```
setwd("E:/workingdirectory/phc6053")

##############################
# USE THE SAS7BDAT PACKAGE TO READ IN THE DATA
##############################
library(sas7bdat)
library(foreign)
df <- read.ssd("E:/workingdirectory/phc6053", "fghm60",
               sascmd="C:/Program Files/SASHome93/SASFoundation/9.3/sas.exe")



##############################
# MAKE SURE WE'RE LOOKING AT THE RIGHT STUFF
##############################
is.data.frame(df)
colnames(df)
summary(df)


##############################
# CREATE THE LOG OF SYSBP
##############################
df$LNSBP <- log(df$SYSBP)


##############################
# CREATE THE CATEGORIZED BMI VARIABLE
##############################
df$BMIGROUPS <- factor(ifelse(df$BMI < 18.5,
                      1,
                      ifelse(df$BMI >= 18.5 &
                                    df$BMI < 25,
                             2,
                             ifelse(df$BMI >= 25 &
                                           df$BMI < 30,
                                    3,
                                    ifelse(df$BMI >= 30,
                                           4,
                                           NA)))),
                      labels=c("Underweight",
                                    "Normal",
                                    "Overweight",
                                    "Obese"))
summary(df$BMIGROUPS)

##############################
# FIX THE SEX COLUMN
##############################
df$SEX <- factor(df$SEX, labels=c("Male", "Female"))

##############################
# CREATE A MALES AND FEMALES DATAFRAME
##############################
males <- df[which(df$SEX == "Male"),]
females <- df[which(df$SEX == "Female"),]

##############################
# NUMERIC SUMMARY OF SYSBP WITHIN EACH SEX
##############################
summary(males$SYSBP)
sd(males$SYSBP)
```

```
summary(females$SYSBP)
sd(females$SYSBP)

##############################
# TWO SAMPLE T-TEST TO COMPARE MEAN OF SYSBP BETWEEN TWO LEVELS OF SEX
# GET P-VALUE FOR TEST FOR EQUALITY OF VARIANCE
# P-VALUE FOR APPROPRIATE T-TEST
# APPROPRIATE 95% CONFIDENCE INTERVAL FOR DIFFERENT BETWEEN MEAN SYSBP
# CONCLUSION
##############################
t.test(df$SYSBP ~ df$SEX)
t.test(df$SYSBP ~ df$SEX, var.equal=TRUE)


##############################
# PROVIDE NUMERIC SUMMARY OF SYSBP WITHIN EACH LEVEL OF BMIGROUPS
# MEAN
# MEDIAN
# SD
##############################
Underweight <- df[which(df$BMIGROUPS == "Underweight"),]
Normal <- df[which(df$BMIGROUPS == "Normal"),]
Overweight <- df[which(df$BMIGROUPS == "Overweight"),]
Obese <- df[which(df$BMIGROUPS == "Obese"),]

partC <- function(x){
  print(c(nrow(x), mean(x$SYSBP), median(x$SYSBP), sd(x$SYSBP)))}
partC(Underweight)
partC(Normal)
partC(Overweight)
partC(Obese)

##############################
# CONDUCT AN ANOVA TO COMPARE MEAN SYSBP BETWEEN FOUR LEVELS OF
# BMIGROUPS
### PROVIDE VALUES IN BASIC ANOVA TABLE
##############################
aov.out <- aov(df$SYSBP ~ df$BMIGROUPS)
summary(aov.out)
aov.out
TukeyHSD(aov.out)

##############################
# PROVIDE SIDE BY SIDE BOXPLOTS OF SYSBP WITHIN EACH SEX
##############################



##############################
# PROVIDE SIDE BY SIDE BOXPLOTS OF SYSBP WITHIN EACH LEVEL
# OF BMIGROUPS
##############################


##############################
# SCATTERPLOT Y=SYSBP X=BMI, with LOESS CURVE
##############################


##############################
# BOXPLOTS
```

```
############################
boxplot(df$SYSBP~df$BMIGROUPS)
boxplot(df$SYSBP~df$SEX)



save.image("E:/workingdirectory/phc6053/ass2.RData")
```