```
library(XLS)
library(xlsx)
library(readxl)

# EXERCISE 1
multi <- read.csv('dataset_multipleRegression.csv')
multi_data <- data.frame(multi)
View(multi_data)

# Using the unemployment rate (UNEM) and number of spring high school graduates (HGRAD),
# predict the fall enrollment (ROLL) for this year by knowing that UNEM=7% and
HGRAD=90,000.
UNEM <- multi_data$UNEM
HGRAD <- multi_data$HGRAD
ROLL <- multi_data$ROLL

frame <- data.frame(UNEM, HGRAD, ROLL)

# Create a linear model
linear_model <- lm(ROLL~UNEM+HGRAD, data = frame)

# Creating a data frame
newFrame <- data.frame(UNEM = c(7), HGRAD = c(90000))

prediction <- predict(linear_model, newdata = newFrame)
prediction

# Repeat and add per capita income (INC) to the model. Predict ROLL if INC=$25,000
INC <- multi_data$INC
frame <- data.frame(UNEM, HGRAD, INC, ROLL)
linear_model <- lm(ROLL~UNEM+HGRAD+INC, data=frame)
newFrame <- data.frame(UNEM = c(7), HGRAD = c(90000), INC = c(25000))
prediction <- predict(linear_model, newdata = newFrame)
prediction

# EXERCISE 2
abalone <- read.csv('abalone.csv')
# Column names
colnames(abalone) <- c("sex", "length", 'diameter', 'height', 'whole_weight', 'shucked_wieght',
'viscera_wieght', 'shell_weight',
                       'rings' )
# summary on abalone
summary(abalone)
# structure of the abalone data
```

```
str(abalone)
# summary of the abalone rings column
summary(abalone$rings)

abalone$rings <- as.numeric(abalone$rings)
abalone$rings <- cut(abalone$rings, br=c(-1,8,11,35), labels = c("young", 'adult', 'old'))
abalone$rings <- as.factor(abalone$rings)
summary(abalone$rings)

# remove the "sex" variable in abalone, because KNN requires all numeric variables for
prediction
# z <- abalone
aba <- abalone
aba$sex <- NULL


# normalize the data using min max normalization
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
aba[1:7] <- as.data.frame(lapply(aba[1:7], normalize))
summary(aba$shucked_wieght)

ind <- sample(2, nrow(aba), replace=TRUE, prob=c(0.7, 0.3))
KNNtrain <- aba[ind==1,]
KNNtest <- aba[ind==2,]
sqrt(2918)
library(class)

KNNpred <- knn(train = KNNtrain[1:7], test = KNNtest[1:7], cl = KNNtrain$rings, k = 55)
KNNpred
table(KNNpred)

# Exercise 3

library(ggplot2) # we will use ggplot2 to visualize the data.

summary(iris)

sapply(iris[,-5], var)

set.seed(300)
k.max <- 12
```

```
wss<- sapply(1:k.max,function(k){kmeans(iris[,3:4],k,nstart = 20,iter.max = 1000)$tot.withinss})
wss # within sum of squares.

icluster <- kmeans(iris[,3:4],3,nstart = 20)

table(iris[,5],icluster$cluster) # We see that only 6 out of the 150 are classified incorrectly!
```