Homework 2 – Question 1

1) Imagine that someone wanted to come up with a total score to summarize each person's view of the oil crisis (Q1-Q20).

a) Explain why it doesn't make sense to just add up all of the numbers.

b) Find the correlation matrix for Q1-Q20 data set and suggest two separate groups of questions that might be added separately.

c) How could these two scores be combined to form a single score?
Homework 2 – Question 2

2) Check whether the data set normsamp.txt is actually multivariate normal.
Adding a Third Variable

```
symbols(birth, heartd, circles=over65, inches=.1)
```
Adding a Third Variable

\begin{verbatim}
library(lattice)
cloud(over65~birth*heartd)
\end{verbatim}

Adding a Third Variable

\begin{verbatim}
coplot(heartd~birth|over65)
\end{verbatim}

Lots of Variables!

\begin{verbatim}
pairs(census[,3:8])
\end{verbatim}
Lots of Variables!

```r
source("http://www.stat.sc.edu/~habing/courses/530faceF03.txt")
faces(census[1:16,3:8],
   substring(census[1:36,1],1,10))
```

Principal Components Analysis

The main idea of Principal Components Analysis is that we would like to come up with new combinations of the original variables that are "easier to work with".

3-D Example

```r
PROC IML;
sigma = {1 .2 2,
         .2 1 2,
         2 2 10};
mu = {0 0 0};
n = 1000;
seed = 91505;
q=NROW(sigma);
MUMAT=REPEAT(mu,n,1);
SROOT=ROOT(sigma);
Z=NORMAL(REPEAT(seed,n,q));
x=Z*SROOT+MUMAT;
CREATE mvnormdata FROM x;
APPEND FROM x;
QUIT;
```
Goal
Find the coefficients (a’s) of the x’s so that:

\[ Y = a_1X_1 + a_2X_2 + \cdots + a_qX_q \]

has the largest possible variance subject to the condition that the length of the coefficient vector is 1.

Example Cont.

```r
library(MASS)
mu<-c(5,0,-1)
sigma<- matrix(c(1,0.2,2,0.2,1,2,2,2,10),
ncol=3,byrow=T)
x<-mvrnorm(n=1000,mu,sigma)
coef<-princomp(x,cor=F)$loadings[,1]
pcl<-princomp(x,cor=F)$scores[,1]
```

Oildata Q’s

```r
coeff<-princomp(sect3,cor=F)$loadings[,1]

Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20
0.14749703 -0.18642164 -0.24432644 0.09980284
0.35633250 -0.17468260 -0.04253903 0.01630308
0.25808813 0.29008088 0.0490861 0.1185841
-0.24074574 0.07216813 0.38062551 0.16292494
-0.23470978 -0.23005868 -0.19195023 -0.11866190
```