Periandri_Anthony_Assignment 4

Anthony Periandri

2025-06-14

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(cluster)
dataset <- read.csv("Pharmaceuticals.csv")
show(dataset)
```

```
##    Symbol                                Name Market_Cap Beta PE_Ratio  ROE  ROA
## 1     ABT                Abbott Laboratories      68.44 0.32     24.7 26.4 11.8
## 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
## 4     AZN                    AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4
## 5     AVE                            Aventis      47.16 0.32     20.1 21.8  7.5
## 6     BAY                           Bayer AG      16.90 1.11     27.9  3.9  1.4
## 7     BMY       Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8 15.1
## 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
## 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
## 12    IVX                   IVAX Corporation       2.60 0.65     19.9 21.4  6.8
## 13    JNJ                  Johnson & Johnson     173.93 0.46     28.4 28.6 16.3
## 14    MRX Medicis Pharmaceutical Corporation       1.20 0.75     28.6 11.2  5.4
## 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
## 18    PHA              Pharmacia Corporation      56.24 0.40     56.5 13.5  5.7
```

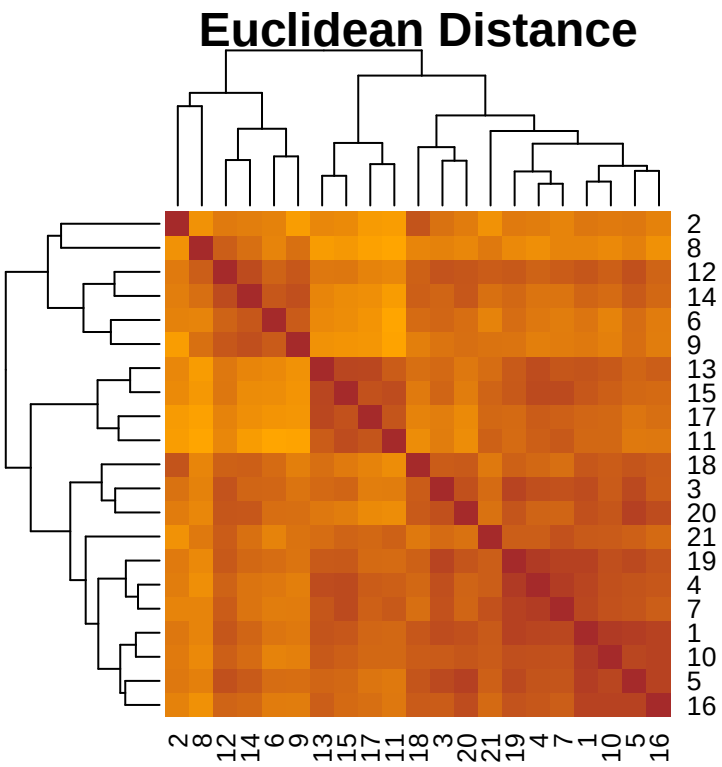
```
## 19    SGP        Schering-Plough Corporation      34.10 0.51     18.9 22.6 13.3
## 20    WPI       Watson Pharmaceuticals, Inc.       3.26 0.24     18.4 10.2  6.8
## 21    WYE                             Wyeth      48.19 0.63     13.1 54.9 13.4
##    Asset_Turnover Leverage Rev_Growth Net_Profit_Margin Median_Recommendation
## 1             0.7     0.42       7.54               16.1          Moderate Buy
## 2             0.9     0.60       9.16                5.5          Moderate Buy
## 3             0.9     0.27       7.05               11.2            Strong Buy
## 4             0.9     0.00      15.00               18.0         Moderate Sell
## 5             0.6     0.34      26.81               12.9          Moderate Buy
## 6             0.6     0.00      -3.17                2.6                 Hold
## 7             0.9     0.57       2.70               20.6         Moderate Sell
## 8             0.6     3.51       6.38                7.5          Moderate Buy
## 9             0.3     1.07      34.21               13.3         Moderate Sell
## 10            0.6     0.53       6.21               23.4                 Hold
## 11            1.0     0.34      21.87               21.1                 Hold
## 12            0.6     1.45      13.99               11.0                 Hold
## 13            0.9     0.10       9.37               17.9          Moderate Buy
## 14            0.3     0.93      30.37               21.3          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
## 18            0.6     0.35      15.00                7.3                 Hold
## 19            0.8     0.00       8.56               17.6                 Hold
## 20            0.5     0.20      29.18               15.1         Moderate Sell
## 21            0.6     1.12       0.36               25.5                 Hold
##       Location Exchange
## 1           US     NYSE
## 2       CANADA     NYSE
## 3           UK     NYSE
## 4           UK     NYSE
## 5       FRANCE     NYSE
## 6      GERMANY     NYSE
## 7           US     NYSE
## 8           US   NASDAQ
## 9      IRELAND     NYSE
## 10          US     NYSE
## 11          UK     NYSE
## 12          US     AMEX
## 13          US     NYSE
## 14          US     NYSE
## 15          US     NYSE
## 16 SWITZERLAND     NYSE
## 17          US     NYSE
## 18          US     NYSE
## 19          US     NYSE
## 20          US     NYSE
## 21          US     NYSE
```

```r
cluster_data <- dataset[, 1:9]
cluster_data <- cluster_data[, sapply(cluster_data, is.numeric)]
normalized_data <- scale(cluster_data)
## Euclidean distance works well with normalized numeric values
euclidean_calc <- dist(normalized_data, method = "euclidean")
print(euclidean_calc)
```

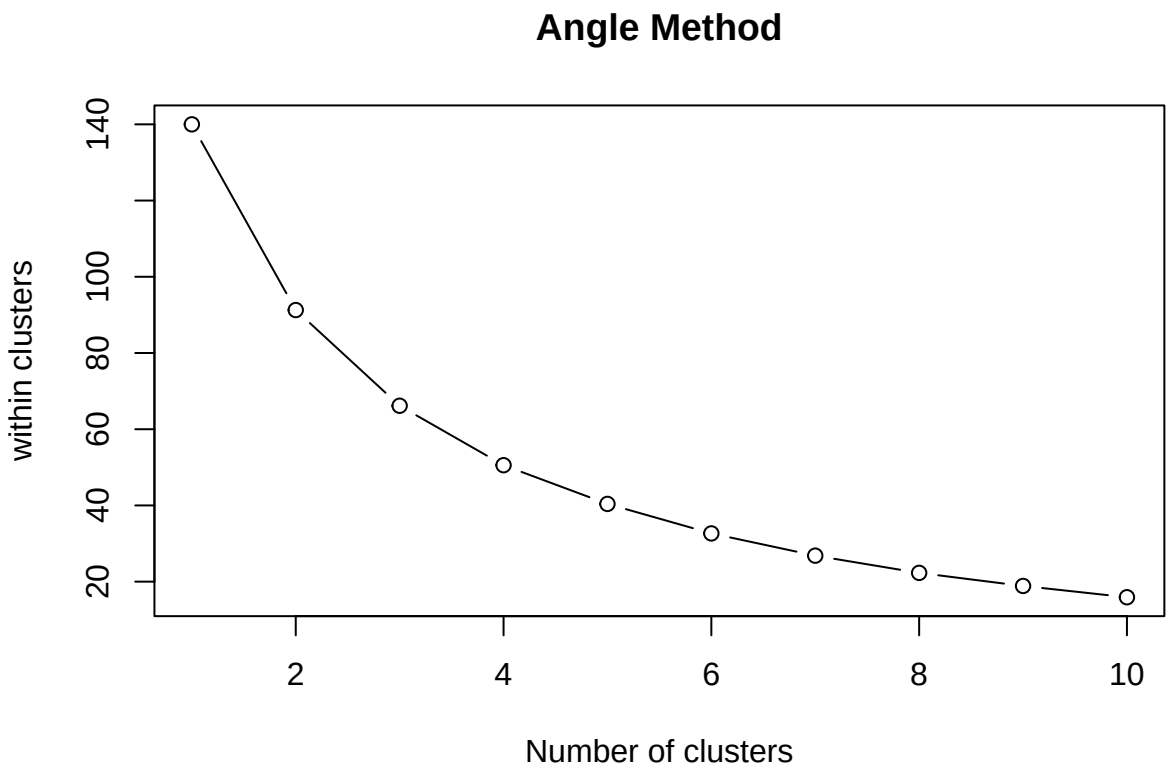
```
##           1         2         3         4         5         6         7
## 2  4.1068779
## 3  1.8750924 3.8442050
## 4  1.4992057 4.4939752 1.9995154
## 5  1.0865518 4.2108035 1.7088476 2.2764541
## 6  4.1019701 4.7031456 3.2634037 4.1770383 3.5881654
## 7  1.6117710 4.8717397 2.1349733 1.0375985 2.3291771 4.4719340
## 8  4.8455221 5.5982924 4.7264409 5.4350598 4.6770919 4.8363796 4.8847038
## 9  4.2440566 6.1865119 3.9881514 4.6053988 3.6879543 2.6709750 4.4218013
## 10 0.8786161 4.3485606 2.6503993 2.1162327 1.5606779 4.7878204 2.1234853
## 11 3.3693538 6.2094802 4.4698816 2.8525325 4.2754310 6.6226673 2.5666239
## 12 2.4381031 4.3689277 2.2295500 3.1026624 2.0656217 3.0545158 2.7519907
## 13 2.3117717 4.9708505 3.4392675 1.8908032 3.1910702 5.1032421 2.4026878
## 14 3.2358725 4.5318281 3.1866549 4.0047608 2.6127360 2.4840891 4.0167606
## 15 2.5109842 5.2315529 3.1997637 1.7488444 3.3602149 5.3392478 1.7822907
## 16 1.3848534 4.7037329 2.7221713 2.5153660 1.3672995 4.3946915 2.8109488
## 17 3.2454860 6.0409901 4.5175593 2.7485384 4.0841838 5.7191527 2.9460904
## 18 2.4915576 2.2918034 2.7727778 3.3389237 2.3492583 3.5258572 3.7535036
## 19 1.2892407 4.3504259 1.3914494 0.9077670 1.6834369 3.6387855 1.3020883
## 20 2.1094659 4.4325141 2.0740103 3.1326842 1.2339469 3.6549616 3.2396296
## 21 2.5997016 5.6243356 3.5218806 2.8272319 2.9466983 4.8139478 2.1329752
##           8         9        10        11        12        13        14
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9  3.8386256
## 10 5.1126730 4.6965593
## 11 6.6232813 6.5305991 3.3246846
## 12 2.8200173 2.4965422 2.9062660 4.9964816
## 13 6.1019654 5.5950961 2.5743618 2.7229222 4.2049809
## 14 3.7108463 2.0257580 3.5614709 6.1390911 1.8223808 4.9272510
## 15 5.9355987 5.7206923 2.8989425 1.9011561 4.1261918 1.5602026 5.2930261
## 16 5.5623021 4.4476591 1.3482583 4.2304456 3.1483279 2.8013002 3.3272885
## 17 6.4244119 5.7955429 3.3430268 2.3426379 4.7872275 1.6155126 5.4688355
## 18 4.9393759 4.6063082 2.7288511 5.3412865 3.0200337 3.7474092 2.8563980
## 19 5.1163733 3.9935735 2.0097643 3.5432665 2.5265885 2.6030034 3.3247189
## 20 5.0094735 3.7186423 2.3920796 5.3187656 2.4253722 4.2056236 2.4770786
## 21 4.2982545 3.9215698 2.6772829 3.0387768 2.7264461 3.5885873 3.7935218
##          15        16        17        18        19        20
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
```

```
## 13
## 14
## 15
## 16 3.4672887
## 17 2.1460637 3.7238123
## 18 4.3213341 2.6882989 4.7817222
## 19 2.5357316 2.2121809 3.4553238 3.0159719
## 20 4.4394940 1.8953465 5.1687765 2.6462369 2.3520165
## 21 3.1783108 3.4832271 3.3186086 4.3010304 2.8095878 3.8625251
```

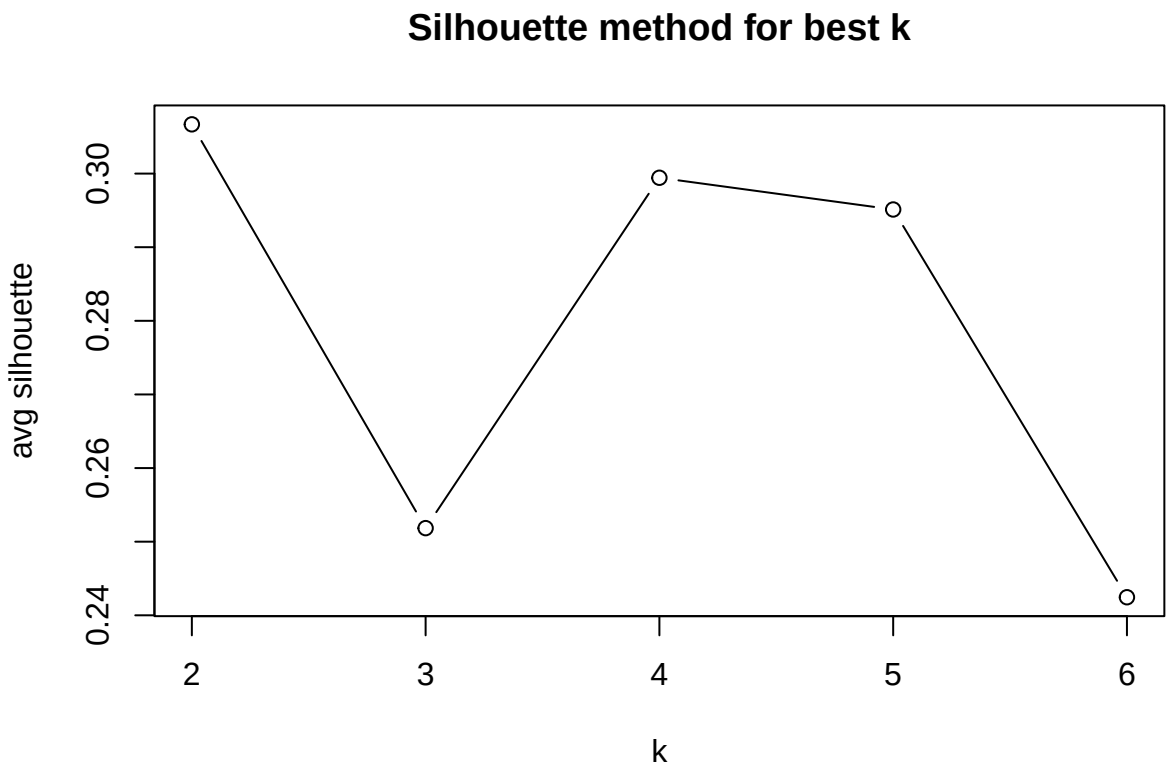
```r
heatmap(as.matrix(euclidean_calc), symm = TRUE,
        col = colorRampPalette(c("brown", "orange"))(100),
        main = "Euclidean Distance")
```




```r
## using angle method to find best k
set.seed(349)
optimal_cluster <- sapply(1:10, function(k){
  kmeans(normalized_data, centers = k, nstart = 25)$tot.withinss
})
plot(1:10, optimal_cluster, type = "b", main = "Angle Method", xlab = "Number of clusters", ylab = "wit
```




```
## using silhouette method to determine best k because angle in angle method was unclear
silhouette_method <- sapply(2:6, function(k) {
  pam(normalized_data, k = k)$silinfo$avg.width
})
plot(2:6, silhouette_method, type = "b", xlab = "k", ylab = "avg silhouette", main = "Silhouette method
```




```
## best k is 3
set.seed(349)
result <- kmeans(normalized_data, centers = 3, nstart = 25)
dataset$cluster <- as.factor(result$cluster)
datasummary <- dataset %>%
  group_by(cluster) %>%
  summarise(across(1:9, mean, .names = "mean_{.col}"))
```

```
## Warning: There were 6 warnings in ‘summarise()‘.
## The first warning was:
## i In argument: ‘across(1:9, mean, .names = "mean_{.col}")‘.
## i In group 1: ‘cluster = 1‘.
## Caused by warning in ‘mean.default()‘:
## ! argument is not numeric or logical: returning NA
## i Run ‘dplyr::last_dplyr_warnings()‘ to see the 5 remaining warnings.
```

```
print(datasummary)
```

```
## # A tibble: 3 x 10
##   cluster mean_Symbol mean_Name mean_Market_Cap mean_Beta mean_PE_Ratio mean_ROE
##   <fct>         <dbl>     <dbl>           <dbl>     <dbl>         <dbl>    <dbl>
## 1 1                NA        NA            4.38     0.888          21.2     15.1
## 2 2                NA        NA           43.7      0.337          32.4     19.0
## 3 3                NA        NA          114.       0.51           19.6     42.1
## # i 3 more variables: mean_ROA <dbl>, mean_Asset_Turnover <dbl>,
```

```
## #   mean_Leverage <dbl>
```

```
cat("cluster vs median recommendation\n")
```

```
## cluster vs median recommendation
```

```
print(table(dataset$cluster, dataset$Median_Recommendation))
```

```
##
##     Hold Moderate Buy Moderate Sell Strong Buy
##   1    2            2             1          0
##   2    4            3             1          1
##   3    3            2             2          0
```

```
cat("cluster vs location\n")
```

```
## cluster vs location
```

```
print(table(dataset$cluster, dataset$Location))
```

```
##
##     CANADA FRANCE GERMANY IRELAND SWITZERLAND UK US
##   1      0      0       1       1           0  0  3
##   2      1      1       0       0           1  1  5
##   3      0      0       0       0           0  2  5
```

```
cat("cluster vs stock exchange\n")
```

```
## cluster vs stock exchange
```

```
print(table(dataset$cluster, dataset$Exchange))
```

```
##
##     AMEX NASDAQ NYSE
##   1    1      1    3
##   2    0      0    9
##   3    0      0    7
```