# Solutions_Q_8

## Question 8.1

The dependent variable would be an indicator of the families' enjoyment on our summer vacation. The independent variables would be average temperature, if we could get to the location easily on a scale (driving, train, airplane), how much money we had saved during the year indicating money would could spend, if other family members could come along (which could be negative for some and positive for others!)

## Question 8.2

For a model with all variables...

The Multiple R-squared: is 0.8031 which indicates is RSS/TSS meaning about 80% of the variance is explained by regression overall variables.

I did not scale the data but did remove terms that had a Pr(>|t|) over 0.05 iteratively as they violated the null hypothesis that the particular coefficient was not zero.

Removing variables still leaves us with 76% R-squared. The final regression was

pred<-lm(Crime~M+Ed+Po1+U2+Ineq+Prob,data=dat)

```
dat=read.csv("~/o.csv", header=TRUE)
pred<-lm(Crime~M+So+Ed+Po1+Po2+LF+M.F+Pop+NW+U1+U2+Wealth+Ineq+Prob+Time,data=dat)

pred<-lm(Crime~M+Ed+Po1+Po2+LF+M.F+Pop+NW+U1+U2+Wealth+Ineq+Prob+Time,data=dat)
pred<-lm(Crime~M+Ed+Po1+Po2+LF+M.F+Pop+NW+U1+U2+Wealth+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+Po2+M.F+Pop+NW+U1+U2+Wealth+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+Po2+M.F+Pop+U1+U2+Wealth+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+M.F+Pop+U1+U2+Wealth+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+M.F+U1+U2+Wealth+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+M.F+U1+U2+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+U1+U2+Ineq+Prob,data=dat)
pred<-lm(Crime~M+Ed+Po1+U2+Ineq+Prob,data=dat)

summary(pred)
```

```
##
## Call:
## lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.68  -78.41  -19.68  133.12  556.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
## M             105.02      33.30   3.154  0.00305 **
## Ed            196.47      44.75   4.390 8.07e-05 ***
## Po1           115.02      13.75   8.363 2.56e-10 ***
## U2             89.37      40.91   2.185  0.03483 *
## Ineq           67.65      13.94   4.855 1.88e-05 ***
```

```
## Prob        -3801.84    1528.10  -2.488  0.01711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 200.7 on 40 degrees of freedom
## Multiple R-squared:  0.7659, Adjusted R-squared:  0.7307
## F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

Doing PCA on the dataset

```
dat=read.csv("~/o.csv", header=TRUE)
pca_mine=prcomp(dat)
summary(pca_mine)
```

```
## Importance of components:
##                             PC1      PC2      PC3     PC4     PC5   PC6
## Standard deviation     982.1164 341.1338 35.46689 7.41067 6.05210 2.266
## Proportion of Variance   0.8912   0.1075  0.00116 0.00005 0.00003 0.000
## Cumulative Proportion    0.8912   0.9987  0.99990 0.99996 0.99999 1.000
##                            PC7   PC8    PC9   PC10   PC11   PC12   PC13
## Standard deviation     1.804 1.319 0.9088 0.7297 0.4199 0.2456 0.2059
## Proportion of Variance 0.000 0.000 0.0000 0.0000 0.0000 0.0000 0.0000
## Cumulative Proportion  1.000 1.000 1.0000 1.0000 1.0000 1.0000 1.0000
##                           PC14    PC15     PC16
## Standard deviation     0.02361 0.01279 0.006995
## Proportion of Variance 0.00000 0.00000 0.000000
## Cumulative Proportion  1.00000 1.00000 1.000000
```

Indicates that almost 90% of the variance is in the first two components (0.9987).

## Not PART OF 8.2 (Question for me to explore)

## Two great examples for the reconstruction

## https://stats.stackexchange.com/questions/229092/how-to-reverse-pca-and-r

## https://stats.stackexchange.com/questions/57467/how-to-perform-dimensior 57478#57478

Some of the variables that were pulled out do to the null hypothesis on the coefficient being zero are present and large in PC1

```
nComp = 2
mu=colMeans(dat)
X = pca_mine$x[,1:nComp] %*% t(pca_mine$rotation[,1:nComp])
X = scale(X, center = -mu, scale = FALSE)
h=X[1,]
```