

Assignment 04

GUDA SAI ROHITH

```
# Loading the necessary libraries
library(cluster)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- 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

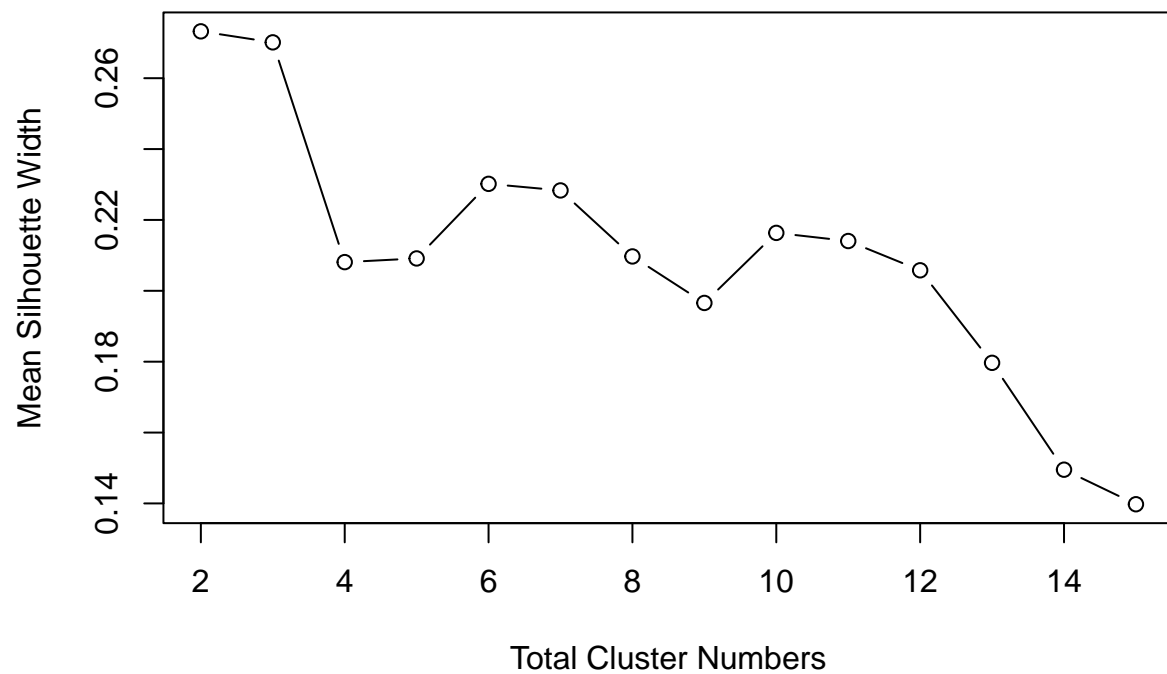
# Read the dataset
df <- read.csv("Pharmaceuticals.csv")

# Part a:

# Select only the numerical variables for clustering
ps <- df[, c("Market_Cap", "Beta", "PE_Ratio", "ROE", "ROA", "Asset_Turnover", "Leverage", "Rev_Growth")]

# Standardize the data
ps <- scale(ps)

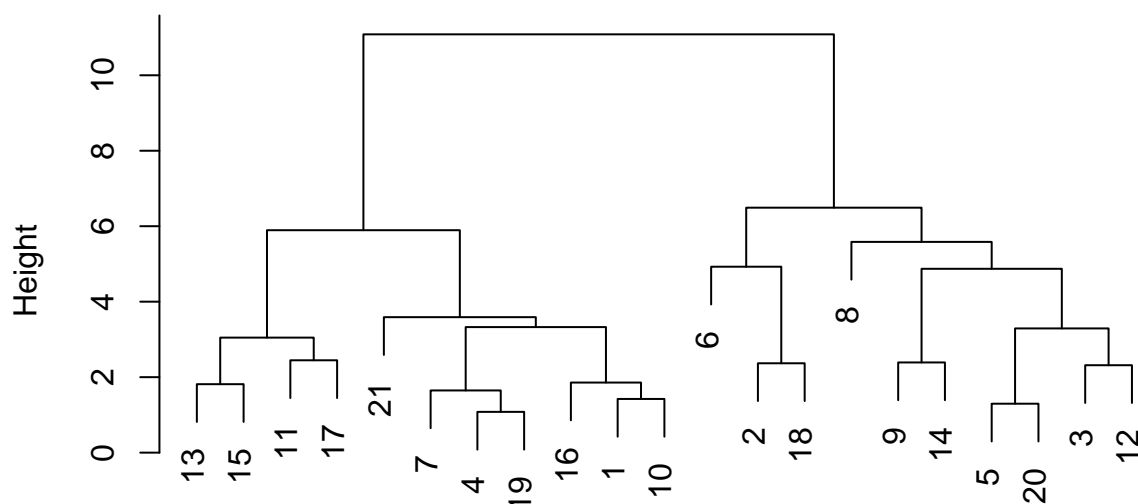
# Determine the number of clusters using the Silhouette Method
sil_width <- numeric(15)
for (i in 2:15) {
  fit <- hclust(dist(ps), method = "ward.D2")
  clus_labels <- cutree(fit, k = i)
  sil_info <- silhouette(clus_labels, dist(ps))
  sil_width[i] <- mean(sil_info[, 3])
}
plot(2:15, sil_width[2:15], type = "b", xlab = "Total Cluster Numbers", ylab = "Mean Silhouette Width")
```



```
# Perform hierarchical clustering (Let's choose k=3 based on the Silhouette Method)
fit <- hclust(dist(ps), method = "ward.D2")
clus_labels <- cutree(fit, k = 3)

# Plot the dendrogram
plot(fit)
```

Cluster Dendrogram



```
dist(ps)
hclust (*, "ward.D2")
```

```
# Part b:
```

```
# Summarize clusters by median
```

```
cluster_summary <- aggregate(ps, by = list(clus_labels), FUN = median)
print("Cluster Summary:")
```

```
## [1] "Cluster Summary:"
```

```
print(cluster_summary)
```

```
##   Group.1 Market_Cap      Beta  PE_Ratio      ROE      ROA Asset_Turnover
## 1      1  0.2762415 -0.2559560 -0.2429088  0.3450295  0.8429577    0.4612656
## 2      2 -0.6953818 -0.4507051  1.9029802 -0.8548399 -0.9422871   -0.4612656
## 3      3 -0.9393967  0.4840907 -0.3287443 -0.7089994 -0.6979905   -0.4612656
##   Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3912841 -0.4354459      0.7474438
## 2 -0.3016910 -0.3811391     -1.5536671
## 3  0.4406517  1.2163867     -0.4259704
```

```
# Part c:
```

```
# Add cluster assignment to the original dataframe
```

```
df$Cluster <- clus_labels
```

```

pattern_median_rec <- aggregate(Median_Recommendation ~ Cluster, data = df, function(z) names(which.max(table(z))))
pattern_location <- aggregate(Location ~ Cluster, data = df, function(z) names(which.max(table(z))))
pattern_exchange <- aggregate(Exchange ~ Cluster, data = df, function(z) names(which.max(table(z))))

print("Patterns in additional variables:")

```

```
## [1] "Patterns in additional variables:"
```

```
print(pattern_median_rec)
```

```
##   Cluster Median_Recommendation
## 1      1                      Hold
## 2      2                      Hold
## 3      3          Moderate Buy
```

```
print(pattern_location)
```

```
##   Cluster Location
## 1      1        US
## 2      2   CANADA
## 3      3        US
```

```
print(pattern_exchange)
```

```
##   Cluster Exchange
## 1      1      NYSE
## 2      2      NYSE
## 3      3      NYSE
```

```
# Part d:
```

```
# Naming the clusters based on interpretation and pattern
```

```
cluster_names <- c("High Growth, Emerging", "Stable, Low Leverage", "Diverse, High Profit")
df$Cluster_Name <- as.factor(cluster_names[df$Cluster])
```

```
# Show the first few rows of df to confirm cluster assignment and naming
```

```
head(df)
```

```
##   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
##	Cluster	Cluster_Name				
## 1	1	High Growth, Emerging				
## 2	2	Stable, Low Leverage				
## 3	3	Diverse, High Profit				
## 4	1	High Growth, Emerging				
## 5	3	Diverse, High Profit				
## 6	2	Stable, Low Leverage				