

## Stat517\_project\_2

Abhinav

September 29, 2018

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.4.4

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

## v ggplot2 3.0.0      v purrr   0.2.5
## v tibble  1.4.2      v dplyr   0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.4.4
## Warning: package 'tibble' was built under R version 3.4.4
## Warning: package 'tidyr' was built under R version 3.4.4
## Warning: package 'readr' was built under R version 3.4.4
## Warning: package 'purrr' was built under R version 3.4.4
## Warning: package 'dplyr' was built under R version 3.4.4
## Warning: package 'stringr' was built under R version 3.4.4
## Warning: package 'forcats' was built under R version 3.4.4

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

library(dendroextras)

## Warning: package 'dendroextras' was built under R version 3.4.4

##
## Attaching package: 'dendroextras'

## The following object is masked from 'package:dplyr':
##
##     slice

library(dendextend)
```

```

## Warning: package 'dendextend' was built under R version 3.4.4

##
## -----
## Welcome to dendextend version 1.8.0
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## Or contact: <tal.galili@gmail.com>
##
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))
## -----

##
## Attaching package: 'dendextend'

## The following object is masked from 'package:dendroextras':
##
##   labels<-

## The following object is masked from 'package:stats':
##
##   cutree

library(cluster)
library(circlize)

## Warning: package 'circlize' was built under R version 3.4.4

## =====
## circlize version 0.4.4
## CRAN page: https://cran.r-project.org/package=circlize
## Github page: https://github.com/jokergoo/circlize
## Documentation: http://jokergoo.github.io/circlize\_book/book/
##
## If you use it in published research, please cite:
## Gu, Z. circlize implements and enhances circular visualization
## in R. Bioinformatics 2014.
## =====

library(mclust)

## Warning: package 'mclust' was built under R version 3.4.4

## Package 'mclust' version 5.4.1
## Type 'citation("mclust")' for citing this R package in publications.

```

```
##
## Attaching package: 'mclust'

## The following object is masked from 'package:purrr':
##
##      map

library(factoextra)

## Warning: package 'factoextra' was built under R version 3.4.4

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

library(MVA)

## Warning: package 'MVA' was built under R version 3.4.4

## Loading required package: HSAUR2

## Warning: package 'HSAUR2' was built under R version 3.4.4

## Loading required package: tools

library(NbClust)
library(seriation)

## Warning: package 'seriation' was built under R version 3.4.4
```

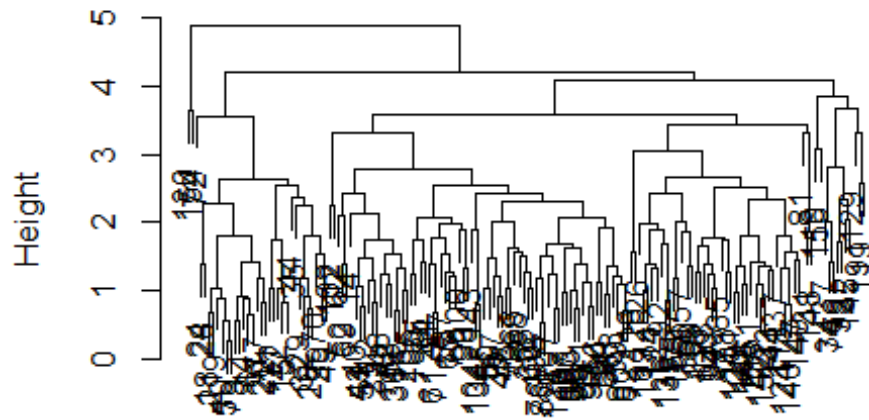
For the year 2015

```
happy1 = read.csv("https://raw.githubusercontent.com/sauchilee/Stat517/master/Data/World_Happiness_2015.csv")
dim(happy1)

## [1] 158 12

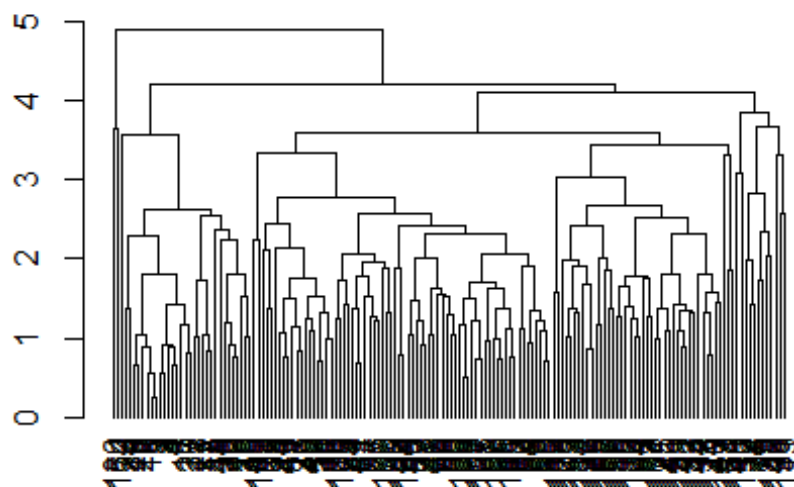
happy1dup<-happy1[,c(2,6:12)]
happy1dup.s = scale(happy1dup[, -1])
#happy1.s
happy1dup.d = dist(happy1dup.s)
#happy1.d
happy1dup.hc.s = hclust(happy1dup.d,method="average")
plot(happy1dup.hc.s)
```

## Cluster Dendrogram



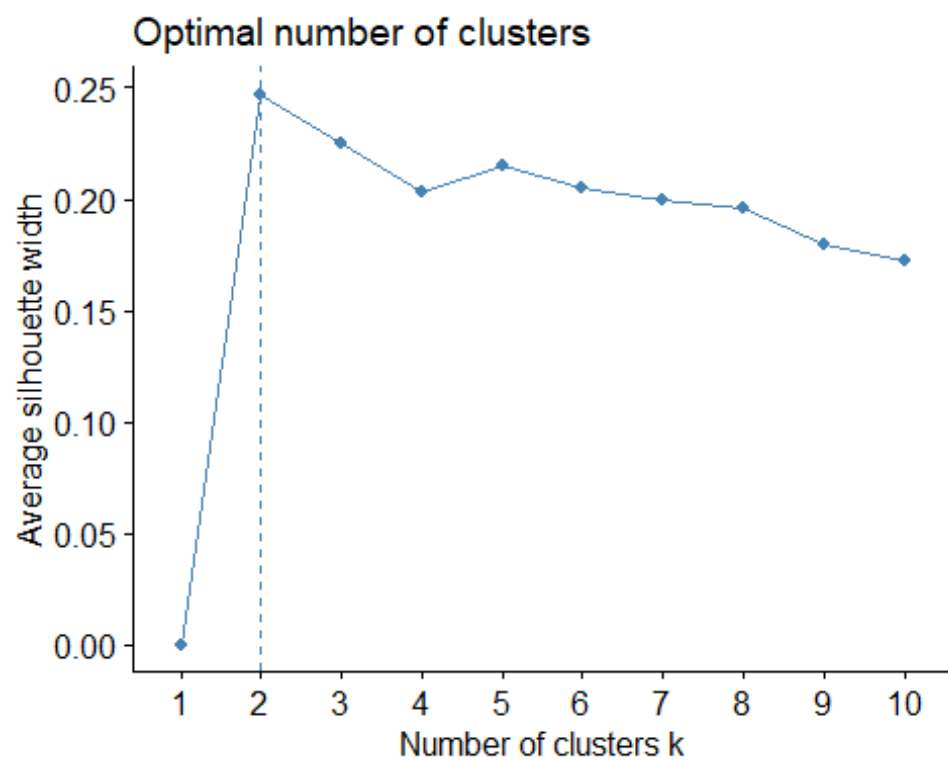
happy1dup.d  
hclust (\*, "average")

```
dend=as.dendrogram(happy1dup.hc.s)
#Labels_colors(dend)=as.numeric(as.factor(happy1dup$Region[happy1dup.hc.s$order]))
plot(dend)
```

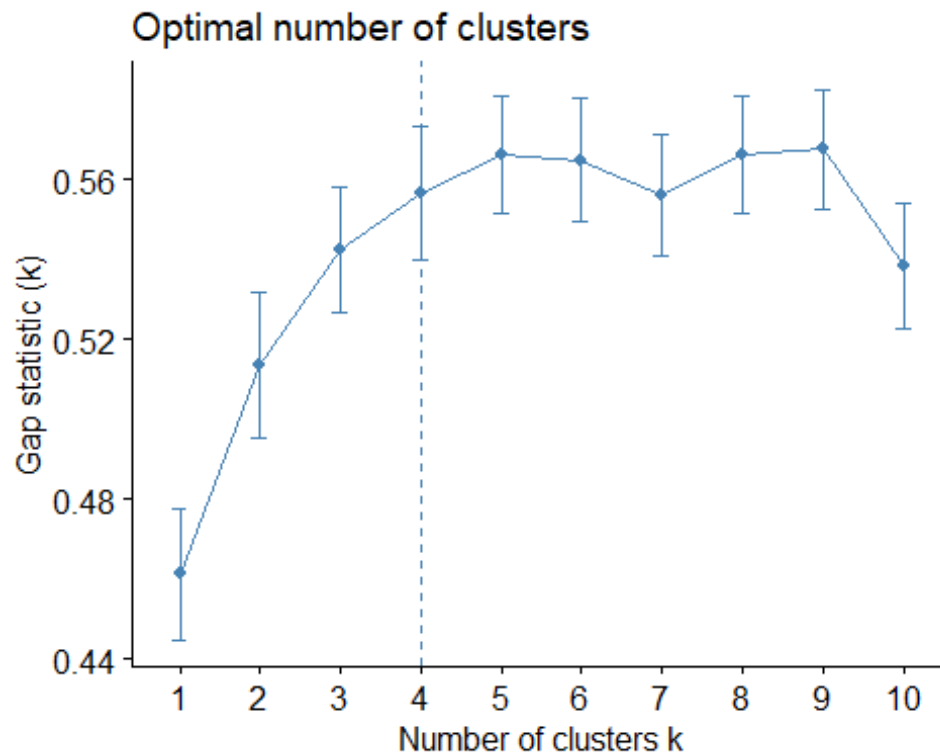


Partition Clustering

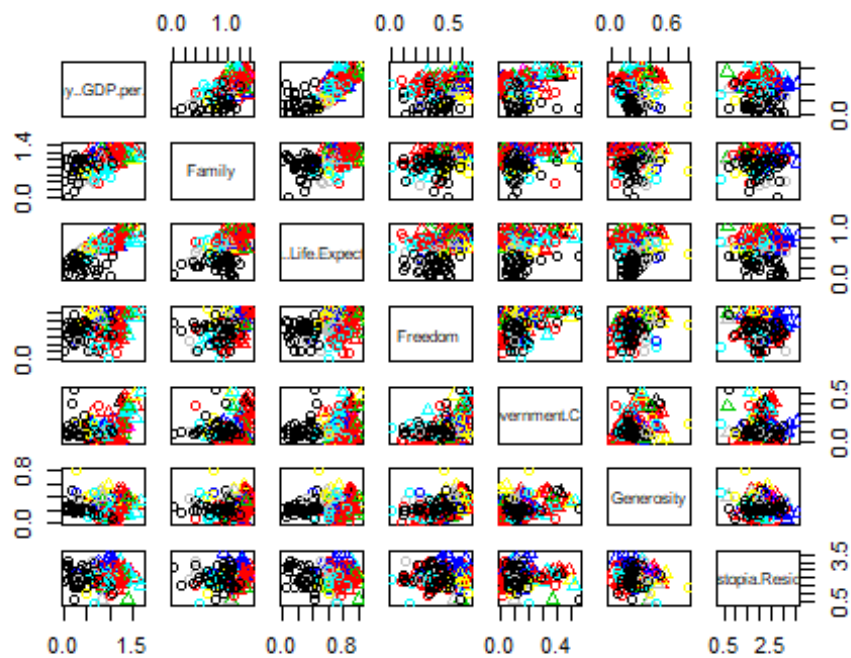
```
set.seed(150)
fviz_nbclust(happy1dup.s, kmeans, method="silhouette")
```



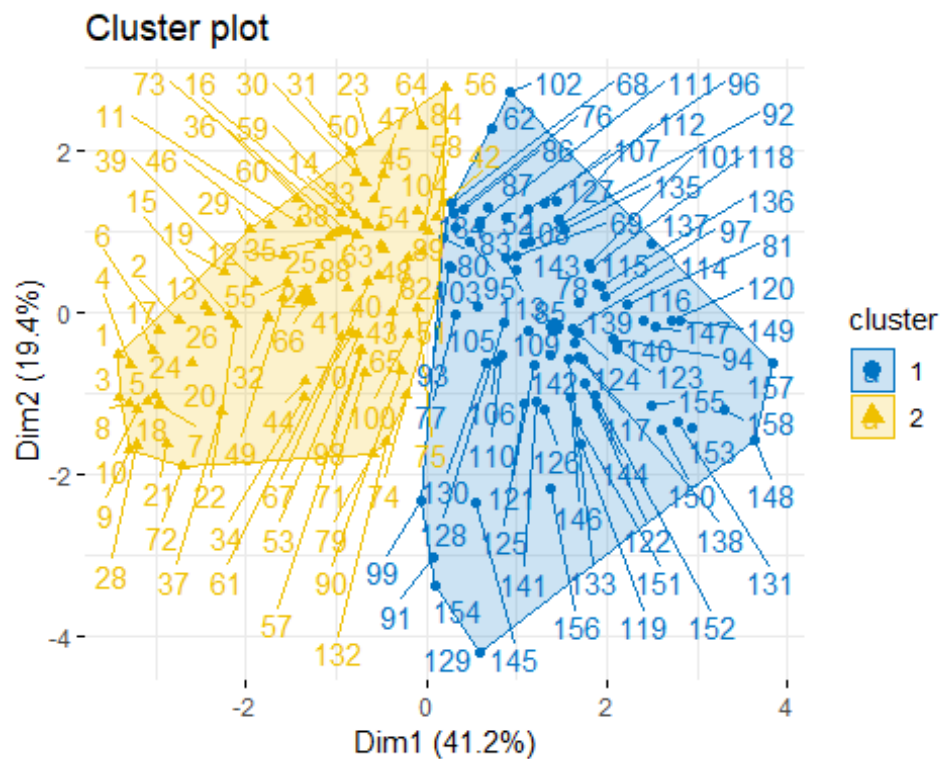
```
fviz_nbclust(happy1dup.s, kmeans, method="gap_stat")
```



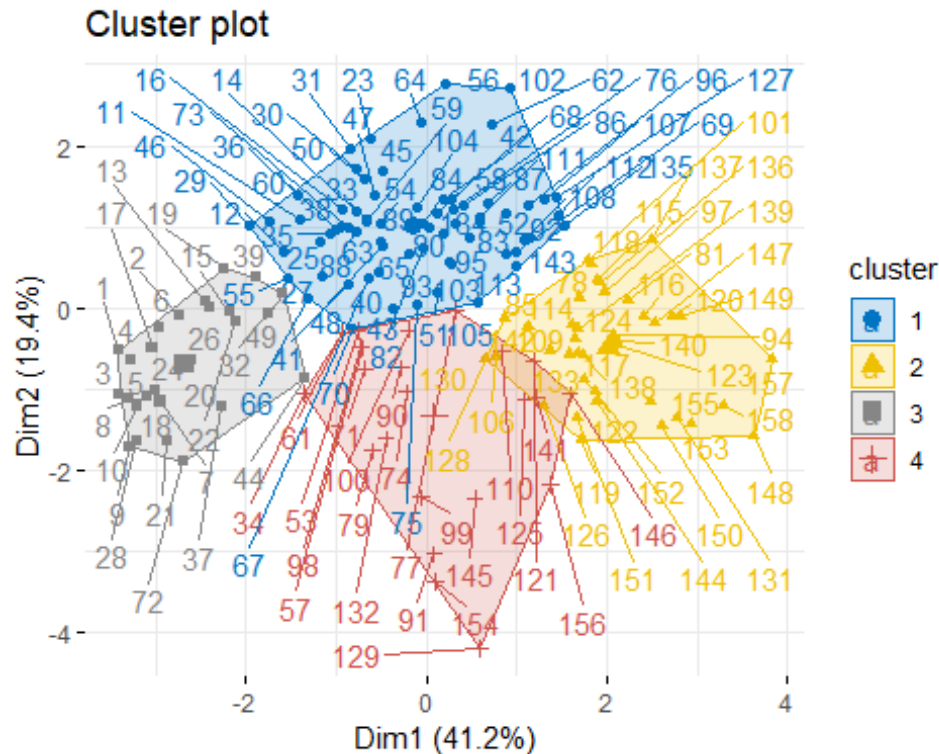
```
happy1dup.k2<-kmeans(happy1dup.s,centers = 2,iter.max=100,nstart=25)  
happy1dup.k4<-kmeans(happy1dup.s,centers = 4,iter.max=100,nstart=25)  
pairs(happy1dup[-1],pch=happy1dup.k2$cluster,col=unclass(happy1dup[,1]))
```



```
fviz_cluster(happy1dup.k2, data=happy1dup.s, ellipse.type = "convex", palette="jco", repel = TRUE, ggtheme = theme_minimal())
```



```
fviz_cluster(happy1dup.k4,data=happy1dup.s,ellipse.type ="convex",palette="jco",repel = TRUE, ggtheme = theme_minimal())
```



It is observed that most of the European countries are clustered in blue and the African countries are clustered in yellow regions.

M Clust:

```
happy1dup.mclust<-Mclust(happy1dup.s)
summary(happy1dup.mclust)

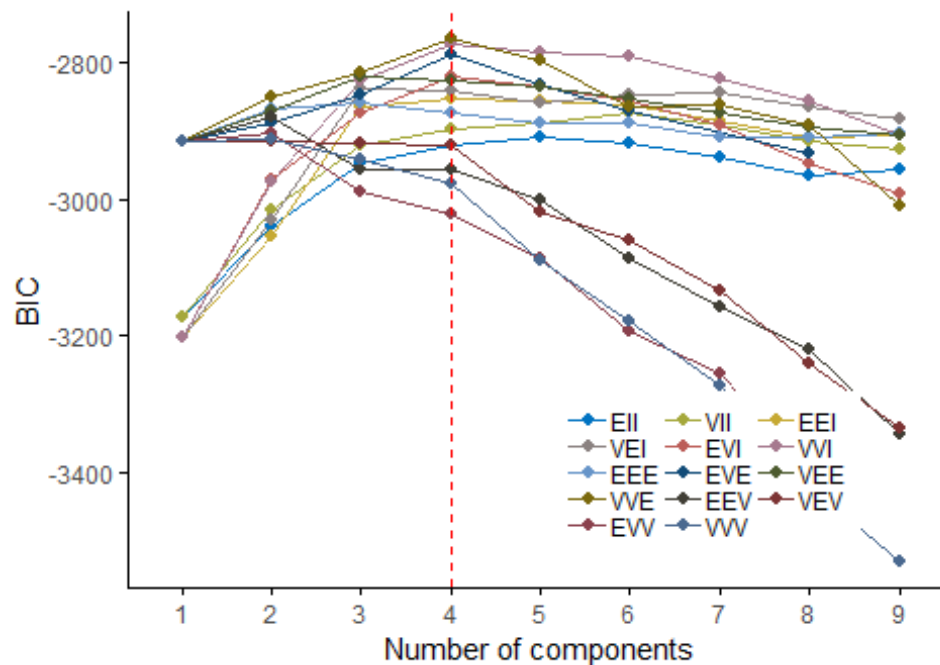
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVE (ellipsoidal, equal orientation) model with 4 components:
##
## log.likelihood    n df      BIC      ICL
##      -1179.659 158 80 -2764.326 -2776.977
##
## Clustering table:
##  1  2  3  4
## 19 34 65 40

fviz_mclust(happy1dup.mclust,"BIC",palette="jco")
```



## Model selection

Best model: VVE | Optimal clusters: n = 4



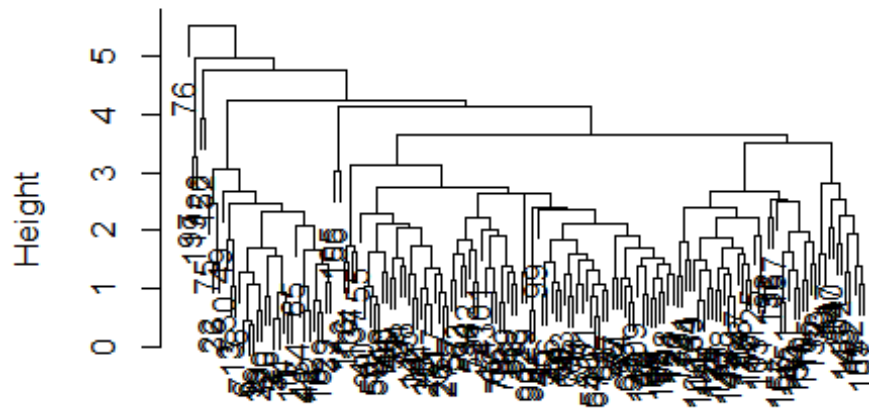
For the year 2016

```
happy2=read.csv("https://raw.githubusercontent.com/sauchilee/Stat517/master/Data/World_Happiness_2016.csv")
dim(happy2)

## [1] 157 13

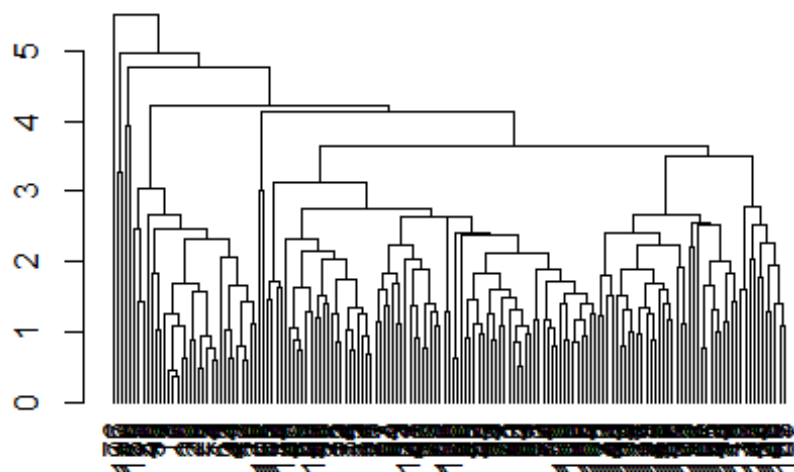
happy2dup<-happy2[,c(2,7:13)]
happy2dup.s = scale(happy2dup[, -1])
#happy1.s
happy2dup.d = dist(happy2dup.s)
#happy1.d
happy2dup.hc.s = hclust(happy2dup.d,method="average")
plot(happy2dup.hc.s)
```

## Cluster Dendrogram

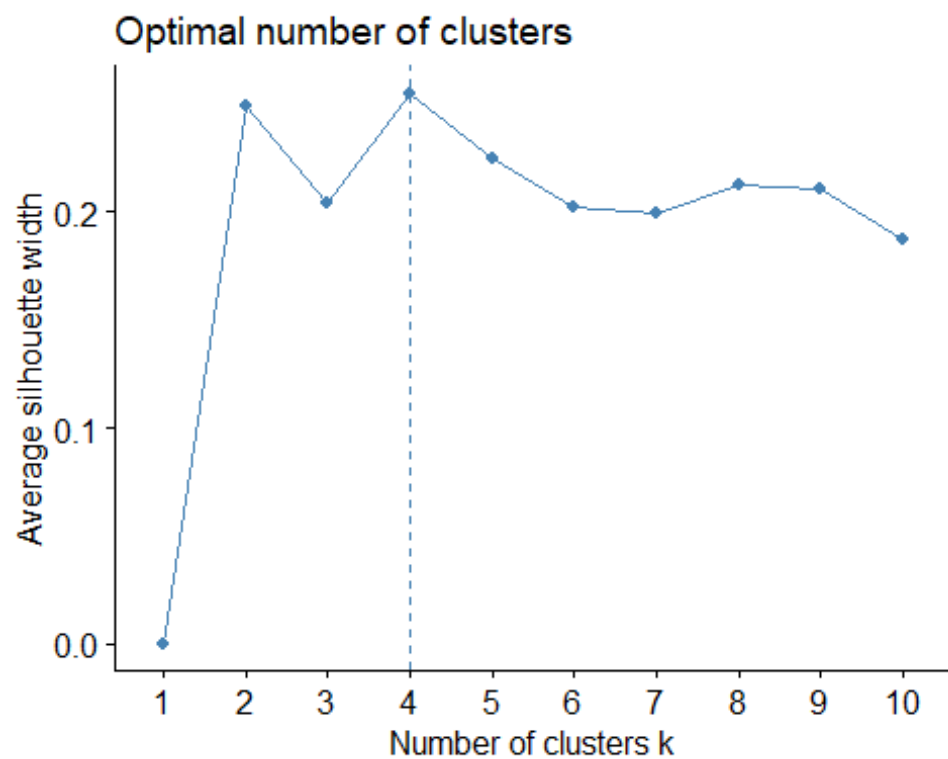


happy2dup.d  
hclust (\*, "average")

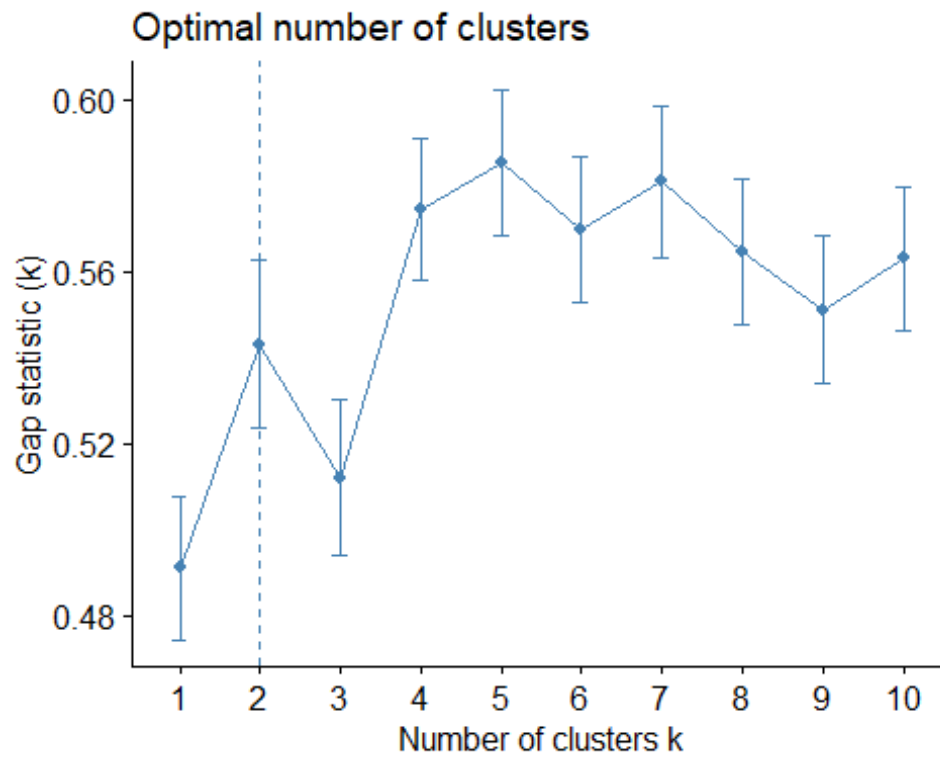
```
dend=as.dendrogram(happy2dup.hc.s)  
#Labels_colors(dend)=as.numeric(as.factor(happy2dup$Region[happy2dup.hc.s$order]))  
plot(dend)
```



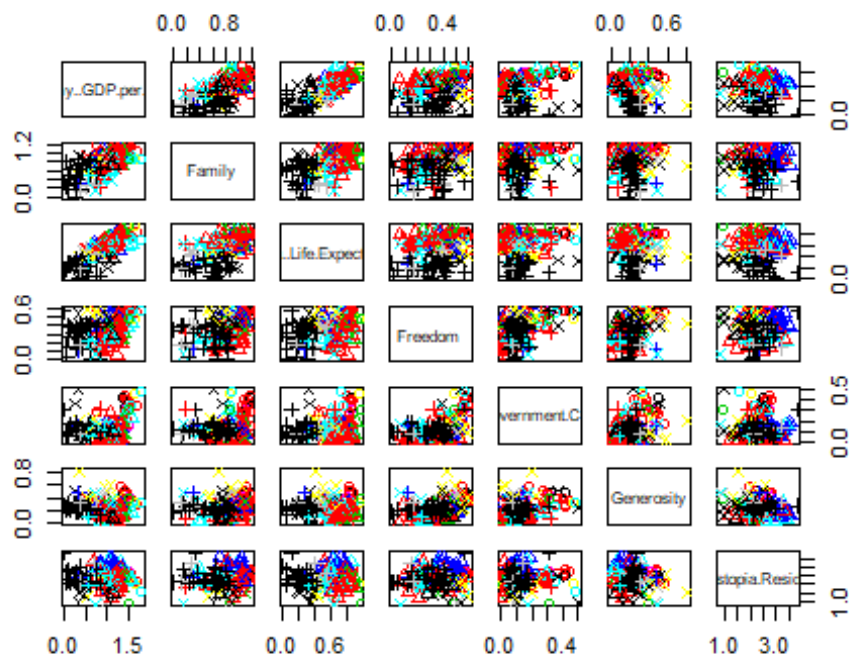
```
set.seed(150)
fviz_nbclust(happy2dup.s, kmeans, method="silhouette")
```



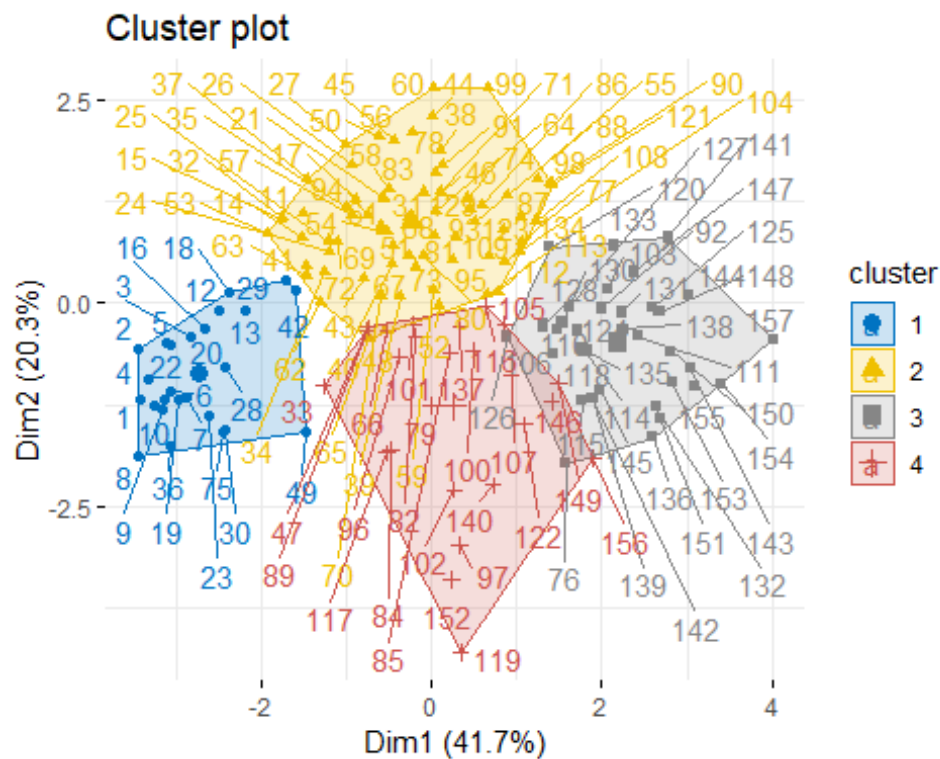
```
fviz_nbclust(happy2dup.s, kmeans, method="gap_stat")
```



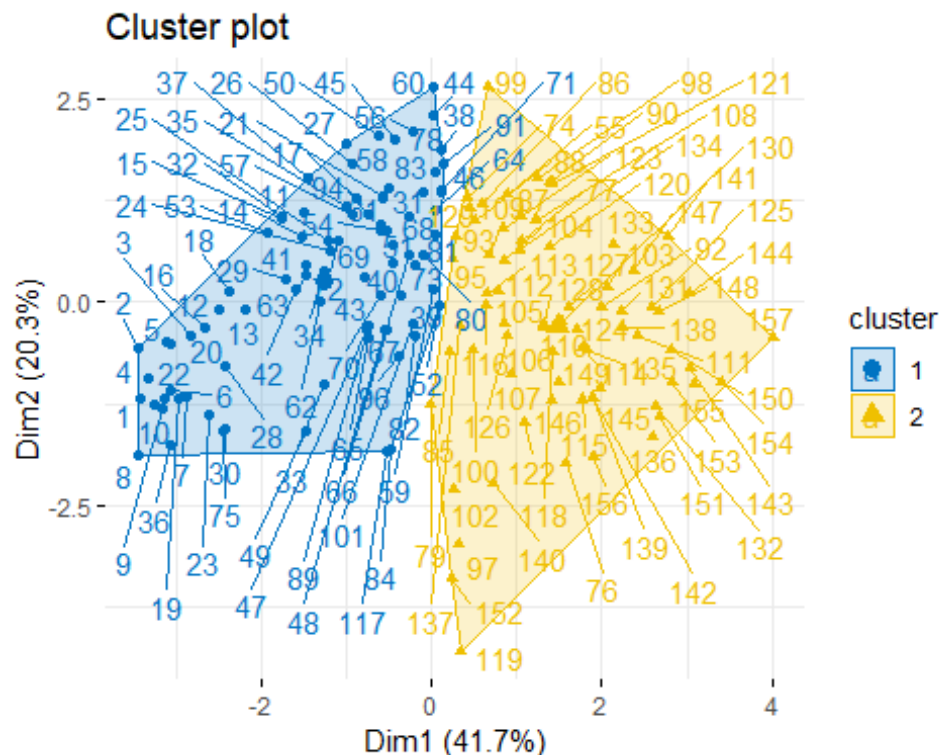
```
happy2dup.k2<-kmeans(happy2dup.s,centers = 4,iter.max=100,nstart=25)  
happy2dup.k4<-kmeans(happy2dup.s,centers = 2,iter.max=100,nstart=25)  
pairs(happy2dup[-1],pch=happy2dup.k2$cluster,col=unclass(happy2dup[,1]))
```



```
fviz_cluster(happy2dup.k2, data=happy2dup.s, ellipse.type = "convex", palette="jco",
repel = TRUE, ggtheme = theme_minimal())
```



```
fviz_cluster(happy2dup.k4,data=happy2dup.s,ellipse.type ="convex",palette="jco",repel = TRUE, ggtheme = theme_minimal())
```



It is observed that

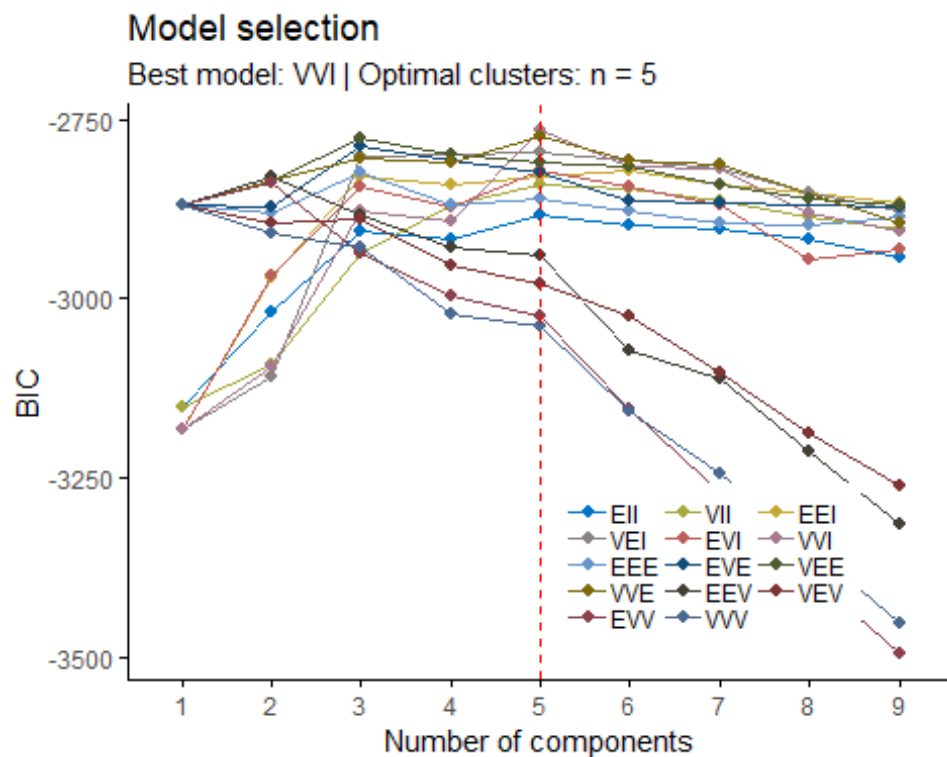
most of the African countries are clustered in black region.

M Clust:

```
happy2dup.mclust<-Mclust(happy2dup.s)
summary(happy2dup.mclust)

## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVI (diagonal, varying volume and shape) model with 5 components:
##
## log.likelihood    n df      BIC      ICL
##      -1196.092 157 74 -2766.347 -2779.408
##
## Clustering table:
##  1  2  3  4  5
## 18 17 51 33 38

fviz_mclust(happy2dup.mclust,"BIC",palette="jco")
```



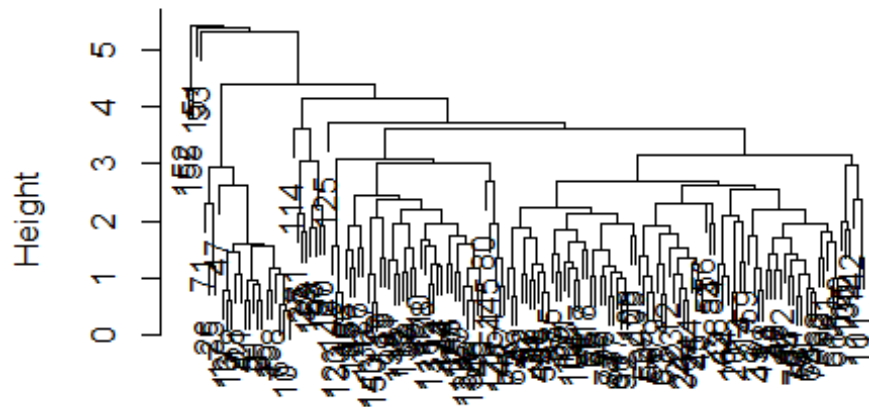
For the year 2017

```
happy3=read.csv("https://raw.githubusercontent.com/sauchilee/Stat517/master/Data/World_Happiness_2017.csv")
dim(happy3)

## [1] 155 12

happy3dup<-happy3[,c(2,6:12)]
happy3dup.s = scale(happy3dup[, -1])
#happy1.s
happy3dup.d = dist(happy3dup.s)
#happy1.d
happy3dup.hc.s = hclust(happy3dup.d,method="average")
plot(happy3dup.hc.s)
```

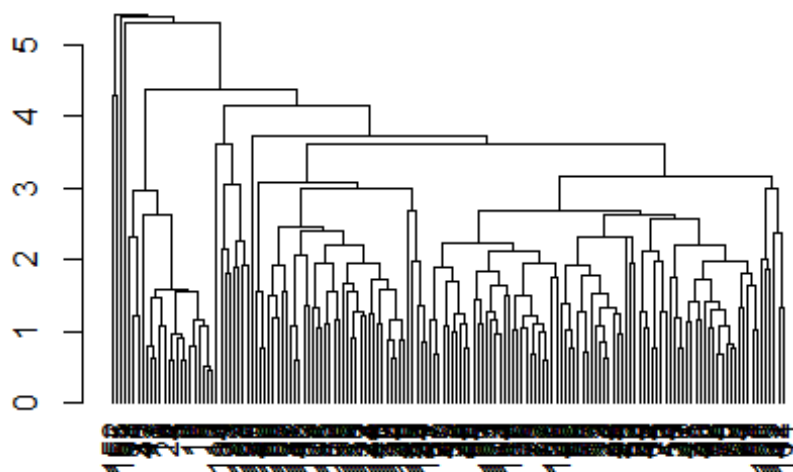
## Cluster Dendrogram



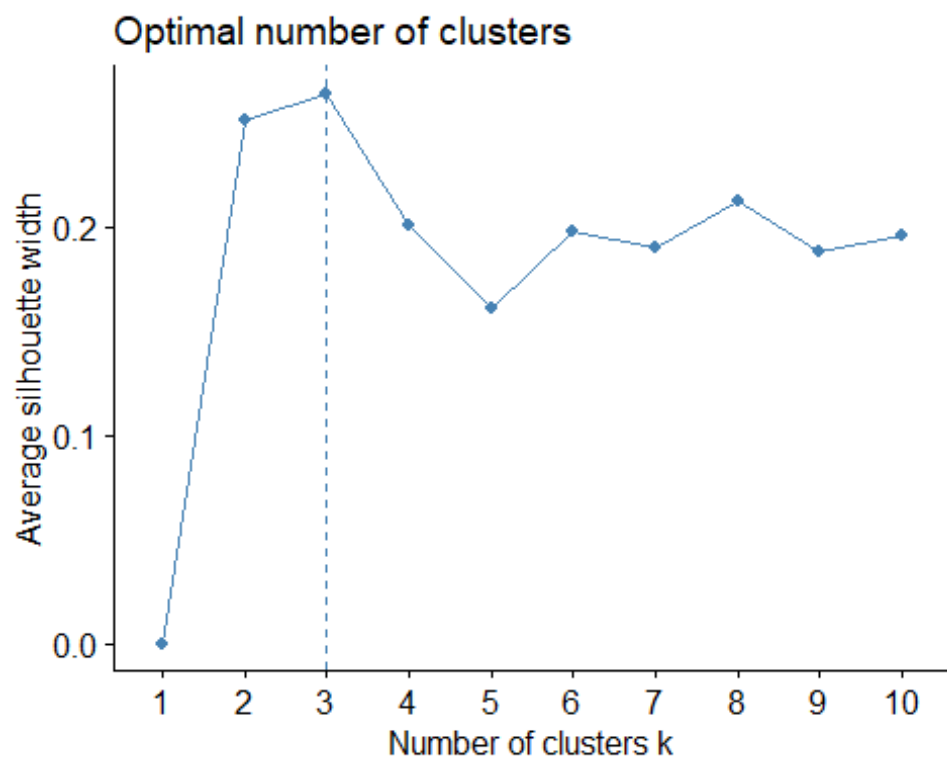
happy3dup.d  
hclust (\*, "average")

```
dend=as.dendrogram(happy3dup.hc.s)
#Labels_colors(dend)=as.numeric(as.factor(happy2dup$Region[happy2dup.hc.s$order]))
plot(dend)
```

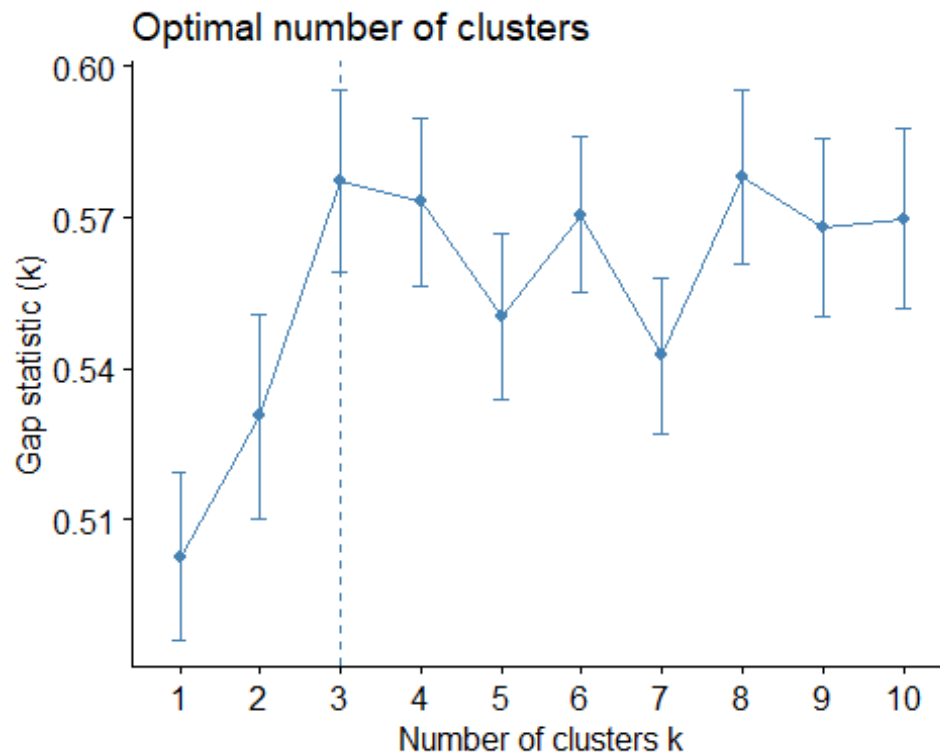




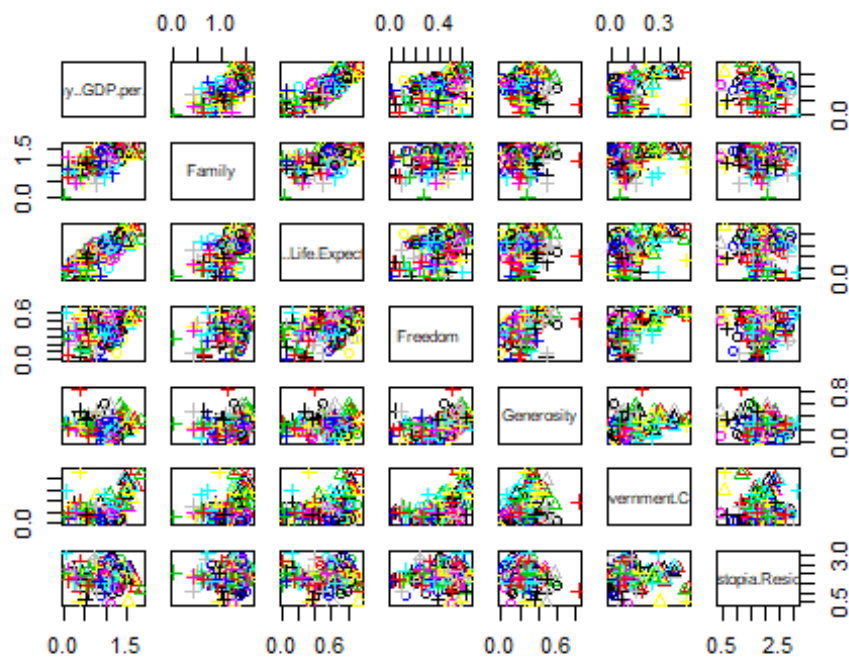
```
set.seed(150)
fviz_nbclust(happy3dup.s, kmeans, method="silhouette")
```



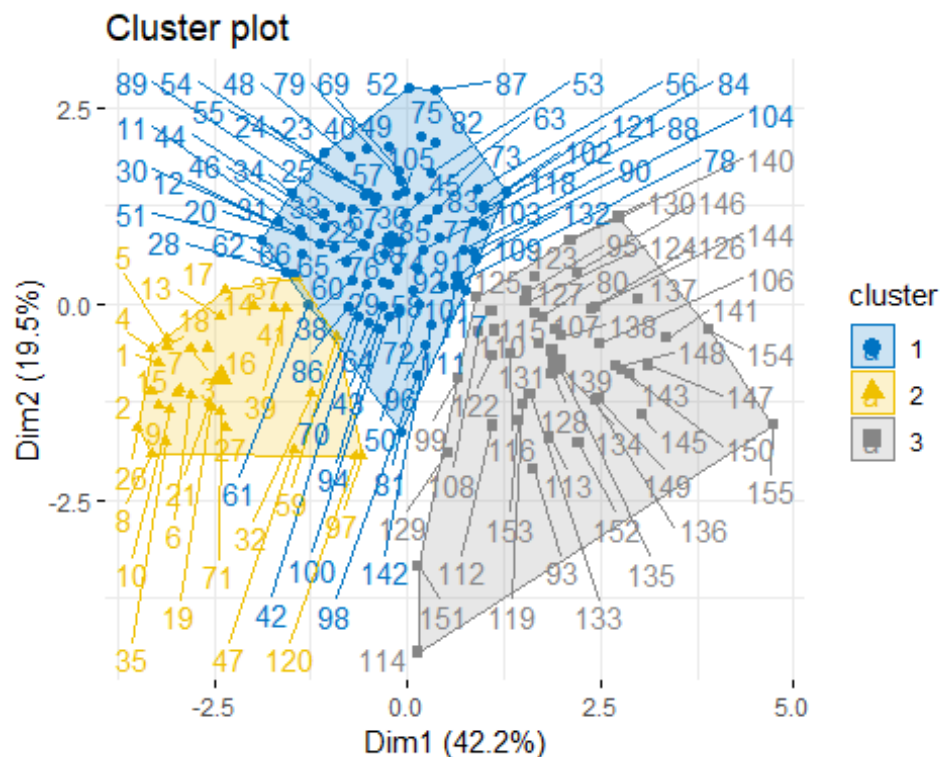
```
fviz_nbclust(happy3dup.s, kmeans, method="gap_stat")
```



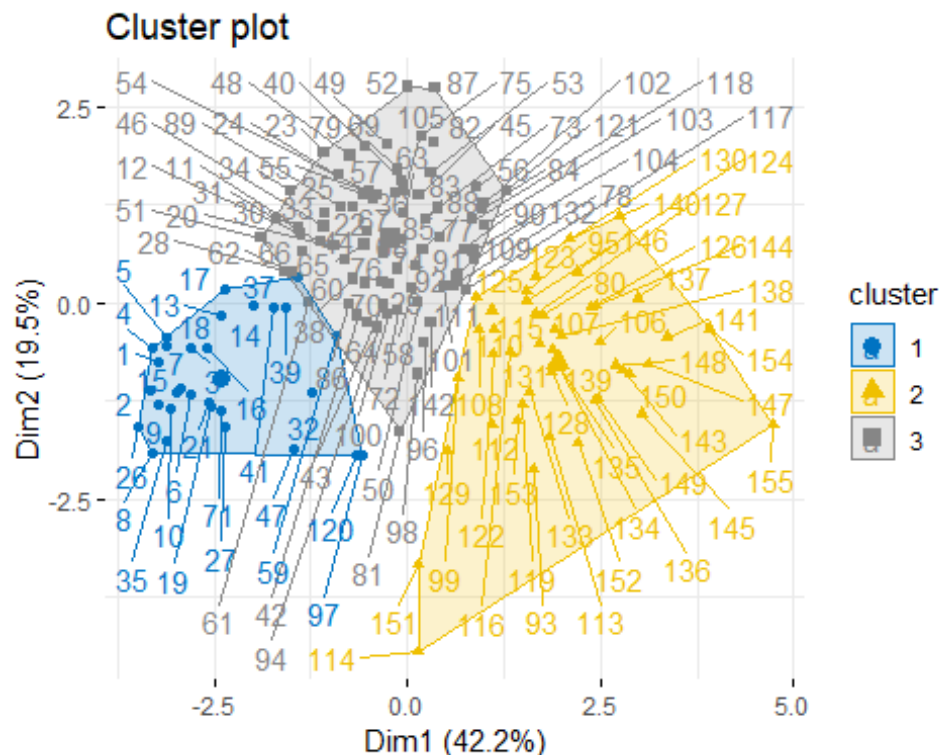
```
happy3dup.k2<-kmeans(happy3dup.s,centers = 3,iter.max=100,nstart=25)  
happy3dup.k4<-kmeans(happy3dup.s,centers = 3,iter.max=100,nstart=25)  
pairs(happy3dup[-1],pch=happy3dup.k2$cluster,col=unclass(happy3dup[,1]))
```



```
fviz_cluster(happy3dup.k2, data=happy3dup.s, ellipse.type = "convex", palette="jco",
repel = TRUE, ggtheme = theme_minimal())
```



```
fviz_cluster(happy3dup.k4,data=happy3dup.s,ellipse.type ="convex",palette="jco",repel = TRUE, ggtheme = theme_minimal())
```



M Clust

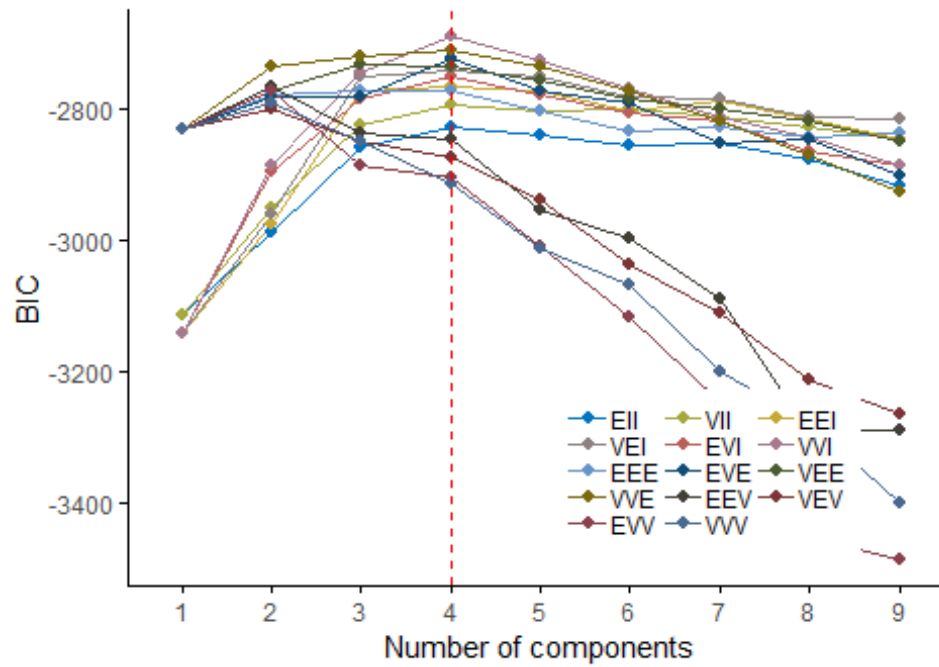
```
happy3dup.mclust<-Mclust(happy3dup.s)
summary(happy3dup.mclust)

## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVI (diagonal, varying volume and shape) model with 4 components:
##
## log.likelihood  n df      BIC      ICL
##      -1196.501 155 59 -2690.564 -2708.876
##
## Clustering table:
##  1  2  3  4
## 17 24 70 44

fviz_mclust(happy3dup.mclust,"BIC",palette="jco")
```

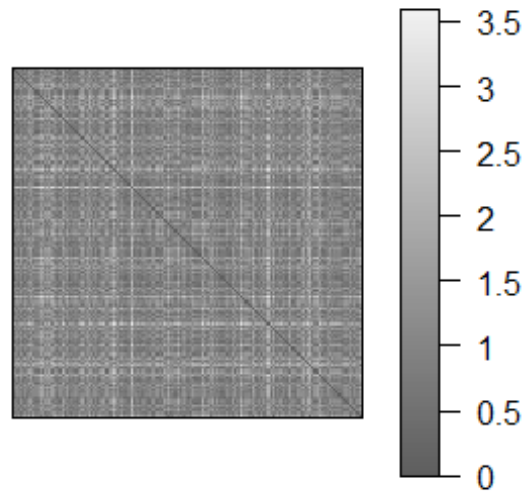
## Model selection

Best model: VVI | Optimal clusters: n = 4

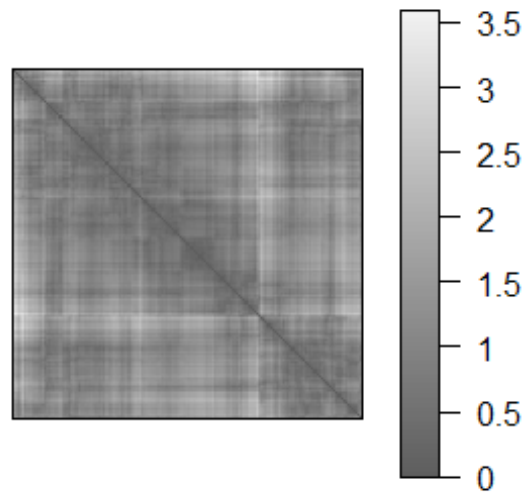


```
## Warning in dist(new1_2015): NAs introduced by coercion
```

original



ordered



```
## [1] 54 46 134 86 104 23 139 89 50 8 122 147 127 106 60 103 27
## [18] 61 98 79 56 28 154 49 76 146 148 117 74 59 7 1 96 26
## [35] 149 80 5 52 128 124 21 95 105 136 57 31 40 102 67 119 111
## [52] 141 142 88 120 10 20 35 145 133 110 83 152 84 94 92 55 3
```

```
## [69] 85 58 99 125 39 62 33 150 155 123 140 113 6 97 130 129 66
## [86] 87 68 63 65 37 82 22 151 101 137 43 29 48 12 11 30 53
## [103] 143 2 18 25 138 126 14 19 116 109 135 158 118 71 13 16 78
## [120] 38 34 77 131 81 156 51 108 70 157 64 112 121 36 107 72 144
## [137] 44 15 100 114 4 153 90 115 24 73 9 45 132 75 17 42 47
## [154] 93 69 32 91 41
```

```
data("happy2dup")
```

```
## Warning in data("happy2dup"): data set 'happy2dup' not found
```

```
new_2016<-as.matrix(happy2dup)
```

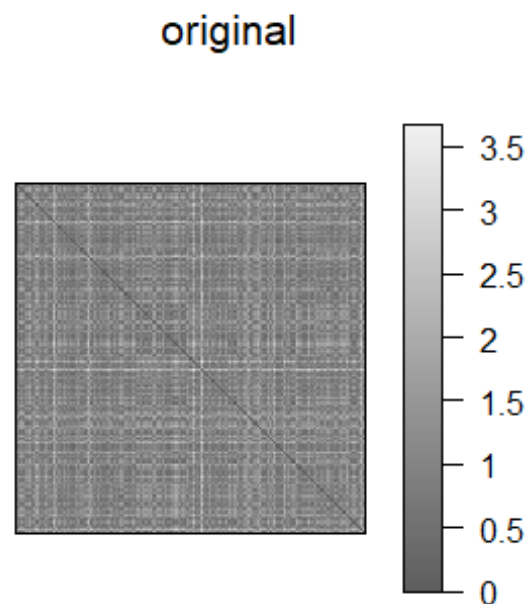
```
new1_2016<-new_2016[sample(seq_len(nrow(new_2016))),]
```

```
dis_2016<-dist(new1_2016)
```

```
## Warning in dist(new1_2016): NAs introduced by coercion
```

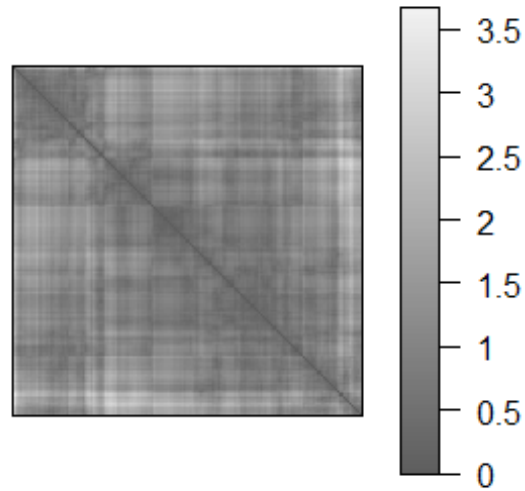
```
ord_2016<-seriate(dis_2016,method="OLO")
```

```
pimage(dis_2016,main="original")
```



```
pimage(dis_2016,ord_2016,main="ordered")
```

ordered



```
get_order(ord_2016)
```

```
## [1] 84 149 123 74 103 22 147 48 141 78 7 100 49 63 140 36 62
## [18] 11 20 67 45 47 39 13 143 85 6 5 108 138 155 87 116 104
## [35] 14 28 110 135 90 102 107 10 97 121 72 34 73 68 126 76 125
## [52] 71 99 58 152 88 153 60 30 24 46 77 96 19 55 64 2 70
## [69] 61 106 56 157 115 59 43 91 112 75 4 16 38 32 131 111 3
## [86] 65 144 134 50 145 101 51 83 79 129 66 35 31 94 44 124 133
## [103] 40 26 86 118 27 98 113 17 119 154 69 12 151 130 52 21 89
## [120] 81 142 8 93 37 15 122 57 146 1 82 29 150 54 23 25 128
## [137] 41 148 42 136 95 120 92 139 137 9 105 80 132 117 33 18 156
## [154] 127 53 114 109
```

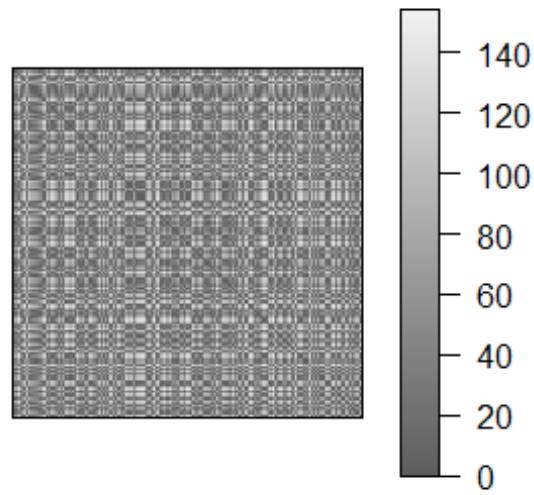
```
data("happy3dup")
```

```
## Warning in data("happy3dup"): data set 'happy3dup' not found
```

```
new_2017<-as.matrix(happy3dup)
new1_2017<-new_2017[sample(seq_len(nrow(new_2017))),]
dis_2017<-dist(new1_2017)
ord_2017<-seriate(dis_2017,method="OLO")
pimage(dis_2017,main="original")
```

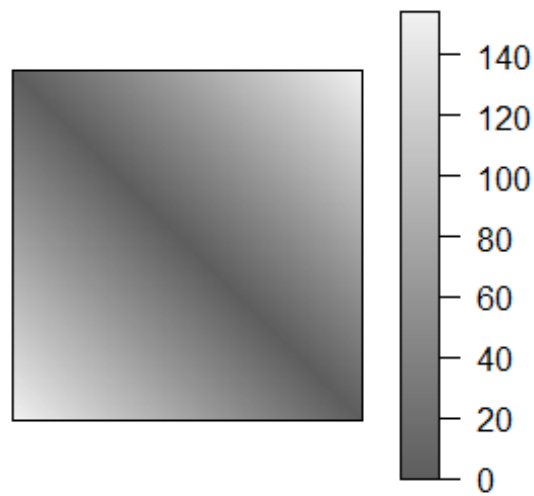


original



```
pimage(dis_2017,ord_2017,main="ordered")
```

ordered



```
get_order(ord_2017)
```

```
## [1] 91 65 39 143 6 112 25 148 128 27 41 129 47 80 145 8 18
## [18] 24 104 142 84 101 155 133 48 135 61 146 138 43 64 16 26 125
## [35] 137 117 33 83 127 113 45 19 111 5 81 149 89 102 9 74 50
## [52] 151 108 34 38 131 92 22 10 93 11 85 54 73 130 20 28 17
## [69] 71 110 121 60 42 49 124 82 4 99 12 72 88 13 77 109 152
## [86] 97 31 144 40 37 62 118 134 136 2 90 96 36 30 1 66 46
## [103] 147 69 32 23 87 141 100 107 122 21 115 154 79 153 150 67 103
## [120] 86 55 52 95 75 140 14 139 35 105 76 119 7 114 94 44 53
## [137] 57 70 120 56 15 59 29 58 78 123 98 51 126 63 68 3 106
## [154] 116 132
```

- 1) Norway tops the global happiness rankings for all three years 2015, 2016, and 2017  
Everytime when the seration analysis is done, the outcome changes randomly. I don't feel that Norway tops the global happiness for all the three years.
- 2) All top ten countries rank highly on all the main features found to support happiness I don't think that the top 10 countries rank on the main features for happiness.
- 3) Happiness is both social and personal I don't think that the happiness is based on both the social and personal. Few countries have good happiness rate for family but poor happiness rate for corruption, but still the countries ranks in the top.
- 4) Unemployment causes a major fall in happiness, and even for those in work the quality of work can cause major variations in happiness Even though there isn't any variable to decide directly about the unemployment, the family's happiness is comparatively less in many countries. The major reason for this could be unemployment and poverty.
- 5) China are no happier than most countries, though richer and longer longevity

Based on Cluster analysis, I believe that China is happier country.

- 6) Much of Africa is struggling

While doing the analysis, I found that most of the African countries fell into a single cluster and were grouped together. Moreover, the happiness rank is less for the African countries. Hence, I think that Africa is struggling.

- 7) Happiness has fallen in America

From the analysis, I understood that happiness is bit stable in America. It hasn't risen or fallen much from 2015-2017.

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