HW 4.2

Run Packages and Import Data

```
library("cluster")
library("knitr")

# built in kmeans function
# function (x, centers, iter.max = 10L, nstart = 1L, algorithm = c("Hartigan-Wong",
# "Lloyd", "Forgy", "MacQueen"), trace = FALSE)

data <- read.table("iris.txt")

# remove species label (categorical RESPONSE variable, should not be used to build model)
nosp_iris <- iris[, -5]
print(nosp_iris)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1            5.1         3.5          1.4         0.2
## 2            4.9         3.0          1.4         0.2
## 3            4.7         3.2          1.3         0.2
## 4            4.6         3.1          1.5         0.2
## 5            5.0         3.6          1.4         0.2
## 6            5.4         3.9          1.7         0.4
## 7            4.6         3.4          1.4         0.3
## 8            5.0         3.4          1.5         0.2
## 9            4.4         2.9          1.4         0.2
## 10           4.9         3.1          1.5         0.1
## 11           5.4         3.7          1.5         0.2
## 12           4.8         3.4          1.6         0.2
## 13           4.8         3.0          1.4         0.1
## 14           4.3         3.0          1.1         0.1
## 15           5.8         4.0          1.2         0.2
## 16           5.7         4.4          1.5         0.4
## 17           5.4         3.9          1.3         0.4
## 18           5.1         3.5          1.4         0.3
## 19           5.7         3.8          1.7         0.3
## 20           5.1         3.8          1.5         0.3
## 21           5.4         3.4          1.7         0.2
## 22           5.1         3.7          1.5         0.4
## 23           4.6         3.6          1.0         0.2
## 24           5.1         3.3          1.7         0.5
## 25           4.8         3.4          1.9         0.2
## 26           5.0         3.0          1.6         0.2
## 27           5.0         3.4          1.6         0.4
## 28           5.2         3.5          1.5         0.2
## 29           5.2         3.4          1.4         0.2
## 30           4.7         3.2          1.6         0.2
## 31           4.8         3.1          1.6         0.2
## 32           5.4         3.4          1.5         0.4
## 33           5.2         4.1          1.5         0.1
## 34           5.5         4.2          1.4         0.2
## 35           4.9         3.1          1.5         0.2
## 36           5.0         3.2          1.2         0.2
## 37           5.5         3.5          1.3         0.2
## 38           4.9         3.6          1.4         0.1
## 39           4.4         3.0          1.3         0.2
## 40           5.1         3.4          1.5         0.2
## 41           5.0         3.5          1.3         0.3
## 42           4.5         2.3          1.3         0.3
## 43           4.4         3.2          1.3         0.2
## 44           5.0         3.5          1.6         0.6
## 45           5.1         3.8          1.9         0.4
## 46           4.8         3.0          1.4         0.3
## 47           5.1         3.8          1.6         0.2
## 48           4.6         3.2          1.4         0.2
## 49           5.3         3.7          1.5         0.2
## 50           5.0         3.3          1.4         0.2
## 51           7.0         3.2          4.7         1.4
## 52           6.4         3.2          4.5         1.5
## 53           6.9         3.1          4.9         1.5
## 54           5.5         2.3          4.0         1.3
## 55           6.5         2.8          4.6         1.5
## 56           5.7         2.8          4.5         1.3
## 57           6.3         3.3          4.7         1.6
## 58           4.9         2.4          3.3         1.0
## 59           6.6         2.9          4.6         1.3
## 60           5.2         2.7          3.9         1.4
## 61           5.0         2.0          3.5         1.0
## 62           5.9         3.0          4.2         1.5
## 63           6.0         2.2          4.0         1.0
## 64           6.1         2.9          4.7         1.4
## 65           5.6         2.9          3.6         1.3
## 66           6.7         3.1          4.4         1.4
## 67           5.6         3.0          4.5         1.5
## 68           5.8         2.7          4.1         1.0
## 69           6.2         2.2          4.5         1.5
## 70           5.6         2.5          3.9         1.1
## 71           5.9         3.2          4.8         1.8
## 72           6.1         2.8          4.0         1.3
## 73           6.3         2.5          4.9         1.5
## 74           6.1         2.8          4.7         1.2
## 75           6.4         2.9          4.3         1.3
## 76           6.6         3.0          4.4         1.4
## 77           6.8         2.8          4.8         1.4
## 78           6.7         3.0          5.0         1.7
## 79           6.0         2.9          4.5         1.5
## 80           5.7         2.6          3.5         1.0
## 81           5.5         2.4          3.8         1.1
## 82           5.5         2.4          3.7         1.0
## 83           5.8         2.7          3.9         1.2
## 84           6.0         2.7          5.1         1.6
## 85           5.4         3.0          4.5         1.5
## 86           6.0         3.4          4.5         1.6
## 87           6.7         3.1          4.7         1.5
## 88           6.3         2.3          4.4         1.3
## 89           5.6         3.0          4.1         1.3
## 90           5.5         2.5          4.0         1.3
## 91           5.5         2.6          4.4         1.2
## 92           6.1         3.0          4.6         1.4
## 93           5.8         2.6          4.0         1.2
## 94           5.0         2.3          3.3         1.0
## 95           5.6         2.7          4.2         1.3
## 96           5.7         3.0          4.2         1.2
## 97           5.7         2.9          4.2         1.3
## 98           6.2         2.9          4.3         1.3
## 99           5.1         2.5          3.0         1.1
## 100          5.7         2.8          4.1         1.3
## 101          6.3         3.3          6.0         2.5
## 102          5.8         2.7          5.1         1.9
## 103          7.1         3.0          5.9         2.1
## 104          6.3         2.9          5.6         1.8
## 105          6.5         3.0          5.8         2.2
## 106          7.6         3.0          6.6         2.1
## 107          4.9         2.5          4.5         1.7
## 108          7.3         2.9          6.3         1.8
## 109          6.7         2.5          5.8         1.8
## 110          7.2         3.6          6.1         2.5
## 111          6.5         3.2          5.1         2.0
## 112          6.4         2.7          5.3         1.9
## 113          6.8         3.0          5.5         2.1
## 114          5.7         2.5          5.0         2.0
## 115          5.8         2.8          5.1         2.4
## 116          6.4         3.2          5.3         2.3
## 117          6.5         3.0          5.5         1.8
## 118          7.7         3.8          6.7         2.2
## 119          7.7         2.6          6.9         2.3
## 120          6.0         2.2          5.0         1.5
## 121          6.9         3.2          5.7         2.3
## 122          5.6         2.8          4.9         2.0
## 123          7.7         2.8          6.7         2.0
## 124          6.3         2.7          4.9         1.8
## 125          6.7         3.3          5.7         2.1
## 126          7.2         3.2          6.0         1.8
## 127          6.2         2.8          4.8         1.8
## 128          6.1         3.0          4.9         1.8
## 129          6.4         2.8          5.6         2.1
## 130          7.2         3.0          5.8         1.6
## 131          7.4         2.8          6.1         1.9
## 132          7.9         3.8          6.4         2.0
## 133          6.4         2.8          5.6         2.2
## 134          6.3         2.8          5.1         1.5
## 135          6.1         2.6          5.6         1.4
## 136          7.7         3.0          6.1         2.3
## 137          6.3         3.4          5.6         2.4
## 138          6.4         3.1          5.5         1.8
## 139          6.0         3.0          4.8         1.8
## 140          6.9         3.1          5.4         2.1
## 141          6.7         3.1          5.6         2.4
## 142          6.9         3.1          5.1         2.3
## 143          5.8         2.7          5.1         1.9
## 144          6.8         3.2          5.9         2.3
## 145          6.7         3.3          5.7         2.5
## 146          6.7         3.0          5.2         2.3
## 147          6.3         2.5          5.0         1.9
## 148          6.5         3.0          5.2         2.0
## 149          6.2         3.4          5.4         2.3
## 150          5.9         3.0          5.1         1.8
```

Set Sed and Run Model

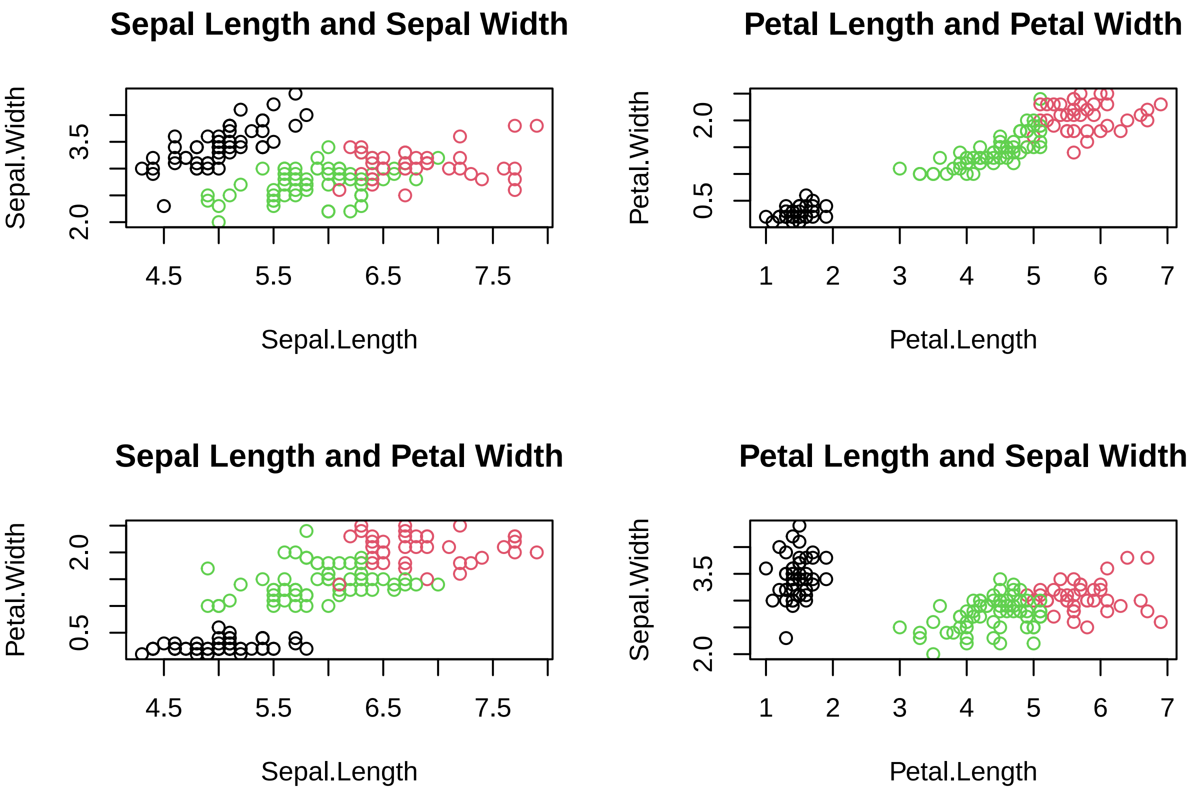
```
# kmeans fit to train
set.seed(3283) # set random seed
kmeansmodel <- kmeans(nosp_iris, centers = 3, nstart = 30, algorithm = "Lloyd")
```

```
## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations

## Warning: did not converge in 10 iterations
```

```
# 3 centers to start to check which combination of predictors is the best (3 types of species)
# using 30 (relatively big number) for number of starts to avoid getting stuck in undesirable local opt
# using default Lloyd algorithm for kmeans
```

Find the Best Combination of Predictors by Graphing

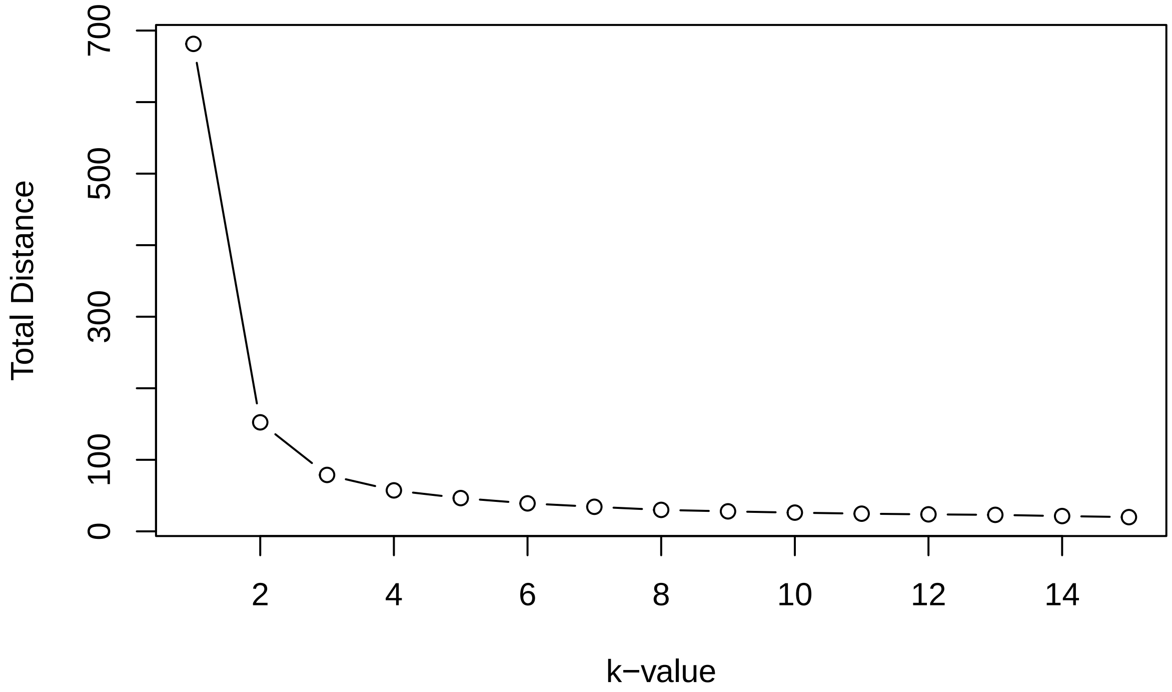
```
# Find best combination of predictors

par(mfrow = c(2, 2)) # set plots next to each other

# sepal length and width
plot(nosp_iris[c("Sepal.Length", "Sepal.Width")],
     col = kmeansmodel$cluster,
     main = "Sepal Length and Sepal Width")

# petal length and width
plot(nosp_iris[c("Petal.Length", "Petal.Width")],
     col = kmeansmodel$cluster,
     main = "Petal Length and Petal Width")

# sepal length and petal width
plot(nosp_iris[c("Sepal.Length", "Petal.Width")],
     col = kmeansmodel$cluster,
     main = "Sepal Length and Petal Width")

# petal length and sepal width
plot(nosp_iris[c("Petal.Length", "Sepal.Width")],
     col = kmeansmodel$cluster,
     main = "Petal Length and Sepal Width")
```



```
# From just looking at it, seems like petal length and petal width has the least overlap
```

Check for Best k (# of Clusters) to Use Using Elbow Method

```
# Best k value (center) to use for petal length and petal width (test out with for loop)
# Using Elbow Method: Check graph of total distance vs k (# clusters) and wherever it stops having dras
# total within distance of kmeans model is denoted with withinss

suppressWarnings ( {
ratios <- list()

for (k in 1:15) {
  kmeansmodel <- kmeans(nosp_iris, centers = k, nstart = 30, algorithm = "Lloyd")
  ratios <- append(ratios, kmeansmodel$tot.withinss)

}

par(mfrow=c(1,1))
plot(1:15, ratios, xlab = "k-value", ylab = "Total Distance", type="b")

})
```



```
# From the graph, seems like the best k to use is 3 (where the elbow bends)
```

Using Silhouette, Sum of Squares, Confusion Matrix for Evaluation

```
# Determine how good this model is with k = 3

# To get a numeric understanding of how good it is, will use the sihouette method
# The sihouette method measures quality of clustering by determing how well each object lies within its
# A high average sihouette width indicates good clustering.
suppressWarnings ( {
kmeansmodel <- kmeans(nosp_iris, centers = 3, nstart = 30, algorithm = "Lloyd")
})
s <- silhouette(kmeansmodel$cluster, dist(nosp_iris))
avg_s <- mean(s[,3]) # 3 is to index towards the sihouette values

print(paste("The average silhouette value of k = 3 is", avg_s, "and as the sihouette value > 0.5, it in
```

```
## [1] "The average silhouette value of k = 3 is 0.55281901235641 and as the sihouette value > 0.5, it i
```

```
# Sum of squares (evaluation)
kmeansmodel$withinss
```

```
## [1] 23.87947 39.82097 15.15100
```

```
# Confusion matrix (evaluation)
table(data$Species, kmeansmodel$cluster)
```

```
##
##               1  2  3
##   setosa      0  0 50
##   versicolor  2 48  0
##   virginica  36 14  0
```