```r
setwd("C:/Users/Min Seong Kim/Dropbox/R_programming/lecture")

# College data: Demographic characteristics, tuition, and more for USA colleges.

# Private: Public/private indicator
# Apps: Number of applications received
# Accept: Number of applicants accepted
# Enroll: Number of new students enrolled
# Top10perc: New students from top 10 % of high school class
# Top25perc: New students from top 25 % of high school class
# F.Undergrad: Number of full-time undergraduates
# P.Undergrad: Number of part-time undergraduates
# Outstate: Out-of-state tuition
# Room.Board: Room and board costs
# Books: Estimated book costs
# Personal: Estimated personal spending
# PhD: Percent of faculty with Ph.D.'s
# Terminal: Percent of faculty with terminal degree
# S.F.Ratio: Student/faculty ratio
# perc.alumni: Percent of alumni who donate
# Expend: Instructional expenditure per student
# Grad.Rate: Graduation rate

college <- read.csv("College.csv")
dim(college)
head(college)
college.2 <- college[,c(-1,-7,-8,-9,-10,-11,-12,-14,-15)]

head(college.2)

reg <- lm(perc.alumni ~ Private + Top10perc, data=college.2)
summary(reg)

# Beyond linear models: Logistic regression modeling
# We can use glm() to fit a logistic regression model. The arguments are very similar to lm()
# The first argument is a formula, of the form Y ~ X1 + X2 + ... + Xp, where Y is the response
# and X1,..., Xp are predictors.
# We must also specify family="binomial" to get logistic regression

table(college.2$Private)

# You may want to see which variables are associated with Private.
par(mfrow=c(3,3), mar=c(4,4,2,0.5))

for (j in c(2:10)) {
  plot(college.2[,1],college.2[,j], xlab="Private",
       main=paste("Boxplot of", colnames(college.2)[j]),
       col="lightblue")
}

logistic <- glm((Private=="Yes") ~ S.F.Ratio + perc.alumni, data=college.2, family="binomial")
summary(logistic)

# Utility functions work as before
# To call coefficients, fitted values, residuals, summarizing, plotting, making predictions,
# the utility functions coef(), fitted(),  residuals(), summary(), plot(), predict() work pretty much just as with lm().

coef(logistic)

p.hat <- fitted(logistic)
y.hat <- round(p.hat)
table(y.hat, y.true=college.2$Private)

# Exercise
```

```
65
66  # You can creat a dummy variable from a continuous variable. For example,
67  college.1$dummy_pc <- as.numeric(college.1$perc.alumni > mean(college.1$perc.alumni))
68
69  # Find the variables that provide good prediction for this dummy varaible.
70
```