

Assignment_4

Dengcheng Chen

2022/3/20

```
library(readr)
library(tidyverse)

## Warning: ³Ï%ü'tidyverse'ÊÇÃAR°±¼4.1.3 À`¼`ÔiμÄ

## -- Attaching packages ----- tidyverse 1.
3.1 --

## v ggplot2 3.3.5      v dplyr  1.0.7
## v tibble  3.1.6      v stringr 1.4.0
## v tidyr   1.2.0      v forcats 0.5.1
## v purrr   0.3.4

## Warning: ³Ï%ü'forcats'ÊÇÃAR°±¼4.1.3 À`¼`ÔiμÄ

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

library(factoextra)

## Warning: ³Ï%ü'factoextra'ÊÇÃAR°±¼4.1.3 À`¼`ÔiμÄ

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

library(ISLR)
Phar <- read.csv("Pharmaceuticals.csv", header=TRUE)
Phar <- Phar[,3:11]
head(Phar)

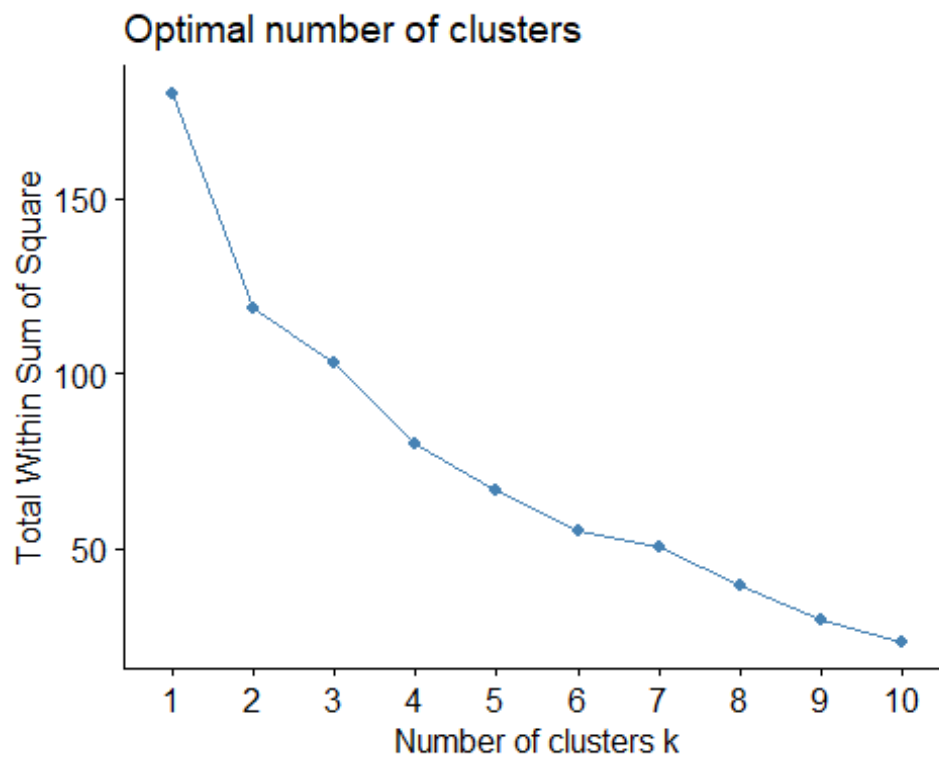
##   Market_Cap Beta PE_Ratio ROE ROA Asset_Turnover Leverage Rev_Growth
## 1      68.44 0.32    24.7 26.4 11.8          0.7      0.42      7.54
## 2       7.58 0.41    82.5 12.9  5.5          0.9      0.60      9.16
## 3       6.30 0.46    20.7 14.9  7.8          0.9      0.27      7.05
## 4      67.63 0.52    21.5 27.4 15.4          0.9      0.00     15.00
## 5      47.16 0.32    20.1 21.8  7.5          0.6      0.34     26.81
## 6      16.90 1.11    27.9  3.9  1.4          0.6      0.00     -3.17
##   Net_Profit_Margin
## 1              16.1
## 2               5.5
## 3              11.2
## 4              18.0
```

```
## 5          12.9
## 6           2.6
```

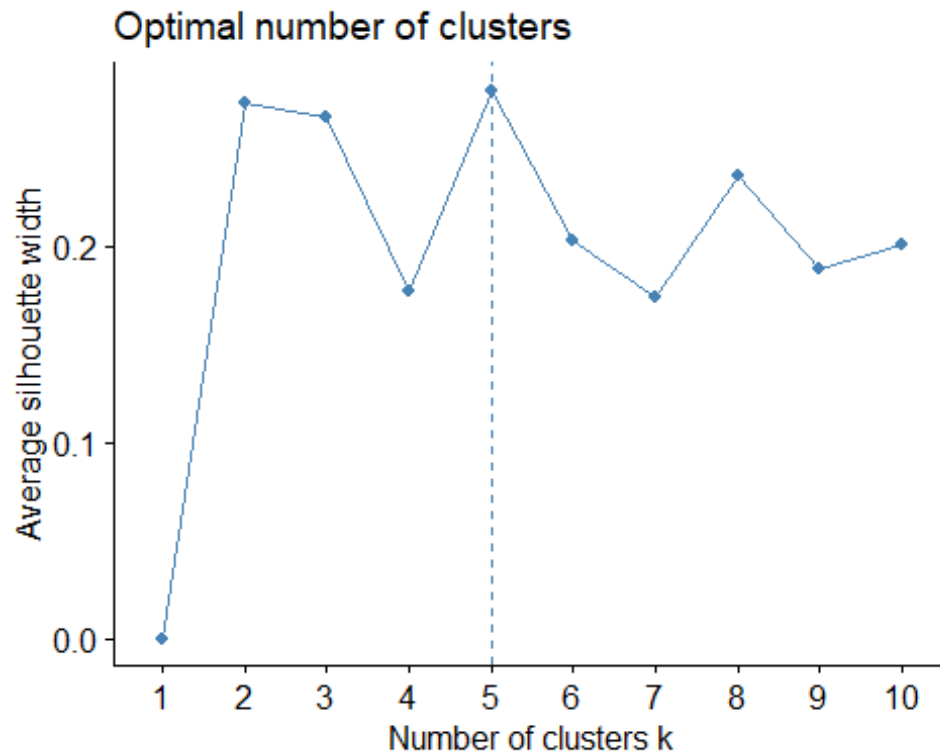
a Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made in conducting the cluster analysis, such as weights for different variables, the specific clustering algorithm(s) used, the number of clusters formed, and so on.

Choose k value

```
Phars <- scale(Phar)
fviz_nbclust(Phars, kmeans, method = "wss")
```



```
fviz_nbclust(Phars, kmeans, method = "silhouette")
```



```
set.seed(123)
k5 <- kmeans(Phars, centers = 5, nstart = 25)
k5$centers
```

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover
## 1	-0.03142211	-0.4360989	-0.31724852	0.1950459	0.4083915	0.1729746
## 2	-0.87051511	1.3409869	-0.05284434	-0.6184015	-1.1928478	-0.4612656
## 3	-0.43925134	-0.4701800	2.70002464	-0.8349525	-0.9234951	0.2306328
## 4	1.69558112	-0.1780563	-0.19845823	1.2349879	1.3503431	1.1531640
## 5	-0.76022489	0.2796041	-0.47742380	-0.7438022	-0.8107428	-1.2684804

```
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.27449312 -0.7041516      0.556954446
## 2  1.36644699 -0.6912914     -1.320000179
## 3 -0.14170336 -0.1168459     -1.416514761
## 4 -0.46807818  0.4671788      0.591242521
## 5  0.06308085  1.5180158     -0.006893899

k5$size

## [1] 8 3 2 4 4

fviz_cluster(k5, data = Phars)
```



b Cluster 1: Row 3,7,10,16,19,21 Cluster 2: Row 6,8,12 Cluster 3: Row 2,18 Cluster 4: Row 11,13,15,17 Cluster 5: Row 5,9,14,20

Cluster 1: low PE_Ratio, low Rev_growth, high Net_Profit_Margin. Cluster 2: highe Beta, high Leverage,lowe Market_Cap. Cluster 3: high PE_ratio, low ROE, low ROA, low Net Profit Margin Cluster 4: high Market_Cap, high ROE, high ROA, high Asset_Turnover Cluster 5: low Market_Cap, low Asset_Turnover, high Rev_growth

C Yes. 1 and 4.

d Cluster 1: high Net_Profit_Margin choice Cluster 2: high Leverage for warriors Cluster 3: bad choice: low return low Net Profit Margin away from me Cluster 4: steady inventing with high Asset_Turnover and high Market_Cap Cluster 5: high Rev_growth for future