project.R

2021-11-23

```
options(scipen = 999)

library(ggplot2)
library(quantmod)

## Loading required package: xts

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: TTR

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

library(xts)
library(rvest)
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --

## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter()         masks stats::filter()
## x dplyr::first()          masks xts::first()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()            masks stats::lag()
## x dplyr::last()           masks xts::last()

library(PerformanceAnalytics)

##
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
##
##     legend

library(corrplot)

## corrplot 0.90 loaded

library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

# Reading Log file data----------------
log <- read.csv(file="logs.csv",encoding="UTF-8")

head(log)

##   X.U.FEFF.DATE         SHW          UNH          KO         JPM
GOOG
## 1       10/1/21  0.13184133  0.182592767  0.08243625  0.03787643
0.112594725
## 2        9/1/21 -0.07714989 -0.061330397 -0.06819389  0.02338236 -
0.083846617
## 3        8/1/21  0.04343201  0.009824836 -0.01262487  0.05990069
0.075735305
## 4        7/1/21  0.06819587  0.033187887  0.06190572 -0.02417385
0.079040126
## 5        6/1/21 -0.03719598 -0.027869495 -0.02134203 -0.05297137
0.039294070
## 6        5/1/21  0.03527218  0.032898770  0.02426822  0.07410005
0.000597457
##            DUK           BA        AMZN         AMT        AAPL
XOM
## 1  0.045291655 -0.058697840  0.02660238  0.06734912  0.05865723
0.09605579
## 2 -0.058900095  0.002004565 -0.05351807 -0.09159053 -0.06664028
0.09519963
## 3 -0.004281195 -0.030819481  0.04303413  0.03313298  0.04092973 -
0.05297903
## 4  0.064728526 -0.054600109 -0.03272227  0.05187670  0.06498235 -
0.08734943
## 5 -0.005485628 -0.030200008  0.06735499  0.05746495  0.10097621
0.09643574
## 6 -0.004667909  0.054244403 -0.07047022  0.00790740 -0.05210718
0.01974137

names(log)
```

```
##  [1] "X.U.FEFF.DATE" "SHW"           "UNH"           "KO"
##  [5] "JPM"           "GOOG"          "DUK"           "BA"
##  [9] "AMZN"          "AMT"           "AAPL"          "XOM"

# Changing variable names to lower cases-----------------
names(log) <- tolower(names(log))

# Converting date into appropriate format----------------
strDates <- c(log$x.u.feff.date)
log$date <- as.Date(strDates, "%m/%d/%y",encoding="UTF-8")
head(log)

##   x.u.feff.date         shw          unh          ko         jpm
goog
## 1       10/1/21  0.13184133  0.182592767  0.08243625  0.03787643
0.112594725
## 2        9/1/21 -0.07714989 -0.061330397 -0.06819389  0.02338236 -
0.083846617
## 3        8/1/21  0.04343201  0.009824836 -0.01262487  0.05990069
0.075735305
## 4        7/1/21  0.06819587  0.033187887  0.06190572 -0.02417385
0.079040126
## 5        6/1/21 -0.03719598 -0.027869495 -0.02134203 -0.05297137
0.039294070
## 6        5/1/21  0.03527218  0.032898770  0.02426822  0.07410005
0.000597457
##            duk           ba        amzn         amt        aapl
xom
## 1  0.045291655 -0.058697840  0.02660238  0.06734912  0.05865723
0.09605579
## 2 -0.058900095  0.002004565 -0.05351807 -0.09159053 -0.06664028
0.09519963
## 3 -0.004281195 -0.030819481  0.04303413  0.03313298  0.04092973 -
0.05297903
## 4  0.064728526 -0.054600109 -0.03272227  0.05187670  0.06498235 -
0.08734943
## 5 -0.005485628 -0.030200008  0.06735499  0.05746495  0.10097621
0.09643574
## 6 -0.004667909  0.054244403 -0.07047022  0.00790740 -0.05210718
0.01974137
##         date
## 1 2021-10-01
## 2 2021-09-01
## 3 2021-08-01
## 4 2021-07-01
## 5 2021-06-01
## 6 2021-05-01
```

```r
# Reading stocks file data----------------
stocks <- read.csv("prices portfolio.csv",encoding="UTF-8")
head(stocks)
```

```
##   X.U.FEFF.DATE    SHW    UNH    KO    JPM    GOOG    DUK     BA    AMZN
AMT
## 1     10/1/2021 316.09 460.47 56.37 168.87 2965.41 101.02 207.03 3372.43
281.97
## 2      9/1/2021 279.27 389.37 52.08 162.71 2665.31  96.65 219.94 3285.04
264.18
## 3      8/1/2021 302.61 414.81 55.89 158.99 2909.24 102.70 219.50 3470.79
290.81
## 4      7/1/2021 290.02 410.78 56.60 150.01 2704.42 103.14 226.48 3327.59
281.49
## 5      6/1/2021 271.50 397.58 53.30 153.72 2506.32  96.87 239.56 3440.16
267.60
## 6      5/1/2021 281.99 408.98 54.47 162.32 2411.56  97.40 247.02 3223.07
253.06
##     AAPL   XOM
## 1 149.58 63.62
## 2 141.29 58.04
## 3 151.38 52.99
## 4 145.43 55.96
## 5 136.56 61.32
## 6 124.03 55.92
```

```r
names(stocks)
```

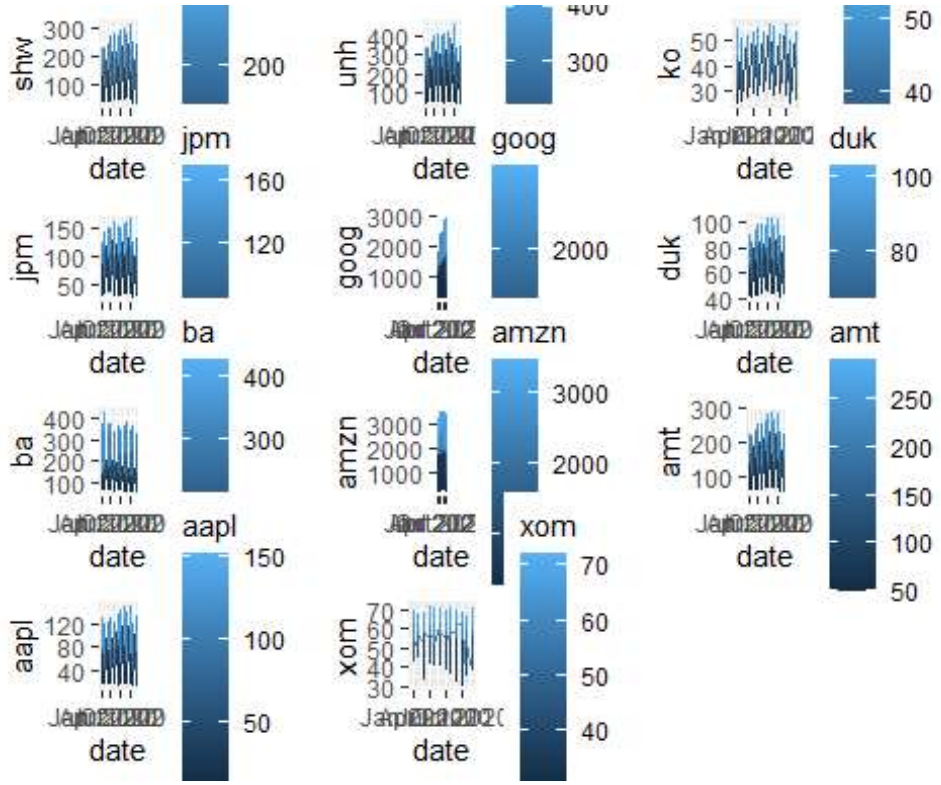
```
##  [1] "X.U.FEFF.DATE" "SHW"           "UNH"           "KO"
##  [5] "JPM"           "GOOG"          "DUK"           "BA"
##  [9] "AMZN"          "AMT"           "AAPL"          "XOM"
```

```r
# Changing variable names to lower cases----------------
names(stocks) <- tolower(names(stocks))

# Converting date into appropriate format---------------

strDates <- c(stocks$x.u.feff.date)
stocks$date <- as.Date(strDates, "%m/%d/%y")
head(stocks)
```

```
##   x.u.feff.date    shw    unh    ko    jpm    goog    duk     ba    amzn
amt
## 1     10/1/2021 316.09 460.47 56.37 168.87 2965.41 101.02 207.03 3372.43
281.97
## 2      9/1/2021 279.27 389.37 52.08 162.71 2665.31  96.65 219.94 3285.04
264.18
## 3      8/1/2021 302.61 414.81 55.89 158.99 2909.24 102.70 219.50 3470.79
290.81
## 4      7/1/2021 290.02 410.78 56.60 150.01 2704.42 103.14 226.48 3327.59
281.49
```

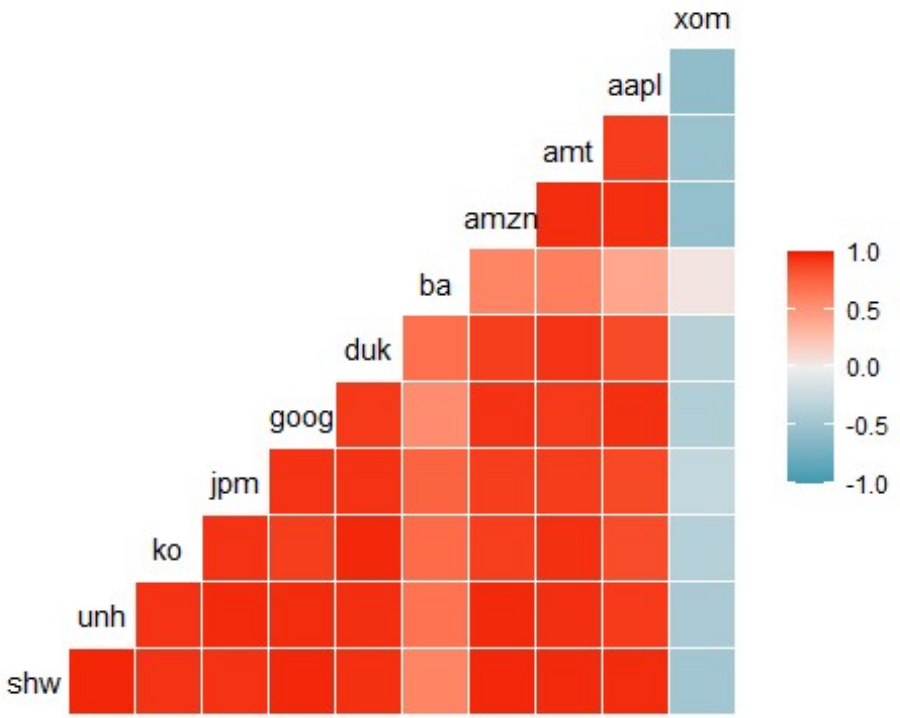
```
## 5      6/1/2021 271.50 397.58 53.30 153.72 2506.32  96.87 239.56 3440.16
267.60
## 6      5/1/2021 281.99 408.98 54.47 162.32 2411.56  97.40 247.02 3223.07
253.06
##     aapl   xom       date
## 1 149.58 63.62 2020-10-01
## 2 141.29 58.04 2020-09-01
## 3 151.38 52.99 2020-08-01
## 4 145.43 55.96 2020-07-01
## 5 136.56 61.32 2020-06-01
## 6 124.03 55.92 2020-05-01

#Subsetting the old dates----------

log = subset(log, select = -c(x.u.feff.date) )
stocks = subset(stocks, select = -c(x.u.feff.date))

#Reading the spy and dia etf data-----

indices <- read.csv("indices.csv")
head(indices)

##        Date    spy    dia
## 1 10/1/2021 358.23 459.25
## 2  9/1/2021 338.29 429.14
## 3  8/1/2021 353.92 451.56
## 4  7/1/2021 349.48 438.51
## 5  6/1/2021 344.95 428.06
## 6  5/1/2021 345.64 420.04

# strDates <- c(indices$Date)
# indices$Date <- as.Date(indices$Date,"%d/%m/%y")
# indices$Date
# indices<-t(indices)

#Converting all values to numerics form
stocks$shw <- as.numeric(stocks$shw)
stocks$unh <- as.numeric(stocks$unh)
stocks$ko <- as.numeric(stocks$ko)
stocks$jpm <- as.numeric(stocks$jpm)
stocks$goog <- as.numeric(stocks$goog)
stocks$duk <- as.numeric(stocks$duk)
stocks$ba <- as.numeric(stocks$ba)
stocks$amzn <- as.numeric(stocks$amzn)
stocks$amt <- as.numeric(stocks$amt)
stocks$aapl <- as.numeric(stocks$aapl)
stocks$xom <- as.numeric(stocks$xom)

require(gridExtra)
```

```
## Loading required package: gridExtra

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

p1 <- ggplot(data = stocks, mapping = aes(x = date, y = shw)) +
  geom_line(alpha = 0.8, aes(color = shw))

p2 <- ggplot(data = stocks, mapping = aes(x = date, y = unh)) +
  geom_line(alpha = 0.9, aes(color = unh))

p3 <- ggplot(data = stocks, mapping = aes(x = date, y = ko)) +
  geom_line(alpha = 0.9, aes(color = ko))

p4 <- ggplot(data = stocks, mapping = aes(x = date, y = jpm)) +
  geom_line(alpha = 0.9, aes(color = jpm))

p5 <-ggplot(data = stocks, mapping = aes(x = date, y = goog)) +
  geom_line(alpha = 0.9, aes(color = goog))

p6 <-ggplot(data = stocks, mapping = aes(x = date, y = duk)) +
  geom_line(alpha = 0.9, aes(color = duk))

p7 <-ggplot(data = stocks, mapping = aes(x = date, y = ba)) +
  geom_line(alpha = 0.9, aes(color = ba))

p8 <- ggplot(data = stocks, mapping = aes(x = date, y = amzn)) +
  geom_line(alpha = 0.9, aes(color = amzn))

p9 <- ggplot(data = stocks, mapping = aes(x = date, y = amt)) +
  geom_line(alpha = 0.9, aes(color = amt))

p10 <- ggplot(data = stocks, mapping = aes(x = date, y = aapl)) +
  geom_line(alpha = 0.9, aes(color = aapl))

p11 <- ggplot(data = stocks, mapping = aes(x = date, y = xom)) +
  geom_line(alpha = 0.9, aes(color = xom))

grid.arrange(p1, p2,p3,p4,p5,p6,p7,p8,p9,p10,p11, ncol=3)
```




```
ggcorr(stocks)
```

```
## Warning in ggcorr(stocks): data in column(s) 'date' are not numeric and
were
## ignored
```




```
# Summary Statistics----------
summary(stocks)
```

```
##       shw              unh              ko              jpm        
##  Min.   : 26.06   Min.   : 41.73   Min.   :24.59   Min.   : 23.51  
##  1st Qu.: 62.98   1st Qu.: 71.94   1st Qu.:32.07   1st Qu.: 46.67  
##  Median : 92.53   Median :139.40   Median :36.66   Median : 64.97  
##  Mean   :117.33   Mean   :169.23   Mean   :38.31   Mean   : 75.09  
##  3rd Qu.:147.68   3rd Qu.:241.27   3rd Qu.:43.81   3rd Qu.: 98.35  
##  Max.   :316.09   Max.   :460.47   Max.   :56.60   Max.   :168.87  
##       goog             duk              ba              amzn        
##  Min.   : 288.9   Min.   : 40.17   Min.   : 55.85   Min.   : 173.1  
##  1st Qu.: 536.3   1st Qu.: 51.89   1st Qu.:109.88   1st Qu.: 333.5  
##  Median : 774.6   Median : 64.04   Median :135.31   Median : 779.5  
##  Mean   : 949.9   Mean   : 65.47   Mean   :179.70   Mean   :1200.4  
##  3rd Qu.:1188.2   3rd Qu.: 77.42   3rd Qu.:245.16   3rd Qu.:1778.3  
##  Max.   :2965.4   Max.   :103.14   Max.   :427.51   Max.   :3470.8  
##       amt              aapl             xom              date        
##  Min.   : 49.24   Min.   : 11.70   Min.   :30.01   Min.   :2020-01-01  
##  1st Qu.: 76.23   1st Qu.: 20.48   1st Qu.:57.88   1st Qu.:2020-03-24  
##  Median :102.25   Median : 28.59   Median :61.70   Median :2020-06-16  
##  Mean   :129.46   Mean   : 44.96   Mean   :59.92   Mean   :2020-06-16  
##  3rd Qu.:189.24   3rd Qu.: 51.34   3rd Qu.:65.62   3rd Qu.:2020-09-08  
##  Max.   :290.81   Max.   :151.38   Max.   :72.12   Max.   :2020-12-01
```

```
summary(log)
```

```
##       shw                unh                  ko                  jpm
##  Min.   :-0.13563   Min.   :-0.12344   Min.   :-0.172743   Min.   :-
0.22461
##  1st Qu.:-0.01317   1st Qu.:-0.01478   1st Qu.:-0.013373   1st Qu.:-
0.01721
##  Median : 0.02428   Median : 0.02194   Median : 0.011439   Median :
0.02705
##  Mean   : 0.02293   Mean   : 0.02194   Mean   : 0.008028   Mean   :
0.01910
##  3rd Qu.: 0.05889   3rd Qu.: 0.05784   3rd Qu.: 0.041044   3rd Qu.:
0.05974
##  Max.   : 0.19773   Max.   : 0.18259   Max.   : 0.082436   Max.   :
0.21352
##       goog               duk                  ba                amzn
##  Min.   :-0.13180   Min.   :-0.109973   Min.   :-0.45466   Min.   :-
0.20219
##  1st Qu.:-0.01769   1st Qu.:-0.019269   1st Qu.:-0.03116   1st Qu.:-
0.03403
##  Median : 0.02106   Median : 0.012173   Median : 0.01590   Median :
0.02612
##  Mean   : 0.02140   Mean   : 0.008923   Mean   : 0.01587   Mean   :
0.02753
##  3rd Qu.: 0.06093   3rd Qu.: 0.042877   3rd Qu.: 0.06198   3rd Qu.:
0.08157
##  Max.   : 0.20192   Max.   : 0.139555   Max.   : 0.45931   Max.   :
0.26890
##       amt                aapl                xom                 date
##  Min.   :-0.12731   Min.   :-0.18404   Min.   :-0.251261   Min.   :2011-
12-01
##  1st Qu.:-0.02028   1st Qu.:-0.02628   1st Qu.:-0.031696   1st Qu.:2014-
05-16
##  Median : 0.01406   Median : 0.02873   Median : 0.005256   Median :2016-
11-01
##  Mean   : 0.01598   Mean   : 0.02474   Mean   : 0.003822   Mean   :2016-
10-31
##  3rd Qu.: 0.04944   3rd Qu.: 0.07685   3rd Qu.: 0.042573   3rd Qu.:2019-
04-16
##  Max.   : 0.16791   Max.   : 0.21438   Max.   : 0.223861   Max.   :2021-
10-01
```

```
summary(indices)
```

```
##      Date                spy              dia
##  Length:120         Min.   :120.2   Min.   :125.0
##  Class :character   1st Qu.:164.8   1st Qu.:190.8
##  Mode  :character   Median :187.8   Median :218.9
##                     Mean   :210.6   Mean   :241.8
##                     3rd Qu.:258.2   3rd Qu.:290.4
##                     Max.   :358.2   Max.   :459.2
```

```r
# Taking the stock prices as on Nov 1 2011-----

initialp <- c(stocks[120,1:11])
initialp <-as.numeric(initialp)

# Investing a sum of 1 Million equally among the stocks in our portfolio
# on Nov 1st 2011
1000000/11
```

```
## [1] 90909.09
```

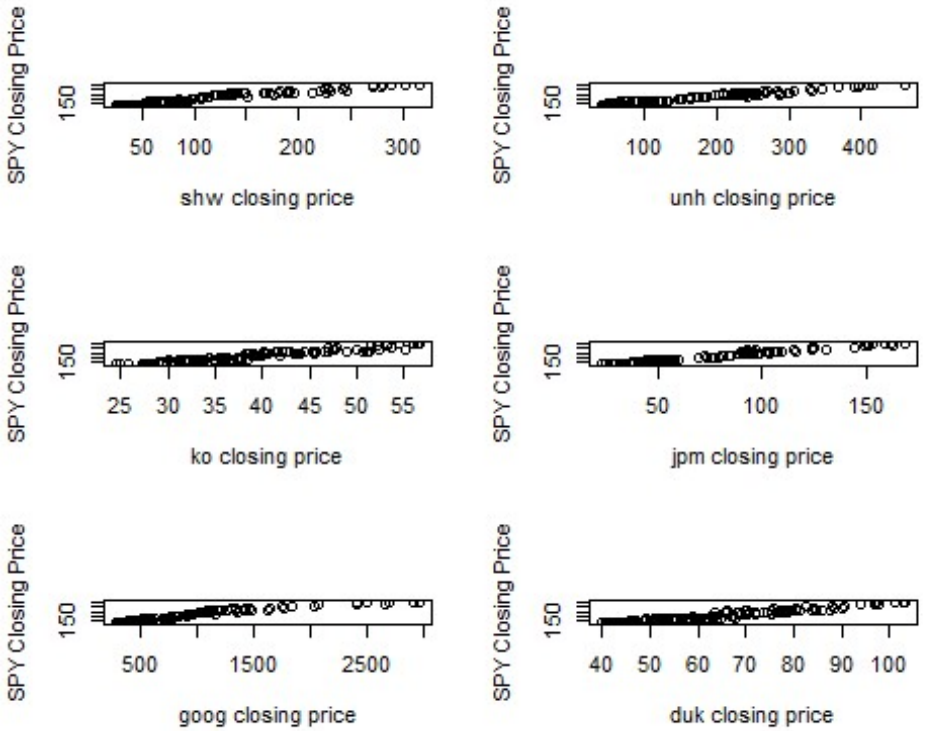
```r
units <- 90909.9/initialp
units
```

```
##  [1] 3488.4843 2178.5262 3697.0272 3866.8609  304.4742 2263.1292 1627.7511
##  [8]  472.7750 1846.2612 7770.0769 1712.3733
```

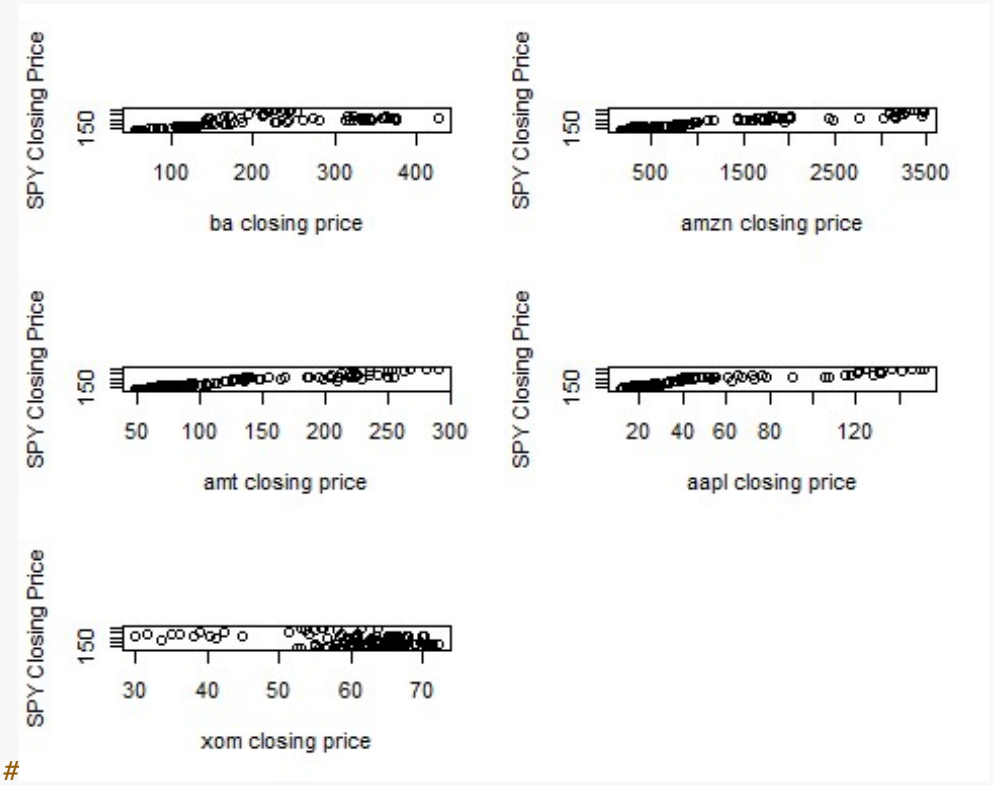
```r
finalp <- c(stocks[1,1:11])
finalp <-as.numeric(finalp)

#Portfolio Returns
sum(finalp*units)
```

```
## [1] 7821905
```

```r
y <- indices$spy
x1 <- stocks$shw
x2 <- stocks$unh
x3 <- stocks$ko
x4 <- stocks$jpm
x5 <- stocks$goog
x6 <- stocks$duk
x7 <- stocks$ba
x8 <- stocks$amzn
x9 <- stocks$amt
x10 <- stocks$aapl
x11 <- stocks$xom

#scatterplots

par(mfrow=c(3, 2))
plot(y ~ x1, xlab = "shw closing price", ylab = "SPY Closing Price")
plot(y ~ x2, xlab = "unh closing price", ylab = "SPY Closing Price")
plot(y ~ x3, xlab = "ko closing price", ylab = "SPY Closing Price")
plot(y ~ x4, xlab = "jpm closing price", ylab = "SPY Closing Price")
plot(y ~ x5, xlab = "goog closing price", ylab = "SPY Closing Price")
plot(y ~ x6, xlab = "duk closing price", ylab = "SPY Closing Price")
```




```
plot(y ~ x7, xlab = "ba closing price", ylab = "SPY Closing Price")
plot(y ~ x8, xlab = "amzn closing price", ylab = "SPY Closing Price")
plot(y ~ x9, xlab = "amt closing price", ylab = "SPY Closing Price")
plot(y ~ x10, xlab = "aapl closing price", ylab = "SPY Closing Price")
plot(y ~ x11, xlab = "xom closing price", ylab = "SPY Closing Price")
```


SPY Closing Price
150
100
200
300
400
ba closing price
SPY Closing Price
150
500
1500
2500
3500
amzn closing price
SPY Closing Price
150
50
100
150
200
250
300
amt closing price
SPY Closing Price
150
20
40
60
80
120
aapl closing price
SPY Closing Price
150
30
40
50
60
70
xom closing price

#