EPIB 613: Graphing Lecture   Nov 22, 2012

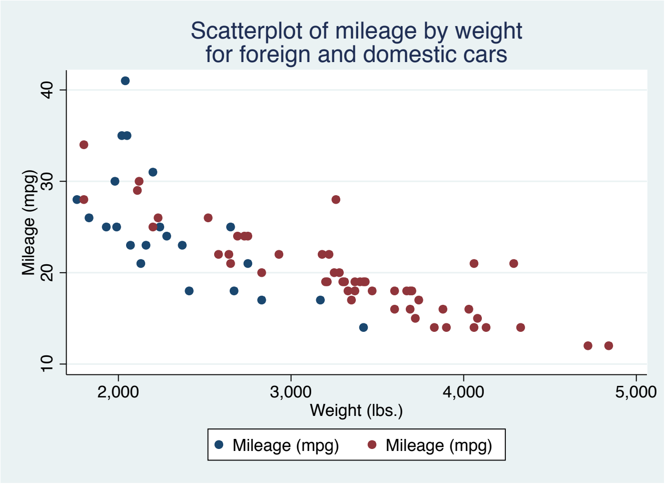
```
sysuse auto
```

```
*A simple scatterplot of MPG by weight
twoway (scatter mpg weight)
```

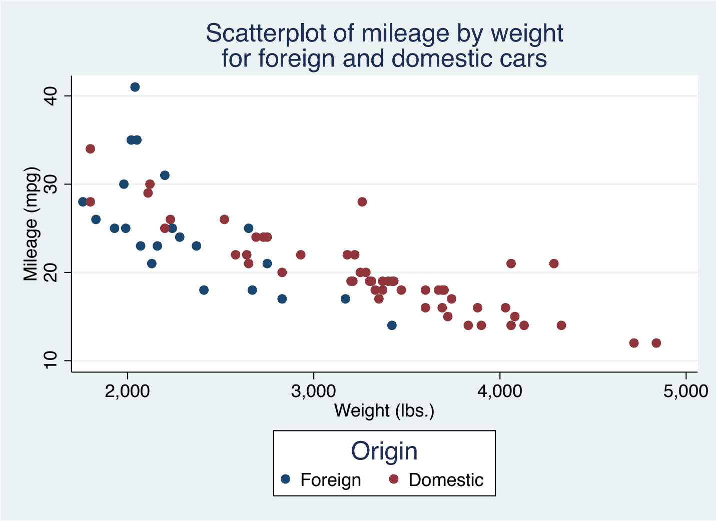
```
*Add a title
twoway (scatter mpg weight), title("Scatterplot of mileage by weight")
```

```
*Restrict to only foreign cars
twoway (scatter mpg weight if foreign==1), title("Scatterplot of mileage by weight for
foreign cars")
```

```
*Put foreign and domestic on the same graph
twoway (scatter mpg weight if foreign==1)(scatter mpg weight if foreign==0),
title("Scatterplot of mileage by weight” “for foreign and domestic cars")
```

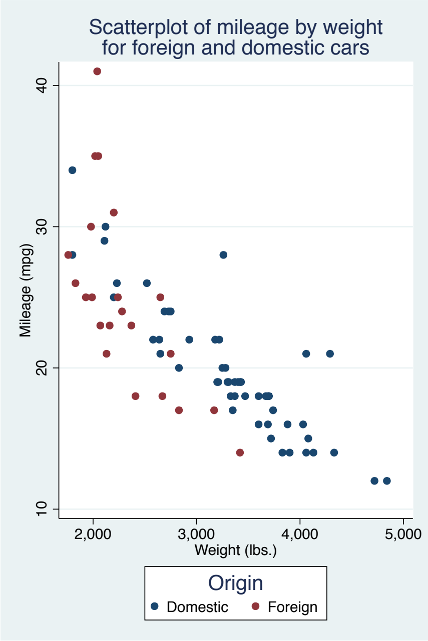


```
*Format the legend so it shows which dots are foreign and which are domestic
twoway (scatter mpg weight if foreign==1)(scatter mpg weight if foreign==0),
title("Scatterplot of mileage by weight for foreign and domestic cars") scale(0.8)
legend(title("Origin") order(1 "Foreign" 2 "Domestic"))
```



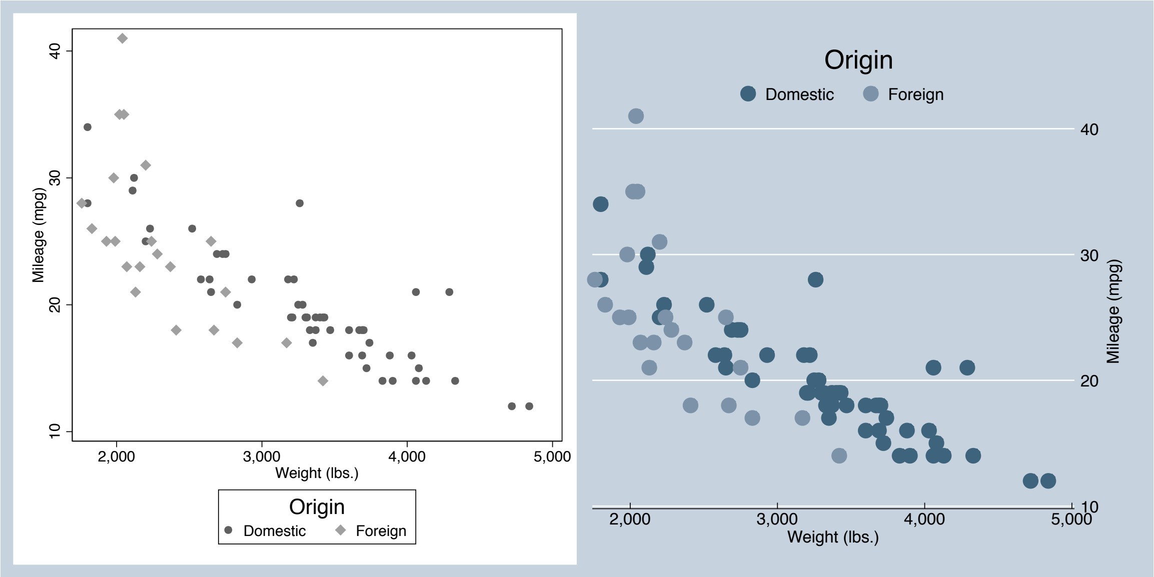
***Change the size and shape of the graph**

```
twoway (scatter mpg weight if foreign==0)(scatter mpg weight if foreign==1), ysize(3)
xsize(2) legend(title("Origin") order(1 "Domestic" 2 "Foreign"))
```



****Change the appearance of the graph**

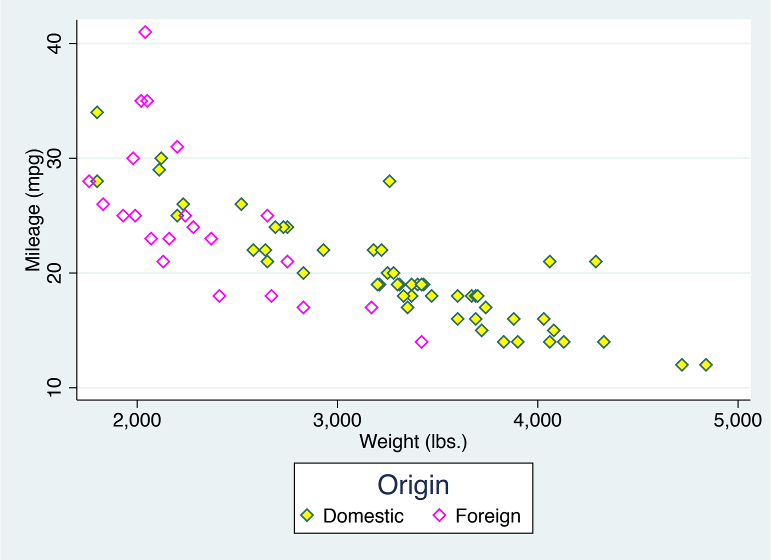
```
graph query, schemes

set scheme s1mono
twoway (scatter mpg weight if foreign==0) (scatter mpg weight if foreign==1),
legend(title("Origin") order(1 "Domestic" 2 "Foreign"))

set scheme economist
twoway (scatter mpg weight if foreign==0)(scatter mpg weight if foreign==1) ,
legend(title("Origin") order(1 "Domestic" 2 "Foreign"))
```



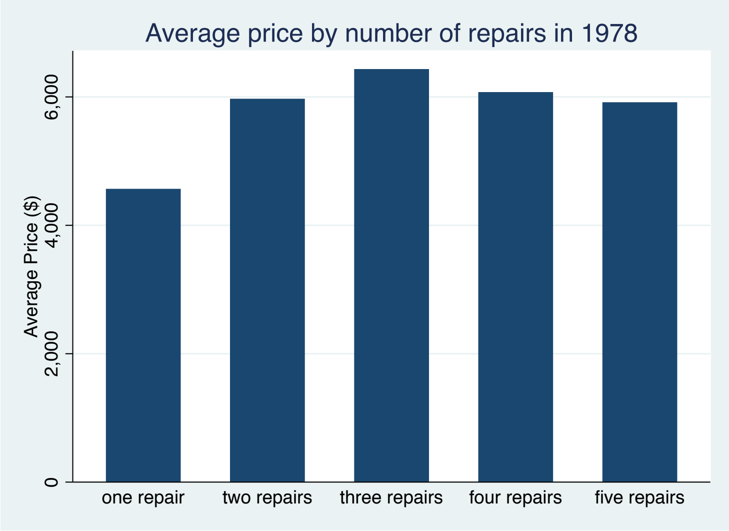
****set scheme back to normal**

```
set scheme s2color
```

```
** Can also change the appearance of symbols
palette symbolpalette
twoway (scatter mpg weight if foreign==0, msymbol(D)) (scatter mpg weight if foreign==1,
msymbol(Dh)), legend(title("Origin") order(1 "Domestic" 2 "Foreign"))

**Change the color of symbols
graph query colorstyle
palette color magenta emerald
twoway (scatter mpg weight if foreign==0, msymbol(D) mcolor(emerald) mfcolor(yellow))
(scatter mpg weight if foreign==1, msymbol(Dh) mcolor(magenta)), legend(title("Origin")
order(1 "Domestic" 2 "Foreign"))
```

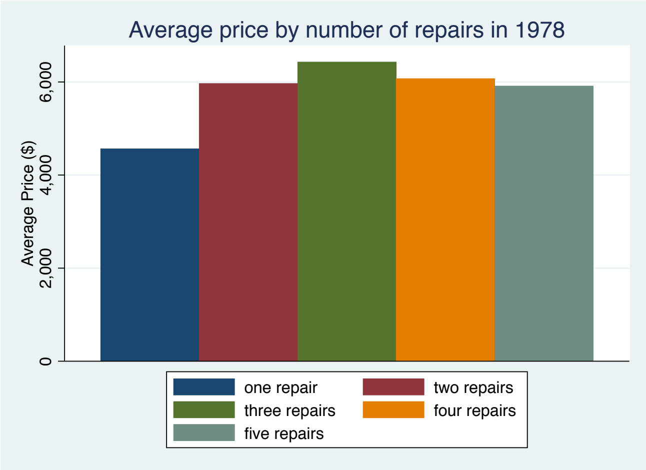


```
**Bar Graphs
graph bar (mean) price, over(rep78)  ytitle("Average Price ($)")
/* Can get mean, median p5, p95, sum, count, min, max*/

**Make the labels more descriptive and add title
graph bar (mean) price, over(rep78, relabel(1 "one repair" 2 "two repairs" 3 "three repairs"
4 "four repairs" 5 "five repairs")) ytitle("Average Price ($)") title("Average price by
number of repairs in 1978")
```



**Another look

```
graph bar (mean) price, over(rep78, relabel(1 "one repair" 2 "two repairs" 3 "three repairs"
4 "four repairs" 5 "five repairs")) ytitle("Average Price ($)") title("Average price by
number of repairs in 1978") asyvars
```

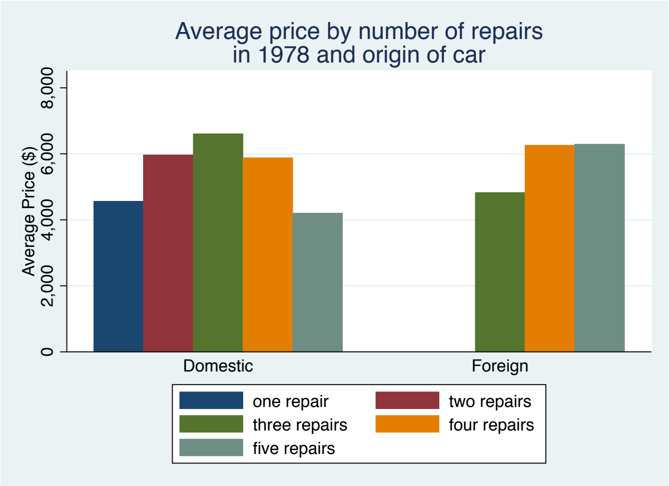


*Stratify by re78 and foreign

```
graph bar (mean) price, over(rep78, relabel(1 "one repair" 2 "two repairs" 3 "three repairs"
4 "four repairs" 5 "five repairs")) asyvars  ytitle("Average Price ($)") title("Average
price by number of repairs in 1978 and origin of car") over(foreign)
```

*Stratify by re78 and foreign- split title onto two lines

```
graph bar (mean) price, over(rep78, relabel(1 "one repair" 2 "two repairs" 3 "three repairs"
4 "four repairs" 5 "five repairs")) asyvars  ytitle("Average Price ($)") title("Average
price by number of repairs” “in 1978 and origin of car") over(foreign) \
```



**Make a stacked bar graph

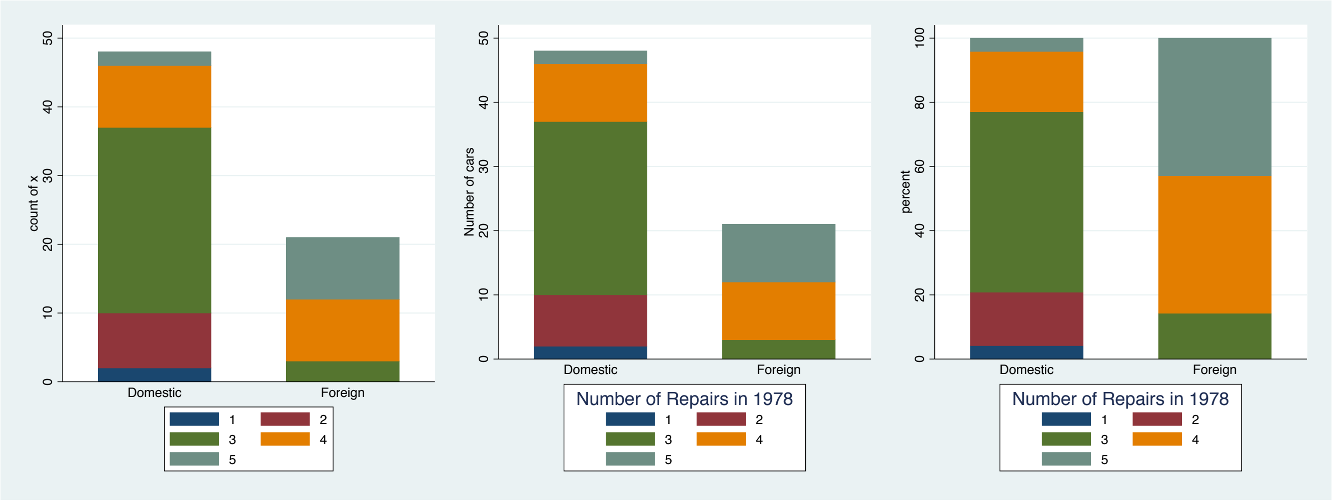
*We want a count of cars by rep78 adn foreign-- since this is individual data, we need to create a variable giving all observations a value of 1

```
gen x=1
graph bar (count) x, over(rep78) over(foreign) asyvars stack
```

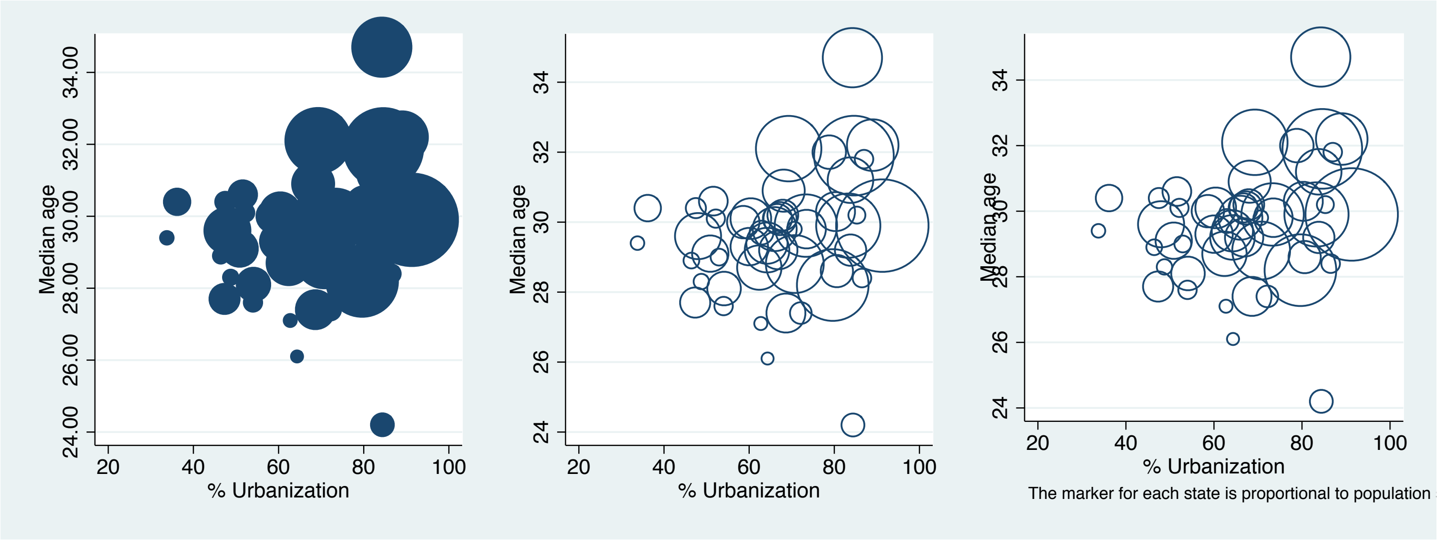
*Add title to legend and change title on y-axis

```
graph bar (count) x, over(rep78) over(foreign) asyvars stack ytitle("Number of cars")
legend(title("Number of Repairs in 1978"))
```

```
*As a percent rather than a count
graph bar (count) x, over(rep78) over(foreign) percent asyvars stack ytitle("Number of
cars") legend(title("Number of Repairs in 1978"))
```



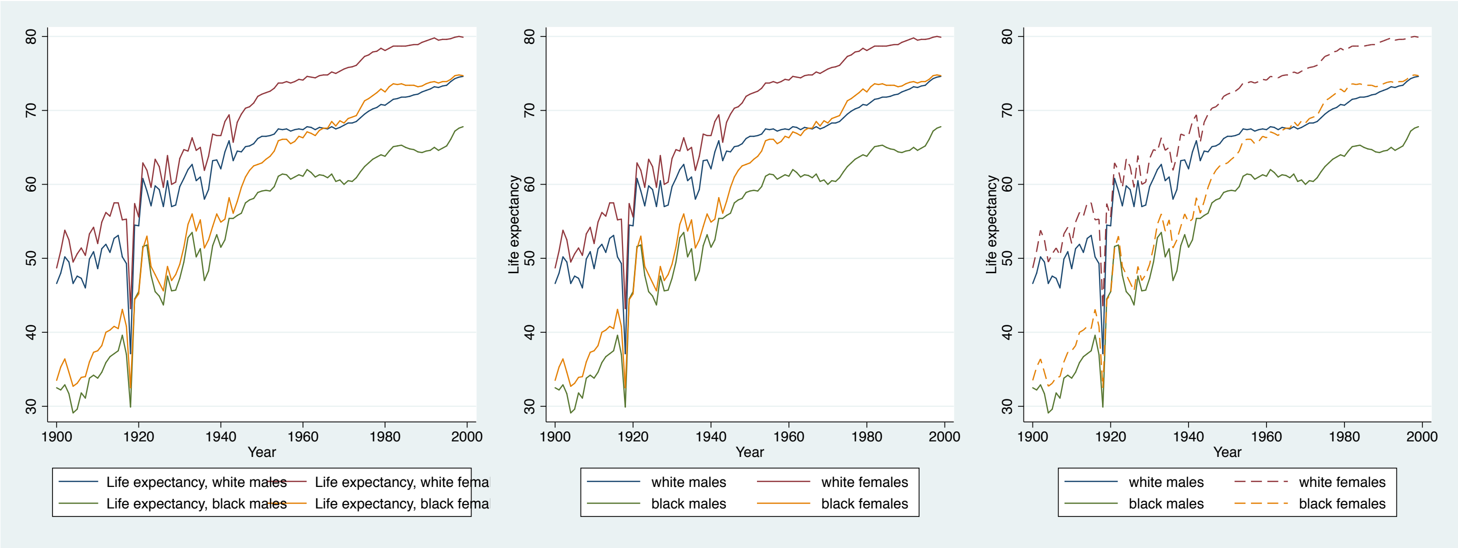
```
**Weighted Scatterplot (weight the dots of the scatterplot by some variable)

sysuse census, clear

*Generate a variable for the percent urban
generate urban = (popurban/pop)*100
label var urban "% Urbanization"

*Scatter plot of median age by percent urban weighted by population size
twoway scatter medage urban [fw=pop]
**Change age format on the y-axis
format medage %9.0f
twoway scatter medage urban [fw=pop], msymbol(Oh) yscale(range(24 35))
twoway scatter medage urban [fw=pop], msymbol(Oh) yscale(range(24 35)) note("The marker for
each state is proportional to population size")
```



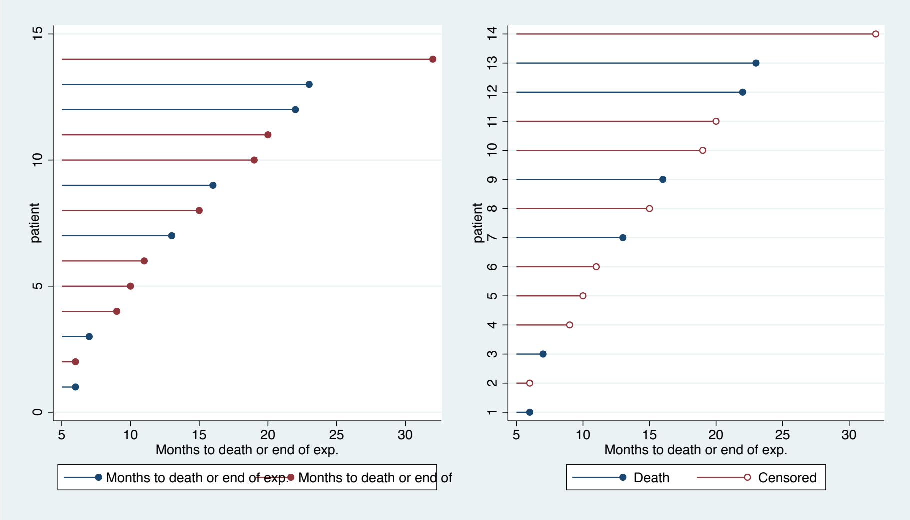
The marker for each state is proportional to population

```
**Line plots
sysuse uslifeexp, clear
twoway line le_wmale le_wfemale le_bmale le_bfemale year
```

```
**Add a title to the y-axis and remove life expectancy from the labels
twoway line le_wmale le_wfemale le_bmale le_bfemale year, ytitle("Life expectancy")
legend(order(1 "white males" 2 "white females" 3 "black males" 4 "black females"))

*Change look of the lines
palette linepalette
twoway line le_wmale le_wfemale le_bmale le_bfemale year, lpattern(solid dash solid dash)
ytitle("Life expectancy") legend(order(1 "white males" 2 "white females" 3 "black males" 4
"black females"))
```

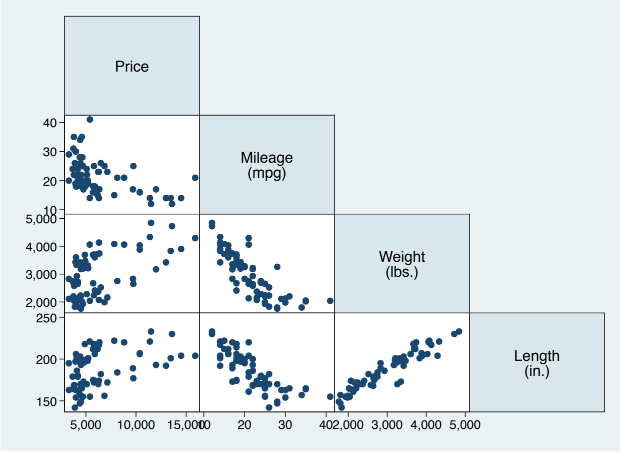


```
**Dropped-line plots (length of time each subject was in the study and whether they died or were censored)
sysuse cancer, clear
**Keep the people who got drug #2
keep if drug==2
**Sort by studytime (order by the length of their follow-up)
sort studytime
**Generate a patient number equal to the observation number that indexes each person
gen patient=_n
twoway (dropline studytime patient if died==1, horizontal)(dropline studytime patient if
died==0, horizontal)

**Remove filling for censored patients, properly label ledgend and create a different scale  for the y axis
twoway (dropline studytime patient if died==1, horizontal)(dropline studytime patient if
died==0, horizontal mfcolor(white)), legend(order(1 "Death" 2 "Censored")) ylabel(1(1)14)
```



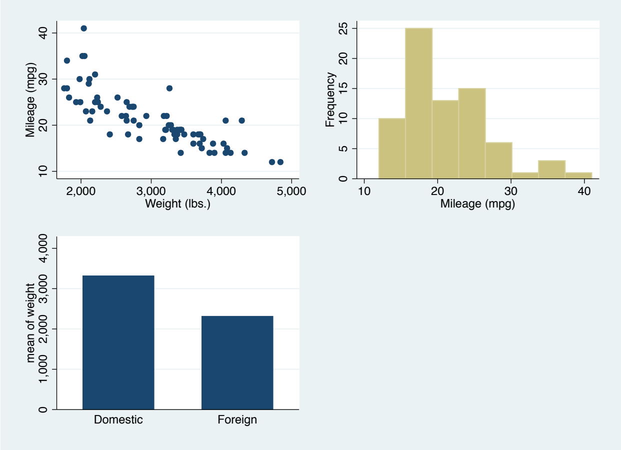
****Correlation plots**

```
sysuse auto, clear
graph matrix price mpg weight length, half
corr price mpg weight length
```



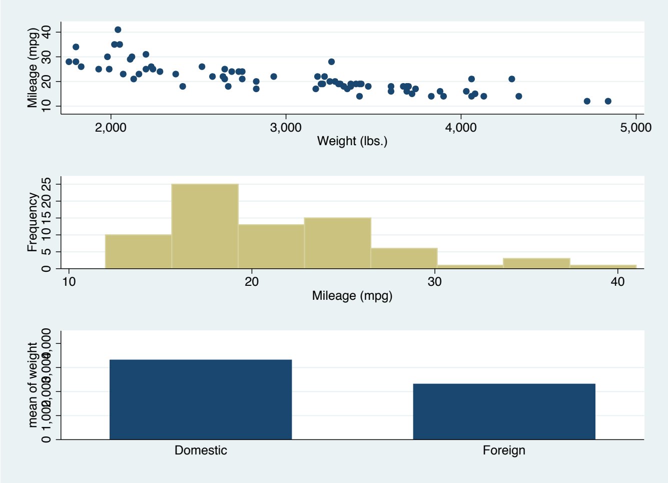
****Combine graphs**

```
twoway scatter mpg weight, name(scatter, replace)
hist mpg, freq name(histogram_mpg, replace)
graph bar weight, over(foreign) name(bar, replace)
graph combine scatter histogram_mpg bar
```



****Arrange them in one column**

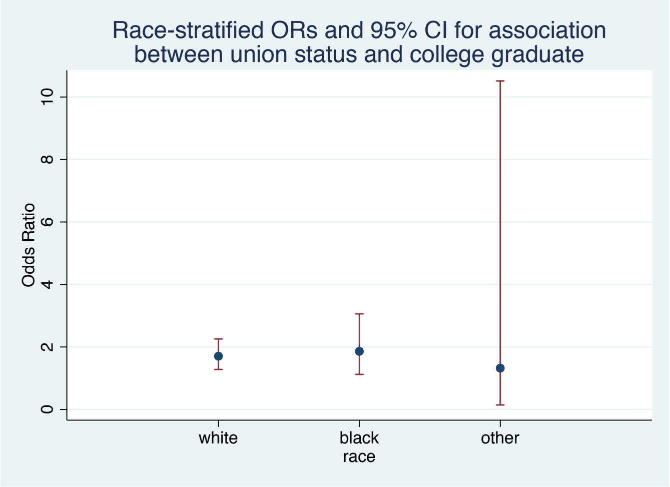
```
graph combine scatter histogram_mpg bar, row(3)
graph combine scatter histogram_mpg bar, col(3)
```



****Other graphs to know about**

****ECLPLOT (useful to plot Odds Ratios and CIs)**

```
sysuse nlsw88, clear
cc union collgrad, by(race)

. cc union collgrad, by(race)

            race |         OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
           white |   1.702344      1.280028   2.256791     37.13673 (exact)
           black |   1.859987      1.121557   3.060468     11.83234 (exact)
           other |       1.32      .1440014   10.51459     1.041667 (exact)
-----------------+-------------------------------------------------
           Crude |    1.64747      1.295668   2.089715              (exact)
    M-H combined |   1.731678      1.368956   2.190507
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(2) =     0.19  Pr>chi2 = 0.9085

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     21.22
                                                Pr>chi2 =    0.0000

. statsby, by(race): cc union collgrad

eclplot or lb_or ub_or race
eclplot or lb_or ub_or race, xscale(range(0 4)) xlabel(1(1)3) ytitle("Odds Ratio")
title("Race-stratified odds ratio and 95% CI for association" "between union status and
college graduate")
```


Race-stratified ORs and 95% CI for association
between union status and college graduate
Odds Ratio
white
black
other
race