Introduction to Stata – Session 3[1]

Tarjei Havnes

[1]ESOP and Department of Economics
University of Oslo

[2]Research department
Statistics Norway

ECON 3150/4150, UiO, 2012

[1]Slides are based largely on Edwin Leuven's hard work.

Before we start

1. In your folder statacourse: auto.dta, country1.dta and country2.dta
   - http://www.uio.no/studier/emner/sv/oekonomi/ECON4150/v12/
2. Go to kiosk.uio.no (Internet Explorer!) and log on using your UIO user name
3. Navigate to Analyse (english: Analysis)
4. Open StataIC 11

Outline

1. Data handling and manipulation
   - Collapse
   - Logging your results
   - Reshaping your data set
   - Appending
   - Merging
   - Reading data in other formats
2. Drawing graphs
   - Basic graphs
   - Customizing your graph
   - Overlaying graphs
   - Saving your graph

Collapse

It is easy to convert the dataset in memory into a dataset of summary statistics

- calculate input for tables or graphs
- create dataset at higher level of aggregation (e.g. from individual to municipality level dataset)

The syntax is

- `collapse [(stat)] [targetvar=]varname ... [if], by(varlist)`
  where *stat* defaults to `mean`, but can be `count`, `sum`, `p34`, `var`, `min`, `max` ...

Tables of summary statistics

```
. u auto
(1978 Automobile Data)
. preserve
. collapse price (p50) medprice = price, by(foreign)
. l, noobs
  +--------------------------------+
  |  foreign     price   medprice |
  |--------------------------------|
  | Domestic   6,072.4    4,782.5 |
  |  Foreign   6,384.7      5,759 |
  +--------------------------------+
. restore

. tab foreign, s(price)
           |           Summary of Price
   Car type |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   Domestic |   6,072.423   3,067.472        5252
    Foreign |   6,384.682    2,562.21        2222
------------+------------------------------------
      Total |   6,165.257   2,929.695        7474

. table foreign, c(m price p50 price)
-------------------------------------
 Car type | mean(price)   med(price)
----------+--------------------------
 Domestic |     6,072.4      4,782.5
  Foreign |     6,384.7        5,759
-------------------------------------
```

Saving your results (logging)

You can save your results to file using -log-

- `log using anauto`

Stata will throw an error when

1. the log file exists
   solution: `log using anauto, replace`
2. the log file is already open
   solution: `close log`
3. when there is no open log
   final solution: `capture close log`

Plain text log file:

- `log using anauto, replace text`

Advice: Always use the same name as the do file

A typical do file (anreg.do)

```
capture log close
log using anreg, replace
set more off

// do stuff here

log close
// always leave one empty line at the end
```

Reshape

(wide form)

id	sex	inc80	inc81	inc82
1	0	5000	5500	6000
2	1	2000	2200	3300
3	0	3000	2000	1000

(long form)

id	year	sex	inc
1	80	0	5000
1	81	0	5500
1	82	0	6000
2	80	1	2000
2	81	1	2200
2	82	1	3300
3	80	0	3000
3	81	0	2000
3	82	0	1000

You can move from wide to long

- `reshape long inc, i(id sex) j(year)`

or from long to wide

- `reshape wide inc, i(id sex) j(year)`

(try it with country2.dta)

Combining datasets vertically (append)

```
. use a
. append using b
```

(a.dta)

x	y
1	1.2
2	2.3
3	0.5

(b.dta)

x	z
6	0.03
12	0.01

(b appended to a)

x	y	z
1	1.2	.
2	2.3	.
3	0.5	.
6	.	0.03
12	.	0.01

Combining datasets horizontally (merge)

```
. use c
. sort id
. merge id using d
```

(c.dta)

id	y
1	1.2
2	2.3
3	0.5

(d.dta)

id	x
1	3.5
2	1.0
6	0.1

(d merged to c)

id	y	x	_merge
1	1.2	3.5	3
2	2.3	1.0	3
3	0.5	.	1
6	.	0.1	2

_merge==1 observation in master only
_merge==2 observation in using only
_merge==3 observation in both master and using

Merge requires both datasets to be sorted on the merge vars

Reading non Stata data

Data does not always come in Stata format

Stata can

- use (and save) datasets in FDA (SAS XPORT) format
  `fdause` (`fdasave`)
- read ASCII data
  - spreadsheet type data files with separators (commas, tabs,...)
    `insheet`
  - text files where data is in fixed colums
    `infix`

Note that Stata can also import data files directly from online sources, without having to first download them.

Documenting - Notes

You can attach notes to the dataset and/or variables

```
. notes _dta : Recovered from Stata distribution

. notes

_dta:
  1.  from Consumer Reports with permission
  2.  Recovered from Stata distribution

. notes rep78 : Mari, why are there missing values?! (Tarjei)

. notes

_dta:
  1.  from Consumer Reports with permission
  2.  Recovered from Stata distribution

rep78:
  1.  Mari, why are there missing values?! (Tarjei)

. notes drop rep78 in 1
  (1 note dropped)
```

Drawing graphs

1. Basic graphs
2. Customizing your graph
3. Overlaying graphs
4. Saving your graph

Basic graphs

The most common graphs are

- scatter plots
- line plots
- histograms

Twoway graphs

Most graphs are twoway graphs

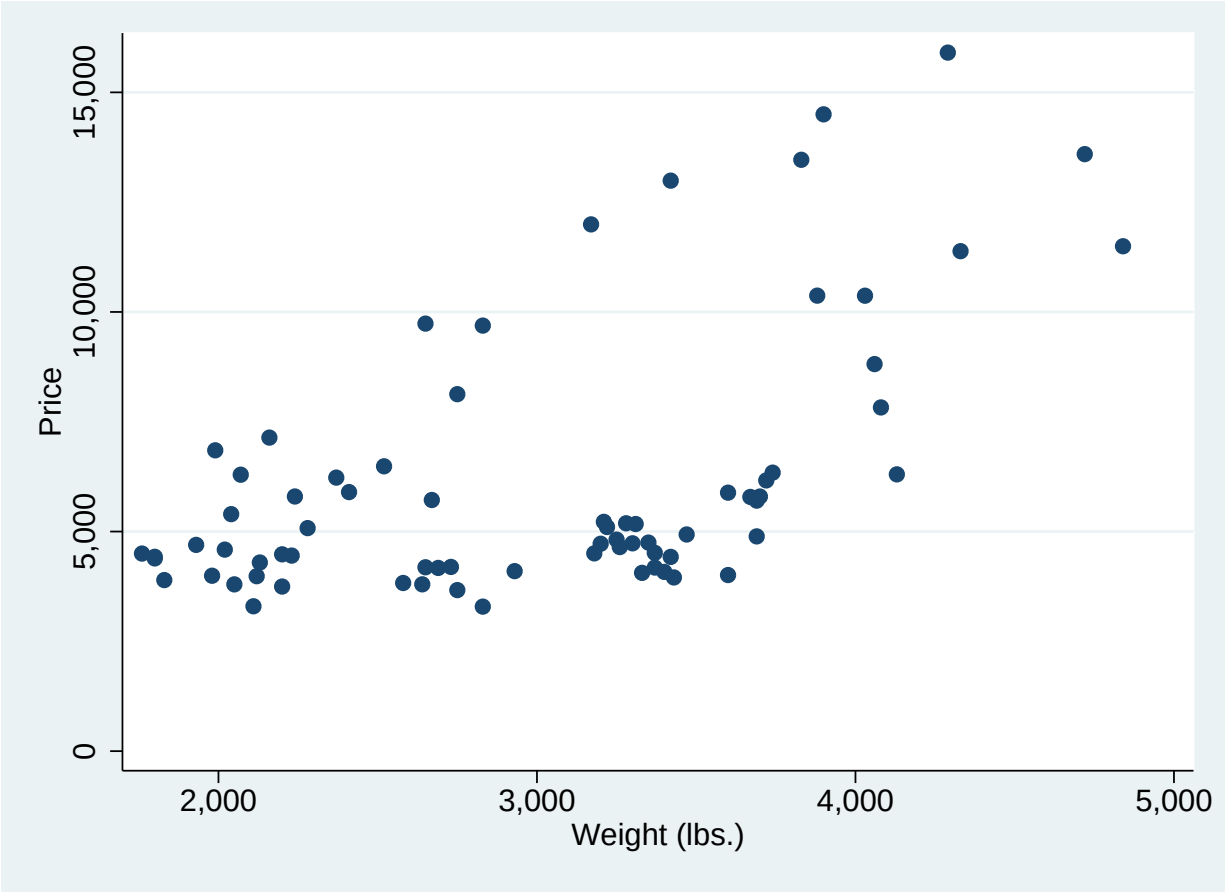
```
twoway plottype varlist [if] [in] [, twoway_options]
```

there are many plottypes (-help twoway-):

plottype	Description
scatter	scatterplot
line	line plot
connected	connected-line plot
bar	bar plot
rarea	range plot with area shading

Scatter plots

```
twoway scatter price weight
```

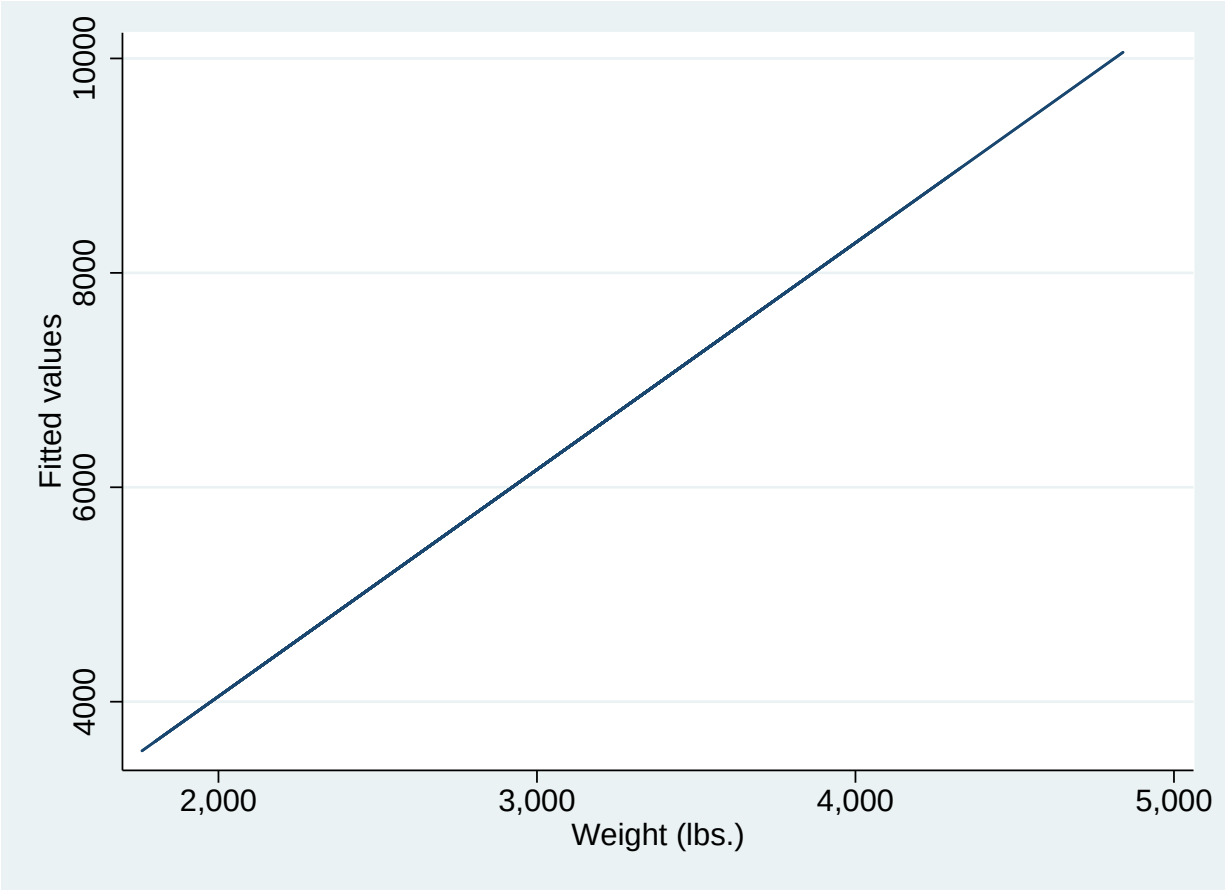


Scatter plots

```
. g price1 = price if foreign==1
. g price0 = price if foreign==0
. twoway scatter price? weight
```

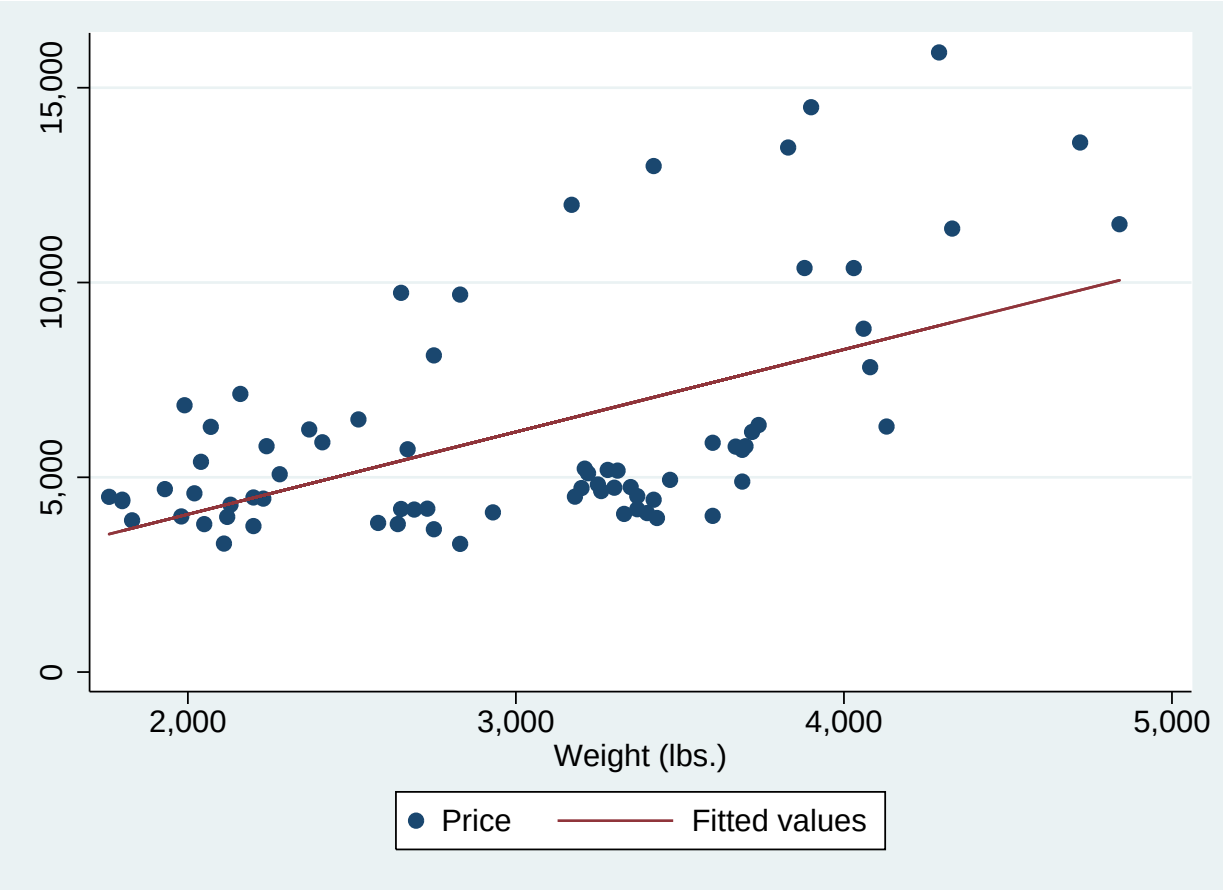
Line plots

```
reg price weight
predict pprice
twoway line pprice weight
```



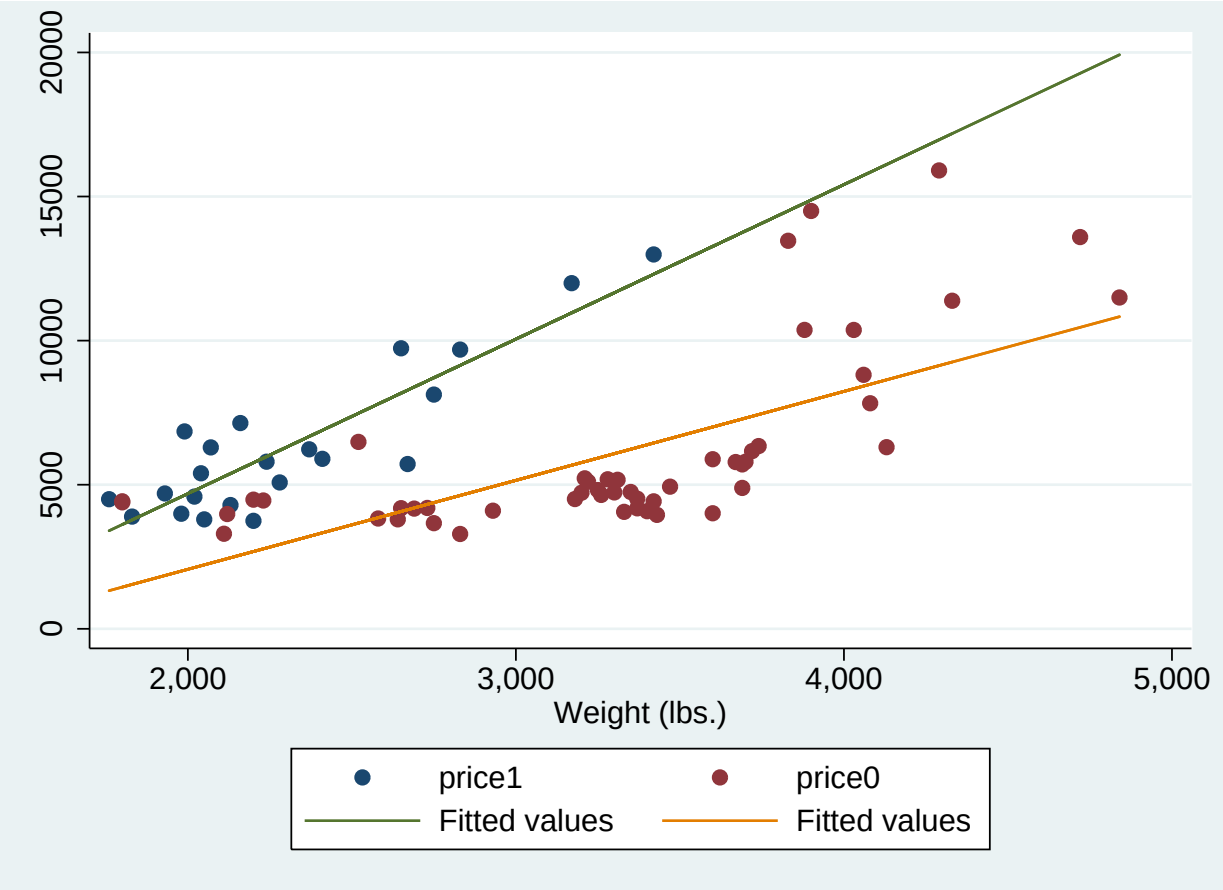
Combining plots

```
twoway (scatter price weight) || (line pprice weight)
```



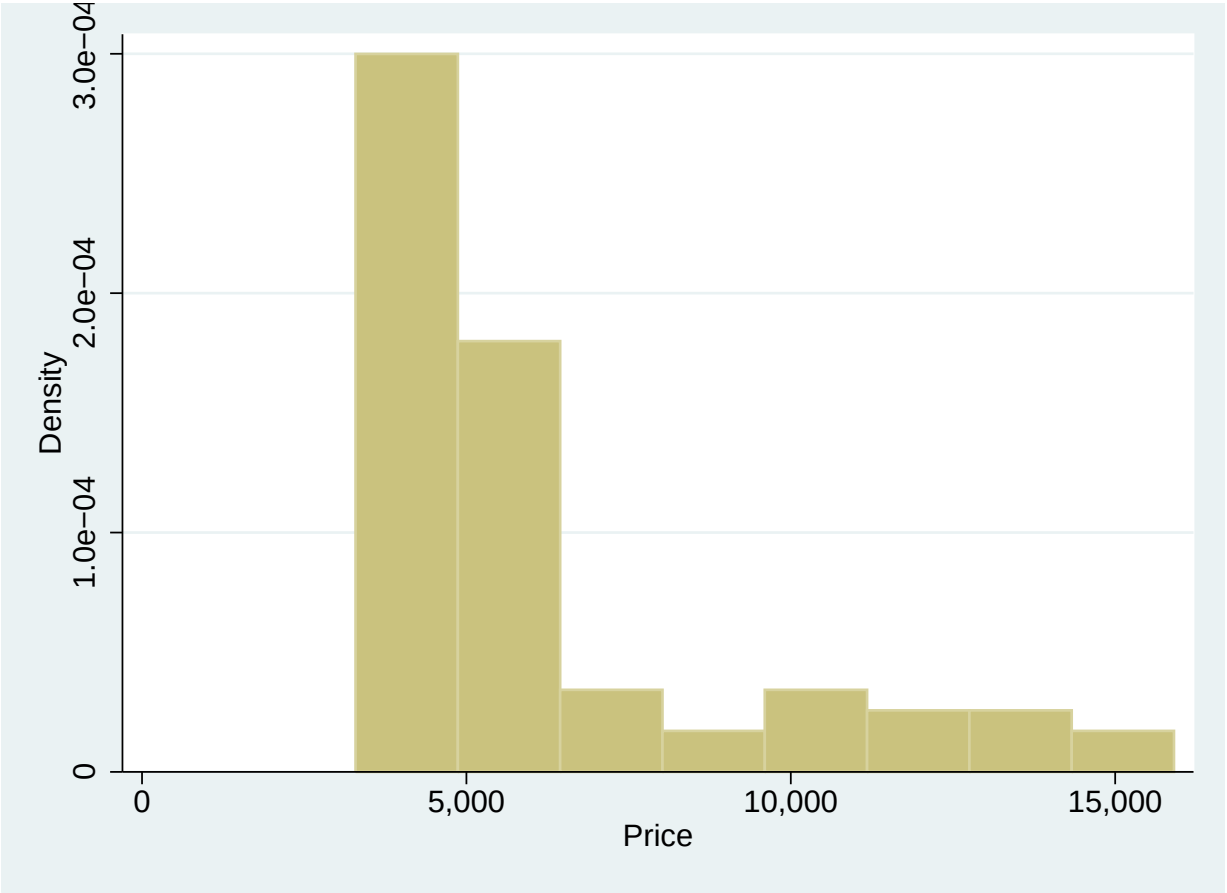
Combining plots

```
twoway (scatter price? weight) || (line pprice? weight)
```



Histograms

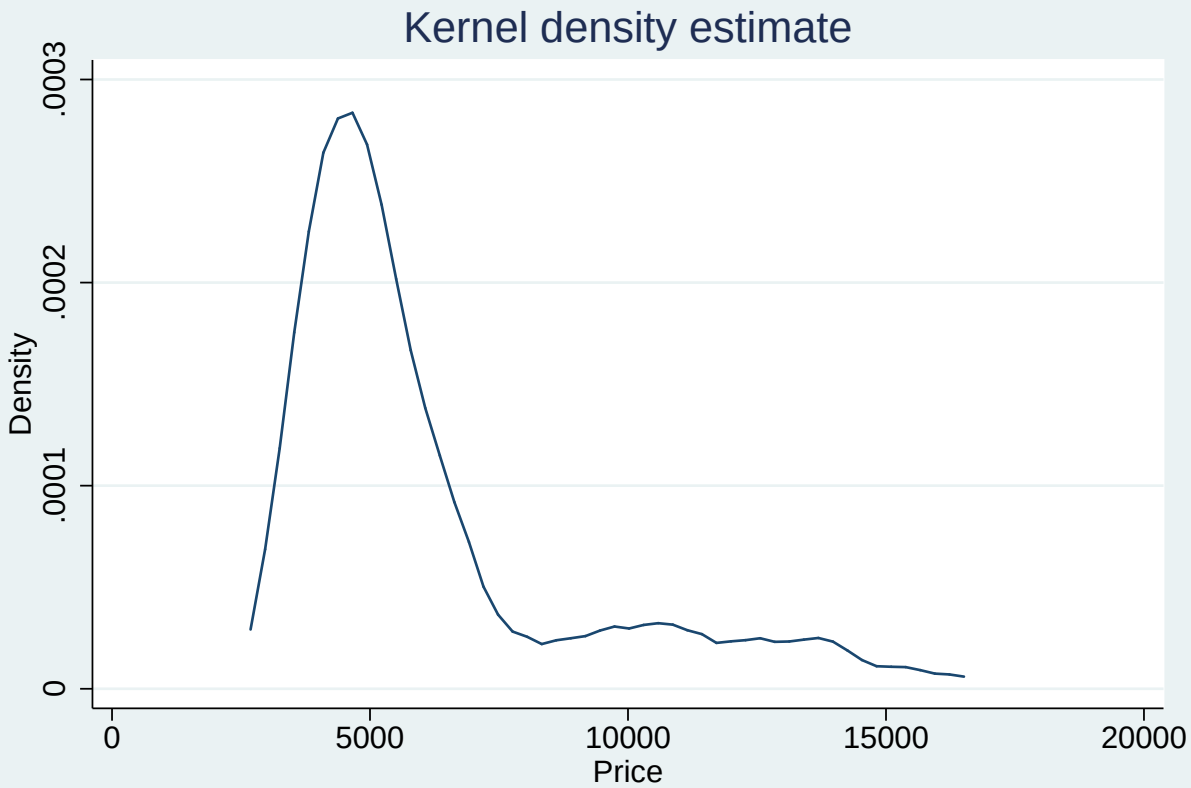
```
hist price
```



tweak the nr of bins with option -bin()-, or the width of the bins with -width()-

Kernel density

```
kdensity price
```



Customizing your graph

There are three ways of customizing the look of your graphs

1. schemes
2. options
3. graph editor

Schemes define an overall look of a graphs, to see what schemes are available

```
graph query, schemes
```

I use -s1mono- as point of departure

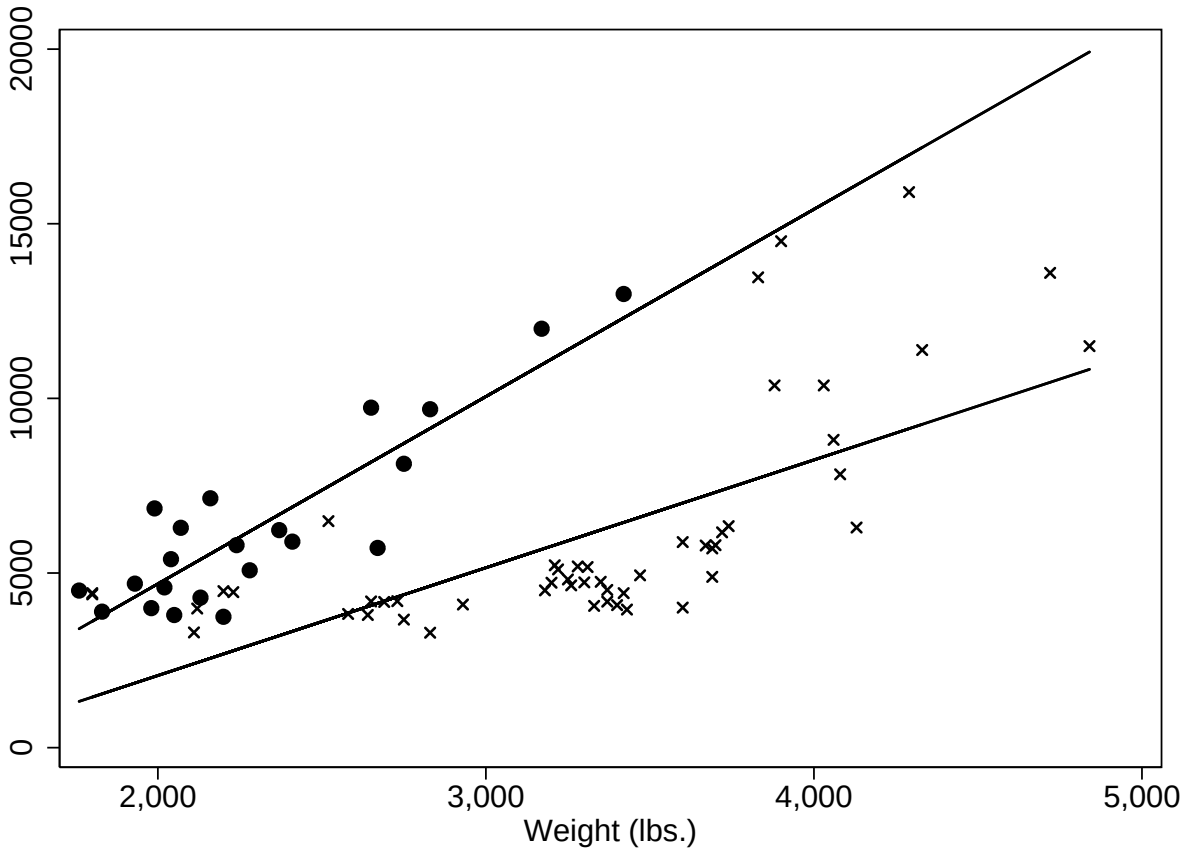
```
scatter price weight, scheme(s1mono)
```

or

```
set scheme s1mono, perm
```

Customizing your graph

```
twoway (scatter price? weight, ///
            msym(O X) mcol(black ..))  ///
    || (line pprice? weight, ///
            lpat(. .) lcol(black ..)) , legend(off)
```



Using the graph editor

The simplest way to fix how your graph appears is to use the graph editor.

1. draw a (simple) version of your graph, including all the plots you want
2. open the graph editor and play around till you figure out how you want it to appear
3. repeat 1, and then record the steps you want from 2 using Tools/Recorder/Begin
4. stop recording, and save to a file, e.g. myfigtype1.grec

Your next graph can then use the same layout by invoking the option play(myfigtype1.grec)

A note about graph size

Try the following and compare the graphs and the size of the graphs on disk

```
use auto
scatter price weight
gr export pricescatter1.eps

use largeauto
scatter price weight
gr export pricescatter2.eps

dir pricescatter*
```

How can you avoid drawing the same point over and over again?

Using collapse to make plot data

You might want to -collapse- your data to

- plot aggregate statistics
- reduce the size of your graph

this may arise if you have micro data (repeated cross-sections, or a panel), and you want to show a trend over time

use -collapse- to calculate means and then plot

```
preserve
collapse yvars, by(xvar)
twoway line yvars xvar
restore
```

Saving your graph

You can save your graph to disk using

```
graph export filename
```

The extension determines the format, e.g.

```
graph export hist.eps
```

if the file exists, use option -replace-

Note: Vector based formats (ps, eps, pdf (MAC, Win in Stata 12), wmf/emf (Win)) give the best quality output. Otherwise use .png

What you have learned...

We have only touched the tip of the iceberg, but you should now know how to

- make basic plots
- overlay twoway plots
- use schemes and basic options
- save your plot
- pay attention to the size of your plots

Don't forget to use the menus!