```r
#
# Problem 3
#

# Download data if it doesn't exist
data <- function() {
    if (!file.exists('./College.csv')) {
        download.file('http://www-bcf.usc.edu/~gareth/ISL/College.csv', destfile =
'./College.csv')
    }
}

#
# Part A:
# Read in the data from the CSV file.
#

# Read in Data
college <- read.csv('./College.csv', header = TRUE)

#
# Part B:
# Look at the data using various functions.
#

# View/edit the data
# fix(college)

# View the data
# View(college)

# Remove first column according to page 55
college <- college[,-1]

#
# Part C:
#

#
# Part I:
# Show a summary of the college data.
#

partI <- function() {
    college.summary <- capture.output(summary(college))
    cat(college.summary,
        file = 'partI.txt',
        sep = '\n')
```

```r
}

#
# Part II:
# Show a scatterplot of the first 10 columns of college.
#

partII <- function() {
    png('partII.png')
    pairs(college[,1:10])
    dev.off()
}

#
# Part III:
# Produce side-by-side boxplots of Outstate vs Private.
#

partIII <- function() {
    pdf('partIII.pdf')
    plot(college$Private, college$Outstate,
         main = 'Out of State Tuition vs Private Colleges',
         xlab = 'Private College',
         ylab = 'Out of State Tuition')
    dev.off()
}

#
# Create a new qualitative variable for Elite colleges. Show various statistics
# for the Elite colleges.
#

partIV <- function() {
    pdf('partIV.pdf')
    Elite <- rep('No', nrow(college))
    Elite[college$Top10perc > 50] <- 'Yes'
    Elite <- as.factor(Elite)
    college <- data.frame(college, Elite)

    # Show number of elite vs non-elite colleges
    summary(college)

    # Show boxplot for Outstate vs Elite
    plot(college$Elite, college$Outstate,
         main = 'Out of State Tuition vs Elite Colleges',
         xlab = 'Elite College',
         ylab = 'Out of State Tuition')
    dev.off()
```

```
}

#
# Part V:
# Show some histograms with differing numbers of bins for a few quantitative
# variables.
#

partV  <- function() {
    pdf('partV.pdf')
    par(mfrow=c(2,2))
    hist(college$Apps, 20,
         main = 'Number of College Applications Recieved',
         xlab = 'Number of Applications Received')
    hist(college$Accept, 10,
         main = 'Number of Applicants Accepted',
         xlab = 'Number of Applicants Accepted')
    hist(college$S.F.Ratio, 10,
         main = 'Student to Faculty Ratio',
         xlab = 'Student to Faculty Ratio')
    hist(college$PhD, 10,
         main = 'Percent of Faculty with a PhD',
         xlab = 'Percent of Faculty with a PhD')
    dev.off()
}

#
# Part VI:
# Continue exploring the data and report what you find.
#

partVI <- function() {
    pdf('partVI.pdf')

    par(mfrow=c(2,1))
    df <- data.frame(x = college$PhD)
    df$y <- college$Books

    T <- lm(y~x,data=df)

    plot(college$PhD, college$Books,
         main = 'Book Costs vs Percent of PhD Professors',
         xlab = 'Percent of PhD Professors',
         ylab = 'Estimated Book Cost')
    abline(T)

    df <- data.frame(x = college$Accept)
    df$y <- college$Top10perc
```

```
    T <- lm(y~x,data=df)

    plot(college$Accept, college$Top10perc,
         main = 'Number of Applications Accepted vs Top 10% New Students',
         xlab = 'Applications Accepted',
         ylab = 'Top 10% New Students')
    abline(T)

    dev.off()
}

# Runs all of the parts of the homework
run <- function() {
    data()
    partI()
    partII()
    partIII()
    partIV()
    partV()
    partVI()
}
```