Assignment 4

library(tidyverse)

## ── 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 3.4.3 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ 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(scales)

##   
## Attaching package: 'scales'  
##   
## The following object is masked from 'package:purrr':  
##   
## discard  
##   
## The following object is masked from 'package:readr':  
##   
## col\_factor

library(factoextra)

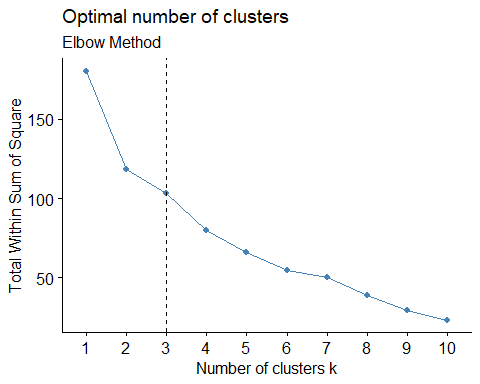
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(cluster)  
pharma\_data <- read.csv("C:\\Users\\Ireland\\Downloads\\Pharmaceuticals.csv")

numerical\_data <- pharma\_data[,3:11]  
  
scaled\_data <- scale(numerical\_data)  
cor\_matrix <- cor(numerical\_data)  
print(cor\_matrix)

## Market\_Cap Beta PE\_Ratio ROE ROA  
## Market\_Cap 1.000000000 -0.31250762 -0.08798317 0.61952576 0.80908852  
## Beta -0.312507620 1.00000000 -0.19716312 -0.20273345 -0.42583638  
## PE\_Ratio -0.087983169 -0.19716312 1.00000000 -0.32205434 -0.29207790  
## ROE 0.619525759 -0.20273345 -0.32205434 1.00000000 0.83168600  
## ROA 0.809088517 -0.42583638 -0.29207790 0.83168600 1.00000000  
## Asset\_Turnover 0.507917513 -0.32069694 0.14974635 0.49612507 0.61977107  
## Leverage -0.408937481 0.40116206 -0.03985770 0.01560562 -0.36535802  
## Rev\_Growth 0.003788982 0.08807135 -0.15499183 -0.01905389 -0.02118403  
## Net\_Profit\_Margin 0.516711077 -0.34546582 -0.46240116 0.63395830 0.74875756  
## Asset\_Turnover Leverage Rev\_Growth Net\_Profit\_Margin  
## Market\_Cap 0.50791751 -0.40893748 0.003788982 0.51671108  
## Beta -0.32069694 0.40116206 0.088071348 -0.34546582  
## PE\_Ratio 0.14974635 -0.03985770 -0.154991834 -0.46240116  
## ROE 0.49612507 0.01560562 -0.019053892 0.63395830  
## ROA 0.61977107 -0.36535802 -0.021184032 0.74875756  
## Asset\_Turnover 1.00000000 -0.30817546 -0.253024565 0.01862763  
## Leverage -0.30817546 1.00000000 -0.021881004 -0.22135214  
## Rev\_Growth -0.25302457 -0.02188100 1.000000000 0.08478937  
## Net\_Profit\_Margin 0.01862763 -0.22135214 0.084789374 1.00000000

fviz\_nbclust(scaled\_data, kmeans, method = "wss") +  
 geom\_vline(xintercept = 3, linetype = 2) +  
 labs(subtitle = "Elbow Method")



# I chose all numeric variables for clustering because they are key financial indicators that reflect the company’s size, valuation, and profitability. I did not apply weights as I considered all selected features to be equally important for this analysis. The Elbow Method helped in determining the optimal number of clusters, ensuring a balance between precision and computational efficiency. As seen in the Elbow Method plot, the optimal number of clusters is three, where the reduction in within-cluster sum of squares starts to diminish, indicating that adding more clusters doesn’t provide significant benefits. #

set.seed(42)  
kmeans\_result <- kmeans(scaled\_data, centers = 3, nstart = 25)  
pharma\_data$Cluster <- as.factor(kmeans\_result$cluster)  
  
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  
## Cluster  
## 1 1  
## 2 2  
## 3 2  
## 4 1  
## 5 3  
## 6 2

numerical\_analysis <- aggregate(numerical\_data, by=list(Cluster=kmeans\_result$cluster), mean)

mode\_stat <- function(x) {  
 tbl <- table(x)  
 name <- names(tbl)[which.max(tbl)]  
 if (is.null(name)) {  
 return(NA)  
 } else {  
 return(name)  
 }  
}  
  
  
non\_numerical\_data <- pharma\_data[,c(12:14, 15)]  
non\_numerical\_analysis <- aggregate(non\_numerical\_data[-4], by=non\_numerical\_data[4], mode\_stat)

print(non\_numerical\_analysis)

## Cluster Median\_Recommendation Location Exchange  
## 1 1 Hold US NYSE  
## 2 2 Hold CANADA NYSE  
## 3 3 Moderate Buy US NYSE

print(numerical\_analysis)

## Cluster Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage  
## 1 1 97.11364 0.4336364 20.95455 35.7 14.954545 0.8000000 0.3254545  
## 2 2 21.75500 0.5950000 46.90000 11.3 5.100000 0.7500000 0.3050000  
## 3 3 9.23500 0.6483333 19.43333 17.3 5.983333 0.4833333 1.2500000  
## Rev\_Growth Net\_Profit\_Margin  
## 1 10.16455 20.17273  
## 2 7.01000 6.65000  
## 3 23.49000 13.51667

# Cluster 1: High-Growth Moderate Profitability - High expected revenue growth and moderate profitability - moderately positive outlook.  
  
# Cluster 2: Stable Large Profitability - Large, stable companies with high profitability and asset efficiency - moderate expected revenue growth, and analysts suggest holding these stocks.  
  
# Cluster 3: Risky Mid Profitability - Mid-sized, highly leveraged companies with high volatility and lower profitability- analysts suggest holding these stocks due to their riskier nature.  
  
cluster\_names <- c("High-Growth Moderate Pofitabilty", "Stable Large Profitability", "Risky Mid Proitability")  
pharma\_data$Cluster\_Name <- factor(pharma\_data$Cluster, labels = cluster\_names)  
  
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  
## Cluster Cluster\_Name  
## 1 1 High-Growth Moderate Pofitabilty  
## 2 2 Stable Large Profitability  
## 3 2 Stable Large Profitability  
## 4 1 High-Growth Moderate Pofitabilty  
## 5 3 Risky Mid Proitability  
## 6 2 Stable Large Profitability