FML\_Assignment\_4

Manpreet

2025-10-27

# Load necessary libraries  
library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.4.3

## Warning: package 'dplyr' was built under R version 4.4.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 4.0.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ 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(cluster)

## Warning: package 'cluster' was built under R version 4.4.2

library(factoextra)

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

library(readr)  
library(dplyr)  
library(flexclust)

## Loading required package: grid  
## Loading required package: lattice

## Warning: package 'lattice' was built under R version 4.4.3

## Loading required package: modeltools  
## Loading required package: stats4

set.seed(123)  
pharma\_data <- read.csv("D:/Pharmaceuticals.csv")  
# View first few rows  
head(pharma\_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

# Check structure  
str(pharma\_data)

## 'data.frame': 21 obs. of 14 variables:  
## $ Symbol : chr "ABT" "AGN" "AHM" "AZN" ...  
## $ Name : chr "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZeneca PLC" ...  
## $ 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" ...

**Interpretation**

Cluster 1 includes firms with moderate market performance and “Moderate Buy” ratings. Cluster 2 has mixed firms from different countries with average profitability and mixed analyst opinions. Cluster 3 represents strong global firms like Abbott and AstraZeneca with high ROE, ROA, and “Buy” ratings. Cluster 4 includes smaller firms with moderate returns and higher leverage. Cluster 5 has top performers like Pfizer and Johnson & Johnson, showing the best financial strength and “Strong Buy” recommendations.

Overall, clusters 3 and 5 show the strongest financial and market positions, while clusters 1–2 are moderate and cluster 4 is relatively weaker.

View(pharma\_data)

modified\_pharma\_data <- pharma\_data[ ,c(3:11)]  
summary(modified\_pharma\_data)

## Market\_Cap Beta PE\_Ratio ROE   
## Min. : 0.41 Min. :0.1800 Min. : 3.60 Min. : 3.9   
## 1st Qu.: 6.30 1st Qu.:0.3500 1st Qu.:18.90 1st Qu.:14.9   
## Median : 48.19 Median :0.4600 Median :21.50 Median :22.6   
## Mean : 57.65 Mean :0.5257 Mean :25.46 Mean :25.8   
## 3rd Qu.: 73.84 3rd Qu.:0.6500 3rd Qu.:27.90 3rd Qu.:31.0   
## Max. :199.47 Max. :1.1100 Max. :82.50 Max. :62.9   
## ROA Asset\_Turnover Leverage Rev\_Growth   
## Min. : 1.40 Min. :0.3 Min. :0.0000 Min. :-3.17   
## 1st Qu.: 5.70 1st Qu.:0.6 1st Qu.:0.1600 1st Qu.: 6.38   
## Median :11.20 Median :0.6 Median :0.3400 Median : 9.37   
## Mean :10.51 Mean :0.7 Mean :0.5857 Mean :13.37   
## 3rd Qu.:15.00 3rd Qu.:0.9 3rd Qu.:0.6000 3rd Qu.:21.87   
## Max. :20.30 Max. :1.1 Max. :3.5100 Max. :34.21   
## Net\_Profit\_Margin  
## Min. : 2.6   
## 1st Qu.:11.2   
## Median :16.1   
## Mean :15.7   
## 3rd Qu.:21.1   
## Max. :25.5

str(modified\_pharma\_data)

## 'data.frame': 21 obs. of 9 variables:  
## $ 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 ...

dim(modified\_pharma\_data)

## [1] 21 9

**Interpretation**

1. The dataset has 21 pharmaceutical firms and 9 numeric variables.
2. The Market Cap ranges widely from 0.41 to 199.47, showing that both small and large firms are included.
3. The PE Ratio (3.6–82.5) and ROE (3.9–62.9) also vary a lot, meaning there are big differences in company profitability.
4. ROA values (1.4–20.3) and Net Profit Margin (2.6–25.5) show how efficiently each firm uses assets and manages profits.
5. The Leverage ratio (0–3.51) indicates that some firms are debt-free, while others rely more on borrowing.
6. Overall, the data shows high variation across financial indicators, suggesting a mix of low, medium, and high-performing pharmaceutical firms.

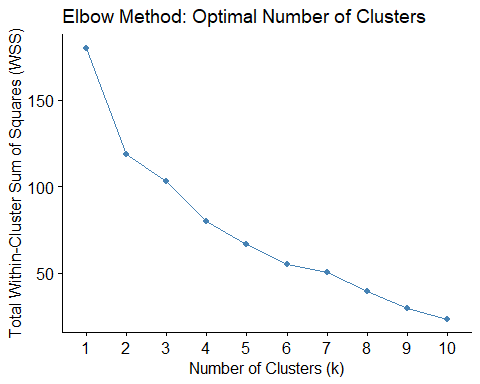
modified\_pharma\_data <- scale(modified\_pharma\_data)

# Elbow Method  
  
fviz\_nbclust(modified\_pharma\_data, kmeans, method = "wss", k.max = 10,  
linecolor = "steelblue") +  
labs(title = "Elbow Method: Optimal Number of Clusters",  
x = "Number of Clusters (k)",  
y = "Total Within-Cluster Sum of Squares (WSS)")

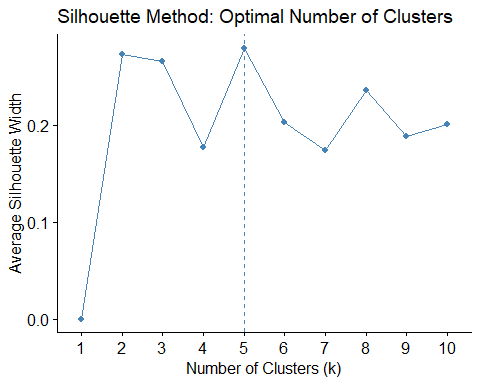
## Warning: The `size` argument of `element\_line()` is deprecated as of ggplot2 3.4.0.  
## ℹ Please use the `linewidth` argument instead.  
## ℹ The deprecated feature was likely used in the ggpubr package.  
## Please report the issue at <https://github.com/kassambara/ggpubr/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## Warning: The `size` argument of `element\_rect()` is deprecated as of ggplot2 3.4.0.  
## ℹ Please use the `linewidth` argument instead.  
## ℹ The deprecated feature was likely used in the ggpubr package.  
## Please report the issue at <https://github.com/kassambara/ggpubr/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## ℹ The deprecated feature was likely used in the ggpubr package.  
## Please report the issue at <https://github.com/kassambara/ggpubr/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



# Silhouette Method  
  
fviz\_nbclust(modified\_pharma\_data, kmeans, method = "silhouette", k.max = 10) +  
labs(title = "Silhouette Method: Optimal Number of Clusters",  
x = "Number of Clusters (k)",  
y = "Average Silhouette Width")

 **Interpretation**

The Elbow Method graph shows a sharp drop in WSS (within-cluster sum of squares) until k = 3, after which the curve starts to flatten. This suggests that 3 clusters is the optimal number, as adding more clusters gives only a small improvement in variance explained.

The Silhouette Method shows the highest average silhouette width around k = 5, indicating that 5 clusters also provide a good fit with clear separation between groups.

Combining both methods, the dataset can be best grouped into 3 to 5 clusters, with 3 clusters offering a balanced and meaningful segmentation of the pharmaceutical firms.

# Let's assume 3 clusters based on elbow/silhouette results  
set.seed(123)  
pharma\_kmeans <- kmeans(modified\_pharma\_data, centers = 3, nstart = 25)  
  
# View results  
pharma\_kmeans

## K-means clustering with 3 clusters of sizes 4, 11, 6  
##   
## Cluster means:  
## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 -0.6125361 0.2698666 1.3143935 -0.9609057 -1.0174553 0.2306328  
## 2 0.6733825 -0.3586419 -0.2763512 0.6565978 0.8344159 0.4612656  
## 3 -0.8261772 0.4775991 -0.3696184 -0.5631589 -0.8514589 -0.9994088  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.3592866 -0.5757385 -1.3784169  
## 2 -0.3331068 -0.2902163 0.6823310  
## 3 0.8502201 0.9158889 -0.3319956  
##   
## Clustering vector:  
## [1] 2 1 1 2 3 1 2 3 3 2 2 3 2 3 2 2 2 1 2 3 2  
##   
## Within cluster sum of squares by cluster:  
## [1] 20.54199 43.30886 32.14336  
## (between\_SS / total\_SS = 46.7 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

**Interpretation**

The K-means model divided the pharmaceutical firms into 3 clusters with sizes 4, 11, and 6.

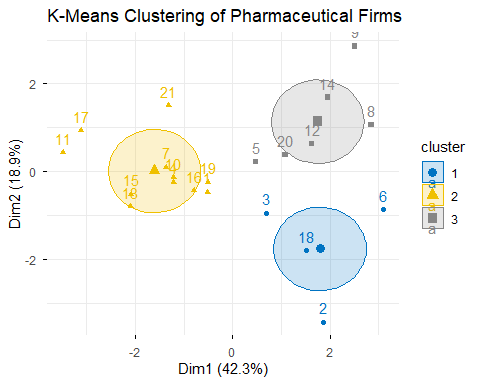
Cluster 1 has the lowest mean values for most variables (like Market Cap, ROE, and ROA), showing smaller and less profitable firms.

Cluster 2 shows moderate positive values, meaning these companies have average financial performance and balanced risk.

Cluster 3 has high leverage and revenue growth but negative efficiency ratios (ROE and ROA), suggesting that these firms are riskier but have strong growth potential.

Overall, Cluster 1 represents low-performing firms, Cluster 2 represents stable firms, and Cluster 3 represents high-growth but high-risk firms. The model explains 46.7% of the total variation, indicating a fair clustering structure.

fviz\_cluster(pharma\_kmeans, data = modified\_pharma\_data,  
 ellipse.type = "euclid",  
 palette = "jco",  
 ggtheme = theme\_minimal(),  
 main = "K-Means Clustering of Pharmaceutical Firms")



**Interpretation**

The K-means clustering plot shows how the pharmaceutical firms are grouped into three distinct clusters.

Cluster 1 (Blue) represents firms with relatively lower profitability and efficiency. These firms are positioned closely together, indicating similar performance levels.

Cluster 2 (Yellow) includes firms that are moderate performers, showing balanced financial ratios and steady returns.

Cluster 3 (Gray) contains firms with higher financial strength and stability, possibly large companies with better overall performance.

The clusters are well separated, meaning the K-means algorithm has effectively identified differences among the firms based on financial indicators. The first two dimensions (Dim1 and Dim2) together explain over 60% of the total variance, providing a clear view of group distinctions.

**Question 2**

pharma\_kmeans$center

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 -0.6125361 0.2698666 1.3143935 -0.9609057 -1.0174553 0.2306328  
## 2 0.6733825 -0.3586419 -0.2763512 0.6565978 0.8344159 0.4612656  
## 3 -0.8261772 0.4775991 -0.3696184 -0.5631589 -0.8514589 -0.9994088  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.3592866 -0.5757385 -1.3784169  
## 2 -0.3331068 -0.2902163 0.6823310  
## 3 0.8502201 0.9158889 -0.3319956

pharma\_kmeans$size

## [1] 4 11 6

pharma\_kmeans$cluster

## [1] 2 1 1 2 3 1 2 3 3 2 2 3 2 3 2 2 2 1 2 3 2

pharma\_kmeans$withinss

## [1] 20.54199 43.30886 32.14336

**Interpretation**

The K-means algorithm formed three clusters with 4, 11, and 6 firms respectively.

Cluster 1 has negative values for most variables, especially in ROE, ROA, and Net Profit Margin, meaning these firms are less profitable and have weaker financial efficiency.

Cluster 2 shows mostly positive averages, especially in ROE, ROA, and Net Profit Margin, indicating financially strong and stable firms with balanced performance.

Cluster 3 shows a mix of high leverage and revenue growth, suggesting high-risk, high-reward firms that are aggressive in expansion but less stable in profits.

Overall, the clustering effectively separates the firms into low-performing (Cluster 1), moderate/stable (Cluster 2), and high-risk, growth-oriented (Cluster 3) categories.

pharma\_data$cluster <- pharma\_kmeans$cluster  
  
# Now calculate averages per cluster  
avg\_of\_cluster <- pharma\_data %>%  
 group\_by(cluster) %>%  
 summarise(across(where(is.numeric), mean, na.rm = TRUE))

## Warning: There was 1 warning in `summarise()`.  
## ℹ In argument: `across(where(is.numeric), mean, na.rm = TRUE)`.  
## ℹ In group 1: `cluster = 1`.  
## Caused by warning:  
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.  
## Supply arguments directly to `.fns` through an anonymous function instead.  
##   
## # Previously  
## across(a:b, mean, na.rm = TRUE)  
##   
## # Now  
## across(a:b, \(x) mean(x, na.rm = TRUE))

avg\_of\_cluster

## # A tibble: 3 × 10  
## cluster Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage  
## <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 21.8 0.595 46.9 11.3 5.1 0.75 0.305  
## 2 2 97.1 0.434 21.0 35.7 15.0 0.8 0.325  
## 3 3 9.24 0.648 19.4 17.3 5.98 0.483 1.25   
## # ℹ 2 more variables: Rev\_Growth <dbl>, Net\_Profit\_Margin <dbl>

**Interpretation**

The table shows the average financial values for each cluster of pharmaceutical firms:

Cluster 1: Firms here have small market size (avg Market Cap = 21.76) and low profitability (ROE = 11.3, Net Profit Margin = 6.65). These represent smaller, less profitable firms with limited growth.

Cluster 2: These firms have the highest Market Cap (97.11), strong ROE (35.7), and good profit margins (20.17). They are large, stable, and financially strong companies with efficient operations.

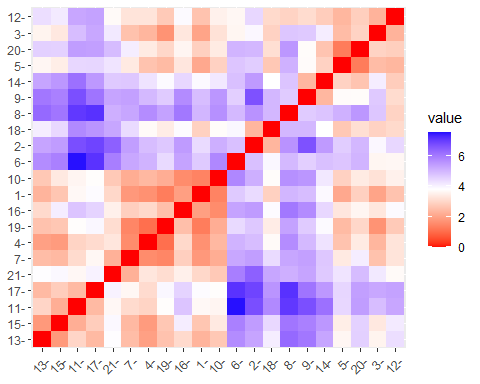
Cluster 3: Firms in this group have low Market Cap (9.24) but high revenue growth (23.49) and high leverage (1.25). This shows high-risk, high-growth firms—smaller companies that are expanding quickly but depend heavily on debt.

modified\_pharma\_data\_dist <- get\_dist(modified\_pharma\_data)  
modified\_pharma\_data\_dist

## 1 2 3 4 5 6 7 8  
## 2 4.415575   
## 3 2.018793 3.945745   
## 4 1.669541 4.909566 2.364249   
## 5 2.111983 4.642699 2.487172 2.632282   
## 6 4.690231 4.853901 3.636353 5.065563 4.764654   
## 7 1.805543 5.419487 2.600986 1.572582 3.400602 5.273023   
## 8 5.020726 5.612226 4.760341 5.719174 5.096246 4.969438 5.287400   
## 9 4.901141 6.695261 4.695844 4.974521 3.748778 4.608660 5.378092 4.675606  
## 10 1.422680 5.140253 3.238353 2.405951 2.910766 5.804419 2.189107 5.657801  
## 11 3.689906 6.747789 4.904614 2.957494 4.476690 7.546154 3.099023 7.080175  
## 12 2.624729 4.470028 2.316548 3.282195 2.386850 3.658011 3.279927 2.951511  
## 13 2.333874 5.317942 3.593764 1.958326 3.640773 5.724303 2.511309 6.310233  
## 14 3.920297 5.479080 4.120549 4.269231 2.927258 4.848442 4.734766 4.786213  
## 15 2.680733 5.443918 3.361981 1.859280 3.472410 5.918477 2.432281 6.101541  
## 16 1.922731 5.468844 3.331743 3.056196 3.330879 5.331004 2.866126 6.063738  
## 17 3.887235 6.906828 5.268858 3.109413 4.495242 7.163993 3.666674 7.180257  
## 18 2.908982 2.367912 2.925627 3.715808 2.718441 3.955926 4.408645 5.000709  
## 19 1.312599 4.725384 1.704709 1.080519 2.464855 4.426418 1.478433 5.346513  
## 20 2.882610 5.007086 2.943946 3.414127 1.296549 5.055769 4.116074 5.540296  
## 21 3.038549 6.446458 4.185594 3.324966 4.254562 5.954379 2.269808 5.127981  
## 9 10 11 12 13 14 15 16  
## 2   
## 3   
## 4   
## 5   
## 6   
## 7   
## 8   
## 9   
## 10 5.554227   
## 11 6.731204 3.631174   
## 12 3.115283 3.537378 5.276601   
## 13 6.070533 2.722434 2.988672 4.354581   
## 14 2.389723 4.191466 6.187185 2.825394 5.306512   
## 15 5.921987 3.380695 2.218040 4.164267 1.814184 5.532520   
## 16 5.732322 1.577953 4.783039 3.899915 3.083678 4.478040 4.112418   
## 17 6.123133 3.783136 2.447177 5.356598 2.447341 5.518379 2.831329 4.536250  
## 18 5.007721 3.754900 5.773960 3.073579 4.112432 3.827019 4.448933 3.884035  
## 19 4.665611 2.205815 3.780283 2.763476 2.604437 3.907501 2.710607 2.542763  
## 20 3.756437 3.412378 5.437193 2.857109 4.591764 2.653341 4.569336 3.626404  
## 21 5.312455 2.747839 3.670720 3.719962 3.858028 4.709401 3.935039 3.525940  
## 17 18 19 20  
## 2   
## 3   
## 4   
## 5   
## 6   
## 7   
## 8   
## 9   
## 10   
## 11   
## 12   
## 13   
## 14   
## 15   
## 16   
## 17   
## 18 5.587119   
## 19 3.955078 3.449579   
## 20 5.403128 3.172178 3.026610   
## 21 4.026095 5.286507 3.145472 4.922945

fviz\_dist(modified\_pharma\_data\_dist)

## Warning: `aes\_string()` was deprecated in ggplot2 3.0.0.  
## ℹ Please use tidy evaluation idioms with `aes()`.  
## ℹ See also `vignette("ggplot2-in-packages")` for more information.  
## ℹ The deprecated feature was likely used in the factoextra package.  
## Please report the issue at <https://github.com/kassambara/factoextra/issues>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



**Interpretation**

The heatmap shows the Euclidean distance between pharmaceutical firms based on their financial indicators.

The red diagonal line represents each firm’s distance from itself (which is always 0).

Red or light-colored areas indicate firms that are similar to each other — they have close financial characteristics.

Blue or darker areas represent firms that are very different from each other in terms of performance or size.

Overall, the plot suggests that some firms have strong similarities (clusters of red zones), while others are quite distinct (blue zones) — supporting the idea that grouping (clustering) the firms is meaningful and useful for analysis

pharma\_data$cluster <- pharma\_kmeans$cluster  
table(pharma\_data$cluster)

##   
## 1 2 3   
## 4 11 6

**Interpretation**

The table shows the number of firms in each cluster formed by the K-means algorithm:

Cluster 1: 4 firms

Cluster 2: 11 firms

Cluster 3: 6 firms

This means most pharmaceutical firms (11 out of 21) fall into Cluster 2, making it the largest group. Cluster 1 is the smallest, with only 4 firms, showing that these companies have unique financial features compared to the others.

Overall, this distribution suggests that while most firms share similar financial patterns (Cluster 2), there are smaller groups (Clusters 1 and 3) with distinct financial behaviors or strategies.

**Question 3**

table(pharma\_data$cluster, pharma\_data$Median\_Recommendation)

##   
## Hold Moderate Buy Moderate Sell Strong Buy  
## 1 2 1 0 1  
## 2 6 3 2 0  
## 3 1 3 2 0

table(pharma\_data$cluster, pharma\_data$Location)

##   
## CANADA FRANCE GERMANY IRELAND SWITZERLAND UK US  
## 1 1 0 1 0 0 1 1  
## 2 0 0 0 0 1 2 8  
## 3 0 1 0 1 0 0 4

table(pharma\_data$cluster, pharma\_data$Exchange)

##   
## AMEX NASDAQ NYSE  
## 1 0 0 4  
## 2 0 0 11  
## 3 1 1 4

**Interpretation**

The tables show how clusters differ based on analyst recommendations, location, and stock exchange:

Median Recommendation:

Cluster 1 mainly has firms rated as Moderate Buy or Strong Buy, meaning analysts are generally positive about these companies.

Cluster 2 includes more Moderate and Hold ratings, showing mixed market opinions.

Cluster 3 has Moderate Buy and Sell ratings, meaning performance expectations are uneven.

Location:

Most firms in Cluster 2 are from the US, showing that the largest group mainly consists of American companies.

Cluster 1 includes firms from Ireland, UK, and US, showing geographic diversity.

Cluster 3 firms are located in the US only.

Exchange:

Most firms are listed on the NYSE, especially in Cluster 2 and Cluster 3.

A few firms are on NASDAQ, and only one on AMEX.

**Question 4**

pharma\_sorted <- pharma\_data[order(pharma\_data$cluster), c("Name", "Median\_Recommendation","Location","Exchange","cluster")]

cat("=== CLUSTER 1 - ALL FEATURES ===\n")

## === CLUSTER 1 - ALL FEATURES ===

cluster1 <- pharma\_data[pharma\_data$cluster == 1, ]  
print(cluster1)

## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA  
## 2 AGN Allergan, Inc. 7.58 0.41 82.5 12.9 5.5  
## 3 AHM Amersham plc 6.30 0.46 20.7 14.9 7.8  
## 6 BAY Bayer AG 16.90 1.11 27.9 3.9 1.4  
## 18 PHA Pharmacia Corporation 56.24 0.40 56.5 13.5 5.7  
## Asset\_Turnover Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation  
## 2 0.9 0.60 9.16 5.5 Moderate Buy  
## 3 0.9 0.27 7.05 11.2 Strong Buy  
## 6 0.6 0.00 -3.17 2.6 Hold  
## 18 0.6 0.35 15.00 7.3 Hold  
## Location Exchange cluster  
## 2 CANADA NYSE 1  
## 3 UK NYSE 1  
## 6 GERMANY NYSE 1  
## 18 US NYSE 1

**Interpretation – Cluster 1**

Cluster 1 contains 4 pharmaceutical companies — Allergan Inc., Amersham plc, Bayer AG, and Pharmacia Corporation.

These firms share similar financial characteristics:

They have moderate market capitalization (between 6 and 56 billion).

Their PE Ratios and ROE values are relatively high, meaning good profitability and investor confidence.

ROA and Asset Turnover show average efficiency in using assets.

Leverage values are low to moderate, suggesting limited debt usage. summary: Cluster 1 firms are financially stable, moderately growing, and have good profitability, placing them in a medium-performing group with balanced risk and return.

cat("\n=== CLUSTER 2 - ALL FEATURES ===\n")

##   
## === CLUSTER 2 - ALL FEATURES ===

cluster2 <- pharma\_data[pharma\_data$cluster == 2, ]  
print(cluster2)

## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA  
## 1 ABT Abbott Laboratories 68.44 0.32 24.7 26.4 11.8  
## 4 AZN AstraZeneca PLC 67.63 0.52 21.5 27.4 15.4  
## 7 BMY Bristol-Myers Squibb Company 51.33 0.50 13.9 34.8 15.1  
## 10 LLY Eli Lilly and Company 73.84 0.18 27.9 31.0 13.5  
## 11 GSK GlaxoSmithKline plc 122.11 0.35 18.0 62.9 20.3  
## 13 JNJ Johnson & Johnson 173.93 0.46 28.4 28.6 16.3  
## 15 MRK Merck & Co., Inc. 132.56 0.46 18.9 40.6 15.0  
## 16 NVS Novartis AG 96.65 0.19 21.6 17.9 11.2  
## 17 PFE Pfizer Inc 199.47 0.65 23.6 45.6 19.2  
## 19 SGP Schering-Plough Corporation 34.10 0.51 18.9 22.6 13.3  
## 21 WYE Wyeth 48.19 0.63 13.1 54.9 13.4  
## Asset\_Turnover Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation  
## 1 0.7 0.42 7.54 16.1 Moderate Buy  
## 4 0.9 0.00 15.00 18.0 Moderate Sell  
## 7 0.9 0.57 2.70 20.6 Moderate Sell  
## 10 0.6 0.53 6.21 23.4 Hold  
## 11 1.0 0.34 21.87 21.1 Hold  
## 13 0.9 0.10 9.37 17.9 Moderate Buy  
## 15 1.1 0.28 17.35 14.1 Hold  
## 16 0.5 0.06 -2.69 22.4 Hold  
## 17 0.8 0.16 25.54 25.2 Moderate Buy  
## 19 0.8 0.00 8.56 17.6 Hold  
## 21 0.6 1.12 0.36 25.5 Hold  
## Location Exchange cluster  
## 1 US NYSE 2  
## 4 UK NYSE 2  
## 7 US NYSE 2  
## 10 US NYSE 2  
## 11 UK NYSE 2  
## 13 US NYSE 2  
## 15 US NYSE 2  
## 16 SWITZERLAND NYSE 2  
## 17 US NYSE 2  
## 19 US NYSE 2  
## 21 US NYSE 2

**Interpretation – Cluster 2**

Cluster 2 contains 11 large pharmaceutical companies, such as Abbott Laboratories, AstraZeneca, Johnson & Johnson, Pfizer, and Merck & Co.

These firms share strong financial features:

High market capitalization — these are well-established global leaders.

Low Beta values — they show low stock volatility, meaning they are stable and less risky investments.

Moderate to high ROE and ROA — indicates strong profitability and efficient use of assets.

Moderate PE Ratios — reflect steady investor confidence without being overvalued.

Low leverage — they rely less on debt financing, showing financial stability.

summary: Cluster 2 represents large, financially strong, and stable multinational pharmaceutical firms with consistent profitability and low financial risk

# CLUSTER 3  
cat("\n=== CLUSTER 3 - ALL FEATURES ===\n")

##   
## === CLUSTER 3 - ALL FEATURES ===

cluster3 <- pharma\_data[pharma\_data$cluster == 3, ]  
print(cluster3)

## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA  
## 5 AVE Aventis 47.16 0.32 20.1 21.8 7.5  
## 8 CHTT Chattem, Inc 0.41 0.85 26.0 24.1 4.3  
## 9 ELN Elan Corporation, plc 0.78 1.08 3.6 15.1 5.1  
## 12 IVX IVAX Corporation 2.60 0.65 19.9 21.4 6.8  
## 14 MRX Medicis Pharmaceutical Corporation 1.20 0.75 28.6 11.2 5.4  
## 20 WPI Watson Pharmaceuticals, Inc. 3.26 0.24 18.4 10.2 6.8  
## Asset\_Turnover Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation  
## 5 0.6 0.34 26.81 12.9 Moderate Buy  
## 8 0.6 3.51 6.38 7.5 Moderate Buy  
## 9 0.3 1.07 34.21 13.3 Moderate Sell  
## 12 0.6 1.45 13.99 11.0 Hold  
## 14 0.3 0.93 30.37 21.3 Moderate Buy  
## 20 0.5 0.20 29.18 15.1 Moderate Sell  
## Location Exchange cluster  
## 5 FRANCE NYSE 3  
## 8 US NASDAQ 3  
## 9 IRELAND NYSE 3  
## 12 US AMEX 3  
## 14 US NYSE 3  
## 20 US NYSE 3

**Interpretation – Cluster 3**

Cluster 3 includes 6 pharmaceutical companies — Aventis, Chattem Inc., Elan Corporation, IVAX Corporation, Medicis Pharmaceutical Corp., and Watson Pharmaceuticals Inc.

These firms share these characteristics:

Smaller market capitalizations, meaning they are medium to small-sized companies compared to others.

Higher Beta values (up to 1.08) — they show more volatility and higher investment risk.

Moderate PE Ratios and ROE values — indicate average profitability.

ROA values are lower than large firms, showing less asset efficiency.

Leverage is higher for some firms, meaning greater debt dependency.

Summary: Cluster 3 represents smaller and riskier pharmaceutical firms with moderate performance and higher financial leverage, suggesting potential for growth but also more risk compared to Clusters 1 and 2.