HEDGE_FUND_EDA

Nandan Soni

2024-12-19

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.5.0     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(ggplot2)
require(caret)
```

```
## Loading required package: caret
```

```
## Warning: package 'caret' was built under R version 4.3.2
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
require(factoextra)
```

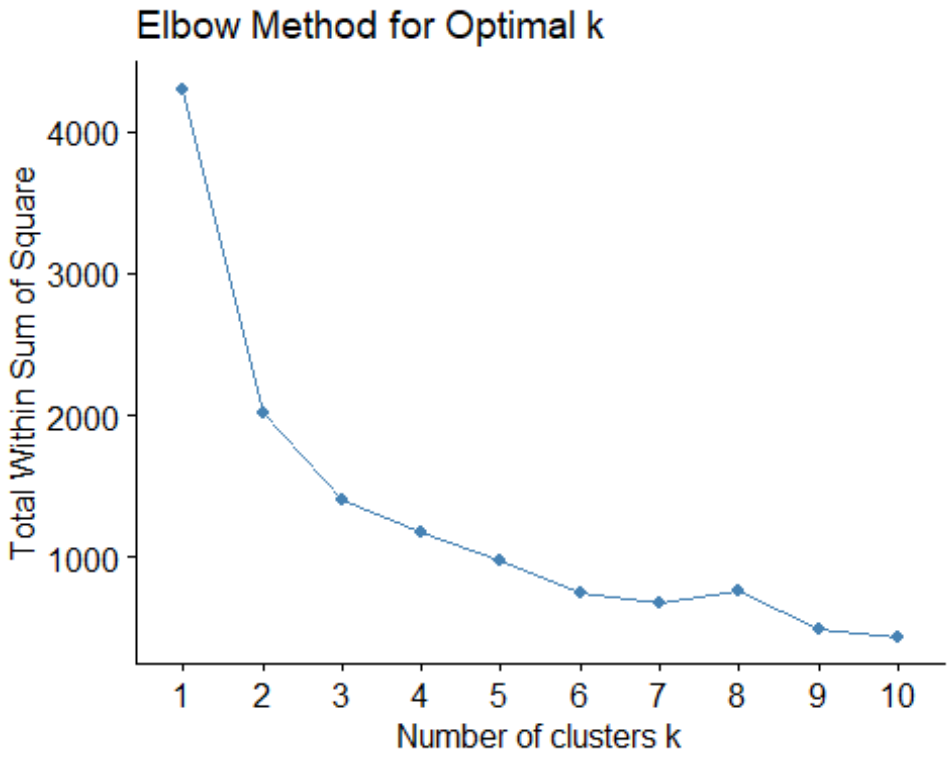
```
## Loading required package: factoextra
```

```
## Warning: package 'factoextra' was built under R version 4.3.3
```

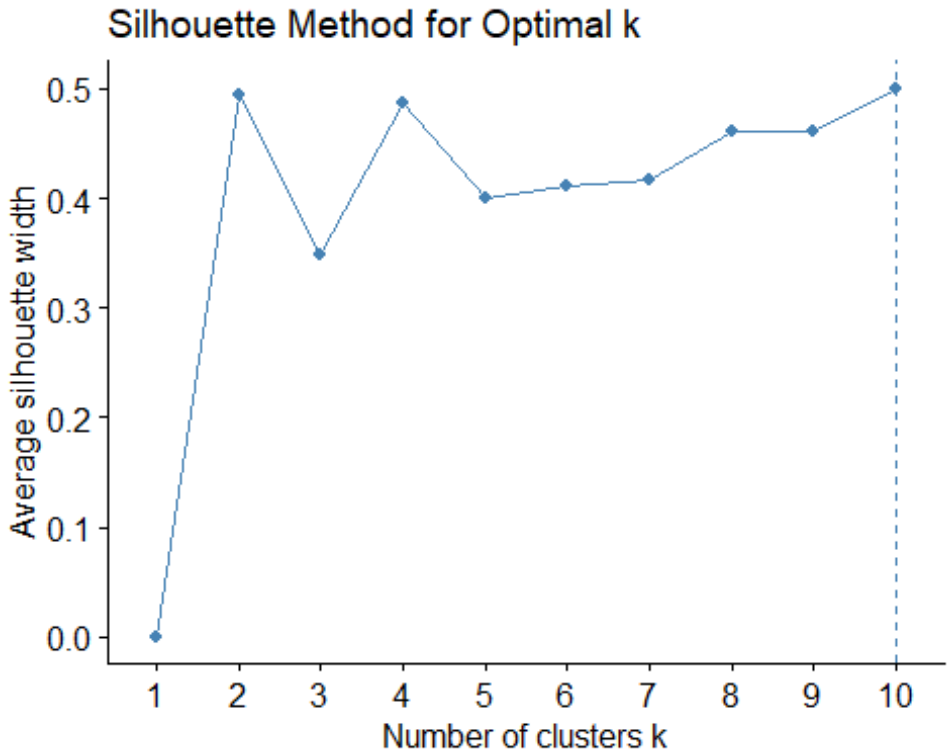
```
## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa
```

```
require(useful)
```

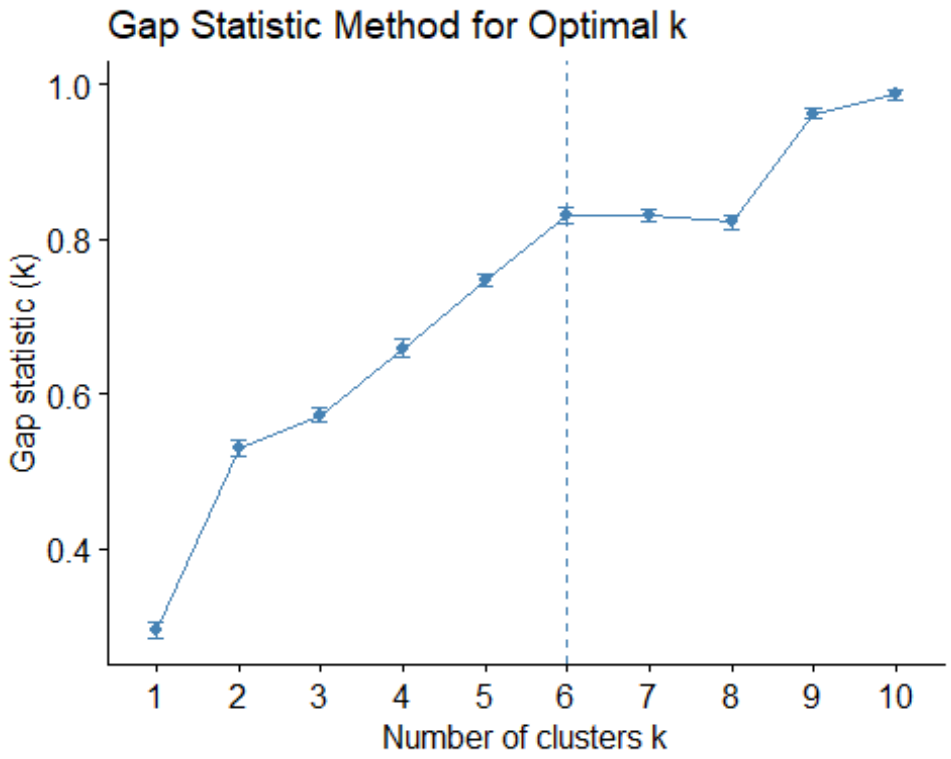
```
## Loading required package: useful

## Warning: package 'useful' was built under R version 4.3.3

library(stats)

setwd("D:/3 rd/401")
data <- read.csv("hedge_fund.csv")

str(data)

## 'data.frame':    478 obs. of  10 variables:
##  $ Date                : int  31048 31079 31107 31138 31168 31199 31229
31260 31291 31321 ...
##  $ M2V                 : num  1.8 1.8 1.8 1.79 1.79 ...
##  $ UNRATE              : num  7.3 7.2 7.2 7.3 7.2 7.4 7.4 7.1 7.1 7.1 ...
##  $ CPIAUCNS            : num  106 106 106 107 107 ...
##  $ PPIACO              : num  103 103 103 103 104 ...
##  $ FEDFUNDS            : num  8.35 8.5 8.58 8.27 7.97 7.53 7.88 7.9 7.92
7.99 ...
##  $ Adj.Close           : num  180 181 181 180 190 ...
##  $ GOLD                : chr  "325.3" "321.9" "345.4" "338.9" ...
##  $ WTISPLC..crude.oil..: num  25.6 27.3 28.2 28.8 27.6 ...
##  $ PERMIT..Units..000s : int  1660 1662 1727 1664 1709 1716 1697 1808 1916
1743 ...

data$GOLD <- as.numeric(gsub("[^0-9.]", "", data$GOLD))

data$PERMIT..Units..000s <- as.numeric(data$PERMIT..Units..000s)

scaled_data <- scale(data)

pca_result <- prcomp(scaled_data, center = TRUE, scale. = TRUE)

data_cluster <- data %>%
  select(-Date) %>%
  na.omit() %>%
  scale()
str(data_cluster)

##  num [1:478, 1:9] 0.0977 0.0977 0.0977 0.0839 0.0839 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:478] "1" "2" "3" "4" ...
##   ..$ : chr [1:9] "M2V" "UNRATE" "CPIAUCNS" "PPIACO" ...
##  - attr(*, "scaled:center")= Named num [1:9] 1.77 5.78 195 160.84 3.44 ...
##   ..- attr(*, "names")= chr [1:9] "M2V" "UNRATE" "CPIAUCNS" "PPIACO" ...
##  - attr(*, "scaled:scale")= Named num [1:9] 0.29 1.67 54.6 45.37 2.76 ...
##   ..- attr(*, "names")= chr [1:9] "M2V" "UNRATE" "CPIAUCNS" "PPIACO" ...

fviz_nbclust(data_cluster, kmeans, method = "wss") +
  labs(title = "Elbow Method for Optimal k")
```




```
fviz_nbclust(data_cluster, kmeans, method = "silhouette") +
  labs(title = "Silhouette Method for Optimal k")
```

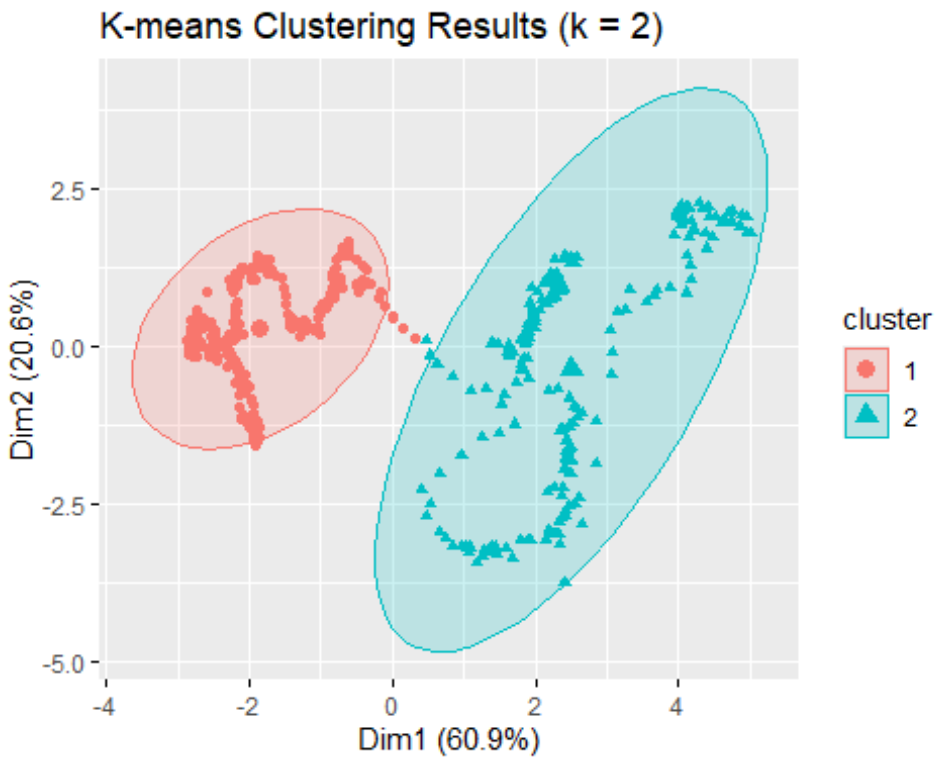


```r
fviz_nbclust(data_cluster, kmeans, method = "gap_stat", nboot = 20) +
  labs(title = "Gap Statistic Method for Optimal k")
```

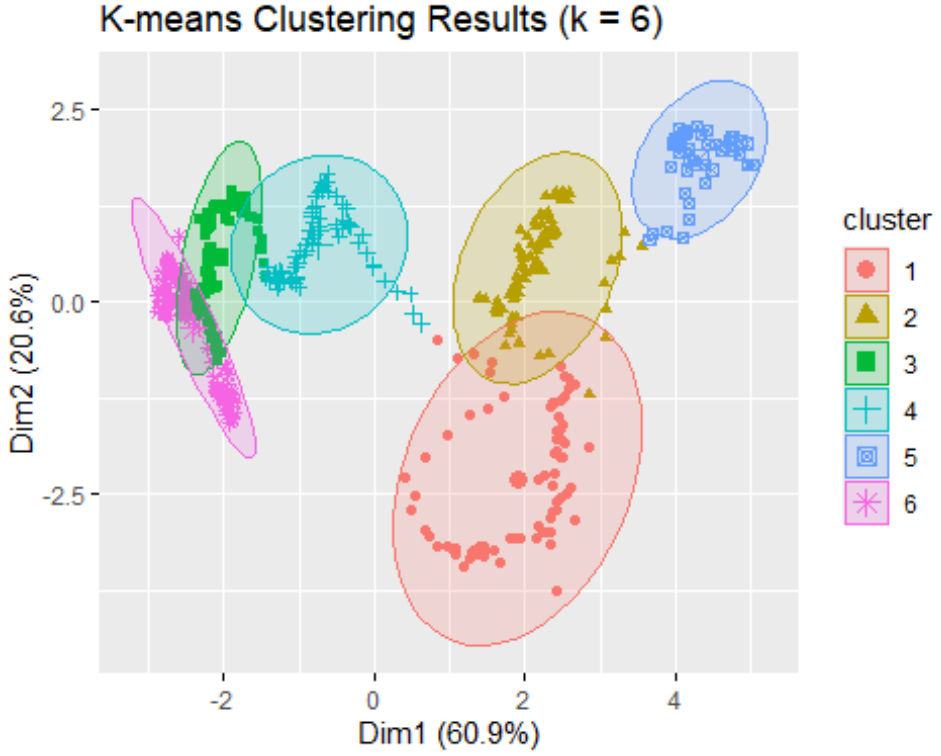



```r
k2 <- kmeans(data_cluster, centers = 2, nstart = 25)
k6 <- kmeans(data_cluster, centers = 6, nstart = 25)
```

```r
fviz_cluster(k2, data = data_cluster, geom = "point", ellipse.type = "norm")
+
  labs(title = "K-means Clustering Results (k = 2)")
```




```r
fviz_cluster(k6, data = data_cluster, geom = "point", ellipse.type = "norm")
+
  labs(title = "K-means Clustering Results (k = 6)")
```



```r
data$Cluster <- k6$cluster


head(data)
```

```
##    Date   M2V UNRATE CPIAUCNS PPIACO FEDFUNDS Adj.Close  GOLD
## 1 31048 1.799    7.3    105.5  103.4     8.35    179.63 325.3
## 2 31079 1.799    7.2    106.0  103.3     8.50    181.18 321.9
## 3 31107 1.799    7.2    106.4  103.1     8.58    180.66 345.4
## 4 31138 1.795    7.3    106.9  103.3     8.27    179.83 338.9
## 5 31168 1.795    7.2    107.3  103.5     7.97    189.55 345.7
## 6 31199 1.795    7.4    107.6  103.3     7.53    191.85 340.4
##   WTISPLC..crude.oil.. PERMIT..Units..000s Cluster
## 1               25.641                1660       6
## 2               27.271                1662       6
## 3               28.238                1727       6
## 4               28.805                1664       6
## 5               27.623                1709       6
## 6               27.143                1716       6
```

```r
ggplot(data, aes(x = M2V, y = Adj.Close, color = factor(Cluster))) +
  geom_point() +
  labs(title = "Gold vs Adjusted Close Price by Cluster",
       color = "Cluster") +
  theme_minimal()
```