

R Notebook

GUOTAI SUN

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
#install.packages("tidyverse")
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.5

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

## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.4       v dplyr 1.0.7
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5
## Warning: package 'tibble' was built under R version 4.0.5
## Warning: package 'tidyr' was built under R version 4.0.5
## Warning: package 'dplyr' was built under R version 4.0.5

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

library(cluster)

#install.packages("factoextra")
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.0.5

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

A <- read_csv("ICE5_Data.csv")

##
## -- Column specification -----
-----
```

```

## cols(
##   id = col_double(),
##   motivation1 = col_double(),
##   motivation2 = col_double(),
##   motivation3 = col_double(),
##   motivation4 = col_double(),
##   motivation5 = col_double()
## )

#remove the id variable, use na.omit to get rid of all the rows contain
ing NAs
AClean <- A %>% na.omit() %>%
  select(-id)

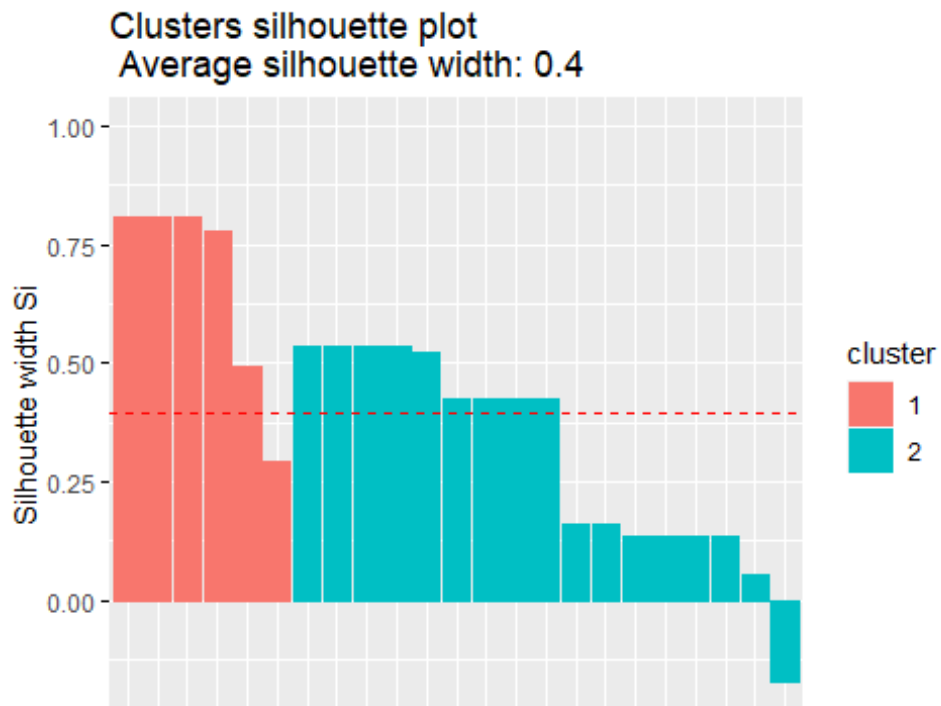
AClean

## # 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

#two-cluster model, k=2
A2CL = kmeans(AClean, centers = 2)
dis2CL = dist(AClean)^2
sil2CL = silhouette(A2CL$cluster, dis2CL)
fviz_silhouette(sil2CL)

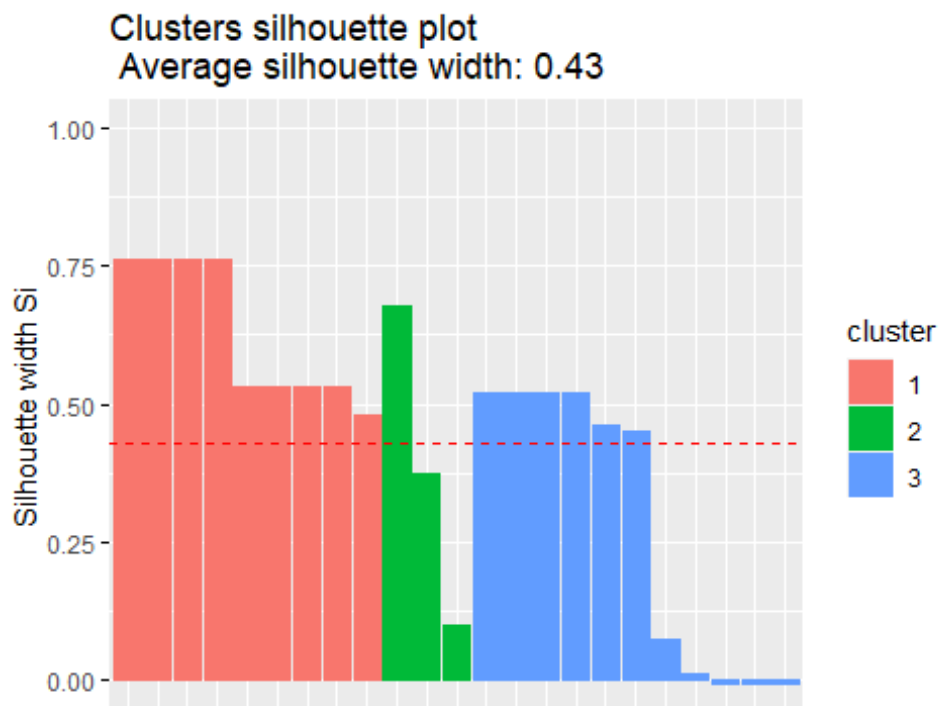
##   cluster size ave.sil.width
## 1       1    6         0.66
## 2       2   17         0.30

```



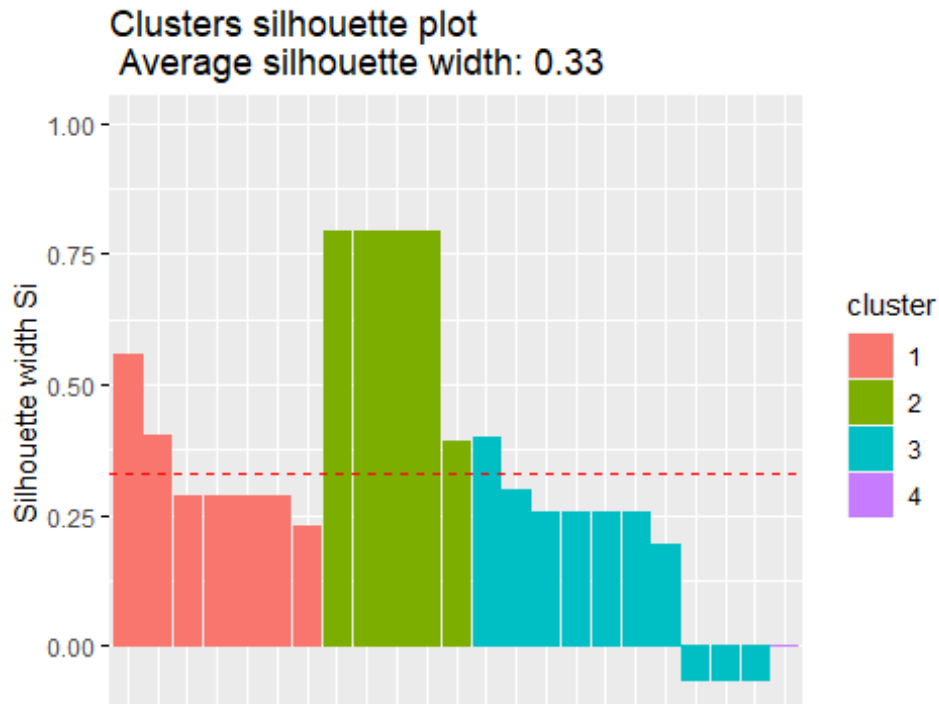
```
#three-cluster model, k=3
A3CL <- kmeans(AClean, centers = 3)
dis3CL = dist(AClean)^2
sil3CL = silhouette(A3CL$cluster, dis3CL)
fviz_silhouette(sil3CL)
```

##	cluster	size	ave.sil.width
## 1	1	9	0.63
## 2	2	3	0.38
## 3	3	11	0.28



```
#four-cluster model, k=4
A4CL <- kmeans(AClean, centers = 4)
dis4CL = dist(AClean)^2
sil4CL = silhouette(A4CL$cluster, dis4CL)
fviz_silhouette(sil4CL)
```

```
##  cluster size ave.sil.width
## 1         1    7         0.33
## 2         2    5         0.71
## 3         3   10         0.17
## 4         4    1         0.00
```



#The silhouette coefficient drops and all four clusters all still have similar intra-cluster similarity. So we will say cluster_n = 3 is a more reasonable classification

#add one more column as cluster

```
AClean %>% mutate(cluster = A3CL$cluster)
```

```
## # A tibble: 23 x 6
```

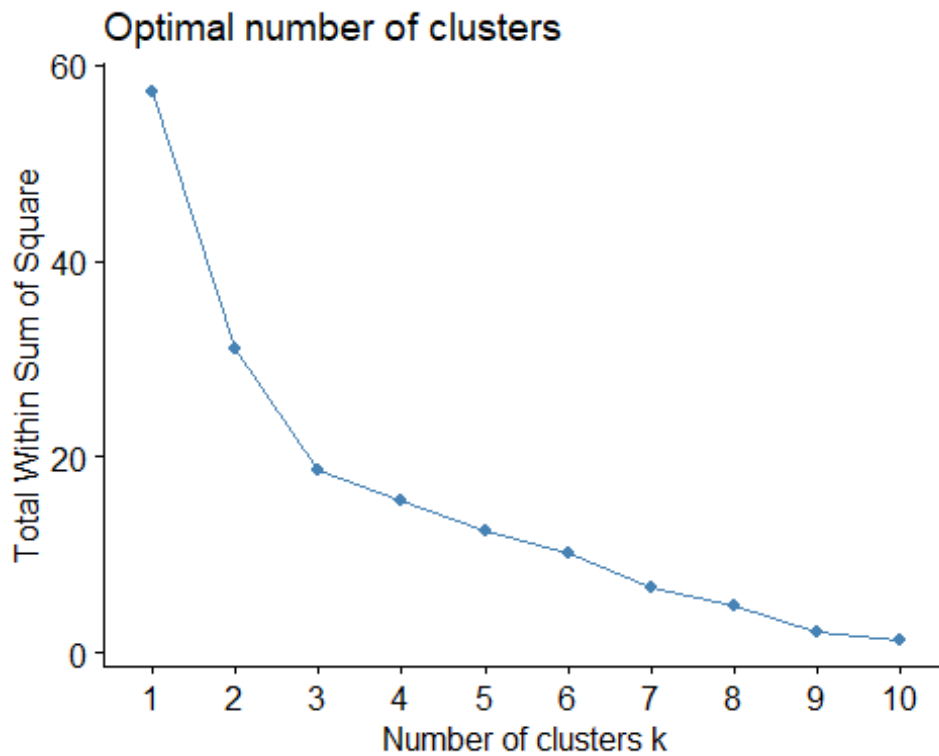
```
##   motivation1 motivation2 motivation3 motivation4 motivation5 cluster
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>   <int>
## 1           2           2           2           2           2       2
## 2           1           2           1           2           2       2
## 3           2           2           2           2           2       2
## 4           2           2           2           2           2       2
## 5           2           1           2           1           2       2
## 6           2           3           2           3           1       3
## 7           1           1           1           1           4       1
## 8           1           1           1           1           3       3
```

```

2
## 9      1      1      1      1      1
3
## 10     1      1      1      1      1
3
## # ... with 13 more rows

fviz_nbclust(AClean, kmeans, method = "wss")

```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.