Homework 2

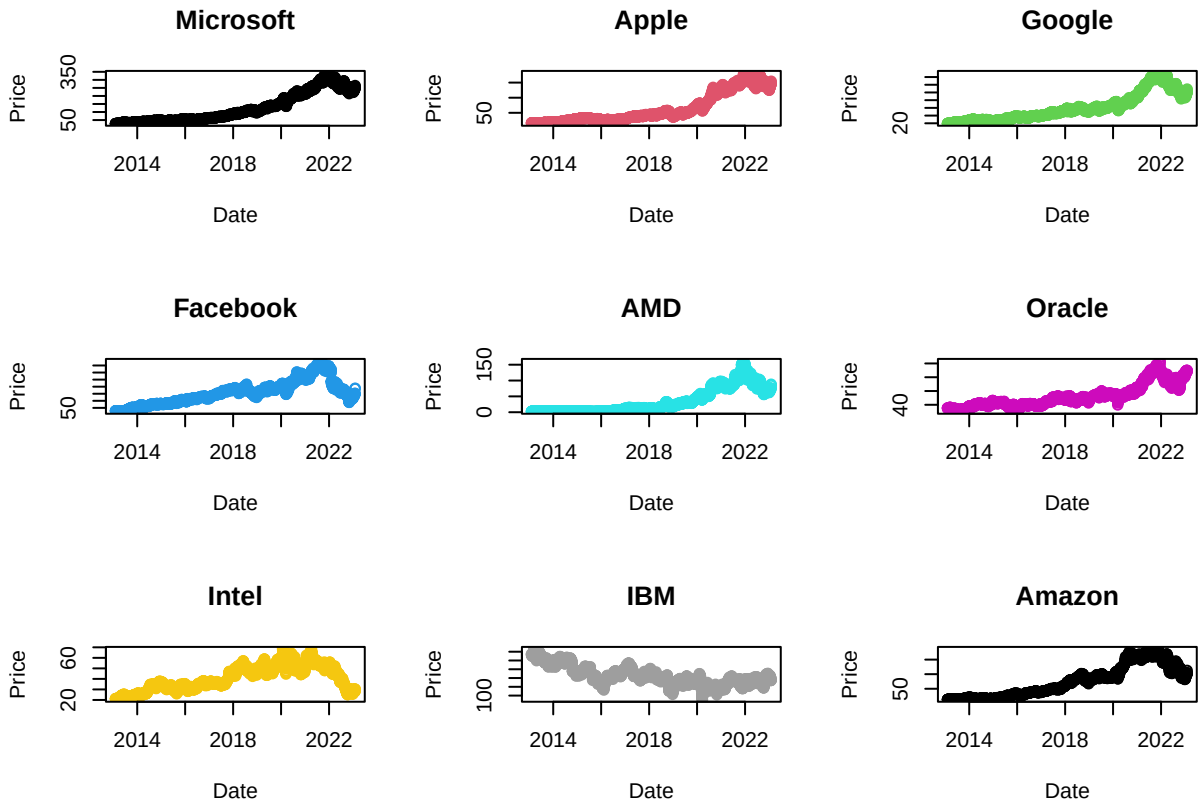
R Homework 2

1 Plotting many time series together. Source: Yahoo fiance. Consider 9 of the most pronominal tech company. Microsoft, Apple Google, Facebook, AMD, Oracle, Intel, IBM, Amazon, 1. Download their closing prices (Wall st) for the past 10 years

```
Microsoft <- read.csv(file = 'MSFT.csv')
Apple <- read.csv(file = 'AAPL.csv')
Google <- read.csv(file = 'GOOG.csv')
Facebook <- read.csv(file = 'META.csv')
AMD <- read.csv(file = 'AMD.csv')
Oracle <- read.csv(file = 'ORCL.csv')
Intel <- read.csv(file = 'INTC.csv')
IBM <- read.csv(file = 'IBM.csv')
Amazon <- read.csv(file = 'AMZN.csv')

Microsoft <- data.frame(x = as.Date(Microsoft$Date), y = Microsoft$Close)
Apple <- data.frame(x = as.Date(Apple$Date), y = Apple$Close)
Google <- data.frame(x = as.Date (Google$Date), y = Google$Close)
Facebook <- data.frame(x = as.Date(Facebook$Date), y = Facebook$Close)
AMD <-  data.frame(x =  as.Date(AMD$Date), y = AMD$Close)
Oracle <- data.frame(x =  as.Date(Oracle$Date), y = Oracle$Close)
Intel <- data.frame(x =  as.Date(Intel$Date), y = Intel$Close)
IBM <- data.frame(x =  as.Date(IBM$Date), y = IBM$Close)
Amazon <- data.frame(x =  as.Date(Amazon$Date), y = Amazon$Close)
```

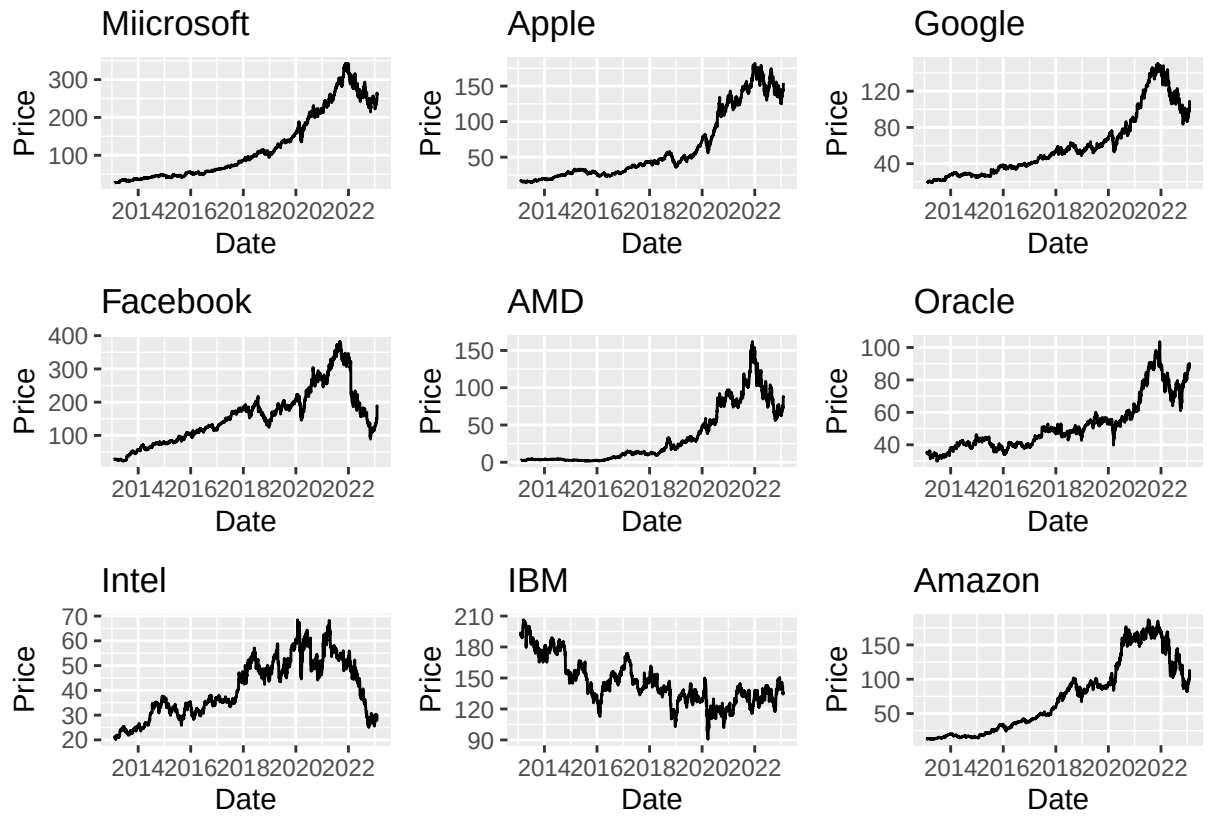
2. Plot (traditional) on a 3X 3 grid layout with proper labeling

```
par(mfrow = c(3,3))
plot(Microsoft,xlab = "Date", ylab = "Price", main = "Microsoft",col = 1)
plot(Apple,xlab = "Date", ylab = "Price", main = "Apple",col = 2)
plot(Google,xlab = "Date", ylab = "Price", main = "Google",col = 3)
plot(Facebook,xlab = "Date", ylab = "Price", main = "Facebook",col = 4)
plot(AMD,xlab = "Date", ylab = "Price", main = "AMD",col = 5)
plot(Oracle,xlab = "Date", ylab = "Price", main = "Oracle",col = 6)
plot(Intel,xlab = "Date", ylab = "Price", main = "Intel",col = 7)
plot(IBM,xlab = "Date", ylab = "Price", main = "IBM",col = 8)
plot(Amazon,xlab = "Date", ylab = "Price", main = "Amazon",col = 9)
```



3. Redo the plotting of part 2 using ggplot2

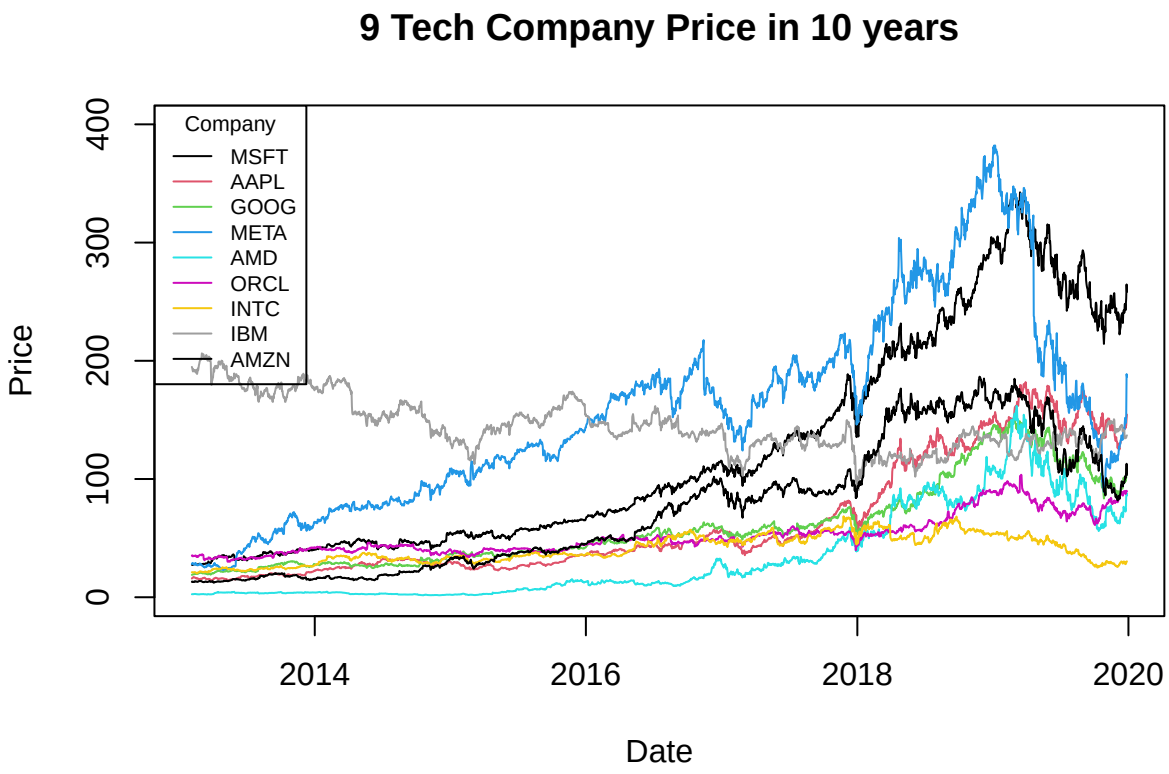
```
library(ggplot2)
library(patchwork)

plot1 <- ggplot(data = Microsoft,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggti
plot2 <- ggplot(data = Apple,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle(
plot3 <- ggplot(data = Google,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle
plot4 <- ggplot(data = Facebook,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtit
plot5 <- ggplot(data = AMD,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle("A
plot6 <- ggplot(data = Oracle,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle
plot7 <- ggplot(data = Intel,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle(
plot8 <- ggplot(data = IBM,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle("I
plot9 <- ggplot(data = Amazon,aes(x = x, y = y)) + geom_line() + xlab("Date") + ylab("Price") + ggtitle

plot1 + plot2 + plot3 + plot4 + plot5 + plot6 + plot7 + plot8 + plot9
```



4.

a. Plot all the 9 on the same plot with proper legend

```
Dates <- seq(as.Date("2013/02/04"), by = "day", length.out = 2519)
Stock_df <- cbind(Microsoft$y,Apple$y,Google$y,Facebook$y,AMD$y, Oracle$y, Intel$y, IBM$y,Amazon$y)
colnames(Stock_df) <- c("Microsoft","Apple","Google","Facebook","AMD","Oracle","Intel","IBM","Amazon")
Stock_df <- cbind.data.frame(x = Dates, Stock_df)
plot(Stock_df$x,Stock_df$Microsoft,type = "l",col = 1, ylim = c(0,400),xlab = "Date", ylab = "Price", ma
lines(Stock_df$x,Stock_df$Apple,type = "l",col=2)
lines(Stock_df$x,Stock_df$Google,type = "l",col=3)
lines(Stock_df$x,Stock_df$Facebook,type = "l",col=4)
lines(Stock_df$x,Stock_df$AMD,type = "l",col=5)
lines(Stock_df$x,Stock_df$Oracle,type = "l",col=6)
lines(Stock_df$x,Stock_df$Intel,type = "l",col=7)
lines(Stock_df$x,Stock_df$IBM,type = "l",col=8)
lines(Stock_df$x,Stock_df$Amazon,type = "l",col=9)
legend("topleft",title= "Company", c("MSFT","AAPL","GOOG","META","AMD","ORCL","INTC","IBM","AMZN"),lty =
```



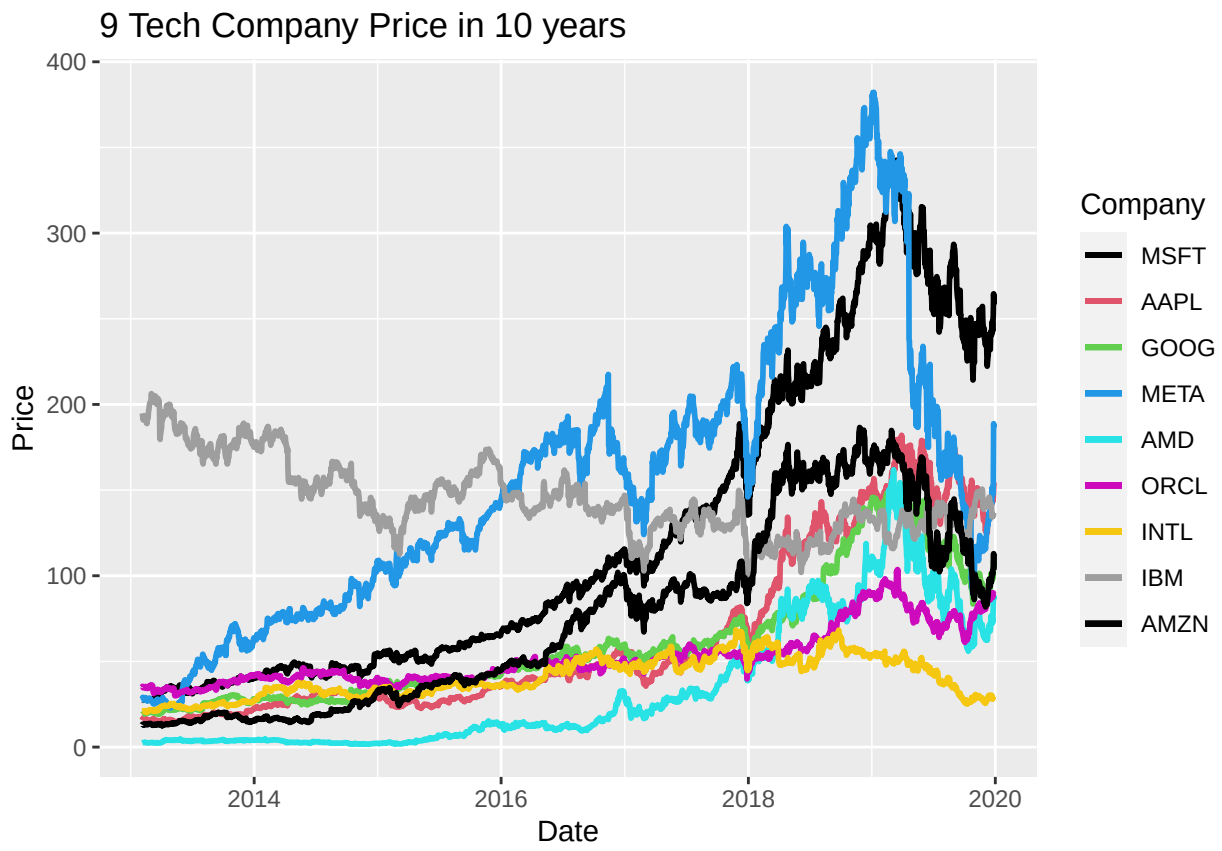
b. Comment on whatever you notice that is worth noticing.

c. Perform the necessary operation to remedy and problem

d. Plot the transformed

5. Redo 4(a) and 4(d) using ggplot2

```
ggplot() +
  geom_line(data = Stock_df, aes(y=Microsoft,x=x,color="MSFT"),size=1) +
  geom_line(data = Stock_df, aes(y=Apple,x=x,color="AAPL"),size=1) +
  geom_line(data = Stock_df, aes(y=Google,x=x,color="GOOG"),size=1) +
  geom_line(data = Stock_df, aes(y= Facebook,x=x,color="META"),size=1) +
  geom_line(data = Stock_df, aes(y=AMD,x=x,color="AMD"),size=1) +
  geom_line(data = Stock_df, aes(y=Oracle,x=x,color="ORCL"),size=1) +
  geom_line(data = Stock_df, aes(y=Intel,x=x,color="INTL"),size=1) +
  geom_line(data = Stock_df, aes(y=IBM,x=x,color="IBM"),size=1) +
  geom_line(data = Stock_df, aes(y=Amazon,x=x,color="AMZN"),size=1) +
  scale_colour_manual(name = "Company", values = c("MSFT" = 1,"AAPL"=2,"GOOG"=3,"META"=4,"AMD"=5,"ORCL"=
  xlab("Date")+
  ylab("Price") +
  ggtitle("9 Tech Company Price in 10 years")
```



2: Writing a function for creating many base learners 1. Let L be a number like 50,100,500, Write an R function that takes in L and D_n (data set) and build L bootstrap realizations of f_hat as linear model Y_i = f(x)+ e_i, i = 1.....n where (x_i,Y_i) are coming from delta_n, Return a list of L models representing f_hat1,f_hat2,.....f_hat_n Each f_hat is built using D_n boot from D_n for l=1,.....L Start with lm() linear model

```
library(cherryblossom)
library(usdata)
library(airports)
library(openintro)
```

```
data("gifted")
Dn <- gifted
L <- 50
set.seed(787)
f <-lm(Dn$score~ .,data = Dn)
mse_f <- rev(anova(f)$"Mean Sq")[1]
f_hat <- list()
mse_hat <- NULL
```

```
bagged.lm <- function(Dn,L){
  for (i in 1: L) {
    boot_Dn <- Dn[sample(L,replace = T),]
    lm(boot_Dn$score~ ., data = boot_Dn)
    f_hat[[i]] <- lm(boot_Dn$score~ ., data = boot_Dn)
  }
```

```
  return(f_hat)
}
```

2. Compute f_hat_l(x)= 1/L sum of f_hat(x_i) where i = 1,. . . .n f_hat (x_i),i = 1,. . . .n

```
set.seed(787)
f_hat <- bagged.lm(Dn,L)
for (i in 1:L) {
  mse_hat[i] <- rev(anova(f_hat[[i]])$"Mean Sq")[1]
}
```

3. Calculate MSE (f_hat(L)) and MSE (f_hat)

```
mse_f <- rev(anova(f)$"Mean Sq")[1]
mse_f_hat <- mean(mse_hat)
```