TSX Trading Simulation

Tan Lin

2023-04-20

```
#Installing Packages and Loading Libraries
#install.packages("ggplot2", repos="https://mirror.csclub.uwaterloo.ca/CRAN/")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.2.3

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

Load Data

```
TSX = read.csv("TSX_Data.csv")
GDP = read.csv("GDP_Data.csv")
```

Data Preparation

Data Types

```
#Converting Text into Dates and Numbers
TSX[,"Date"] = as.Date(TSX[,"Date"])
TSX[,2:7] = lapply(TSX[,2:7],as.numeric)

GDP$Date = as.Date(GDP$Date)
GDP$GDP = as.numeric(GDP$GDP)


#Check If Converted
class(TSX[1,1])
```

```
## [1] "Date"
```

```r
class(TSX[1,2])
```

```
## [1] "numeric"
```

```r
class(TSX[1,3])
```

```
## [1] "numeric"
```

```r
class(TSX[1,4])
```

```
## [1] "numeric"
```

```r
class(TSX[1,5])
```

```
## [1] "numeric"
```

```r
class(TSX[1,6])
```

```
## [1] "numeric"
```

```r
class(TSX[1,7])
```

```
## [1] "numeric"
```

NA values

```r
# Check for NA Values
apply(sapply(TSX,is.na),2,sum)
```

```
##      Date      Open      High       Low     Close Adj.Close    Volume
##         0        45        45        45        45        45        45
```

```r
# Removing records with NA
TSX = na.omit(TSX)
```

```r
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
```

```
TSX = left_join(TSX,GDP,by="Date")
TSX$GDP[1] = GDP$GDP[1]
for (i in 1:length(TSX$Date))
  {
  if(is.na(TSX$GDP[i])){
    TSX$GDP[i] = TSX$GDP[i-1]
  }
}
```

Train and Test Data

```
# Selecting Dates before 2020 as Training Data
TSX_train = TSX[TSX[,"Date"] < "2020-01-01",]
TSX_train = TSX_train[TSX_train[,"Date"] >= "2000-01-01",]

# Selecting the Dates after 2020 as Testing Data
TSX_test = TSX[TSX[,"Date"] >= "2020-01-01",]
```

#Working on TSX Train

```
dataset = TSX[,c("Date","Close","GDP")]
size = length(dataset$Date)
dataset$diff_p = 0
dataset$GDP_diff_last[1] = 0
for(i in 1:size){
  #Calculate Change and Percent Change
  if(i > 1)
  {
    dataset[i,"diff"] = dataset[i,"Close"] - dataset[i-1,"Close"]
    dataset[i,"diff_p"] = dataset[i,"diff"] / dataset[i-1,"Close"]
    dataset$GDP_diff[i] = dataset$GDP[i] - dataset$GDP[i-1]
    if(dataset$GDP_diff[i] == 0){
      dataset$GDP_diff_last[i] = dataset$GDP_diff_last[i-1]
    }else{
      dataset$GDP_diff_last[i] = dataset$GDP_diff[i]
    }
  }
    #Calculate 7 day Variance
  if(i > 8){
    dataset[i,"mean7"] = mean(dataset[(i-7):i,"Close"])
    dataset[i,"var7"] = var(dataset[(i-7):i,"Close"])
    dataset[i,"high7"] = max(dataset[(i-7):i,"Close"])
    dataset[i,"low7"] = min(dataset[(i-7):i,"Close"])
    dataset[i,"trend7"] = mean(dataset[(i-7):i,"diff"])
  }
  #Calculate 30 day variance
  if(i > 31){
    dataset[i,"mean30"] = mean(dataset[(i-30):i,"Close"])
    dataset[i,"var30"] = var(dataset[(i-30):i,"Close"])
    dataset[i,"high30"] = max(dataset[(i-30):i,"Close"])
    dataset[i,"low30"] = min(dataset[(i-30):i,"Close"])
```

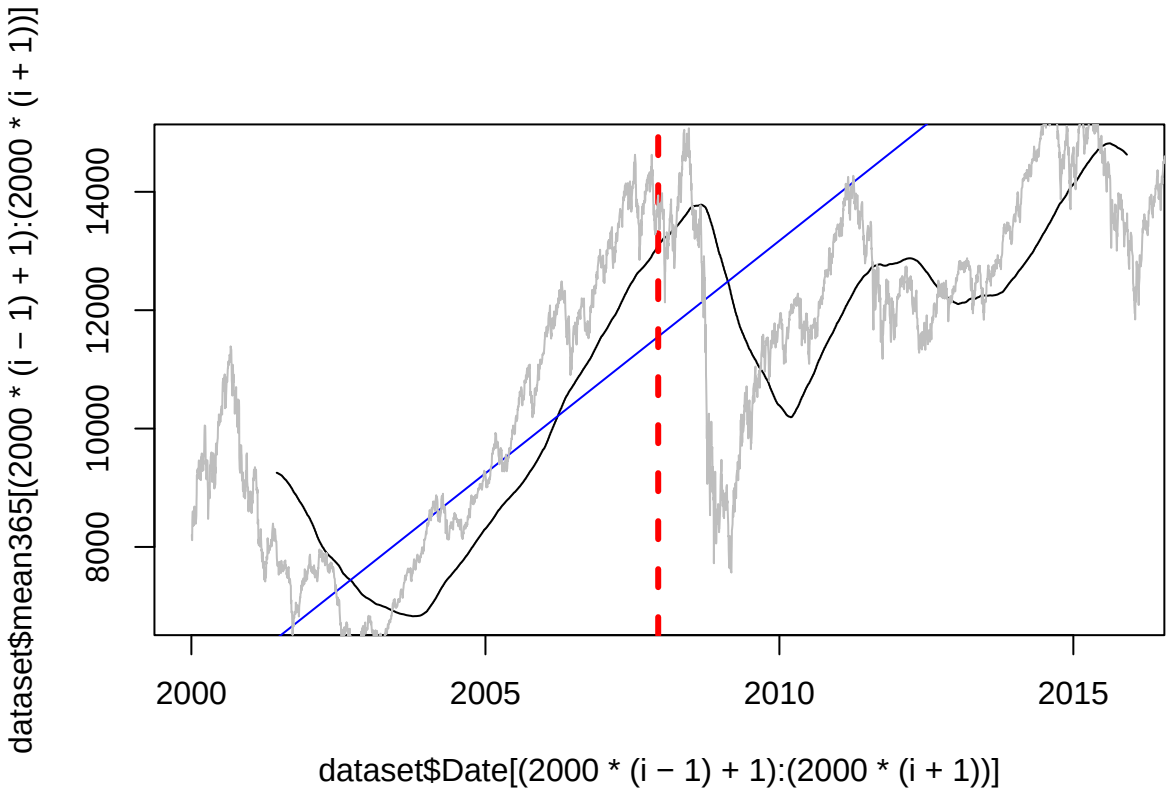
```
    dataset[i,"trend30"] = mean(dataset[(i-30):i,"diff"])
  }
    #Calculate 365 day variance
  if(i > 366){
    dataset[i,"mean365"] = mean(dataset[(i-365):i,"Close"])
    dataset[i,"var365"] = var(dataset[(i-365):i,"Close"])
    dataset[i,"high365"] = max(dataset[(i-365):i,"Close"])
    dataset[i,"low365"] = min(dataset[(i-365):i,"Close"])
    dataset[i,"trend365"] = mean(dataset[(i-365):i,"diff"])
  }
}
```

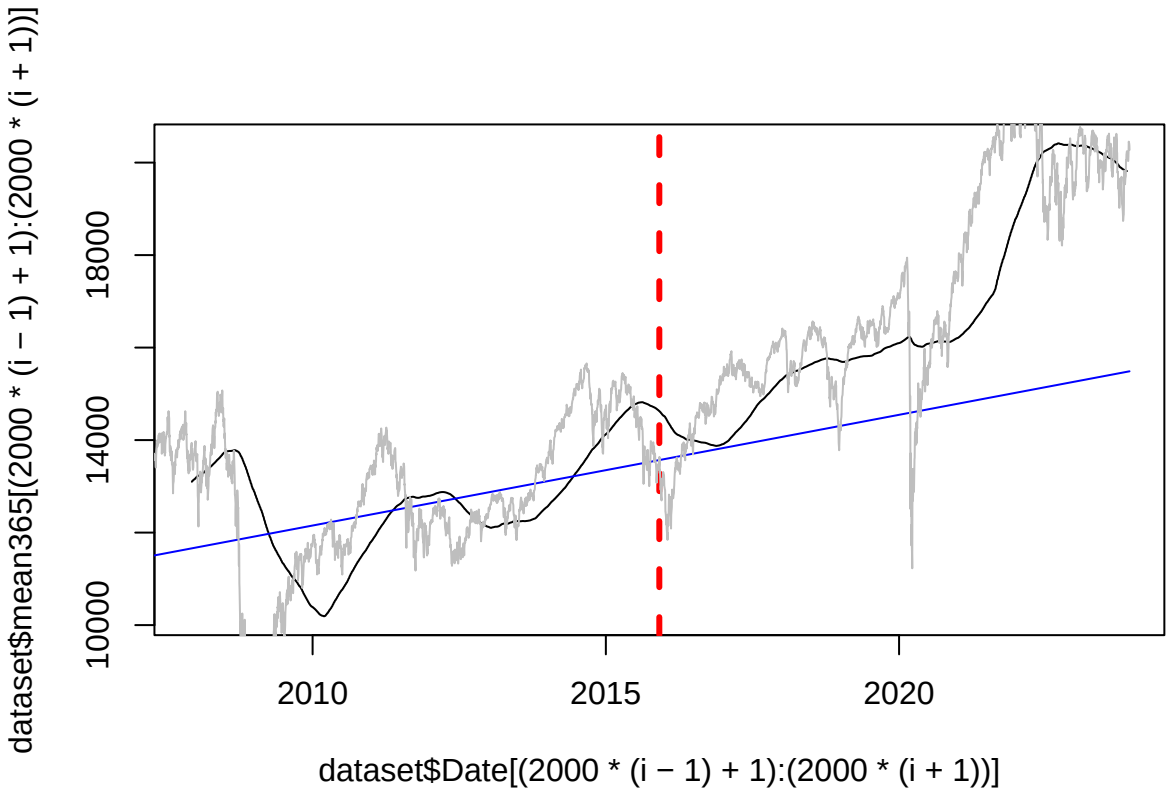
Examine Average Returns Base on Period of Variance

```
values = data.frame(qt=rep(0,20),var7=rep(0,20),var30=rep(0,20))
for(i in 0:19)
{
  qt = 0.05*i
  values[i+1,"qt"] = qt
  values[i+1,"var7"] = mean(dataset[dataset$var7 > quantile(dataset$var7, qt, na.rm=TRUE), "diff"], na.
  values[i+1,"var30"] = mean(dataset[dataset$var30 > quantile(dataset$var30, qt, na.rm=TRUE), "diff"], n
}

# Medium 30 Day Variance and Low 7 Day Variance
mean(dataset[dataset$var30[dataset$var30[dataset$var7 < quantile(dataset$var7, 0.25, na.rm=TRUE)] > quar
```

```
## [1] 2.069268
```

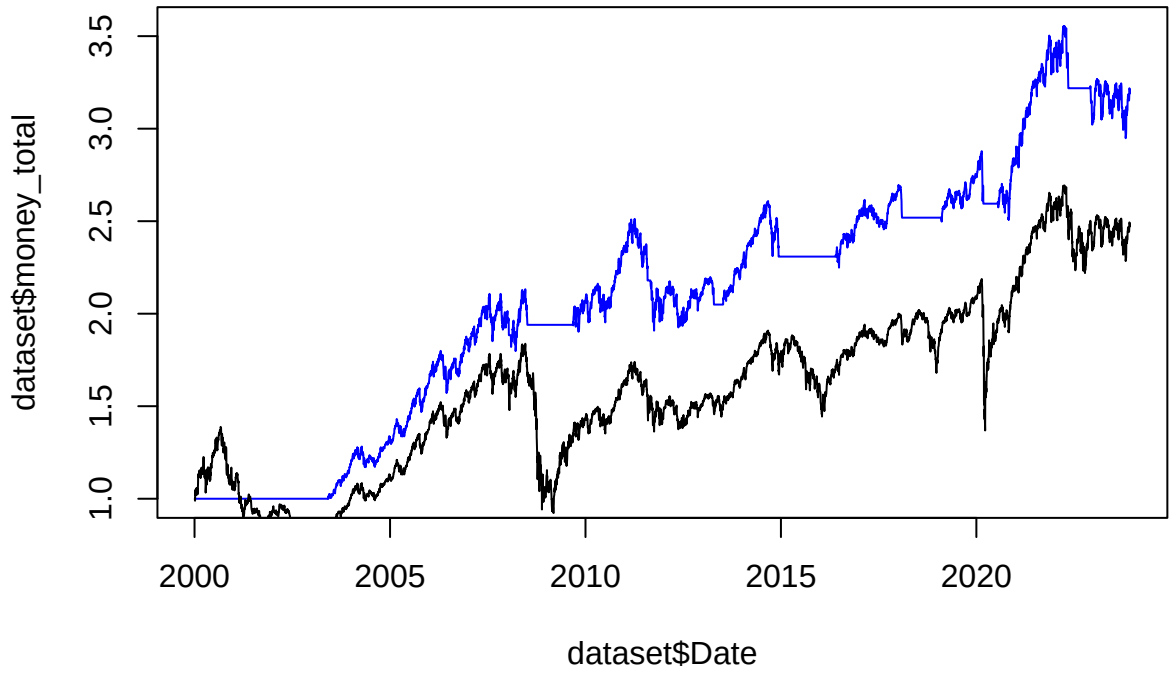
```
for(i in 1:2)
{
  fit = lm(mean365 ~ Date, data = dataset[(2000*(i-1)+1):(2000*i),])
  plot(dataset$Date[(2000*(i-1)+1):(2000*(i+1))], dataset$mean365[(2000*(i-1)+1):(2000*(i+1))],type="l"
  lines(dataset$Date, as.numeric(dataset$Date)*fit$coefficients[[2]]+fit$coefficients[[1]], col="blue")
  abline(v=dataset$Date[2000*i], col="red",lwd = 3, lty=2)
  lines(dataset$Date,dataset$Close,col="gray")
}
```


dataset$mean365[(2000 * (i − 1) + 1):(2000 * (i + 1))]
8000
10000
12000
14000
2000
2005
2010
2015
dataset$Date[(2000 * (i − 1) + 1):(2000 * (i + 1))]

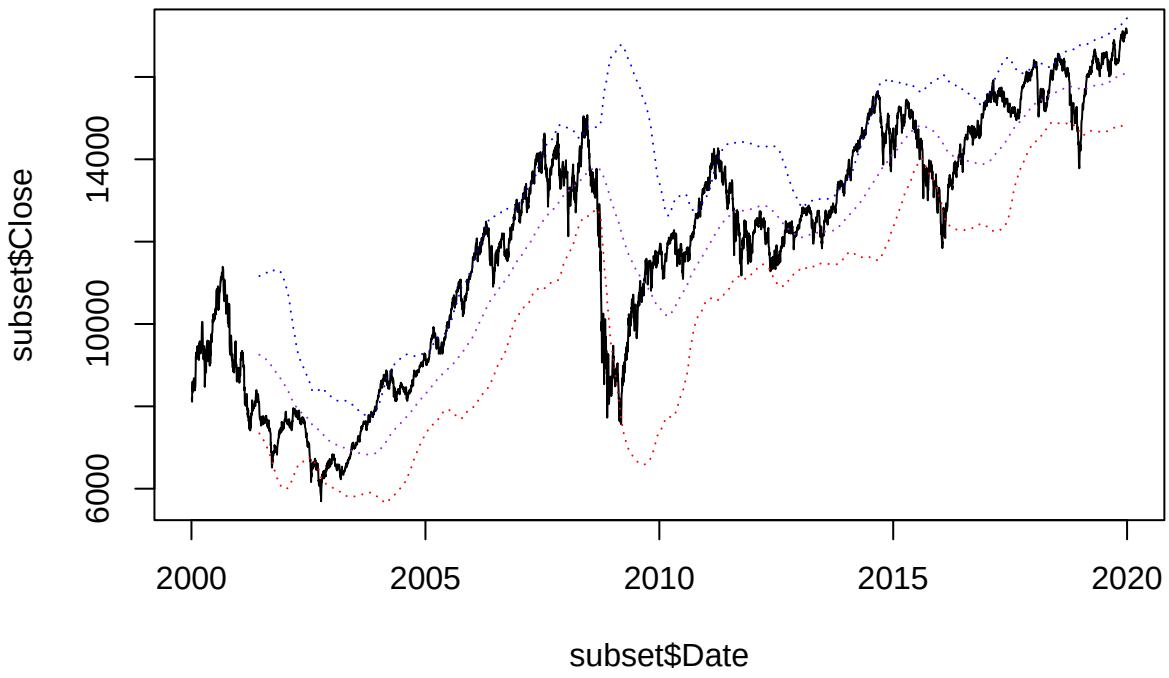



```
#Simulated Trading
dataset$money_out[1] = 1
dataset$money_in[1] = 0
dataset$money_total[1] = dataset$money_in[1] + dataset$money_out[1]

crossed = 0
for(i in 2:size){
  dataset$money_in[i] = dataset$money_in[i-1] * dataset$Close[i] / dataset$Close[i-1]
  dataset$money_out[i] = dataset$money_out[i-1]
  dataset$money_total[i] = dataset$money_in[i] + dataset$money_out[i]

  if(i > 367)
  {
    if(dataset$Close[i] > dataset$mean365[i] + sqrt(dataset$var365[i]))
    {
      crossed = 1;
    }
    if(dataset$Close[i] < dataset$mean365[i] - sqrt(dataset$var365[i]))
    {
      crossed = -1;
    }
    if(dataset$Close[i] < dataset$mean365[i] & crossed == 1){
      dataset$money_in[i] = 0
      dataset$money_out[i] = dataset$money_total[i]
      crossed = 0
    }
```

```
    else
    if(dataset$Close[i] > dataset$mean365[i] & crossed == -1){
      dataset$money_in[i] = dataset$money_total[i]
      dataset$money_out[i] = 0
      crossed = 0
    }
  }


}
plot(dataset$Date,dataset$money_total,type="l",col="blue")
lines(dataset$Date,dataset$Close/dataset$Close[1],col="black")
```

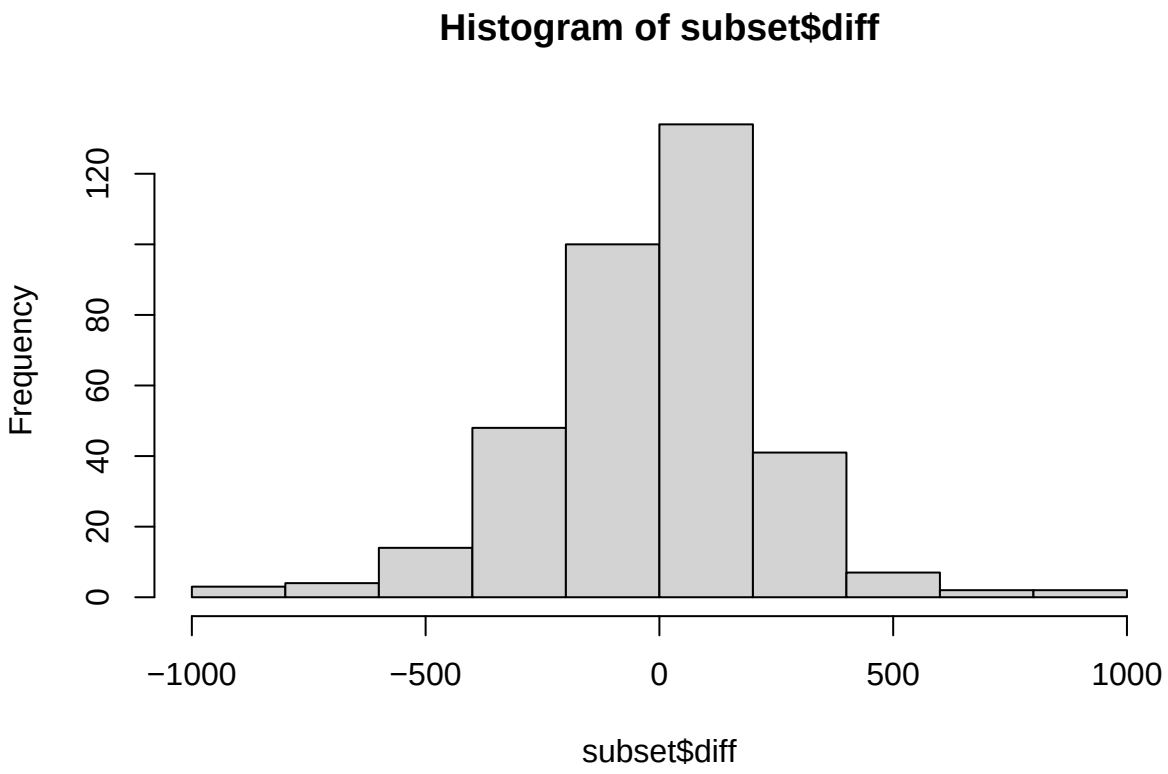



```
subset = dataset[dataset$Date > "2000-01-01" & dataset$Date < "2020-01-01",]
plot(subset$Date,subset$Close,type="l")
lines(subset$Date,subset$mean365,col="purple",lty = "dotted")
lines(subset$Date,subset$mean365-1.96*sqrt(subset$var365),col="red",lty = "dotted")
lines(subset$Date,subset$mean365+1.96*sqrt(subset$var365),col="blue",lty = "dotted")
```

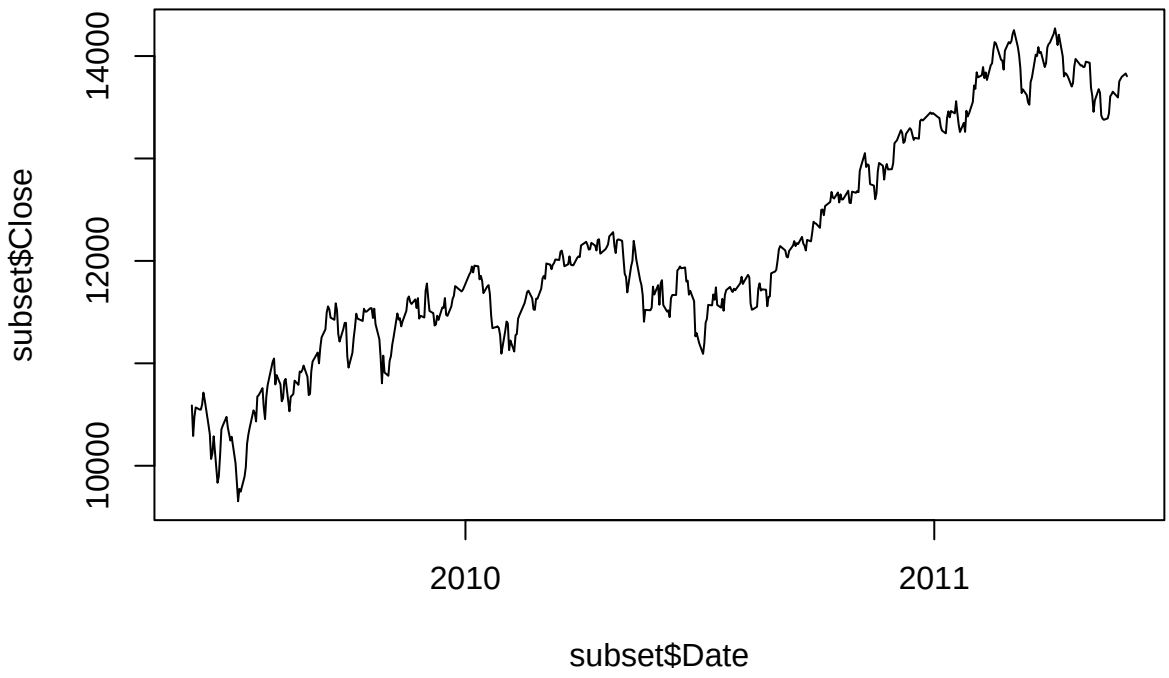



```
subset = dataset[dataset$Date > "2008-01-01" & dataset$Date < "2009-06-01",]
plot(subset$Date,subset$Close, type="l")
```

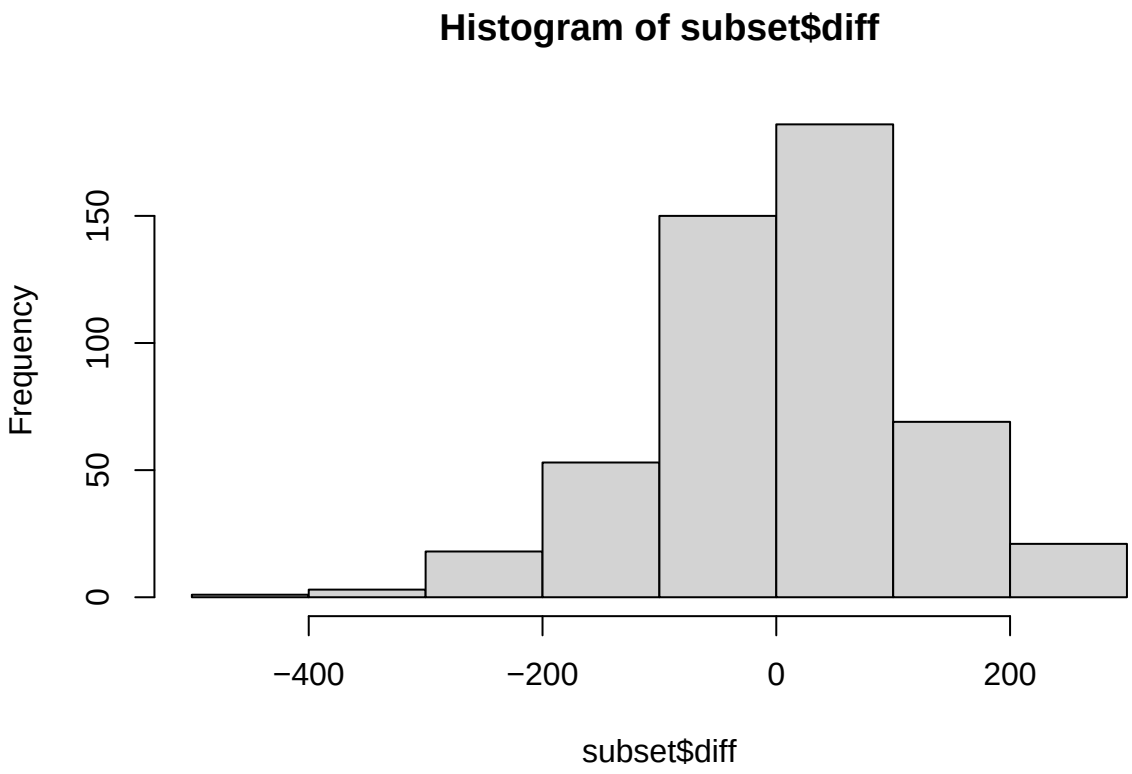

```
hist(subset$diff)
```



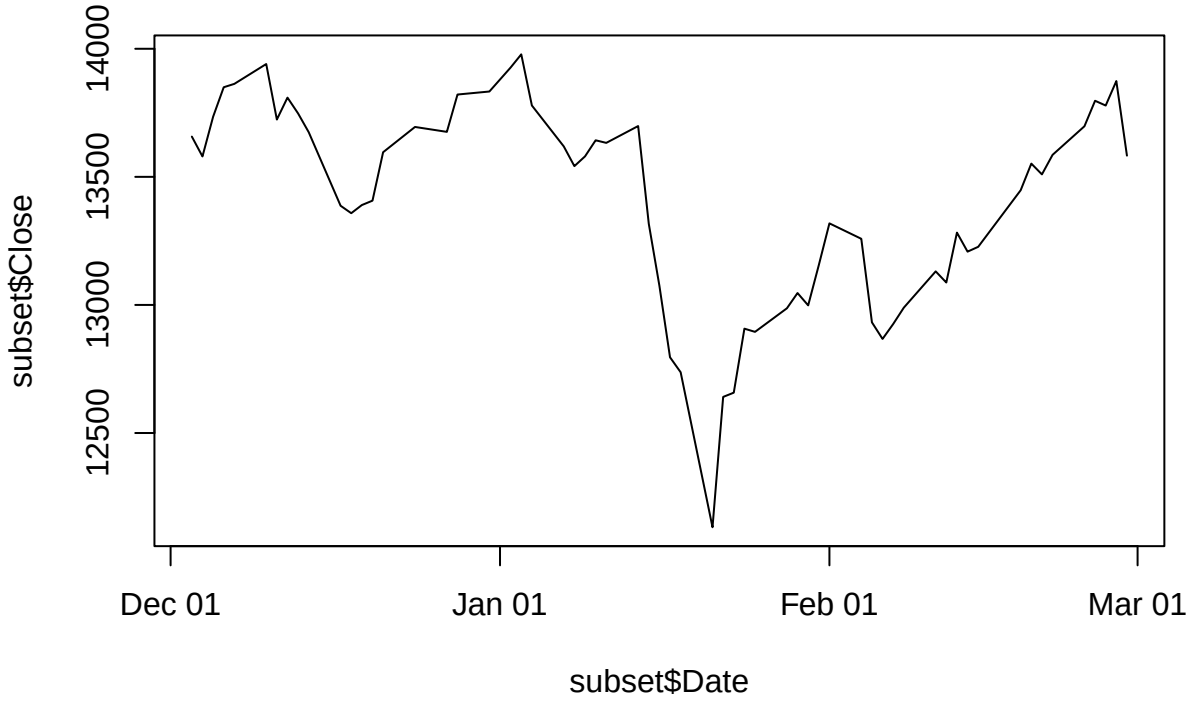

```
subset = dataset[dataset$Date > "2009-06-01" & dataset$Date < "2011-06-01",]
plot(subset$Date,subset$Close, type="l")
```



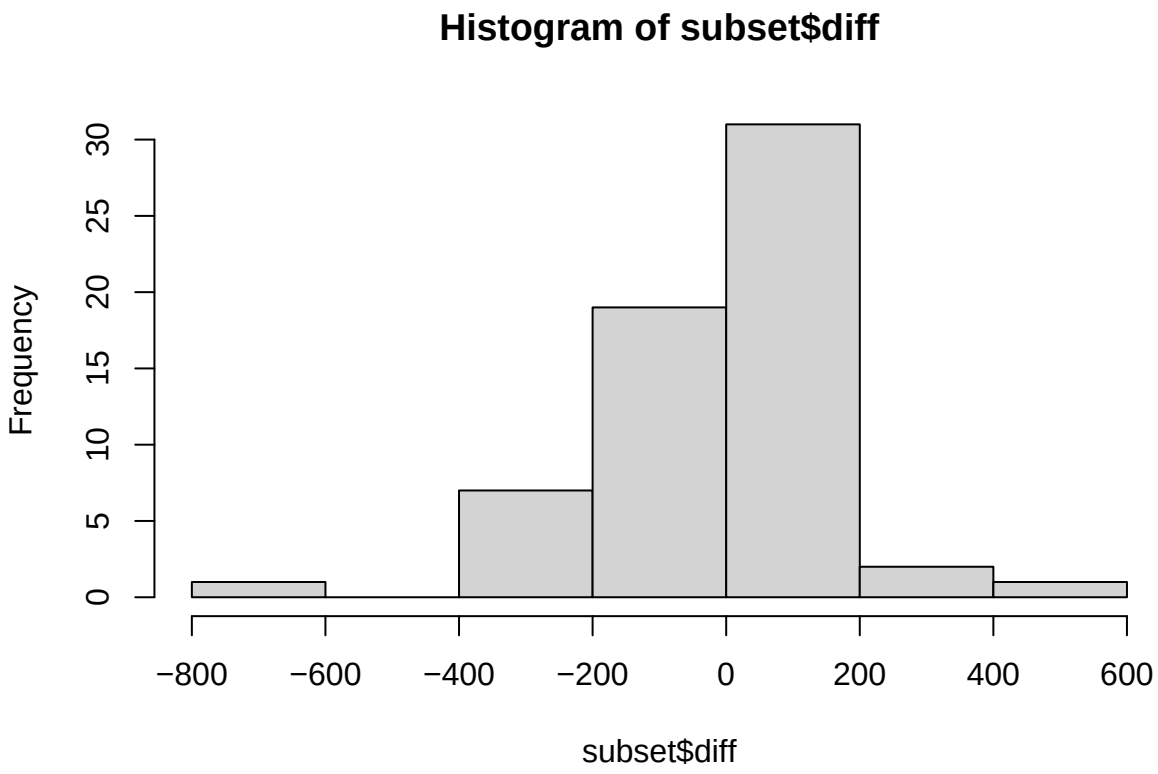

```
hist(subset$diff)
```



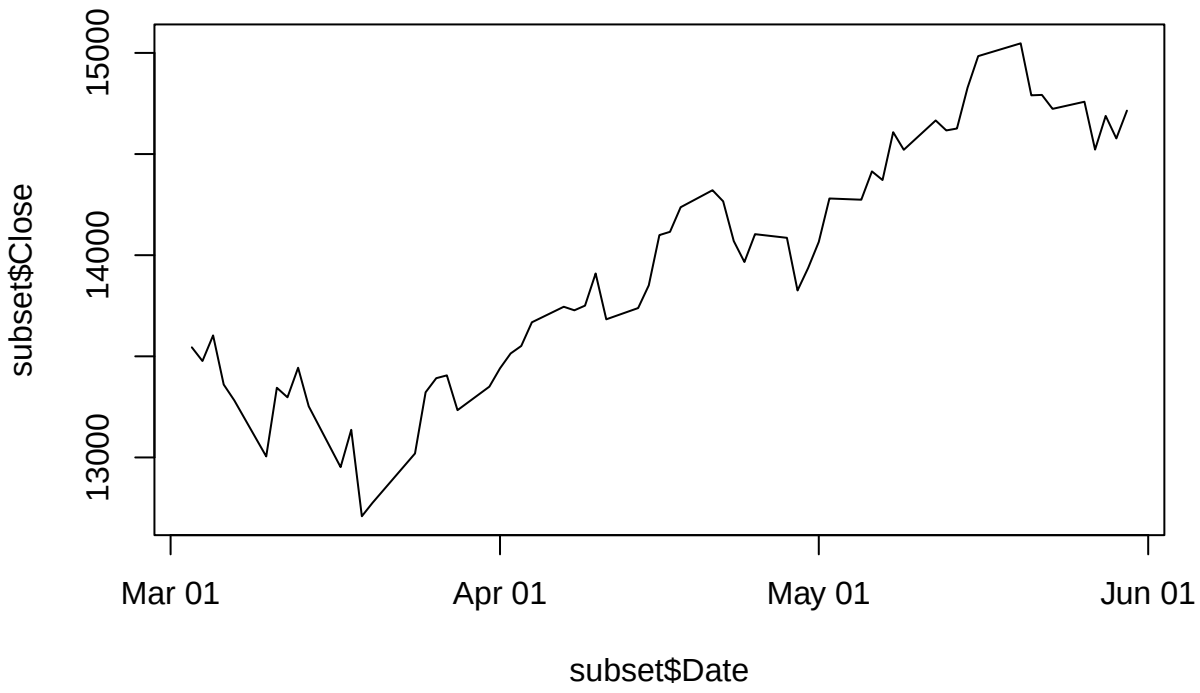

```
subset = dataset[dataset$Date > "2007-12-01" & dataset$Date < "2008-03-01",]
plot(subset$Date,subset$Close, type="l")
```



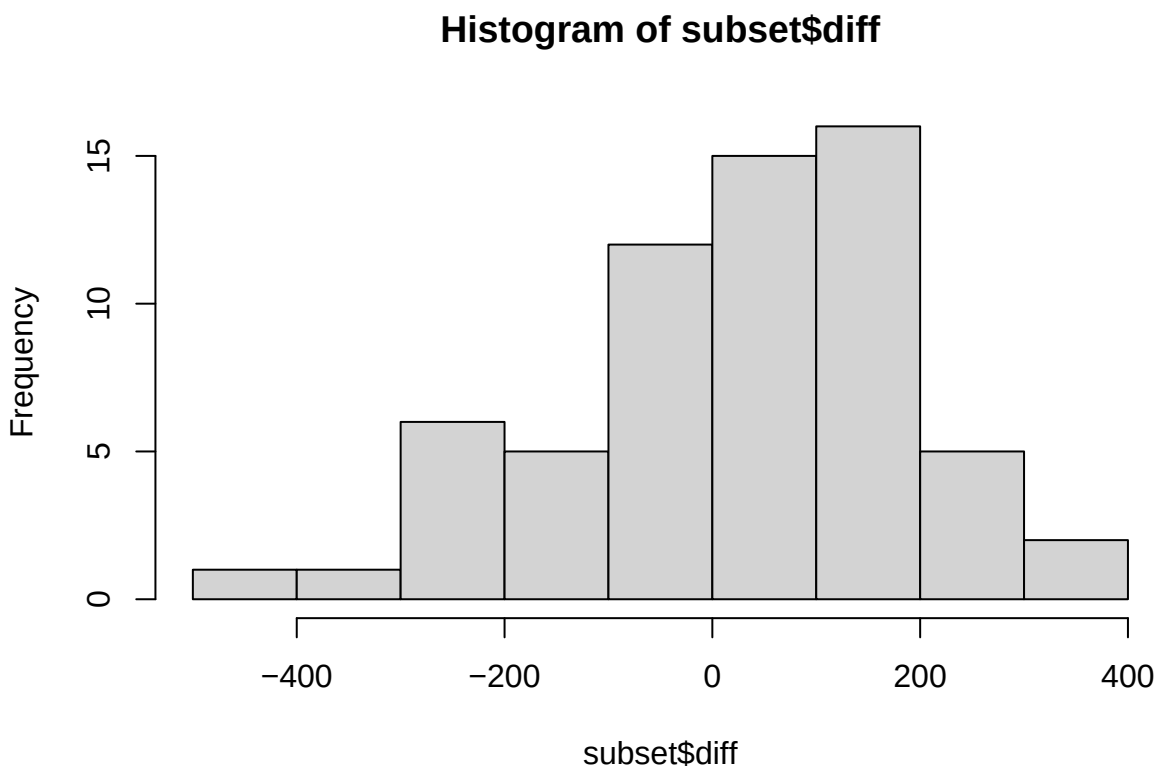

```
hist(subset$diff)
```



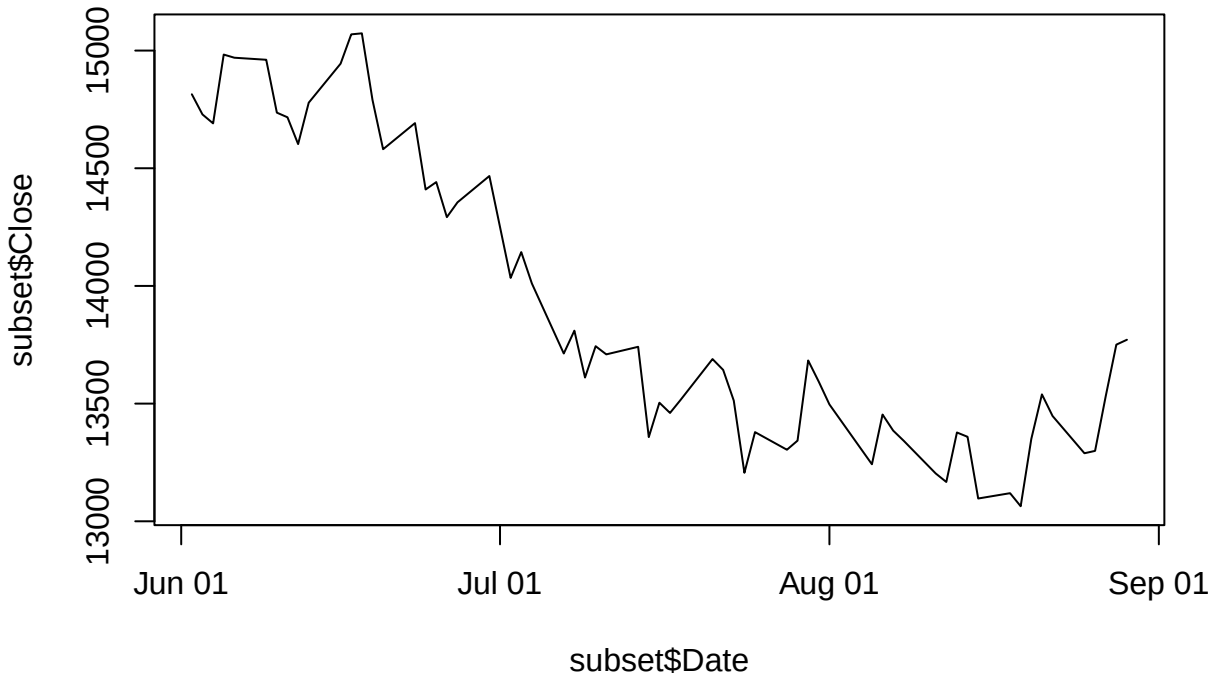

```
subset = dataset[dataset$Date > "2008-03-01" & dataset$Date < "2008-06-01",]
plot(subset$Date,subset$Close, type="l")
```



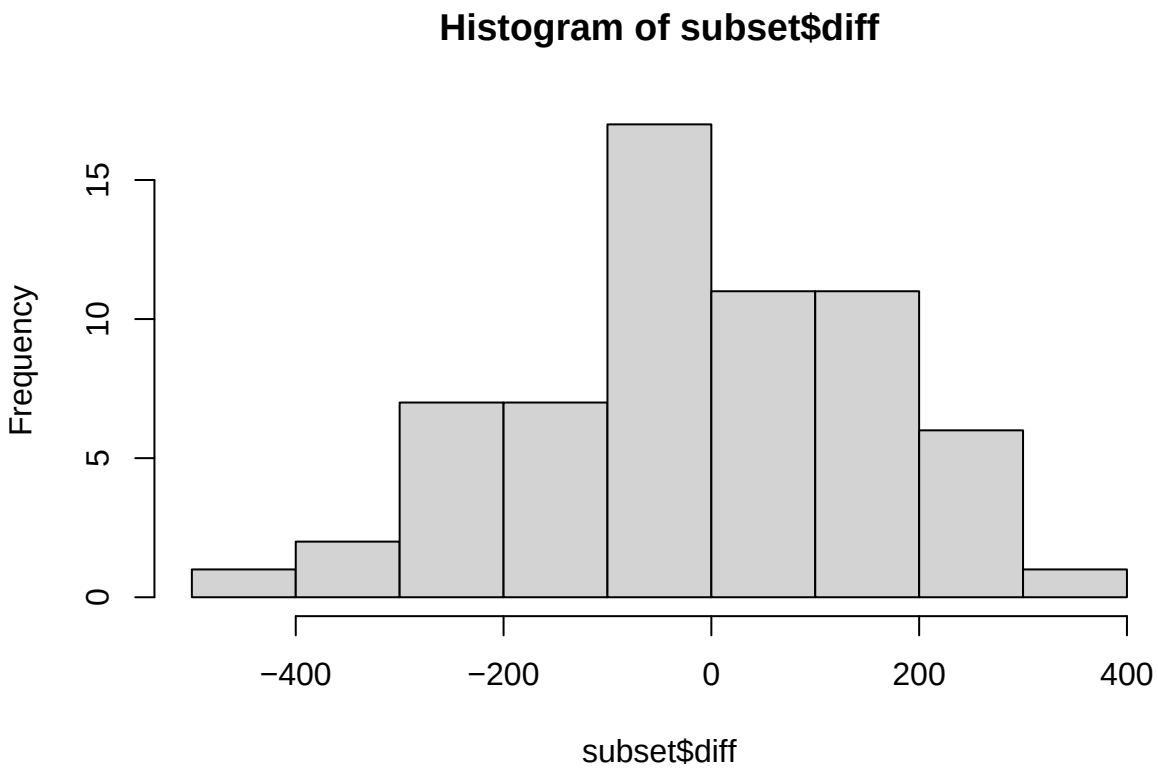

```r
hist(subset$diff)
```



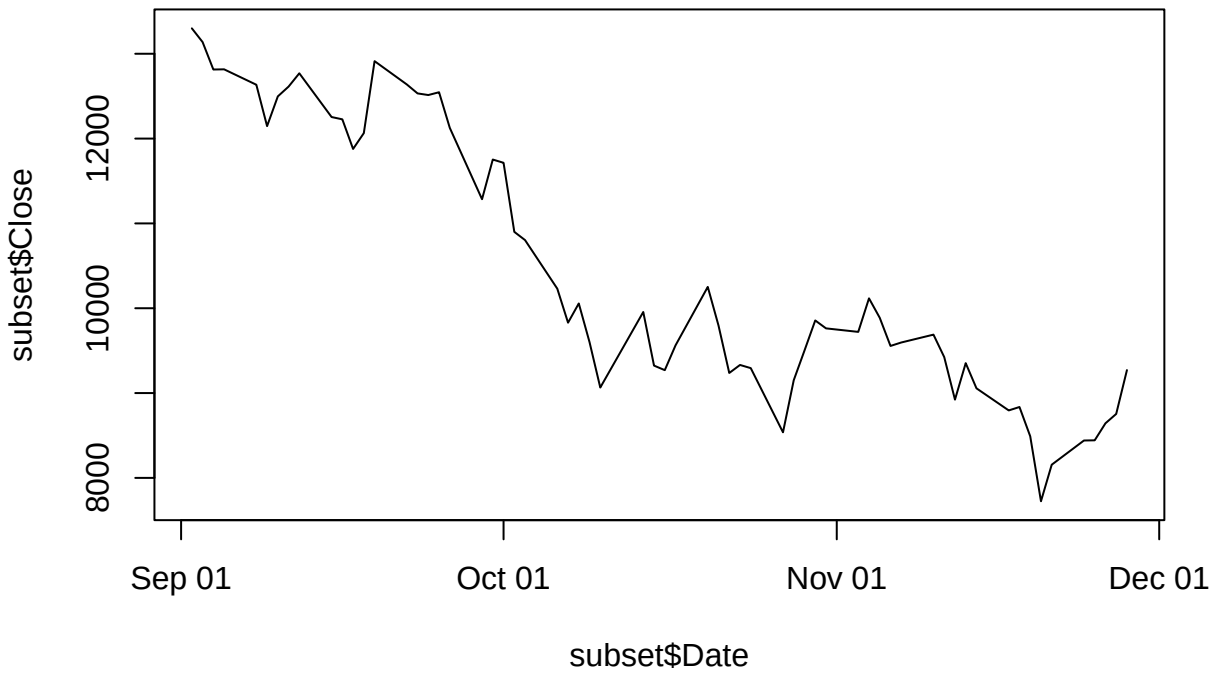

```
subset = dataset[dataset$Date > "2008-06-01" & dataset$Date < "2008-09-01",]
plot(subset$Date,subset$Close, type="l")
```



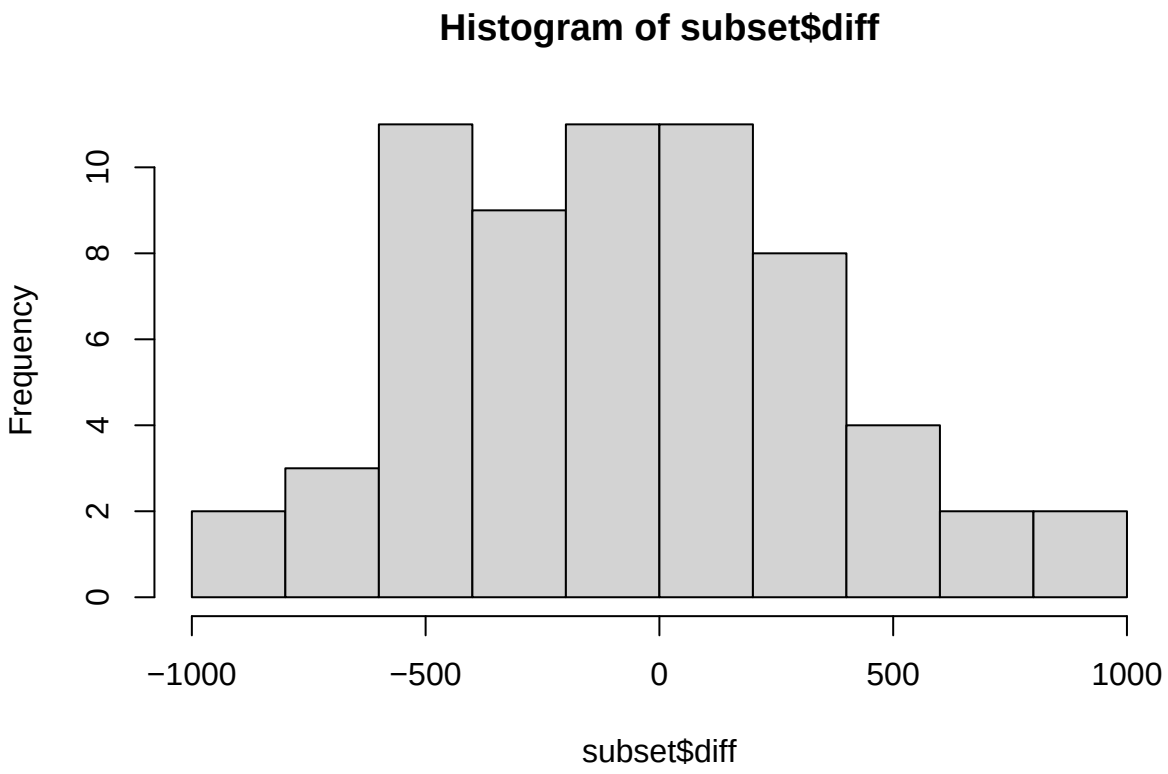

```r
hist(subset$diff)
```



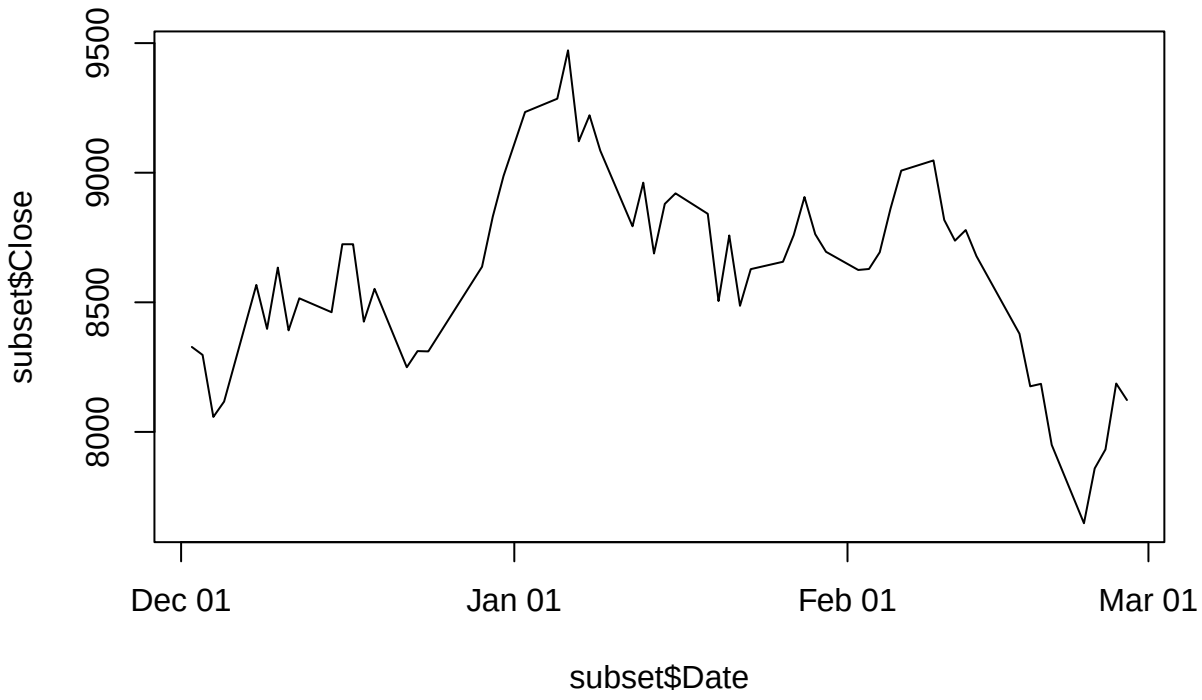

```
subset = dataset[dataset$Date > "2008-09-01" & dataset$Date < "2008-12-01",]
plot(subset$Date,subset$Close, type="l")
```



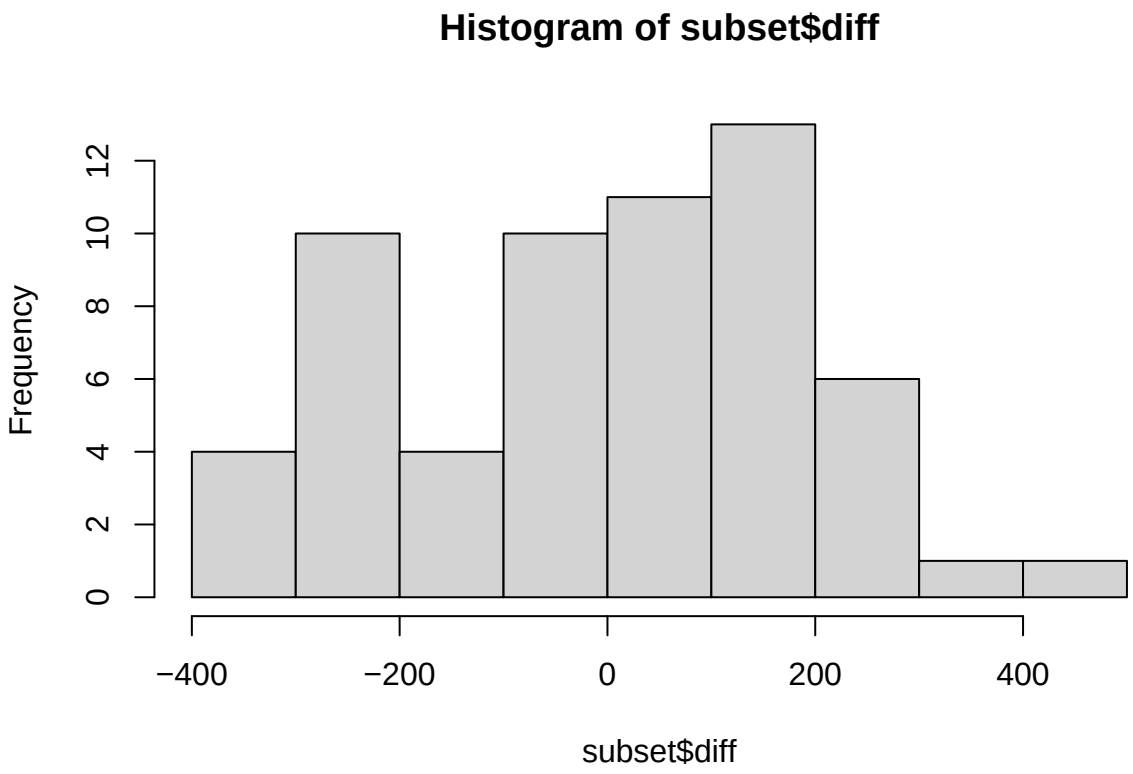

```
hist(subset$diff)
```



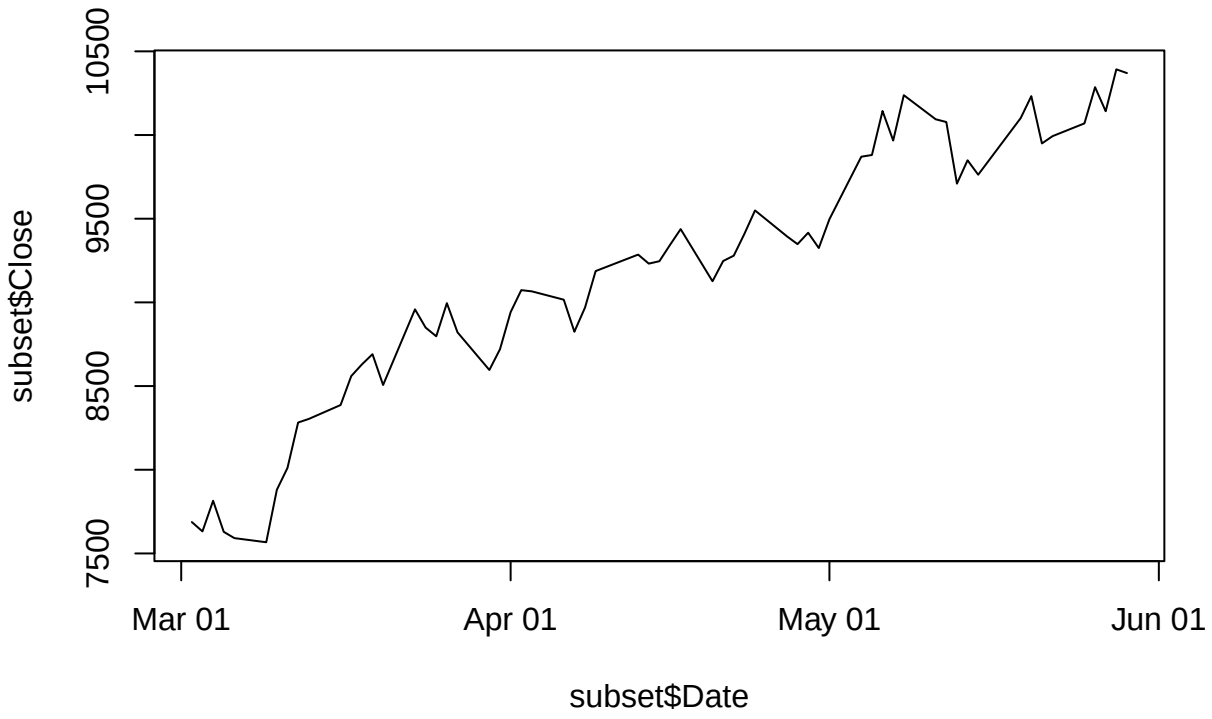

```
subset = dataset[dataset$Date > "2008-12-01" & dataset$Date < "2009-03-01",]
plot(subset$Date,subset$Close, type="l")
```



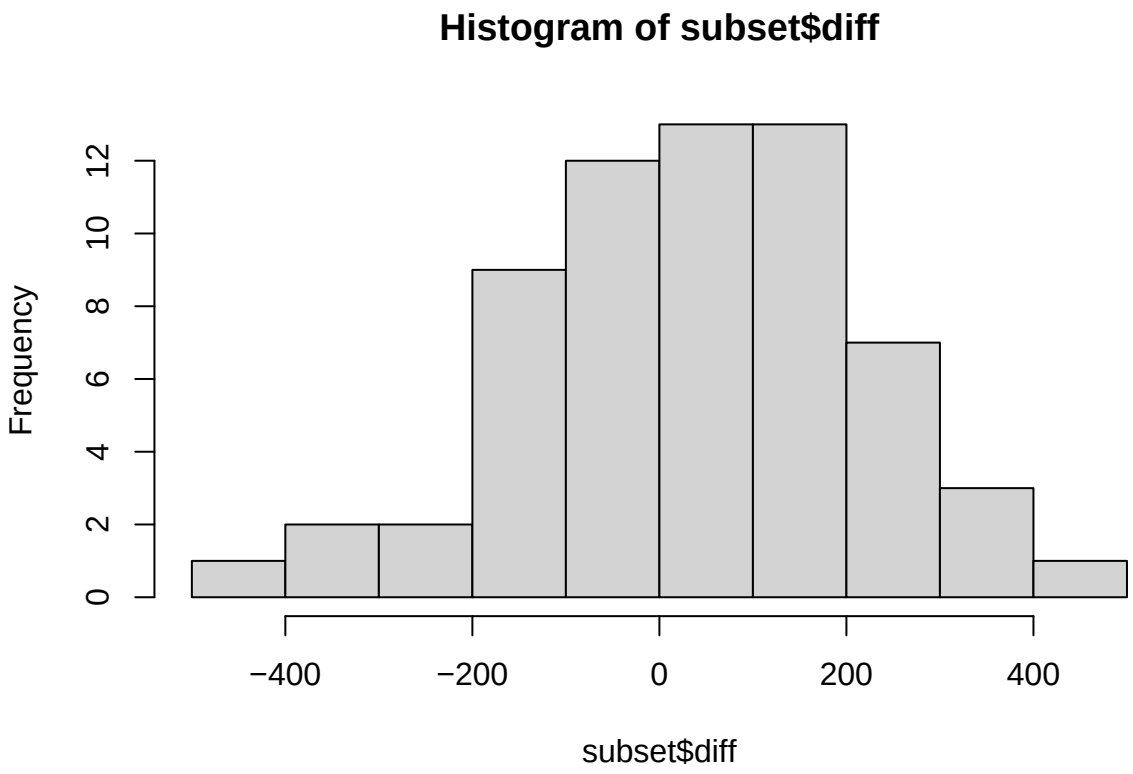

```
hist(subset$diff)
```



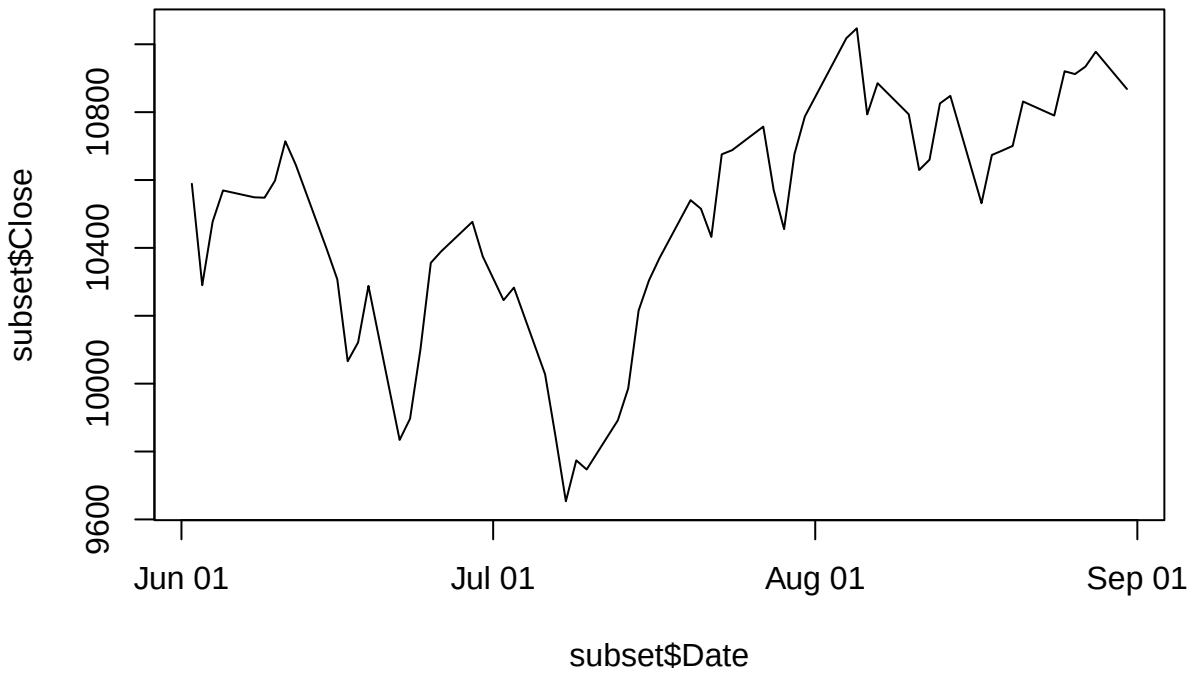

```
subset = dataset[dataset$Date > "2009-03-01" & dataset$Date < "2009-06-01",]
plot(subset$Date,subset$Close, type="l")
```



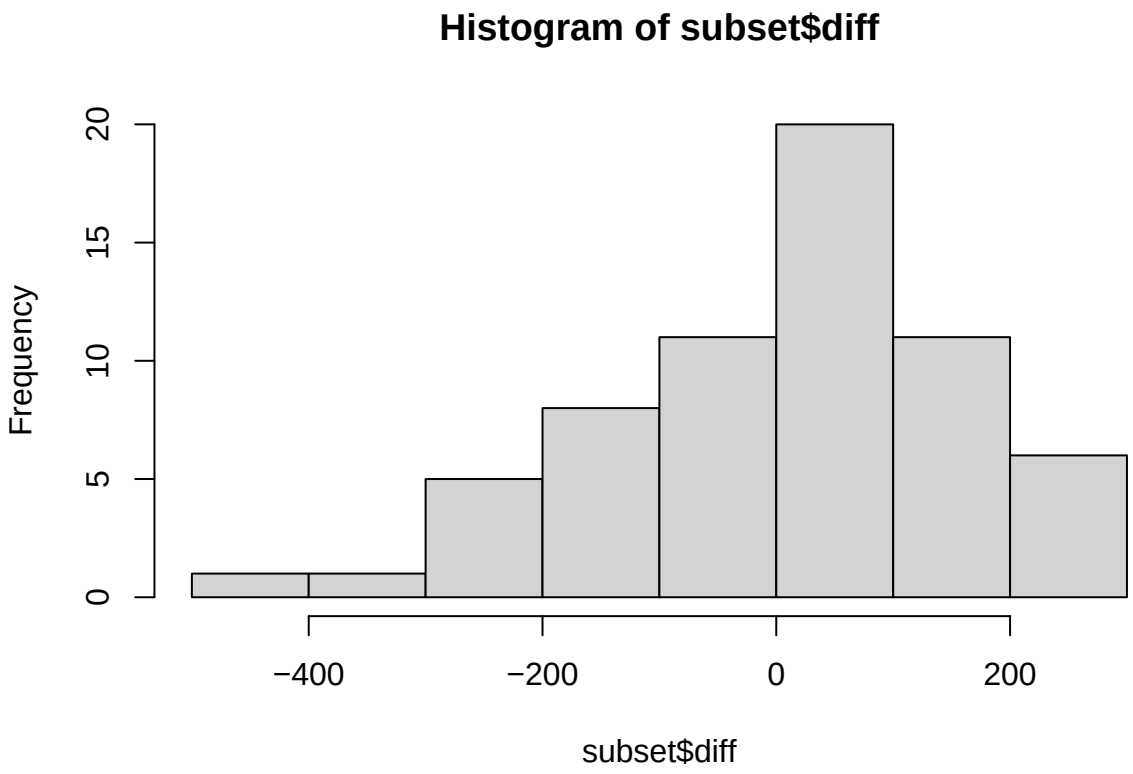

```r
hist(subset$diff)
```



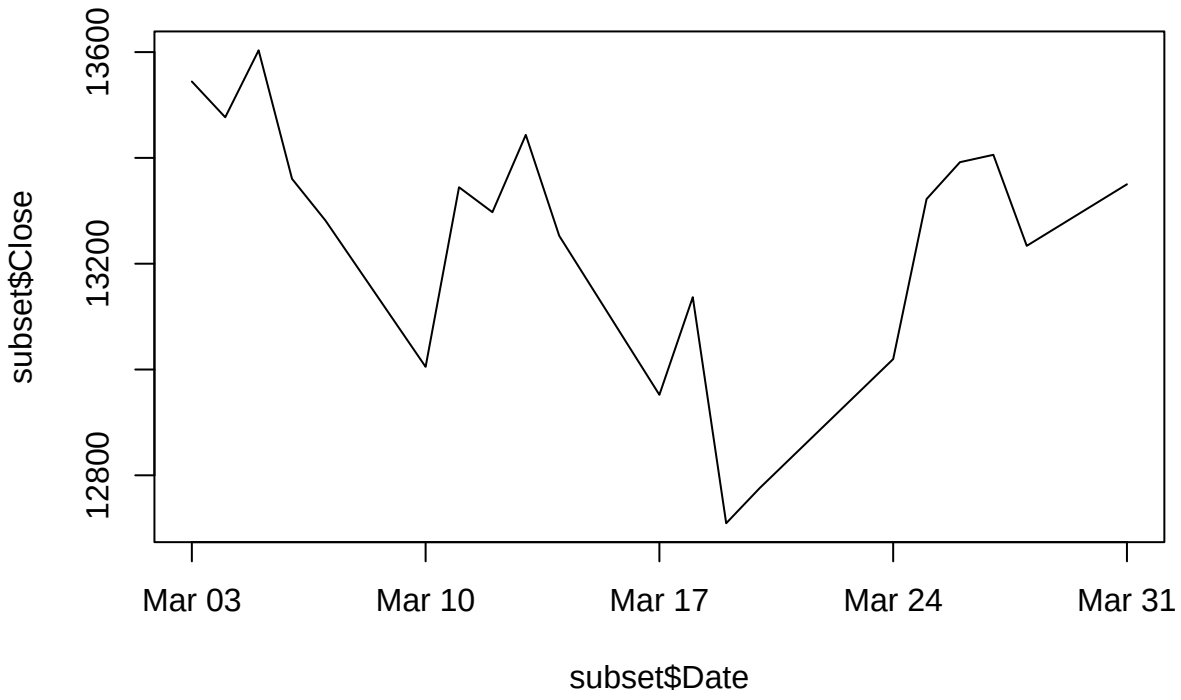

```
subset = dataset[dataset$Date > "2009-06-01" & dataset$Date < "2009-09-01",]
plot(subset$Date,subset$Close, type="l")
```



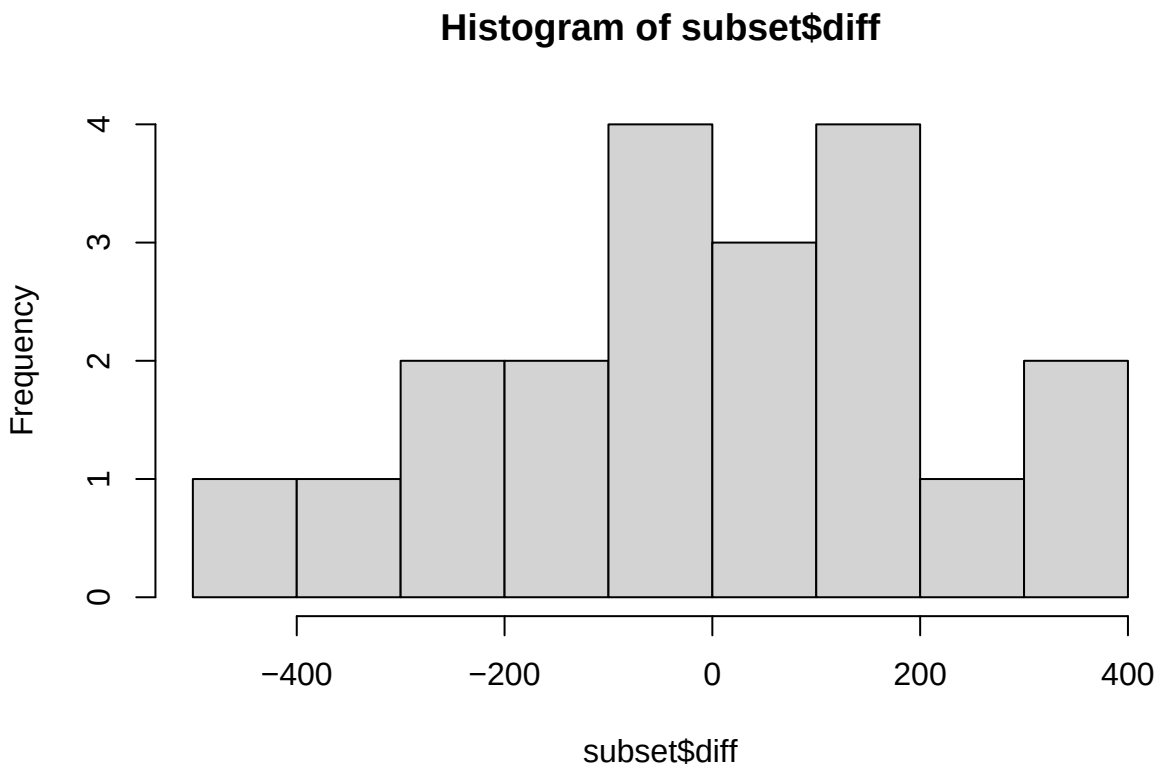

```
hist(subset$diff)
```



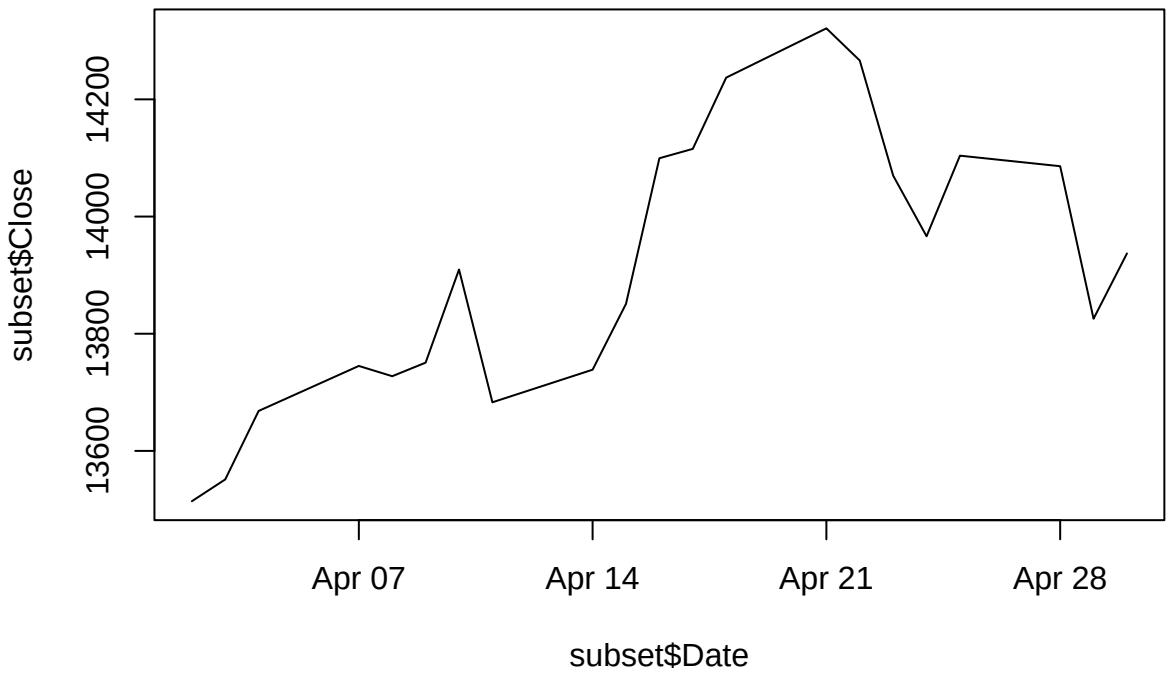

```
subset = dataset[dataset$Date > "2008-03-01" & dataset$Date < "2008-04-01",]
plot(subset$Date,subset$Close, type="l")
```



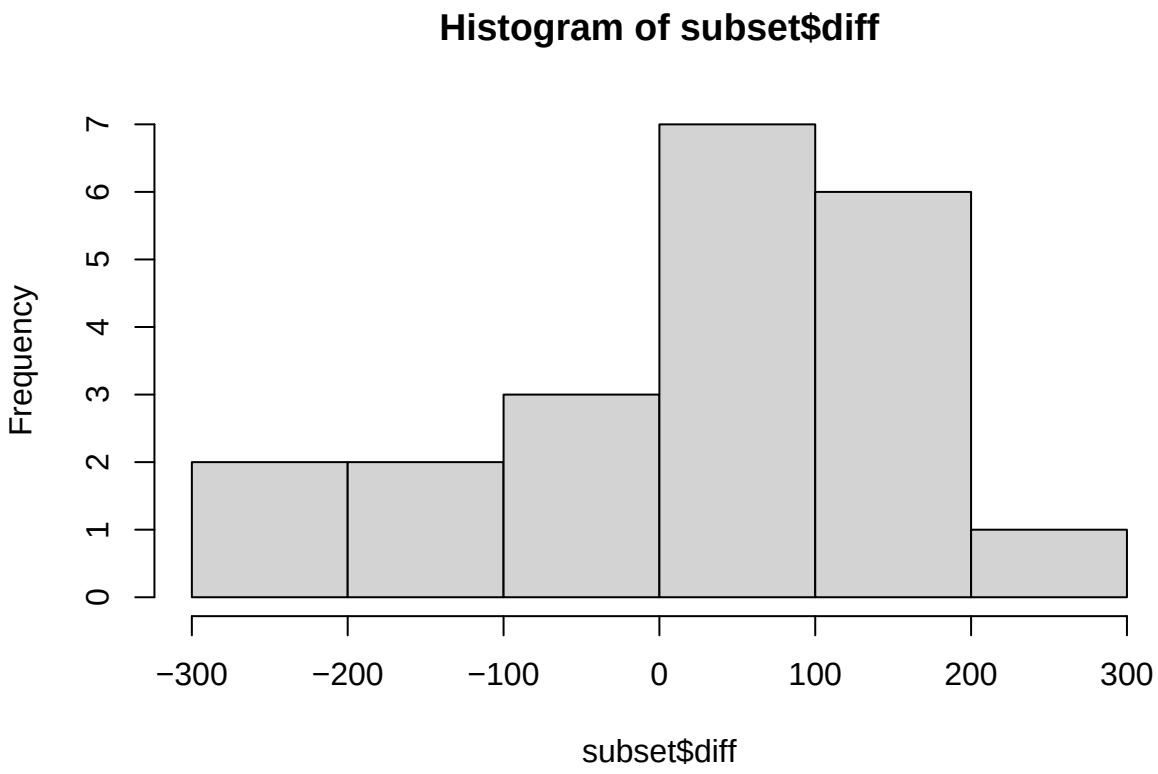

```
hist(subset$diff)
```



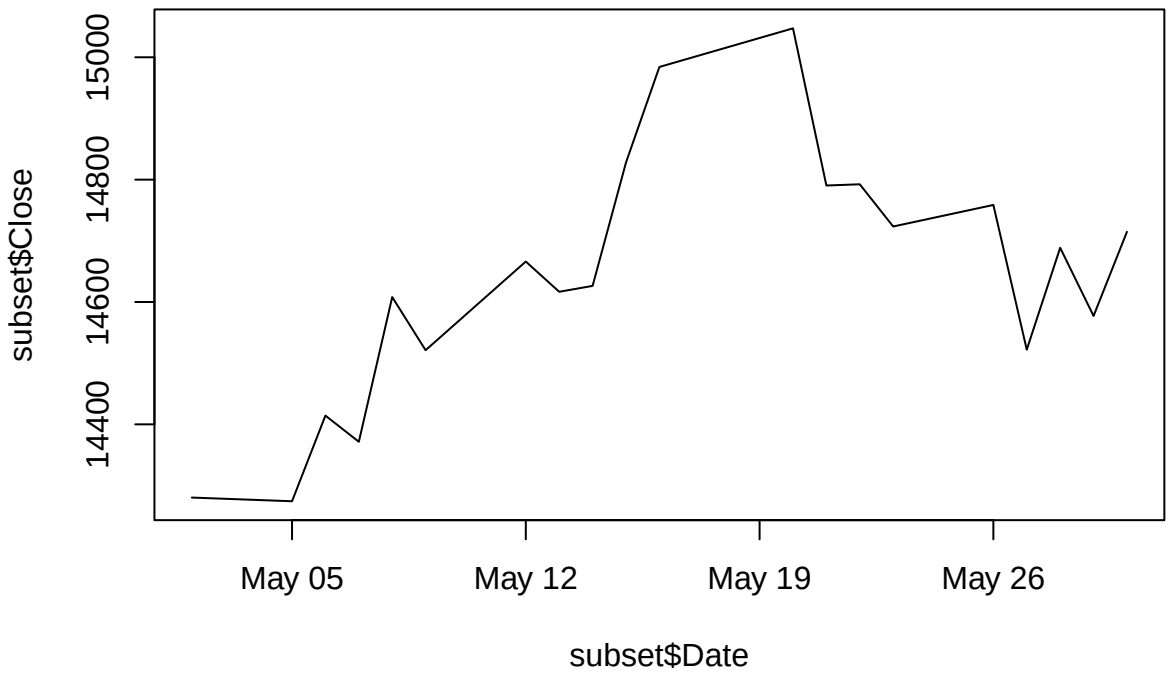

```
subset = dataset[dataset$Date > "2008-04-01" & dataset$Date < "2008-05-01",]
plot(subset$Date,subset$Close, type="l")
```



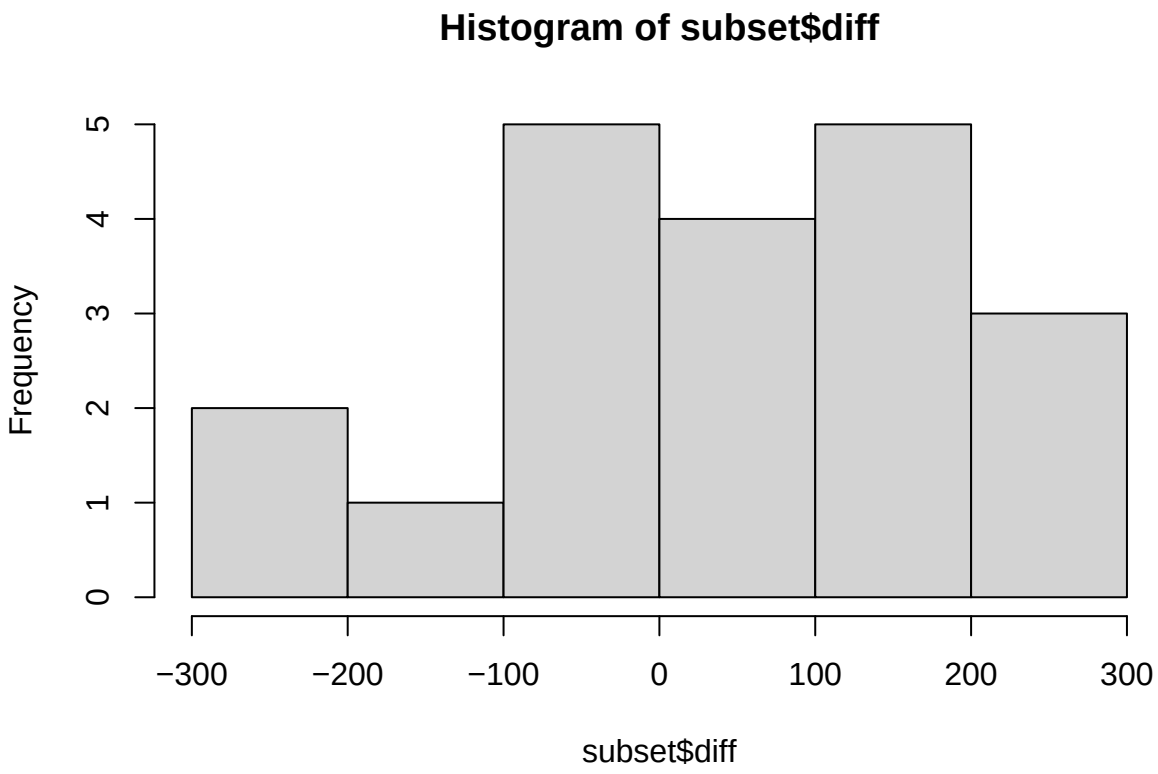

```
hist(subset$diff)
```



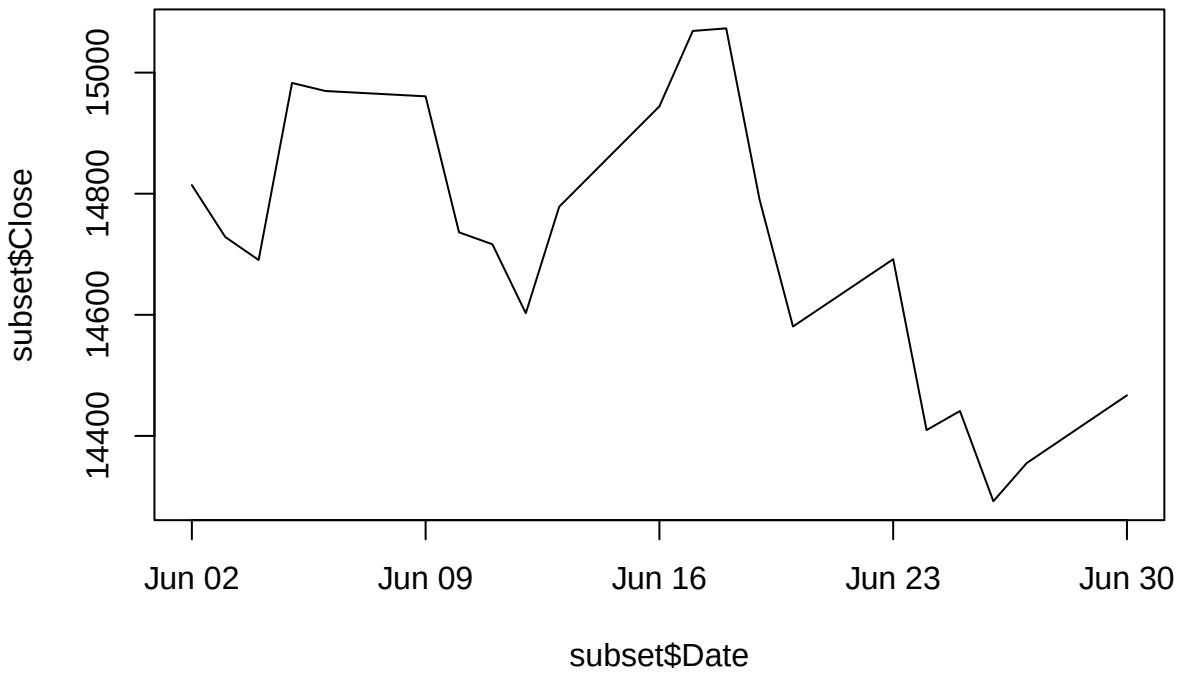

```
subset = dataset[dataset$Date > "2008-05-01" & dataset$Date < "2008-06-01",]
plot(subset$Date,subset$Close, type="l")
```



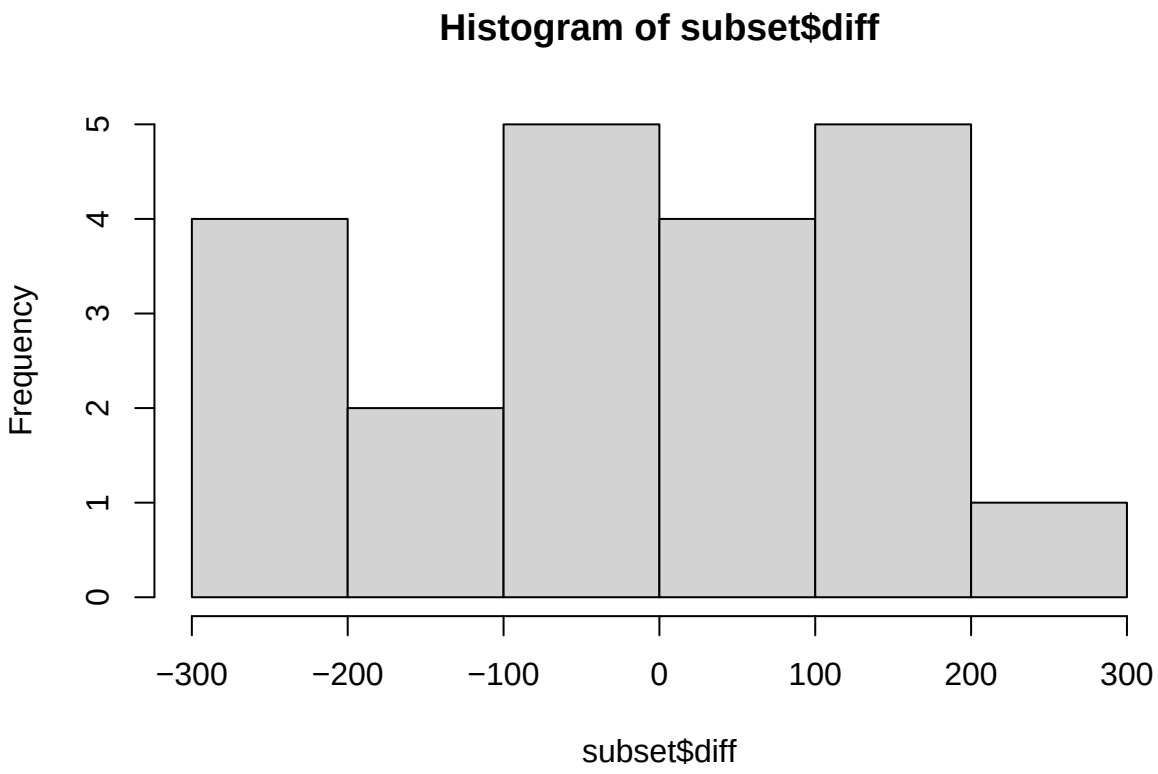

```
hist(subset$diff)
```




```
subset = dataset[dataset$Date > "2008-06-01" & dataset$Date < "2008-07-01",]
plot(subset$Date,subset$Close, type="l")
```

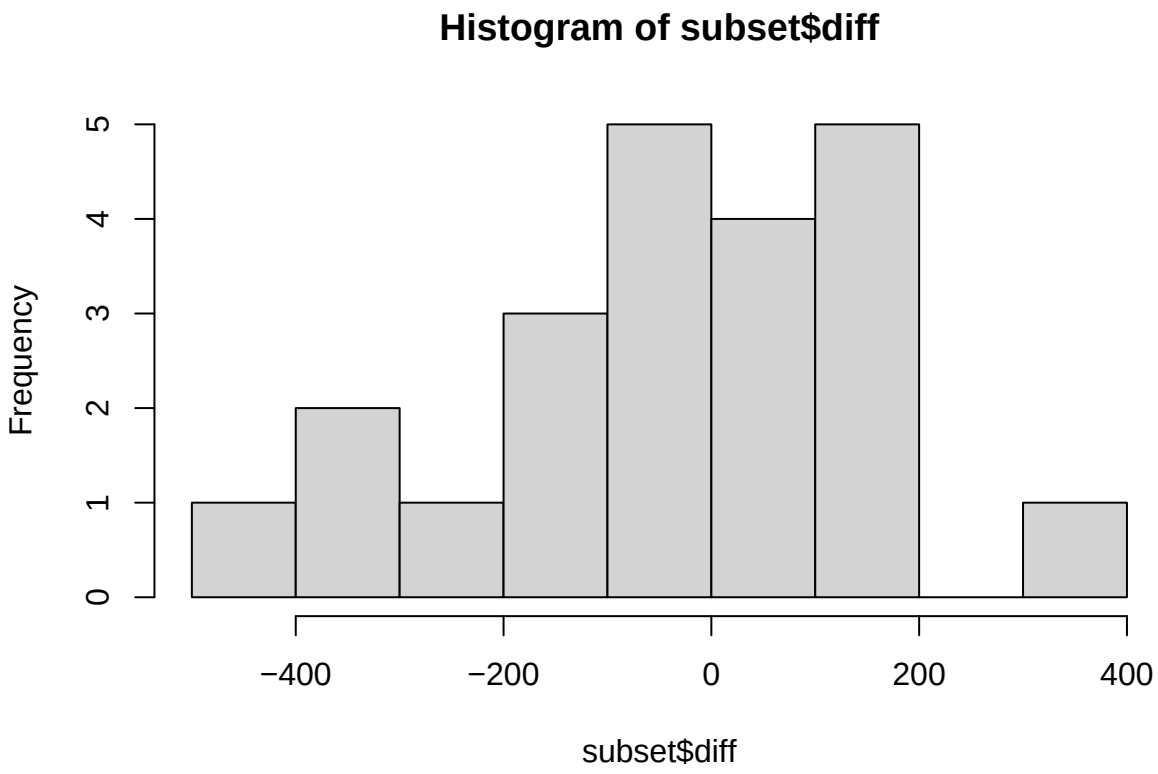



```
hist(subset$diff)
```

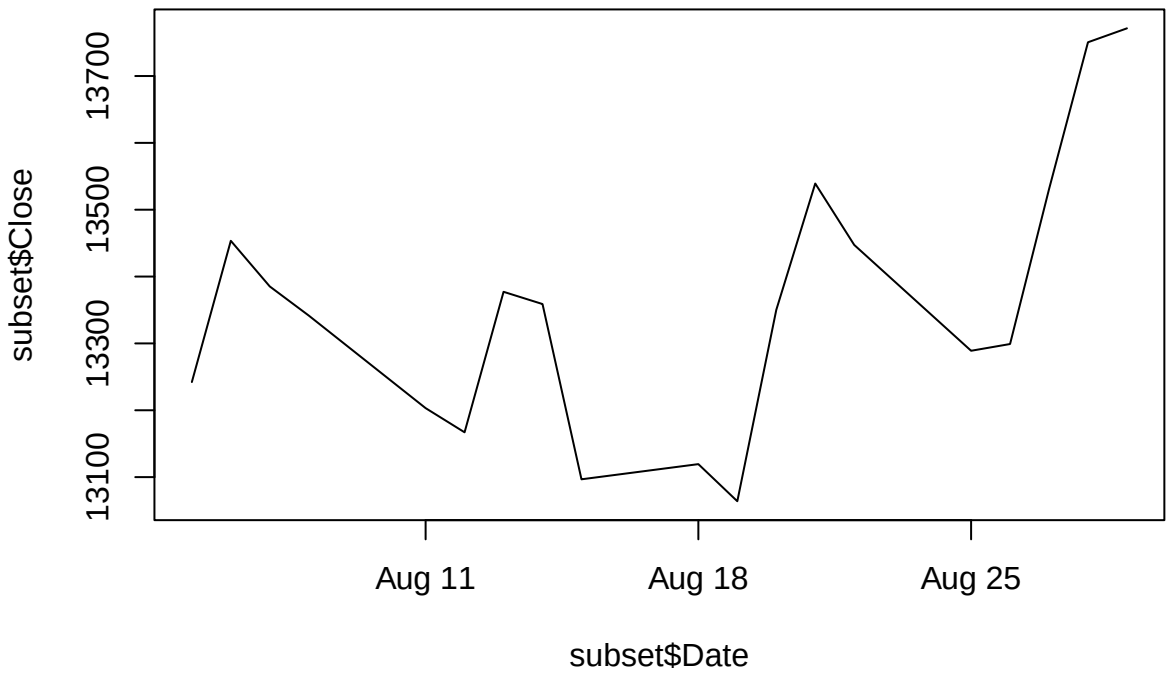



```
subset = dataset[dataset$Date > "2008-07-01" & dataset$Date < "2008-08-01",]
plot(subset$Date,subset$Close, type="l")
```

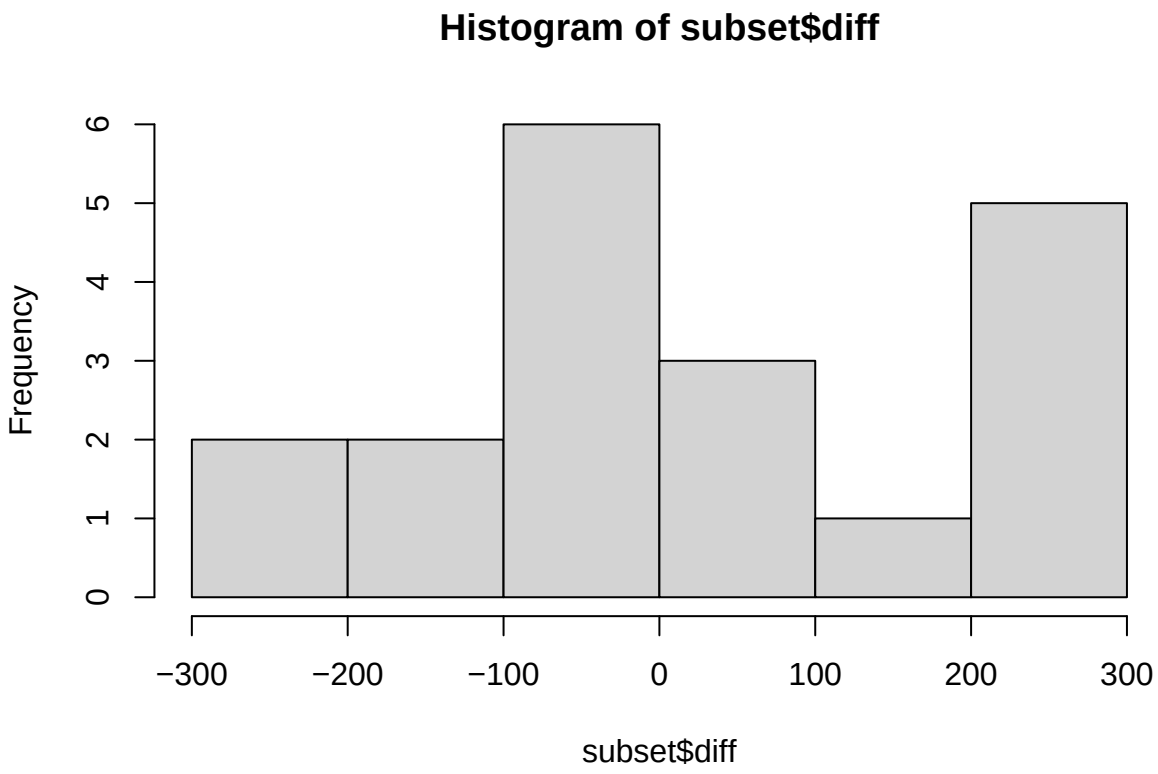

```
hist(subset$diff)
```



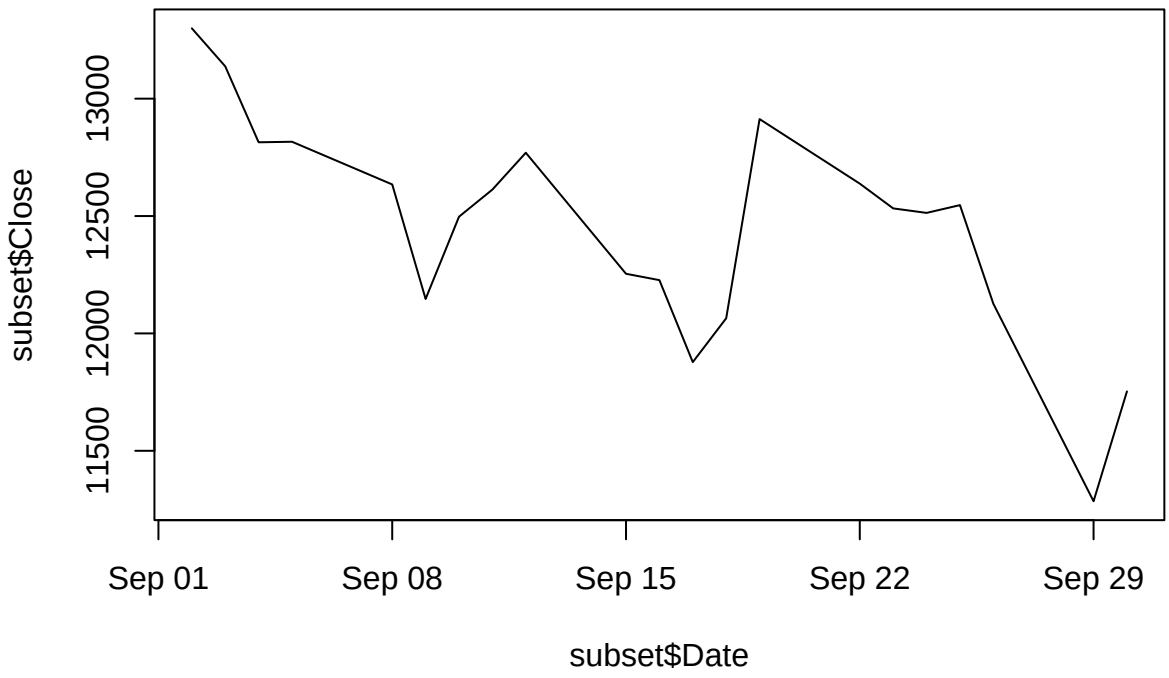

```
subset = dataset[dataset$Date > "2008-08-01" & dataset$Date < "2008-09-01",]
plot(subset$Date,subset$Close, type="l")
```



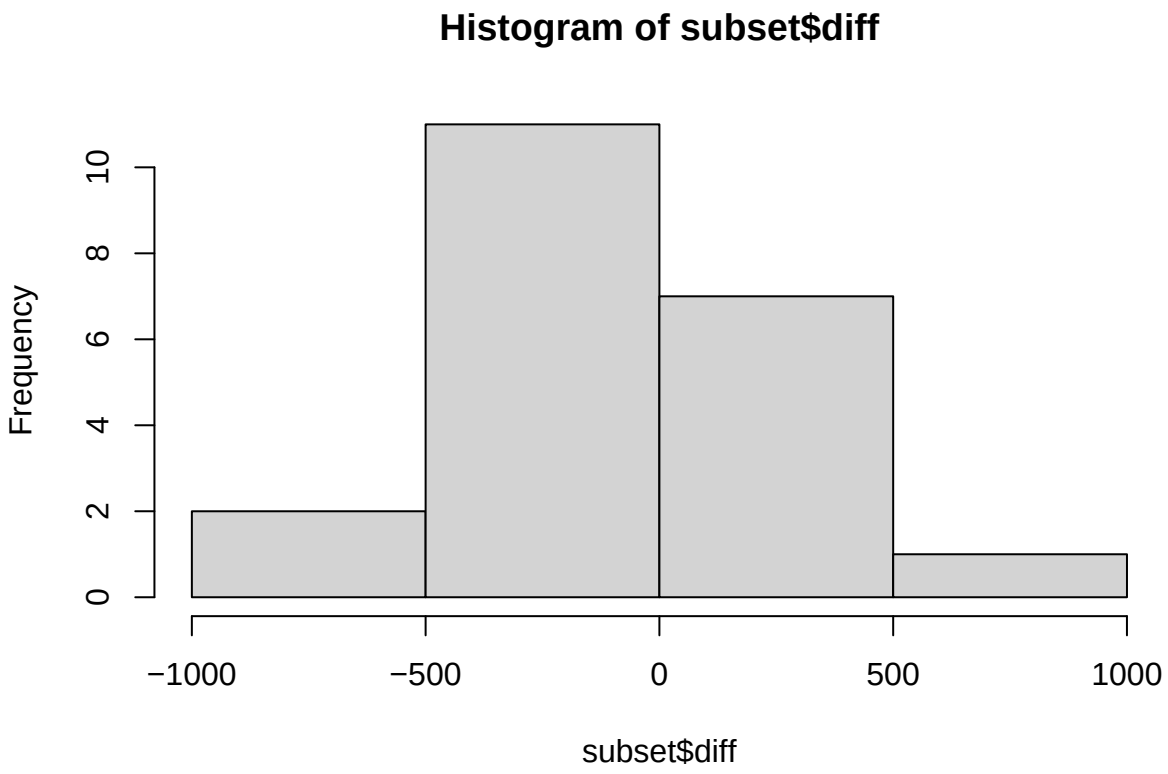

```
hist(subset$diff)
```



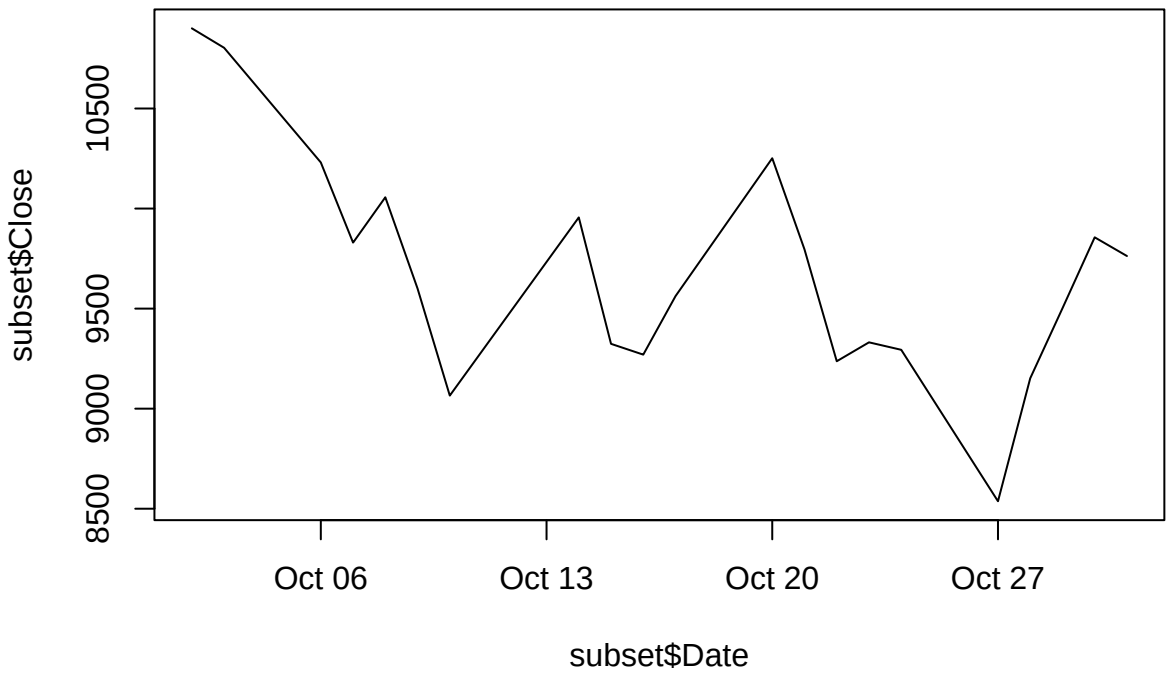

```
subset = dataset[dataset$Date > "2008-09-01" & dataset$Date < "2008-10-01",]
plot(subset$Date,subset$Close, type="l")
```



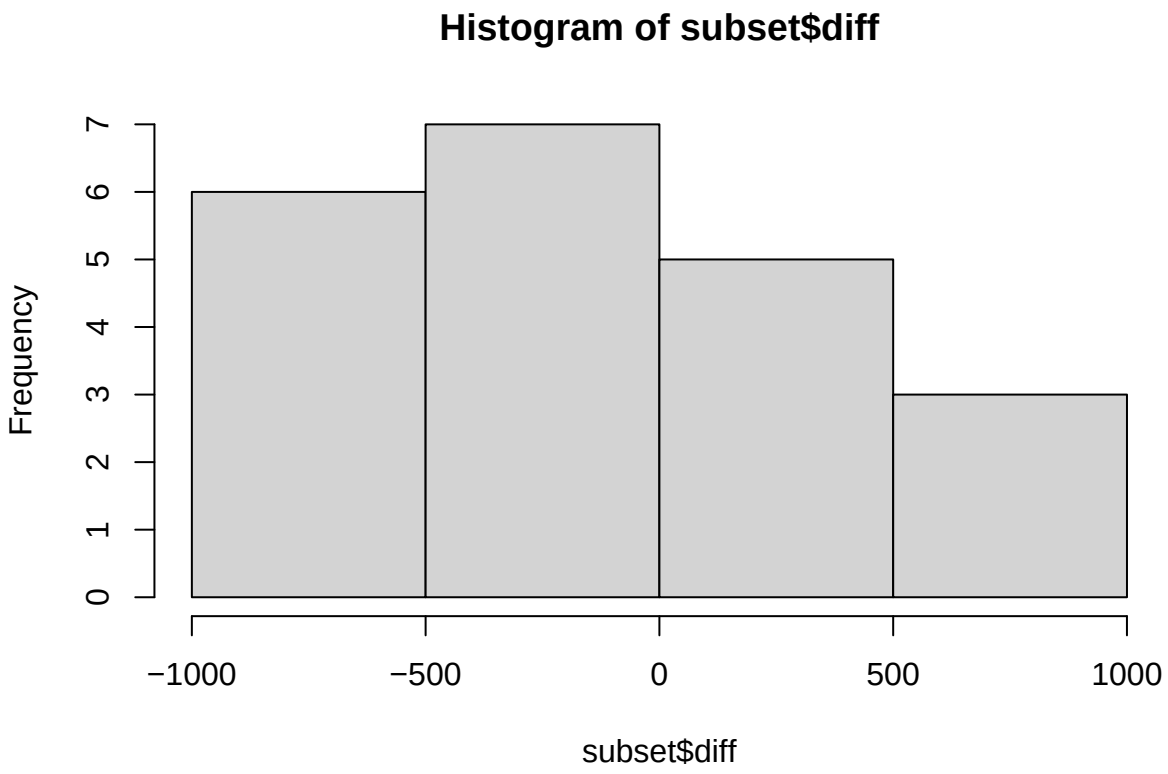

```
hist(subset$diff)
```



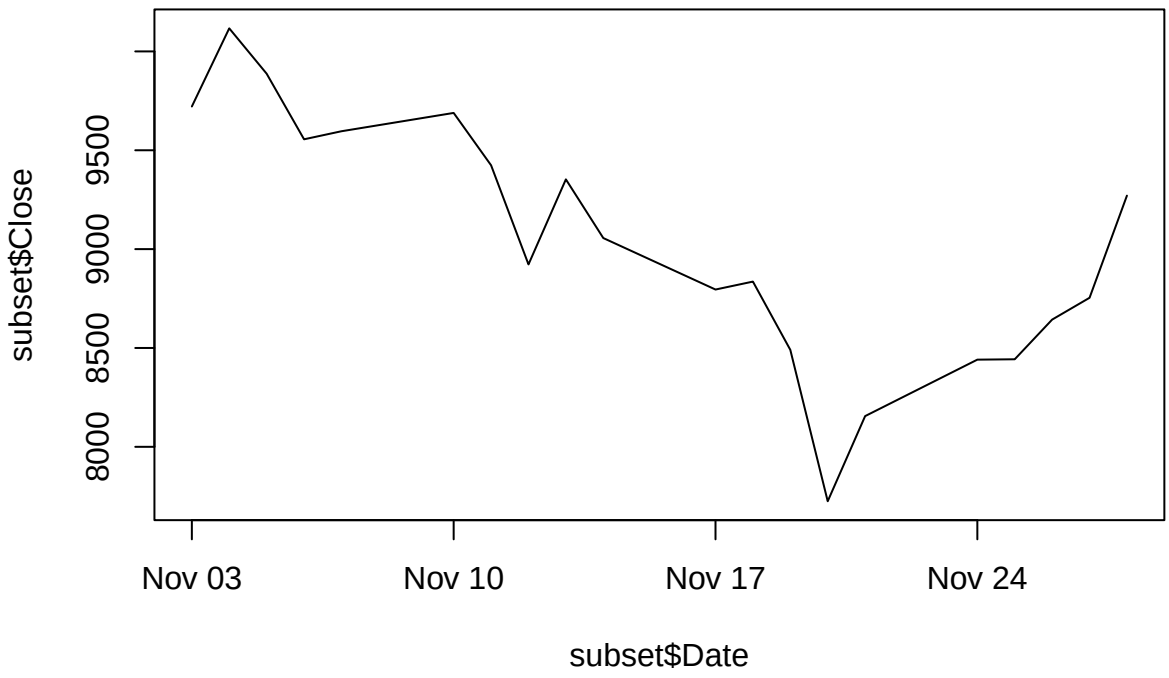

```
subset = dataset[dataset$Date > "2008-10-01" & dataset$Date < "2008-11-01",]
plot(subset$Date,subset$Close, type="l")
```



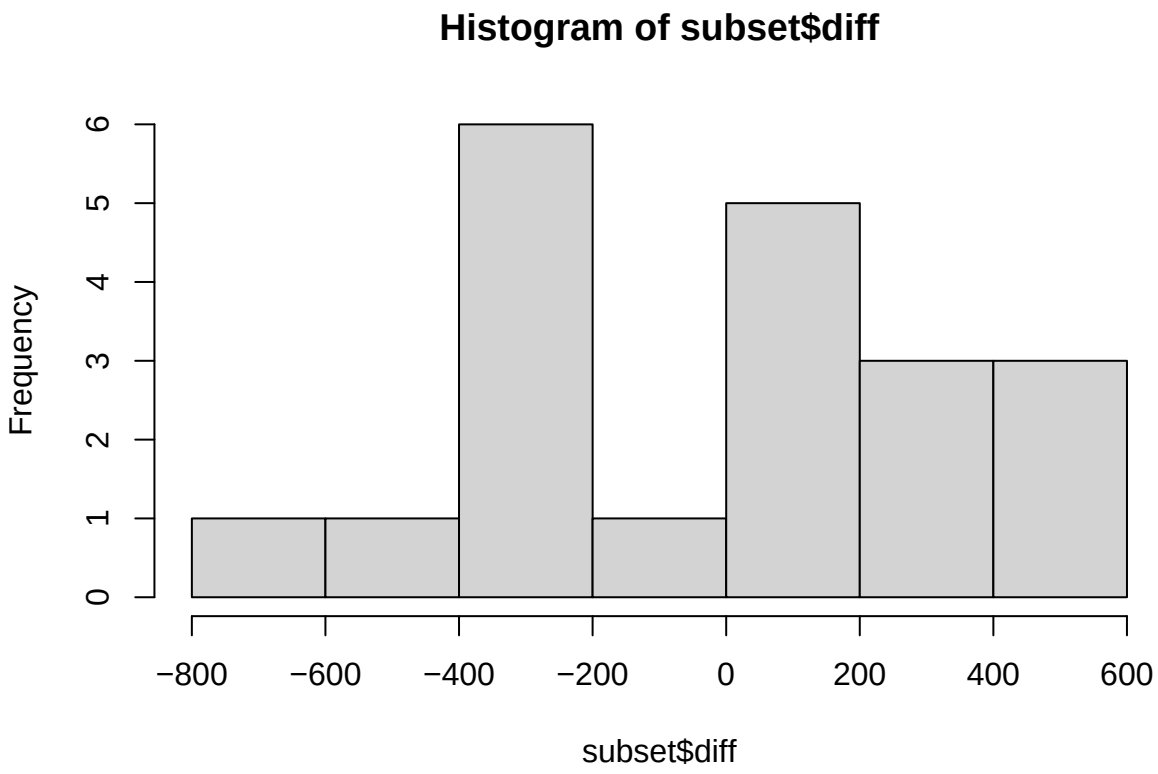

```
hist(subset$diff)
```



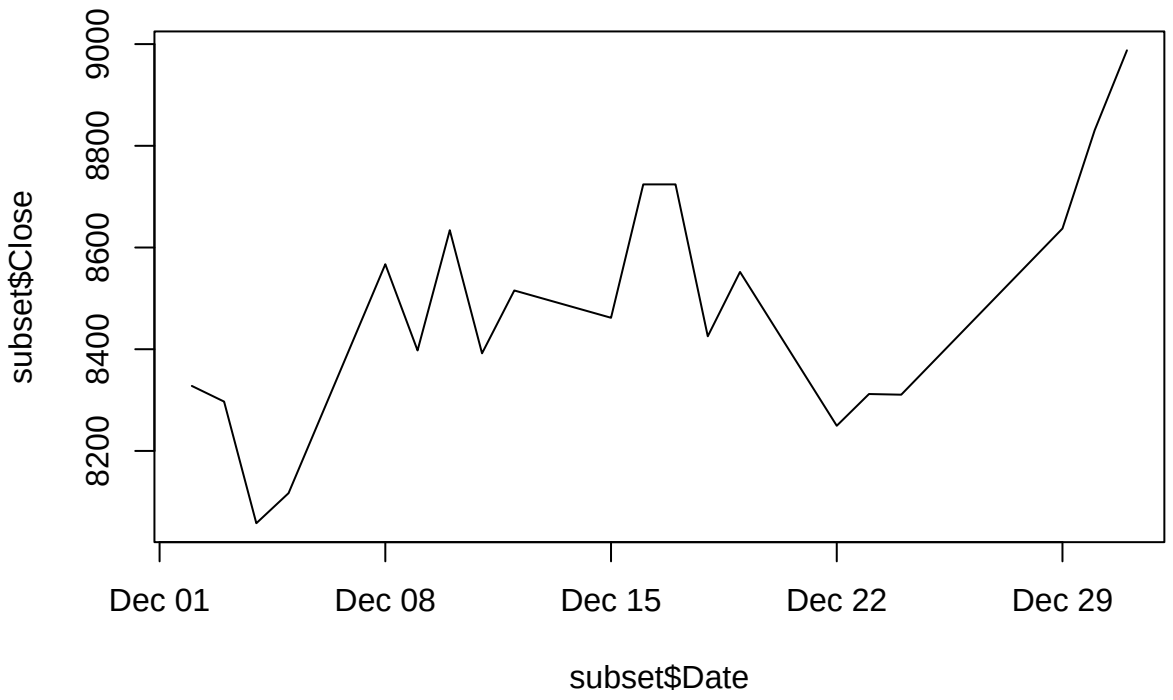

```
subset = dataset[dataset$Date > "2008-11-01" & dataset$Date < "2008-12-01",]
plot(subset$Date,subset$Close, type="l")
```



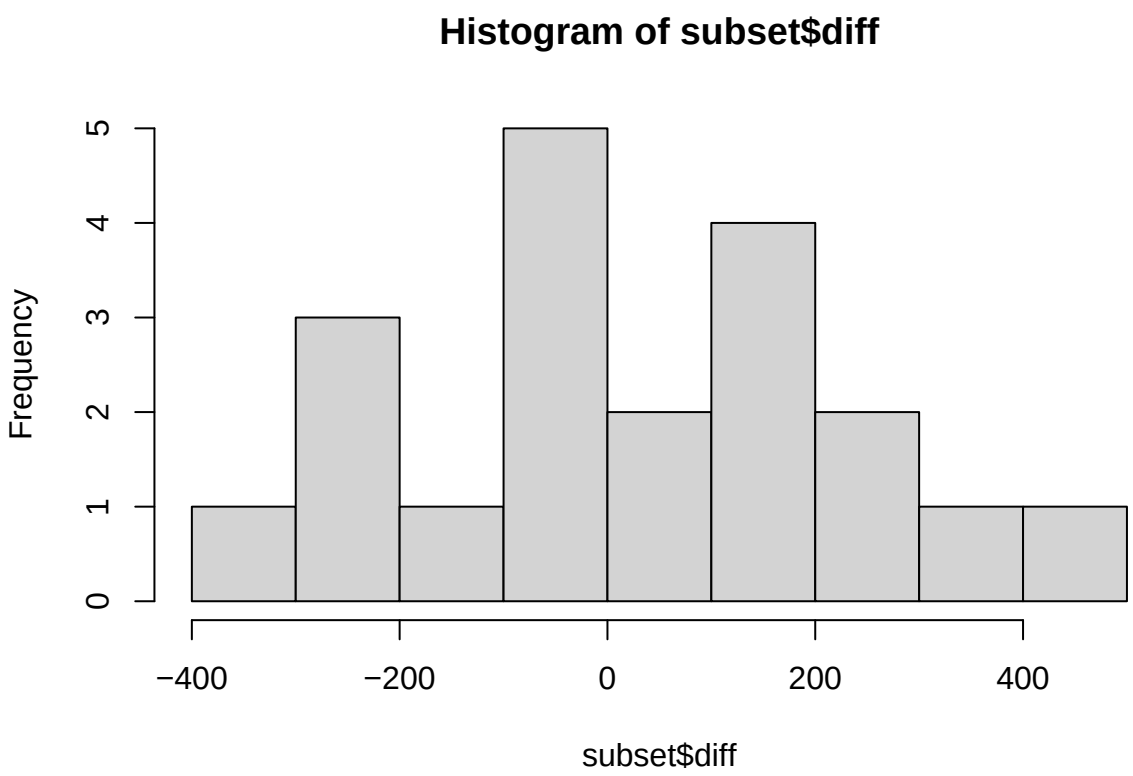

```
hist(subset$diff)
```




```
subset = dataset[dataset$Date > "2008-12-01" & dataset$Date < "2009-01-01",]
plot(subset$Date,subset$Close, type="l")
```




```
hist(subset$diff)
```


Histogram of subset$diff
Frequency
0
1
2
3
4
5
−400
−200
0
200
400
subset$diff