Luke Beebe
Assignment 4

The following pictures are the regression coefficients from the most accurate models predicting MPG between USA, Germany and Japan.

```
> leaps.lars(both=T,xmat=Auto.mat.usa2nd,yvec=Auto.mat.usa[,1],ncheck=5,int=F)
[1] 1
[1] "Press= 1048.0470934059"
[1] "MPSE= 4.53700040435454"
[1] "Cp= 7.78048836274499"
             Intercept            displacement                  weight
          2.344534e+02           -4.656436e-01            3.314722e-02
          acceleration                    year         cylindersweight
         -6.841730e+00           -4.514745e+00            5.360273e-04
 cylindersacceleration          cylindersyear displacementdisplacement
          2.319231e-01           -6.928483e-02            1.228204e-04
      displacementyear  horsepoweracceleration              weightyear
          5.278304e-03           -3.096444e-03           -5.351085e-04
      accelerationyear                yearyear
          7.166771e-02            3.237503e-02
[1] "LEAPS CORRELATION 0.955326000142054"
> leaps.lars(both=T,xmat=Auto.mat.germany2nd,yvec=Auto.mat.germany[,1],ncheck=5,int=F)
[1] 1
[1] "Press= 528.995880853274"
[1] "MPSE= 9.28062948865393"
[1] "Cp= -0.213330316766488"
      Intercept       horsepower           weight             year  cylindersweight
   3.850112e+02     2.364405e+00    -7.229320e-02    -1.017014e+01     8.212186e-03
  cylindersyear displacementyear   horsepoweryear     weightweight       weightyear
  -3.168925e-01    -1.053403e-03    -3.274818e-02    -6.778729e-06     9.297024e-04
       yearyear
   8.568144e-02
[1] "LEAPS CORRELATION 0.932441018514054"
> leaps.lars(both=T,xmat=Auto.mat.japan2nd,yvec=Auto.mat.japan[,1],ncheck=5,int=F)
[1] 1
[1] "Press= 751.265259710616"
[1] "MPSE= 10.434239718203"
[1] "Cp= 5.23147464526383"
               Intercept               cylinders      cylinderscylinders
           -1.142973e+02            5.421060e+01           -5.529733e+00
displacementacceleration  horsepoweracceleration          horsepoweryear
           -1.099489e-02           -2.665846e-02            4.829747e-03
        accelerationyear
            3.137588e-02
[1] "LEAPS CORRELATION 0.882593532687259"
```

Now, here are the coefficients used for the whole model disregarding origins.

```
> leaps.lars(both=T,xmat=Auto.mat2nd,yvec=Auto.mat[,1],ncheck=5,int=F)
[1] 1
[1] "Press= 2965.80019718738"
[1] "MPSE= 7.82533033558675"
[1] "Cp= 5.48237932178478"
              Intercept            displacement            acceleration
           3.672205e+02           -1.671374e-01           -6.112244e+00
                   year      cylinderscylinders         cylindersweight
          -7.848981e+00           -3.463469e-01            1.436651e-03
 displacementhorsepower        displacementyear          horsepoweryear
           2.187828e-04            1.652216e-03           -1.542533e-03
             weightyear acceleratioacceleration        accelerationyear
          -1.596952e-04            4.104226e-02            6.029007e-02
               yearyear
           5.241731e-02
[1] "LEAPS CORRELATION 0.94044225990666"
```

Upon looking at the beta coefficients, I see that *year* is negatively correlated with MPG, but *yearyear* is positively correlated with MPG. *yearyear* is year*year, which means it's year^2, so the values multiplying with the smaller beta coefficient end up being larger -- meaning year is actually positively correlated with MPG.

Because our data is sorted with combined data, i.e. *displacementyear* (2nd order matrix)…, it is hard to tell from a glance which independent variables have more bearing on the model than others when a lot of the beta coefficients work in inverse ways like *yearyear* and *year*.

So, I re-inputted the matrix as a 1st order matrix to parse out independent variables' influence on each model. These were my results for USA, Germany, Japan:

```
   cylinders displacement   horsepower       weight acceleration         year
-0.997485969 -0.005917714 -0.010321260 -0.003423843 -0.447890391  0.625054647
[1] "LARS CORRELATION 0.922642771751773"
> leaps.lars(both=T,xmat=Auto.mat.germany[,-c(1,8)],yvec=Auto.mat.germany[,1],ncheck=5,int=F)
[1] 1
[1] "Press= 934.841417724904"
[1] "MPSE= 14.6068971519516"
[1] "Cp= 3.44682356342066"
   Intercept       weight acceleration         year
-44.53576635  -0.00889359   0.65862449   1.09307398
[1] "LEAPS CORRELATION 0.848025506034959"
```

```
> leaps.lars(both=T,xmat=Auto.mat.japan[,-c(1,8)],yvec=Auto.mat.japan[,1],ncheck=5,int=F)
[1] 1
[1] "Press= 1020.39165512687"
[1] "MPSE= 13.7890764206334"
[1] "Cp= 3.6141142306042"
   Intercept displacement       weight acceleration         year
-25.51321316   0.06138308  -0.01366633   0.36850321   0.95626150
```

The models are less accurate, but it is clearer which variable has more impact. Throughout the data, it seems Year has the strongest positive correlation with MPG, while the other variables are mostly negatively correlated.

USA: Year is positively correlated; Cylinders, displacement, horsepower, weight, and acceleration are negatively correlated.

Germany: Year and acceleration are positively correlated; Weight is negatively correlated (weight has the largest values from 3,000-4,000).

Japan: Year, acceleration and displacement are positively correlated; Weight is negatively correlated (Again, weight has the largest values in the dataset ranging from 3,000 to 4,000)

It makes sense that Year would be positively correlated with MPG and weight be negatively correlated with MPG across all data.

It's interesting that the USA model, which was most accurate, takes cylinders and horsepower, whereas Japan and Germany don't. Because of the negative correlation, it makes me think of the muscle cars America has made famous-mustangs, stingrays, etc.-medium bodied gas guzzlers.

The German model seems quite basic in terms of inputs compared to the other models.

Japan is interesting because it has a positive correlation with displacement, which makes me wonder if they had a different method of engine displacement.