CHRISTOPHER PRENER, PH.D.
FALL, 2016

WEEK 13
LECTURE 14

QUANTITATIVE ANALYSIS

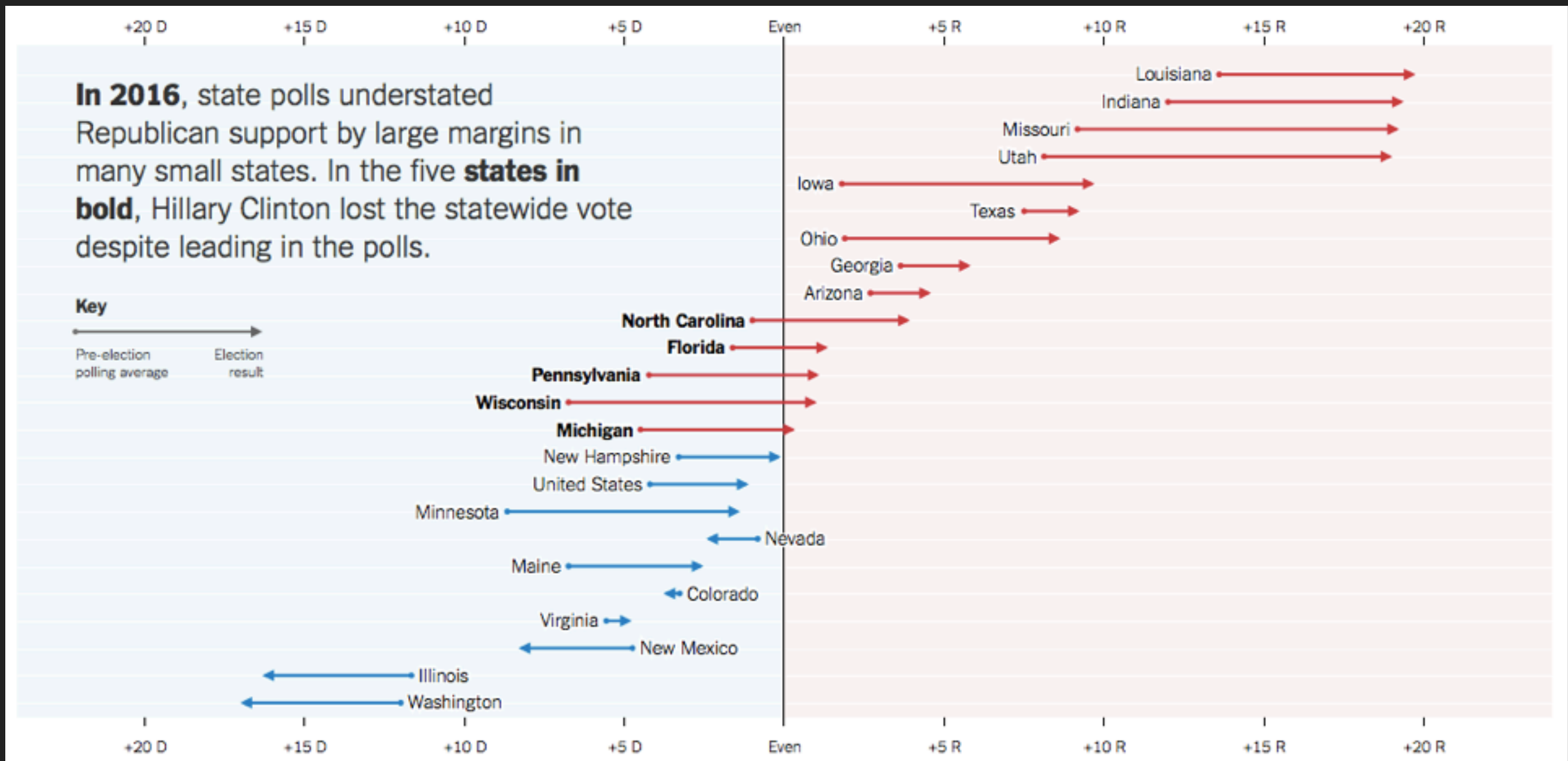
CORRELATIONS (2)

AGENDA

1. Follow-up
2. Scatterplots
3. Pearson's *r*
4. Power Analysis
5. Cronbach's *alpha*
6. Automation and Replication

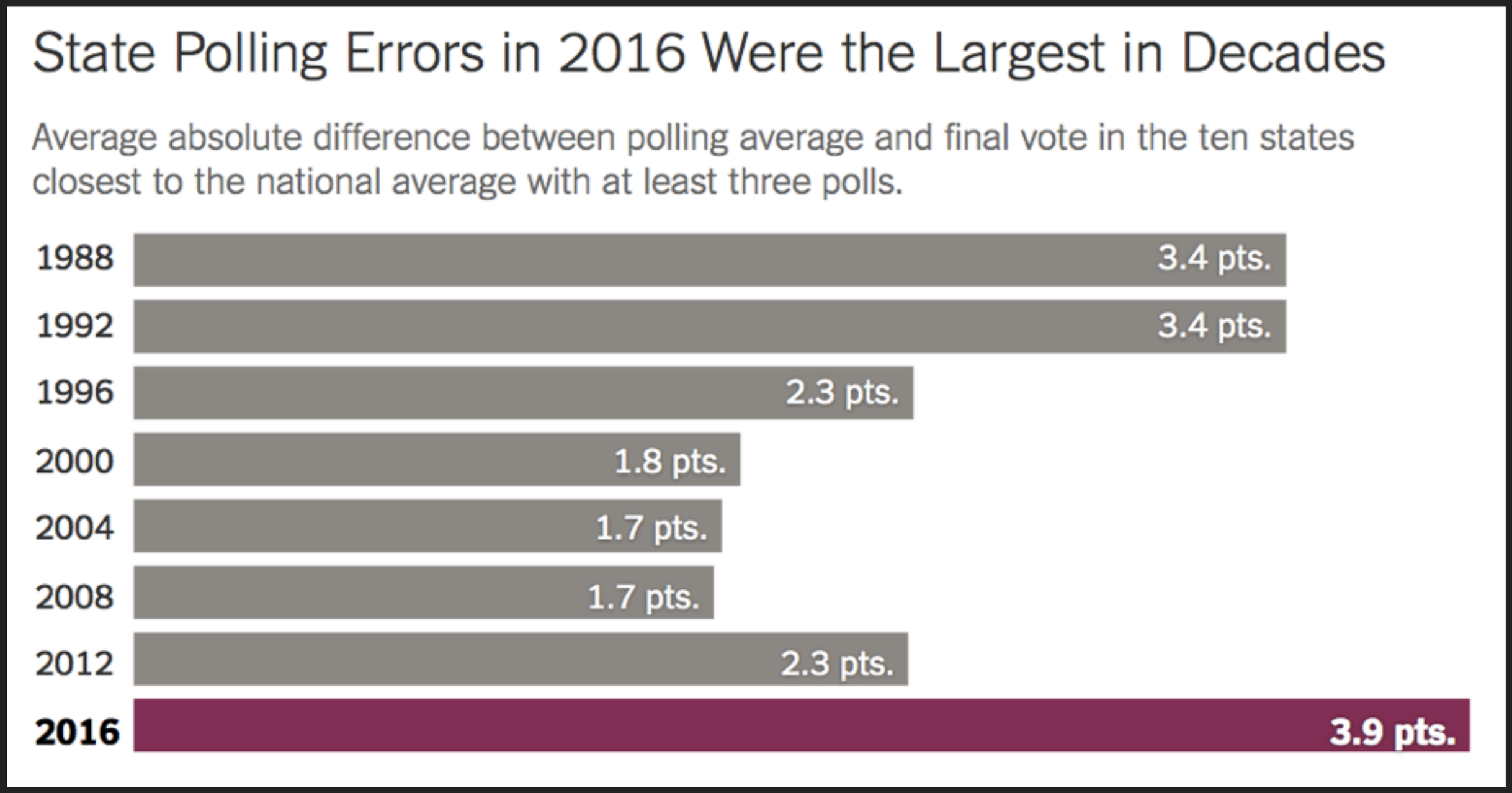
1 FOLLOW-UP

STATE-LEVEL POLLING ERRORS

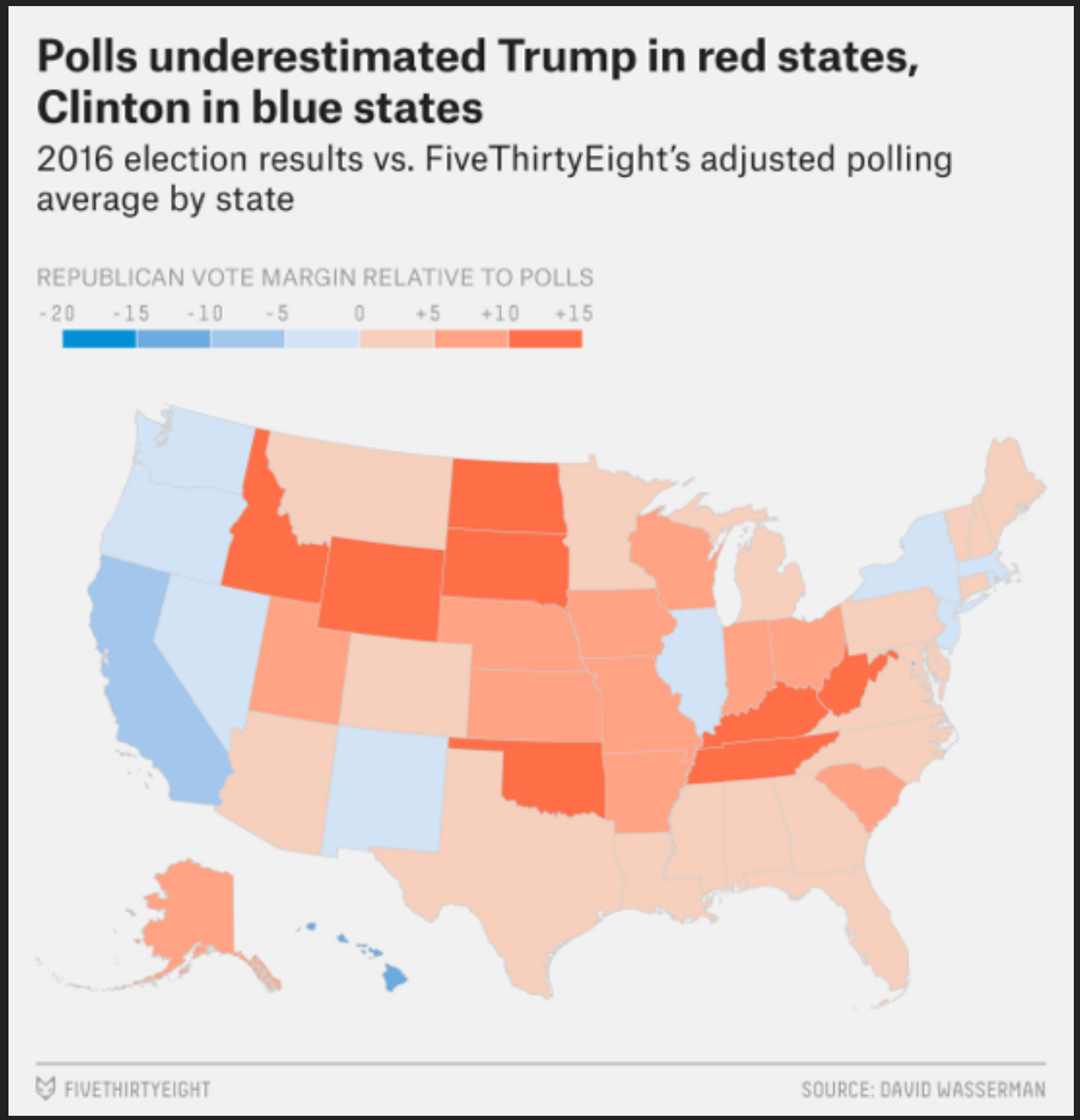


Arrows show simple average of state polls conducted in the three weeks before Election Day in states with at least three polls, compared with the projected final election result.

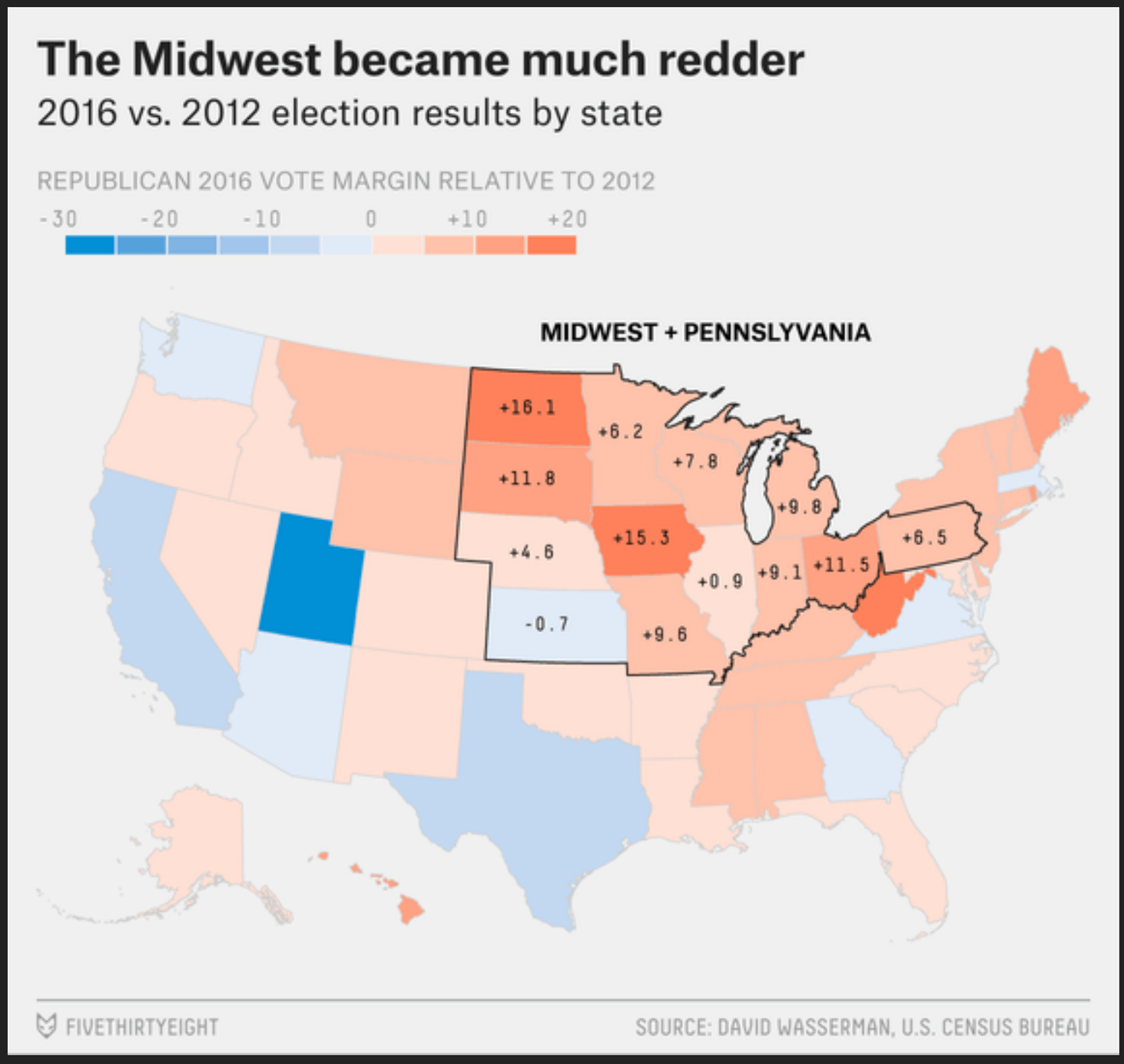
STATE-LEVEL POLLING ERRORS



STATE-LEVEL POLLING ERRORS



STATE-LEVEL POLLING ERRORS ARE CORRELATED



UNDECIDED = UNCERTAINTY

Voters who decided in the final week went strongly for Trump

STATE	VOTE SHARE OF THOSE WHO DECIDED THE WEEK BEFORE THE ELECTION: CLINTON	TRUMP
Wisconsin	30%	59%
Minnesota	31	53
Utah	19	41
Iowa	34	54
Pennsylvania	37	54
Florida	38	55
Maine	33	49
New Hampshire	37	52
Michigan	39	50
North Carolina	41	49
New Mexico	41	46
Ohio	43	46
Virginia	45	42
Nevada	45	40
Georgia	52	42

The exit poll did not provide a breakout of voters who decided in the last week in Colorado or Arizona, because the sample size was too small.

SOURCE: NATIONAL EXIT POLL

2 SCATTER PLOTS

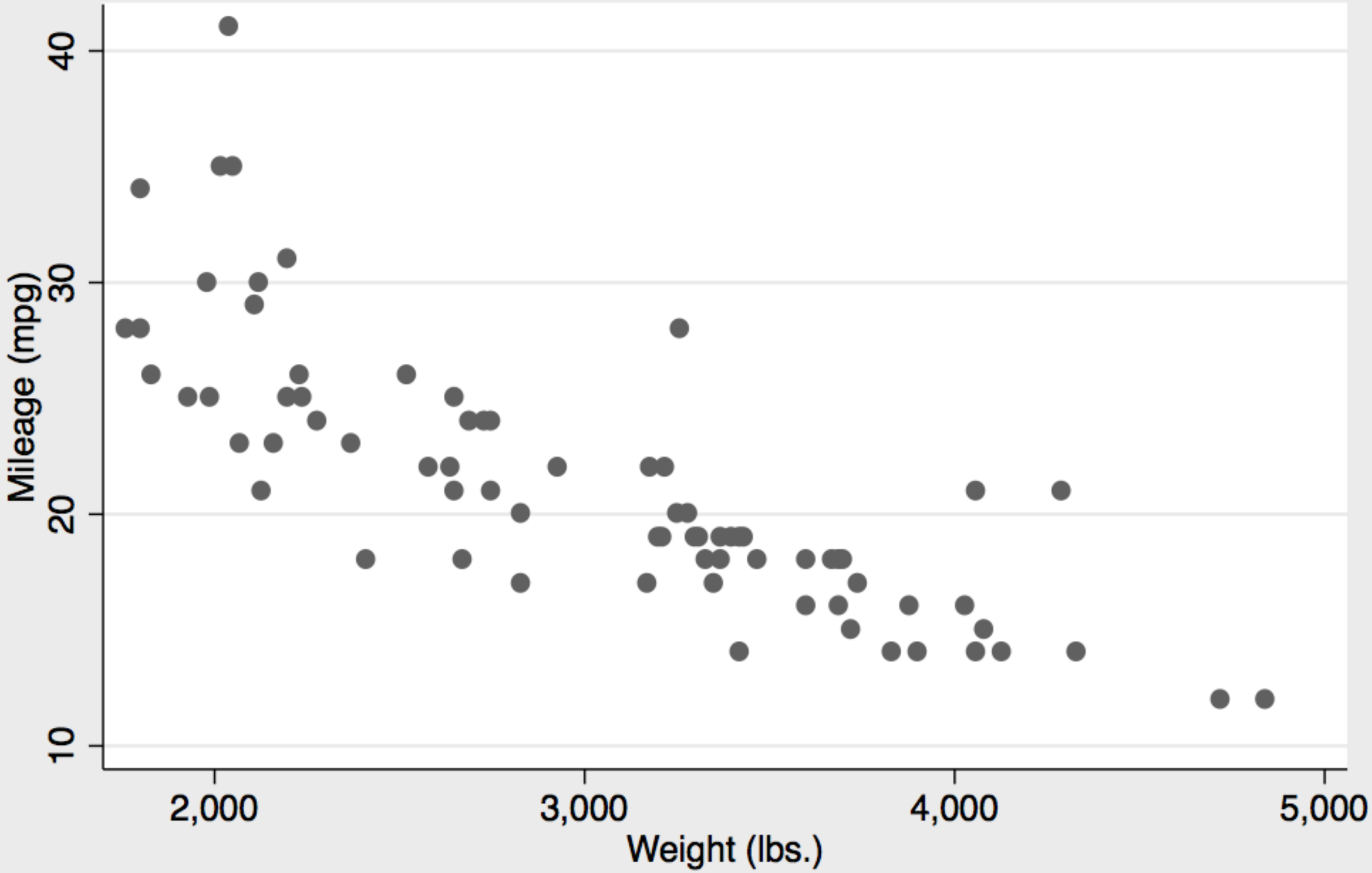
SCATTERPLOTS

```
graph twoway scatter yvar xvar [, STANDARD OPTIONS]
```

```
. graph twoway scatter mpg weight
```

Miles per Gallon and Vehicle Weight

1978 Automobiles



SCATTERPLOTS WITH A REGRESSION LINE

```
graph twoway (lfit yvar xvar) (scatter yvar xvar) [, STD OPTIONS]
```

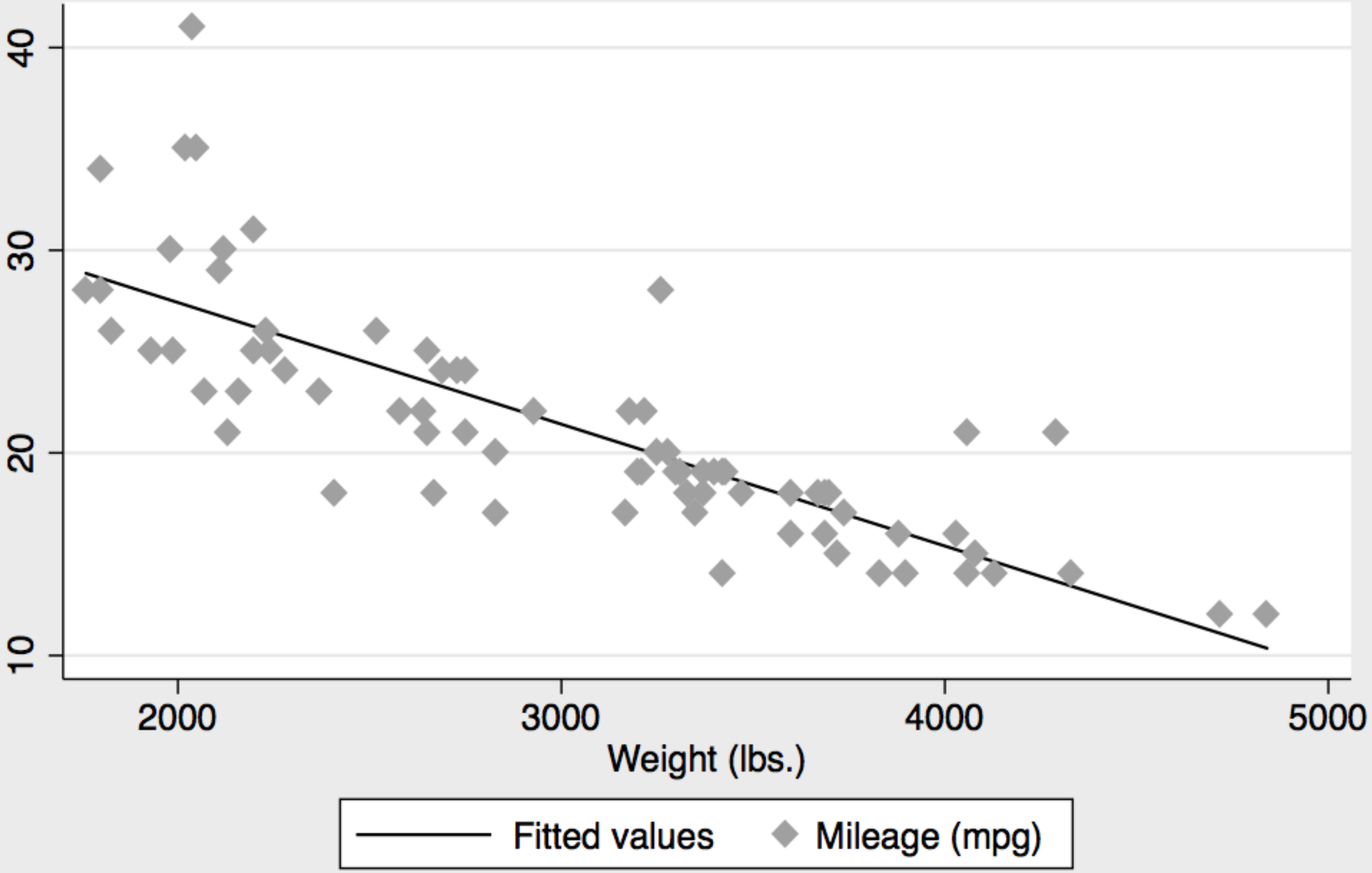
```
. graph twoway (lfit mpg weight) (scatter mpg weight)
```

DEFINITION

STRAIGHT LINE THAT IS DRAWN THROUGH THE CENTER MASS OF THE POINTS ON A SCATTERPLOT – USED TO ILLUSTRATE TRENDS IN A DATASET

Miles per Gallon and Vehicle Weight

1978 Automobiles



SCATTERPLOTS WITH A REGRESSION LINE

```
graph twoway (lfitci yvar xvar) (scatter yvar xvar) [, STD OPTIONS]
```

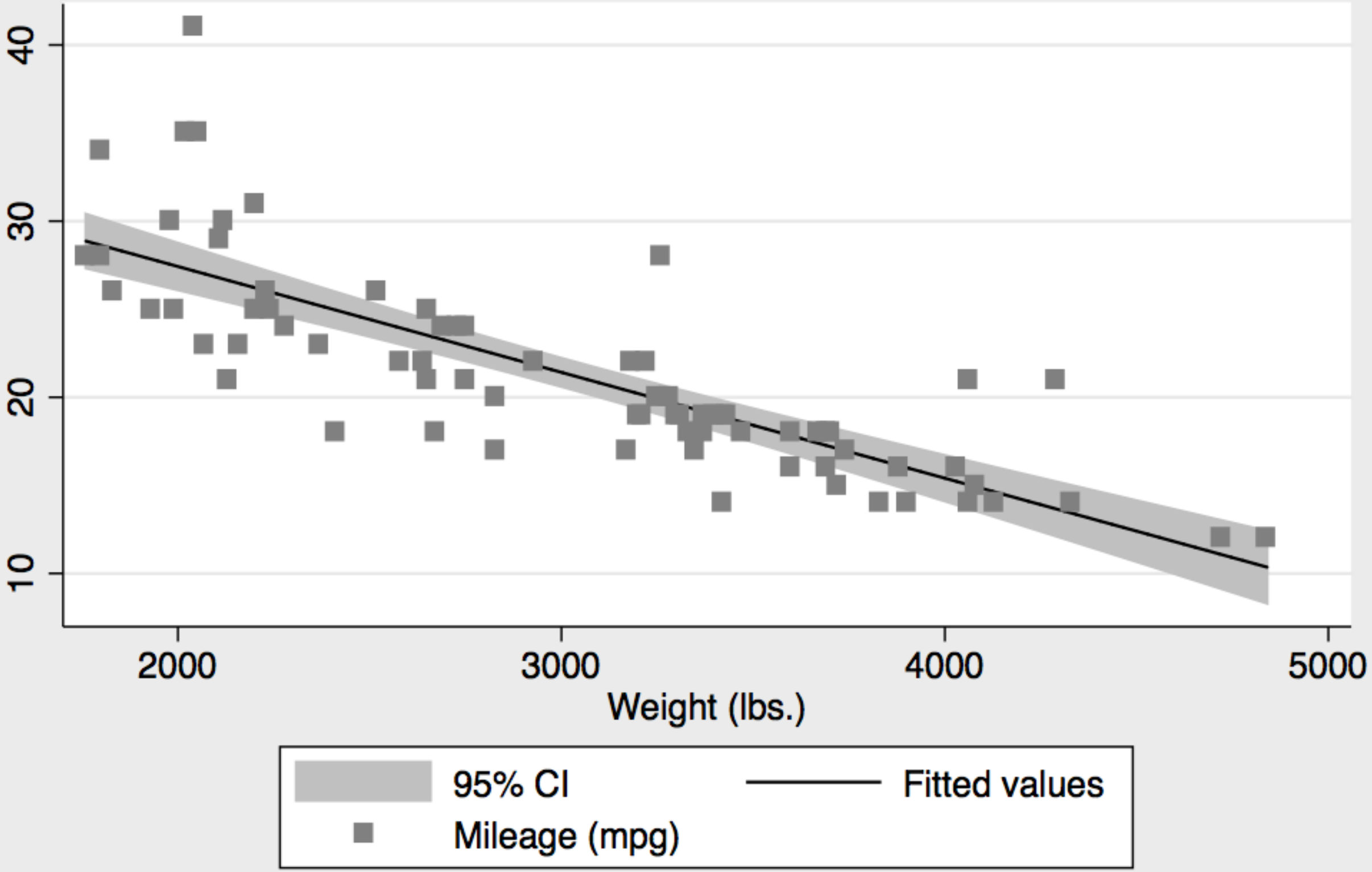
```
. graph twoway (lfitci mpg weight) (scatter mpg weight)
```

DEFINITION

THE TREND IS AN ESTIMATE, SO IT HAS A CORRESPONDING CONFIDENCE INTERVAL THAT REPRESENTS THE UNCERTAINTY OF THE ESTIMATE

Miles per Gallon and Vehicle Weight

1978 Automobiles



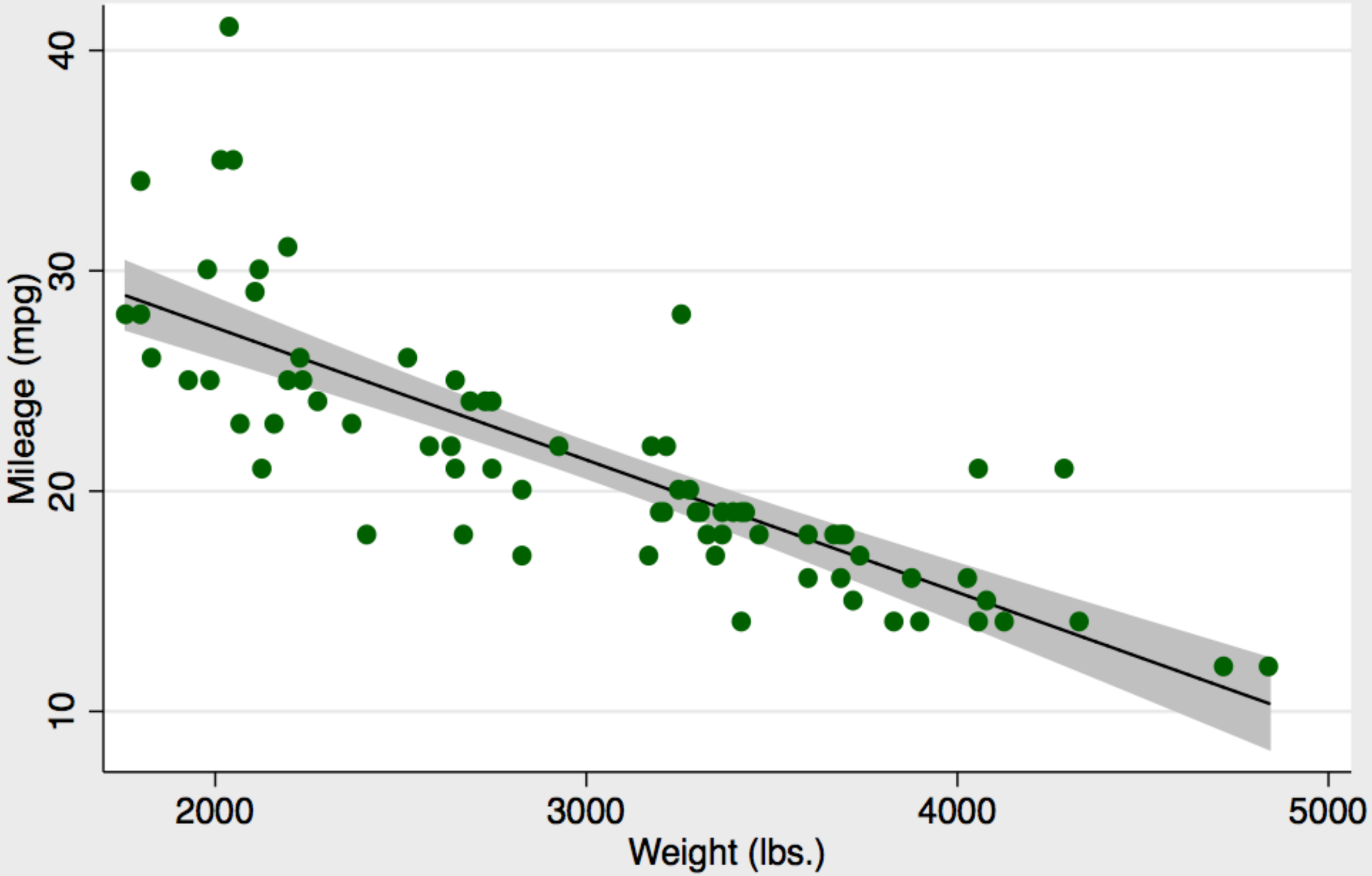
CLEANING UP SCATTERPLOTS

```
// options:

// - turn legend off - legend(off)

// - change symbol - msymbol(circle)
```

```
. graph twoway (lfitci mpg weight) ///
    (scatter mpg weight, msymbol(circle) mcolor(dkgreen)), ///
    legend(off)
```

Miles per Gallon and Vehicle Weight

1978 Automobiles



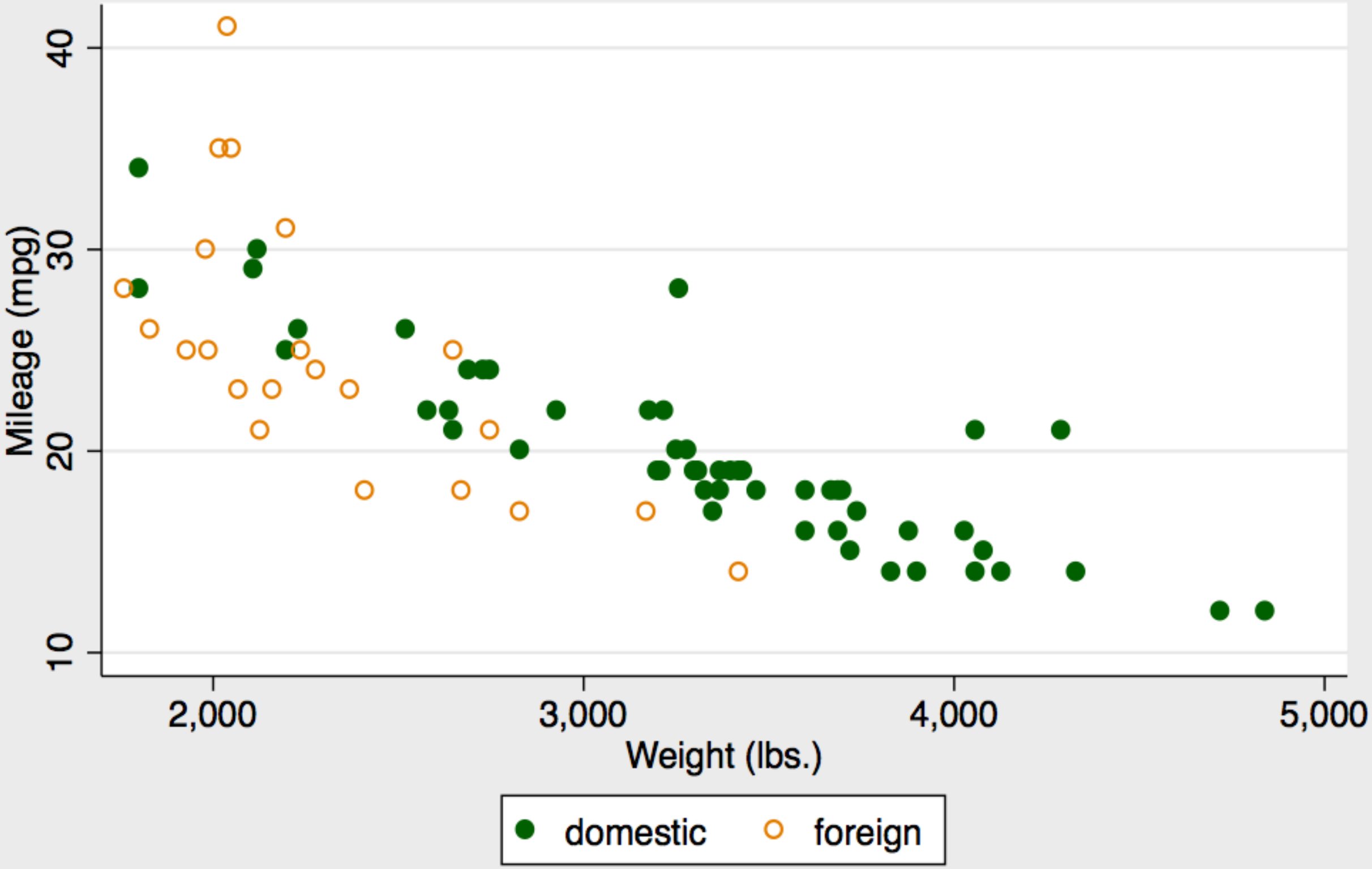
FACTORING

```
// use if qualifier to facet plot; use symbols and color to

// differentiate between groups
```

```
. graph twoway (scatter mpg weight if foreign==0, ///
    msymbol(circle) mcolor(dkgreen)) ///
    (scatter mpg weight if foreign==1, ///
    msymbol(circle_hollow) mlcolor(dkorange))
```

Miles per Gallon and Vehicle Weight

1978 Automobiles



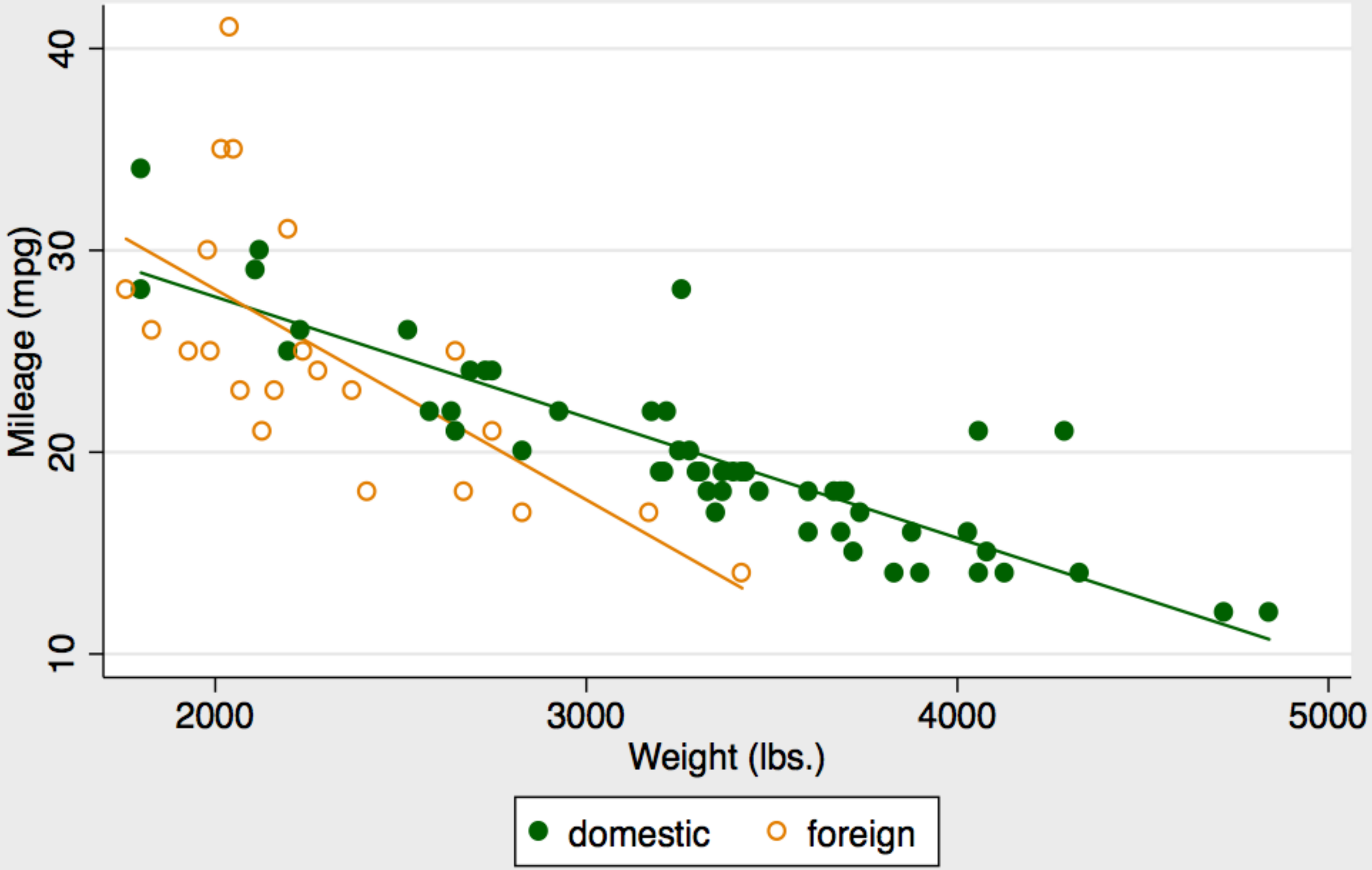
FACTORING WITH REGRESSION LINES

```
// add two new layers, one for reach line; use symbols and color to

// differentiate between groups; edit the legend for clarity
```

```
. graph twoway (lfit mpg weight if foreign==0, lcolor(dkgreen)) ///
    (lfit mpg weight if foreign==1, lcolor(dkorange) lpattern(solid)) ///
    (scatter mpg weight if foreign==0, msymbol(circle) ///
    mcolor(dkgreen)) ///
    (scatter mpg weight if foreign==1, msymbol(circle_hollow)
    mlcolor(dkorange)), legend(order(3 "domestic" 4 "foreign"))
```

Miles per Gallon and Vehicle Weight

1978 Automobiles

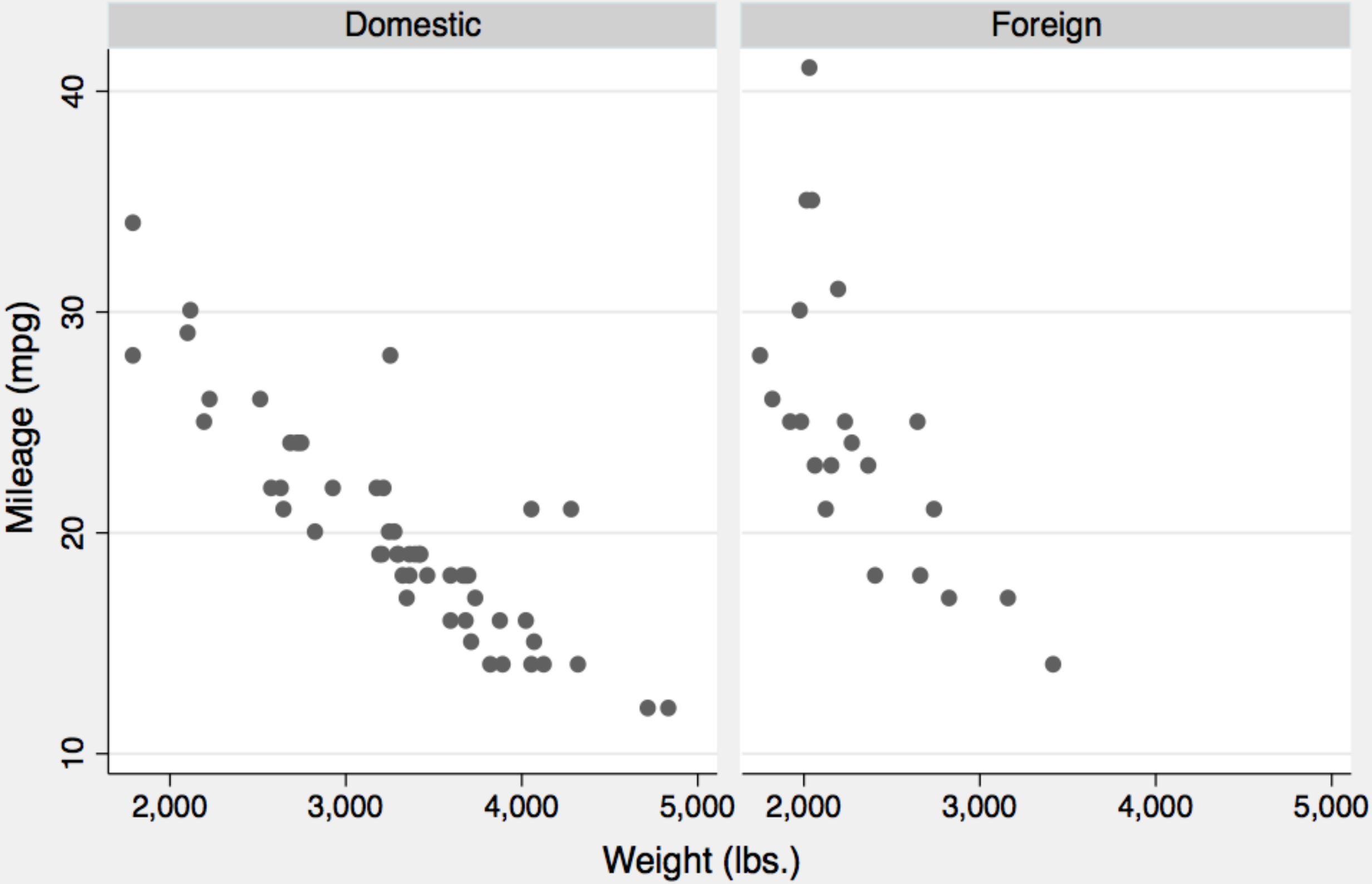


FACETING

```
// use the by() option with title() and subtitle() specified as

// suboptions within by() to adjust their positioning appropriately
```

```
. graph twoway scatter mpg weight, ///
    by(foreign, title("Miles per Gallon and Vehicle Weight"))
```

Miles per Gallon and Vehicle Weight



Graphs by Car type

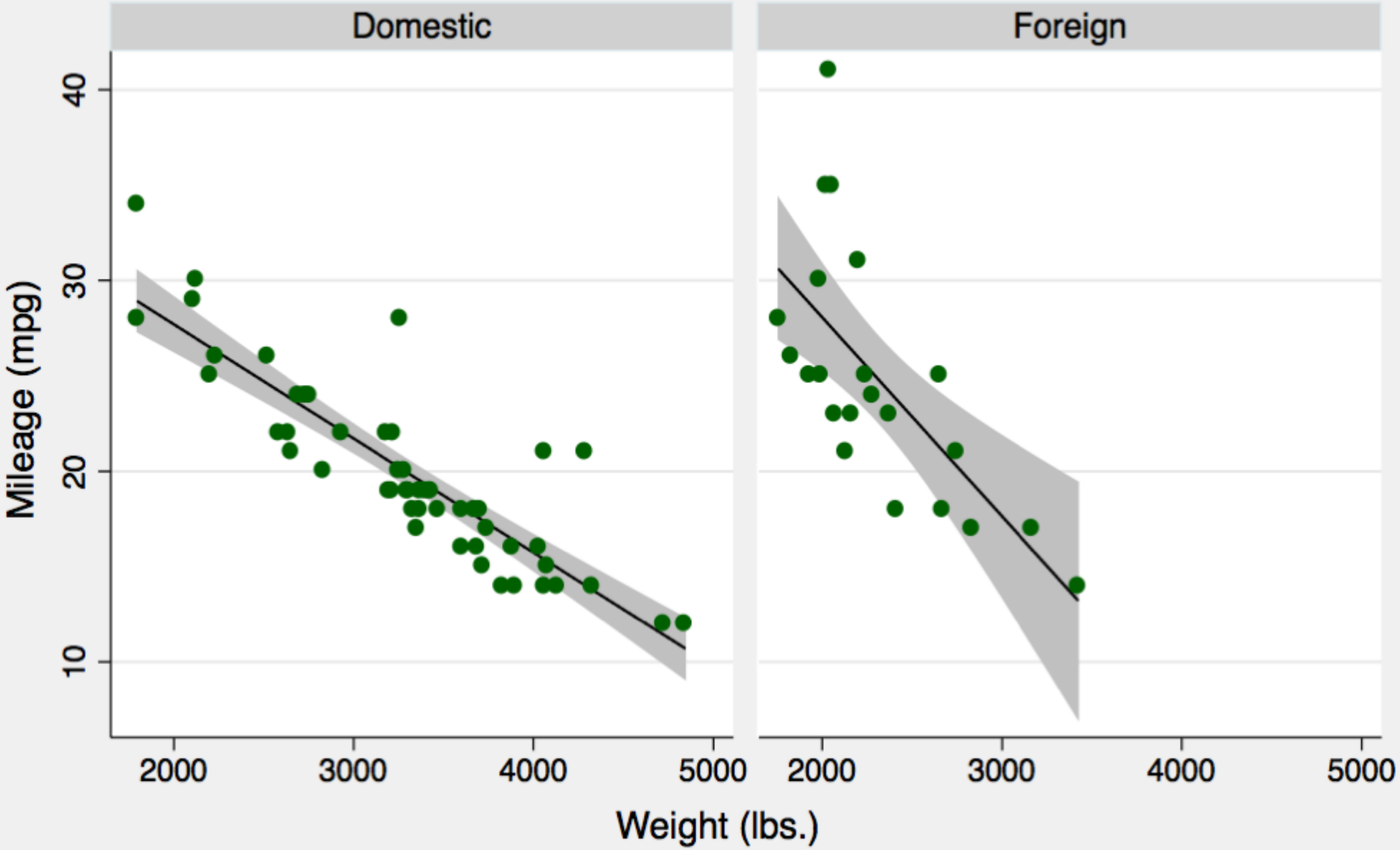
FACETING WITH REGRESSION LINES

```
// use the by() option with title() and subtitle() specified as

// suboptions within by() to adjust their positioning appropriately
```

```
. graph twoway (lfitci mpg weight) ///
    (scatter mpg weight, msymbol(circle) mcolor(dkgreen)), ///
    by(foreign, title("Miles per Gallon and Vehicle Weight") ///
    subtitle("1978 Automobiles") legend(off))
```

Miles per Gallon and Vehicle Weight

1978 Automobiles



Graphs by Car type

3 PEARSON'S R

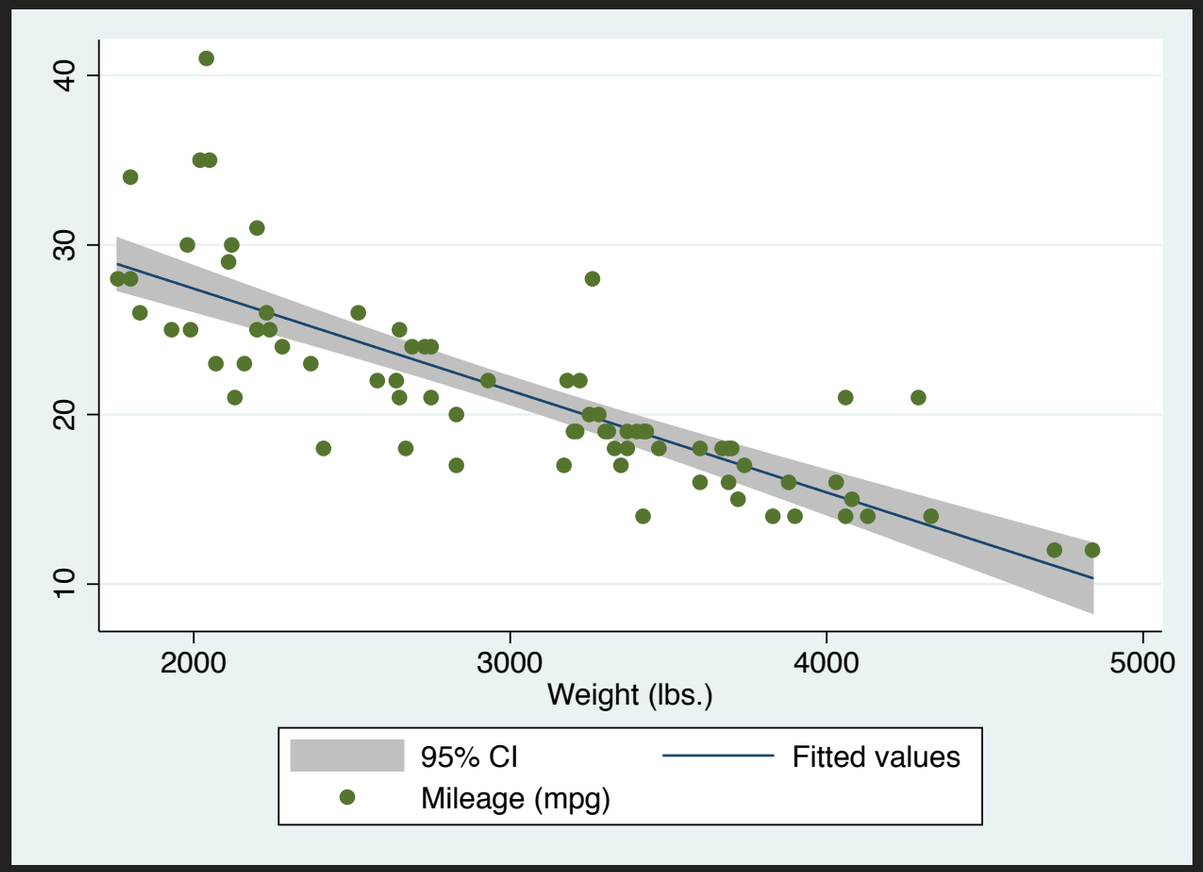
PAIRWISE VS. LISTWISE DELETION

```
. corr mpg weight rep78
(obs=69)
             |      mpg   weight    rep78
-------------+---------------------------
         mpg |   1.0000
      weight |  -0.8055   1.0000
       rep78 |   0.4023  -0.4003   1.0000

. pwcorr mpg weight rep78
             |      mpg   weight    rep78
-------------+---------------------------
         mpg |   1.0000
      weight |  -0.8072   1.0000
       rep78 |   0.4023  -0.4003   1.0000
```

```
. list mpg weight rep78 in 1/10
     +---------------------+
     | mpg   weight   rep78 |
     |---------------------|
  1. |  22    2,930       3 |
  2. |  17    3,350       3 |
  3. |  22    2,640       . |
  4. |  20    3,250       3 |
  5. |  15    4,080       4 |
     |---------------------|
  6. |  18    3,670       3 |
  7. |  26    2,230       . |
  8. |  20    3,280       3 |
  9. |  16    3,880       3 |
 10. |  19    3,400       3 |
     +---------------------+
```

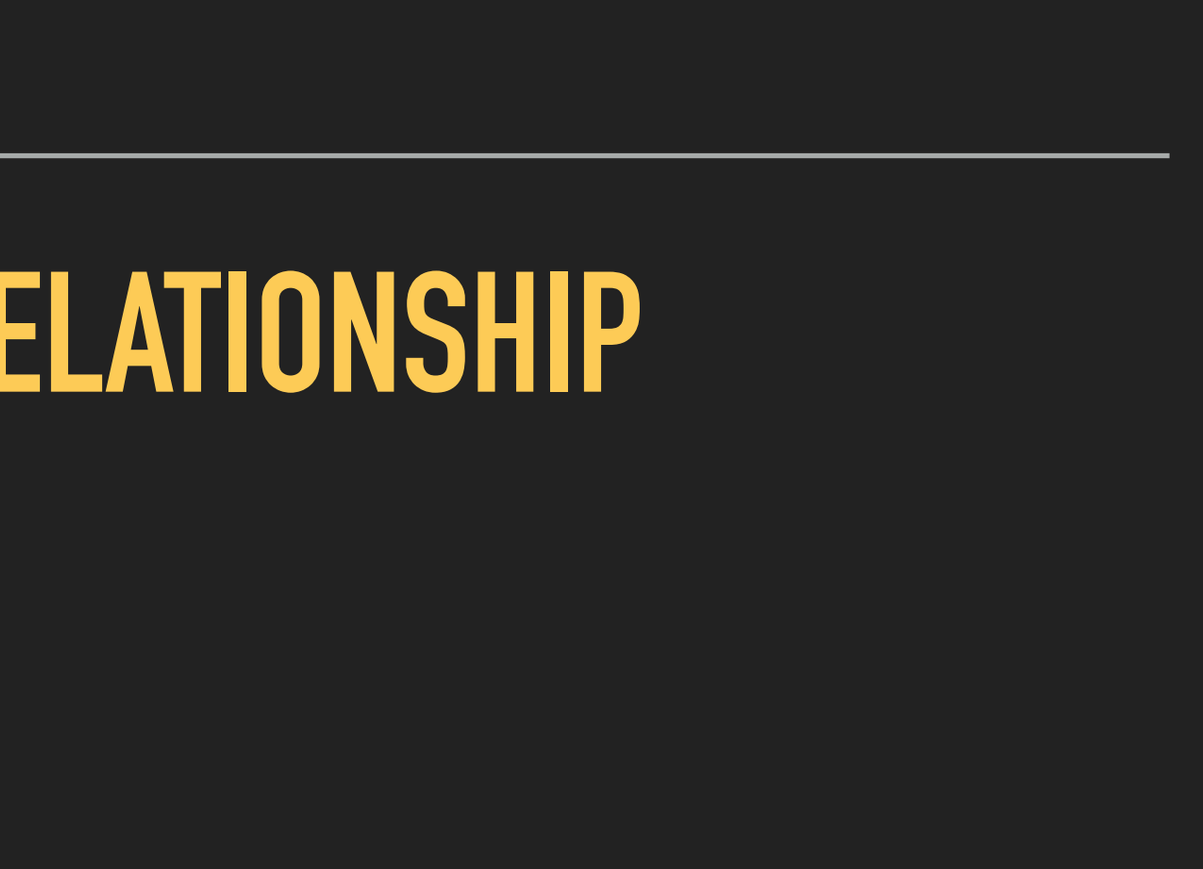
CONTINUOUS-CONTINUOUS RELATIONSHIP

```
pwcorr varlist [, listwise sig]

. pwcorr mpg weight foreign, listwise sig

             |      mpg   weight  foreign
-------------+---------------------------
         mpg |   1.0000
             |
             |
      weight |  -0.8072   1.0000
             |   0.0000
             |
     foreign |   0.3934  -0.5928   1.0000
             |   0.0005   0.0000
             |
```



CONTINUOUS-BINARY RELATIONSHIP

```
. pwcorr mpg weight foreign, listwise sig

             |      mpg   weight  foreign
-------------+---------------------------
         mpg |   1.0000
             |
             |
      weight |  -0.8072   1.0000
             |   0.0000
             |
     foreign |   0.3934  -0.5928   1.0000
             |   0.0005   0.0000
             |
```



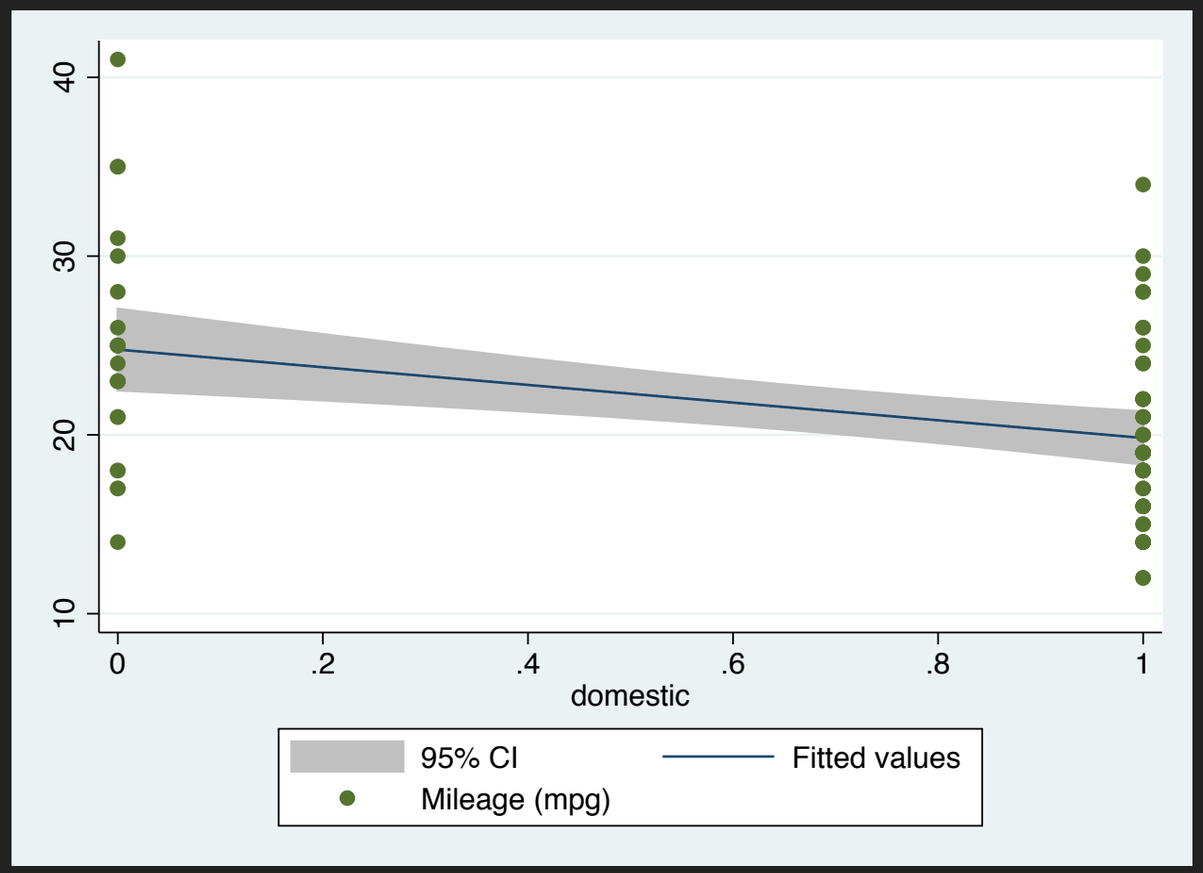
CONTINUOUS-BINARY RELATIONSHIP

```
. pwcorr mpg weight foreign, listwise sig

             |      mpg   weight  foreign
-------------+---------------------------
         mpg |   1.0000
             |
             |
      weight |  -0.8072   1.0000
             |   0.0000
             |
     foreign |   0.3934  -0.5928   1.0000
             |   0.0005   0.0000
             |
```

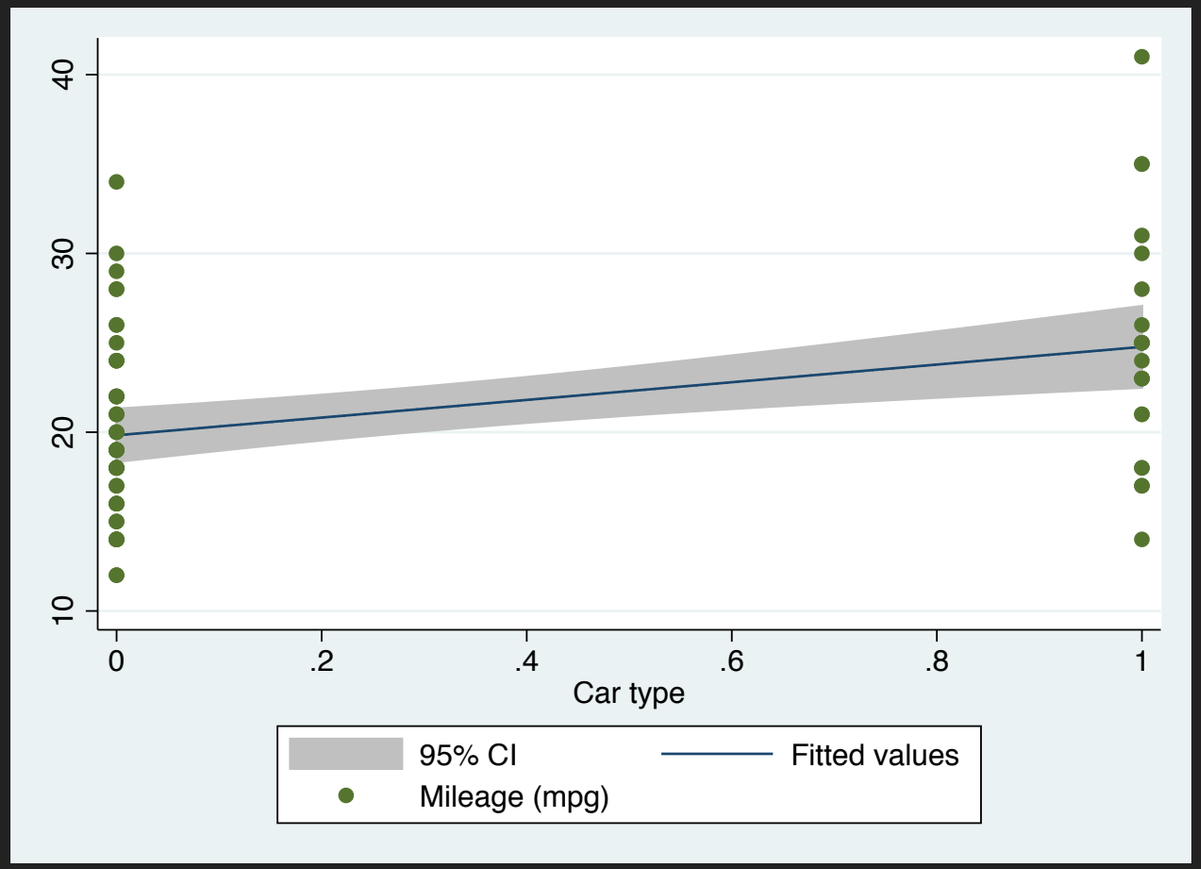
```
. generate domestic = foreign

. recode domestic 0=1 1=0
(domestic: 74 changes made)

. label define domestic ///
    0 foreign 1 domestic

. label values domestic domestic

. label variable "Car type, domestic"
```

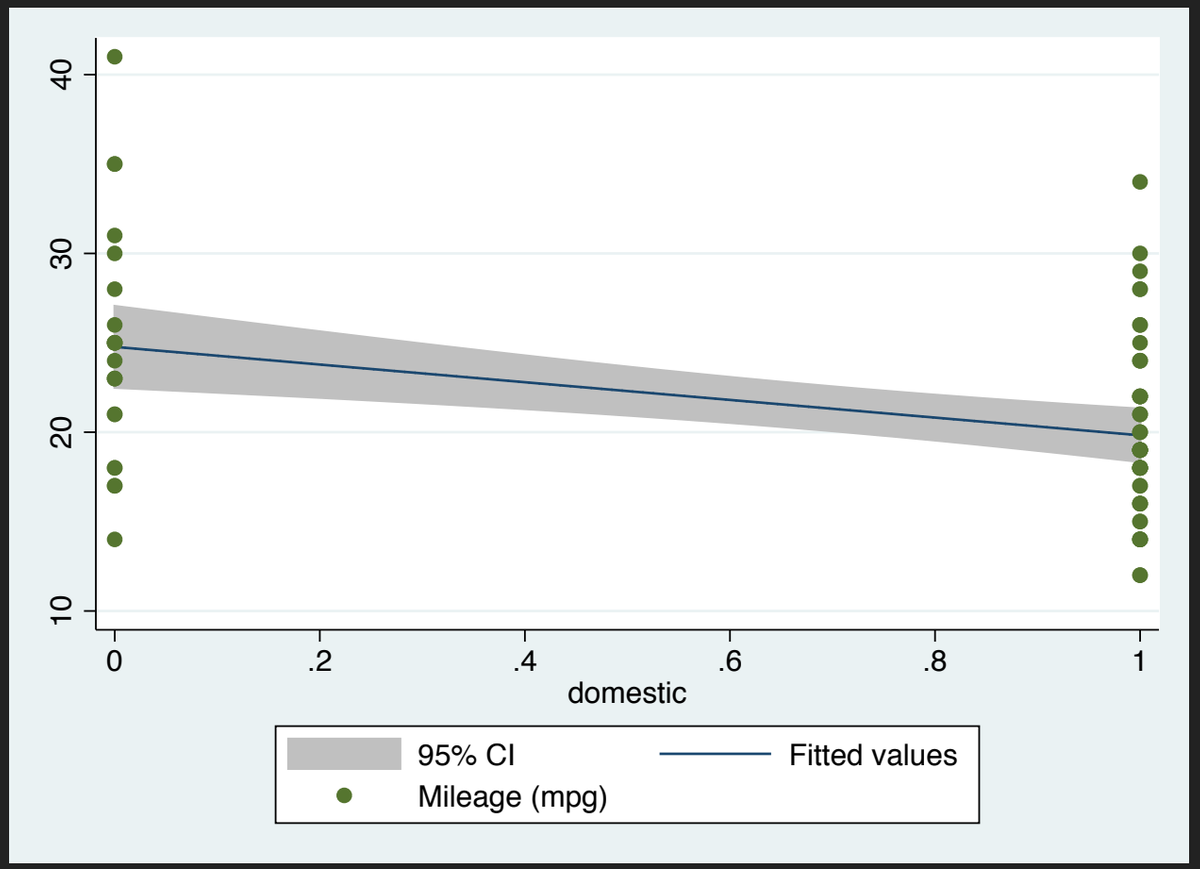
CONTINUOUS-BINARY RELATIONSHIP

```
. pwcorr mpg weight foreign domestic, listwise sig

             |      mpg   weight  foreign domestic
-------------+------------------------------------
         mpg |   1.0000
             |
             |
      weight |  -0.8072   1.0000
             |   0.0000
             |
     foreign |   0.3934  -0.5928   1.0000
             |   0.0005   0.0000
             |
    domestic |  -0.3934   0.5928  -1.0000
             |   0.0005   0.0000   1.0000
             |
```



CONTINUOUS–BINARY RELATIONSHIP





REPORTING CORRELATIONS

```
. pwcorr mpg weight foreign, listwise sig

             |      mpg   weight  foreign
-------------+---------------------------
         mpg |   1.0000
             |
             |
      weight |  -0.8072   1.0000
             |   0.0000
             |
     foreign |   0.3934  -0.5928   1.0000
             |   0.0005   0.0000
             |
```

REPORTING CORRELATIONS

	mpg	weight	foreign
mpg	1.00	-	-
weight	-.807***	1.00	-
foreign	.393***	-.593***	1.00

* - p < .05; ** - p < .01; *** - p < .001

4 POWER ANALYSIS

KEY ASSUMPTIONS

- What is the null hypothesis? $r = 0$
- What effect size is meaningful to calculate? $r = ?$
- How much power do I want to have? β=0.8 or β=0.9 are common
- What alpha level am I using? typically α=0.05

ESTIMATING SAMPLE SIZE

```
power onecorrelation nullVal effectSize [, power(val)]
```

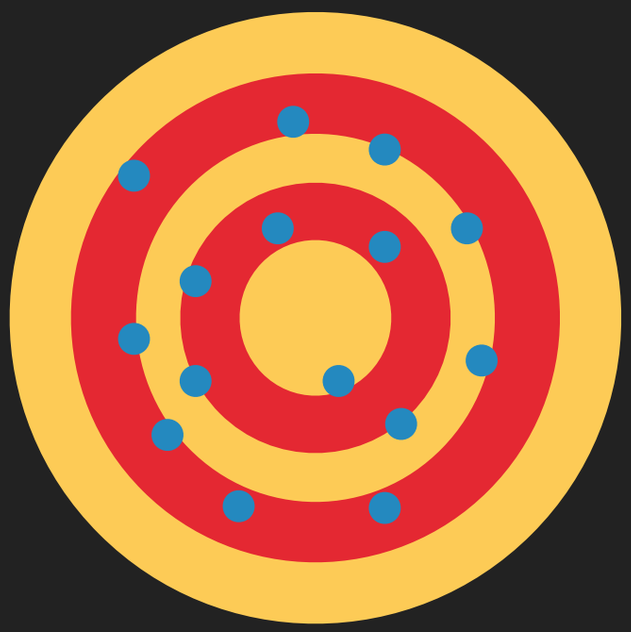
```
. power onecorrelation 0 .3, power(.8)

Performing iteration ...

Estimated sample size for a one-sample correlation test
Fisher's z test
Ho: r = r0  versus  Ha: r != r0

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =    0.3000
           r0 =    0.0000
           ra =    0.3000

Estimated sample size:

            N =        85
```

5 CRONBACH'S ALPHA

RELIABILITY VS VALIDITY

Reliable, Not Valid

Valid, Not Reliable



Not Reliable or Valid

Valid and Reliable

THEORY

What is the size of the car?

DEFINITION

THE PROPORTION OF THE OBSERVED VARIANCE THAT REPRESENTS THE TRUE VARIANCE. INTERPRETED AS A MEASURE OF HOW RELIABLY A GROUP OF VARIABLES MEASURE A SINGLE FACTOR (LATENT CONSTRUCT).

THEORY

What is the size of the car?

```
alpha varlist [, std detail generate(scaleVar)]

. alpha weight length turn displacement, std

Test scale = mean(standardized items)

Average interitem correlation:      0.8624
Number of items in the scale:            4
Scale reliability coefficient:      0.9616
```

DEFINITION

THE PROPORTION OF THE OBSERVED VARIANCE THAT REPRESENTS THE TRUE VARIANCE. INTERPRETED AS A MEASURE OF HOW RELIABLY A GROUP OF VARIABLES MEASURE A SINGLE FACTOR (LATENT CONSTRUCT).

THEORY

What is the size of the car?

```
. alpha weight length turn displacement, std

Test scale = mean(standardized items)

Average interitem correlation:     0.8624
Number of items in the scale:           4
Scale reliability coefficient:     0.9616
```

INTERPRETATION

NO HARD RULES, BUT $\alpha < .7$ IS OFTEN CONSIDERED ADEQUATE.

FULL OUTPUT

```
. alpha weight length turn displacement, std item
Test scale = mean(standardized items)

                                                              average
                          item-test     item-rest       interitem
Item          |  Obs  Sign   correlation   correlation     correlation     alpha
--------------+------------------------------------------------------------------
weight        |   74    +       0.9763       0.9569          0.8254      0.9341
length        |   74    +       0.9624       0.9319          0.8430      0.9416
turn          |   74    +       0.9236       0.8645          0.8920      0.9612
displacement  |   74    +       0.9258       0.8683          0.8892      0.9601
--------------+------------------------------------------------------------------
Test scale    |                                              0.8624      0.9616
------------------------------------------------------------------------------
```

FULL OUTPUT

```
. alpha weight length turn displacement, std item
Test scale = mean(standardized items)

                                                              average
                        item-test     item-rest       interitem
Item          |  Obs  Sign   correlation   correlation     correlation     alpha
--------------+------------------------------------------------------------------
weight        |   74    +        0.9763        0.9569          0.8254      0.9341
length        |   74    +        0.9624        0.9319          0.8430      0.9416
turn          |   74    +        0.9236        0.8645          0.8920      0.9612
displacement  |   74    +        0.9258        0.8683          0.8892      0.9601
--------------+------------------------------------------------------------------
Test scale    |                                               0.8624      0.9616
------------------------------------------------------------------------------
```

alpha = alpha with all items in scale

FULL OUTPUT

```
. alpha weight length turn displacement, std item
Test scale = mean(standardized items)
                                                            average
                             item-test     item-rest       interitem
Item         |  Obs  Sign   correlation   correlation    correlation     alpha
-------------+-----------------------------------------------------------------
weight       |   74    +      0.9763        0.9569          0.8254      0.9341
length       |   74    +      0.9624        0.9319          0.8430      0.9416
turn         |   74    +      0.9236        0.8645          0.8920      0.9612
displacement |   74    +      0.9258        0.8683          0.8892      0.9601
-------------+-----------------------------------------------------------------
Test scale   |                                              0.8624      0.9616
-------------------------------------------------------------------------------
```

item-test = correlation between item and the test scale as a whole

FULL OUTPUT

```
. alpha weight length turn displacement, std item
Test scale = mean(standardized items)

                                                              average
                          item-test     item-rest       interitem
Item          |  Obs  Sign   correlation   correlation     correlation     alpha
--------------+------------------------------------------------------------------
weight        |   74    +       0.9763       0.9569          0.8254       0.9341
length        |   74    +       0.9624       0.9319          0.8430       0.9416
turn          |   74    +       0.9236       0.8645          0.8920       0.9612
displacement  |   74    +       0.9258       0.8683          0.8892       0.9601
--------------+------------------------------------------------------------------
Test scale    |                                              0.8624       0.9616
------------------------------------------------------------------------------
```

item-rest = correlation between item and the test scale as a whole (calculated without item)

FULL OUTPUT

```
. alpha weight length turn displacement, std item
Test scale = mean(standardized items)

                                                    average
                          item-test     item-rest    interitem
Item          |  Obs  Sign   correlation   correlation     correlation     alpha
--------------+-----------------------------------------------------------------
weight        |   74    +        0.9763        0.9569          0.8254    0.9341
length        |   74    +        0.9624        0.9319          0.8430    0.9416
turn          |   74    +        0.9236        0.8645          0.8920    0.9612
displacement  |   74    +        0.9258        0.8683          0.8892    0.9601
--------------+-----------------------------------------------------------------
Test scale    |                                                0.8624    0.9616
-----------------------------------------------------------------------------
```

average interitem correlation = mean of correlations

between individual items excluding given items

FULL OUTPUT

```
. alpha weight length turn displacement, std item
Test scale = mean(standardized items)

                                                              average
                          item-test     item-rest       interitem
Item          |  Obs  Sign   correlation   correlation     correlation    alpha
--------------+------------------------------------------------------------------
weight        |   74    +        0.9763        0.9569          0.8254    0.9341
length        |   74    +        0.9624        0.9319          0.8430    0.9416
turn          |   74    +        0.9236        0.8645          0.8920    0.9612
displacement  |   74    +        0.9258        0.8683          0.8892    0.9601
--------------+------------------------------------------------------------------
Test scale    |                                               0.8624    0.9616
------------------------------------------------------------------------------
```

alpha = alpha if given item was not included in scale

FINAL SCALE CONSTRUCTION

```
. alpha weight length turn displacement, std generate(size)

Test scale = mean(standardized items)

Average interitem correlation:     0.8624
Number of items in the scale:          4
Scale reliability coefficient:     0.9616
```

Variables

	Name	Label
	make	Make and Model
	price	Price
	mpg	Mileage (mpg)
	rep78	Repair Record 1978
	headroom	Headroom (in.)
	trunk	Trunk space (cu. ft.)
	weight	Weight (lbs.)
	length	Length (in.)
	turn	Turn Circle (ft.)
	displacement	Displacement (cu. in.)
	gear_ratio	Gear Ratio
	foreign	Car type
	size	mean(standardized...

6 AUTOMATION AND REPLICATION

DEFINING SETS OF VARIABLES

```
. local ctrlSize “weight length displacement”

. local ctrlEfficient “mpg gear_ratio”

. local ctrlCon “price weight length displacement mpg”

. local ctrlAll “price weight length displacement mpg gear_ratio”
```

LOOPS FOR DATA ANALYSIS

```
. local ctrlCon “price weight length displacement mpg”


. foreach var in `ctrlAll’ {
    ttest `ctrlCon’, by(foreign)
}
```

IMPORTANT CONSIDERATIONS

BASELINE STATISTICS

DOCUMENT PROCESS AND RESULTS

VERSION CONTROL
STATA & GITHUB

KEEP FULL REPOSITORIES

AUTOMATE AS MUCH AS POSSIBLE

DOCUMENT DETAILS

Document produced by Christopher Prener, Ph.D for the Saint Louis University course SOC 5050: QUANTITATIVE ANALYSIS - APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository README.md file for additional details.