> evals <- read.csv("C:/Users/jsunez/Desktop/DSS 665/Week 3/evals.csv",header=T)

> attach(evals)

The following object is masked from juul:

age

The following objects are masked from evals (pos = 4):

age, bty\_avg, cls\_did\_eval, cls\_level, cls\_perc\_eval, cls\_profs, cls\_students, Course, ethnicity,

gender, language, pic\_color, pic\_outfit, rank, score

> #Problem1

> print("Summary of Variable 'score':"); summary(score);

[1] "Summary of Variable 'score':"

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.300 3.800 4.300 4.175 4.600 5.000

> print(paste("Standard Deviation:", sd(score), sep = " "))

[1] "Standard Deviation: 0.543864521209176"

> hist(score, xlab = "Eval Score", ylab = "Frequency", col = "green", main = "Histogram of Evaluation Scores")

> #Problem2

> plot(score, bty\_avg, pch=17, xlim=c(0,5), ylim=c(0,10), col="blue", main = "Scores vs Beauty Plot", xlab = "Score", ylab = "Beauty")

> library(ggplot2)

> dat <- data.frame(score, bty\_avg)

> ggplot(dat, aes(x=score, y=bty\_avg)) + geom\_point(color="red") + geom\_smooth(method = lm) +

+ labs(x="Score", y="Beauty Average", title="Beauty Average vs. Score")

> #Problem3

> gndr <- table(gender)

> gndr

gender

female male

195 268

> barplot(gndr, main="Gender", xlab="Gender", col = c("pink", "blue"))

> #Problem4

> boxplot(score~gender, main="Boxplot of Score by Gender", col=c("pink", "blue"))

> par(mfrow=c(2,1))

> hist(score[gender=="male"], xlab = "Male", col = "blue", main = "Histogram of Male Scores")

> hist(score[gender=="female"], xlab = "Female", col = "pink", main = "Histogram of Female Scores")

> t.test(score[gender=="female"], score[gender=="male"])

Welch Two Sample t-test

data: score[gender == "female"] and score[gender == "male"]

t = -2.7507, df = 398.7, p-value = 0.006218

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.24264375 -0.04037194

sample estimates:

mean of x mean of y

4.092821 4.234328

> t.test(score~gender, data=evals)

Welch Two Sample t-test

data: score by gender

t = -2.7507, df = 398.7, p-value = 0.006218

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.24264375 -0.04037194

sample estimates:

mean in group female mean in group male

4.092821 4.234328

> #Problem5

> rank.levels <- table(rank, cls\_level)

> par(mfrow=c(1,1))

> barplot(prop.table(t(rank.levels),2),beside = T, legend.text = colnames(rank.levels), col=c(2:3), main = "Professor Ranking and Course Level")