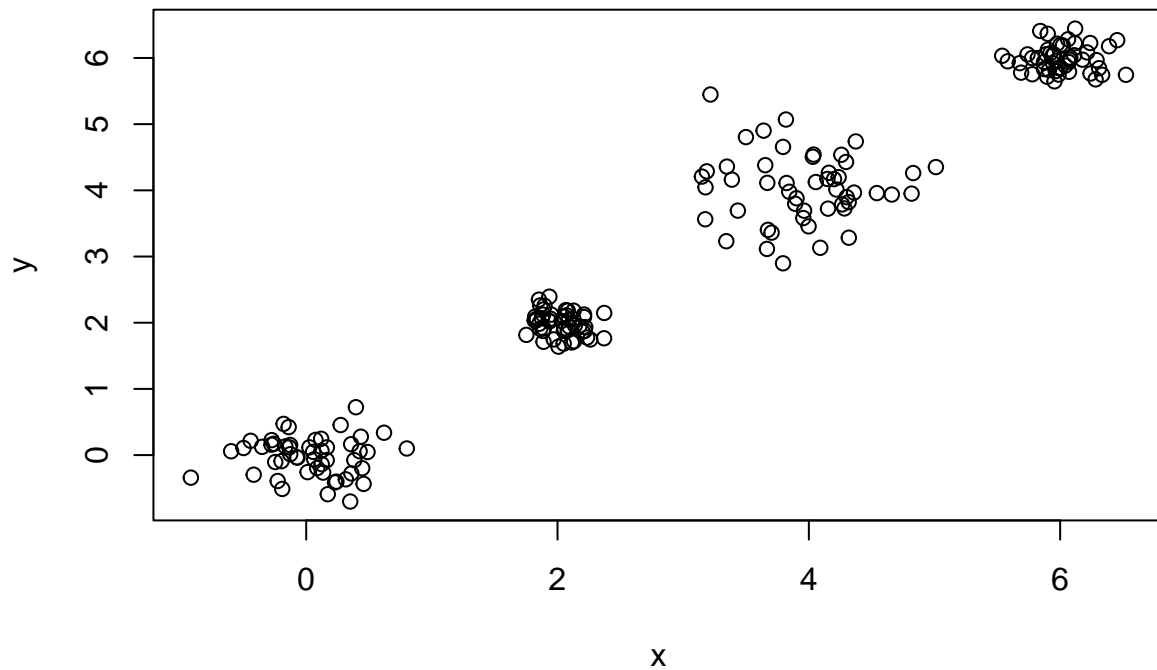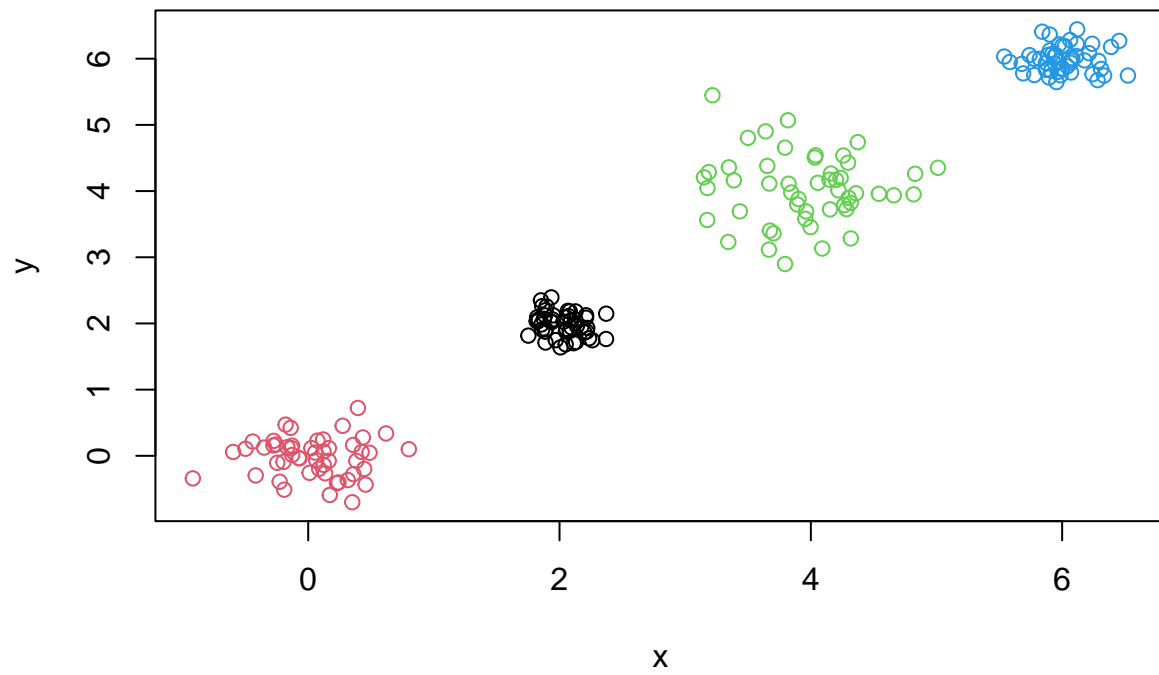# ICE5

Shaowei He

11/9/2021
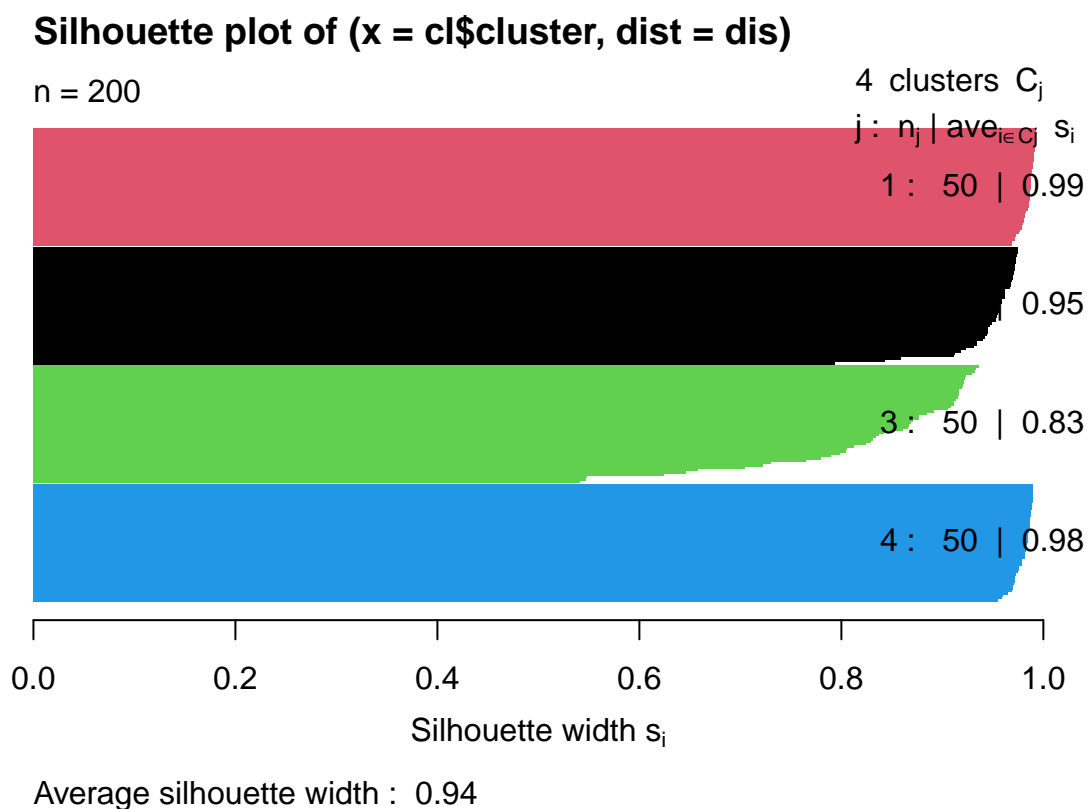
```r
twoDData<-rbind(matrix(rnorm(100, mean = 0, sd = 0.3), ncol = 2),
          matrix(rnorm(100, mean = 2, sd = 0.2), ncol = 2),
          matrix(rnorm(100, mean = 4, sd = 0.5), ncol = 2),
          matrix(rnorm(100, mean = 6, sd = 0.2), ncol = 2))
colnames(twoDData)<-c("x","y")
plot(twoDData)
```



```r
cl<-kmeans(twoDData, centers = 4)
plot(twoDData, col = cl$cluster)
```

```
library(cluster)
dis=dist(twoDData)^2
sil=silhouette(cl$cluster, dis)
plot(sil,col = cl$cluster)
```

Silhouette plot of (x = cl$cluster, dist = dis)

n = 200

4 clusters C_j
$j : n_j \mid ave_{i \in C_j} \; s_i$

1 : 50 | 0.99

0.95

3 : 50 | 0.83

4 : 50 | 0.98

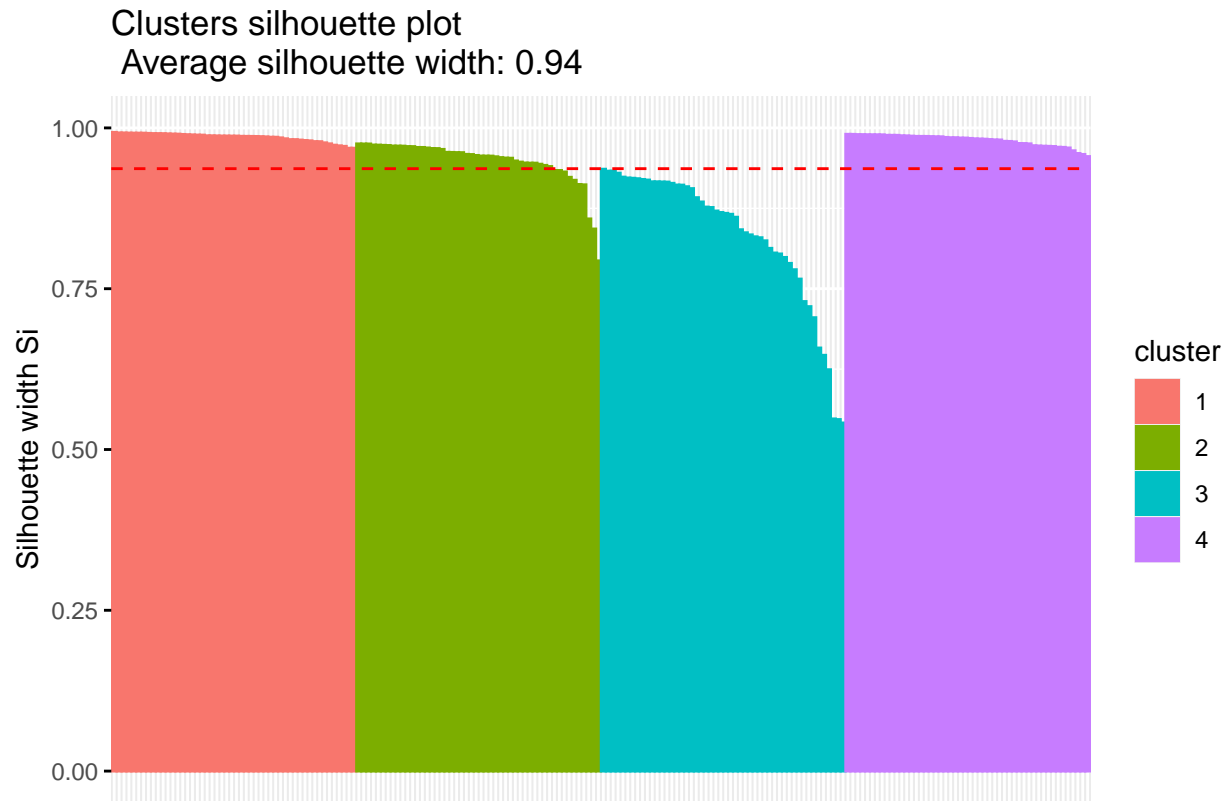Silhouette width s_i

Average silhouette width : 0.94

```
library(factoextra)
```

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

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

```
fviz_silhouette(sil)
```

```
##   cluster size ave.sil.width
## 1       1   50          0.99
## 2       2   50          0.95
## 3       3   50          0.83
## 4       4   50          0.98
```

Clusters silhouette plot
Average silhouette width: 0.94


```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
setwd("~/Desktop")
motivation<-read_csv("~/Desktop/ICE5_Data.csv")
```

```
## Rows: 38 Columns: 6
```

```
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## dbl (6): id, motivation1, motivation2, motivation3, motivation4, motivation5
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
motivationClean <- motivation %>% na.omit() %>% select(-id)
motivationClean
```

```
## # A tibble: 23 x 5
##    motivation1 motivation2 motivation3 motivation4 motivation5
##          <dbl>       <dbl>       <dbl>       <dbl>       <dbl>
##  1           2           2           2           2           2
##  2           1           2           1           2           2
##  3           2           2           2           2           2
##  4           2           2           2           2           2
##  5           2           1           2           1           2
##  6           2           3           2           3           1
##  7           1           1           1           1           4
##  8           1           1           1           1           3
##  9           1           1           1           1           1
## 10           1           1           1           1           1
## # ... with 13 more rows
```
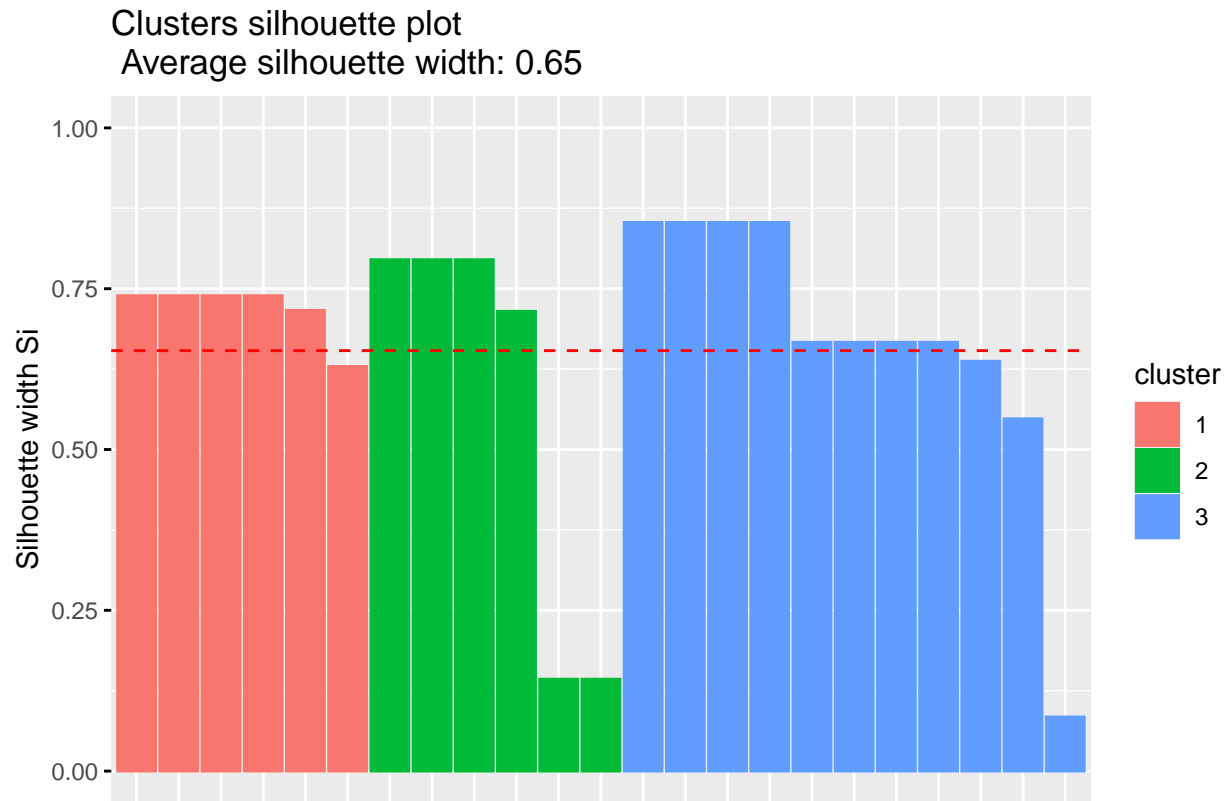
```
motivation2CL <- kmeans(motivationClean, centers = 2)
dis2CL = dist(motivationClean)^2
sil2CL = silhouette(motivation2CL$cluster, dis2CL)
fviz_silhouette(sil2CL)
```

```
##   cluster size ave.sil.width
## 1       1   11          0.72
## 2       2   12          0.44
```
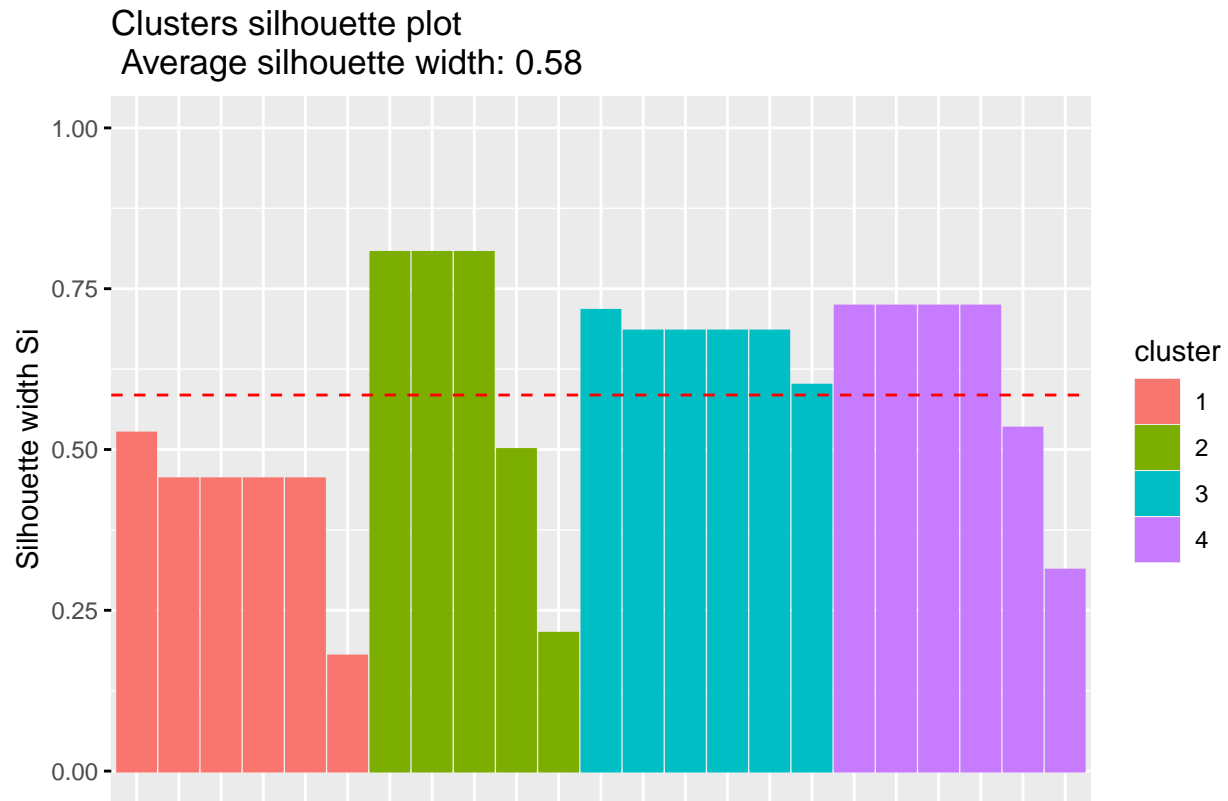
Clusters silhouette plot
Average silhouette width: 0.58
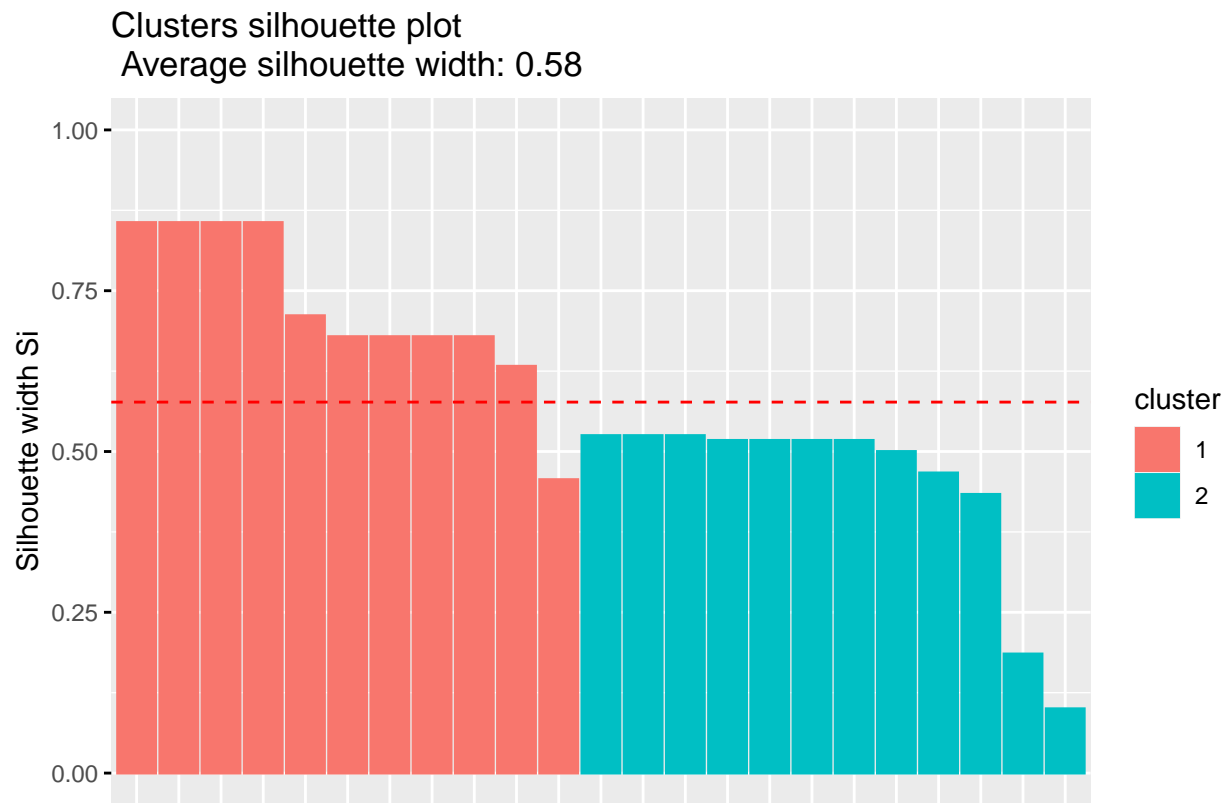
```
motivation3CL <- kmeans(motivationClean, centers = 3)
dis3CL = dist(motivationClean)^2
sil3CL = silhouette(motivation3CL$cluster, dis3CL)
fviz_silhouette(sil3CL)
```

```
##   cluster size ave.sil.width
## 1       1    6          0.72
## 2       2    6          0.56
## 3       3   11          0.67
```

Clusters silhouette plot
Average silhouette width: 0.65


```
motivation4CL <- kmeans(motivationClean, centers = 4)
dis4CL = dist(motivationClean)^2
sil4CL = silhouette(motivation4CL$cluster, dis4CL)
fviz_silhouette(sil4CL)
```

```
##   cluster size ave.sil.width
## 1       1    6          0.42
## 2       2    5          0.63
## 3       3    6          0.68
## 4       4    6          0.62
```

Clusters silhouette plot
Average silhouette width: 0.58


```
motivationClean %>% mutate(cluster = motivation3CL$cluster)
```

```
## # A tibble: 23 x 6
##    motivation1 motivation2 motivation3 motivation4 motivation5 cluster
##          <dbl>       <dbl>       <dbl>       <dbl>       <dbl>   <int>
##  1           2           2           2           2           2       1
##  2           1           2           1           2           2       2
##  3           2           2           2           2           2       1
##  4           2           2           2           2           2       1
##  5           2           1           2           1           2       2
##  6           2           3           2           3           1       1
##  7           1           1           1           1           4       3
##  8           1           1           1           1           3       3
##  9           1           1           1           1           1       2
## 10           1           1           1           1           1       2
## # ... with 13 more rows
```

```
fviz_nbclust(motivationClean, kmeans, method = "wss")
```

Optimal number of clusters