STATS 411 - Final Report

Betty Hwang, Katie Jo, Devin Reeh, Mariano Aloiso

2024-03-14

Exploring Meta Stock Price Trends with Multivariate Statistical Analysis

```
library(ggplot2)
library(corrplot)
```

Load data

```
data <- read.csv("meta_2014_2023.csv")

# Show header
head(data)
```

```
##         date  open  high   low close   volume    rsi_7   rsi_14     cci_7
## 1 2014-01-02 54.83 55.22 54.19 54.71 43195500 51.91747 58.07782 -64.31212
## 2 2014-01-03 55.02 55.65 54.53 54.56 38246200 50.60499 57.38762 -40.05473
## 3 2014-01-06 54.42 57.26 54.05 57.20 68852600 67.48392 65.22152  43.90775
## 4 2014-01-07 57.70 58.55 57.22 57.92 77207400 70.67258 67.00319 150.62014
## 5 2014-01-08 57.60 58.41 57.23 58.23 56682400 72.04942 67.76880 107.79594
## 6 2014-01-09 58.65 58.96 56.65 57.22 92253300 61.13924 62.66706  67.34889
##      cci_14  sma_50   ema_50 sma_100  ema_100     macd bollinger TrueRange
## 1 -13.51710 50.2818 50.74095 47.6654 46.91456 1.828901   53.2450  1.030002
## 2 -17.36125 50.3194 50.89072 47.8288 47.06690 1.687987   53.5420  1.120003
## 3  42.36473 50.4254 51.13815 48.0306 47.26878 1.768947   53.9850  3.209999
## 4 117.88698 50.5348 51.40411 48.2433 47.48097 1.869653   54.4840  1.349998
## 5 101.41519 50.6604 51.67181 48.4600 47.69507 1.951977   54.9535  1.180000
## 6  72.50647 50.8002 51.88939 48.6614 47.88477 1.913662   55.3020  2.309997
##      atr_7   atr_14 next_day_close
## 1 1.652052 1.710739          54.56
## 2 1.576045 1.668543          57.20
## 3 1.809467 1.778647          57.92
## 4 1.743829 1.748030          58.23
## 5 1.663282 1.707456          57.22
## 6 1.755670 1.750495          57.94
```

```
# Show data summary
summary(data)
```

```
##      date                open            high             low        
##  Length:2516        Min.   : 54.02   Min.   : 54.94   Min.   : 51.85  
##  Class :character   1st Qu.:115.79   1st Qu.:117.45   1st Qu.:114.01  
##  Mode  :character   Median :170.12   Median :172.11   Median :168.22  
##                     Mean   :178.04   Mean   :180.33   Mean   :175.83  
##                     3rd Qu.:220.30   3rd Qu.:221.83   3rd Qu.:216.49  
##                     Max.   :381.68   Max.   :384.33   Max.   :378.81  
##      close            volume              rsi_7            rsi_14     
##  Min.   : 53.53   Min.   :  5467500   Min.   :14.08   Min.   :21.93  
##  1st Qu.:115.56   1st Qu.: 15631750   1st Qu.:43.51   1st Qu.:46.36  
##  Median :170.25   Median : 21062750   Median :55.36   Median :54.68  
##  Mean   :178.13   Mean   : 26170097   Mean   :54.34   Mean   :54.05  
##  3rd Qu.:219.87   3rd Qu.: 30220075   3rd Qu.:66.15   3rd Qu.:62.65  
##  Max.   :382.18   Max.   :232316600   Max.   :93.58   Max.   :86.07  
##      cci_7              cci_14              sma_50           ema_50      
##  Min.   :-233.33   Min.   :-422.48   Min.   : 50.28   Min.   : 50.74  
##  1st Qu.: -68.97   1st Qu.: -57.37   1st Qu.:115.55   1st Qu.:115.48  
##  Median :  27.72   Median :  35.25   Median :170.47   Median :170.87  
##  Mean   :  13.86   Mean   :  21.79   Mean   :175.35   Mean   :175.39  
##  3rd Qu.:  95.15   3rd Qu.: 102.73   3rd Qu.:210.96   3rd Qu.:216.29  
##  Max.   : 233.33   Max.   : 418.50   Max.   :363.75   Max.   :362.96  
##     sma_100          ema_100             macd            bollinger     
##  Min.   : 47.67   Min.   : 46.91   Min.   :-29.0463   Min.   : 53.24  
##  1st Qu.:112.84   1st Qu.:112.55   1st Qu.: -0.8819   1st Qu.:114.90  
##  Median :170.37   Median :171.02   Median :  1.0321   Median :169.46  
##  Mean   :172.71   Mean   :172.83   Mean   :  0.7878   Mean   :177.03  
##  3rd Qu.:203.54   3rd Qu.:214.06   3rd Qu.:  2.8551   3rd Qu.:216.49  
##  Max.   :351.03   Max.   :347.21   Max.   : 15.6235   Max.   :373.42  
##    TrueRange          atr_7            atr_14        next_day_close  
##  Min.   : 0.530   Min.   : 1.096   Min.   : 1.276   Min.   : 53.53  
##  1st Qu.: 2.087   1st Qu.: 2.292   1st Qu.: 2.379   1st Qu.:115.79  
##  Median : 3.905   Median : 4.305   Median : 4.364   Median :170.26  
##  Mean   : 5.010   Mean   : 4.998   Mean   : 4.981   Mean   :178.25  
##  3rd Qu.: 6.620   3rd Qu.: 7.373   3rd Qu.: 7.531   3rd Qu.:220.22  
##  Max.   :87.250   Max.   :22.275   Max.   :16.665   Max.   :382.18  
```

```r
# Count null values by column
colSums(is.na(data))
```

```
##           date           open           high            low          close 
##              0              0              0              0              0 
##         volume          rsi_7         rsi_14          cci_7         cci_14 
##              0              0              0              0              0 
##         sma_50         ema_50        sma_100        ema_100           macd 
##              0              0              0              0              0 
##      bollinger      TrueRange          atr_7         atr_14 next_day_close 
##              0              0              0              0              0 
```

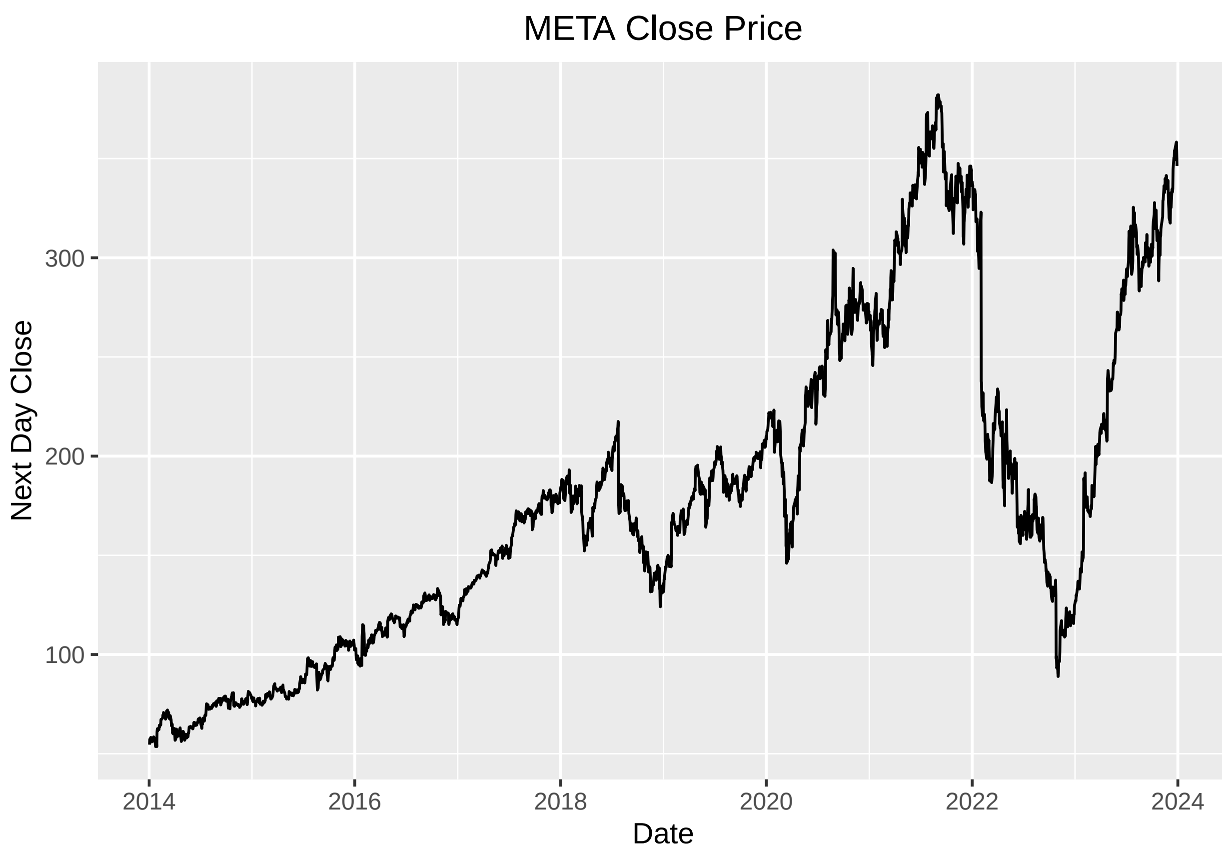
```r
# The data contains 2516 observations and 20 columns
ncol(data) #20 columns
```

```
## [1] 20
```

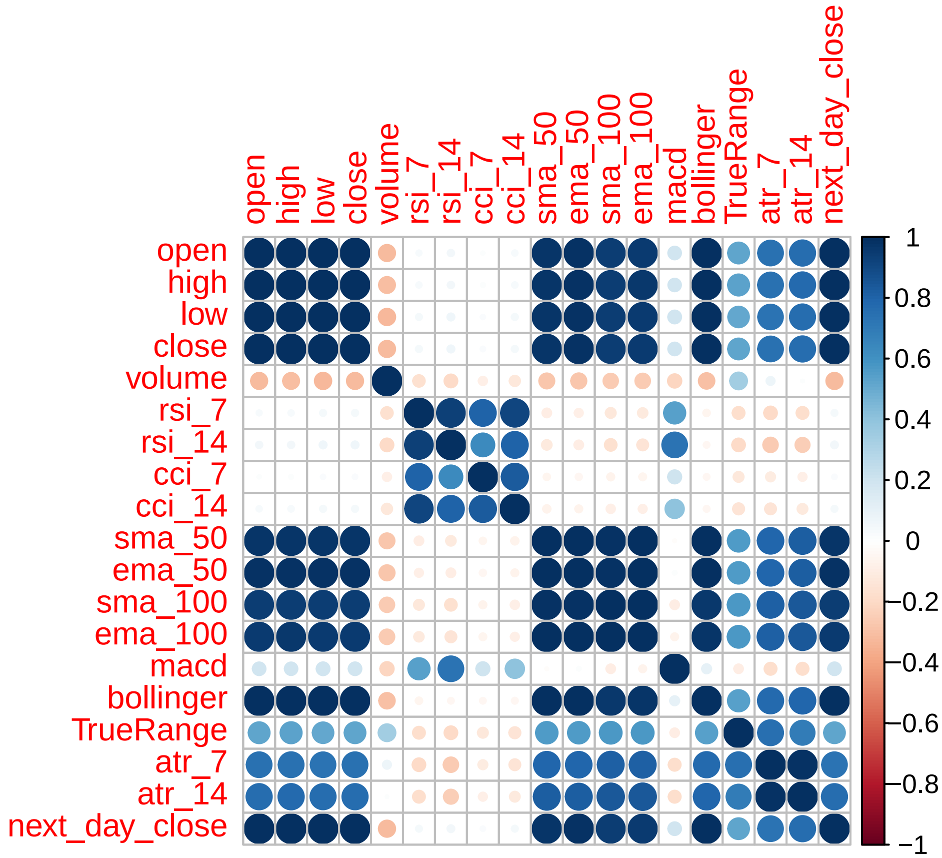
```
nrow(data) #2516 obs
```

```
## [1] 2516
```

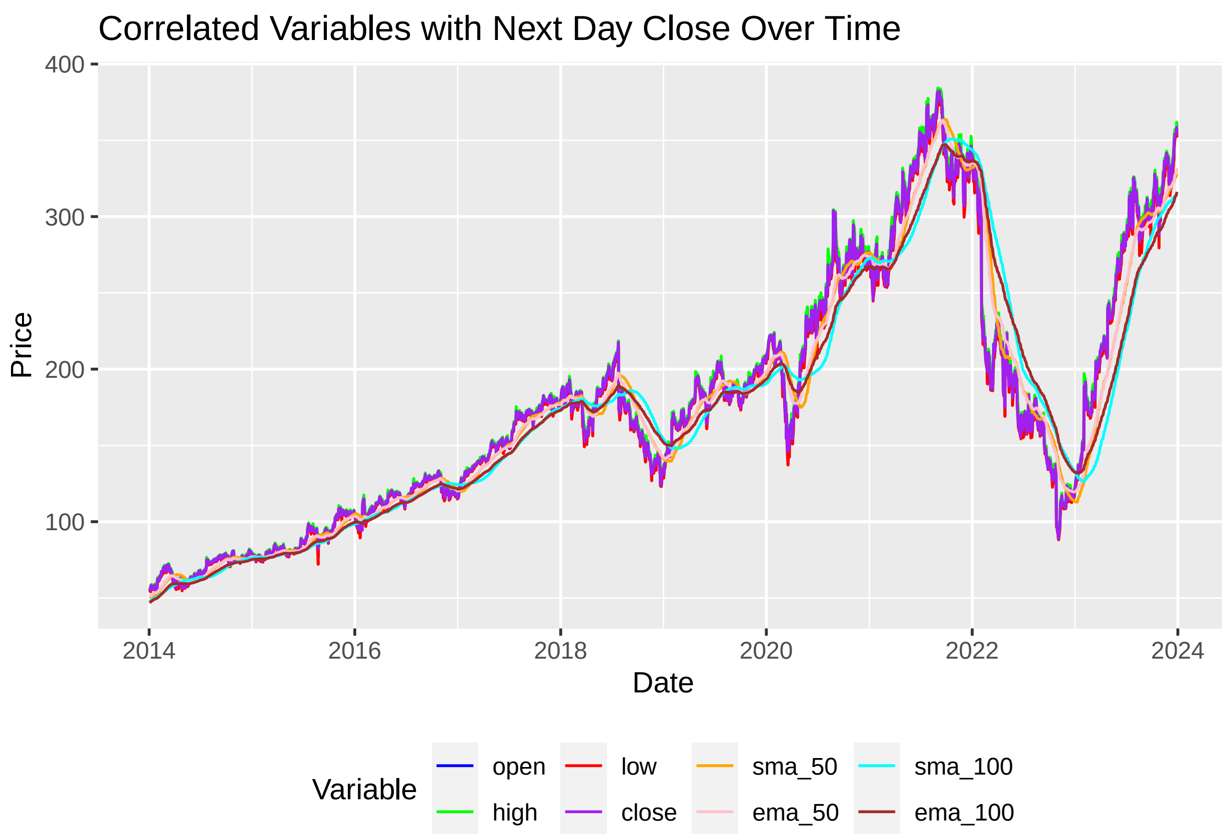
```
# Show the stock price trend
data <- data[order(data$date),]
data$date <- as.Date(data$date)

ggplot(data, aes(x = date, y = next_day_close)) +
  geom_line() +
  labs(x = 'Date', y = 'Next Day Close', title = 'META Close Price') +
  theme(plot.title = element_text(hjust = 0.5))
```




```
# Plot correlation matrix: open, high, low, close, sma_50, ema_50, sma_100, and ema_100 are highly corr
num_cols <- data[, sapply(data, is.numeric)]
cor_mat <- cor(num_cols)
corrplot(cor_mat)
```




```r
# Show trends for those highly correlated variables
correlated_vars <- c("open", "high", "low", "close", "sma_50", "ema_50", "sma_100", "ema_100")
data_melted <- reshape2::melt(data, id.vars = "date", measure.vars = correlated_vars)
ggplot(data_melted, aes(x = date, y = value, color = variable)) +
  geom_line() +
  scale_color_manual(values = c("open" = "blue", "high" = "green", "low" = "red", "close" = "purple",
                                "sma_50" = "orange", "ema_50" = "pink", "sma_100" = "cyan", "ema_100" =
  labs(title = "Correlated Variables with Next Day Close Over Time",
       x = "Date",
       y = "Price",
       color = "Variable") +
  theme(legend.position = "bottom")
```




```r
# Add a next_day_volume column. This will be used for canonical analysis
data$next_day_volume <- c(data$volume[-1], NA)

# Remove last row
data <- data[-nrow(data), ]
```

Baseline model

Run the analysis on all the variables. This gives a baseline understanding of the underlying structure in your data.

```r
# Select all columns except next day variables and date
data_baseline <- data[, -c(1, 20, 21)]
```

Baseline - Principal component analysis

```r
library(scales)
```

```r
# Conduct PCA
pca_baseline <- prcomp(data_baseline, scale = TRUE)
```

```r
# Show summary
summary(pca_baseline)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5    PC6     PC7
## Standard deviation     3.2642 2.0057 1.2193 0.96099 0.58200 0.4930 0.41442
## Proportion of Variance 0.5919 0.2235 0.0826 0.05131 0.01882 0.0135 0.00954
## Cumulative Proportion  0.5919 0.8154 0.8980 0.94934 0.96815 0.9817 0.99120
##                            PC8     PC9    PC10    PC11    PC12    PC13    PC14
## Standard deviation     0.31585 0.16786 0.12793 0.09386 0.04908 0.03310 0.03020
## Proportion of Variance 0.00554 0.00157 0.00091 0.00049 0.00013 0.00006 0.00005
## Cumulative Proportion  0.99674 0.99831 0.99922 0.99971 0.99984 0.99990 0.99995
##                           PC15    PC16     PC17     PC18
## Standard deviation     0.02475 0.01221 0.009917 0.004573
## Proportion of Variance 0.00003 0.00001 0.000010 0.000000
## Cumulative Proportion  0.99999 0.99999 1.000000 1.000000
```

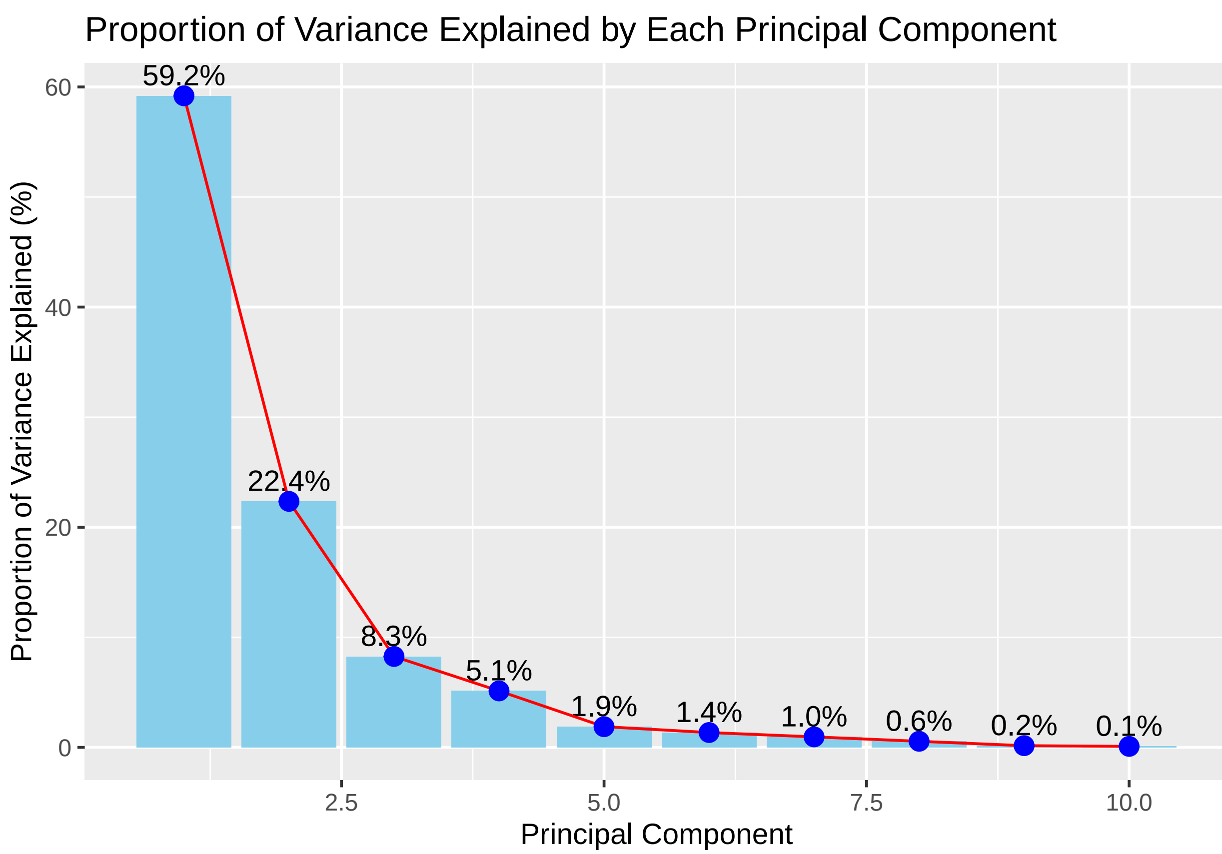
```r
# Show loadings
print(pca_baseline$rotation)
```

```
##                    PC1          PC2         PC3         PC4         PC5
## open       0.299142541  0.079376589  0.08429949  0.06509714 -0.12206576
## high       0.299688672  0.079381973  0.07423025  0.06359149 -0.11409956
## low        0.298384819  0.085453973  0.08949939  0.05816795 -0.12856537
## close      0.298875062  0.084838355  0.08030970  0.05830526 -0.12208017
## volume    -0.070895131 -0.161869535 -0.63974783  0.30716709 -0.58496326
## rsi_7     -0.028117054  0.476632187 -0.15838092 -0.05465801  0.05822534
## rsi_14    -0.032157461  0.471047040 -0.03273489  0.21840279  0.08561523
## cci_7     -0.018515015  0.389270515 -0.27766376 -0.43355861 -0.14493158
## cci_14    -0.021371790  0.448481672 -0.21699817 -0.21208801  0.01130093
## sma_50     0.303630202  0.002547041  0.05384655 -0.06474049 -0.14744318
## ema_50     0.304275028  0.010396789  0.05757111 -0.04151524 -0.13131642
## sma_100    0.300821665 -0.021542052  0.02296046 -0.12988812 -0.09421283
## ema_100    0.302856859 -0.013193644  0.02815054 -0.10248955 -0.08218711
## macd       0.007319387  0.336940174  0.17998004  0.69847126  0.12279974
## bollinger  0.302490686  0.032499274  0.08003039  0.02767302 -0.14031856
## TrueRange  0.192420256 -0.110292068 -0.46486396  0.28176451  0.20564328
## atr_7      0.260546700 -0.101028557 -0.29716157 -0.01683914  0.46625669
## atr_14     0.266899640 -0.085185558 -0.25014041 -0.07314103  0.46641802
##                     PC6         PC7         PC8          PC9         PC10
## open       0.0061266487 -0.04109820  0.04518709  0.110229247  0.095310882
## high       0.0001339708 -0.04440817  0.03407884  0.103859062  0.113876909
## low        0.0125928604 -0.04438955  0.03261424  0.103246766  0.114605331
## close      0.0057554918 -0.04202985  0.01324322  0.095392415  0.157980990
## volume     0.3464314159  0.03829759 -0.02140045 -0.026801358 -0.013767969
## rsi_7      0.0158672329  0.29575291 -0.33530337 -0.359008341  0.629341107
## rsi_14     0.0460927876  0.35623286 -0.42032683  0.373046474 -0.514475684
## cci_7     -0.1345758946 -0.69453667 -0.19477443  0.018650464 -0.149642919
## cci_14     0.0350637900  0.25876710  0.79040309  0.073419196 -0.078601619
## sma_50    -0.0174495640  0.06123617 -0.01699899  0.121936466  0.114817200
## ema_50    -0.0107956667  0.04752686 -0.02110794  0.029689228 -0.012158080
## sma_100   -0.0109201499  0.14601696 -0.04040457 -0.531146667 -0.336117887
## ema_100   -0.0027076753  0.11308718 -0.04343605 -0.377386529 -0.318922816
```

```
## macd       0.1161155430 -0.41425844  0.18930996 -0.292776169 -0.049158958
## bollinger -0.0080246652 -0.02783680 -0.01955962  0.286408513  0.049680446
## TrueRange -0.7781252206  0.04141443  0.03592546 -0.009881897 -0.002746544
## atr_7      0.3079744849 -0.09991906 -0.03458497  0.234870068  0.101767866
## atr_14     0.3789691121 -0.05674259 -0.01900676 -0.139640111 -0.084098042
##                   PC11         PC12          PC13         PC14          PC15
## open      -0.082295396 -0.341857323 -4.422873e-02 -0.027053174 -7.047871e-01
## high      -0.075683581 -0.310543178 -1.232902e-01 -0.071227671  4.790994e-02
## low       -0.085953127 -0.266226856 -1.053141e-01 -0.020514276  1.218224e-02
## close     -0.082413678 -0.194840488 -1.837303e-01 -0.103102919  6.895308e-01
## volume    -0.009445717 -0.001369986 -4.627709e-03  0.001048157  2.542047e-06
## rsi_7      0.035197567  0.091854951 -4.356586e-02 -0.021060000 -5.676864e-02
## rsi_14    -0.040461225 -0.084902860  3.840807e-02  0.012452835  3.001148e-02
## cci_7     -0.004370004 -0.005858532  2.435212e-02  0.014578956  1.257962e-03
## cci_14     0.020746374  0.031932055 -1.672523e-05 -0.002668306  1.555188e-02
## sma_50    -0.048619304  0.136747622  6.007940e-01  0.569372472  7.733769e-02
## ema_50     0.038013557  0.271489196  2.068165e-01  0.062928298 -1.409352e-02
## sma_100    0.237302648 -0.160502844  3.770328e-01 -0.468678731  3.852858e-02
## ema_100    0.120554089  0.150702110 -5.898318e-01  0.489003201 -3.814500e-02
## macd       0.081438976  0.137574149  1.022069e-01  0.076725230  5.926731e-03
## bollinger  0.038477397  0.696533289 -1.682184e-01 -0.428659028 -1.083862e-01
## TrueRange -0.064502941  0.004474726  1.663241e-03 -0.004799785  2.297161e-04
## atr_7      0.658076109 -0.092332527 -1.284363e-02  0.049023483  7.399439e-03
## atr_14    -0.670755160  0.087441953  3.626041e-02 -0.057739135 -7.423052e-03
##                    PC16          PC17          PC18
## open      -0.0835126538  0.4721318387  5.580538e-03
## high      -0.6089273333 -0.5990531020  5.741586e-03
## low        0.7852367242 -0.3769153605  1.294453e-02
## close     -0.0561762290  0.5242173274 -1.168697e-02
## volume    -0.0002997814 -0.0003701757  3.075844e-04
## rsi_7     -0.0014142009 -0.0147056255  1.794172e-03
## rsi_14     0.0006743092  0.0071655383 -9.863609e-04
## cci_7     -0.0019933777  0.0036130355 -1.697993e-03
## cci_14    -0.0007023757  0.0038674039  4.199738e-04
## sma_50    -0.0363201948  0.0080800781  3.647045e-01
## ema_50    -0.0082418562 -0.0172391643 -8.710501e-01
## sma_100    0.0173722203  0.0036267143  1.375383e-01
## ema_100   -0.0177389648  0.0066810788  7.178360e-02
## macd      -0.0080930476 -0.0002468950  4.977775e-03
## bollinger -0.0064013869 -0.0244732592  2.894664e-01
## TrueRange  0.0151960028  0.0019800478 -8.041543e-05
## atr_7      0.0114402199  0.0056052653 -1.072616e-03
## atr_14    -0.0013506612 -0.0027490132  5.430876e-04
```

```r
# Scree plot
pca_summ <- summary(pca_baseline)
pve <- pca_summ$importance[2,] * 100

pve_data <- data.frame(PC = 1:10, PVE = pve[1:10]) # Let's look at the first 10 PCs

ggplot(pve_data, aes(x = PC, y = PVE)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  geom_text(aes(label = sprintf("%.1f%%", PVE)), vjust = -0.5) +
  geom_line(aes(group = 1), color = "red") +
```

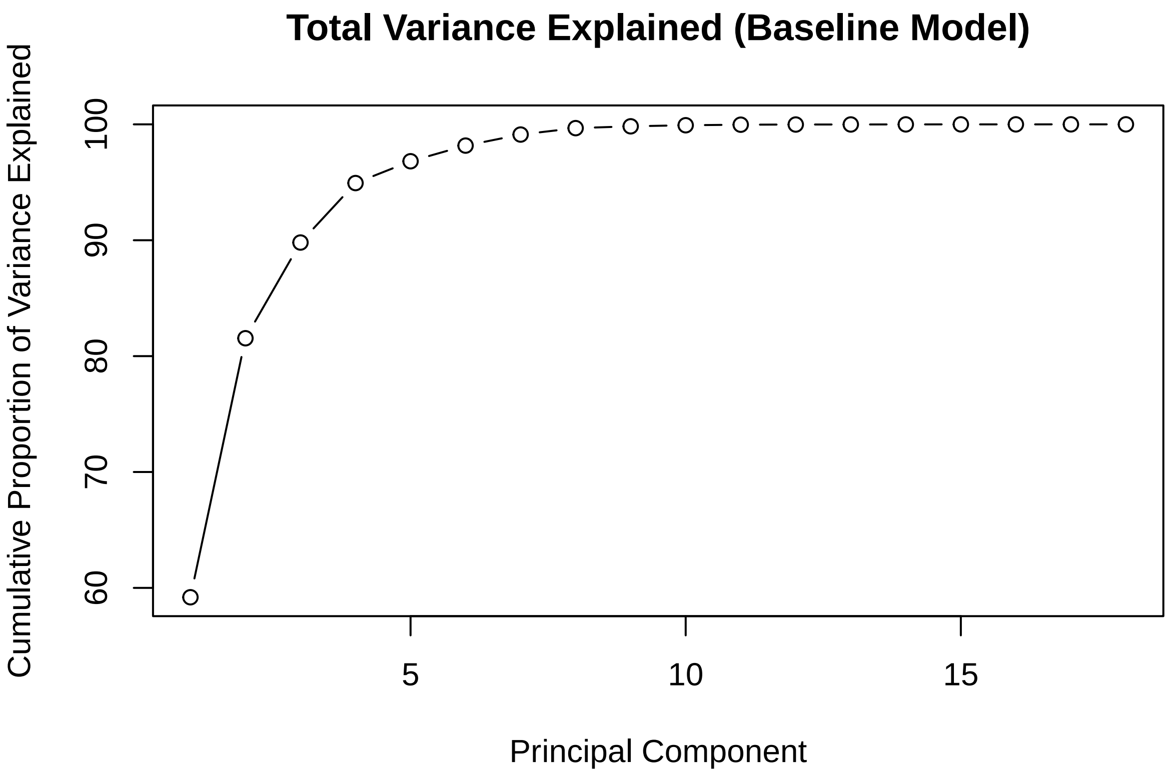
```r
  geom_point(color = "blue", size = 3) +
  labs(x = "Principal Component", y = "Proportion of Variance Explained (%)",
       title = "Proportion of Variance Explained by Each Principal Component")
```




```r
# Calculate cumulative proportion of variance explained
cumulative_variance <- cumsum(pca_baseline$sdev^2) / sum(pca_baseline$sdev^2) * 100
cumulative_variance
```

```
##  [1]  59.19354  81.54321  89.80313  94.93368  96.81550  98.16581  99.11994
##  [8]  99.67417  99.83070  99.92163  99.97057  99.98395  99.99004  99.99511
## [15]  99.99851  99.99934  99.99988 100.00000
```

```r
# Plot cumulative proportion of variance explained
plot(cumulative_variance,
     type = "b",
     xlab = "Principal Component",
     ylab = "Cumulative Proportion of Variance Explained",
     main = "Total Variance Explained (Baseline Model)")
```



Baseline - Factor analysis

```
library(psych)
```

Conduct factor analysis using 2 factors and varimax rotation.

```
# Specify the number of factors
min_num_factors <- 2
max_num_factors <- 2

# Create an empty list to store factor analysis results
fa_results <- list()

# Loop through each number of factors
for (i in min_num_factors:max_num_factors) {
    # Conduct factor analysis
    fa_result <- fa(data_baseline, nfactors = i, rotate = "varimax")

    # Store the factor analysis result in the list
    fa_results[[paste("fa_baseline_", i, sep = "")]] <- fa_result
}
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
```

```
## The estimated weights for the factor scores are probably incorrect.  Try a
## different factor score estimation method.
```

```
# Print factor loadings, scores, and summary for each factor analysis
for (i in min_num_factors:max_num_factors) {
    cat("Factor analysis for", i, "factor(s):\n")
    print(fa_results[[paste("fa_baseline_", i, sep = "")]]$loadings)
    print(fa_results[[paste("fa_baseline_", i, sep = "")]])
    cat("\n")
}
```

```
## Factor analysis for 2 factor(s):
##
## Loadings:
##           MR1    MR2
## open       0.981  0.140
## high       0.983  0.140
## low        0.979  0.152
## close      0.981  0.151
## volume    -0.220 -0.252
## rsi_7             0.983
## rsi_14            0.957
## cci_7             0.717
## cci_14            0.881
## sma_50     0.994
## ema_50     0.996
## sma_100    0.981
## ema_100    0.990
## macd              0.587
## bollinger  0.991
## TrueRange  0.591 -0.198
## atr_7      0.827 -0.206
## atr_14     0.850 -0.178
##
##                   MR1   MR2
## SS loadings    10.578 3.789
## Proportion Var  0.588 0.211
## Cumulative Var  0.588 0.798
## Factor Analysis using method =  minres
## Call: fa(r = data_baseline, nfactors = i, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##             MR1   MR2   h2     u2 com
## open       0.98  0.14 0.98 0.0176 1.0
## high       0.98  0.14 0.99 0.0132 1.0
## low        0.98  0.15 0.98 0.0185 1.0
## close      0.98  0.15 0.98 0.0151 1.0
## volume    -0.22 -0.25 0.11 0.8885 2.0
## rsi_7     -0.07  0.98 0.97 0.0274 1.0
## rsi_14    -0.09  0.96 0.92 0.0760 1.0
## cci_7     -0.04  0.72 0.52 0.4836 1.0
## cci_14    -0.05  0.88 0.78 0.2212 1.0
## sma_50     0.99 -0.02 0.99 0.0126 1.0
## ema_50     1.00  0.00 0.99 0.0072 1.0
## sma_100    0.98 -0.07 0.97 0.0324 1.0
```

```
## ema_100    0.99 -0.05 0.98 0.0180 1.0
## macd       0.03  0.59 0.35 0.6547 1.0
## bollinger  0.99  0.04 0.98 0.0170 1.0
## TrueRange  0.59 -0.20 0.39 0.6119 1.2
## atr_7      0.83 -0.21 0.73 0.2733 1.1
## atr_14     0.85 -0.18 0.75 0.2451 1.1
##
##                         MR1  MR2
## SS loadings           10.58 3.79
## Proportion Var         0.59 0.21
## Cumulative Var         0.59 0.80
## Proportion Explained   0.74 0.26
## Cumulative Proportion  0.74 1.00
##
## Mean item complexity =  1.1
## Test of the hypothesis that 2 factors are sufficient.
##
## df null model =  153  with the objective function =  70.96 with Chi Square =  177904.9
## df of  the model are 118  and the objective function was  29.66
##
## The root mean square of the residuals (RMSR) is  0.07
## The df corrected root mean square of the residuals is  0.08
##
## The harmonic n.obs is  2515 with the empirical chi square  3693.62  with prob <  0
## The total n.obs was  2515  with Likelihood Chi Square =  74319.18  with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.458
## RMSEA index =  0.5  and the 90 % confidence intervals are  0.497 0.503
## BIC =  73395.23
## Fit based upon off diagonal values = 0.99
```

Baseline - Canonical correlation analysis

```r
library(CCA)
library(candisc)
```

Conduct canonical correlation analysis using the package `CCA`. Split the data into two sets: one set for next day variables and another for technical indicators (e.g., RSI, CCI, moving averages).

```r
# List respective columns
col_price <- c("next_day_close", "next_day_volume")
col_indicators <- c("open", "high", "low", "close", "volume",
                    "rsi_7", "rsi_14", "cci_7", "cci_14", "sma_50",
                    "ema_50", "sma_100", "ema_100", "macd",
                    "bollinger", "TrueRange", "atr_7", "atr_14")

# Select the variables for each group
group_1 <- scale(data[, col_price])
group_2 <- scale(data_baseline[, col_indicators])

# Run canonical correlation analysis
cca_baseline <- cancor(group_1, group_2,
```

```r
                     set.names = c("Next Day Variables", "Technical Indicators"))

# Canonical correlation analysis results
coef(cca_baseline, type = "both", standardize = TRUE)
```

```
## [[1]]
##                         Xcan1      Xcan2
## next_day_close  -1.002511575 -0.3259485
## next_day_volume -0.008028792 -1.0541382
##
## [[2]]
##                    Ycan1        Ycan2
## open       0.1681268035 -0.415107622
## high      -0.1137606189 -5.189434624
## low       -0.1988873884  4.552820375
## close     -0.7746714252  0.985554366
## volume    -0.0035367724 -1.136583032
## rsi_7      0.0213488982 -0.327037887
## rsi_14    -0.0112990123  0.214461307
## cci_7     -0.0081769620  0.055221181
## cci_14    -0.0032896089  0.092914862
## sma_50    -0.2156634256  0.135495963
## ema_50     0.3600572754  0.450706382
## sma_100   -0.0895934797  0.059782243
## ema_100    0.0288245180 -0.013220686
## macd      -0.0161373306  0.003677684
## bollinger -0.1598557146 -0.969719013
## TrueRange -0.0009032911  0.473113419
## atr_7      0.0259904560  0.204428361
## atr_14    -0.0320405376 -0.457593630
```

```r
# Print structure
cca_baseline$structure
```

```
## $X.xscores
##                       Xcan1       Xcan2
## next_day_close  -0.9999710  0.00761623
## next_day_volume  0.3091995 -0.95099719
##
## $Y.xscores
##                 Xcan1        Xcan2
## open      -0.99767679  0.001234955
## high      -0.99810882 -0.003314195
## low       -0.99810941  0.007331797
## close     -0.99838932  0.002491691
## volume     0.31302890 -0.655135292
## rsi_7     -0.04301149  0.136447684
## rsi_14    -0.05632449  0.171213830
## cci_7     -0.02604512  0.076066120
## cci_14    -0.04371651  0.112518293
## sma_50    -0.97556123 -0.032375330
## ema_50    -0.98075702 -0.028965600
## sma_100   -0.94529521 -0.043174207
```

```
## ema_100   -0.95695427 -0.040734135
## macd      -0.19366584  0.151378836
## bollinger -0.99080396 -0.018464342
## TrueRange -0.52536831 -0.288393850
## atr_7     -0.74139003 -0.253439584
## atr_14    -0.76885180 -0.220696570
##
## $X.yscores
##                       Ycan1        Ycan2
## next_day_close  -0.9984068  0.005450189
## next_day_volume  0.3087158 -0.680535452
##
## $Y.yscores
##                 Ycan1        Ycan2
## open      -0.99923988  0.001725758
## high      -0.99967260 -0.004631339
## low       -0.99967318  0.010245636
## close     -0.99995353  0.003481951
## volume     0.31351934 -0.915502377
## rsi_7     -0.04307888  0.190675393
## rsi_14    -0.05641274  0.239258471
## cci_7     -0.02608593  0.106296690
## cci_14    -0.04378500  0.157235865
## sma_50    -0.97708968 -0.045242092
## ema_50    -0.98229361 -0.040477250
## sma_100   -0.94677624 -0.060332712
## ema_100   -0.95845357 -0.056922895
## macd      -0.19396926  0.211540556
## bollinger -0.99235629 -0.025802531
## TrueRange -0.52619142 -0.403008750
## atr_7     -0.74255159 -0.354162788
## atr_14    -0.77005639 -0.308406884
```

Subset Data

Subset the data to include only the close price and the long term technical indicators

```r
# Select columns
data_subset <- data[, c("rsi_14", "cci_14", "sma_50", "volume",
                        "ema_50", "sma_100", "ema_100", "bollinger",
                        "macd", "TrueRange", "atr_14")]
```

Subset - Principal component analysis

```r
# Conduct PCA
pca_subset <- prcomp(data_subset, scale = TRUE)

# Show summary
summary(pca_subset)
```

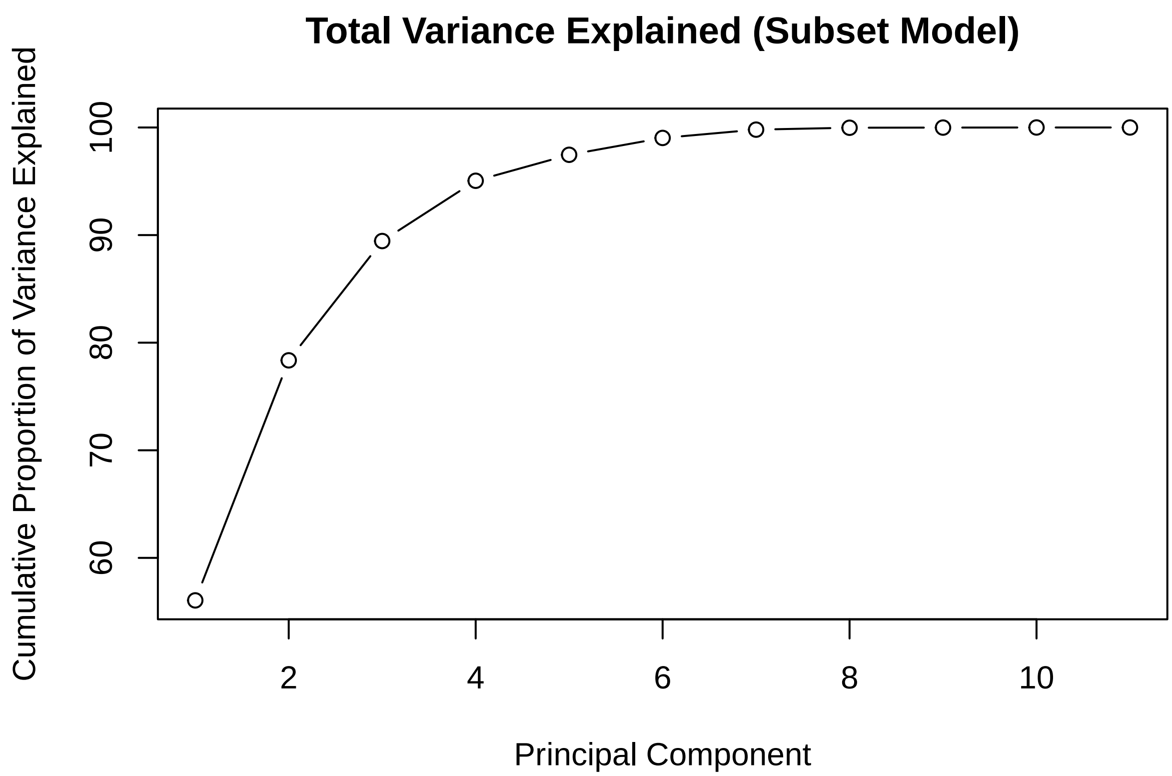
```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.4831 1.5666 1.1043 0.78494 0.51545 0.41487 0.29160
## Proportion of Variance 0.5605 0.2231 0.1109 0.05601 0.02415 0.01565 0.00773
## Cumulative Proportion  0.5605 0.7836 0.8945 0.95048 0.97464 0.99028 0.99801
##                            PC8     PC9    PC10     PC11
## Standard deviation     0.13905 0.03903 0.03141 0.004654
## Proportion of Variance 0.00176 0.00014 0.00009 0.000000
## Cumulative Proportion  0.99977 0.99991 1.00000 1.000000
```

```r
# Show loadings
print(pca_subset$rotation)
```

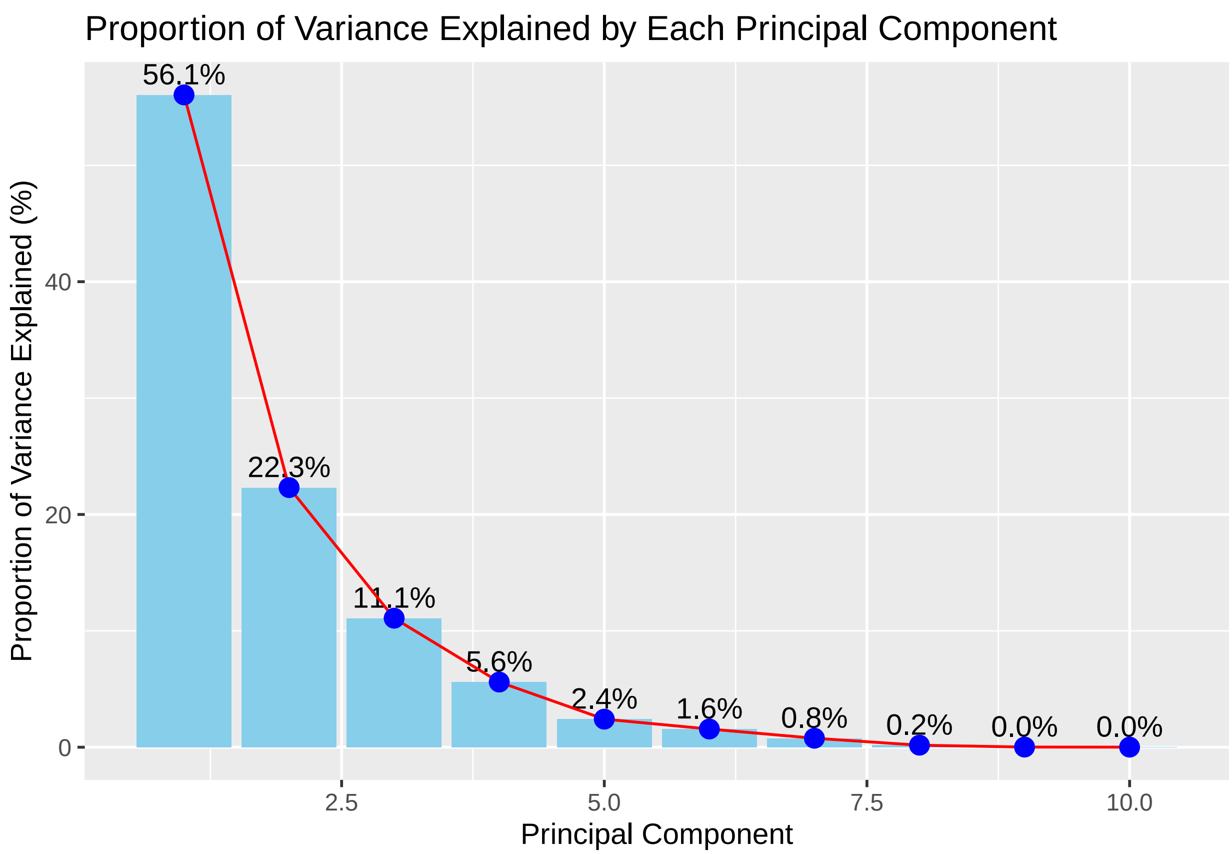
```
##                   PC1         PC2         PC3          PC4         PC5
## rsi_14    -0.08068720  0.58850326  0.21668012 -0.051896264 -0.00447393
## cci_14    -0.05545127  0.50332816  0.25599969 -0.645721184  0.03618535
## sma_50     0.39665377  0.06829797 -0.06260431 -0.005486598 -0.16545759
## volume    -0.07949934 -0.29891760  0.72518613  0.017595906 -0.61392809
## ema_50     0.39651648  0.08120439 -0.05717658  0.019997790 -0.16036041
## sma_100    0.39768853  0.02952925 -0.06240998 -0.085465302 -0.11898997
## ema_100    0.39885420  0.04368611 -0.05671717 -0.054037436 -0.11654202
## bollinger  0.39034485  0.11703249 -0.05030042  0.102663798 -0.17850941
## macd      -0.03353360  0.50712044  0.13829777  0.718895650 -0.06839148
## TrueRange  0.26486076 -0.14113097  0.54085046  0.158348213  0.68986287
## atr_14     0.35867728 -0.07840933  0.18205227 -0.130238124  0.16875063
##                   PC6         PC7          PC8          PC9          PC10
## rsi_14     0.04838092  0.77125976  0.014107858 -0.009324623  0.0046541359
## cci_14     0.03723906 -0.50751764  0.020976884  0.019497899  0.0019349704
## sma_50     0.14874401 -0.01573963  0.305074619 -0.682892970 -0.2889985683
## volume     0.02284109  0.01269017 -0.022622857  0.002419322 -0.0039268500
## ema_50     0.12741129 -0.01647817  0.163405183 -0.103986763  0.0235362230
## sma_100    0.08995585  0.03054606 -0.637567963 -0.171744134  0.5929399923
## ema_100    0.06934300  0.03257429 -0.425861310  0.417959394 -0.6725065997
## bollinger  0.13438099 -0.04143782  0.511201796  0.560916163  0.3332428323
## macd      -0.20318399 -0.36644784 -0.146864843 -0.057180202 -0.0249215053
## TrueRange  0.34038810 -0.02334023  0.003508886 -0.002313713  0.0003072101
## atr_14    -0.87708942  0.09145440  0.091201610 -0.016822617  0.0192386647
##                    PC11
## rsi_14    -8.572196e-05
## cci_14     9.871750e-05
## sma_50     3.759914e-01
## volume     2.330994e-04
## ema_50    -8.673075e-01
## sma_100    1.379203e-01
## ema_100    7.145173e-02
## bollinger  2.867374e-01
## macd       7.876558e-03
## TrueRange -2.856776e-04
## atr_14    -4.305275e-04
```

```r
# Calculate cumulative proportion of variance explained
cumulative_variance <- cumsum(pca_subset$sdev^2) / sum(pca_subset$sdev^2) * 100
```

```r
# Plot cumulative proportion of variance explained
plot(cumulative_variance,
     type = "b",
     xlab = "Principal Component",
     ylab = "Cumulative Proportion of Variance Explained",
     main = "Total Variance Explained (Subset Model)")
```




```r
# Scree plot
pca_summ <- summary(pca_subset)
pve <- pca_summ$importance[2,] * 100

pve_data <- data.frame(PC = 1:10, PVE = pve[1:10]) # Let's look at the first 10 PCs

ggplot(pve_data, aes(x = PC, y = PVE)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  geom_text(aes(label = sprintf("%.1f%%", PVE)), vjust = -0.5) +
  geom_line(aes(group = 1), color = "red") +
  geom_point(color = "blue", size = 3) +
  labs(x = "Principal Component", y = "Proportion of Variance Explained (%)",
       title = "Proportion of Variance Explained by Each Principal Component")
```



Subset - Factor analysis

Find appropriate number of factors for factor analysis by looking at the eigenvalue and scree plot.

```
cor(data_subset)
```

```
##                rsi_14      cci_14      sma_50      volume      ema_50
## rsi_14     1.00000000  0.80980621 -0.11456017 -0.19936541 -0.09550981
## cci_14     0.80980621  1.00000000 -0.06846389 -0.12897655 -0.06101446
## sma_50    -0.11456017 -0.06846389  1.00000000 -0.27252777  0.99902066
## volume    -0.19936541 -0.12897655 -0.27252777  1.00000000 -0.27770537
## ema_50    -0.09550981 -0.06101446  0.99902066 -0.27770537  1.00000000
## sma_100   -0.16622839 -0.08711913  0.98633288 -0.25263984  0.98645723
## ema_100   -0.14586069 -0.08085205  0.99137320 -0.25878502  0.99210676
## bollinger -0.04299090 -0.04428141  0.99141875 -0.29117626  0.99453112
## macd       0.73697427  0.40883388 -0.01250303 -0.21545767  0.01681154
## TrueRange -0.19726212 -0.14921496  0.56070513  0.34250540  0.56175124
## atr_14    -0.24089290 -0.11874302  0.82113510  0.01037979  0.82069251
##               sma_100     ema_100   bollinger        macd   TrueRange
## rsi_14    -0.16622839 -0.14586069 -0.04299090  0.73697427 -0.19726212
## cci_14    -0.08711913 -0.08085205 -0.04428141  0.40883388 -0.14921496
## sma_50     0.98633288  0.99137320  0.99141875 -0.01250303  0.56070513
## volume    -0.25263984 -0.25878502 -0.29117626 -0.21545767  0.34250540
## ema_50     0.98645723  0.99210676  0.99453112  0.01681154  0.56175124
```

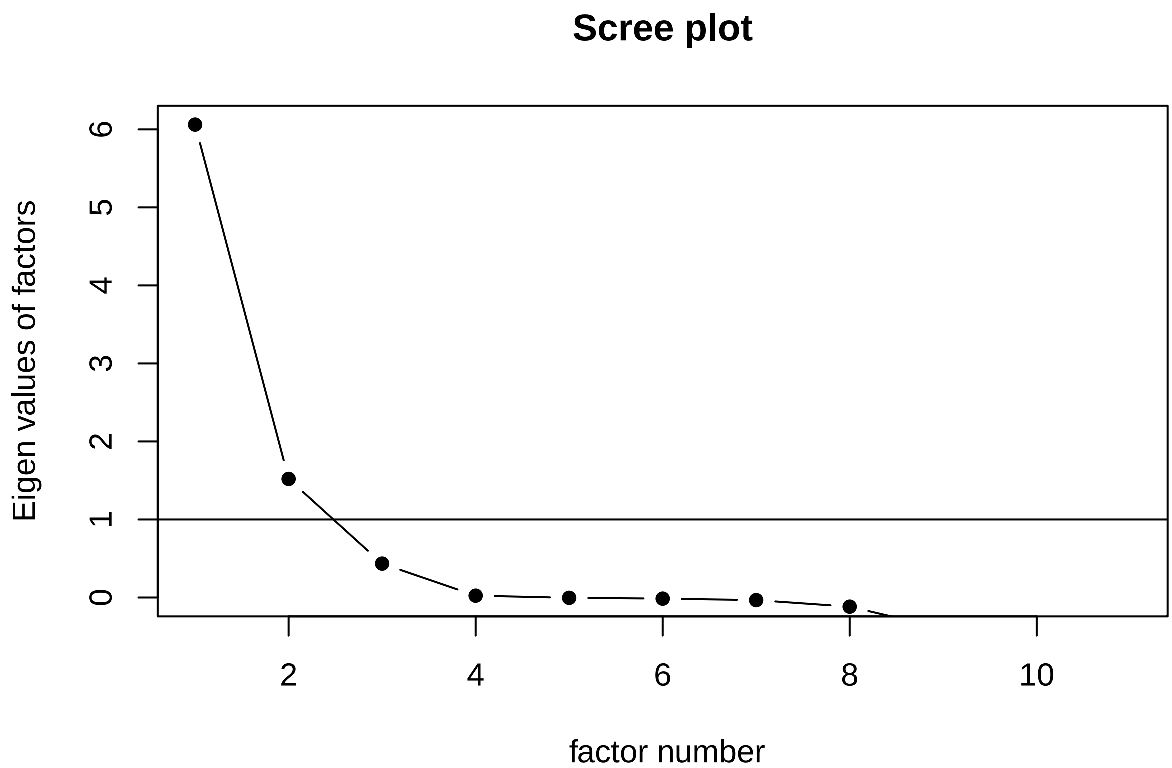
```
## sma_100    1.00000000  0.99790022  0.96538468 -0.09397867  0.57306313
## ema_100    0.99790022  1.00000000  0.97547488 -0.06172866  0.57613319
## bollinger  0.96538468  0.97547488  1.00000000  0.10026294  0.54901664
## macd      -0.09397867 -0.06172866  0.10026294  1.00000000 -0.09277482
## TrueRange  0.57306313  0.57613319  0.54901664 -0.09277482  1.00000000
## atr_14     0.84700396  0.84917110  0.79358526 -0.17422594  0.69961845
##                atr_14
## rsi_14    -0.24089290
## cci_14    -0.11874302
## sma_50     0.82113510
## volume     0.01037979
## ema_50     0.82069251
## sma_100    0.84700396
## ema_100    0.84917110
## bollinger  0.79358526
## macd      -0.17422594
## TrueRange  0.69961845
## atr_14     1.00000000
```

```r
ev <- eigen(cor(data_subset))
ev$values
```

```
##  [1] 6.165579e+00 2.454134e+00 1.219462e+00 6.161276e-01 2.656842e-01
##  [6] 1.721180e-01 8.502858e-02 1.933504e-02 1.523329e-03 9.865121e-04
## [11] 2.165938e-05
```

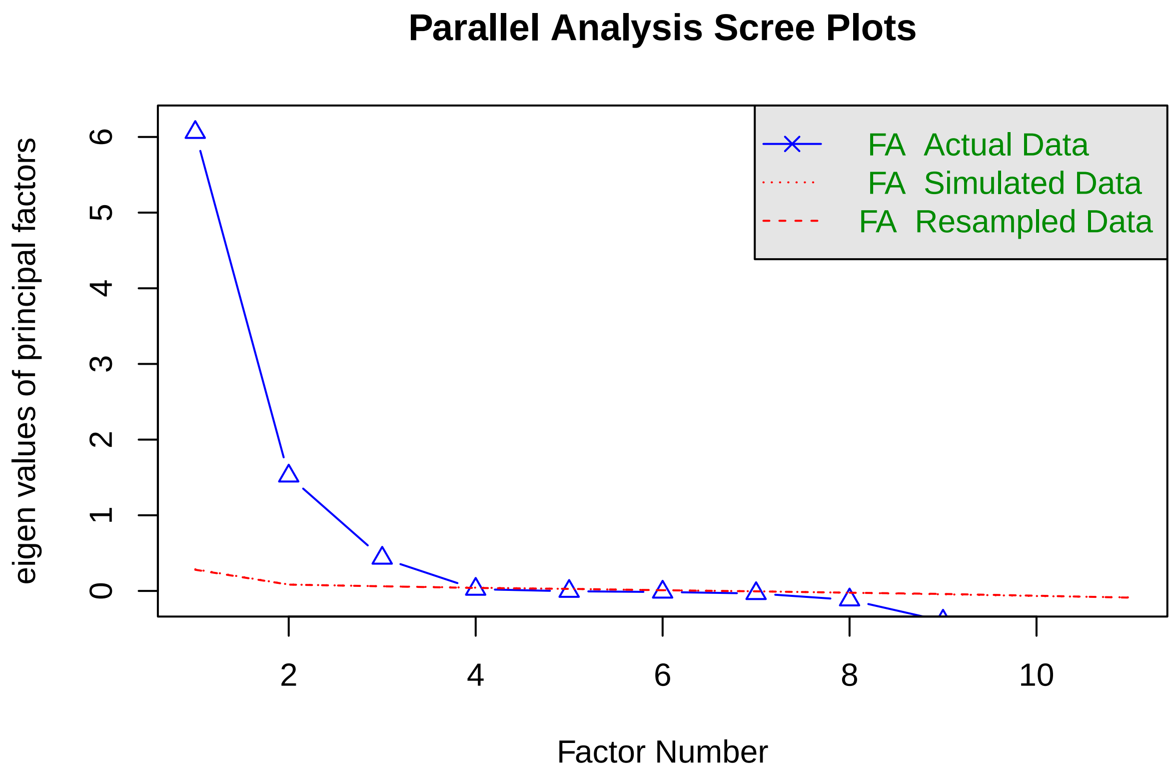
```r
scree(data_subset, pc = FALSE)
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect.  Try a
## different factor score estimation method.
```




```r
fa.parallel(data_subset, fa ="fa")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect.  Try a
## different factor score estimation method.
```




```
## Parallel analysis suggests that the number of factors =  3  and the number of components =  NA
```

Use 3 factors according to parallel analysis.

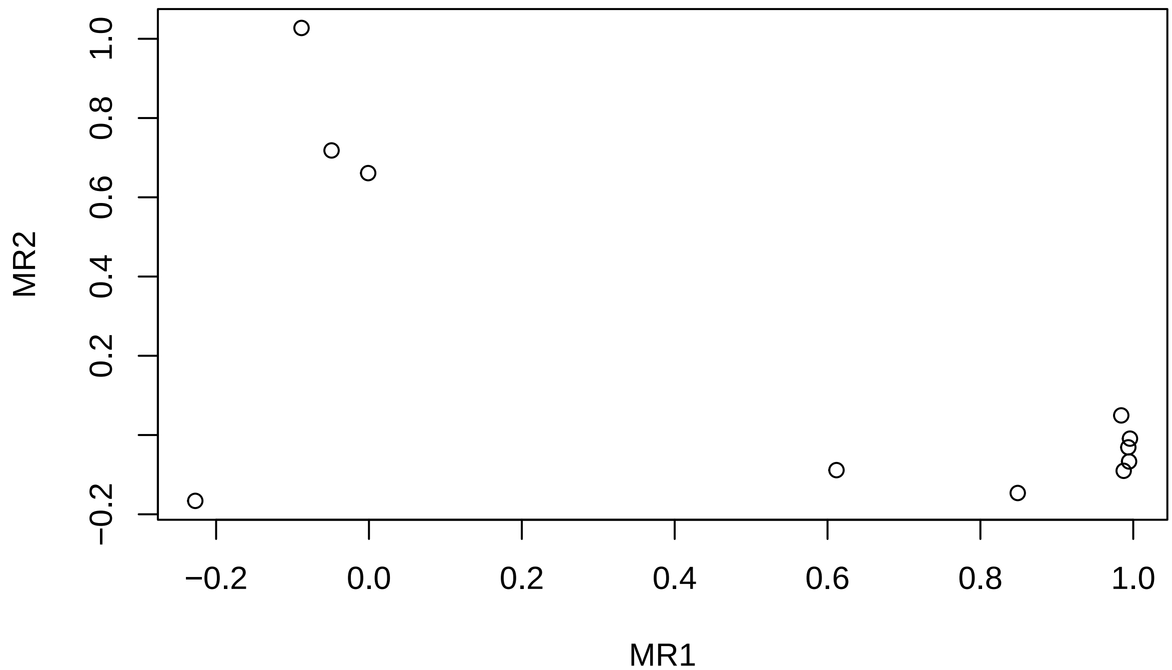
```r
# Specify the number of factors
num_factors <- 3

# Conduct factor analysis
fa_result <- fa(data_subset,
                nfactors = num_factors,
                rotate = "varimax")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect.  Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected.  Examine the results carefully
```

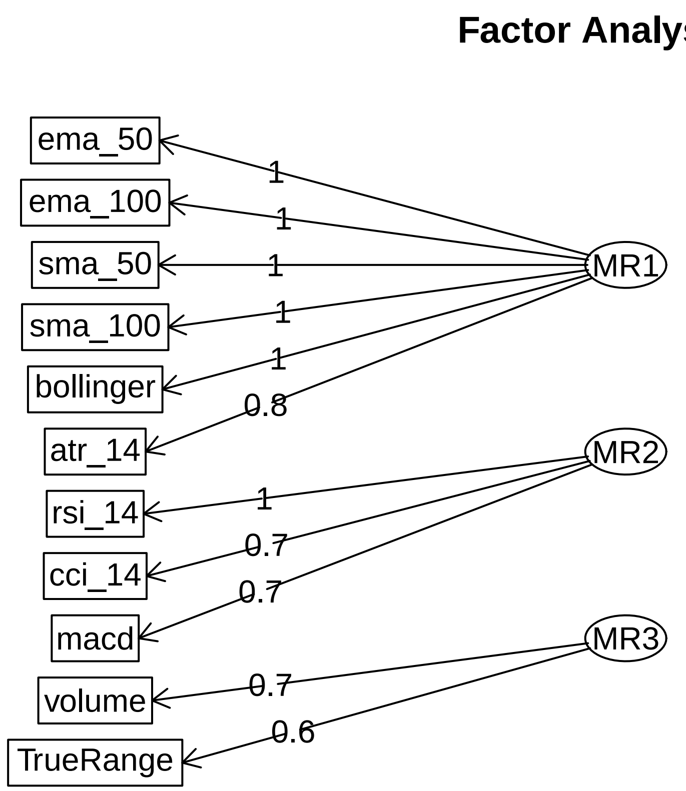
```r
# Print factor loadings, scores, and summary
load <- fa_result$loadings
names(data_subset)
```

```
##  [1] "rsi_14"     "cci_14"     "sma_50"     "volume"     "ema_50"     "sma_100"
##  [7] "ema_100"    "bollinger" "macd"       "TrueRange" "atr_14"
```

```r
plot(load)
```




```r
fa.diagram(load)
```




```r
round(fa_result$loadings,3)
```

```
##
## Loadings:
##           MR1    MR2    MR3
## rsi_14            1.027
## cci_14            0.718
## sma_50     0.994
## volume    -0.227 -0.166  0.725
## ema_50     0.996
## sma_100    0.987
## ema_100    0.995
## bollinger  0.984
## macd              0.661 -0.120
## TrueRange  0.612         0.634
## atr_14     0.849 -0.146  0.260
##
##                  MR1   MR2   MR3
## SS loadings    6.070 2.080 1.042
## Proportion Var 0.552 0.189 0.095
## Cumulative Var 0.552 0.741 0.836
```

Use 2 factors for comparison, as simplified model.

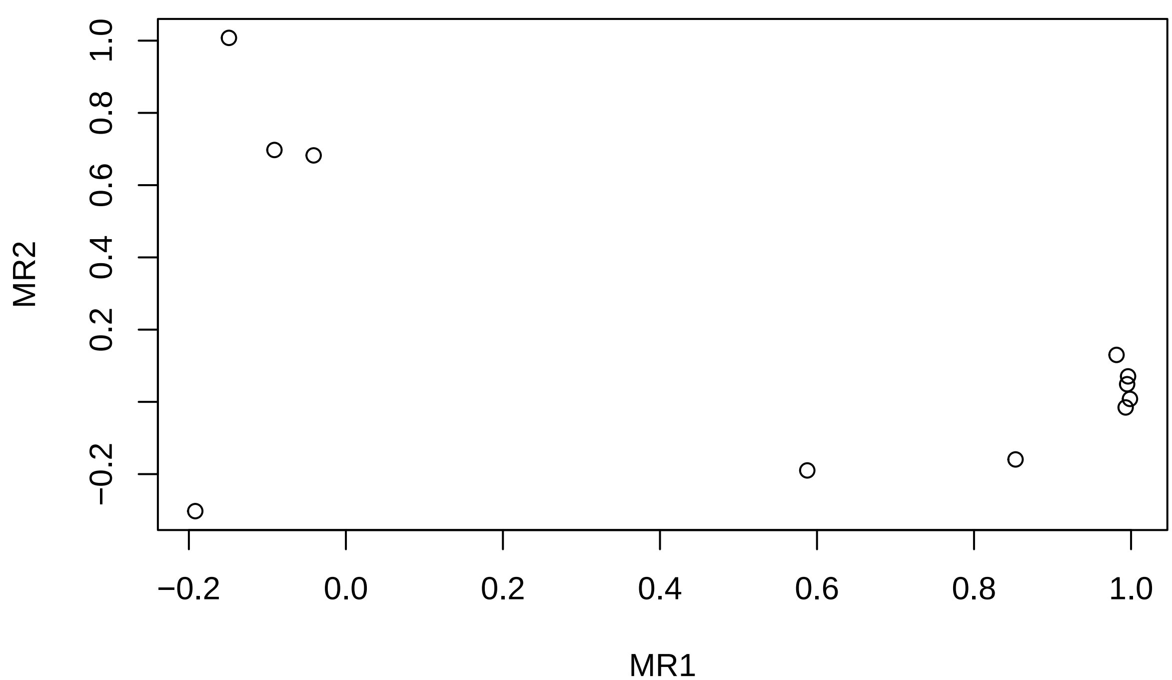
```r
# Specify the number of factors
num_factors <- 2

# Conduct factor analysis
fa_result <- fa(data_subset,
                nfactors = num_factors,
                rotate = "varimax")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect.  Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected.  Examine the results carefully
```

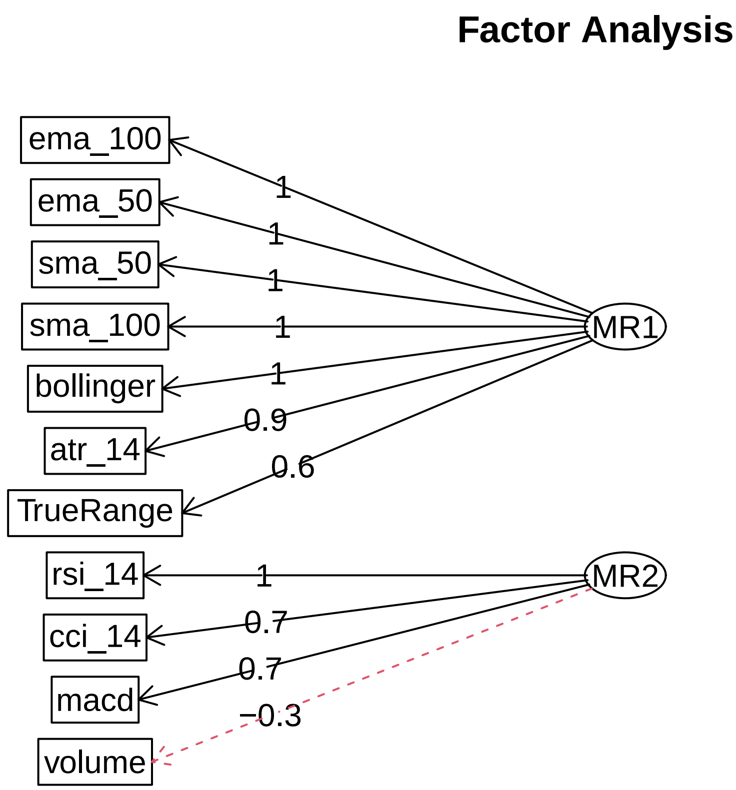
```r
# Print factor loadings, scores, and summary
load <- fa_result$loadings
names(data_subset)
```

```
##  [1] "rsi_14"     "cci_14"     "sma_50"     "volume"     "ema_50"     "sma_100"
##  [7] "ema_100"    "bollinger" "macd"       "TrueRange" "atr_14"
```

```r
plot(load)
```



```r
fa.diagram(load)
```




```r
round(fa_result$loadings, 3)
```

```
##
## Loadings:
##           MR1    MR2
## rsi_14    -0.149  1.008
## cci_14            0.697
## sma_50     0.995
## volume    -0.192 -0.302
## ema_50     0.996
## sma_100    0.993
## ema_100    0.999
## bollinger  0.982  0.130
## macd              0.682
## TrueRange  0.588 -0.190
## atr_14     0.853 -0.159
##
##                  MR1   MR2
## SS loadings    6.073 2.144
## Proportion Var 0.552 0.195
## Cumulative Var 0.552 0.747
```

Subset - Canonical correlation analysis

```
# List respective columns
col_price <- c("next_day_close", "next_day_volume")
col_indicators <- c("rsi_14", "cci_14", "sma_50", "volume",
                    "ema_50", "sma_100", "ema_100", "bollinger",
                    "macd", "TrueRange", "atr_14")

# Select the variables for each group
group_1 <- scale(data[, col_price])
group_2 <- scale(data_subset[, col_indicators])

# Run canonical correlation analysis
cca_subset <- cancor(group_1, group_2,
                    set.names = c("Next Day Variables", "Technical Indicators"))

# Canonical correlation analysis results
coef(cca_subset, type = "both", standardize = TRUE)
```

```
## [[1]]
##                        Xcan1      Xcan2
## next_day_close  -1.002388702 -0.3263262
## next_day_volume -0.007631641 -1.0541412
##
## [[2]]
##                   Ycan1        Ycan2
## rsi_14    -0.0031605524 -0.009916673
## cci_14    -0.0377966311  0.035179060
## sma_50    -0.6615046149  0.102545124
## volume     0.0005383673 -1.135749729
## ema_50     0.0047169750 -0.273677375
## sma_100   -0.0472684824 -0.042110755
## ema_100    0.0481893523  0.201802563
## bollinger -0.3228315722 -0.276784911
## macd      -0.1553543190 -0.021536216
## TrueRange  0.0063797972  0.380752223
## atr_14    -0.0127302485 -0.342499900
```

```
# Print structure
cca_subset$structure
```

```
## $X.xscores
##                      Xcan1        Xcan2
## next_day_close  -0.9999738  0.007239486
## next_day_volume  0.3095578 -0.950880633
##
## $Y.xscores
##                 Xcan1       Xcan2
## rsi_14    -0.05638899  0.17119260
## cci_14    -0.04375890  0.11250181
## sma_50    -0.97554896 -0.03274287
## volume     0.31327571 -0.65501731
```

```
## ema_50    -0.98074604 -0.02933510
## sma_100   -0.94527888 -0.04353035
## ema_100   -0.95693885 -0.04109467
## bollinger -0.99079693 -0.01883763
## macd      -0.19372286  0.15130586
## TrueRange -0.52525962 -0.28859176
## atr_14    -0.76876860 -0.22098622
##
## $X.yscores
##                      Ycan1        Ycan2
## next_day_close  -0.9976381  0.005155729
## next_day_volume  0.3088347 -0.677186585
##
## $Y.yscores
##                Ycan1       Ycan2
## rsi_14    -0.05652102  0.24038238
## cci_14    -0.04386135  0.15797093
## sma_50    -0.97783297 -0.04597635
## volume     0.31400916 -0.91975135
## ema_50    -0.98304221 -0.04119128
## sma_100   -0.94749201 -0.06112372
## ema_100   -0.95917929 -0.05770363
## bollinger -0.99311664 -0.02645111
## macd      -0.19417641  0.21245816
## TrueRange -0.52648938 -0.40523000
## atr_14    -0.77056848 -0.31030077
```