```r
library(readr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v purrr     1.0.2
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(caret)
```

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

```r
library(knitr)
library(class)
library(ggplot2)
library(ggcorrplot)
library(dplyr)
library(e1071)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(factoextra)
```

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

```r
library(cluster)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```r
library(pander)
library(kernlab)
```

```
##
## Attaching package: 'kernlab'
##
## The following object is masked from 'package:purrr':
##
##     cross
##
## The following object is masked from 'package:ggplot2':
##
##     alpha
```

```r
library(tidyr)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
data = read.csv("/Users/srikanthgembali/Downloads/Pharmaceuticals.csv")
head(data)
```

```
##   Symbol                Name Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover
## 1    ABT Abbott Laboratories      68.44 0.32     24.7 26.4 11.8            0.7
## 2    AGN      Allergan, Inc.       7.58 0.41     82.5 12.9  5.5            0.9
## 3    AHM        Amersham plc       6.30 0.46     20.7 14.9  7.8            0.9
## 4    AZN     AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4            0.9
## 5    AVE             Aventis      47.16 0.32     20.1 21.8  7.5            0.6
## 6    BAY           Bayer AG       16.90 1.11     27.9  3.9  1.4            0.6
##   Leverage Rev_Growth Net_Profit_Margin Median_Recommendation Location Exchange
## 1     0.42       7.54              16.1          Moderate Buy       US     NYSE
## 2     0.60       9.16               5.5          Moderate Buy   CANADA     NYSE
## 3     0.27       7.05              11.2            Strong Buy       UK     NYSE
## 4     0.00      15.00              18.0         Moderate Sell       UK     NYSE
## 5     0.34      26.81              12.9          Moderate Buy   FRANCE     NYSE
## 6     0.00      -3.17               2.6                  Hold  GERMANY     NYSE
```

```r
str(data)
```

```
## 'data.frame':    21 obs. of  14 variables:
##  $ Symbol             : chr  "ABT" "AGN" "AHM" "AZN" ...
##  $ Name               : chr  "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZeneca PL
##  $ Market_Cap         : num  68.44 7.58 6.3 67.63 47.16 ...
##  $ Beta               : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...
##  $ PE_Ratio           : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...
##  $ ROE                : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...
```

```
##  $ ROA                  : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...
##  $ Asset_Turnover       : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
##  $ Leverage             : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...
##  $ Rev_Growth           : num  7.54 9.16 7.05 15 26.81 ...
##  $ Net_Profit_Margin    : num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...
##  $ Median_Recommendation: chr  "Moderate Buy" "Moderate Buy" "Strong Buy" "Moderate Sell" ...
##  $ Location             : chr  "US" "CANADA" "UK" "UK" ...
##  $ Exchange             : chr  "NYSE" "NYSE" "NYSE" "NYSE" ...
```

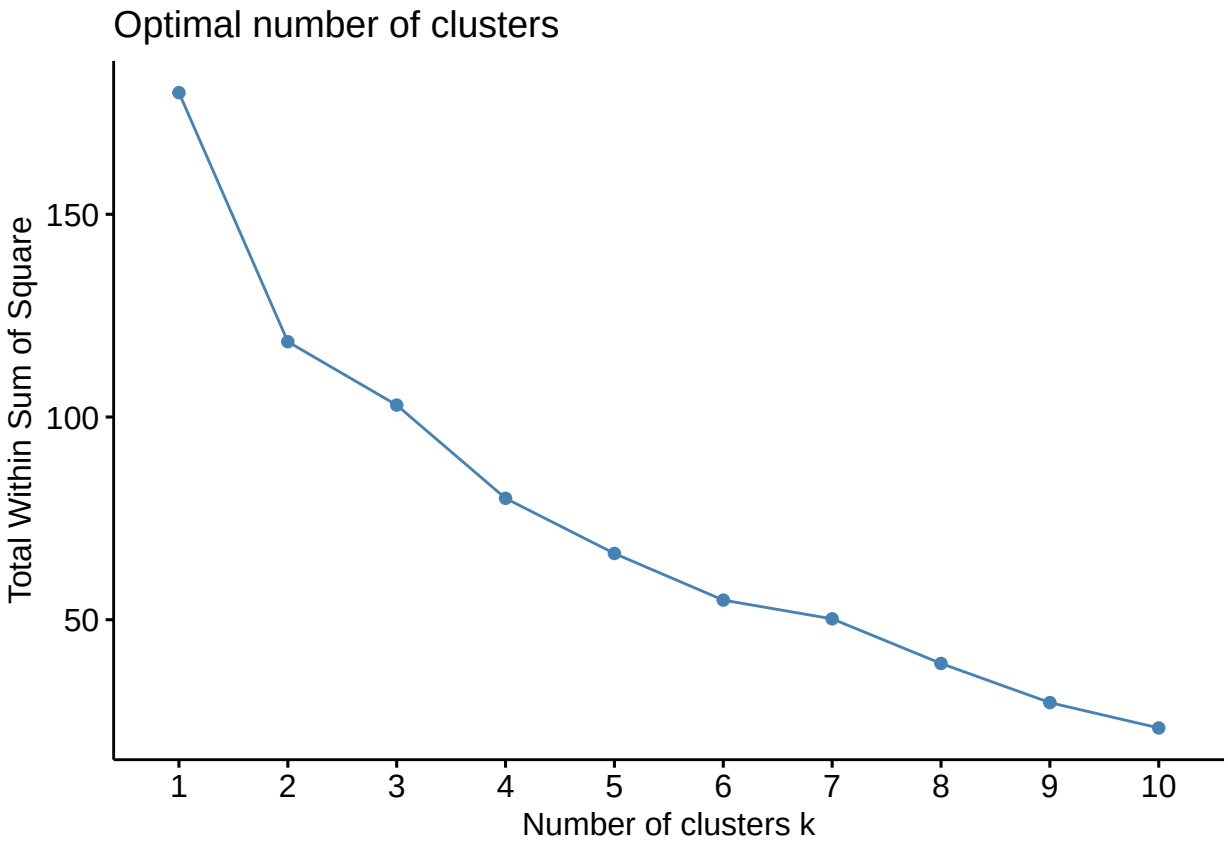
a. Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made in conducting the cluster analysis, such as weights for different variables, the specific clustering algorithm(s) used, the number of clusters formed, and so on.

```
set.seed(123)
a = data[, 3:11] #dropping remaining columns
scale_a = scale(a) #standardize the data
head(scale_a)
```

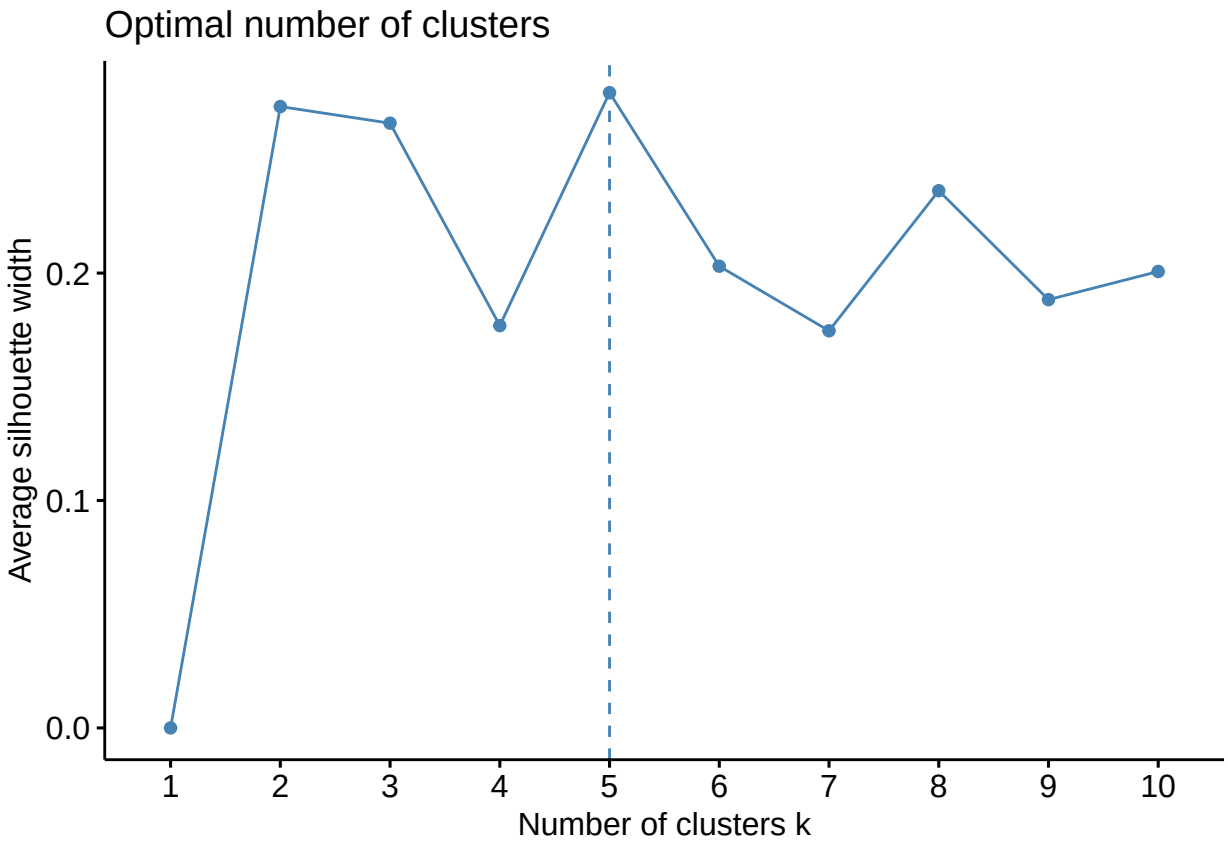
```
##      Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## [1,]  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121  -5.121077e-16
## [2,] -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871   9.225312e-01
## [3,] -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700   9.225312e-01
## [4,]  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259   9.225312e-01
## [5,] -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461  -4.612656e-01
## [6,] -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612  -4.612656e-01
##        Leverage Rev_Growth Net_Profit_Margin
## [1,] -0.2120979 -0.5277675        0.06168225
## [2,]  0.0182843 -0.3811391       -1.55366706
## [3,] -0.4040831 -0.5721181       -0.68503583
## [4,] -0.7496565  0.1474473        0.35122600
## [5,] -0.3144900  1.2163867       -0.42597037
## [6,] -0.7496565 -1.4971443       -1.99560225
```

```
# Finding optimal number of clusters using Elbow method and Silhouette method

Elbow = fviz_nbclust(scale_a, kmeans, method = "wss")
plot(Elbow)
```




```
silhouette = fviz_nbclust(scale_a, kmeans, method = "silhouette")
plot(silhouette)
```

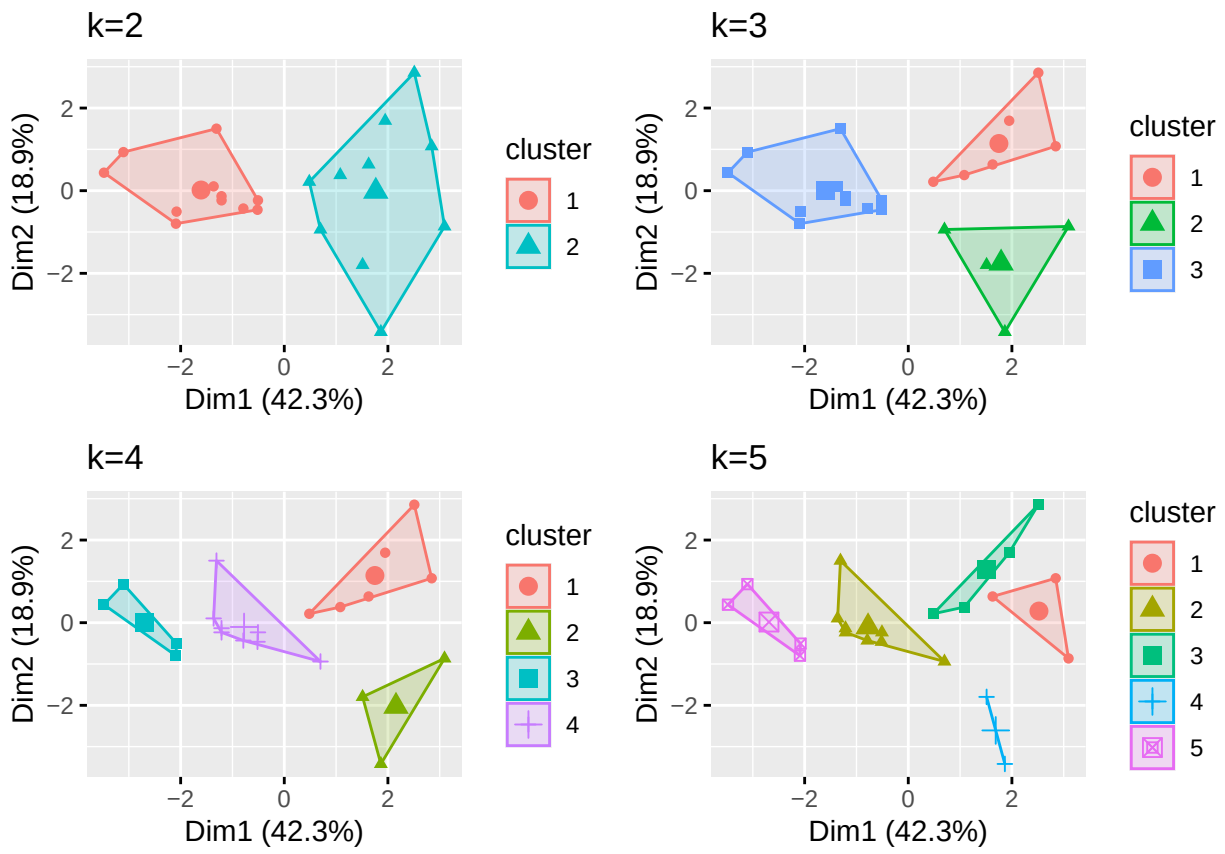


Elbow method show that best K = 2, whereas Silhouette method show best K = 5. We can plot all clusters from 2 to 5 to check which cluster is not overlapping

```
k2 = kmeans(scale_a,centers = 2,nstart=25)
k3 = kmeans(scale_a,centers = 3,nstart=25)
k4 = kmeans(scale_a,centers = 4,nstart=25)
k5 = kmeans(scale_a,centers = 5,nstart=25)

p1 = fviz_cluster(k2,geom = "point", data=scale_a)+ggtitle("k=2")
p2 = fviz_cluster(k3,geom = "point", data=scale_a)+ggtitle("k=3")
p3 = fviz_cluster(k4,geom = "point", data=scale_a)+ggtitle("k=4")
p4 = fviz_cluster(k5,geom = "point", data=scale_a)+ggtitle("k=5")

grid.arrange(p1,p2,p3,p4)
```



We can consider best k = 5, as the cluster are not overlapping by considering all the firms.

b. Interpret the clusters with respect to the numerical variables used in forming the clusters.

```
k = 5
b = kmeans(scale_a, centers = k)
cluster_membership = b$cluster
data$Cluster = cluster_membership
centroids = b$centers
print(centroids)
```

```
##    Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1  1.69558112 -0.1780563 -0.1984582  1.2349879  1.3503431   1.153164e+00
## 2 -0.66114002 -0.7233539 -0.3512251 -0.6736441 -0.5915022  -1.537552e-01
## 3 -0.52462814  0.4451409  1.8498439 -1.0404550 -1.1865838  -3.330669e-16
## 4  0.08926902 -0.4618336 -0.3208615  0.3260892  0.5396003   6.589509e-02
## 5 -0.96247577  1.1949250 -0.3639982 -0.5200697 -0.9610792  -1.153164e+00
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.4680782  0.4671788         0.5912425
## 2 -0.4040831  0.6917224        -0.4005718
## 3 -0.3443544 -0.5769454        -1.6095439
## 4 -0.2559803 -0.7230135         0.7343816
## 5  1.4773718  0.7120120        -0.3688236
```

```
group_b = data[, c("Name", "Cluster")]
group_b
```

```
##                                 Name Cluster
## 1                 Abbott Laboratories       4
## 2                      Allergan, Inc.       3
## 3                        Amersham plc       2
## 4                     AstraZeneca PLC       4
## 5                             Aventis       2
## 6                           Bayer AG        3
## 7        Bristol-Myers Squibb Company       4
## 8                          Chattem, Inc       5
## 9               Elan Corporation, plc       5
## 10              Eli Lilly and Company       4
## 11                GlaxoSmithKline plc       1
## 12                   IVAX Corporation       5
## 13                  Johnson & Johnson       1
## 14 Medicis Pharmaceutical Corporation       5
## 15                  Merck & Co., Inc.       1
## 16                        Novartis AG       4
## 17                         Pfizer Inc       1
## 18              Pharmacia Corporation       3
## 19        Schering-Plough Corporation       4
## 20       Watson Pharmaceuticals, Inc.       2
## 21                             Wyeth       4
```

```
grouped_data <- group_b %>% group_by(Cluster) %>%
  summarise(Names = paste(Name, collapse = ", "))
grouped_data
```

```
## # A tibble: 5 x 2
##   Cluster Names
##     <int> <chr>
## 1       1 GlaxoSmithKline plc, Johnson & Johnson, Merck & Co., Inc., Pfizer Inc
## 2       2 Amersham plc, Aventis, Watson Pharmaceuticals, Inc.
## 3       3 Allergan, Inc., Bayer AG, Pharmacia Corporation
## 4       4 Abbott Laboratories, AstraZeneca PLC, Bristol-Myers Squibb Company, E~
## 5       5 Chattem, Inc, Elan Corporation, plc, IVAX Corporation, Medicis Pharma~
```

c. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters)

```
c = data %>% select(c(12,13,14)) %>% mutate(Cluster = b$cluster)
c
```

```
##    Median_Recommendation    Location Exchange Cluster
## 1           Moderate Buy          US     NYSE       4
## 2           Moderate Buy      CANADA     NYSE       3
## 3             Strong Buy          UK     NYSE       2
## 4          Moderate Sell          UK     NYSE       4
## 5           Moderate Buy      FRANCE     NYSE       2
```

```
## 6                   Hold     GERMANY     NYSE       3
## 7          Moderate Sell          US     NYSE       4
## 8           Moderate Buy          US   NASDAQ       5
## 9          Moderate Sell     IRELAND     NYSE       5
## 10                  Hold          US     NYSE       4
## 11                  Hold          UK     NYSE       1
## 12                  Hold          US     AMEX       5
## 13          Moderate Buy          US     NYSE       1
## 14          Moderate Buy          US     NYSE       5
## 15                  Hold          US     NYSE       1
## 16                  Hold SWITZERLAND     NYSE       4
## 17          Moderate Buy          US     NYSE       1
## 18                  Hold          US     NYSE       3
## 19                  Hold          US     NYSE       4
## 20         Moderate Sell          US     NYSE       2
## 21                  Hold          US     NYSE       4
```

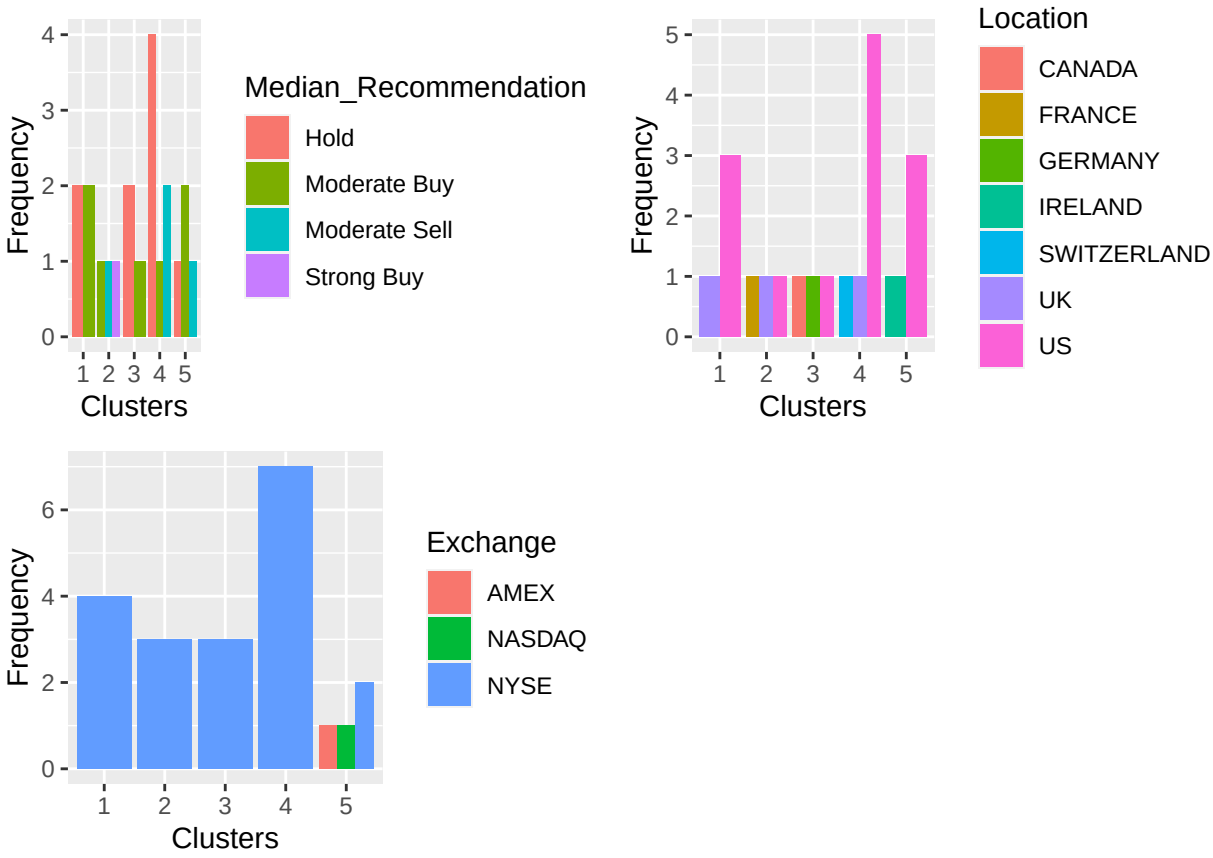
```
library(cowplot)

Median_Recom <- ggplot(data, mapping = aes(factor(Cluster), fill=Median_Recommendation)) + geom_bar(posi

Location <- ggplot(data, mapping = aes(factor(Cluster), fill=Location)) + geom_bar(position = 'dodge') +

Exchange <- ggplot(data, mapping = aes(factor(Cluster), fill=Exchange)) + geom_bar(position = 'dodge') +

plot_grid(Median_Recom, Location, Exchange)
```



Cluster 1 has a combination of US and UK firms listed in NYSE, and they have Moderate buy and hold indicating that it has potential for appreciation or growth, but with some cause or reservations.

Cluster 2 has mix of US, UK and France firms listed on NYSE, and they have Moderate Buy, Moderate sell and Strong Buy indicating varying levels of analyst confidence and market sentiment towards these securities.

Cluster 3 has companies from US, Canada and Germany listed on NYSE, and they have high Hold than Moderate Buy suggesting that analysts or financial institutions may view these stocks as having relatively stable performance potential in the near term, but with less optimism for significant appreciation in value.

Cluster 4 has many companies from US, mix from UK and Switzerland listed on NYSE and they have high hold, better Moderate Sell and lower Moderate Buy that may suggest a prevailing sentiment among analysts or institutions that these stocks are currently overvalued or facing potential headwinds in the near term.

Cluster 5 has firms from US and Germany listed on all exchanges NYSE, NASDAQ and AMEX and they have high Moderate but and Moderate Sell suggesting a mixed sentiment among analysts regarding the performance and prospects of these firms.

d. Provide an appropriate name for each cluster using any or all of the variables in the dataset.

Cluster 1: "Transatlantic Growth Prospects" Cluster 1 comprises a blend of firms from both the United States and the United Kingdom, listed on the New York Stock Exchange (NYSE). Analysts have assigned a mixture of 'Moderate Buy' and 'Hold' recommendations to these securities, reflecting a cautious yet optimistic outlook for potential appreciation or growth.

Cluster 2: "Cross-Continental Market Sentiment" Cluster 2 showcases a diverse mix of companies from the US, UK, and France, all listed on the NYSE. Analyst recommendations within this cluster range from 'Moderate Buy' to 'Moderate Sell' and 'Strong Buy,' indicating varying levels of analyst confidence and market sentiment towards these securities.

Cluster 3: "Stable Performers Across Borders" Cluster 3 features companies hailing from the US, Canada, and Germany, listed primarily on the NYSE. Analysts have issued a higher proportion of 'Hold' recommendations compared to 'Moderate Buy,' suggesting a perception of relative stability in performance potential for these stocks in the short term, albeit with less anticipation for significant value appreciation.

Cluster 4: "Valuation Concerns Amidst Global Presence" Cluster 4 comprises a substantial number of companies from the US, alongside a mix from the UK and Switzerland, all listed on the NYSE. Notably, analyst recommendations lean towards 'Hold,' with a slightly more favorable disposition towards 'Moderate Sell' than 'Moderate Buy.' This pattern suggests prevailing concerns among analysts or institutions regarding overvaluation or potential headwinds in the near term.

Cluster 5: "Mixed Signals in Global Markets" Cluster 5 is characterized by firms originating from the US and Germany, with listings across all major exchanges including the NYSE, NASDAQ, and AMEX. Analyst sentiment within this cluster reflects a blend of 'High Moderate Buy' and 'Moderate Sell,' indicating a divergence of opinion regarding the performance and prospects of these companies amidst global market dynamics.