R Notebook

```
library(readr)
Pharmaceuticals <- read_csv("Downloads/Pharmaceuticals.csv")
```

```
## Rows: 21 Columns: 14-- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): Symbol, Name, Median_Recommendation, Location, Exchange
## dbl (9): Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev...
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
library(factoextra)
```

```
## Loading required package: ggplot2
```

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

```
library(ISLR)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v stringr   1.5.1
## v forcats   1.0.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

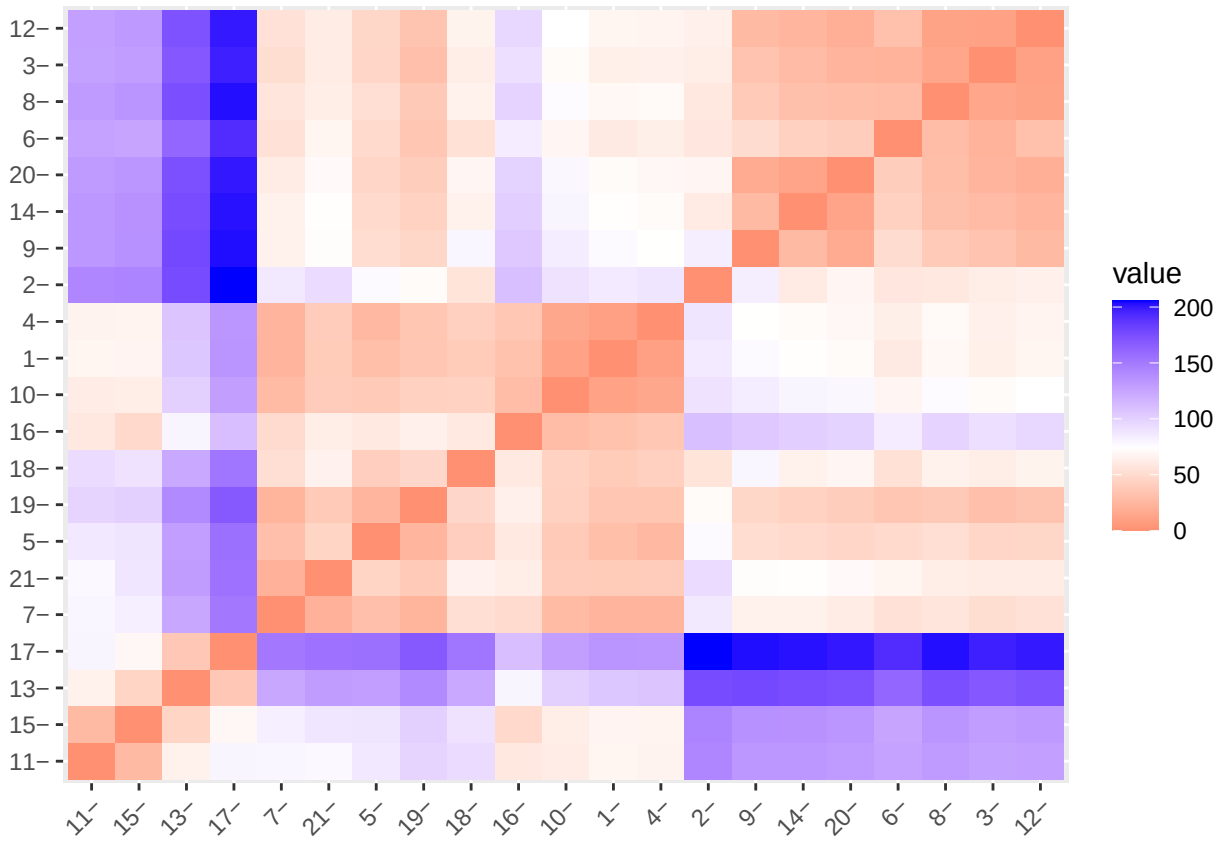
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(flexclust)
```

```
## Loading required package: grid
## Loading required package: lattice
## Loading required package: modeltools
## Loading required package: stats4
```

```
Pharmaceuticals_only_numbers <- Pharmaceuticals[,-1]
Pharmaceuticals_only_numbers <- Pharmaceuticals_only_numbers[,-1]
Pharmaceuticals_only_numbers <- Pharmaceuticals_only_numbers[,-12]
Pharmaceuticals_only_numbers <- Pharmaceuticals_only_numbers[,-11]
Pharmaceuticals_only_numbers <- Pharmaceuticals_only_numbers[,-10]
```

```
pharm <- scale(Pharmaceuticals_only_numbers)
distance <- get_dist(Pharmaceuticals_only_numbers)
fviz_dist(distance)
```




```
Pharmclust <- kmeans(Pharmaceuticals_only_numbers, centers = 4, nstart = 25)
Pharmclust$size
```

```
## [1] 2 8 9 2
```

```
Pharmclust$centers
```

```
##   Market_Cap      Beta PE_Ratio      ROE      ROA Asset_Turnover  Leverage
## 1  186.70000 0.5550000 26.00000 37.10000 17.75000      0.8500000 0.1300000
## 2    4.87875 0.6937500 28.45000 14.21250  5.38750      0.5875000 1.0037500
## 3   60.39778 0.3966667 24.24444 27.81111 11.87778      0.6888889 0.3766667
## 4  127.33500 0.4050000 18.45000 51.75000 17.65000      1.0500000 0.3100000
##   Rev_Growth Net_Profit_Margin
## 1  17.455000           21.5500
## 2  15.896250           10.9375
## 3   8.832222           18.2000
## 4  19.610000           17.6000
```

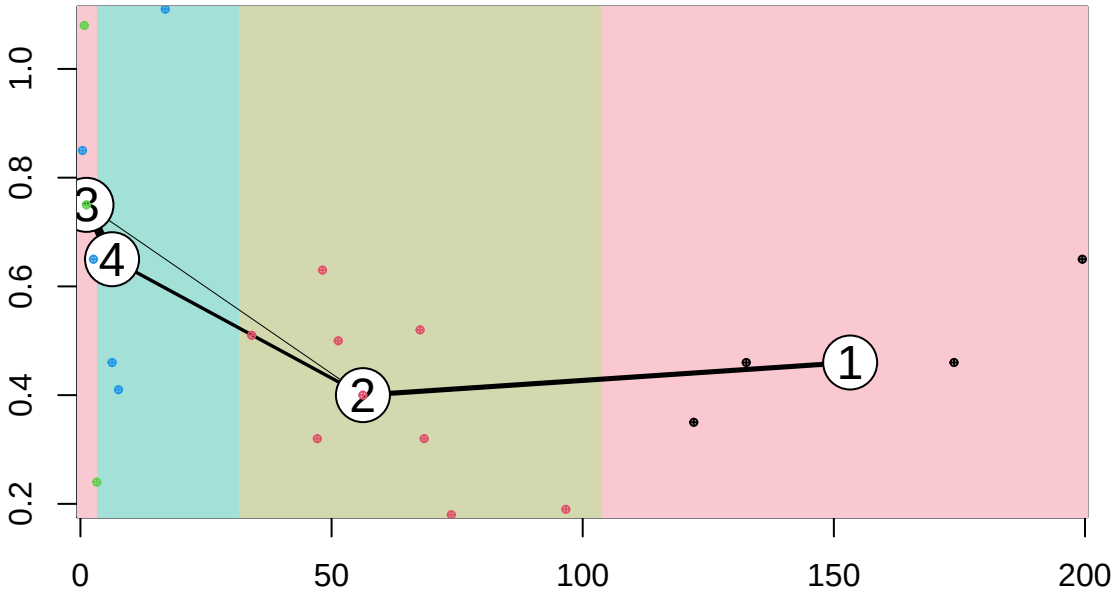
```
# The fviz_cluster package can't be used on my computer, I can't install the correct packages (ggpubr)
```

```
set.seed(123)
pharma_cluster <- kcca(Pharmaceuticals_only_numbers, k = 4, kccaFamily("kmedians"))
pharma_cluster
```

```
## kcca object of family 'kmedians'
##
## call:
## kcca(x = Pharmaceuticals_only_numbers, k = 4, family = kccaFamily("kmedians"))
##
```

```
## cluster sizes:
##
## 1 2 3 4
## 4 9 3 5
```

```
cluster_index <- predict(pharma_cluster)
image(pharma_cluster)
points(Pharmaceuticals_only_numbers, col = cluster_index, pch = 10, cex = .5)
```




```
# By looking at this cluster, we can see that the clustering
# seems to be based off of the market capitalization of each of these companies.
# The Y axis is based off of the beta
```

```
# Cluster 1 is market cap > 100
# Cluster 2 is middle market cap
# Cluster 3 is very small market cap
# Cluster 4 is small market cap
# The 1st cluster all of those variables have hold or moderately buy as a recommendation.
```