Homework

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(tidyr)  
library(ggplot2)

my\_stocks <- read.csv("stocks.csv",header=TRUE, stringsAsFactors = FALSE)

**Your job first, is to ‘know your data’. Run whatever code you feel necessary to get to know your data. Add more code blocks as needed, leave this part in your document when you submit it**

str(my\_stocks)

## 'data.frame': 65020 obs. of 9 variables:  
## $ exchange : chr "NYSE" "NYSE" "NYSE" "NYSE" ...  
## $ stock\_symbol : chr "XL" "XL" "XL" "XL" ...  
## $ date : chr "2/8/10" "2/5/10" "2/4/10" "2/3/10" ...  
## $ stock\_price\_open : num 16.5 16.4 17 17.2 16.9 ...  
## $ stock\_price\_high : num 16.9 16.6 17 17.3 17.5 ...  
## $ stock\_price\_low : num 16.3 15.9 16.3 17.1 16.8 ...  
## $ stock\_price\_close : num 16.5 16.5 16.4 17.1 17.3 ...  
## $ stock\_volume : int 4793200 4760900 6716100 2657900 4282200 3258200 4546200 4069700 3339600 4608300 ...  
## $ stock\_price\_adj\_close: num 16.5 16.5 16.4 17.1 17.3 ...

summary(my\_stocks)

## exchange stock\_symbol date stock\_price\_open  
## Length:65020 Length:65020 Length:65020 Min. : 0.16   
## Class :character Class :character Class :character 1st Qu.: 14.39   
## Mode :character Mode :character Mode :character Median : 25.47   
## Mean : 33.41   
## 3rd Qu.: 47.63   
## Max. :162.75   
## stock\_price\_high stock\_price\_low stock\_price\_close stock\_volume   
## Min. : 0.17 Min. : 0.16 Min. : 0.16 Min. : 0   
## 1st Qu.: 14.53 1st Qu.: 14.22 1st Qu.: 14.39 1st Qu.: 3200   
## Median : 25.55 Median : 25.40 Median : 25.48 Median : 192550   
## Mean : 33.73 Mean : 33.09 Mean : 33.42 Mean : 2094209   
## 3rd Qu.: 48.00 3rd Qu.: 47.25 3rd Qu.: 47.63 3rd Qu.: 2178000   
## Max. :164.00 Max. :161.38 Max. :162.50 Max. :142985300   
## stock\_price\_adj\_close  
## Min. : 0.22   
## 1st Qu.: 6.09   
## Median : 12.23   
## Mean : 17.73   
## 3rd Qu.: 19.53   
## Max. :171.50

head(my\_stocks)

## exchange stock\_symbol date stock\_price\_open stock\_price\_high  
## 1 NYSE XL 2/8/10 16.47 16.85  
## 2 NYSE XL 2/5/10 16.38 16.55  
## 3 NYSE XL 2/4/10 17.02 17.02  
## 4 NYSE XL 2/3/10 17.25 17.34  
## 5 NYSE XL 2/2/10 16.93 17.52  
## 6 NYSE XL 2/1/10 16.75 17.09  
## stock\_price\_low stock\_price\_close stock\_volume stock\_price\_adj\_close  
## 1 16.29 16.51 4793200 16.51  
## 2 15.91 16.46 4760900 16.46  
## 3 16.31 16.41 6716100 16.41  
## 4 17.09 17.14 2657900 17.14  
## 5 16.80 17.33 4282200 17.33  
## 6 16.64 16.88 3258200 16.88

tail(my\_stocks)

## exchange stock\_symbol date stock\_price\_open stock\_price\_high  
## 65015 NYSE XFR 5/13/04 22.90 22.90  
## 65016 NYSE XFR 5/12/04 22.60 22.70  
## 65017 NYSE XFR 5/11/04 22.80 22.90  
## 65018 NYSE XFR 5/10/04 23.00 23.00  
## 65019 NYSE XFR 5/7/04 23.55 23.55  
## 65020 NYSE XFR 5/6/04 24.00 24.00  
## stock\_price\_low stock\_price\_close stock\_volume stock\_price\_adj\_close  
## 65015 22.9 22.90 400 15.50  
## 65016 22.6 22.70 400500 15.37  
## 65017 22.6 22.65 2600 15.33  
## 65018 22.3 22.70 3800 15.37  
## 65019 23.2 23.20 2900 15.71  
## 65020 23.7 23.70 2200 16.05

**There are 65020 observations and 9 variables. The head() and tail()are showing the first and last 6 rows of the data set. The summary table is showing statistical measures for all numeric and integer columns.**

**How many different stocks are there (by stock symbol) (5)?**

my\_stocks %>% count(stock\_symbol)

## stock\_symbol n  
## 1 XAA 3644  
## 2 XCJ 511  
## 3 XCO 1006  
## 4 XEC 1757  
## 5 XEL 6146  
## 6 XFB 1486  
## 7 XFD 1594  
## 8 XFH 1594  
## 9 XFJ 1594  
## 10 XFL 1594  
## 11 XFP 1594  
## 12 XFR 1450  
## 13 XIN 543  
## 14 XJT 1955  
## 15 XKE 1619  
## 16 XKK 1594  
## 17 XKN 1594  
## 18 XKO 2019  
## 19 XL 4677  
## 20 XOM 10122  
## 21 XRM 1189  
## 22 XRX 8353  
## 23 XTO 4182  
## 24 XVF 1717  
## 25 XVG 1486

**There are 25 diffrent stocks in the dataset.**

**What are those stocks (5)?**

table(my\_stocks$stock\_symbol)

##   
## XAA XCJ XCO XEC XEL XFB XFD XFH XFJ XFL XFP XFR XIN   
## 3644 511 1006 1757 6146 1486 1594 1594 1594 1594 1594 1450 543   
## XJT XKE XKK XKN XKO XL XOM XRM XRX XTO XVF XVG   
## 1955 1619 1594 1594 2019 4677 10122 1189 8353 4182 1717 1486

**Which stock had the highest single gain in a day, which day was it (10)?**

my\_stocks["Profit\_or\_Loss"]<-my\_stocks$stock\_price\_close - my\_stocks$stock\_price\_open

subset(my\_stocks,Profit\_or\_Loss==max(Profit\_or\_Loss))

## exchange stock\_symbol date stock\_price\_open stock\_price\_high  
## 10582 NYSE XOM 10/13/08 64.9 73.74  
## stock\_price\_low stock\_price\_close stock\_volume stock\_price\_adj\_close  
## 10582 63.88 73.08 64864300 70.56  
## Profit\_or\_Loss  
## 10582 8.18

**The (NYSE) XOM had the highest single gain (8.18) in a day and the day was 2008-10-13.**

**Which stock had the highest single loss in a day, which day was it (10)?**

subset(my\_stocks,Profit\_or\_Loss==min(Profit\_or\_Loss))

## exchange stock\_symbol date stock\_price\_open stock\_price\_high  
## 10584 NYSE XOM 10/9/08 77.6 78.45  
## stock\_price\_low stock\_price\_close stock\_volume stock\_price\_adj\_close  
## 10584 67.47 68 68771800 65.65  
## Profit\_or\_Loss  
## 10584 -9.6

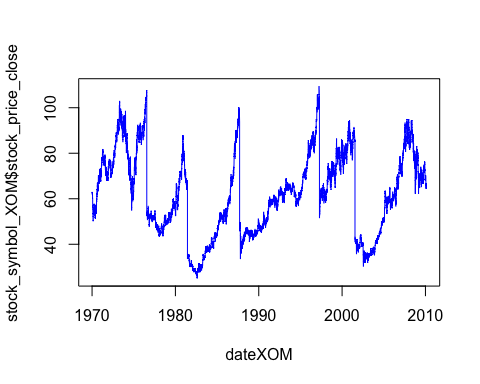
**The (NYSE) XOM had the highest single loss (-9.6) in a day and the day was 2008-10-09**

**Create a plot of closing price for XOM over time (5)**

stock\_symbol\_XOM<-my\_stocks[my\_stocks$stock\_symbol=="XOM",]

dateXOM<-as.Date(stock\_symbol\_XOM$date,format = "%m/%d/%y")

plot(dateXOM,stock\_symbol\_XOM$stock\_price\_close,type = "l",col="blue")

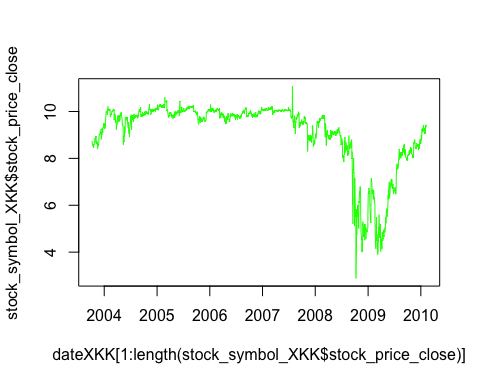


**Create a plot of closing price for XKK over time (5)**

stock\_symbol\_XKK<-my\_stocks[my\_stocks$stock\_symbol=="XKK",]

dateXKK<-as.Date(stock\_symbol\_XKK$date,format = "%m/%d/%y")

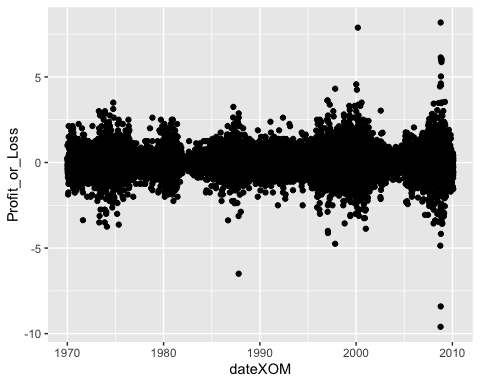
plot(dateXKK[1:length(stock\_symbol\_XKK$stock\_price\_close)],stock\_symbol\_XKK$stock\_price\_close,type="l",col="green")



**Create a plot of the change in price per day for both XOM and XKK (10)**

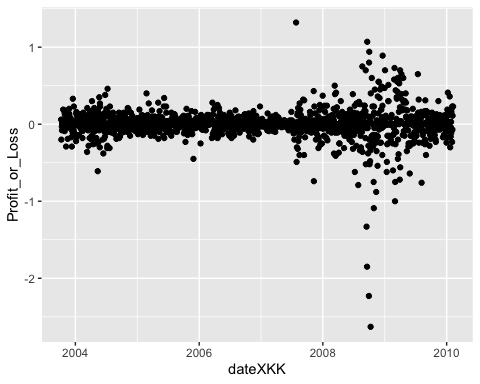
stock\_XOM<-filter(my\_stocks,stock\_symbol==c("XOM"))

ggplot(data = stock\_XOM,aes(x=dateXOM,y=Profit\_or\_Loss))+geom\_point()



stock\_XKK<-filter(my\_stocks,stock\_symbol==c("XKK"))

ggplot(data = stock\_XKK,aes(dateXKK,y=Profit\_or\_Loss))+geom\_point()



This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

plot(cars)



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.