1. The ==penguins== dataset included in the ==palmerpenguins== package provides the size measurements for adult foraging penguins near Palmer Station, Antarctica.
   a) Access the data and determine its dimension.
   b) How many species of penguins are provided in the dataset?
   c) Extract the variables bill_length_mm and flipper_length_mm and the corresponding species.
   d) Calculate standard scores of both variables bill_length_mm and flipper_length_mm
   e) Determine the beat value of k to choose k means cluster.
   f) Create the number of clusters per the recommendation of part (e) and determine the size of the in each cluster.
   g) Calculate the center of each cluster
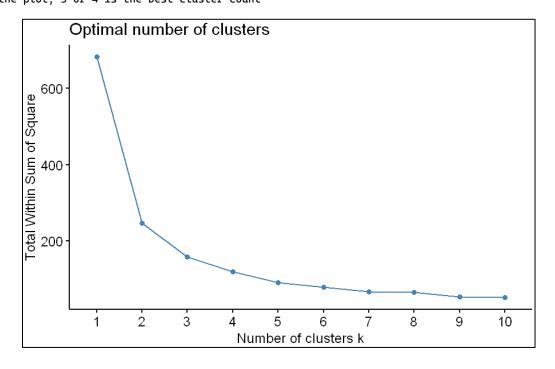   h) Assess the performance of the cluster analysis.

```
> # Lab 9
> # 1
> # a
> install.packages("palmerpenguins")
> library(palmerpenguins)
> data(penguins, package = "palmerpenguins")
> head(penguins)
# A tibble: 6 × 8
  species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g sex      year
  <fct>   <fct>            <dbl>         <dbl>             <int>        <int> <fct>   <int>
1 Adelie  Torgersen         39.1          18.7               181         3750 male     2007
2 Adelie  Torgersen         39.5          17.4               186         3800 female   2007
3 Adelie  Torgersen         40.3          18                 195         3250 female   2007
4 Adelie  Torgersen         NA            NA                 NA           NA NA        2007
5 Adelie  Torgersen         36.7          19.3               193         3450 female   2007
6 Adelie  Torgersen         39.3          20.6               190         3650 male     2007
> dim(penguins)
[1] 344   8
```

```
> # b
> names(penguins)
[1] "species"         "island"          "bill_length_mm"   "bill_depth_mm"
[5] "flipper_length_mm" "body_mass_g"       "sex"              "year"
> attach(penguins)
> table(species)
species
   Adelie Chinstrap    Gentoo
      152        68       124
> cat("There are 3 species of penguins in the given dataset")
There are 3 species of penguins in the given dataset
```

```
> # c
> new = penguins[,c(1,3,5)]
> head(new)
# A tibble: 6 × 3
  species bill_length_mm flipper_length_mm
  <fct>          <dbl>             <int>
1 Adelie          39.1               181
```

```
2 Adelie             39.5              186
3 Adelie             40.3              195
4 Adelie             NA                NA
5 Adelie             36.7              193
6 Adelie             39.3              190
> dim(new)
[1] 344    3
> names(new)
[1] "species"         "bill_length_mm"    "flipper_length_mm"
> cleandata <- na.omit(new)
> head(cleandata)
# A tibble: 6 × 3
  species bill_length_mm flipper_length_mm
  <fct>            <dbl>             <int>
1 Adelie            39.1               181
2 Adelie            39.5               186
3 Adelie            40.3               195
4 Adelie            36.7               193
5 Adelie            39.3               190
6 Adelie            38.9               181
> dim(cleandata)
[1] 342    3



> # also
> needed_data <- cleandata[,c(2,3)]
> head(needed_data)
# A tibble: 6 × 2
  bill_length_mm flipper_length_mm
           <dbl>             <int>
1           39.1               181
2           39.5               186
3           40.3               195
4           36.7               193
5           39.3               190
6           38.9               181
> dim(needed_data)
[1] 342    2
> class = cleandata$species
> head(class)
[1] Adelie Adelie Adelie Adelie Adelie Adelie
Levels: Adelie Chinstrap Gentoo
> length(class)
[1] 342
> class
  [1] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [10] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [19] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [28] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [37] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [46] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [55] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [64] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [73] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [82] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
 [91] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
[100] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
[109] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
[118] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
[127] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
[136] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
[145] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Gentoo    Gentoo
[154] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
[163] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
[172] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
[181] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
```

```
[190] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[199] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[208] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[217] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[226] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[235] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[244] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[253] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[262] Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo     Gentoo
[271] Gentoo     Gentoo     Gentoo     Gentoo     Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[280] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[289] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[298] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[307] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[316] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[325] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
[334] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
Levels: Adelie Chinstrap Gentoo
```

```r
> # d
> data1 <- scale(needed_data)
> head(data1)
     bill_length_mm flipper_length_mm
[1,]     -0.8832047        -1.4162715
[2,]     -0.8099390        -1.0606961
[3,]     -0.6634077        -0.4206603
[4,]     -1.3227986        -0.5628905
[5,]     -0.8465718        -0.7762357
[6,]     -0.9198375        -1.4162715
```

```r
> # e
> install.packages("factoextra")
> library(factoextra)
> fviz_nbclust(data1, kmeans, method="wss")
> cat("From the plot, 3 or 4 is the best cluster count")
From the plot, 3 or 4 is the best cluster count
```


Optimal number of clusters

```
> # f
> result = kmeans(data1,3)
> result$size
[1]  64 127 151
> result
K-means clustering with 3 clusters of sizes 64, 127, 151

Cluster means:
  bill_length_mm flipper_length_mm
1      0.9367029        -0.3695463
2      0.6656141         1.1461115
3     -0.9568342        -0.8073192

Clustering vector:
  [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3
 [46] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [91] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3
[136] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[181] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[226] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[271] 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 3 1 1 1 1 1 1 1 3 1 3 1 1 1 1 2 1 1 2
[316] 1 1 1 1 1 1 2 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 2 1

Within cluster sum of squares by cluster:
[1] 29.51333 66.51303 61.32751
 (between_SS / total_SS =  76.9 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
> cat("For 3 clusters: 127 151  64")
For 3 clusters: 127 151  64
>
>
>
> result1 = kmeans(data1,4)
> result1$size
[1]  62 149  50  81
> result1
K-means clustering with 4 clusters of sizes 62, 149, 50, 81

Cluster means:
  bill_length_mm flipper_length_mm
1      0.9330747        -0.3896908
2     -0.9623709        -0.8234867
3      1.2356382         1.5250485
4      0.2933423         0.8717028

Clustering vector:
  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
 [46] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 [91] 4 2 2 2 4 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 2 2 2 2 2 2
[136] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 3 4 3 4 4 4 4 4 4 4 3 4 4 4 3 4 3 4 3 3 4 4 4 4 4 4 4 4 4
[181] 3 3 4 4 3 3 3 4 4 4 4 4 3 4 3 3 4 4 3 4 4 4 3 4 3 4 4 4 4 4 3 4 4 4 3 4 3 4 3 4 3 4 4 3 4
[226] 4 3 4 3 4 4 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 3 4 1 3 3 3 4 3 4 4 4 3 4 4 3 3 4 3 4 3 4 3 4
[271] 4 3 4 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 3 1 1 4
[316] 1 1 1 1 1 1 4 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 3 4 1 3 1

Within cluster sum of squares by cluster:
[1] 26.08107 58.01789 16.93993 17.38010
 (between_SS / total_SS =  82.6 %)

Available components:
```

```
[1] "cluster"      "centers"     "totss"       "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"        "ifault"
> cat("For 4 clusters: 41 116 116  69")
For 4 clusters: 41 116 116  69




> # g
> # For 3 clusters
> result$centers
  bill_length_mm flipper_length_mm
1      0.9367029        -0.3695463
2      0.6656141         1.1461115
3     -0.9568342        -0.8073192
> result$cluster
  [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3
 [46] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [91] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3
[136] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[181] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[226] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[271] 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 3 1 1 1 1 1 1 1 3 1 3 1 1 1 1 2 1 1 2
[316] 1 1 1 1 1 1 2 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 2 1
> result$size
[1]  64 127 151
>
>
>
> # For 4 clusters
> result1$centers
  bill_length_mm flipper_length_mm
1      0.9330747        -0.3896908
2     -0.9623709        -0.8234867
3      1.2356382         1.5250485
4      0.2933423         0.8717028
> result1$cluster
  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
 [46] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 [91] 4 2 2 2 4 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 2 2 2 2 2 2
[136] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 3 4 3 4 4 4 4 4 4 4 3 4 4 4 3 4 3 4 3 3 4 4 4 4 4 4 4 4 4
[181] 3 3 4 4 3 3 3 4 4 4 4 4 3 4 3 3 4 4 3 4 4 4 3 4 3 4 4 4 4 4 3 4 4 4 3 4 3 4 3 4 3 4 4 4 3 4
[226] 4 3 4 3 4 4 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 3 4 1 3 3 3 4 3 4 4 4 3 4 4 3 3 4 3 4 3 4 3 4 3 4
[271] 4 3 4 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 3 1 1 4
[316] 1 1 1 1 1 1 4 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 3 4 1 3 1
> result1$size
[1]  62 149  50  81




> # h
> result$cluster
  [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3
3
 [49] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3
 [97] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3
[145] 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2
[193] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2
[241] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1
[289] 1 1 1 1 1 1 3 1 3 1 1 1 1 1 1 1 1 3 1 3 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 3 1 1 1 1 1 1
1
[337] 1 1 1 1 2 1
```

```
> length(result$cluster)
[1] 342
> table(result$cluster, class)
   class
    Adelie Chinstrap Gentoo
  1      4        59      1
  2      1         4    122
  3    146         5      0
> table(class)
class
   Adelie Chinstrap    Gentoo
      151        68       123
> cat("Thus, we can see that: FOR 3 CLUSTER SYSTEM
+ 1. Out of 151 actual Adelie species, 146 were correctly identified and rest 5 were wrongly identi
fied.
+ 2. Out of 68 actual Chinstrap species, 59 were correctly identified and rest 9 were wrongly ident
ified.
+ 3. Out of 123 actual Gentoo species, 122 were correctly identified and rest 1 is wrongly identifi
ed.")
Thus, we can see that: FOR 3 CLUSTER SYSTEM
1. Out of 151 actual Adelie species, 146 were correctly identified and rest 5 were wrongly identifi
ed.
2. Out of 68 actual Chinstrap species, 59 were correctly identified and rest 9 were wrongly identif
ied.
3. Out of 123 actual Gentoo species, 122 were correctly identified and rest 1 is wrongly identified
.
```