

Homework 11

Peyton Hall

04/26/2025

Question 1

```
library(readxl)
shopping_mall_survey <- read_excel("~/Desktop/STAT 301/Week 14/shopping mall survey.xlsx")
shopping_mall_survey
```

```
## # A tibble: 10 x 6
##   ID      Question1 Question2 Question3 Question4 Question5
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 A          6          4          7          3          3
## 2 B          2          3          1          4          4
## 3 C          7          2          6          4          3
## 4 D          4          6          4          5          6
## 5 E          1          3          2          2          4
## 6 F          6          4          6          3          4
## 7 G          5          3          6          3          4
## 8 H          7          3          7          4          4
## 9 I          2          4          3          3          3
## 10 J         3          5          3          6          6
```

```
# a)
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
q1df <- scale(shopping_mall_survey[,c("Question1", "Question2", "Question3", "Question4", "Question5")])
q1dist <- dist(q1df, method = "euclidean")
modell1 <- hclust(q1dist, method = "complete")
fviz_dend(modell1, main = "Optimal number of clusters")
```

```
## Warning: The '<scale>' argument of 'guides()' cannot be 'FALSE'. Use "none" instead as
## of ggplot2 3.3.4.
```

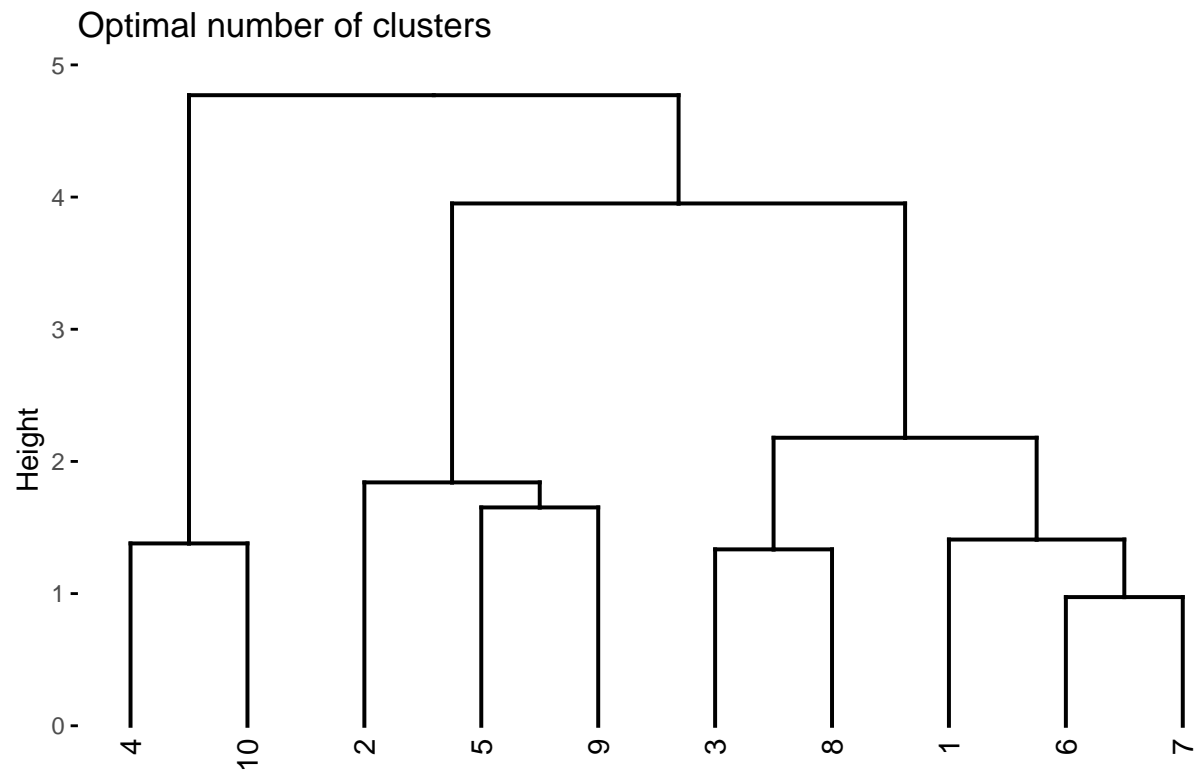
```
## i The deprecated feature was likely used in the factoextra package.
```

```
## Please report the issue at <https://github.com/kassambara/factoextra/issues>.
```

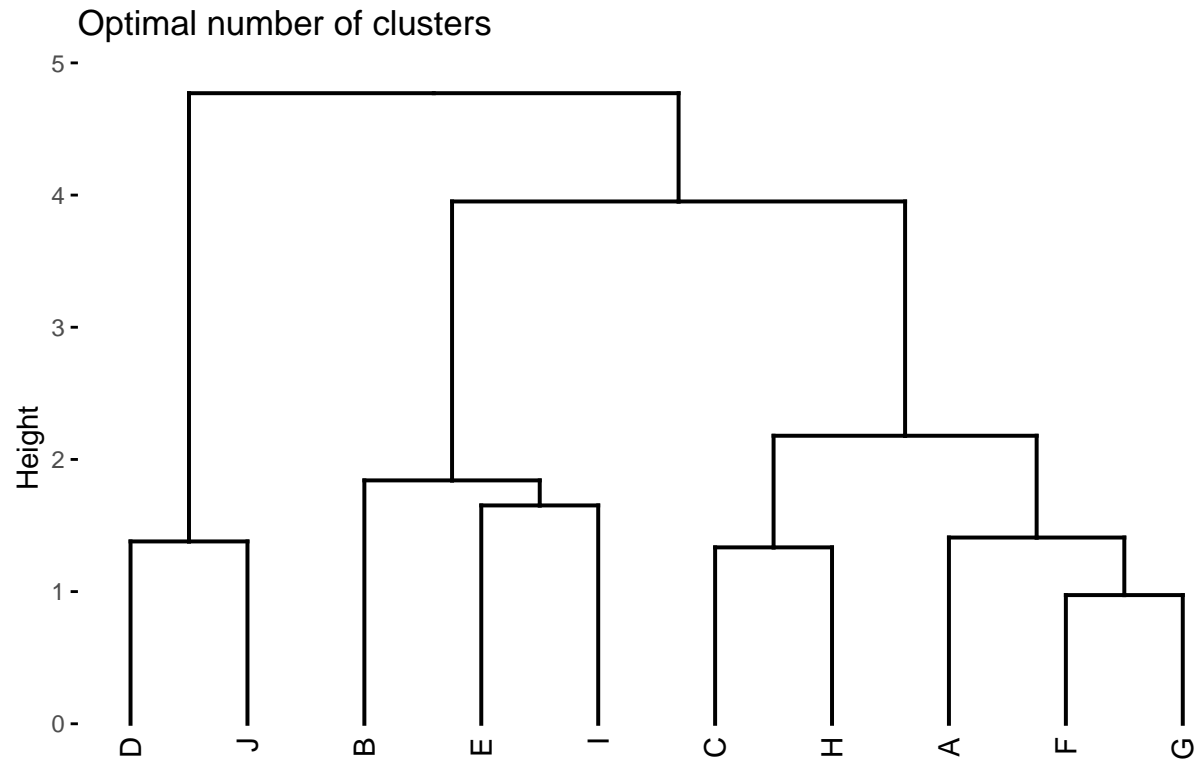
```
## This warning is displayed once every 8 hours.
```

```
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
```

```
## generated.
```



```
# b)
# centers = # of clusters, nstart = # of individuals
# model1 <- kmeans(q1df, centers = 3, nstart = 10)
# model1$cluster
model1$labels <- c("A","B","C","D","E","F","G","H","I","J")
fviz_dend(model1, main = "Optimal number of clusters")
```



Question 2

```
library(readxl)
personal_test_scores <- read_excel("~/Desktop/STAT 301/Week 14/personal teste scores.xlsx")
personal_test_scores
```

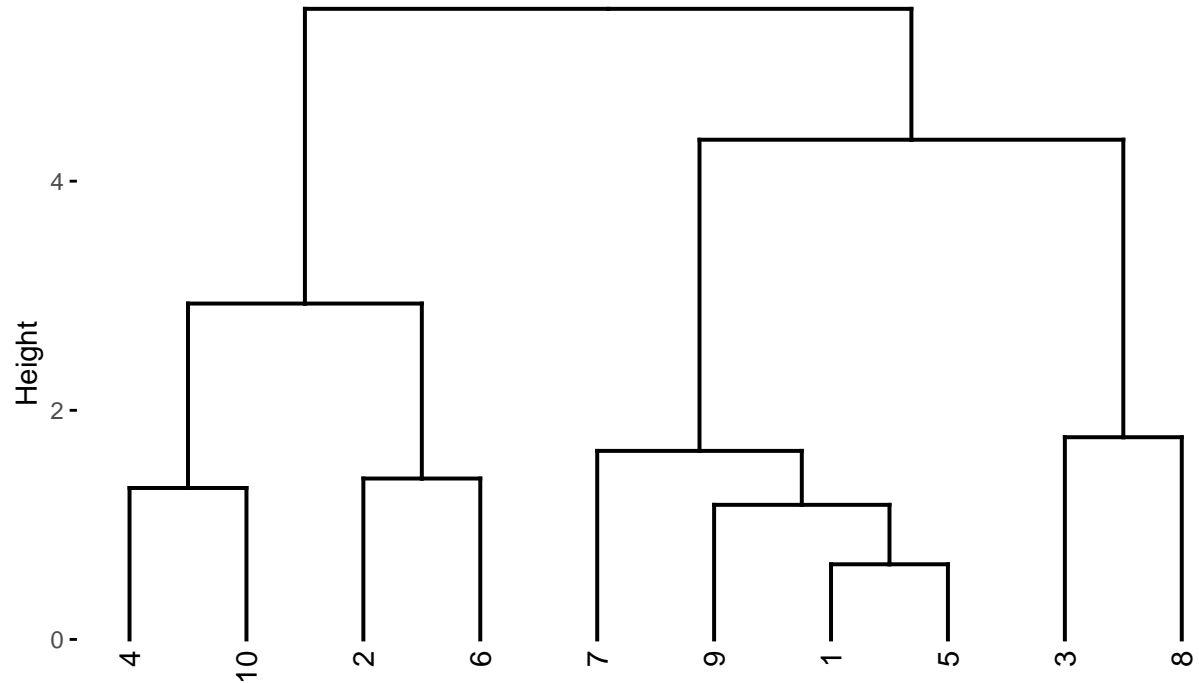
```
## # A tibble: 10 x 7
##   Participant Anxiety Agoraphobia Arachnophobia Adventure Extraversion
##   <chr>          <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 1             71          68          80          44          54
## 2 2             39          30          41          77          90
## 3 3             46          55          45          50          46
## 4 4             33          33          39          57          64
## 5 5             74          75          90          45          55
## 6 6             39          47          48          91          87
## 7 7             66          70          69          54          44
## 8 8             33          40          36          31          37
## 9 9             85          75          93          45          50
## 10 10           45          35          44          70          66
## # i 1 more variable: Sociability <dbl>
```

a) & b)

```
library(factoextra)
q2df <- scale(personal_test_scores[,c("Anxiety", "Agoraphobia", "Arachnophobia", "Adventure", "Extraversion")])
q2dist <- dist(q2df, method = "euclidean")
```

```
model2 <- hclust(q2dist, method = "complete")
fviz_dend(model2, main = "Optimal number of clusters")
```

Optimal number of clusters

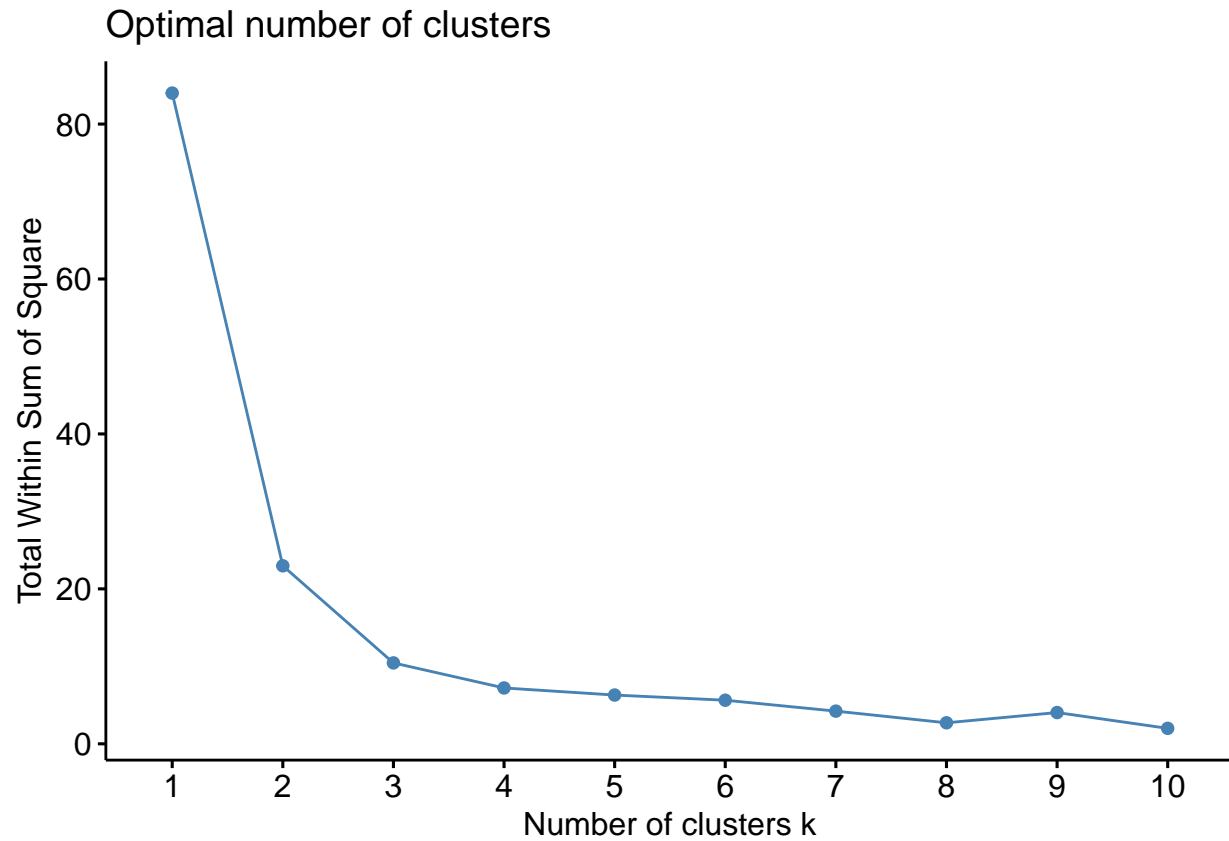


Question 3

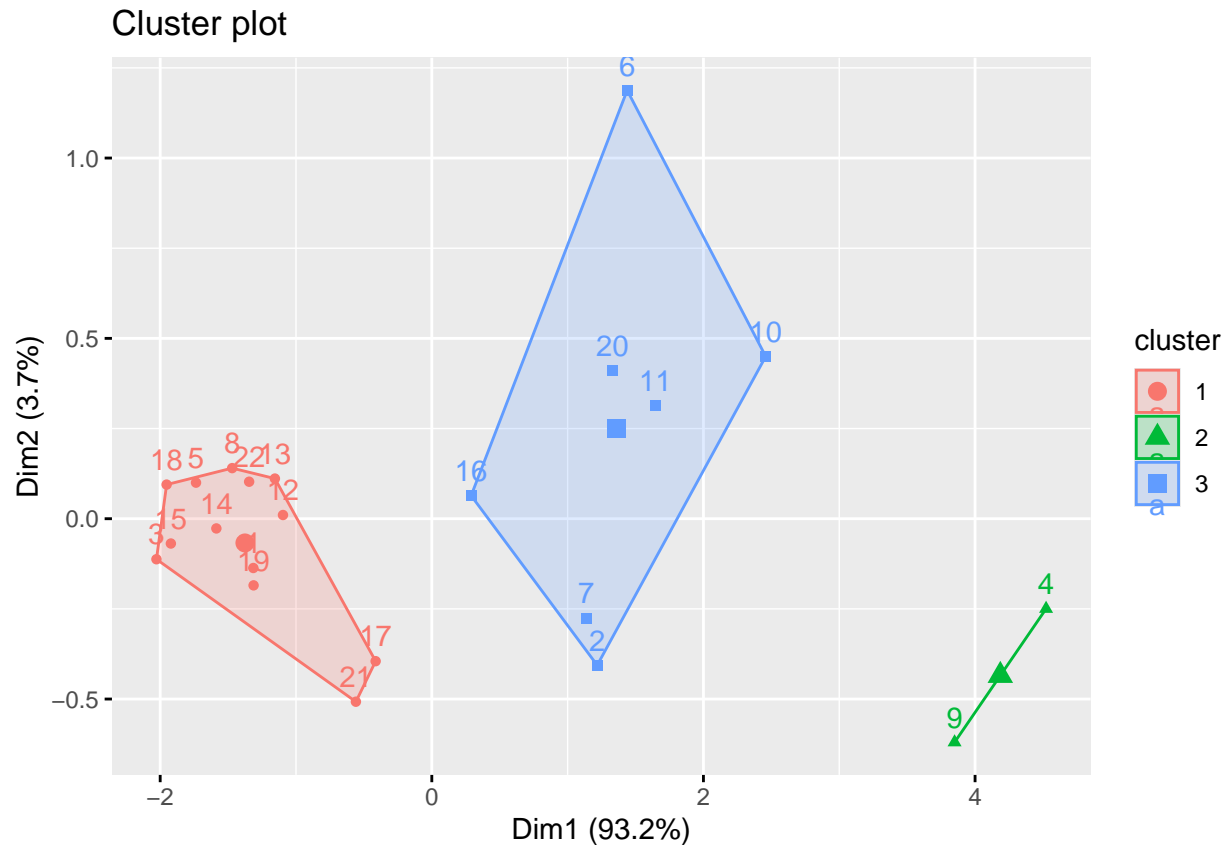
```
library(readxl)
student_grades <- read_excel("~/Desktop/STAT 301/Week 14/student grades.xlsx")
student_grades
```

```
## # A tibble: 22 x 5
##   studentID Math Science Reading Art
##   <chr>      <dbl>   <dbl>   <dbl> <dbl>
## 1 Jack      94      82     87    89
## 2 Tom       46      67     33    72
## 3 Mary     98      97     93   100
## 4 Sandy     14       5      7    24
## 5 Phil     86      97     95    95
## 6 Henry     34      32     75    66
## 7 Sam      69      44     59    55
## 8 Vicky     85      90     96    89
## 9 Jerry     24      26     15    22
## 10 Jacelin  25      33     45    52
## # i 12 more rows
```

```
# a)
library(factoextra)
q3df <- scale(student_grades[,c("Math", "Science", "Reading", "Art")])
fviz_nbclust(q3df, kmeans, method = "wss")
```



```
# b)
model3 <- kmeans(q3df, centers = 3, nstart = 22) # 3 clusters, 22 observations
fviz_cluster(model3, data = q3df)
```



```
model3$cluster
```

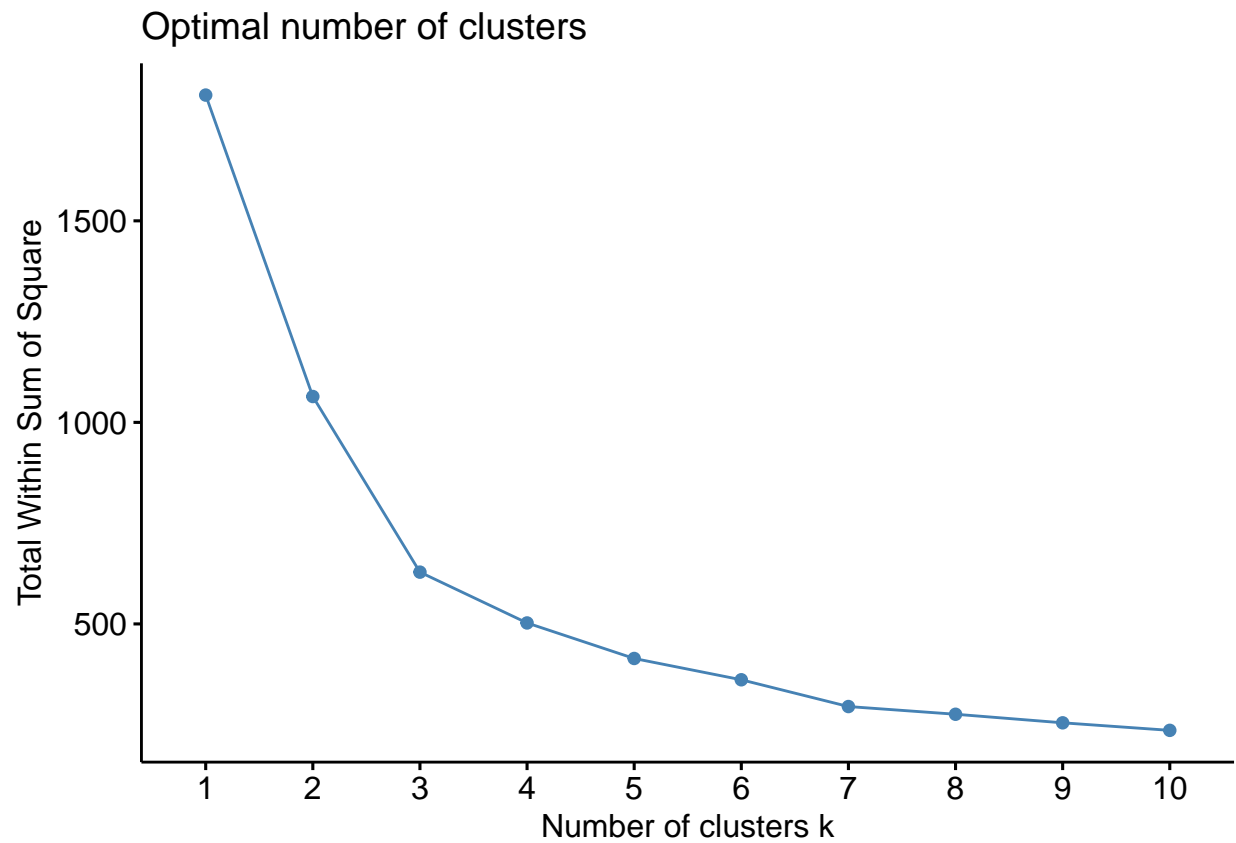
```
## [1] 1 3 1 2 1 3 3 1 2 3 3 1 1 1 3 1 1 3 1 1
```

Question 4

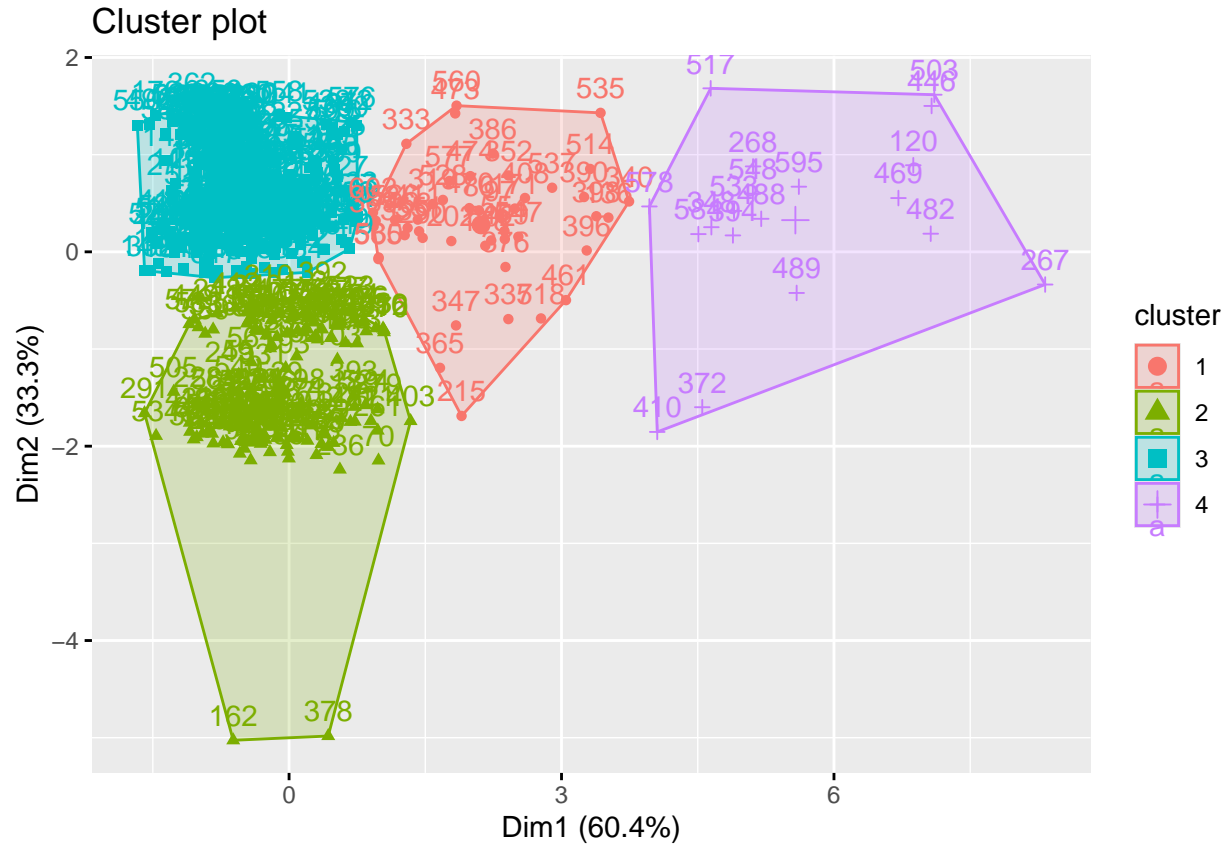
```
library(readxl)
nyc_flights <- read_excel("~/Desktop/STAT 301/Week 14/NYCflights.xlsx")
nyc_flights
```

```
## # A tibble: 605 x 6
##   year month   day dep_delay arr_delay air_time
##   <dbl> <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1  2013     1     1         2        11     227
## 2  2013     1     1         4        20     227
## 3  2013     1     1         2        33     160
## 4  2013     1     1        -1       -18     183
## 5  2013     1     1        -6       -25     116
## 6  2013     1     1        -4        12     150
## 7  2013     1     1        -5        19     158
## 8  2013     1     1        -3       -14      53
## 9  2013     1     1        -3        -8     140
## 10 2013     1     1        -2         8     138
## # i 595 more rows
```

```
# a)
library(factoextra)
q4df <- scale(nyc_flights[,c(4:6)])
fviz_nbclust(q4df, kmeans, method = "wss")
```



```
# b)
model4 <- kmeans(q4df, centers = 4, nstart = 50)
fviz_cluster(model4, data = q4df)
```



```
model4$cluster
```

```
## [1] 2 2 3 3 3 3 3 3 3 3 3 3 2 2 2 3 2 3 3 3 3 3 3 3 2 3 3 3 2 3 3 2 3
## [38] 2 3 3 3 3 2 3 3 3 3 3 3 2 2 2 3 3 2 2 3 3 3 2 3 3 3 2 3 3 3 3 2 3 3 3
## [75] 3 3 3 2 2 2 2 2 2 3 3 1 3 2 3 2 3 2 2 2 1 3 2 3 3 3 2 3 3 3 3 3 2 3
## [112] 3 3 3 3 3 3 3 3 4 3 3 2 3 3 3 3 3 2 3 3 3 2 3 3 1 2 3 2 2 3 3 3 3 2 2 3
## [149] 3 3 3 2 3 3 2 3 3 3 2 3 3 2 3 3 2 3 3 2 3 3 1 3 2 2 3 3 2 1 3 3 3 2 3 2 3
## [186] 3 3 3 2 3 3 3 2 3 3 2 3 3 3 2 3 1 3 3 3 3 3 3 3 2 3 3 3 3 1 2 3 3 3 3 2 2
## [223] 3 3 3 3 3 2 3 3 3 2 3 2 2 2 3 2 3 3 3 3 3 3 3 3 2 3 2 2 3 3 3 3 3 3 3 2 3
## [260] 3 2 3 3 3 2 3 4 4 3 2 2 3 3 3 2 3 3 3 3 3 3 3 3 2 3 3 2 3 3 3 3 1 2 3 3 2 3 3
## [297] 3 3 3 3 2 3 3 3 3 3 3 2 3 2 3 2 3 3 2 3 3 3 1 3 2 3 3 3 3 3 3 3 3 3 2 3 1
## [334] 2 3 2 1 3 2 1 3 3 2 3 3 3 1 4 2 3 3 1 3 3 1 2 3 3 3 3 3 3 3 3 2 1 3 2 3 1 3
## [371] 3 4 3 3 2 1 2 2 2 3 3 3 3 3 3 1 3 3 3 1 3 2 2 4 3 1 3 1 3 2 3 3 2 3 3 2 3
## [408] 1 3 4 1 2 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 2 3 3 3 3 3
## [445] 3 4 3 2 2 3 3 1 3 3 3 3 3 3 3 1 2 1 3 3 3 3 2 2 1 4 1 3 3 1 1 2 3 3 3 2 2 2
## [482] 4 2 2 3 3 3 4 4 3 3 3 3 3 2 3 3 3 3 3 3 3 1 4 3 2 3 3 3 3 3 2 3 3 1 3 3 4 1
## [519] 3 3 3 3 2 2 3 3 2 1 2 1 2 3 4 2 1 2 1 3 3 2 2 3 3 2 3 3 1 4 3 3 1 3 3 1 2
## [556] 2 3 3 2 1 2 2 2 2 1 1 3 2 3 2 2 3 3 2 2 3 1 4 2 2 3 3 2 4 3 3 3 3 2 3 2 3
## [593] 2 2 4 2 3 2 2 3 3 1 3 3 3
```