

# ClusterAnalysis

Akhila

07/05/2019

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'tinytex' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'magrittr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## Warning: package 'hclust' is not available (for R version 3.6.0)
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'ggdendro' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## Warning: package 'agnes' is not available (for R version 3.6.0)
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'ggloop' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'factoextra' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Installing package into 'C:/Users/AKMANI/Documents/R/win-library/3.6'
## (as 'lib' is unspecified)
```

```
## package 'cluster' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'cluster'
```

```
##
## The downloaded binary packages are in
## C:\Users\AKMANI\AppData\Local\Temp\Rtmp8QpDWp\downloaded_packages
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method      from
##   [.quosures   rlang
##   c.quosures   rlang
##   print.quosures rlang
```

```
## Registered S3 method overwritten by 'rvest':
##   method      from
##   read_xml.response xml2
```

```
## -- Attaching packages -----
## ----- tidyverse 1.2.1 -----
```

```
## v ggplot2 3.1.1      v purrr  0.3.2
## v tibble  2.1.1      v dplyr  0.8.0.1
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

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

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##   set_names
```

```
## The following object is masked from 'package:tidyr':
##
##   extract
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

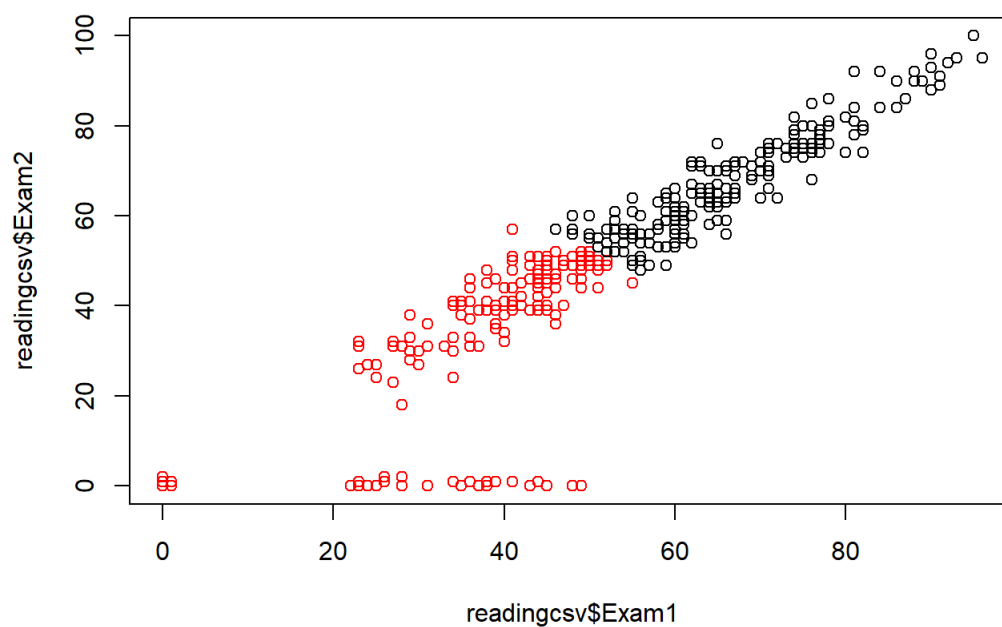
```
## [1] "C:/Users/AKMANI/Desktop/DKE_OVGU/Semester 3/Visual Analytics/Exercise/Exercise 03"
```

```
## 'data.frame': 395 obs. of 2 variables:  
## $ Exam1: num 28 27 41 71 52 76 59 23 92 74 ...  
## $ Exam2: num 31 31 51 75 50 75 53 31 94 75 ...
```

```
for (k in 2:8){  
  tryCatch({  
    print(k)  
    dfCluster <- kmeans(readingcsv,centers = k, iter.max = 10)  
  },  
  warning=function(w) {done <- FALSE})  
  plot(x=readingcsv$Exam1,  
       y=readingcsv$Exam2,  
       col=as.numeric(dfCluster$clust),main=paste("K =",k) )  
}
```

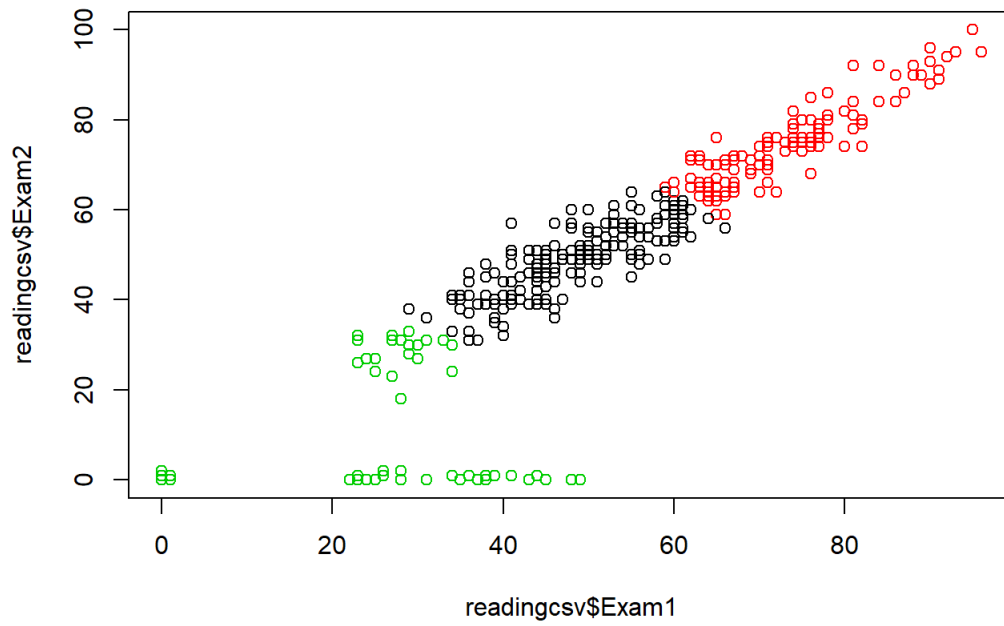
```
## [1] 2
```

**K = 2**



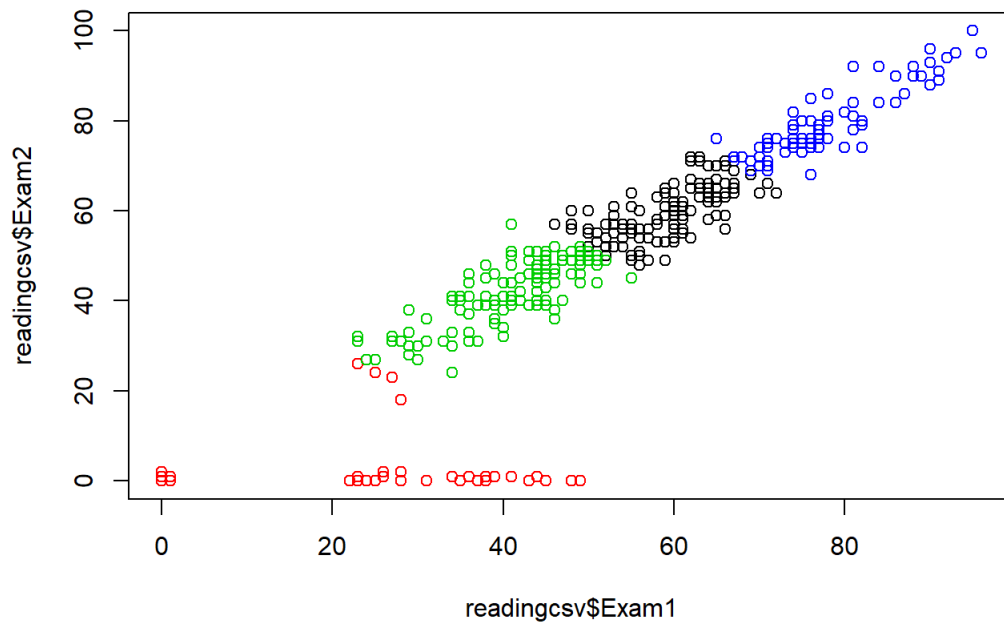
```
## [1] 3
```

**K = 3**



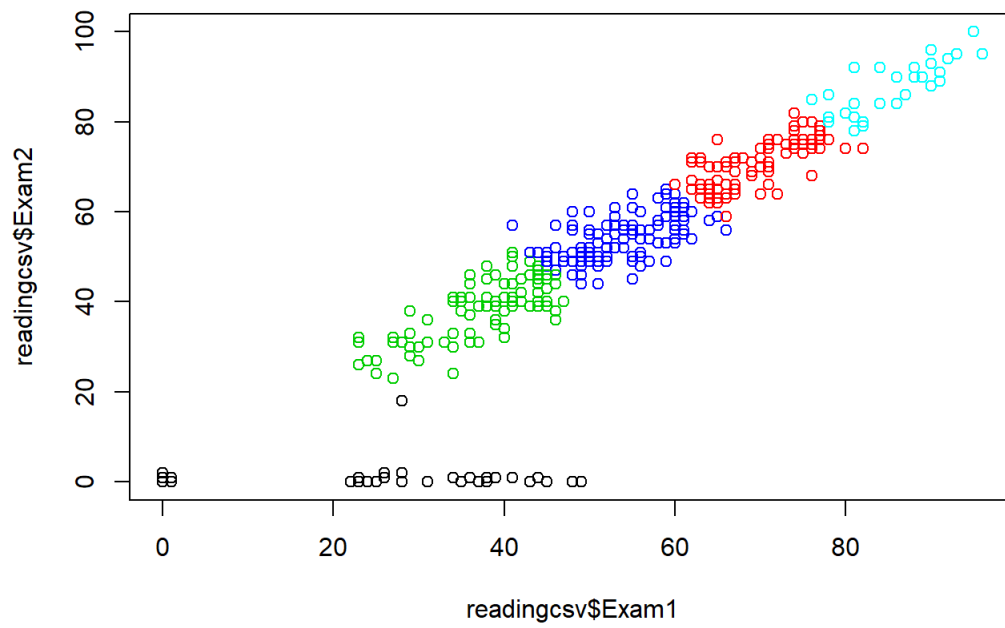
```
## [1] 4
```

**K = 4**



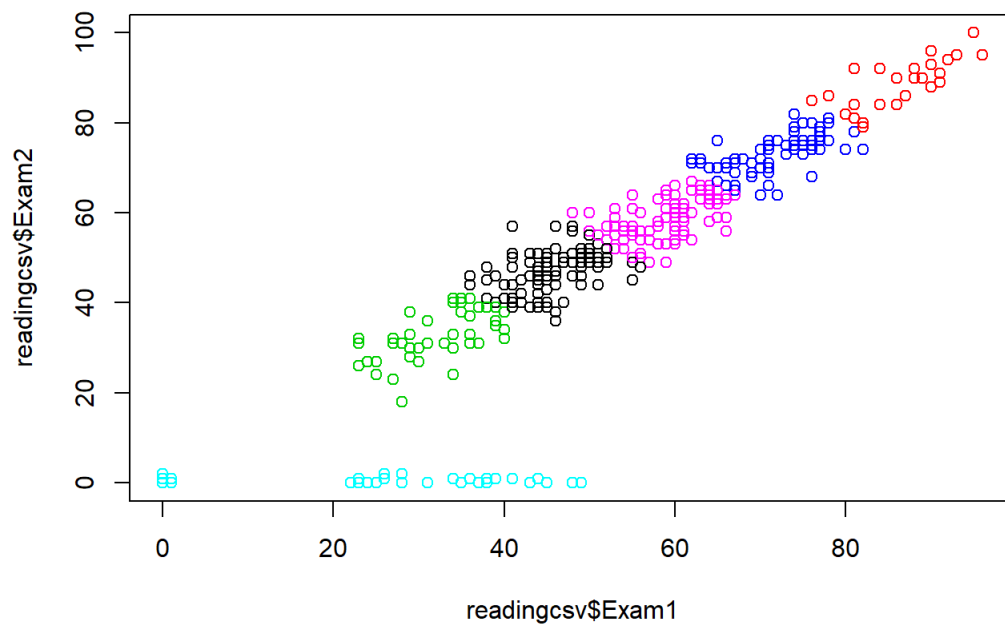
```
## [1] 5
```

**K = 5**



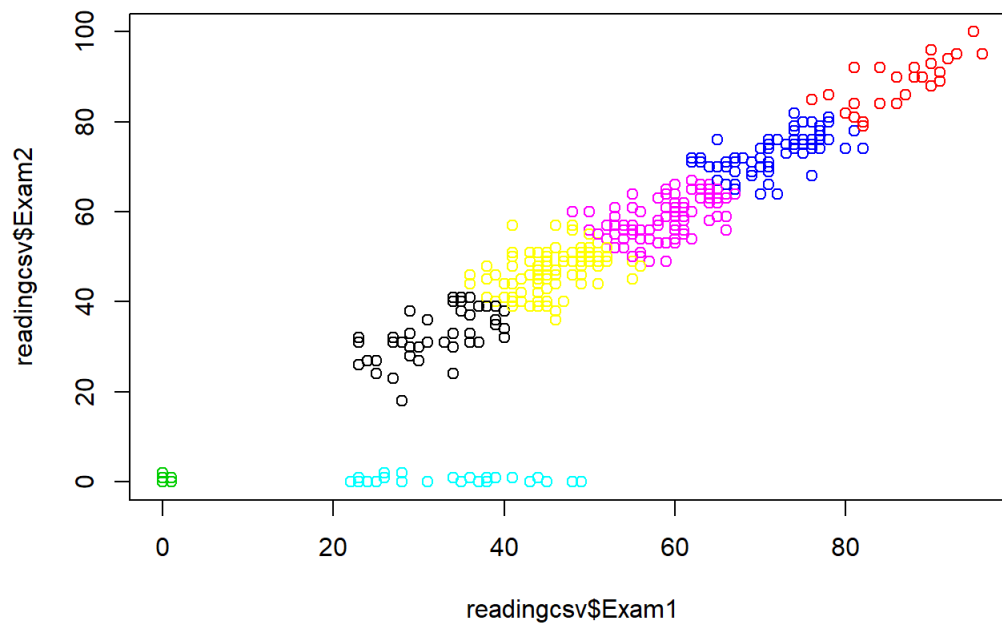
```
## [1] 6
```

**K = 6**



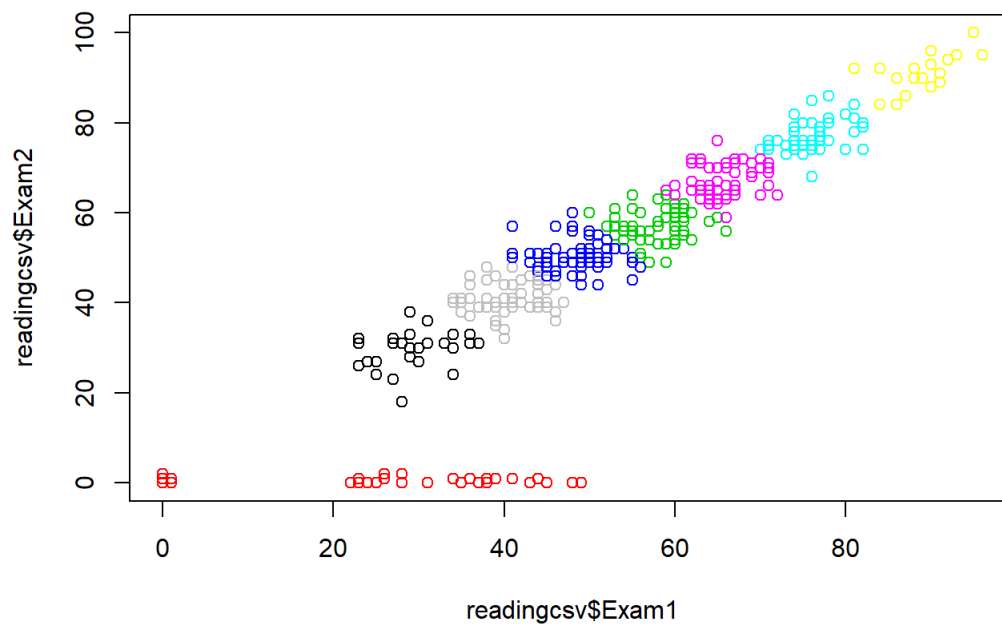
```
## [1] 7
```

**K = 7**



```
## [1] 8
```

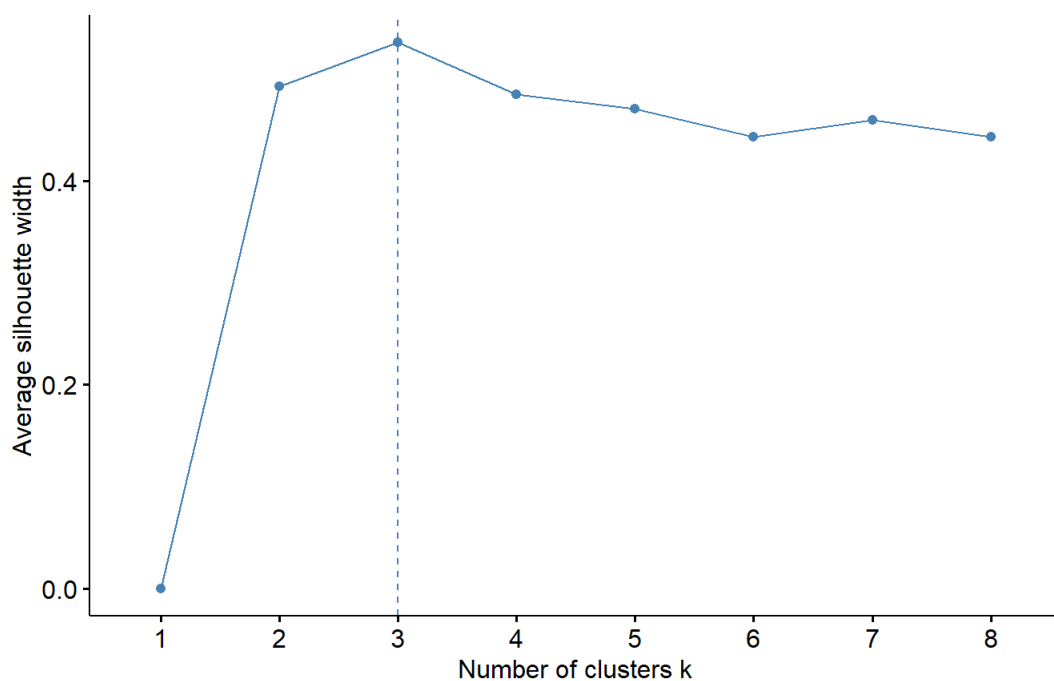
**K = 8**



```
fviz_nbclust(readingcsv, kmeans, method = "silhouette", k.max = 8)+labs(subtitle = "Silhouette method")
```

## Optimal number of clusters

Silhouette method

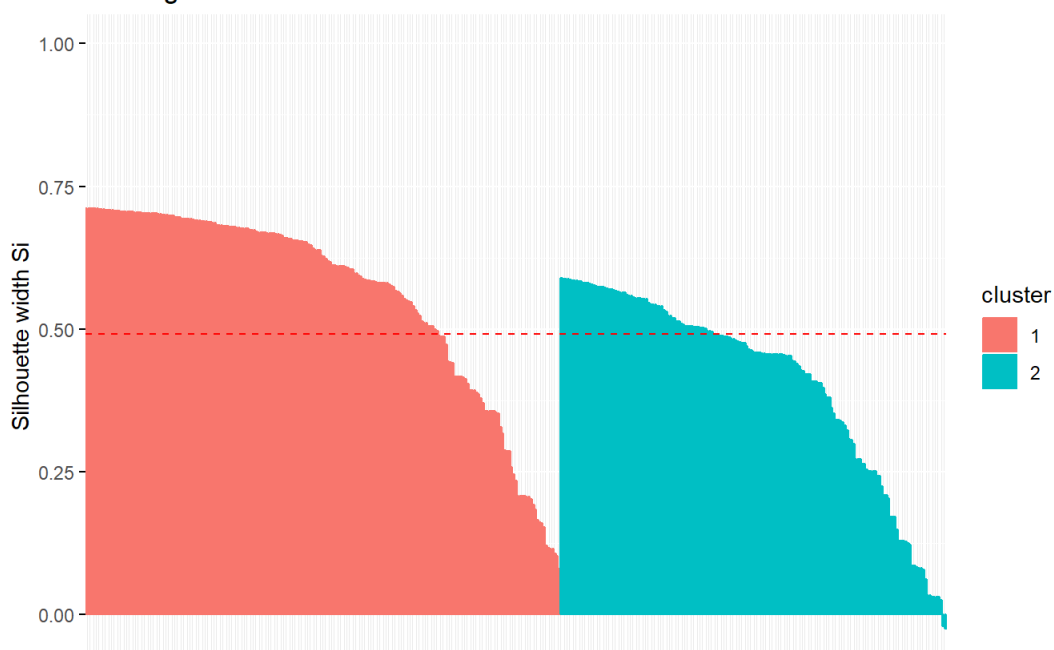


```
for (k in 2:8){  
  tryCatch({  
    print(k)  
    dfCluster <- eclust(readingscsv, "kmeans", k = k, nstart = 25, graph = FALSE)  
  },  
  warning=function(w) {done <- FALSE})  
  print(fviz_silhouette(dfCluster))  
}
```

```
## [1] 2  
##   cluster size ave.sil.width  
## 1      1  218         0.56  
## 2      2  177         0.41
```

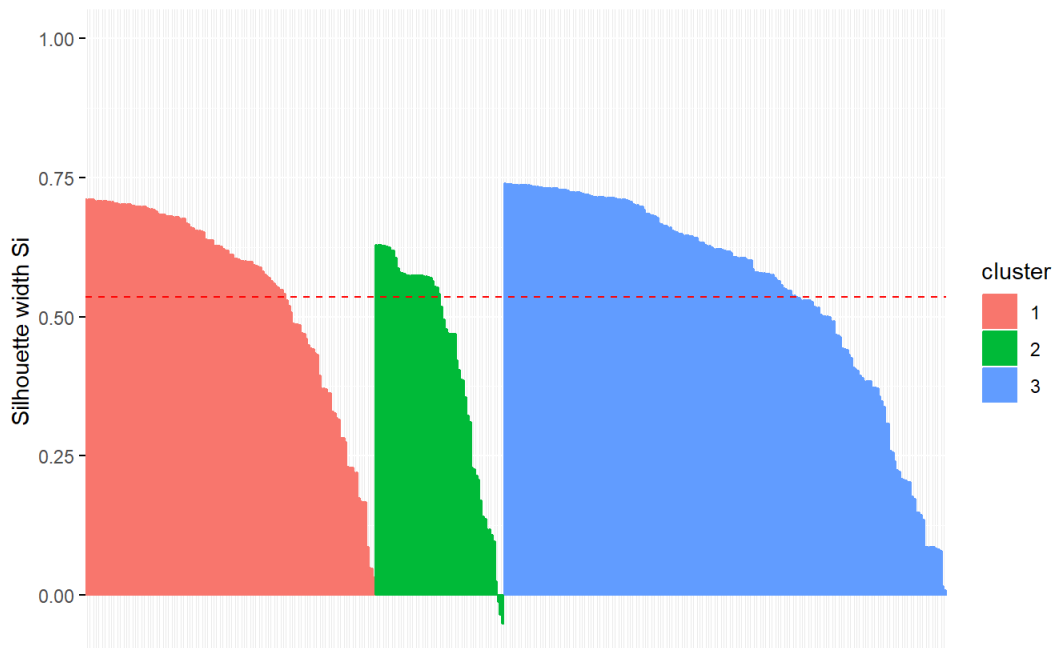
## Clusters silhouette plot

Average silhouette width: 0.49



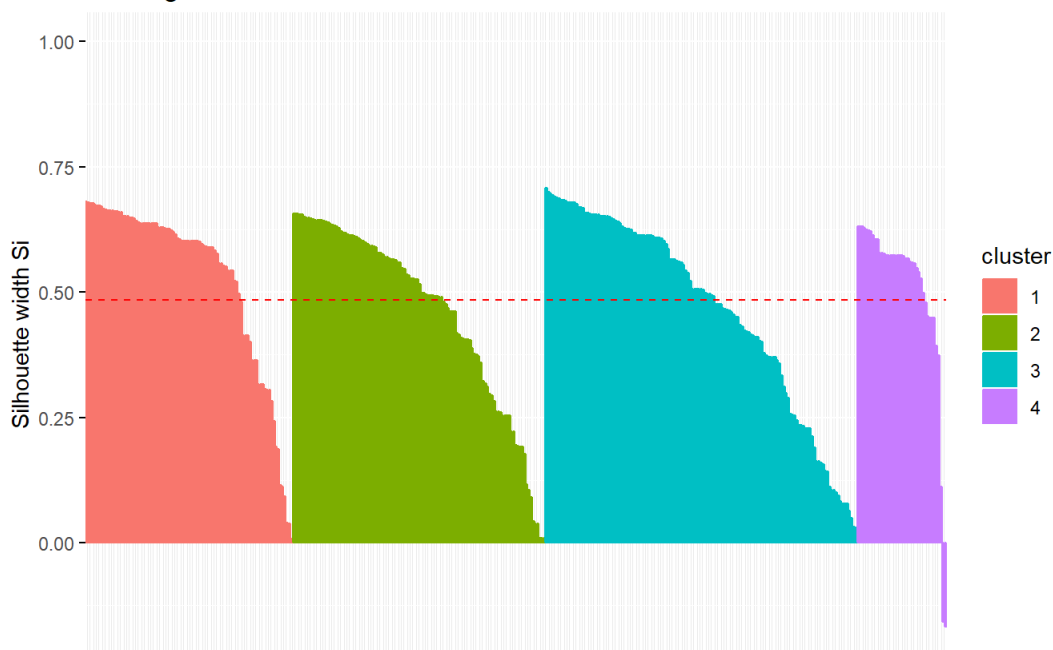
```
## [1] 3
##   cluster size ave.sil.width
## 1      1  133      0.55
## 2      2   59      0.43
## 3      3  203      0.55
```

Clusters silhouette plot  
Average silhouette width: 0.54



```
## [1] 4
##   cluster size ave.sil.width
## 1      1   95      0.53
## 2      2  116      0.46
## 3      3  143      0.47
## 4      4   41      0.51
```

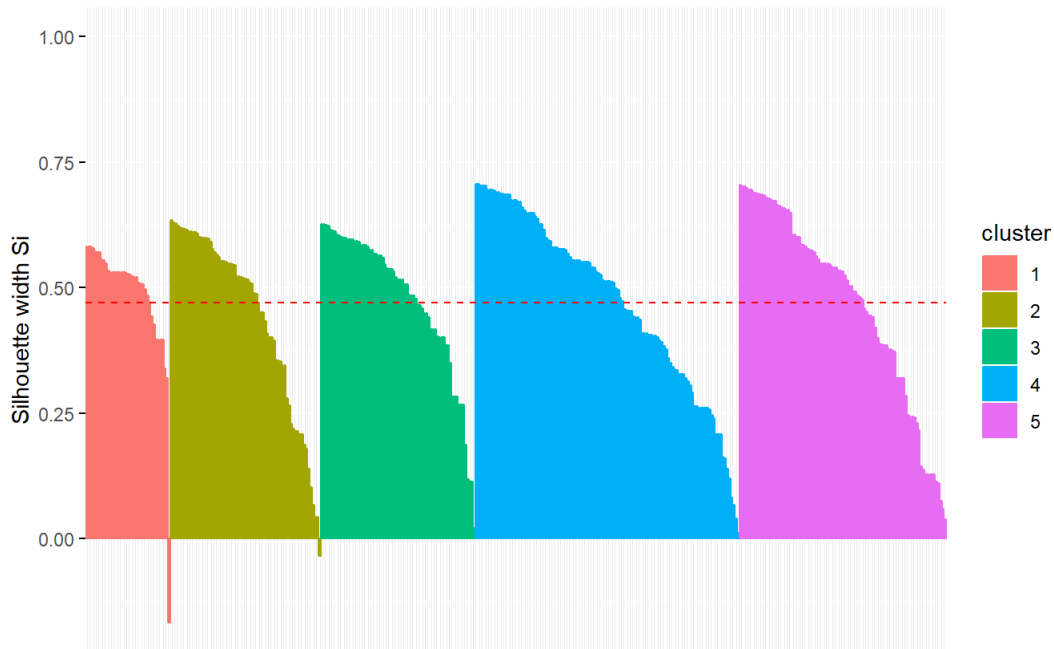
Clusters silhouette plot  
Average silhouette width: 0.48





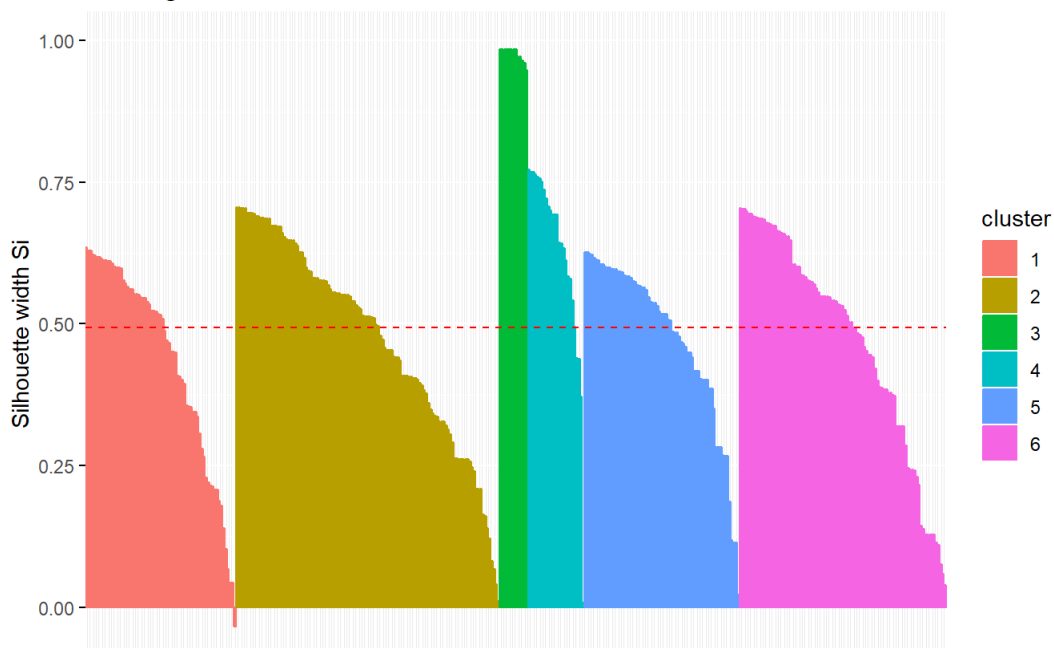
```
## [1] 5
##   cluster size ave.sil.width
## 1      1   39      0.49
## 2      2   69      0.44
## 3      3   71      0.48
## 4      4  121      0.47
## 5      5   95      0.47
```

Clusters silhouette plot  
Average silhouette width: 0.47



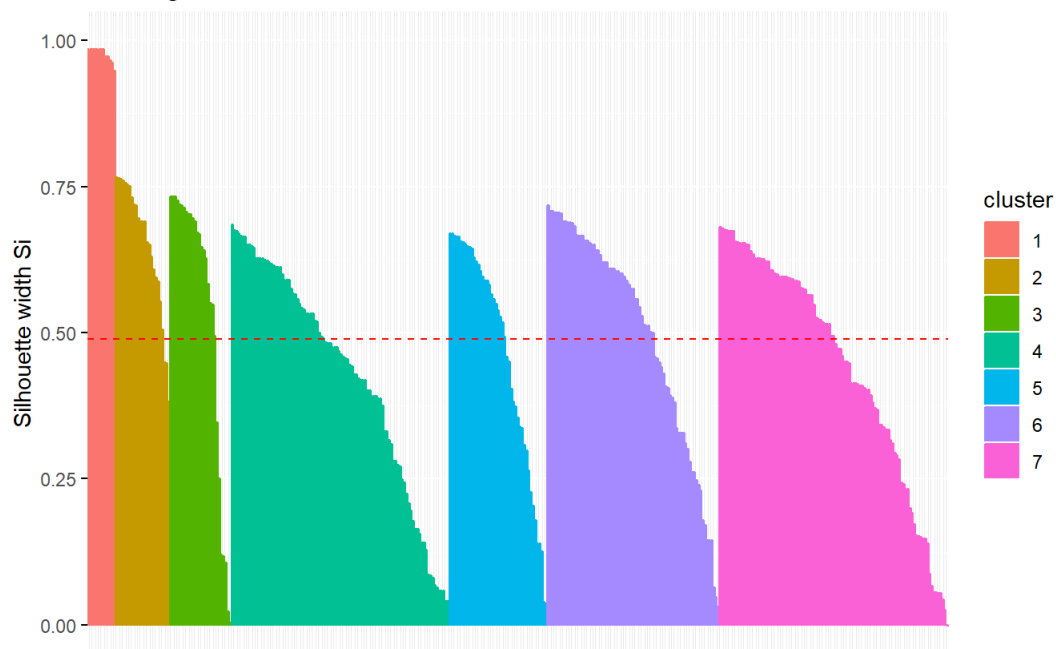
```
## [1] 6
##   cluster size ave.sil.width
## 1      1   69      0.44
## 2      2  121      0.47
## 3      3   13      0.98
## 4      4   26      0.63
## 5      5   71      0.48
## 6      6   95      0.47
```

Clusters silhouette plot  
Average silhouette width: 0.49



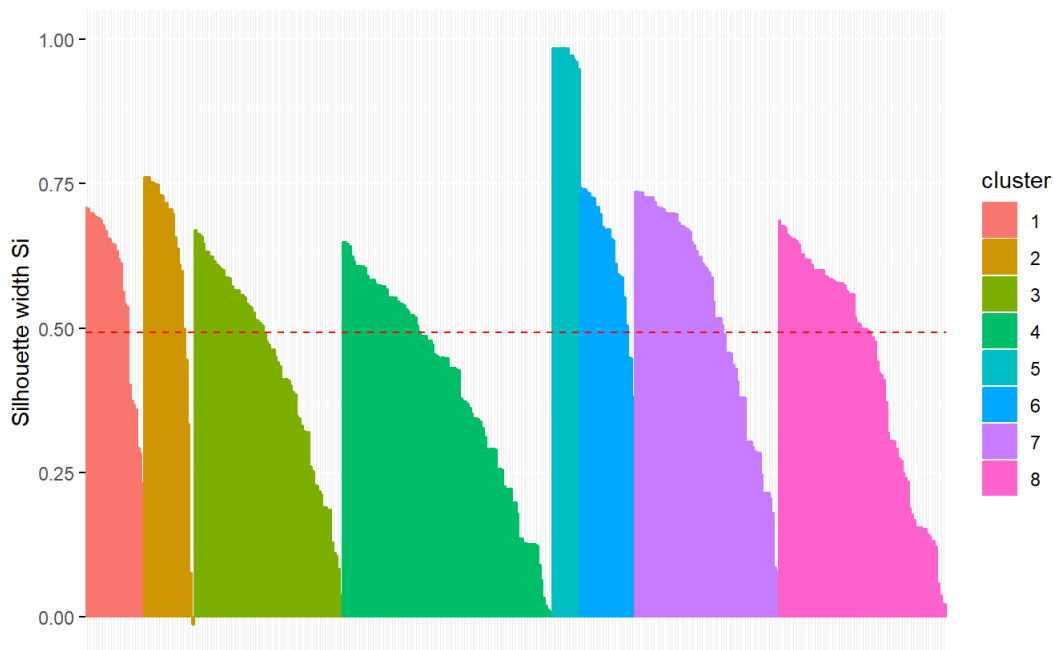
```
## [1] 7
##   cluster size ave.sil.width
## 1      1   13      0.98
## 2      2   25      0.65
## 3      3   28      0.53
## 4      4  100      0.43
## 5      5   45      0.46
## 6      6   79      0.50
## 7      7  105      0.44
```

Clusters silhouette plot  
Average silhouette width: 0.49



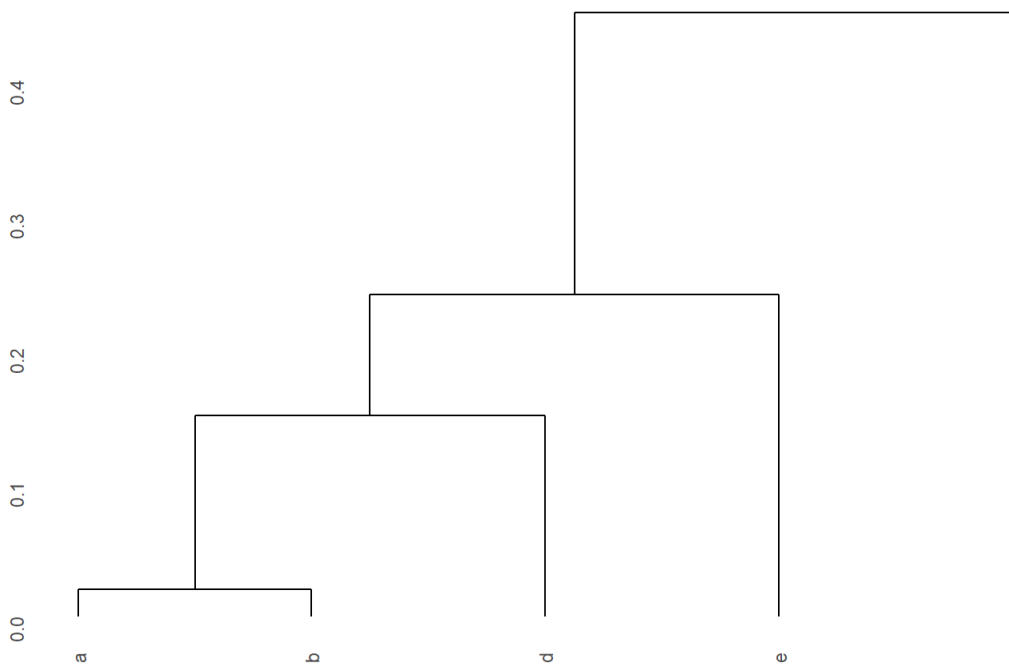
```
## [1] 8
##   cluster size ave.sil.width
## 1      1   27      0.57
## 2      2   23      0.61
## 3      3   68      0.44
## 4      4   96      0.40
## 5      5   13      0.98
## 6      6   25      0.64
## 7      7   66      0.53
## 8      8   77      0.43
```

Clusters silhouette plot  
Average silhouette width: 0.49



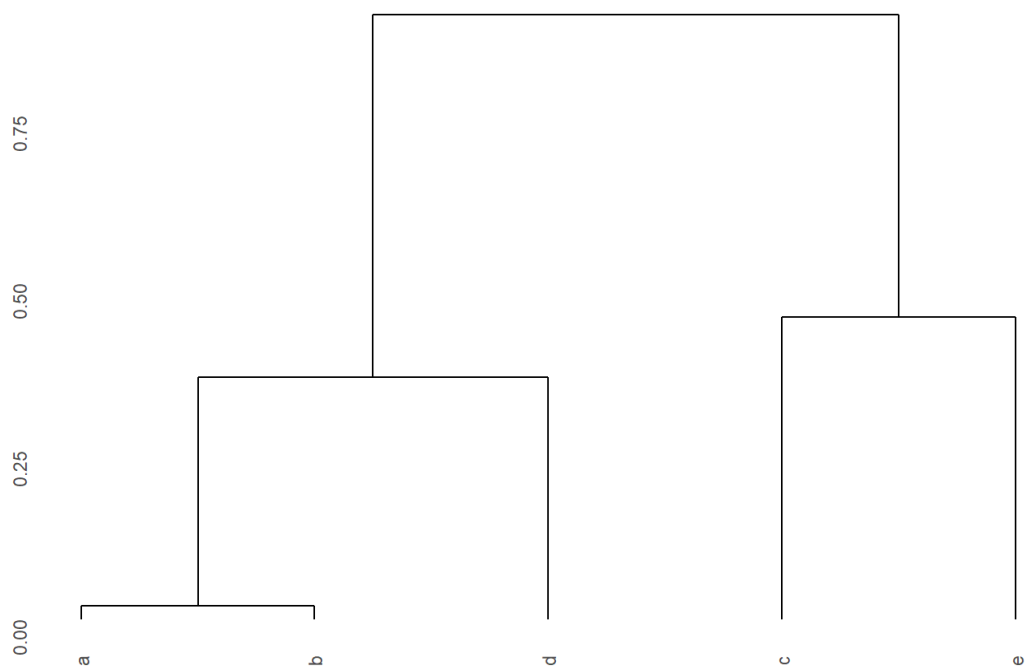
```
matrixdist <- matrix(
  c(0.00,0.02,0.9,0.36,0.53,0.02,0.00,0.65,0.15,0.24,0.9,0.65,0.00,0.59,0.45,
    0.36,0.15,0.59,0.00,0.56,0.53,0.24,0.450,0.56,0.00),
  nrow=5,
  ncol=5,byrow = "TRUE")
rownames(matrixdist) <- letters[seq( from = 1, to = 5 )]
colnames(matrixdist) <- letters[seq( from = 1, to = 5 )]
#Single Linkage
hclust_single <- hclust(as.dist(matrixdist), method = "single")
dend1 <- as.dendrogram(hclust_single)
dend2 <- reorder(dend1,10:1)
dend3 <- rev(dend2)
ggdendrogram(dend3, rotate=FALSE, size=0.5) + labs(title="Single Linkage")
```

Single Linkage



```
#Complete Linkage
hclust_single <- hclust(as.dist(matrixdist), method = "complete")
dend1 <- as.dendrogram(hclust_single)
dend2 <- reorder(dend1,10:1)
dend3 <- rev(dend2)
ggdendrogram(dend3, rotate=FALSE, size=0.5) + labs(title="Complete Linkage")
```

## Complete Linkage



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.