```
> library(faraway)
> if (FALSE)
+ {"
+ data on 38 driver measurements regarding seat positioning.
+ age(yr), weight(lb), height(cm), height in shoes(cm), seated height,
+ arm length, thigh length, lower leg length, hipcenter - horizontal distance of
the midpoint
+ of the hips from a fixed location in the car in millimeters
+ "}
>
> data(seatpos, package="faraway")
> head(seatpos)
  Age Weight HtShoes    Ht Seated  Arm Thigh  Leg hipcenter
1  46    180   187.2 184.9   95.2 36.1  45.3 41.3  -206.300
2  31    175   167.5 165.5   83.8 32.9  36.5 35.9  -178.210
3  23    100   153.6 152.2   82.9 26.0  36.6 31.0   -71.673
4  19    185   190.3 187.4   97.3 37.4  44.1 41.0  -257.720
5  23    159   178.0 174.1   93.9 29.5  40.1 36.9  -173.230
6  47    170   178.7 177.0   92.4 36.0  43.2 37.4  -185.150
> nrow(seatpos)
[1] 38
> seatpos
   Age Weight HtShoes    Ht Seated  Arm Thigh  Leg hipcenter
1   46    180   187.2 184.9   95.2 36.1  45.3 41.3  -206.300
2   31    175   167.5 165.5   83.8 32.9  36.5 35.9  -178.210
3   23    100   153.6 152.2   82.9 26.0  36.6 31.0   -71.673
4   19    185   190.3 187.4   97.3 37.4  44.1 41.0  -257.720
5   23    159   178.0 174.1   93.9 29.5  40.1 36.9  -173.230
6   47    170   178.7 177.0   92.4 36.0  43.2 37.4  -185.150
7   30    137   165.7 164.6   87.7 32.5  35.6 36.2  -164.750
8   28    192   185.3 182.7   96.9 35.8  39.9 43.1  -270.920
9   23    150   167.6 165.0   91.4 29.4  35.5 33.4  -151.780
10  29    120   161.2 158.7   85.2 26.6  31.0 32.8  -113.880
11  47    143   171.9 169.1   87.8 32.9  39.2 36.9  -196.150
12  41    107   155.7 152.5   82.9 29.6  32.7 31.1  -125.550
13  51    227   179.8 177.2   91.7 31.1  41.4 40.2  -203.610
14  30    147   164.9 162.7   88.0 27.7  33.6 33.8  -163.220
15  22    178   177.2 176.4   94.1 31.1  41.0 36.6  -204.110
16  67    166   177.1 175.3   89.4 36.7  40.1 39.2  -186.800
17  25    153   173.4 171.2   85.0 33.1  45.2 38.4  -228.350
18  65    113   162.6 158.7   85.2 31.1  35.7 32.5  -103.850
19  22    142   167.3 164.6   90.4 29.5  36.5 34.0  -105.690
20  21    130   172.5 170.5   89.7 29.9  35.8 35.6  -137.360
21  20    145   168.4 166.3   87.9 30.3  34.6 38.5  -133.080
22  33    293   201.2 198.4  101.6 39.6  44.2 43.1  -279.150
23  24    180   187.6 185.3   92.6 34.9  39.9 41.8  -185.870
24  39    117   152.8 150.2   79.4 28.9  34.8 30.2   -30.950
25  58    150   169.2 166.4   86.2 33.0  37.9 35.7  -196.550
26  22    171   184.1 181.6   95.4 33.7  41.8 39.2  -205.610
27  21    125   165.8 163.4   85.0 31.0  36.4 35.3   -94.502
28  23    160   166.4 164.3   86.2 29.1  36.6 31.6  -125.840
29  21    157   177.0 175.5   91.4 34.4  41.6 36.4  -222.500
```

```
30  40    115   153.8 151.6   80.3 27.5  37.6 31.7  -102.200
31  59    168   155.2 153.0   84.4 34.1  35.6 34.6   -47.520
32  47    175   176.6 175.8   90.9 34.5  45.5 37.4  -183.550
33  72    186   177.7 175.0   90.1 38.3  39.7 37.7  -118.050
34  34    115   155.2 152.2   82.0 28.9  32.9 32.6  -148.670
35  19    150   172.2 169.9   89.4 34.0  39.7 38.1  -268.320
36  41    121   166.3 164.1   86.5 31.5  45.1 33.8  -117.000
37  21    154   172.0 170.4   90.0 29.5  36.8 37.5  -201.510
38  56    158   173.8 171.5   90.0 36.1  39.2 35.5  -176.450
>
> #save graph in pdf
>
pdf(file="C:/Users/jmard/OneDrive/Desktop/RegressionMethodsSpring2020/Ridge_Lasso
/collin_seatpos_Figure.pdf")
>
> #Full Model
> lmod <- lm(hipcenter ~ ., seatpos)
> summary(lmod)

Call:
lm(formula = hipcenter ~ ., data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-73.827 -22.833  -3.678  25.017  62.337

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.43213  166.57162   2.620   0.0138 *
Age           0.77572    0.57033   1.360   0.1843
Weight        0.02631    0.33097   0.080   0.9372
HtShoes      -2.69241    9.75304  -0.276   0.7845
Ht            0.60134   10.12987   0.059   0.9531
Seated        0.53375    3.76189   0.142   0.8882
Arm          -1.32807    3.90020  -0.341   0.7359
Thigh        -1.14312    2.66002  -0.430   0.6706
Leg          -6.43905    4.71386  -1.366   0.1824
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.72 on 29 degrees of freedom
Multiple R-squared:  0.6866,    Adjusted R-squared:  0.6001
F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05

> #the results show R2 is fairly high but none of the variables is significant
> #is this due to collinearity?
>
> #Look at the pairwise relationship of variables
> # Basic Scatterplot Matrix
> pairs(~hipcenter+Age+Weight+HtShoes+Ht+Seated+Arm+Thigh+Leg,data=seatpos,
+     main="Simple Scatterplot Matrix")
>
```

```
> #now look at the pairwise correlations
> round(cor(seatpos[,-9]),2)
          Age Weight HtShoes    Ht Seated  Arm Thigh   Leg
Age      1.00   0.08   -0.08 -0.09  -0.17 0.36  0.09 -0.04
Weight   0.08   1.00    0.83  0.83   0.78 0.70  0.57  0.78
HtShoes -0.08   0.83    1.00  1.00   0.93 0.75  0.72  0.91
Ht      -0.09   0.83    1.00  1.00   0.93 0.75  0.73  0.91
Seated  -0.17   0.78    0.93  0.93   1.00 0.63  0.61  0.81
Arm      0.36   0.70    0.75  0.75   0.63 1.00  0.67  0.75
Thigh    0.09   0.57    0.72  0.73   0.61 0.67  1.00  0.65
Leg     -0.04   0.78    0.91  0.91   0.81 0.75  0.65  1.00
> #there are some pairwise correlations >= .8
>
> #now look at eigenvalues  Ax=lambda*x
> #eigenvalue:  Ax=lambda*x  x is an eigenvector of matrix A and lambda is an
eigenvalue of A
> x <- model.matrix(lmod)[,-1]
> XtX=t(x) %*% x  #X'X
> XtX
            Age    Weight   HtShoes        Ht   Seated       Arm     Thigh
Leg
Age      55992.0  210188.0  229159.2  226000.2 118719.2  43858.40  51999.20
48510.70
Weight  210188.0  967776.0 1025820.6 1012219.9 531130.1 193637.70 231544.80
217994.30
HtShoes 229159.2 1025820.6 1120823.9 1105811.9 581221.9 210860.72 252912.60
237450.11
Ht      226000.2 1012219.9 1105811.9 1091019.1 573431.0 208041.23 249545.18
234278.18
Seated  118719.2  531130.1  581221.9  573431.0 301577.6 109280.44 131091.79
123081.00
Arm      43858.4  193637.7  210860.7  208041.2 109280.4  39859.14  47646.17
44713.42
Thigh    51999.2  231544.8  252912.6  249545.2 131091.8  47646.17  57336.29
53583.93
Leg      48510.7  217994.3  237450.1  234278.2 123081.0  44713.42  53583.93
50399.28
> e <- eigen(t(x) %*% x)  # eigenvalues of X'X
> e$val
[1] 3.653671e+06 2.147948e+04 9.043225e+03 2.989526e+02 1.483948e+02 8.117397e+01
5.336194e+01 7.298209e+00
> sqrt(e$val[1]/e$val)
[1]   1.00000  13.04226  20.10032 110.55123 156.91171 212.15650 261.66698
707.54911
> #some of the square roots of the largest eigenvalue to the other eigenvalues
(condition number) are large
> #indicates collinearity exists in several linear combinations of the Xs.
>
> #now look at Age regressed on the other Xs and compute VIF(Age)
> summary(lm(x[,1] ~ x[,-1]))$r.squared
[1] 0.4994823
> 1/(1-0.49948)
```

```
[1] 1.997922
>
> #now look at all the VIFs
> require(faraway)
> vif(x)  #VIFs
      Age    Weight    HtShoes        Ht    Seated       Arm     Thigh
  Leg
  1.997931  3.647030 307.429378 333.137832  8.951054  4.496368  2.762886
6.694291
>
> #measure hipcenter is difficult - see what happens if we add a random
perturbation to the size of the response
> lmod1 <- lm(hipcenter+10*rnorm(38) ~ ., seatpos)  #adds 10*standard normal to
each response
> summary(lmod)

Call:
lm(formula = hipcenter ~ ., data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-73.827 -22.833  -3.678  25.017  62.337

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.43213  166.57162   2.620   0.0138 *
Age           0.77572    0.57033   1.360   0.1843
Weight        0.02631    0.33097   0.080   0.9372
HtShoes      -2.69241    9.75304  -0.276   0.7845
Ht            0.60134   10.12987   0.059   0.9531
Seated        0.53375    3.76189   0.142   0.8882
Arm          -1.32807    3.90020  -0.341   0.7359
Thigh        -1.14312    2.66002  -0.430   0.6706
Leg          -6.43905    4.71386  -1.366   0.1824
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.72 on 29 degrees of freedom
Multiple R-squared:  0.6866,    Adjusted R-squared:  0.6001
F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05

> summary(lmod1)

Call:
lm(formula = hipcenter + 10 * rnorm(38) ~ ., data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-67.233 -23.008  -5.235  24.483  75.550

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 501.60501  177.28877   2.829  0.00838 **
Age           0.84578    0.60702   1.393  0.17411
Weight        0.09375    0.35226   0.266  0.79202
HtShoes      -3.60763   10.38054  -0.348  0.73070
Ht            1.37028   10.78163   0.127  0.89974
Seated        0.04811    4.00393   0.012  0.99050
Arm          -1.81769    4.15113  -0.438  0.66472
Thigh        -1.26797    2.83117  -0.448  0.65758
Leg          -6.05400    5.01715  -1.207  0.23731
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.15 on 29 degrees of freedom
Multiple R-squared:  0.6797,    Adjusted R-squared:  0.5913
F-statistic: 7.691 on 8 and 29 DF,  p-value: 1.742e-05

> #R2 and SE are similar in the two models

>
> #consider just the correlations of the length variables
> round(cor(x[,3:8]),2)
        HtShoes   Ht Seated  Arm Thigh  Leg
HtShoes    1.00 1.00   0.93 0.75  0.72 0.91
Ht         1.00 1.00   0.93 0.75  0.73 0.91
Seated     0.93 0.93   1.00 0.63  0.61 0.81
Arm        0.75 0.75   0.63 1.00  0.67 0.75
Thigh      0.72 0.73   0.61 0.67  1.00 0.65
Leg        0.91 0.91   0.81 0.75  0.65 1.00
> #choose only 1 of these since they are all highly correlated.  Pick Ht
>
> lmod2 <- lm(hipcenter ~ Age + Weight + Ht, seatpos)
> summary(lmod2)

Call:
lm(formula = hipcenter ~ Age + Weight + Ht, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-91.526 -23.005   2.164  24.950  53.982

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 528.297729 135.312947   3.904 0.000426 ***
Age           0.519504   0.408039   1.273 0.211593
Weight        0.004271   0.311720   0.014 0.989149
Ht           -4.211905   0.999056  -4.216 0.000174 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.49 on 34 degrees of freedom
Multiple R-squared:  0.6562,    Adjusted R-squared:  0.6258
F-statistic: 21.63 on 3 and 34 DF,  p-value: 5.125e-08
```

Height changed sign

```
> #R2 and adjusted R2 are similar to the Full Model but fewer predictors used
>
> dev.off()
null device
          1
>
```