Homework 2

Mitch Oshima

January 20, 2016

```
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
sp_ids = unique(iris$Species)

output = matrix(0, nrow=length(sp_ids), ncol=ncol(iris)-1)
rownames(output) = sp_ids
colnames(output) = names(iris[ , -ncol(iris)])

for(i in seq_along(sp_ids)) {
    iris_sp = subset(iris, subset=Species == sp_ids[i], select=-Species)
    for(j in 1:(ncol(iris_sp))) {
        x = 0
        y = 0
        if (nrow(iris_sp) > 0) {
            for(k in 1:nrow(iris_sp)) {
                x = x + iris_sp[k, j]
                y = y + 1
            }
            output[i, j] = x / y
        }
    }
}
output
```

```
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa            5.006       3.428        1.462       0.246
## versicolor        5.936       2.770        4.260       1.326
## virginica         6.588       2.974        5.552       2.026
```

1.) The output object is a 3x4 matrix where setosa, veriscolor, and virginiaca are the rows and sepal.Length, sepal.width, petal.length, and petal.width are columns. The values that are stored in the matrix are the average values for each species.

2.) Before the loop is initialized the list sp_ids is created to house the unique species types contained in the matrix iris. The object output is also created which is a matrix that uses sp_ids as the rows and the other colums from iris as the columns. The for loop is described in the pseudocode below.

Loop through the list sp_ids subset iris when the species in sp_ids equals Species and select all the data except for the Species column. Store this data in iris_sp.

```
Loop through each column of iris_sp
    and set X and Y equal to 0.

    If the number of rows in iris_sp is greater than 0 then,
        Loop through the rows of iris_sp
            add each value of each row for each species and store that value                in X while

            Divide X by Y to get the average for of the column for each                     species ar
```

3.) I would rename "output" as "means" or "averages" so that the user knows what the values in the matrix represents. I would rename "X" as "sum" or "total" and "Y" as "count" so that it is more clear that X is the sum of each variable and Y is the number of observations that are being summed.

```
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
sp_ids = unique(iris$Species)

averages = matrix(0, nrow=length(sp_ids), ncol=ncol(iris)-1)
rownames(averages) = sp_ids
colnames(averages) = names(iris[ , -ncol(iris)])

for(i in seq_along(sp_ids)) {
    iris_sp = subset(iris, subset=Species == sp_ids[i], select=-Species)
    for(j in 1:(ncol(iris_sp))) {
        total = 0
        count = 0
        if (nrow(iris_sp) > 0) {
            for(k in 1:nrow(iris_sp)) {
                total = total + iris_sp[k, j]
                count = count + 1
            }
            averages[i, j] = total / count
        }
    }
}
averages
```

```
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa            5.006       3.428        1.462       0.246
## versicolor        5.936       2.770        4.260       1.326
## virginica         6.588       2.974        5.552       2.026
```

4.)

```r
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```r
sp_ids = unique(iris$Species)

averages = matrix(0, nrow=length(sp_ids), ncol=ncol(iris)-1)
rownames(averages) = sp_ids
colnames(averages) = names(iris[ , -ncol(iris)])

for(i in seq_along(sp_ids)) {
  iris_sp = subset(iris, subset=Species == sp_ids[i], select=-Species)
  for(j in ncol(iris_sp)) {
    x = mean(iris_sp[,j])
  }
  averages[i,j] = x
}
averages
```

```
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa                0           0            0       0.246
## versicolor            0           0            0       1.326
## virginica             0           0            0       2.026
```

5.)

```r
x = c(1,2,3,4,5,6,7,8,9,10)
y = c()
for( i in x) {
  tot = sum(x[1:i])
  y = c(y, tot)
}
y
```

```
##  [1]  1  3  6 10 15 21 28 36 45 55
```

6.)

```r
x = c(1,2,3,4,5,6,7,8,9,10)
y = c()
for( i in x) {
  tot = sum(x[1:i])
  y = c(y, tot)
  if(y[i]>10) {
```

```
    y[i]="NA"
  }
}
y
```

```
##  [1] "1"  "3"  "6"  "10" "NA" "NA" "NA" "NA" "NA" "NA"
```

7.)

```
myFunc = function(Vector){
  y = c()
  for( i in length(Vector)) {
    tot = sum(Vector[1:i])
    y = c(y, tot)
    if(y[i]>10) {
      y[i]="na"
    }
  }
  y
}
```