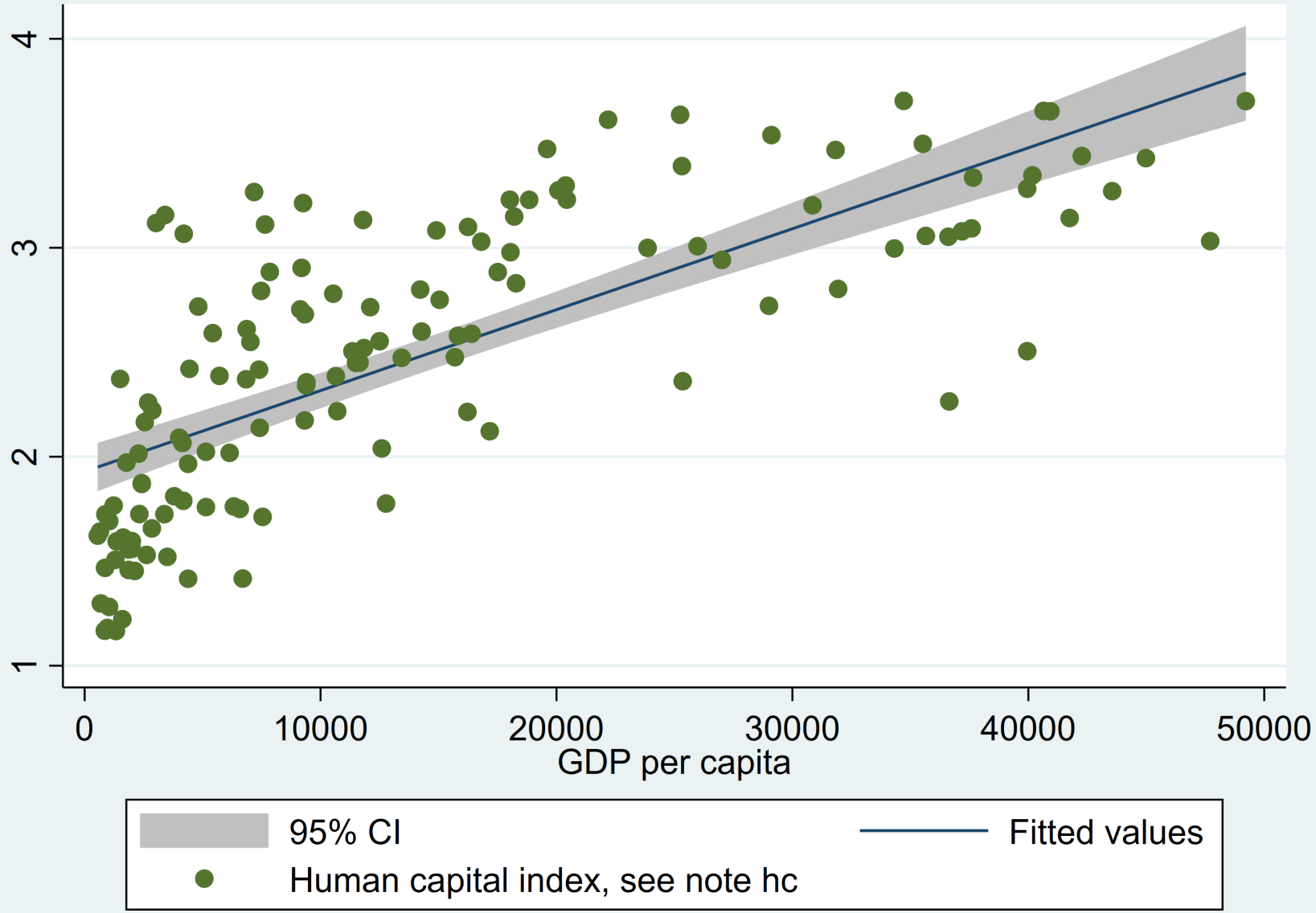


STATA Command: graph twoway scatter

```
1 graph twoway (scatter hc gdp_per) (lfitci hc gdp_per)
```

- **lfitci**: draws the confidence region as a gray ban

Graph Example

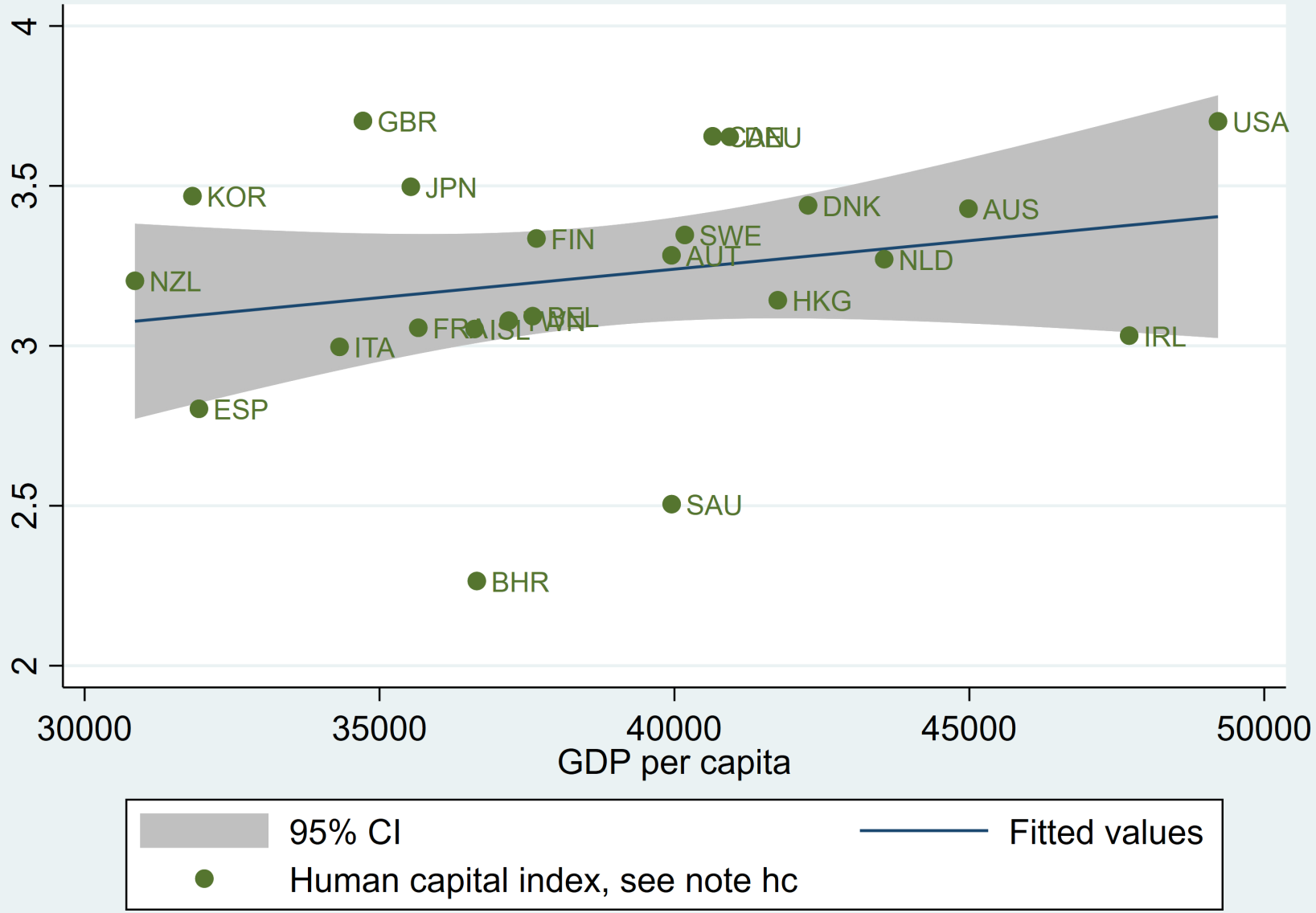


STATA Command: graph twoway scatter

```
1  graph twoway (lfitci hc gdp_per) (scatter hc gdp_per,
       mlabel(countrycode) )
```

- It is also possible to label the points with the values of a variable, using the **mlabel(varname)** option

Graph Example



graph twoway line: Line Plots

STATA Command: graph twoway line

```
1 sysuse uslifeexp, clear
2 graph twoway (line le_wmale le_bmale year , clcolor(
      blue red)) ///
3 , title("U.S. Life Expectancy") subtitle("Males") ///
4 legend(order(1 "white" 2 "black") ring(0) pos(5)) ///
5 yscale(log range(25 80))
```

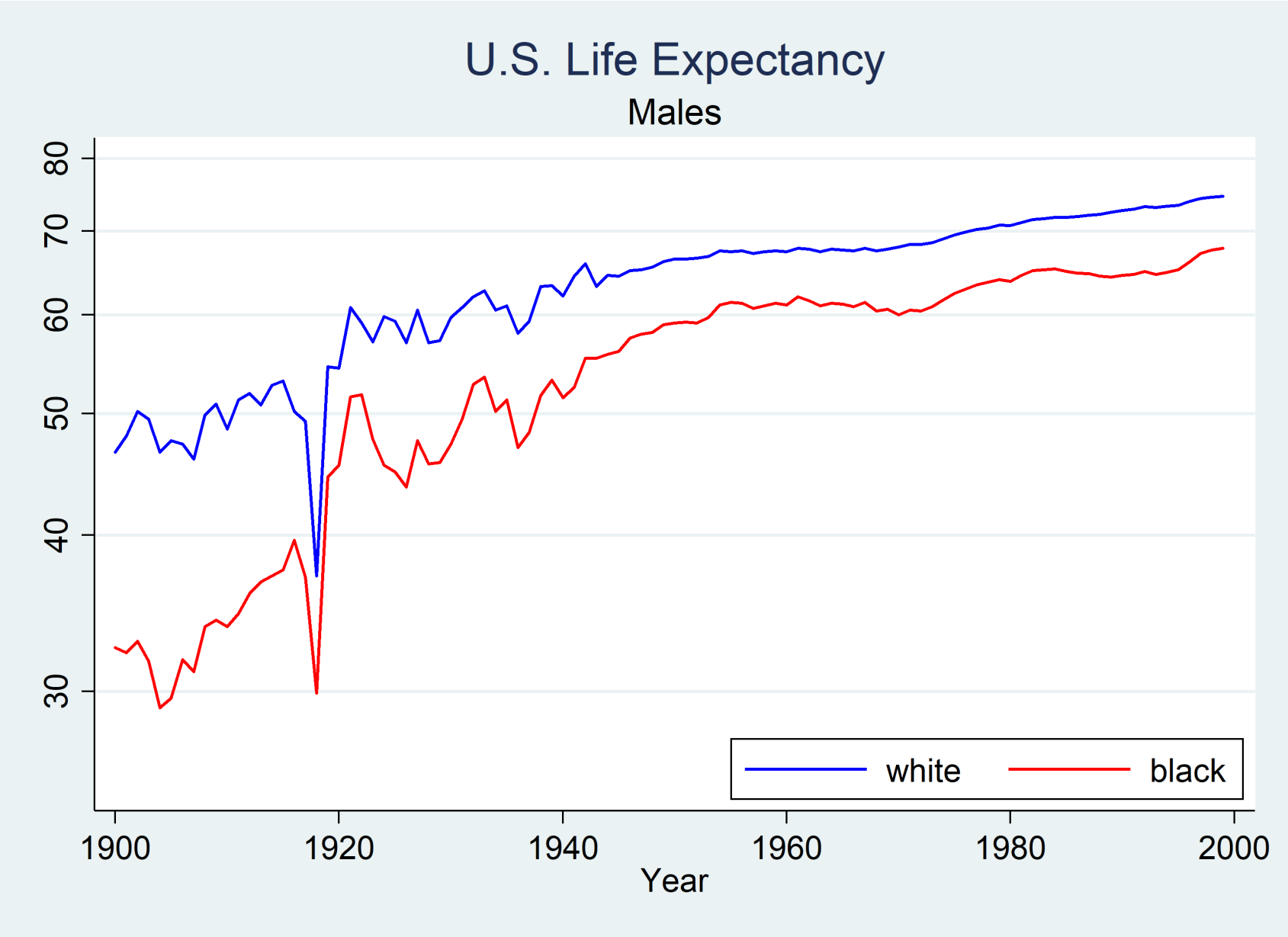
- There are options that apply to all two-way graphs, including titles, labels, and legends
- Stata graphs can have a **title()** and **subtitle()**, usually at the top
- Have a **legend()**, **note()** and **caption()**, usually at the bottom

STATA Command: graph twoway line

```
1 sysuse uslifeexp, clear
2 graph twoway (line le_wmale le_bmale year , clcolor(
      blue red)) ///
3 , title("U.S. Life Expectancy") subtitle("Males") ///
4 legend(order(1 "white" 2 "black") ring(0) pos(5)) ///
5 yscale(log range(25 80))
```

- Use **pos(5)** to place legend near the 5 o'clock position
- Use **ring(0)** to move the legend inside the plotting area
- Use the **clcolor()** option to specify color
- Use **yscale()** to choose arithmetic, log, or reversed scales

Graph Example

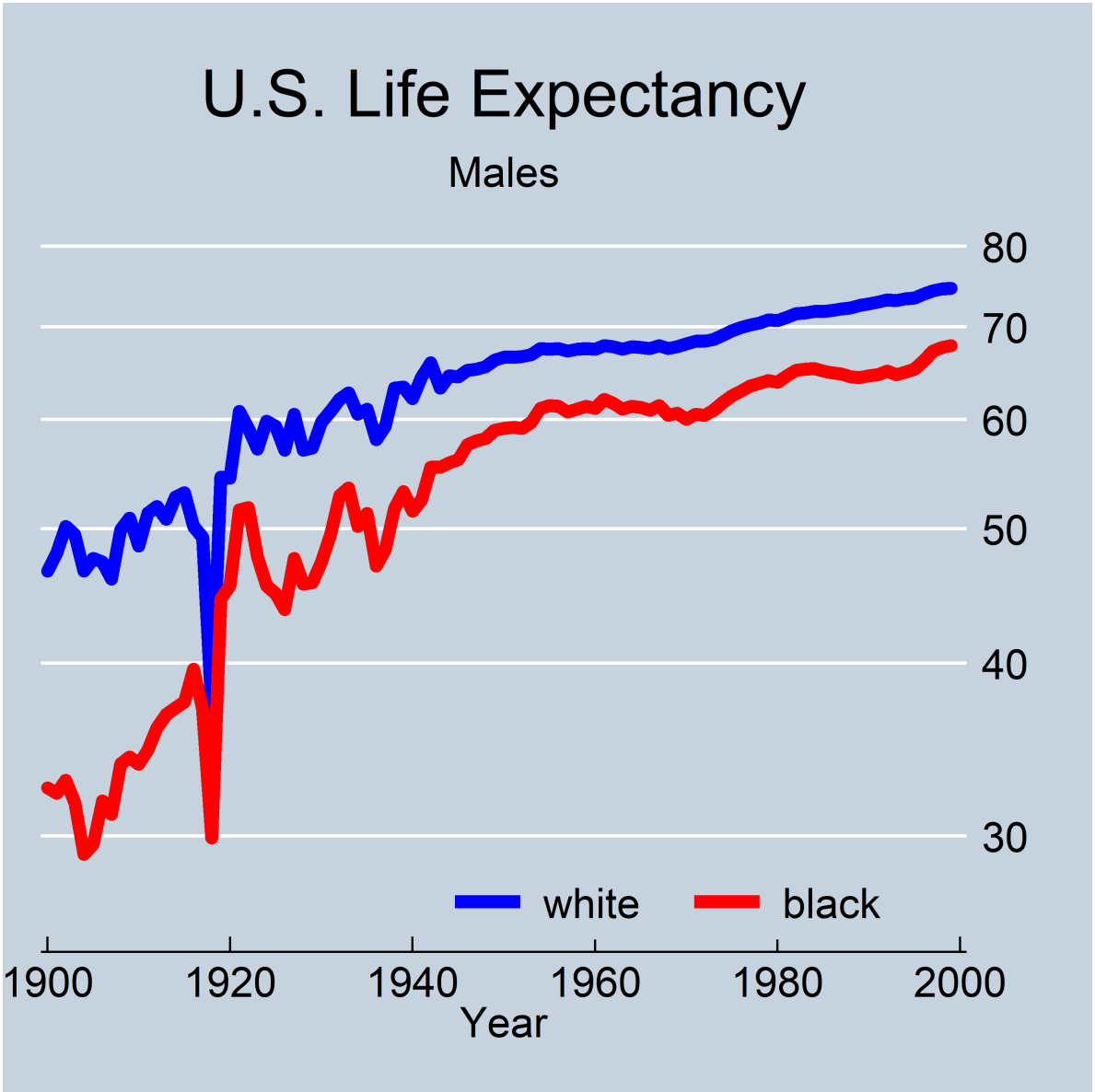


STATA Command: graph twoway line

```
1 graph display, scheme(economist)
```

- You can also redisplay the (last) graph using a different scheme with **graph display, scheme()**.
- To see a list of available schemes type **graph query, schemes**
- **economist** for the style used in The Economist

Graph Example



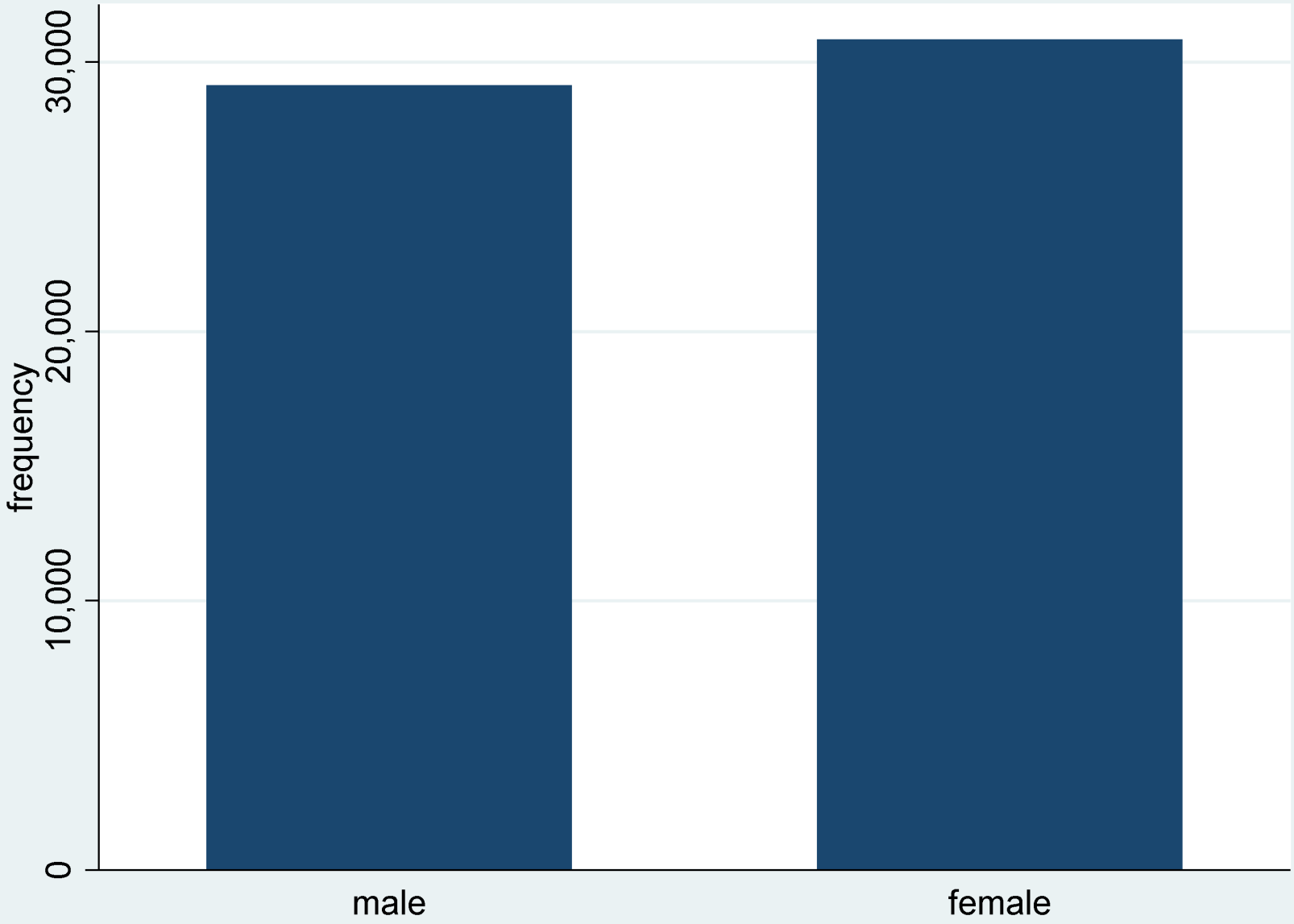
graph bar: Bar Graph

STATA Command: graph bar

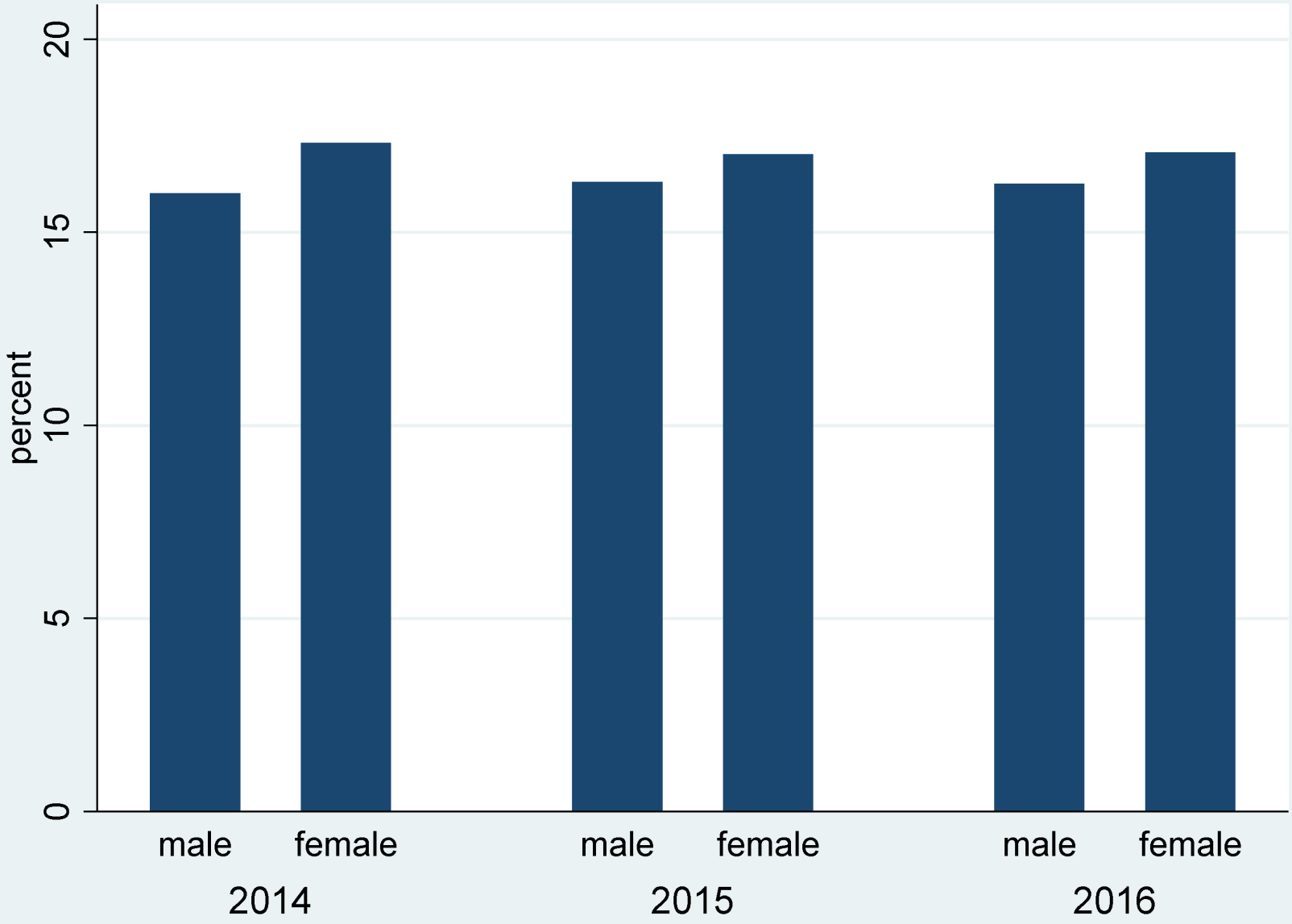
```
1 graph bar (count), over(sex)
2 graph hbar (percent), over(sex) over(year)
```

- **graph bar**: bar plot
- option **over(varname)**: draw different bars based on specific categorical variable

Graph Example



Graph Example

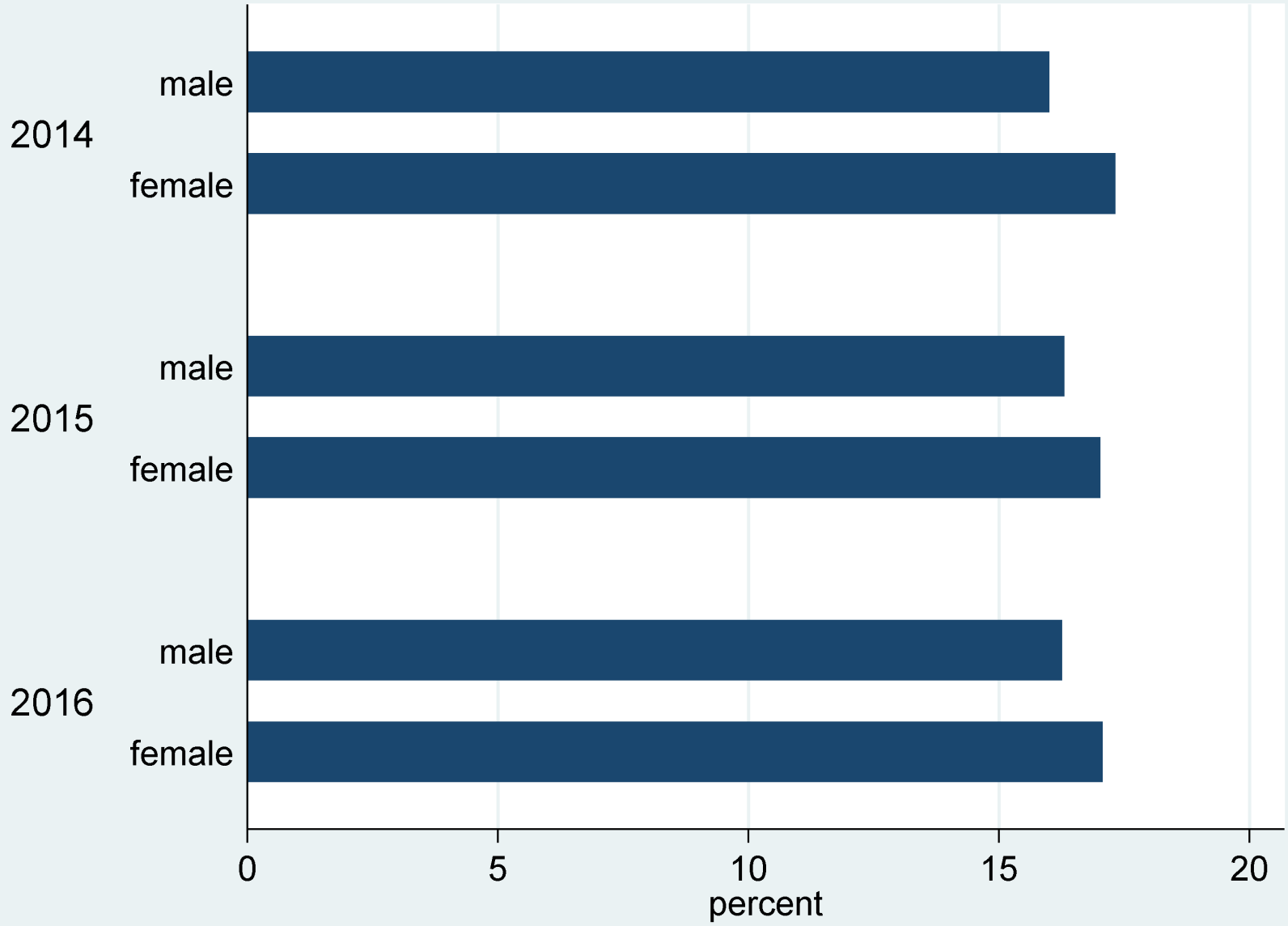


STATA Command: graph bar

```
1 graph hbar (percent), over(sex) over(year)
```

- **hbar**: horizontal bar plot
  - y is presented horizontally, and x vertically

Graph Example

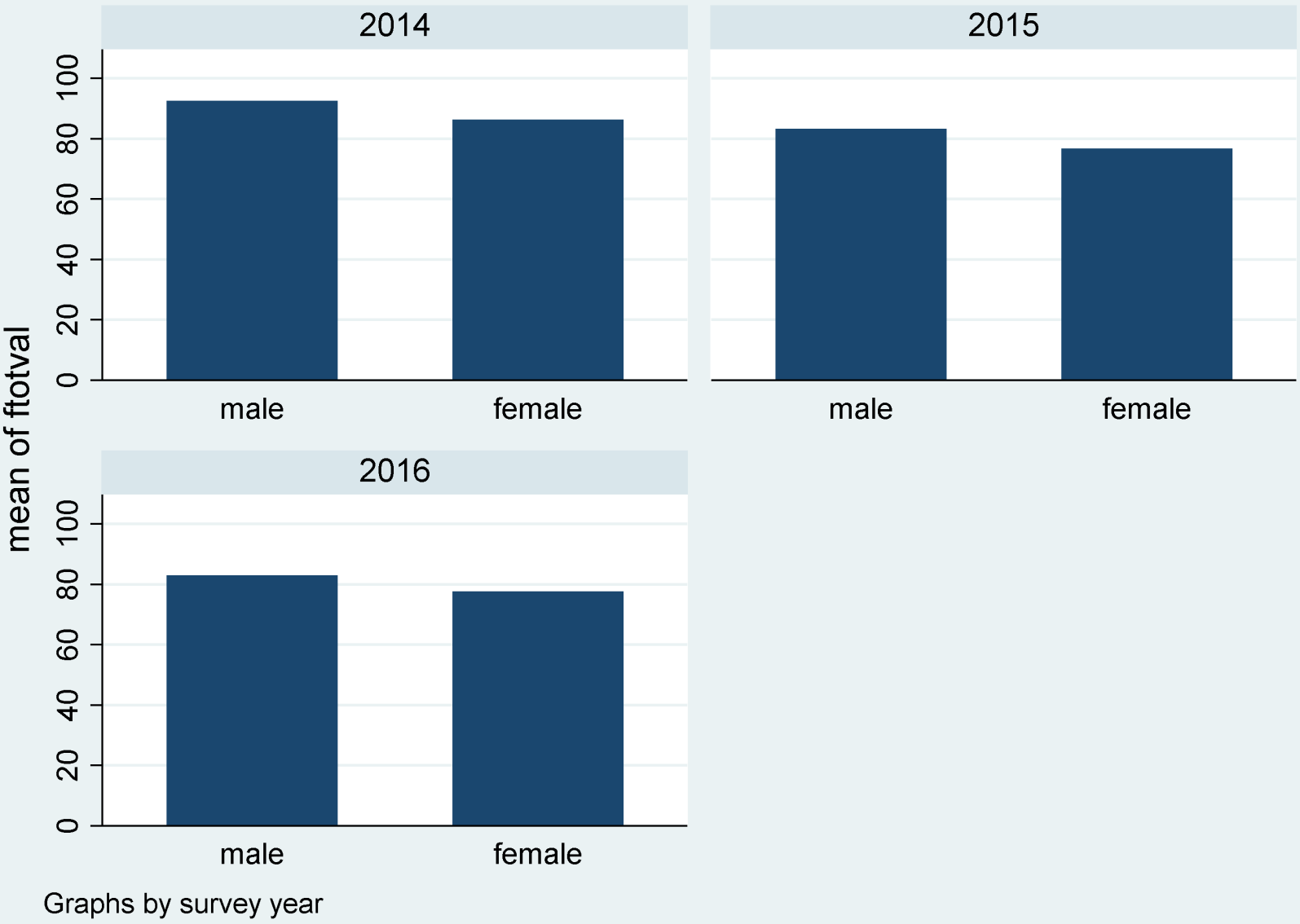


STATA Command: graph bar

```
1 graph bar (mean) ftotval, over(sex) by(year)
```

- option **by(varname)**: draw bar graph by specific categorical variable separately

Graph Example

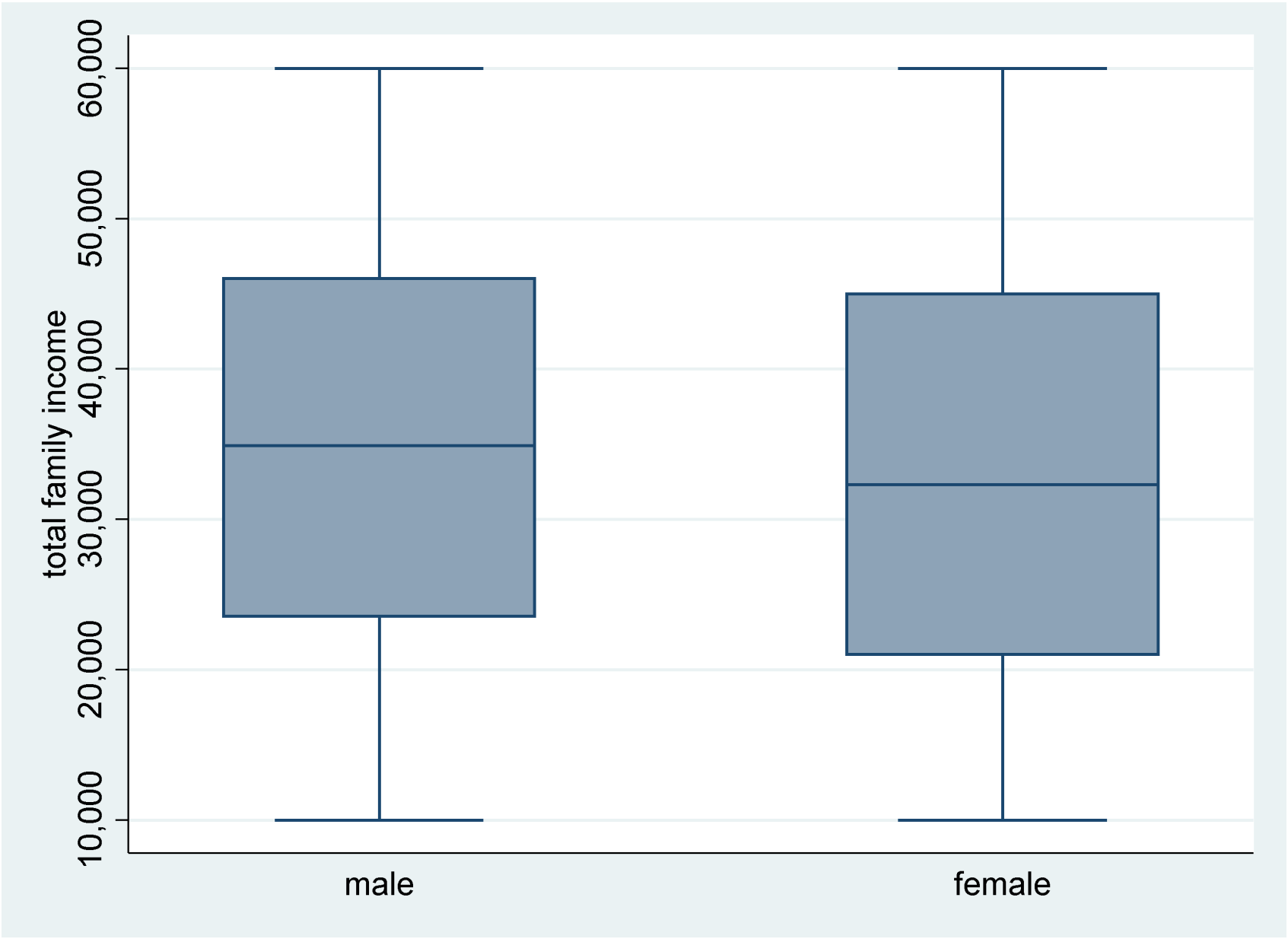


STATA Command: graph box

```
1  graph box ftotval if ftotval< 60000 & ftotval>10000,
       over(sex)
```

- **graph box**: box plot

Graph Example



Customizing Appearance and Export Your Graph

Customizing Appearance

```
graph twoway (connected  rgdp_hour_cg81  year if
    ind_code =="IND_SER", msymbol(O) mcolor(blue)) (
    connected c_paid_w_hour_cg81   year if ind_code
    =="IND_SER",msymbol(T) mcolor(red)),  legend(
    position(10) ring(0) col(1)  label(1 "real GDP
    per hour (labor productivity) ") label(2 "real
    wage")  ) xtitle(year) ytitle(accumulated growth
    rate(%),angle(vertical)) xlabel(1981(4)2014,grid)
     ylabel(0(100)400,angle(hor) labsize(medium))
    graphregion(color(white)) legend(on region(lcolor
    (white))) ytick(#10) xtick(#33)
```

- **msymbol()**: specify marker symbol
- **mcolor()**: specify marker color
- **legend()**: specify legend (position, color, label)

Customizing Appearance

```
graph twoway (connected  rgdp_hour_cg81  year if
    ind_code =="IND_SER", msymbol(O) mcolor(blue)) (
    connected c_paid_w_hour_cg81   year if ind_code
    =="IND_SER",msymbol(T) mcolor(red)),  legend(
    position(10) ring(0) col(1)  label(1 "real GDP
    per hour (labor productivity) ") label(2 "real
    wage")  ) xtitle(year) ytitle(accumulated growth
    rate(%),angle(vertical)) xlabel(1981(4)2014,grid)
     ylabel(0(100)400,angle(hor) labsize(medium))
    graphregion(color(white)) legend(on region(lcolor
    (white))) ytick(#10) xtick(#33)
```

- **xlabel()**: specify x-axis label
- **ylabel()**: specify y-axis label
- **graphregion()**: specify graphregion (color)
- **ytick()**: specify number of tick in y-axis

Export Your Graph

```
1 cd "$pic"
2 graph export f1.png,replace width(3000)
```

- **graph export**: generate your graph (.png can be used in Latex)

Graph Example