评分：_____

SHANGHAI UNIVERSITY

COURSE PAPER

Homework 1

Information	student1	student2
Academy	Management	Management
Specialty	Management Science	Information Science
Student ID	18121216	19121255
Name	Lulu Shi	Jingwen Zhang
Course	Machine Learning in Business	
Teacher	Si Zhang **|** Liyang Xiao	

Listed are the **Answers** to the **Questions.**

1) **Define a new variable calling dat to import the data set and summarize the data.**

   \>setwd("D:/Desktop/AAA 上海大学文件/BBB 课程学习资料/B-2020-2021/CLASS 冬/ML_in_Business/work_1_ddl1221/homework 1")

   \>dat=read.csv('Stock_Data.csv')

   \>summary(dat)

   (the output of summary(dat) is in **H1_output_1.csv**)

   What are the highest AC price of MSFT and the average trading volume of CitiGroup Inc.?

   the highest AC price of MSFT **:** 51.610

   the average trading volume of CitiGroup Inc.: 8678513

Reminder: After importing, you need to use the command dat$Date=as.Date(dat$Date,

format='%d/%m/%Y') to formate the date variable.

   \>dat$Date=as.Date(dat$Date,format='%d/%m/%Y')

2) **Plot the historical AC prices of Ford. (You should have date on the x-axis.)**

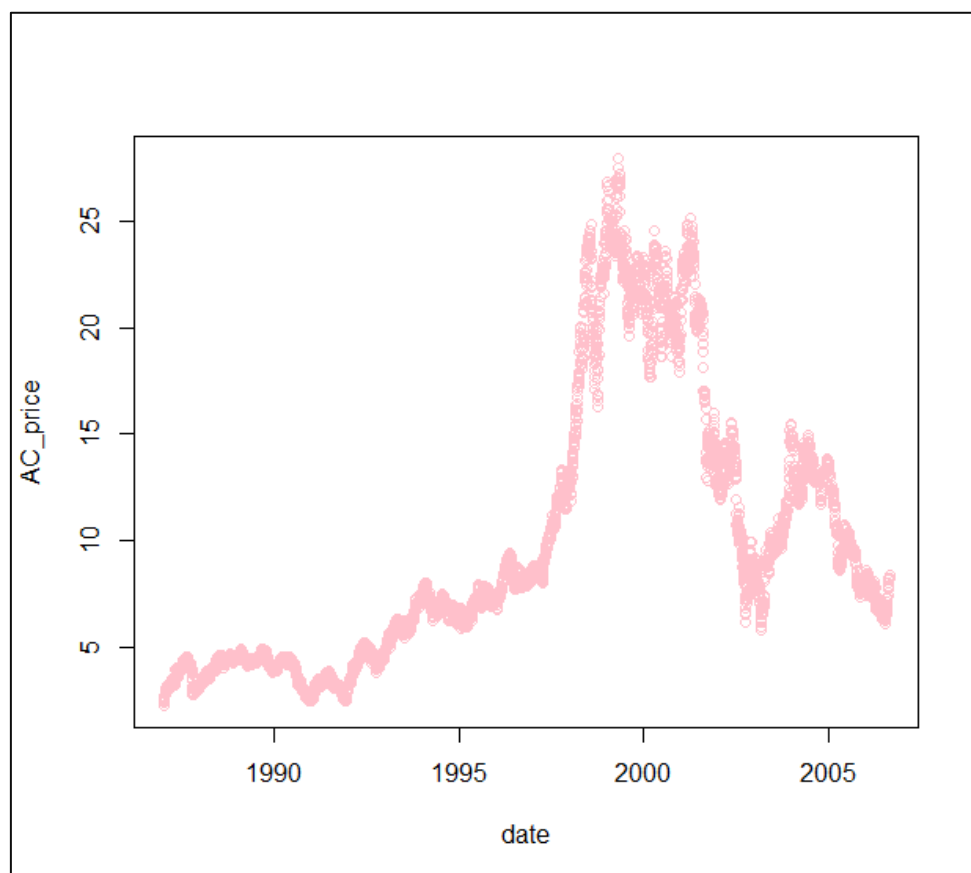   \>plot(dat$Date,dat$F_AC, xlab='date',ylab='AC_price',col='pink')

FIGURE 1.1

Meanwhile, to beautify the figure, we used ggplot2 package to draw it :

```
>library(ggplot2)
> ggplot(dat,aes(x=Date, y=F_AC)) +
  geom_point(alpha=0.1,size=1.0,shape=21,fill='white',stroke=2,color='red')+
  theme(axis.title.x =element_text(size=13), axis.title.y=element_text(size=13))
```
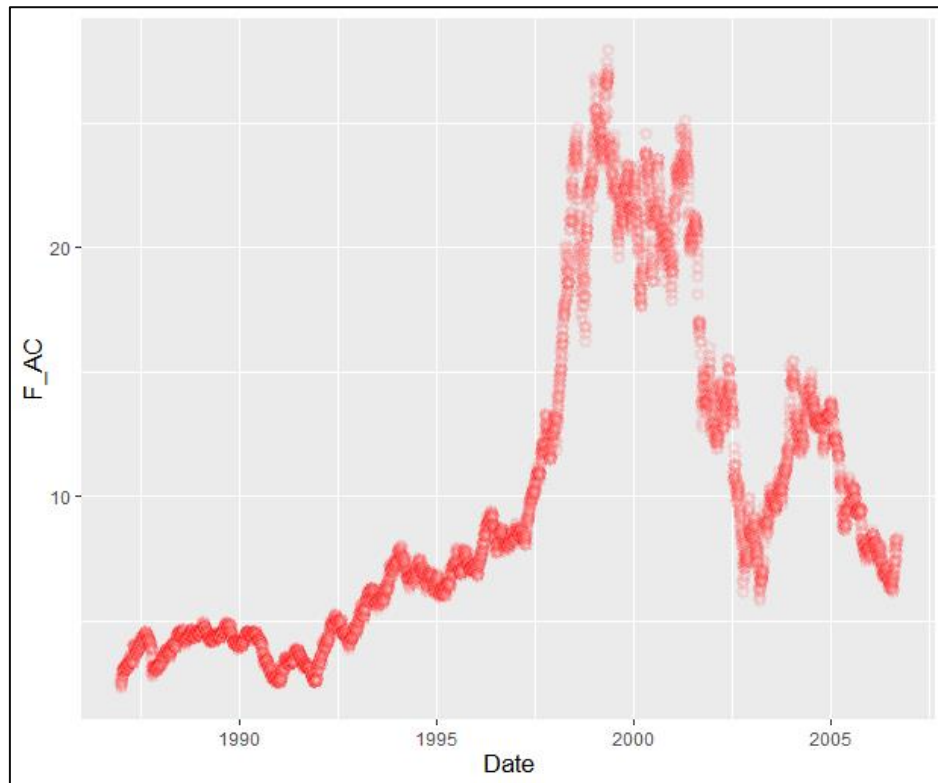

FIGURE 1.2

3) Define a variable calling prePriceF to store AC prices of Ford from 02/01/1987 to 31/08/2006 and a variable calling curPriceF to store AC prices of Ford from 05/01/1987 to 01/09/2006.

```
>start1= which(dat$Date=='1987-01-02')
>end1= which(dat$Date=='2006-08-31')
>prePriceF=dat$F_AC[start1:end1]
>start2= which(dat$Date=='1987-01-05')
>end2= which(dat$Date=='2006-09-01')
>curPriceF=dat$F_AC[start2:end2]
```

4) Calculate daily returns of Ford from 05/01/1987 to 01/09/2006 and make a histogram of the returns. What does that histogram look like?

```
>Return=NULL
>for(i in 1:4962){
  Return=c(Return,curPriceF[i]/prePriceF[i]-1)
}
>hist(Return,main='Return')
```
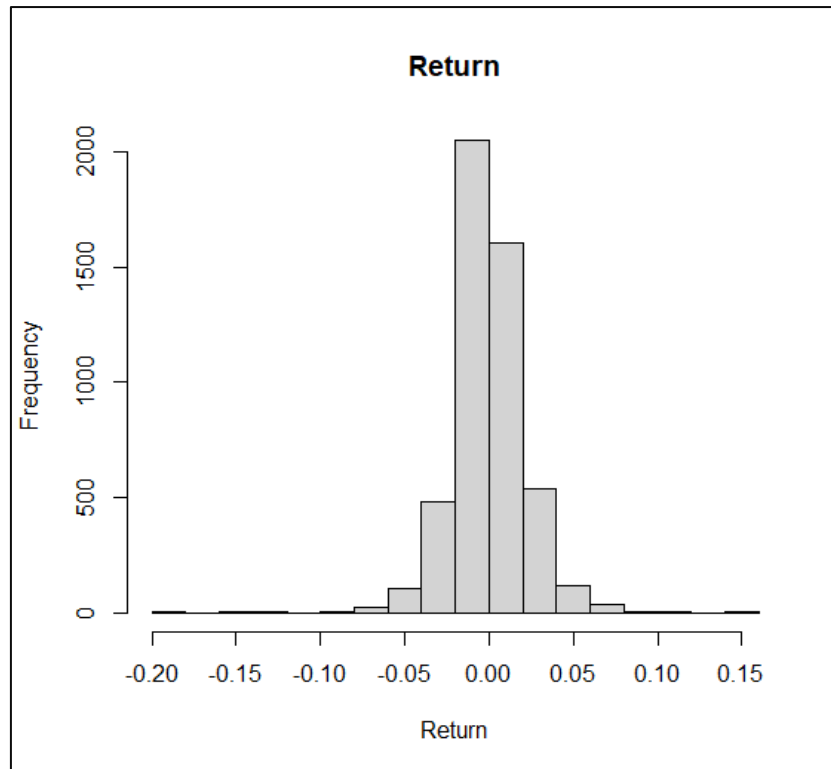
FIGURE 1.3

Meanwhile, to beautify the figure, we used ggplot2 package to draw it :

```
>library(ggplot2)
>return = Return
>r = data.frame(return)
>ggplot(r,aes(x=Return))+
  geom_histogram(fill='red',color='black',alpha=0.3)+
  theme(axis.title.x =element_text(size=18), axis.title.y=element_text(size=18))
```
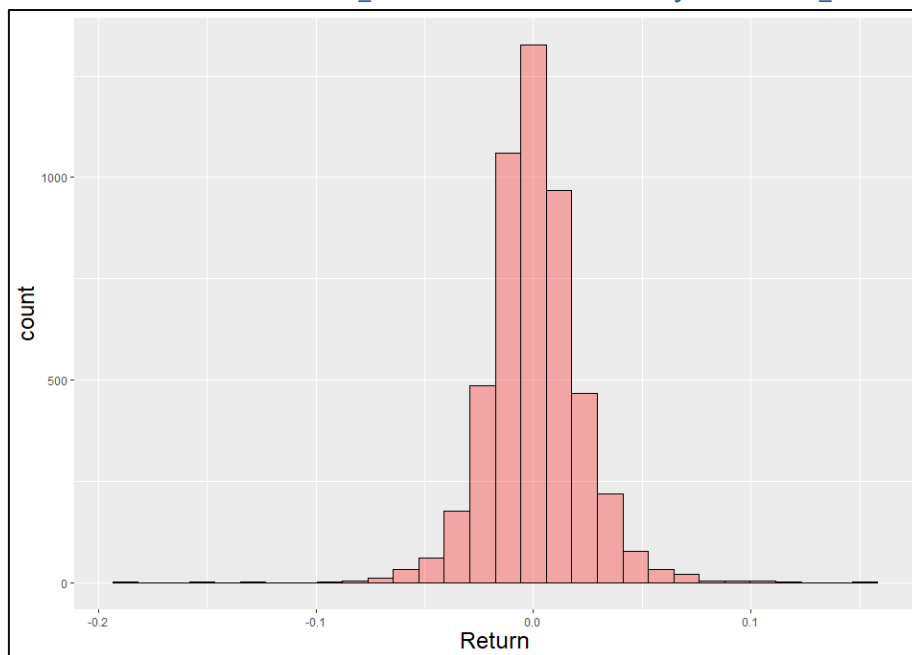

FIGURE 1.4

Code Parts   (source code : ML_h1.R)

```r
>setwd("D:/ML_in_Business/homework 1")
# declare the work path.
>dat=read.csv('Stock_Data.csv')
# read the stock data to R.
>summary(dat)
# summarize the dat to get statistics result.
>dat$Date=as.Date(dat$Date,format='%d/%m/%Y')
# to format the Date column.
>plot(dat$Date,dat$F_AC, xlab='date',ylab='AC_price',col='pink')
# draw the figure of scatter points.
>start1= which(dat$Date=='1987-01-02')
# set the start date of prePriceF.
>end1= which(dat$Date=='2006-08-31')
# set the end date of prePriceF.
>prePriceF=dat$F_AC[start1:end1]
# index the previous F_AC column from start to end.
>start2= which(dat$Date=='1987-01-05')
# set the start date of curPriceF.
>end2= which(dat$Date=='2006-09-01')
# set the end date of curPriceF.
>curPriceF=dat$F_AC[start2:end2]
# index the current F_AC column sfrom start to end.
>Return=NULL
>for(i in 1:4962){
   Return=c(Return,curPriceF[i]/prePriceF[i]-1)
}
# use Return function to calculate the returns of F_AC.
>hist(Return,main='Return')
# draw the figure to histogram.
>library(ggplot2)
>r = data.frame(Return)
# to transform numeric Return to a dataframe , which is the must in ggplot().
>ggplot(r,aes(x=Return))+
   geom_histogram(fill='red',color='black',alpha=0.3)+
   theme(axis.title.x =element_text(size=18), axis.title.y=element_text(size=18))
# use ggplot2 to draw the figure, the principle is the composition of layers.
```