NumberOfClusters

Nazanin Zounemat Kermani

January 25, 2018

```
pathTodata = "C:/Users/nz1413/Desktop/"
PathToReport = "C:/Users/nz1413/Desktop/AnalysisTag/DiseaseSubTypeDiscovery/Final/report/"

data <- t(read.table(paste(pathTodata, "/dataTAC/sputum_508genes.txt", sep = ""),  header = TRUE))
dataScaled = scale(data)
```

```
library(factoextra)
```

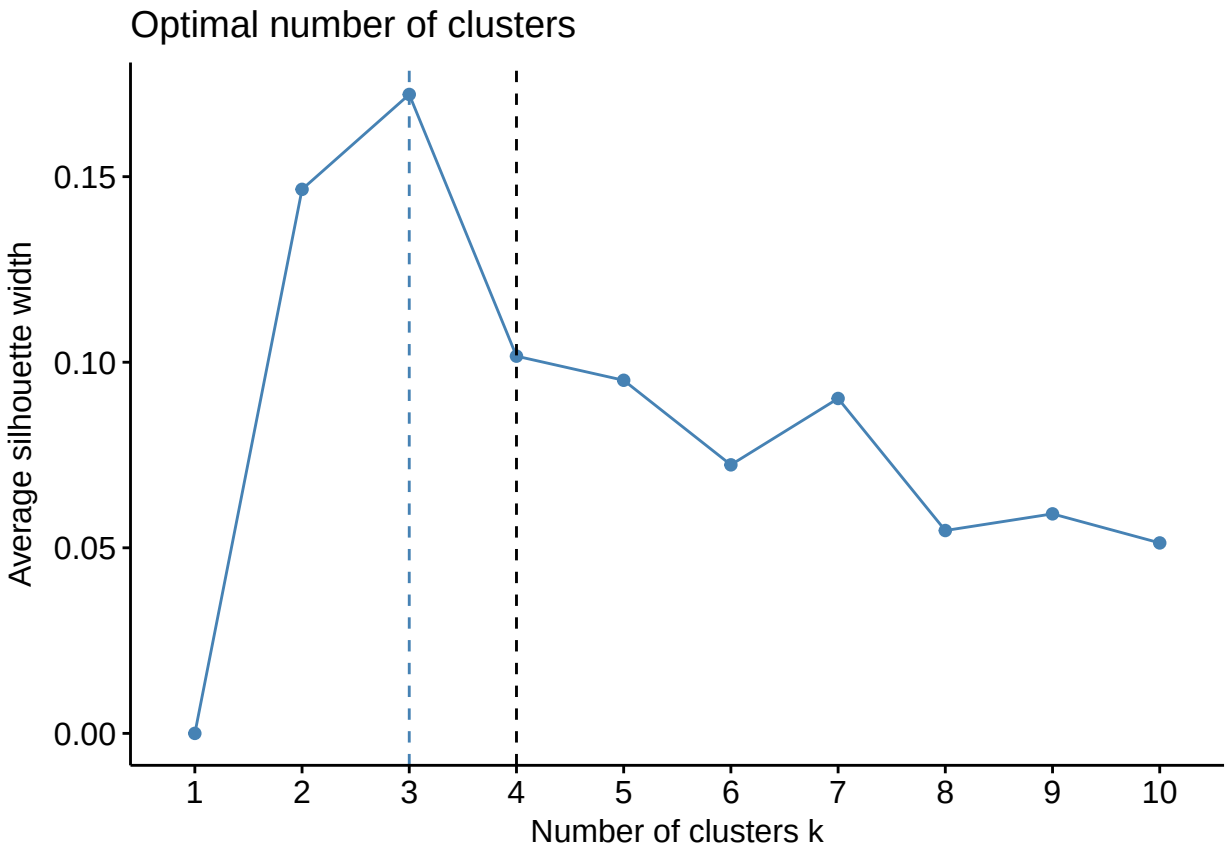
```
## Loading required package: ggplot2
```

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

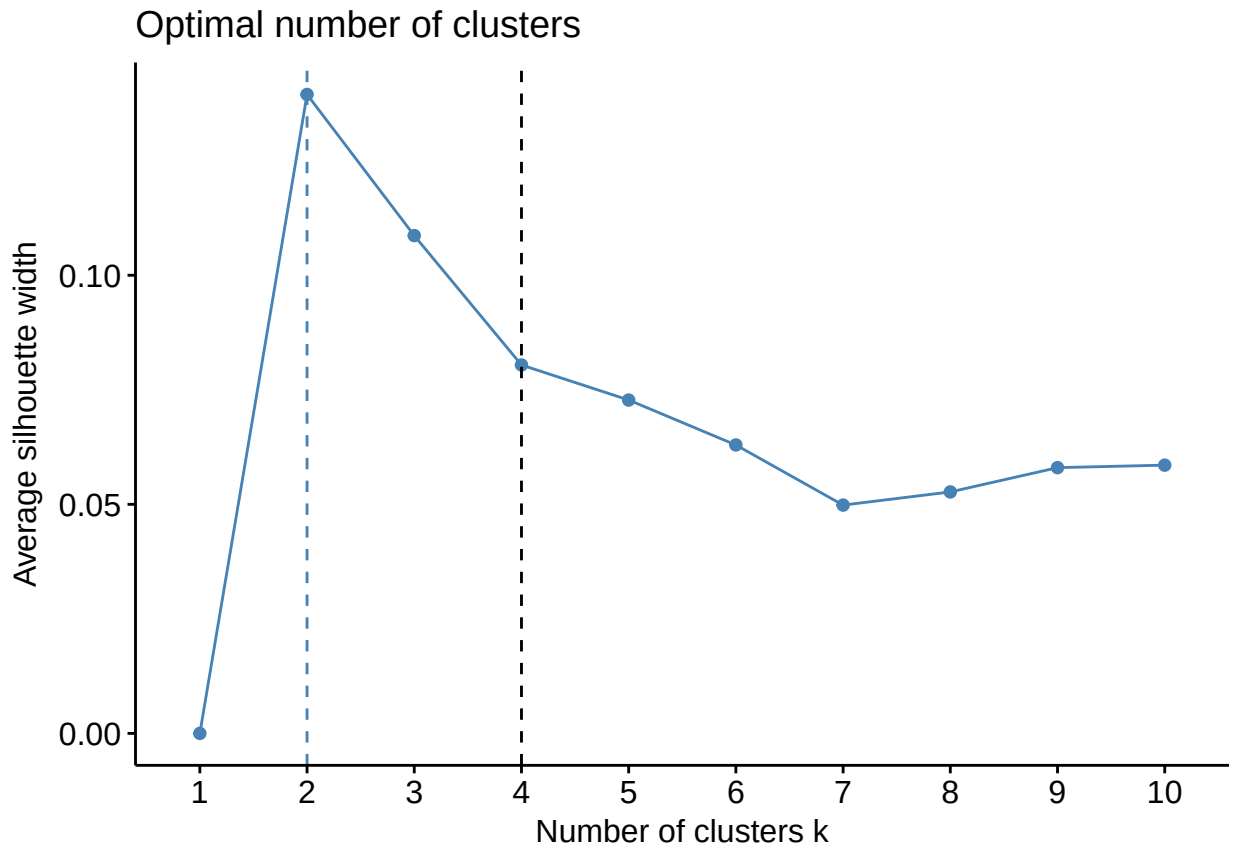
```
Methods= c("silhouette", "wss","gap_stat")
Clustering.method = c("kmeans", "hcut","pam")
pdf(paste(PathToReport , "numberofClusters.pdf", sep=""))
set.seed(123)
for(emethod in Methods)
{
  # for average silhouette width
fviz_nbclust(dataScaled, FUNcluster = kmeans, nstart=25,  method = emethod,  print.summary = TRUE)+
  geom_vline(xintercept = 4, linetype = 2)
}

 hc_method = "complete"

for(emethod in Methods)
{
  # for average silhouette width
fviz_nbclust(dataScaled, FUNcluster = hcut, nstart=25,  method = emethod,  print.summary = TRUE)+
  geom_vline(xintercept = 4, linetype = 2)
}

for(emethod in Methods)
{
  # for average silhouette width
fviz_nbclust(dataScaled, FUNcluster = cluster::pam,   method = emethod,  print.summary = TRUE)
}
dev.off()
```

```
## pdf
##   2
```

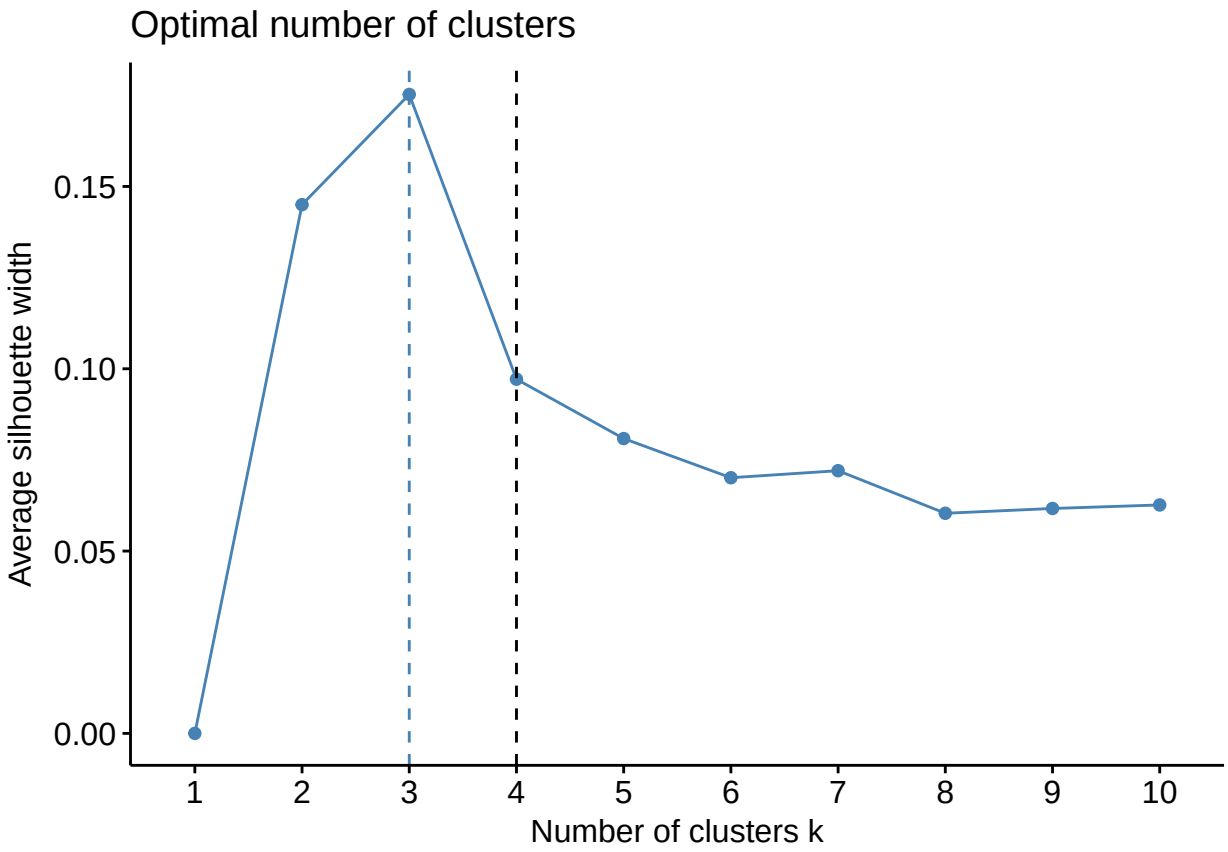
```
# for average silhouette width
fviz_nbclust(dataScaled, kmeans, nstart=25,  method = "silhouette")+
  geom_vline(xintercept = 4, linetype = 2)
```



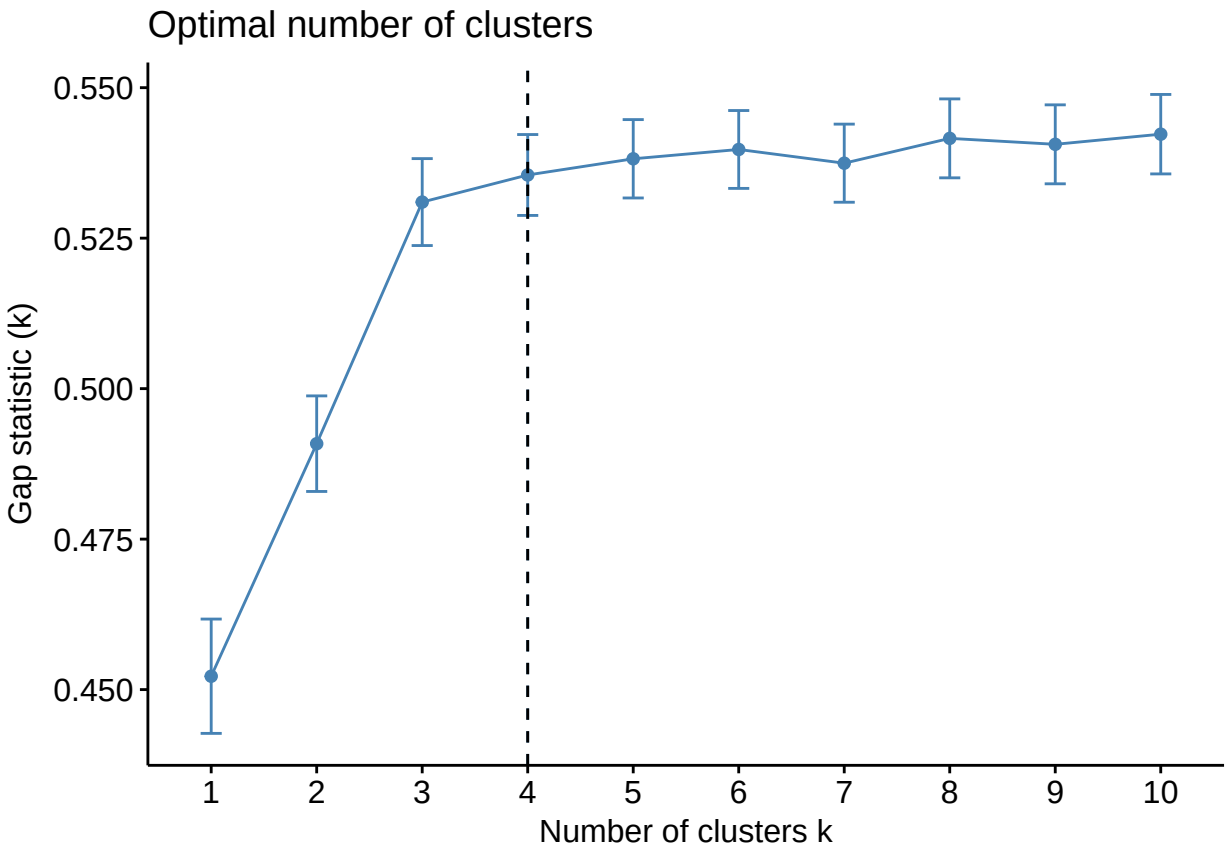

```
fviz_nbclust(dataScaled, cluster::pam, method = "silhouette")+
  geom_vline(xintercept = 4, linetype = 2)
```



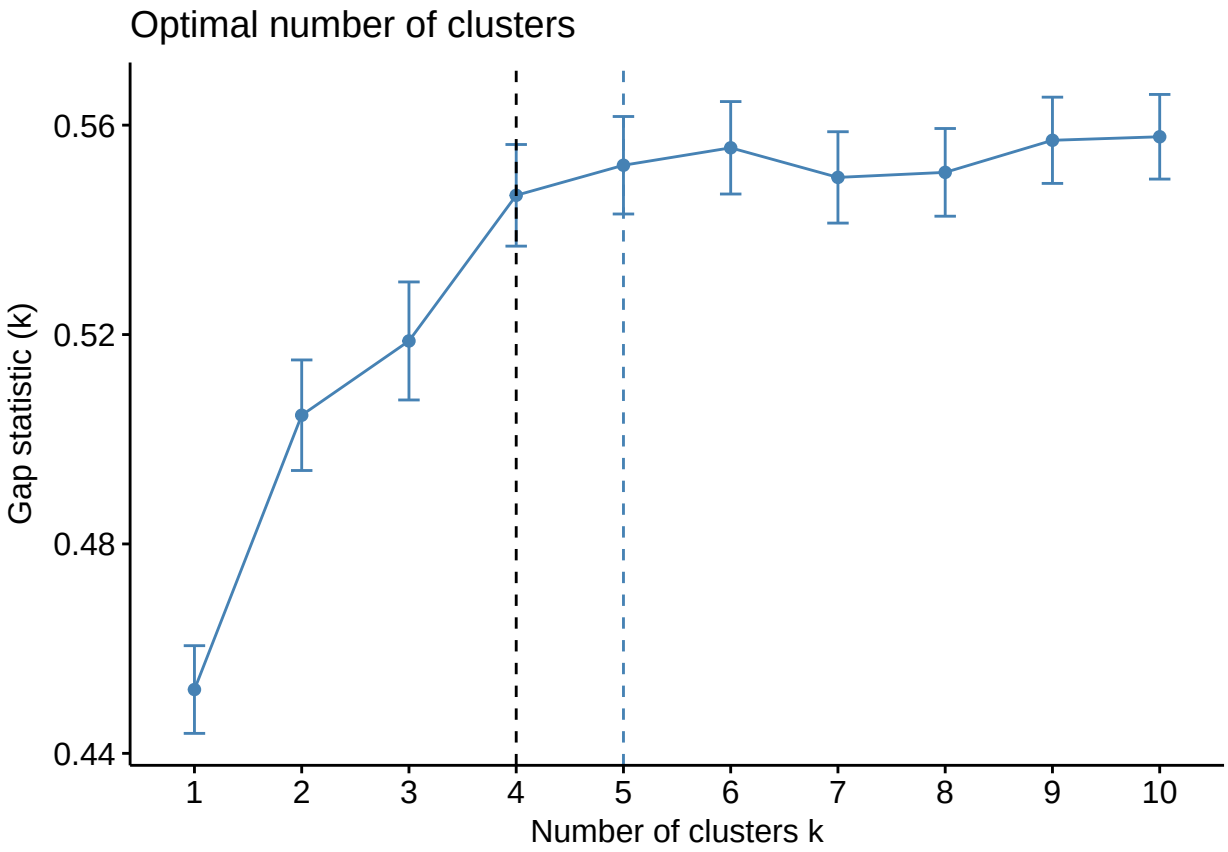

```
fviz_nbclust(dataScaled, hcut, method = "silhouette")+
  geom_vline(xintercept = 4, linetype = 2)
```



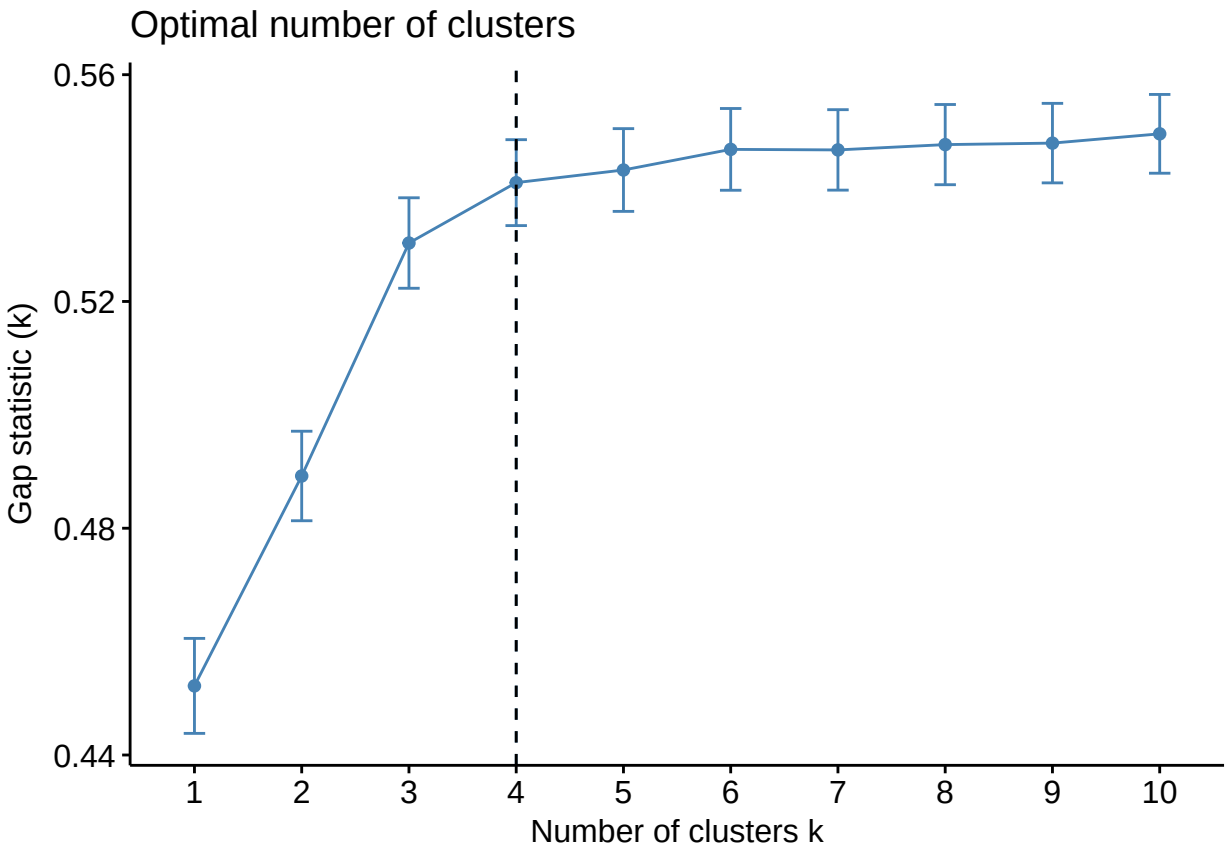

```
# for gap statistics
fviz_nbclust(dataScaled, kmeans,nstart=25, method = "gap_stat")+
  geom_vline(xintercept = 4, linetype = 2)
```




```
fviz_nbclust(dataScaled, cluster::pam, method = "gap_stat")+
  geom_vline(xintercept = 4, linetype = 2)
```




```
fviz_nbclust(dataScaled, hcut, method = "gap_stat")+
  geom_vline(xintercept = 4, linetype = 2)
```


Optimal number of clusters
Gap statistic (k)
0.56
0.52
0.48
0.44
1
2
3
4
5
6
7
8
9
10
Number of clusters k