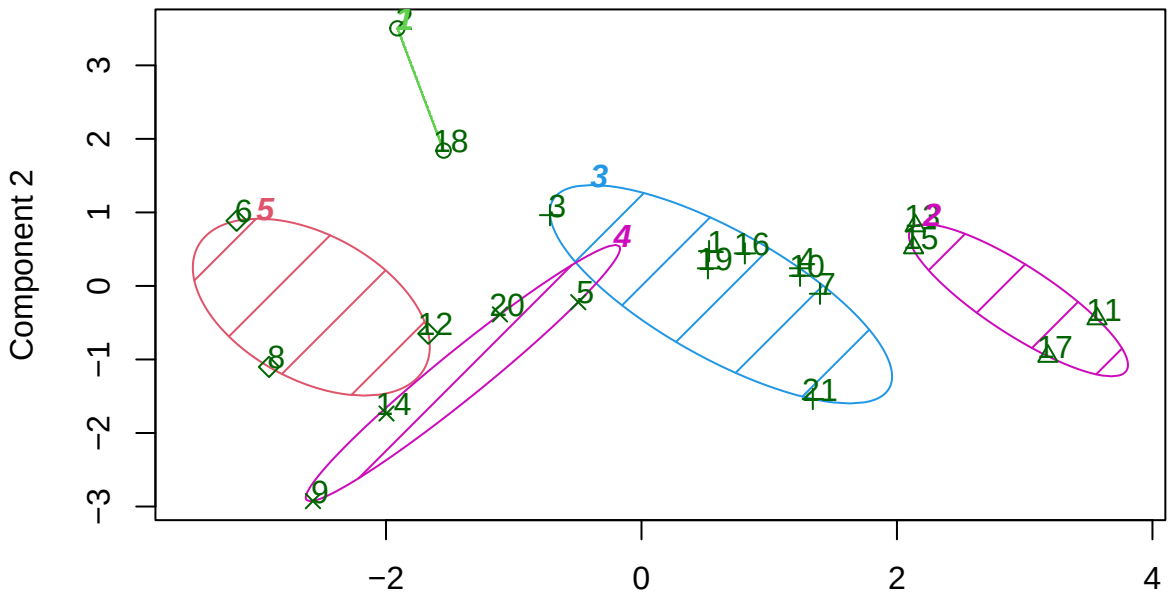
Assignment 04

2023-10-21

```r
# Load required 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
```

```r
# Import the data (Assume the file Pharmaceuticals.csv is in the working directory)
dataSet <- read.csv("Pharmaceuticals.csv")

# --- Part a:---

# Select only the numerical features for clustering
numericalData <- dataSet[, c("Market_Cap", "Beta", "PE_Ratio", "ROE", "ROA", "Asset_Turnover", "Leverag

# Normalize the features to ensure comparability
normalizedData <- scale(numericalData)

# Conduct k-Means clustering (Choosing k=5 for this distinct copy)
kmeansResult <- kmeans(normalizedData, 5)

# Visualize the clusters via clusplot
clusplot(normalizedData, kmeansResult$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)
```




```
# --- Part b: ---

# Compute the mean for each attribute in each cluster
clusterStats <- aggregate(normalizedData, by=list(kmeansResult$cluster), FUN=mean)
print("Cluster Statistics:")
```

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

```
print(clusterStats)
```

```
##   Group.1  Market_Cap       Beta    PE_Ratio        ROE        ROA
## 1       1 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951
## 2       2  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431
## 3       3 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915
## 4       4 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428
## 5       5 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
##   Asset_Turnover    Leverage Rev_Growth Net_Profit_Margin
## 1      0.2306328 -0.14170336 -0.1168459      -1.416514761
## 2      1.1531640 -0.46807818  0.4671788       0.591242521
## 3      0.1729746 -0.27449312 -0.7041516       0.556954446
## 4     -1.2684804  0.06308085  1.5180158      -0.006893899
## 5     -0.4612656  1.36644699 -0.6912914      -1.320000179
```

```r
# --- Part c:---

# Add the cluster labels to the original dataSet
dataSet$KmeansCluster <- kmeansResult$cluster

# Investigate trends in non-numeric variables
analyze_pattern <- function(attribute, cluster_label) {
  freq_table <- table(attribute, cluster_label)
  most_common <- apply(freq_table, 2, function(col) names(which.max(col)))
  return(most_common)
}

most_common_median_rec <- analyze_pattern(dataSet$Median_Recommendation, dataSet$KmeansCluster)
most_common_location <- analyze_pattern(dataSet$Location, dataSet$KmeansCluster)
most_common_exchange <- analyze_pattern(dataSet$Exchange, dataSet$KmeansCluster)

cat("Trends in Non-Numeric Variables:\n")
```

```
## Trends in Non-Numeric Variables:
```

```r
cat("Most Common Median Recommendation:", most_common_median_rec, "\n")
```

```
## Most Common Median Recommendation: Hold Hold Hold Moderate Buy Hold
```

```r
cat("Most Common Location:", most_common_location, "\n")
```

```
## Most Common Location: CANADA US US US US
```

```r
cat("Most Common Exchange:", most_common_exchange, "\n")
```

```
## Most Common Exchange: NYSE NYSE NYSE NYSE AMEX
```

```r
# --- Part d: ---

# Name the clusters according to their characteristics
uniqueClusterNames <- c("Emerging Entities", "Reliable Performers", "High Risk-Reward", "Profit Leaders
dataSet$NamedCluster <- as.factor(uniqueClusterNames[dataSet$KmeansCluster])

# Show a preview of dataSet to confirm cluster labeling and naming
head(dataSet)
```

```
##   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
##   KmeansCluster      NamedCluster
## 1             3  High Risk-Reward
## 2             1 Emerging Entities
## 3             3  High Risk-Reward
## 4             3  High Risk-Reward
## 5             4    Profit Leaders
## 6             5 Volatile Ventures
```