

## Similarity Part 3: Clustering

Team: DLK190005, SCK190106, LDN190002 , ASA190005,

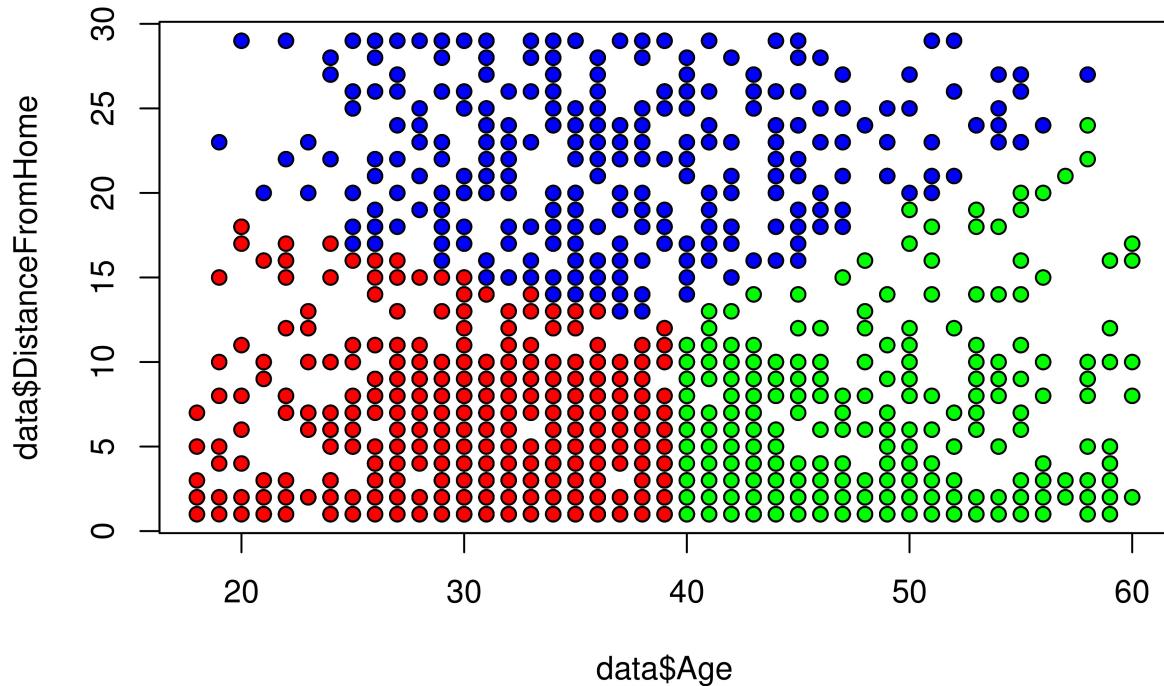
Date: 09/25/2022

HomeWork 4 -Similarity

K-Means-Cluster

```
data <- read.csv(file="C:/Users/Abed/Desktop/employeeData.csv",header=TRUE,)  
data <- na.omit(data)  
  
set.seed(1234)  
e <- data[c(1,5,6,11)]  
  
cluster = kmeans(e,3,nstart=20)  
  
table(cluster$cluster,data$Attrition)  
  
##  
##      No   Yes  
## 1 1752  428  
## 2 1094  140  
## 3  831  137  
  
plot(data$Age,data$DistanceFromHome,pch=21,bg=c("red","green","blue")  
[unclass(cluster$cluster)],main = "Company Data")
```

## Company Data



## Hierarchial Clustering

```
subset <- data[1:50,]
library(flexclust)

## Loading required package: grid

## Loading required package: lattice

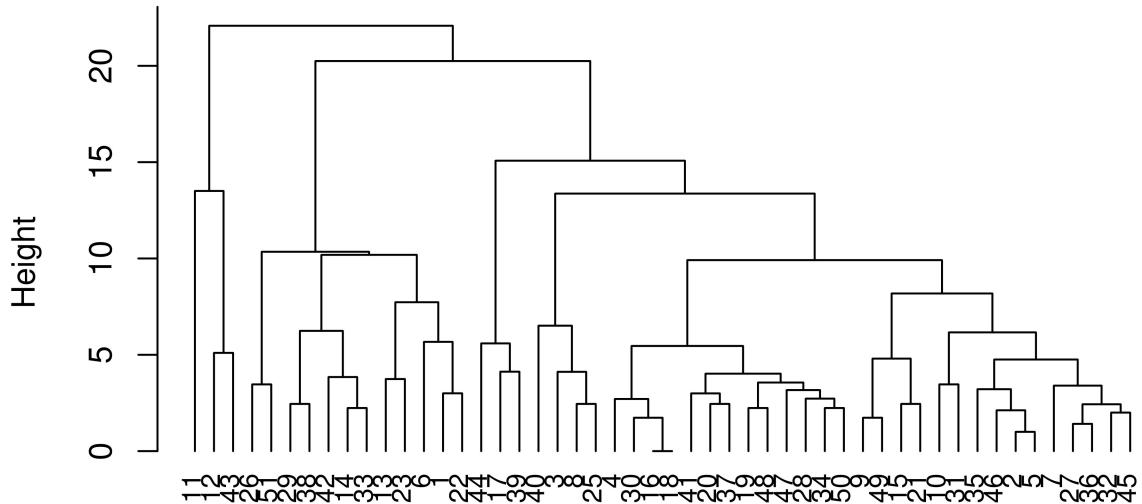
## Loading required package: modeltools

## Loading required package: stats4

c <- subset[c(1,5,6,11)]

d <- dist(c)
avgFit <- hclust(d,method="average")
plot(avgFit,hang =-1,cex=.8,main="Hierarchial Clustering")
```

## Hierachial Clustering



```
d  
hclust (*, "average")
```

```
### Model Based-Clustering
```

```
library(mclust)
```

```
## Package 'mclust' version 5.4.10  
## Type 'citation("mclust")' for citing this R package in publications.
```

```
df <- scale(c[,-1])  
mc <- Mclust(df)  
summary(mc)
```

```
## -----  
## Gaussian finite mixture model fitted by EM algorithm  
## -----  
##  
## Mclust VEV (ellipsoidal, equal shape) model with 5 components:  
##  
## log-likelihood  n df          BIC          ICL  
##             -99.51038 50 41 -359.4137 -359.4397  
##  
## Clustering table:  
##   1  2  3  4  5  
##   8 10 12 15  5
```

```

library(factoextra)

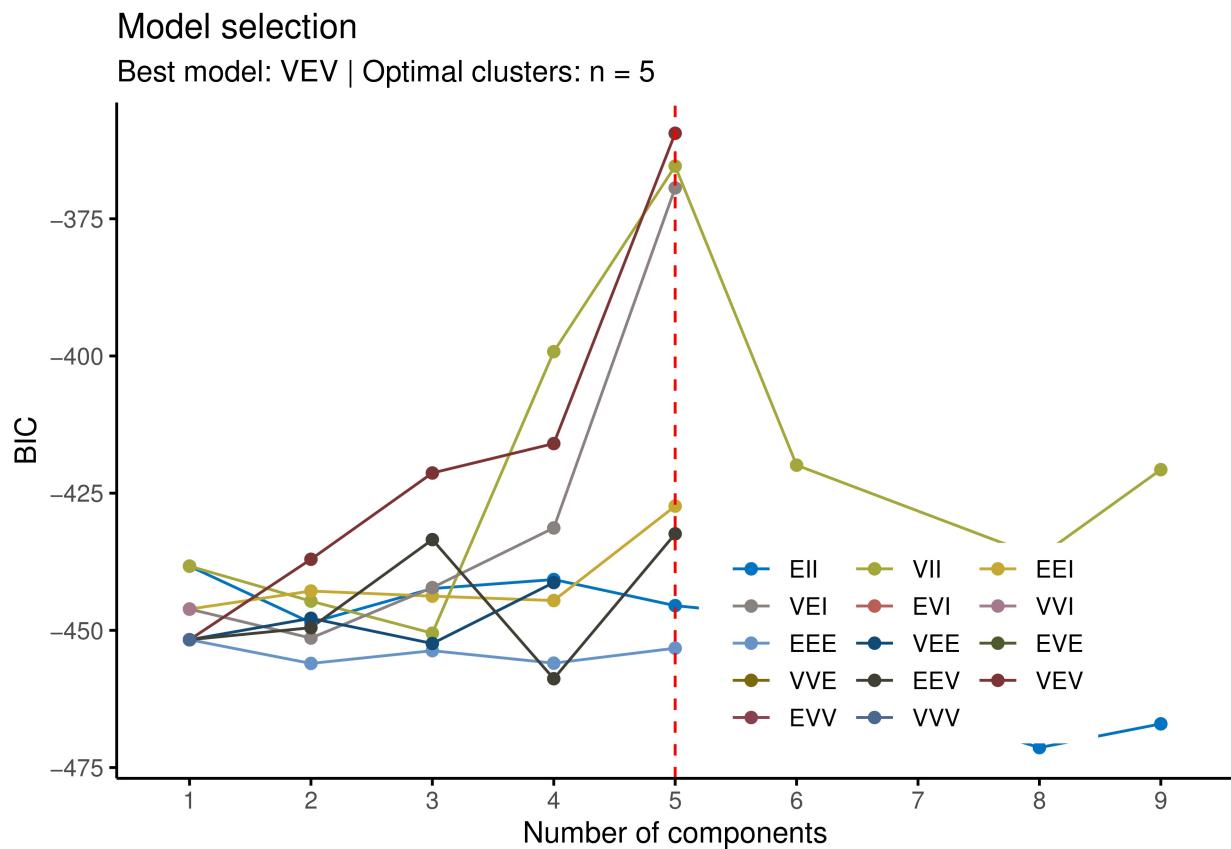
## Loading required package: ggplot2

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

fviz_mclust(mc,"BIC",palette = "jco")

## Warning: `gather_()` was deprecated in tidyverse 1.2.0.
## Please use `gather()` instead.

```



## Analysis

For K-means-Clustering showed that there is no real correlation between Age and distance commuting from and to work. Hierarchical clustering showed that individuals of the same age, tend to have the same level of work and other attributes. Model Based Clustering showed