**PSG COLLEGE OF TECHNOLOGY**

**DEPARTMENT OF APPLIED MATHEMATICS AND**

**COMPUTATIONAL SCIENCES**

****

**PROJECT ON:** APPLIED STATISTICS AND R PROGRAMMING

**FACULTY:**

DR. SENTHILKUMAR M

MR. KUMARESAN M

**PROJECT TITLE: “**CASE STUDY 5”

**PROJECT MEMBERS:**

* Anchala C – 20PD03
* Rhini R – 20PD21
* Sivani U – 20PD29
* Vijayasri R – 20PD38

**DATA:** Casestudy6.txt

# Case Study-5: *Stock Prices 1*

A basic rule of thumb for investors in the stock market is to ``diversify''; that is to spread one's money across stocks which are likely to behave differently in response to various conditions in the market. Risk to the investor is reduced because, under a given set of circumstances, some stocks in the portfolio will rise while others fall. How can one determine which stocks are similar and which are not for the purpose of diversification?

The [data](http://www.stat.ucla.edu/cases/stocks1/Stock1.data) provided are daily stock prices from January 1988 through October 1989, for ten aerospace companies. Given this information, the first step toward answering the question posed above is to reformulate the question in terms of these data. For example, two stocks may be considered similar if they maintain approximately the same level, vary to a similar degree, or tend to move up and down in related ways over some relevant time period. An initial analysis might use some graphical techniques to examine these aspects of the data.

Describe the following using statistical software:

1. Make histograms of these price series
2. Time plots:

Another simple tool for comparing price series over time is the univariate time plot. [Plot](http://www.stat.ucla.edu/cases/stocks1/Stocktime.lsp) stock price on day for each of the ten companies for which price series is provided. Are the Y axis scales the same for all plots? What advantages are there in making all scales the same? What are the disadvantages? Look at the overall shapes of the plots.

1. It might also be useful to have one or two numbers that capture relevant characteristics of a stock's behavior. Mean and variance are two descriptive statistics often used to summarize data. Compute the [means](http://www.stat.ucla.edu/cases/stocks1/Stockmean.lsp) of stock prices for Companies A through J. Which company has the highest mean price? The lowest? Find the means on the histograms. Does this mean that the company with the higher mean is a better investment than the company with the lower mean? Describe the histograms of the companies with the highest and lowest means.
2. Make histograms of these price series

**CODE:**

#loading the required packages for the case study

library(tidyverse)

library(lubridate)

library(ggpmisc)

#load data from csv file to a dataframe

data<- read.csv(file='C:\\Users\\RHINI\\Downloads\\Dataset\_R1.csv',header=T)

data%>%glimpse()

#plotting histograms of stock prices for each of the 10 companies

ggplot(data,aes(x=A))+ geom\_histogram(bins=80,color='black',fill='aliceblue')

ggplot(data,aes(x=B))+ geom\_histogram(bins=80,color='black',fill='aliceblue')

ggplot(data,aes(x=C))+ geom\_histogram(bins=80,color='black',fill='aliceblue')

ggplot(data,aes(x=D))+ geom\_histogram(bins=80,color='black',fill='aliceblue')

ggplot(data,aes(x=E))+ geom\_histogram(bins=80,color='black',fill='aliceblue')

ggplot(data,aes(x=F))+ geom\_histogram(bins=80,color='black',fill='thistle2')

ggplot(data,aes(x=G))+ geom\_histogram(bins=80,color='black',fill='thistle2')

ggplot(data,aes(x=H))+ geom\_histogram(bins=80,color='black',fill='thistle2')

ggplot(data,aes(x=I))+ geom\_histogram(bins=80,color='black',fill='thistle2')

ggplot(data,aes(x=J))+ geom\_histogram(bins=80,color='black',fill='thistle2')

#comparison of distibution of stock prices of all companies in a single plot

ggplot()+

geom\_histogram(aes(x=data$A,fill='A'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$B,fill='B'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$C,fill='C'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$D,fill='D'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$E,fill='E'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$F,fill='F'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$G,fill='G'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$H,fill='H'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$I,fill='I'),bins=30,alpha=0.5)+

geom\_histogram(aes(x=data$J,fill='J'),bins=30,alpha=0.5)+

xlab("Stock prices of 10 companies(with each color representing a company)")+

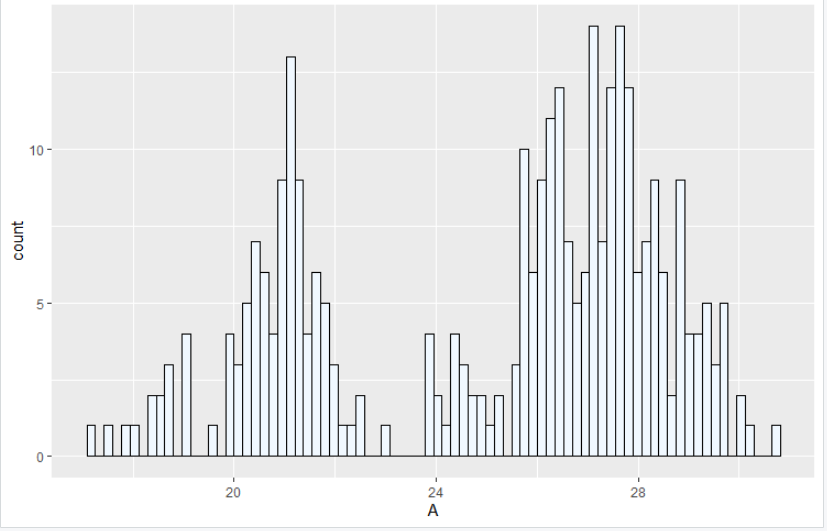
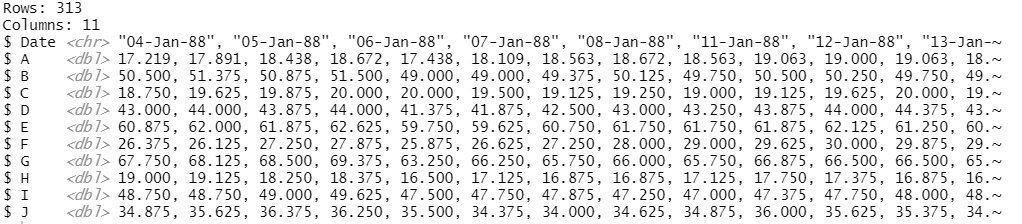
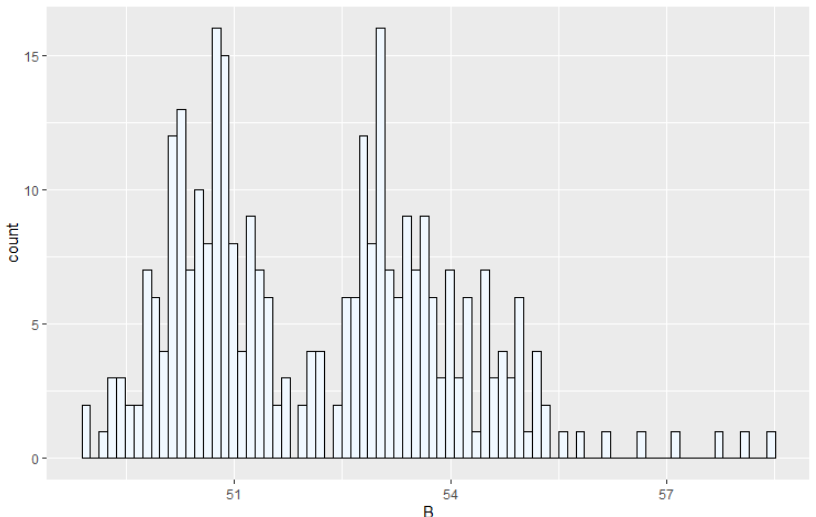
ylab("Count")+

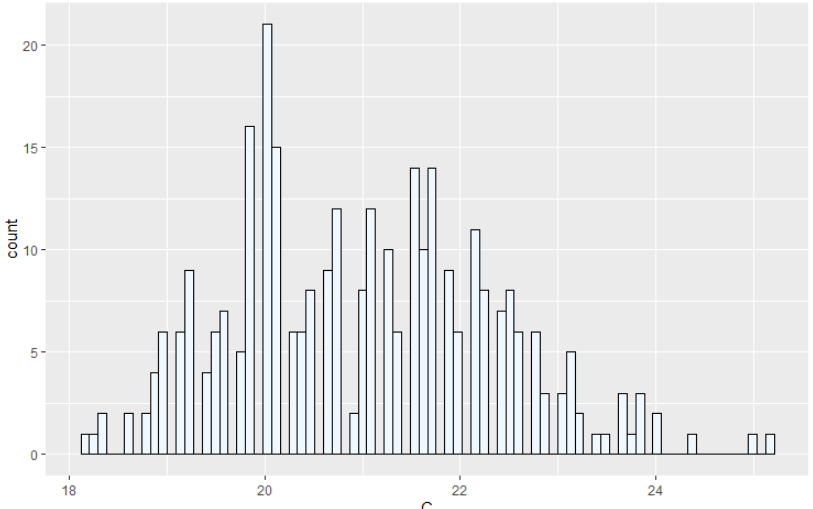
ggtitle("Distribution of stock prices of 10 companies")+

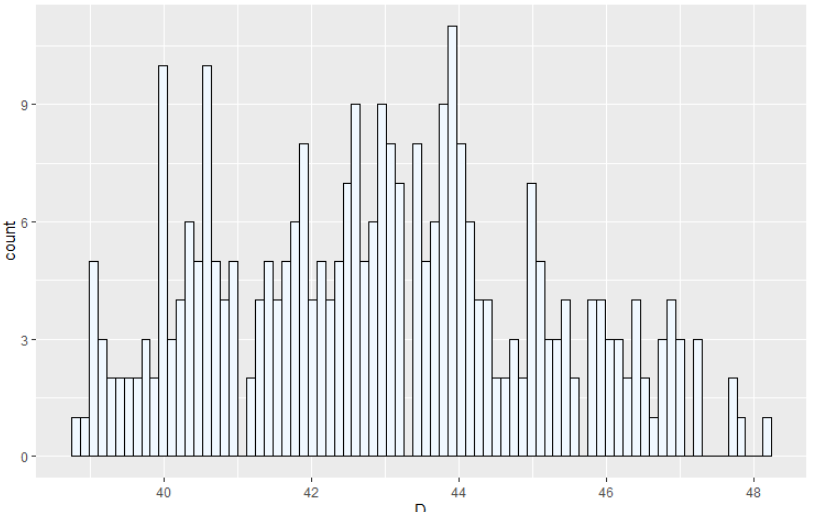
theme(plot.title=element\_text(hjust=0.5))+

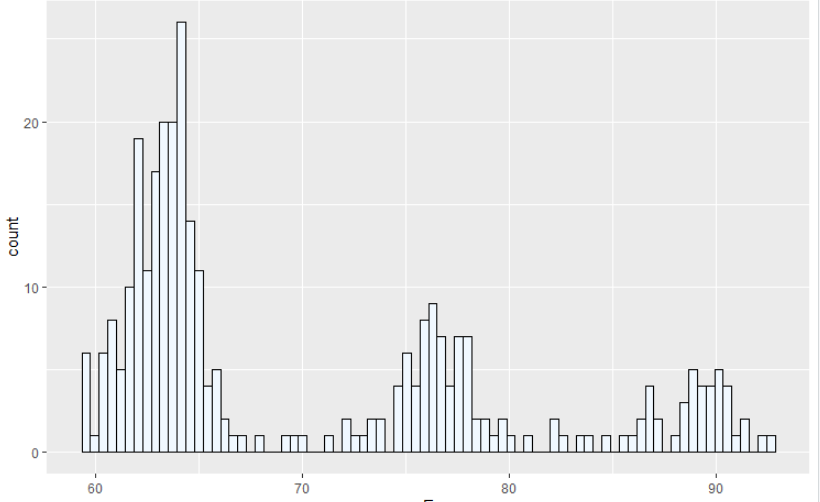
scale\_fill\_manual(values=c('salmon','navy','deepskyblue','gold','deeppink','blue','orange','red','green','yellow'),name="Company")

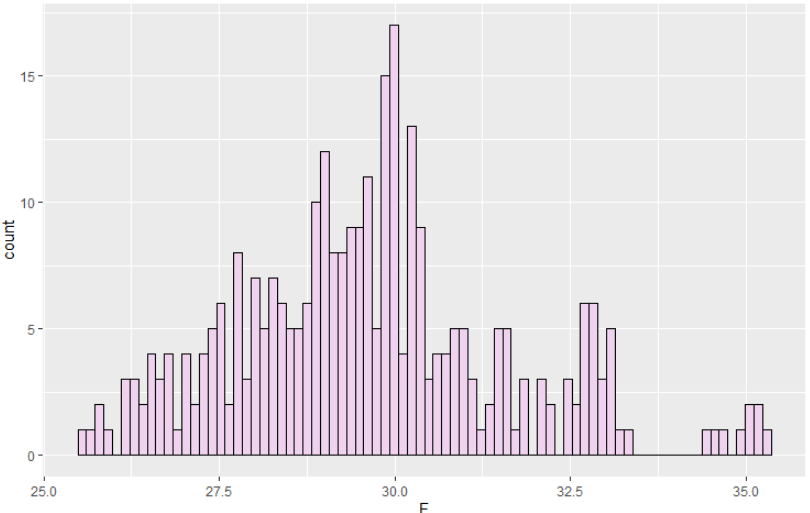
**OUTPUT:**

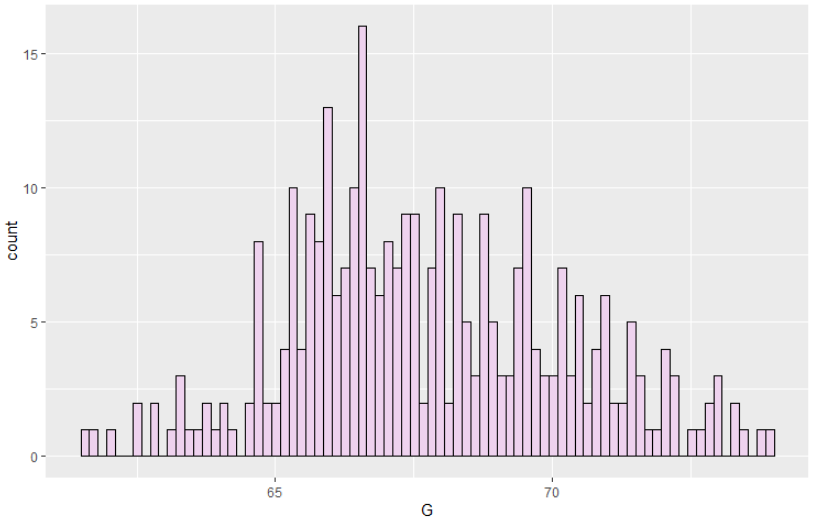
 

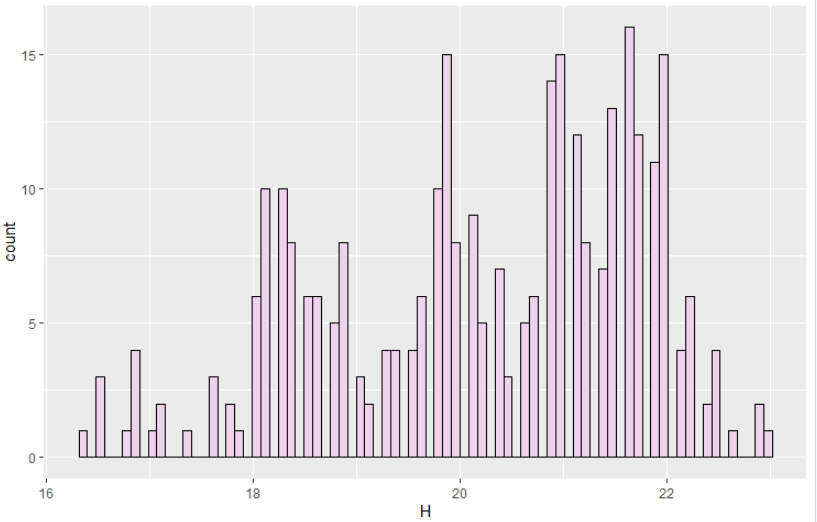


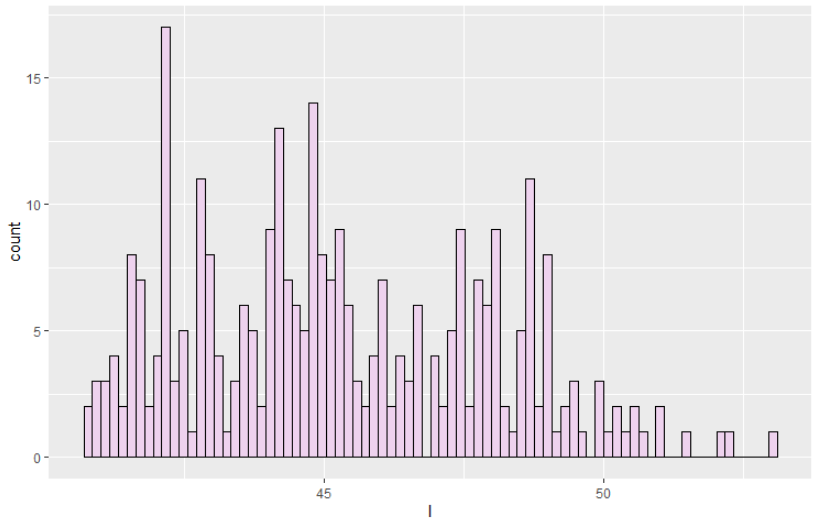


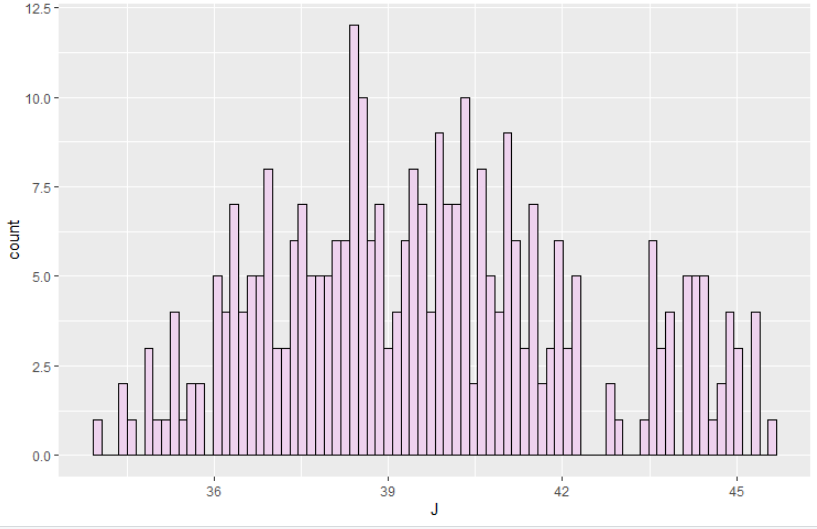


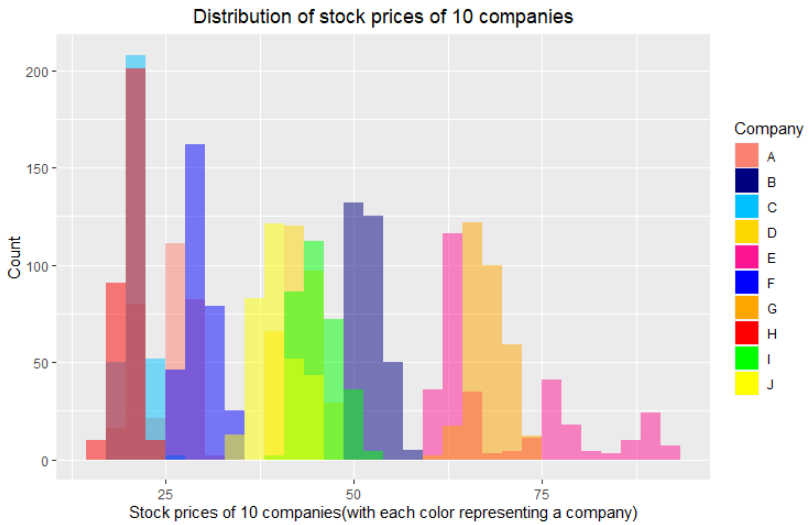












2.Univariate time-series plot

**CODE:**

#mutating the Date column to the default Date format used in R

data <- data%>%

mutate(Date=dmy(Date))

data%>%glimpse()

#plotting the univariate time-series plot for stock prices of all companies separately

ggplot(data,aes(x=Date,y=A))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=B))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=C))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=D))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=E))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=F))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=G))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=H))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=I))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

ggplot(data,aes(x=Date,y=J))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

stat\_smooth(color='red',method="loess")

#limiting the time-axix(x-axis) to a duration of one month

#and also pointing the peaks and valleys of stock prices of company A at that time period

min<-as.Date("1988-10-21")

max<-as.Date("1988-11-21")

ggplot(data,aes(x=Date,y=A))+

geom\_line(color='black',size=2)+

scale\_x\_date(date\_labels="%b/%Y")+

scale\_x\_date(limits=c(min,max))+

stat\_peaks(color='green')+

stat\_peaks(geom='text',color='green',vjust=-0.5,x.label.fmt="%d%a")+

stat\_valleys(color='blue')+

stat\_valleys(geom='text',color='blue',vjust=1.5,hjust=1,x.label.fmt="%d%a")

#making the y axis same for all time-series plots to make comparison easy

ggplot(data,aes(x=Date,y=A))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=B))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=C))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=D))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=E))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=F))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=G))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=H))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

ggplot(data,aes(x=Date,y=I))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

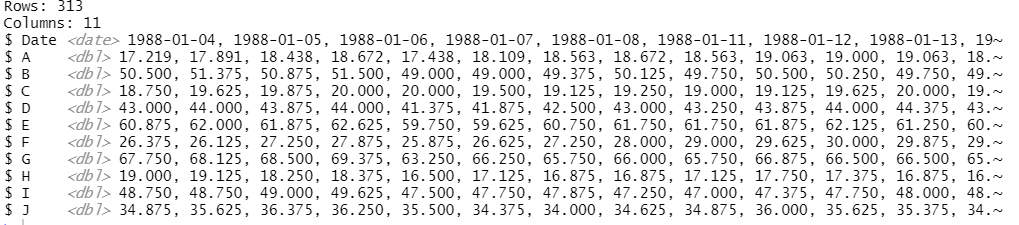
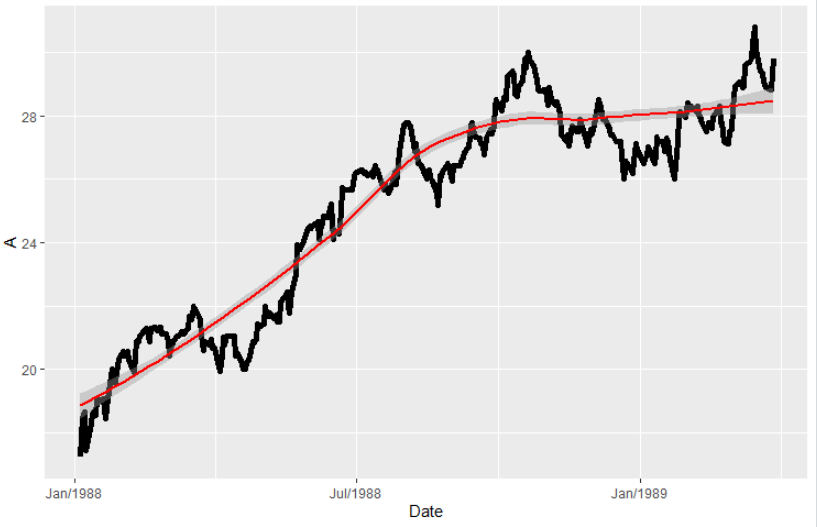
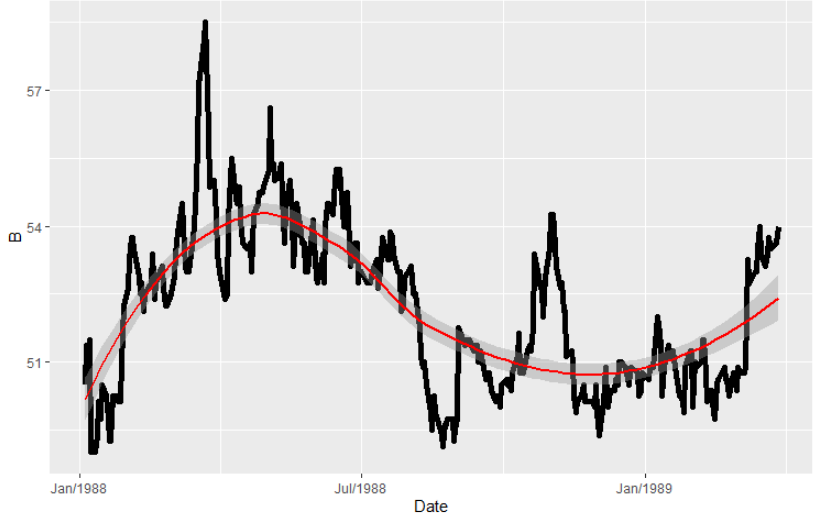
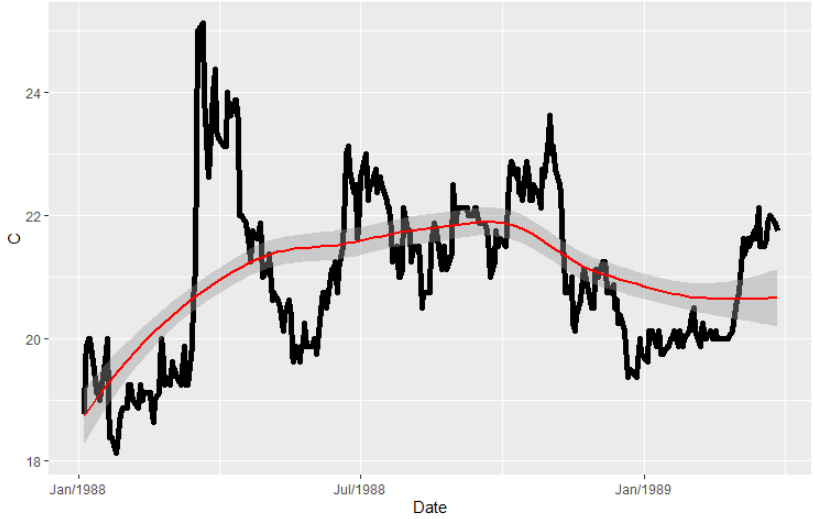
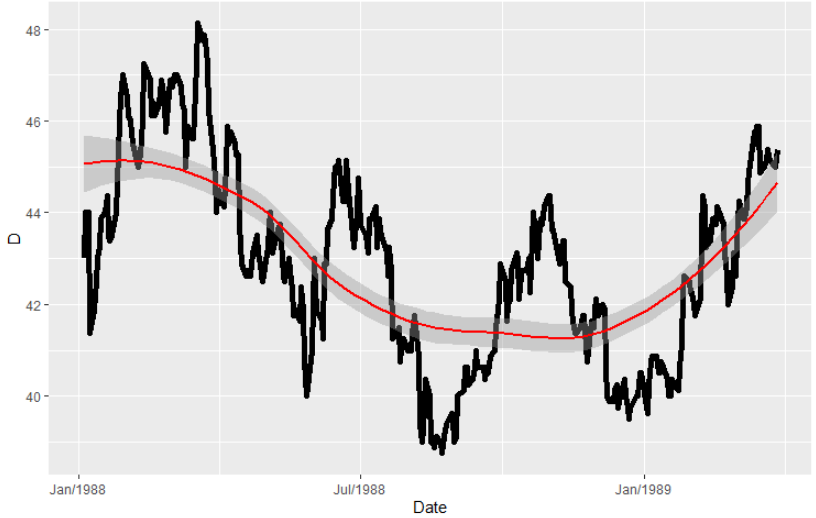
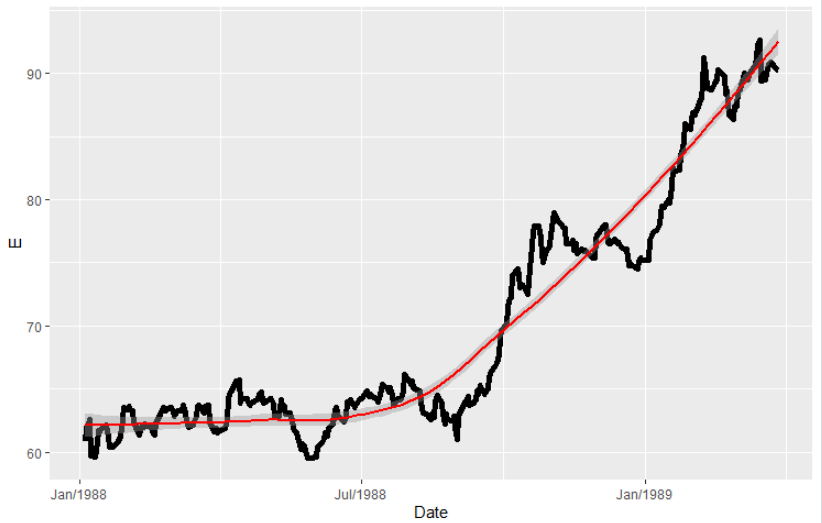
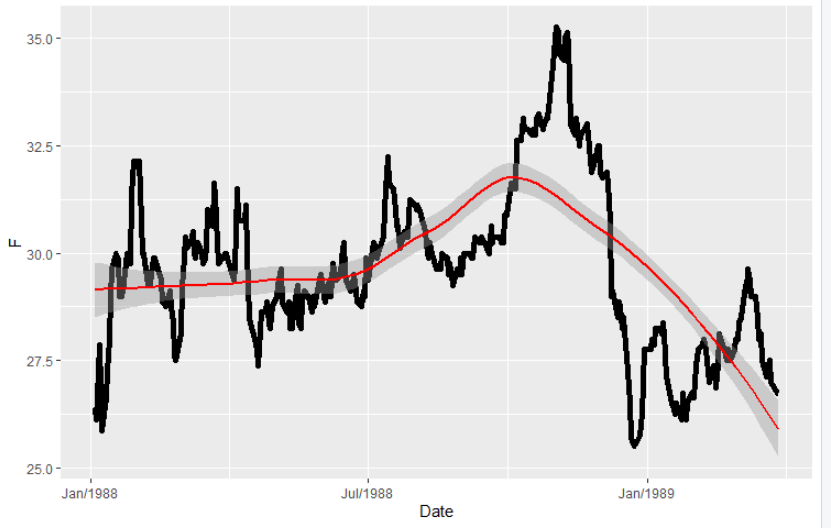
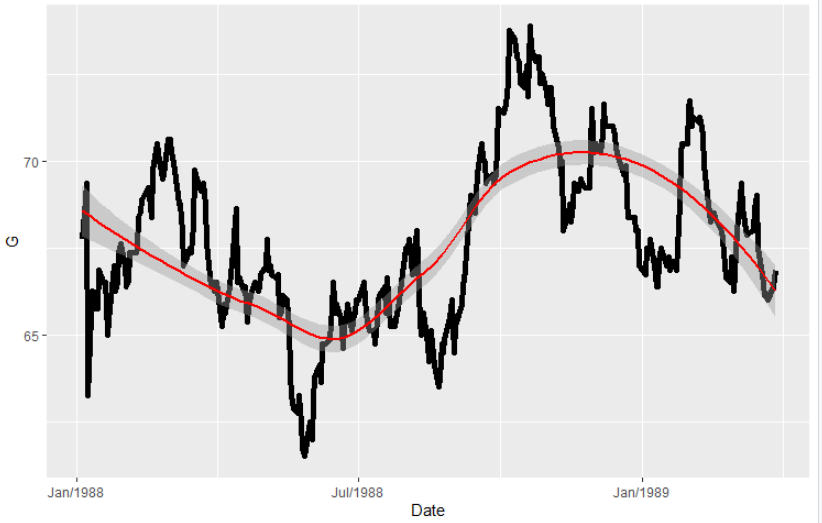
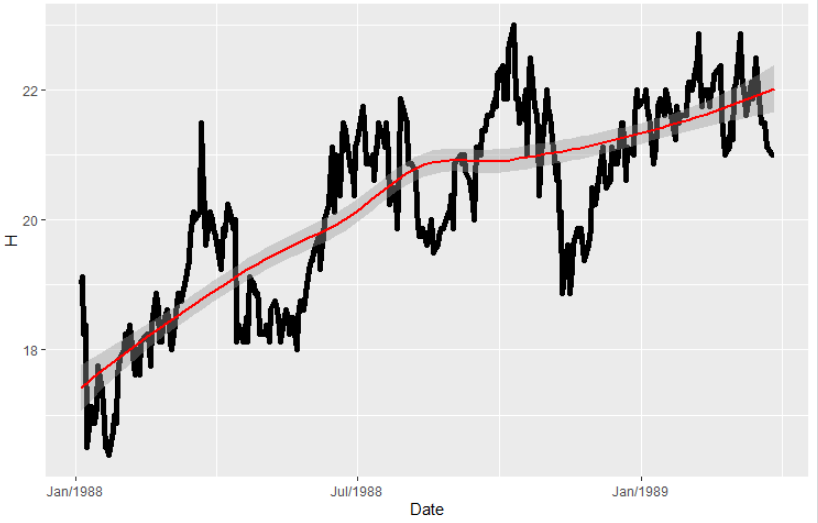
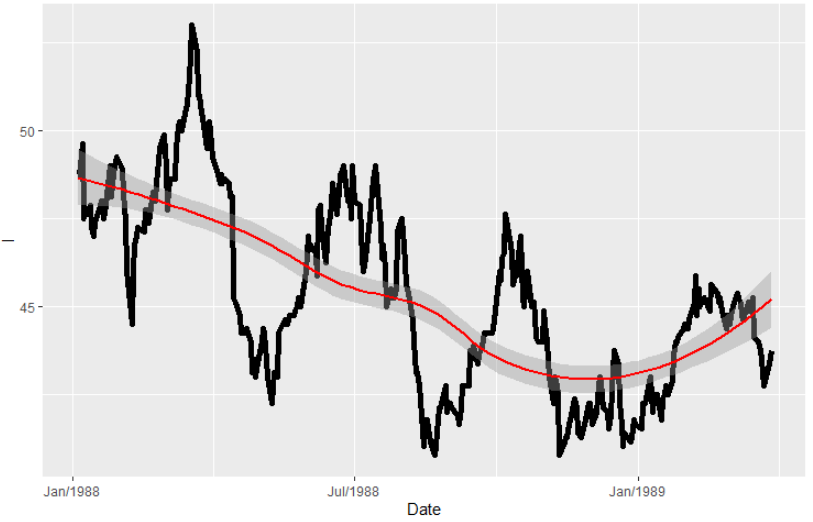
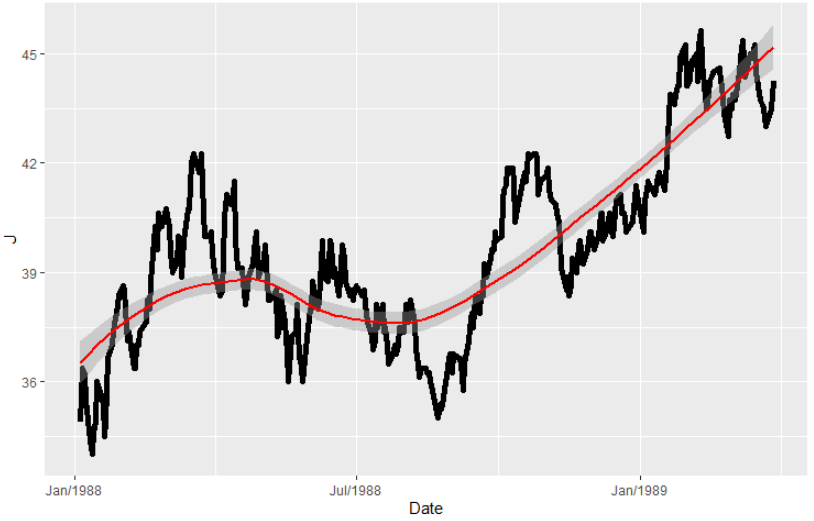
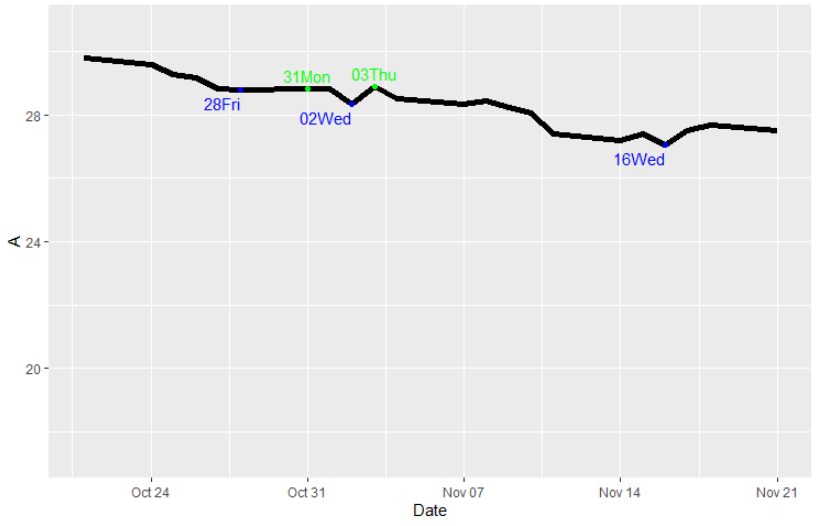
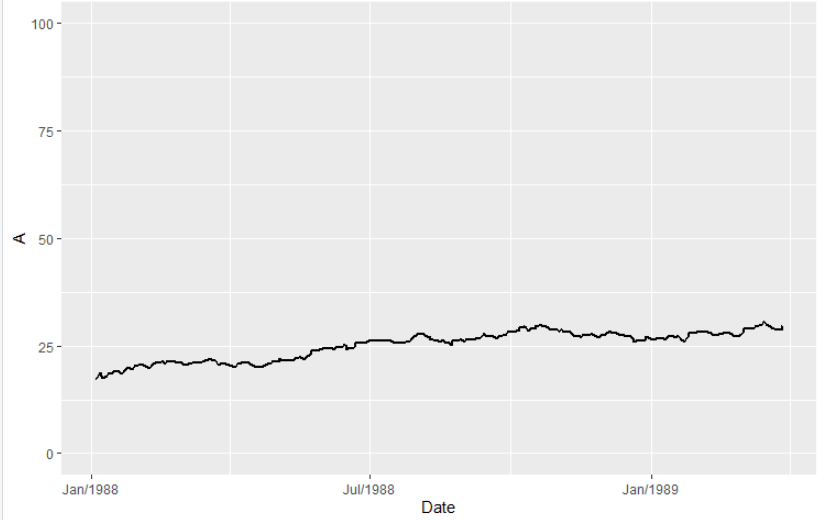
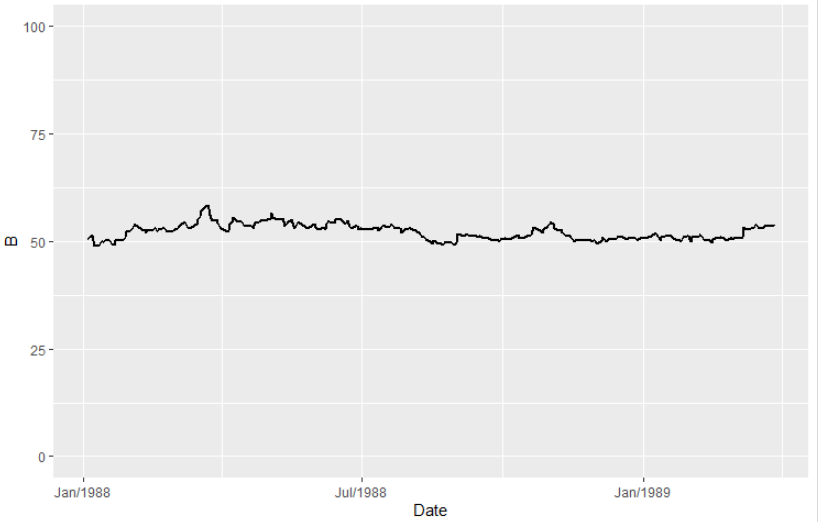
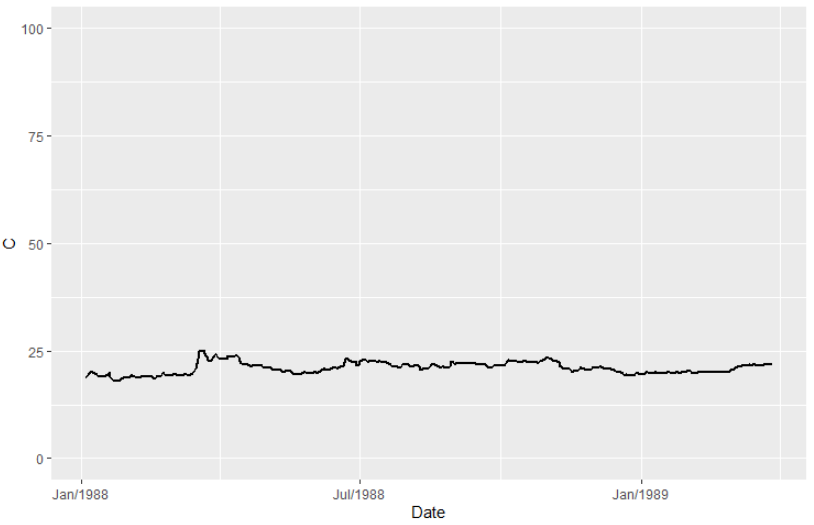
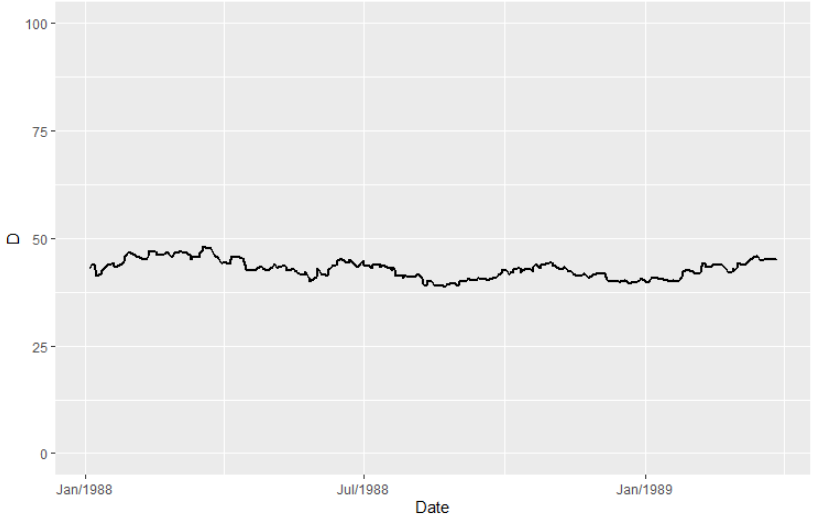
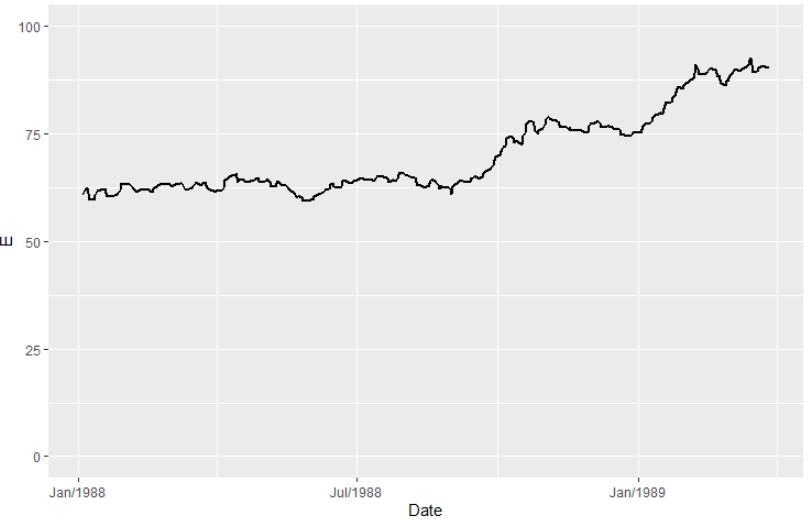
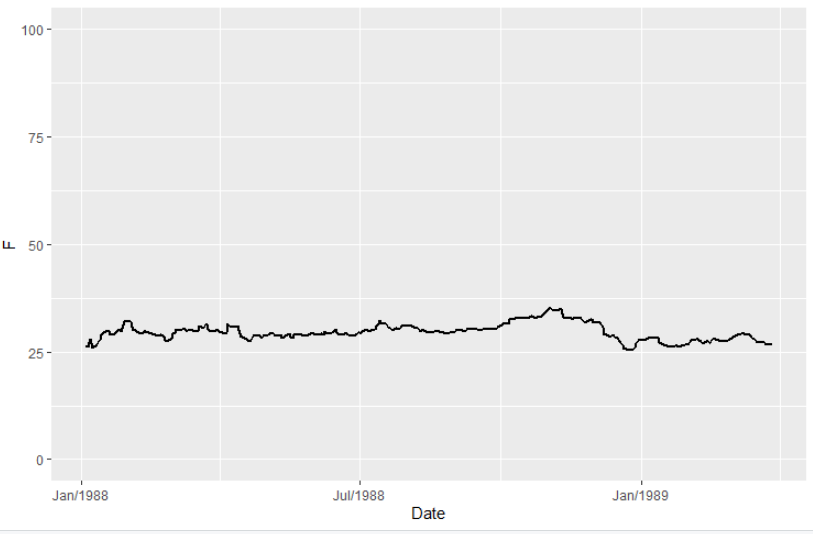
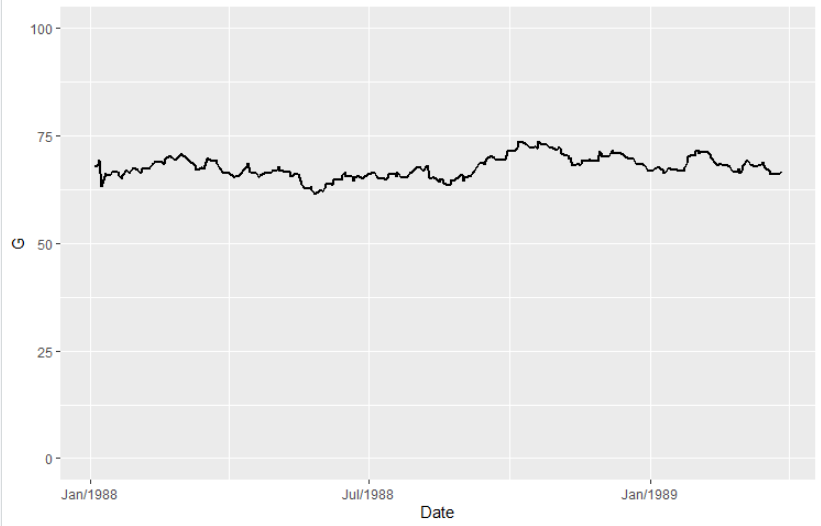
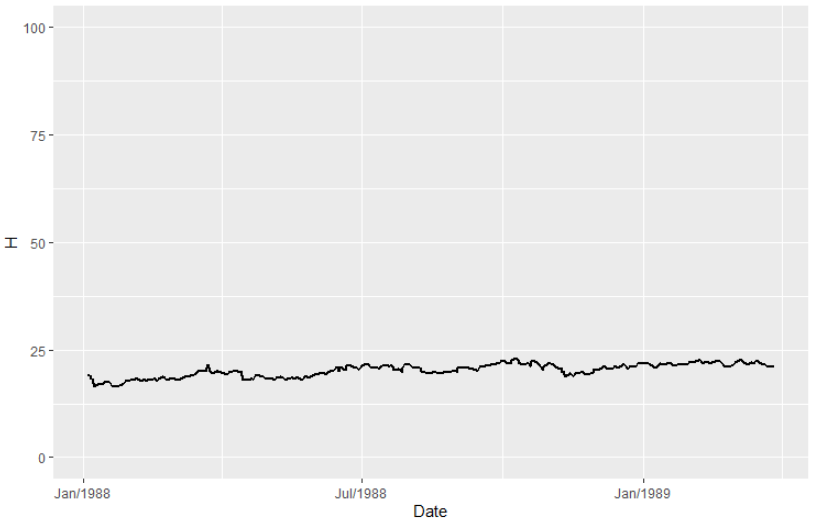
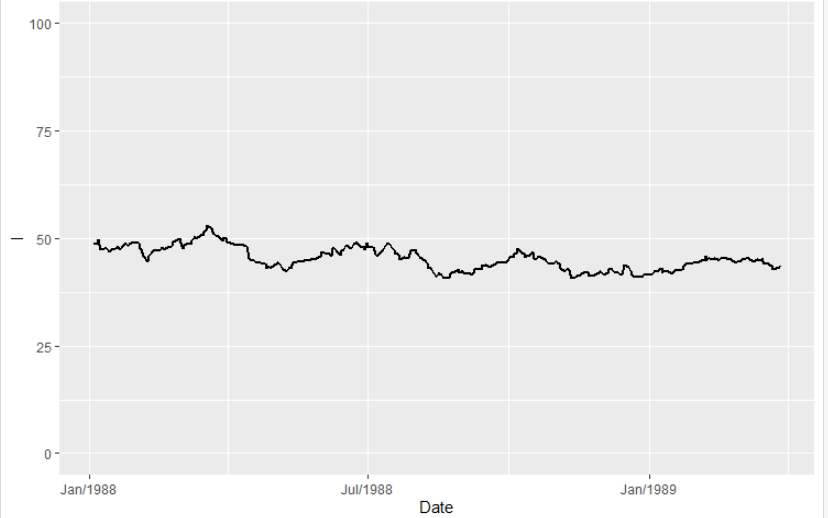
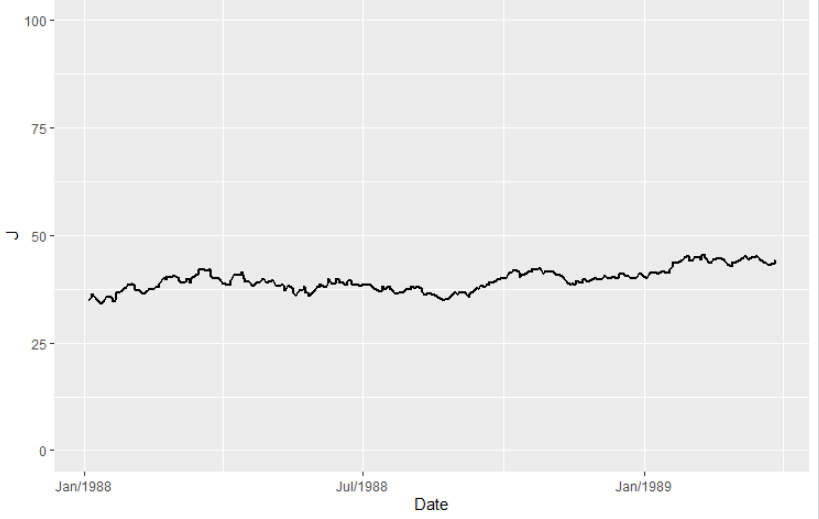
ggplot(data,aes(x=Date,y=J))+

geom\_line(color='black',size=1)+

scale\_x\_date(date\_labels="%b/%Y")+

ylim(0,100)

**OUTPUT:**

3.Calculate the mean stock prices and plot on histogram

**CODE:**

#calculating and printing the mean of stock prices of all companies one-by-one

mean(data$A,na.rm=T)

mean(data$B,na.rm=T)

mean(data$C,na.rm=T)

mean(data$D,na.rm=T)

mean(data$E,na.rm=T)

mean(data$F,na.rm=T)

mean(data$G,na.rm=T)

mean(data$H,na.rm=T)

mean(data$I,na.rm=T)

mean(data$J,na.rm=T)

#printing the histogram of stock prices of all companies together along with the mean stock price of each company

ggplot()+

geom\_histogram(aes(x=data$A,fill='A'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$A,na.rm=T)),color="salmon",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$B,fill='B'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$B,na.rm=T)),color="navy",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$C,fill='C'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$C,na.rm=T)),color="deepskyblue",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$D,fill='D'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$D,na.rm=T)),color="gold",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$E,fill='E'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$E,na.rm=T)),color="deeppink",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$F,fill='F'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$F,na.rm=T)),color="blue",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$G,fill='G'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$G,na.rm=T)),color="orange",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$H,fill='H'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$H,na.rm=T)),color="red",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$I,fill='I'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$I,na.rm=T)),color="green",linetype="dashed",size=1)+

geom\_histogram(aes(x=data$J,fill='J'),bins=30,alpha=0.5)+geom\_vline(aes(xintercept=mean(data$J,na.rm=T)),color="yellow",linetype="dashed",size=1)+

xlab("Stock prices of 10 companies(with each color representing a company)")+

ylab("Count")+

ggtitle("Distribution of stock prices of 10 companies")+

theme(plot.title=element\_text(hjust=0.5))+

scale\_fill\_manual(values=c('salmon','navy','deepskyblue','gold','deeppink','blue','orange','red','green','yellow'),name="Company")

#histogram of companies with highest and lowest mean

ggplot(data,aes(x=E))+ geom\_histogram(bins=80,color='black',fill='aliceblue')+ geom\_vline(aes(xintercept=mean(E,na.rm=T)),color="blue",linetype="dashed",size=1)

ggplot(data,aes(x=H))+ geom\_histogram(bins=80,color='black',fill='thistle2')+ geom\_vline(aes(xintercept=mean(H,na.rm=T)),color="blue",linetype="dashed",size=1)

**OUTPUT:**

