CSC425 HW1

Problem 1

Load Library

```r
library(tseries)
library(fBasics)
```

```
## Loading required package: timeDate

## Loading required package: timeSeries

##

## Rmetrics Package fBasics

## Analysing Markets and calculating Basic Statistics

## Copyright (C) 2005-2014 Rmetrics Association Zurich

## Educational Software for Financial Engineering and Computational Science

## Rmetrics is free software and comes with ABSOLUTELY NO WARRANTY.

## https://www.rmetrics.org --- Mail to: info@rmetrics.org
```

```r
library(zoo)
```

```
##
## Attaching package: 'zoo'

## The following object is masked from 'package:timeSeries':
##
##     time<-

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Import Data

```r
# Set Working Directory
setwd("~/Desktop/CSC425/hwork1")
# Load data with now variable names into the data frame
dat = read.table("crudeoil_w0416.csv", header = T, sep = ',')
head(dat, 20)
```

```
##         date price
## 1  02-Jan-04 32.68
## 2  09-Jan-04 33.89
## 3  16-Jan-04 34.51
## 4  23-Jan-04 35.45
## 5  30-Jan-04 33.61
## 6  06-Feb-04 33.41
## 7  13-Feb-04 33.88
## 8  20-Feb-04 35.54
```

```
## 9  27-Feb-04 36.08
## 10 05-Mar-04 36.67
## 11 12-Mar-04 36.44
## 12 19-Mar-04 37.78
## 13 26-Mar-04 36.65
## 14 02-Apr-04 35.23
## 15 09-Apr-04 35.70
## 16 16-Apr-04 37.39
## 17 23-Apr-04 37.32
## 18 30-Apr-04 37.31
## 19 07-May-04 39.24
## 20 14-May-04 40.37
```

a) Create a Time plot for the time series of spot prices.

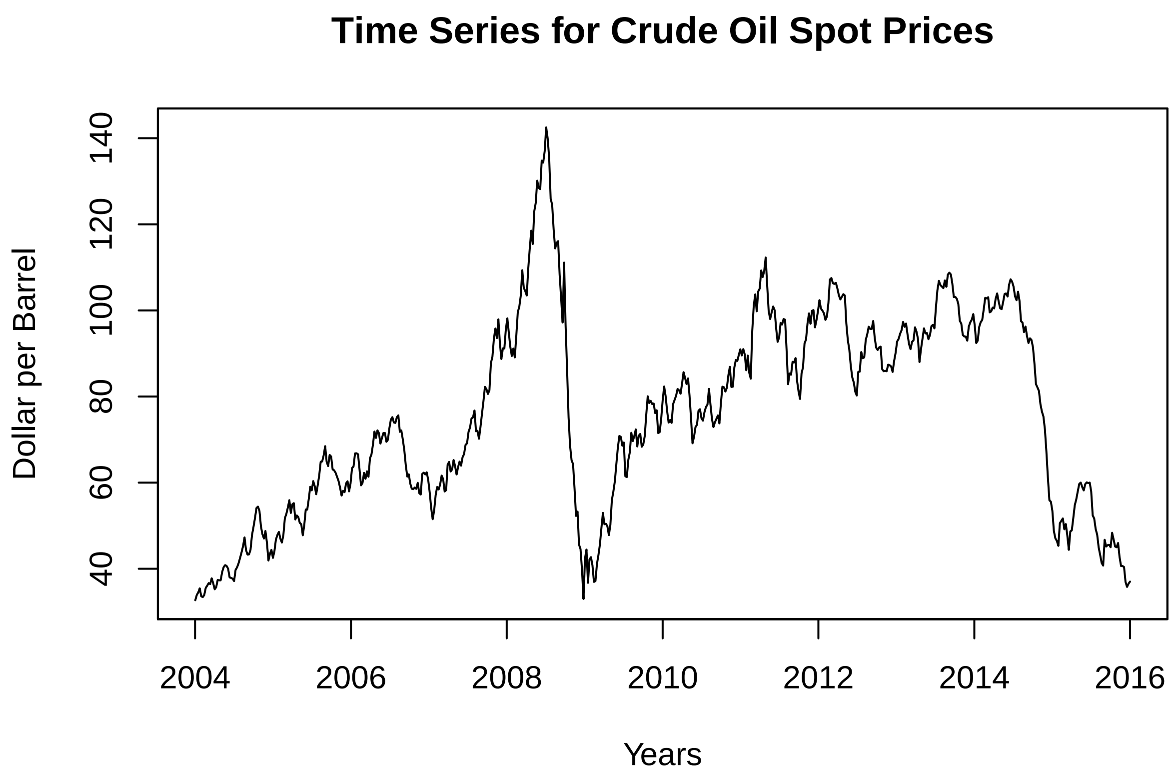
```r
# Time series
TSprice = zoo(dat[,2], as.Date(as.character(dat[,1]), format = "%d-%b-%y"))
head(TSprice,20)
```

```
## 2004-01-02 2004-01-09 2004-01-16 2004-01-23 2004-01-30 2004-02-06
##      32.68      33.89      34.51      35.45      33.61      33.41
## 2004-02-13 2004-02-20 2004-02-27 2004-03-05 2004-03-12 2004-03-19
##      33.88      35.54      36.08      36.67      36.44      37.78
## 2004-03-26 2004-04-02 2004-04-09 2004-04-16 2004-04-23 2004-04-30
##      36.65      35.23      35.70      37.39      37.32      37.31
## 2004-05-07 2004-05-14
##      39.24      40.37
```

```r
tail(TSprice,20)
```

```
## 2015-08-21 2015-08-28 2015-09-04 2015-09-11 2015-09-18 2015-09-25
##      41.34      40.73      46.73      45.16      45.48      45.57
## 2015-10-02 2015-10-09 2015-10-16 2015-10-23 2015-10-30 2015-11-06
##      45.00      48.36      46.82      45.16      44.99      45.98
## 2015-11-13 2015-11-20 2015-11-27 2015-12-04 2015-12-11 2015-12-18
##      42.70      40.62      40.63      40.40      36.93      35.78
## 2015-12-25 2016-01-01
##      36.53      37.02
```

```r
# plot
plot(TSprice, xlab = "Years", ylab = "Dollar per Barrel", main = "Time Series for Crude Oil Spot Prices
```



In general, the changes of oil prices goes up and down from January 2004 to January 2016. In particuluar, there is a much larger scale of fluctuation during 2008 due to the financial crisis. After the year, its prices are stablized between $70 and $110 from 2009 to 2015. Another highlighted point is occured in 2015 that its price rapidly decreased.

b) Compute the percentage change rate of spot prices

```
rate = (TSprice - lag(TSprice, -1)) / lag(TSprice,-1)
head(rate,20)
```

```
##    2004-01-09    2004-01-16    2004-01-23    2004-01-30    2004-02-06
##  0.0370257038  0.0182944821  0.0272384816 -0.0519040903 -0.0059506099
##    2004-02-13    2004-02-20    2004-02-27    2004-03-05    2004-03-12
##  0.0140676444  0.0489964581  0.0151941474  0.0163525499 -0.0062721571
##    2004-03-19    2004-03-26    2004-04-02    2004-04-09    2004-04-16
##  0.0367727772 -0.0299100053 -0.0387448840  0.0133409026  0.0473389356
##    2004-04-23    2004-04-30    2004-05-07    2004-05-14    2004-05-21
## -0.0018721583 -0.0002679528  0.0517287590  0.0287971458  0.0116423086
```

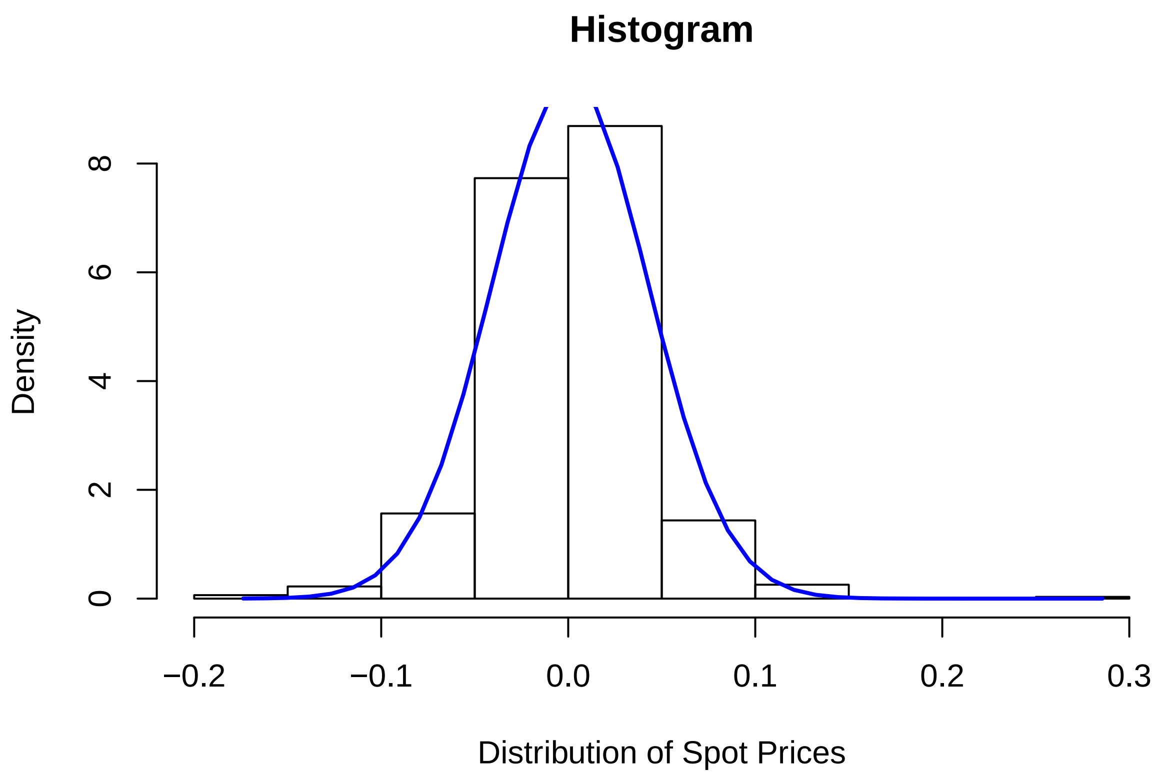
```
tail(rate,20)
```

```
##    2015-08-21    2015-08-28    2015-09-04    2015-09-11    2015-09-18
## -0.0430555556 -0.0147556846  0.1473115640 -0.0335972609  0.0070859167
##    2015-09-25    2015-10-02    2015-10-09    2015-10-16    2015-10-23
##  0.0019788918 -0.0125082291  0.0746666667 -0.0318444996 -0.0354549338
##    2015-10-30    2015-11-06    2015-11-13    2015-11-20    2015-11-27
## -0.0037643933  0.0220048900 -0.0713353632 -0.0487119438  0.0002461841
##    2015-12-04    2015-12-11    2015-12-18    2015-12-25    2016-01-01
```

```
## -0.0056608417 -0.0858910891 -0.0311399946  0.0209614310  0.0134136326
```

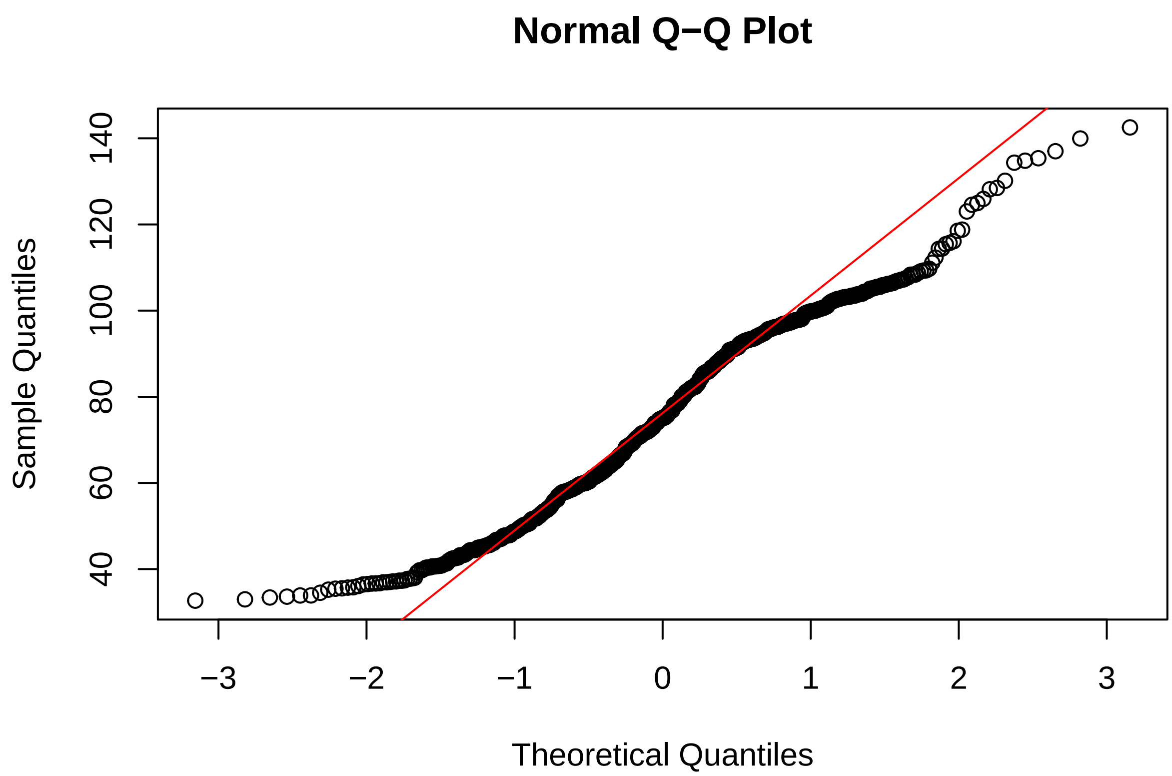
C) Analyze the distribution of rate using a histogram and a normal quantile plot.

```r
#histogram
par(mfcol = c(1,1))
hist(rate, xlab = "Distribution of Spot Prices", prob = TRUE, main = "Histogram")
#add approximating normal density curve
xfit <- seq(min(rate), max(rate), length = 40)
yfit <- dnorm(xfit, mean = mean(rate), sd = sd(rate))
lines(xfit, yfit, col = "blue", lwd = 2)
```



Based on the graph above, the distribution is little bit skewed to the right.

```r
#Normal quantile plot
qqnorm(TSprice)
qqline(TSprice, col = 2)
```



Based on the qq plot, although every single black circle does not meet with the red line, it looks like it is close to a normal distribution.

d) Test the null Hypothesis of perfect symmetry for the distribution of rate at 5% significance level.

```
skew_test = skewness(rate) / sqrt(6/length(rate))
skew_test
```

```
## [1] 2.642198
## attr(,"method")
## [1] "moment"
```

Since the absolute value of skew_test is 2.642198 which is larger than 1.96, we can reject the null hypothesis of a symmetric distribution.

```
# p-value
2 * (1 - pnorm(abs(skew_test)))
```

```
## [1] 0.008236978
## attr(,"method")
## [1] "moment"
```

Since the p-value of the test is 0.008 which is less than 0.05, we can reject the null hypothesis

e) Test the null hypothesis of excess kurtosis equal to zero (normal tails) at 5% significnace level.

```
k_stat = kurtosis(rate) / sqrt(24 / length(rate))
k_stat
```

```
## [1] 23.63411
## attr(,"method")
## [1] "excess"
```

since the absolute value of k__stat is 23.63411 which is larger than 1.96, we can reject the null hypothesis of normal tails.

```
# p-value
2 * (1-pnorm(abs(k_stat)))
```

```
## [1] 0
## attr(,"method")
## [1] "excess"
```

since the p-value of the Kurtosis test is 0.0001005426, we can reject the null hypothesis

f) Test the hypothesis of normality for the distribution of rate using teh Jarque-Bera test at 5% level.

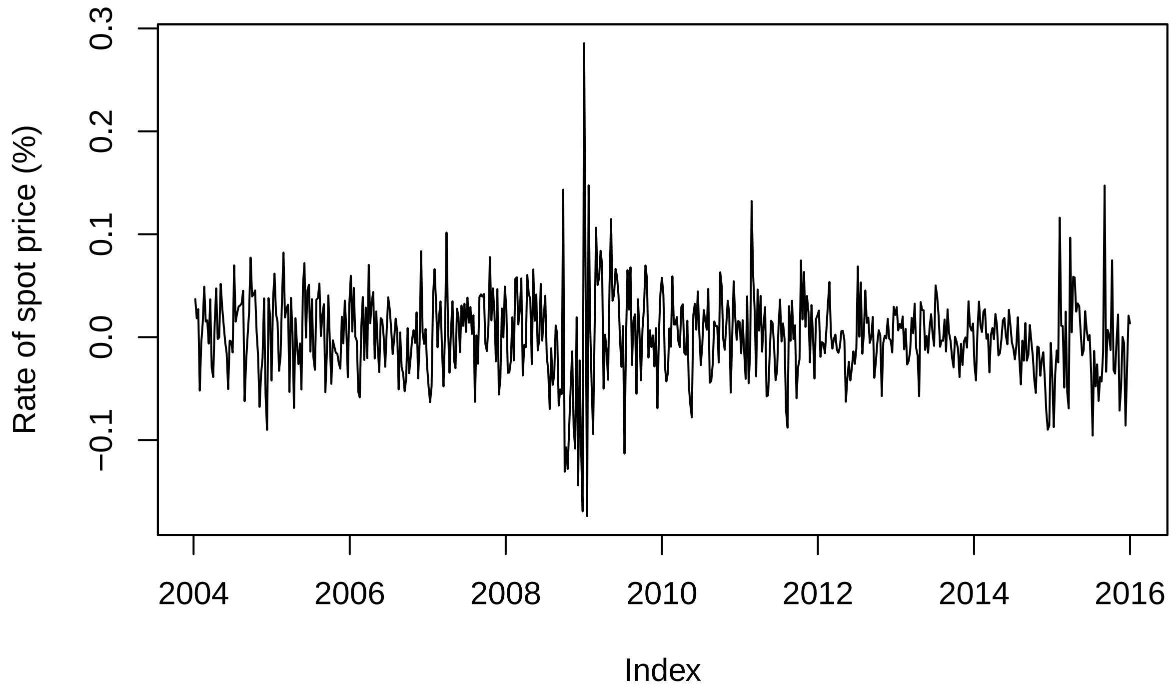
```
normalTest(rate, method = c("jb"))
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 571.4988
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Wed Sep 27 14:12:27 2017 by user:
```

since the p-value of the JB is 2.2e-16 which is less than 0.05, we can reject the null hypothesis of normal distribution.

g) Create a time plot for the time series of rate.

```
plot(rate, ylab = 'Rate of spot price (%)')
```



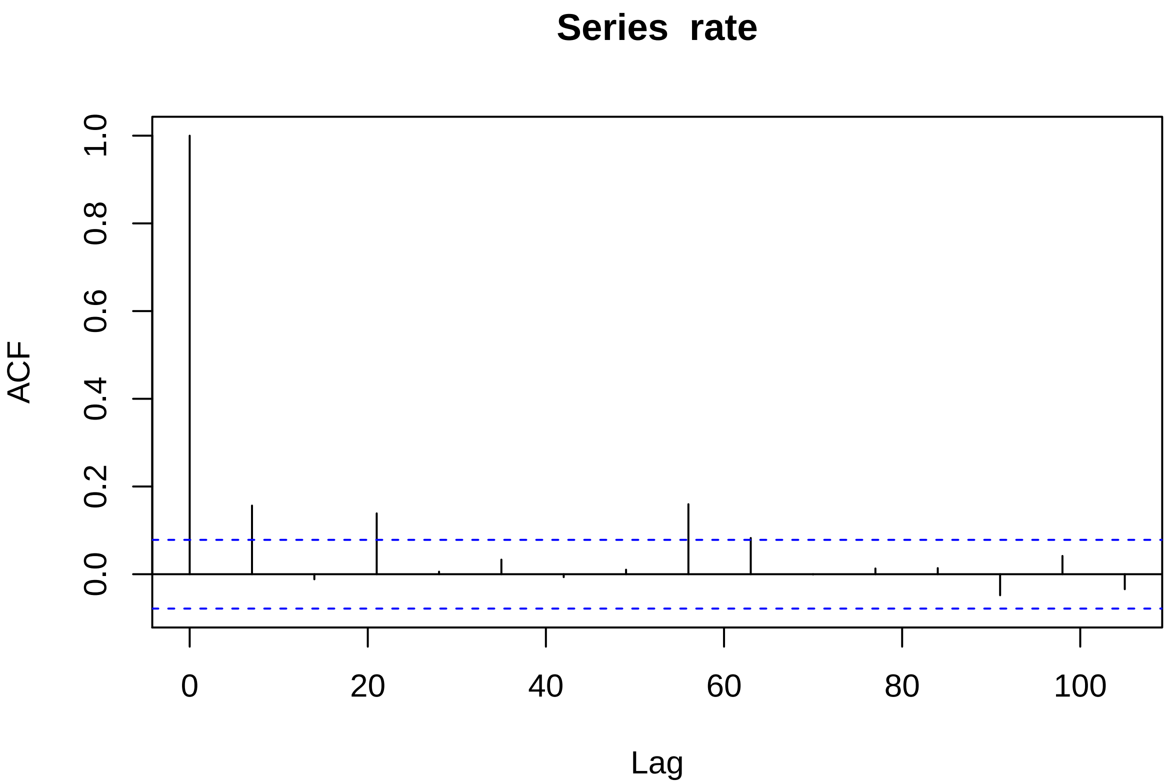
Based on the time plot above, the plot has a fluctuation over the time. I do not see clear seasonality in this plot. It has a more likely random variation. Meanwhile, there is a sudden shift in 2009. This is because of the effect from the financial crisis in 2008.

h) compute and plot the first 15 lags of the autocorrelation function and discuss if the series shows evidence of serial correlation.

```
# print acf to console
acf(rate, lag = 15, plot = F)
```

```
##
## Autocorrelations of series 'rate', by lag
##
##      0      7     14     21     28     35     42     49     56     63
##  1.000  0.157 -0.012  0.139  0.006  0.033 -0.007  0.010  0.160  0.083
##     70     77     84     91     98    105
##  0.000  0.013  0.014 -0.048  0.042 -0.034
```

```
# plot acf
acf(rate, lag = 15, plot = T)
```



The autocorrelation plot shows that most of the bar are not statistically significant, except first, second, fourth, and nineth lag-15. Which means that they are indepedent.

Problem 2

Load the data

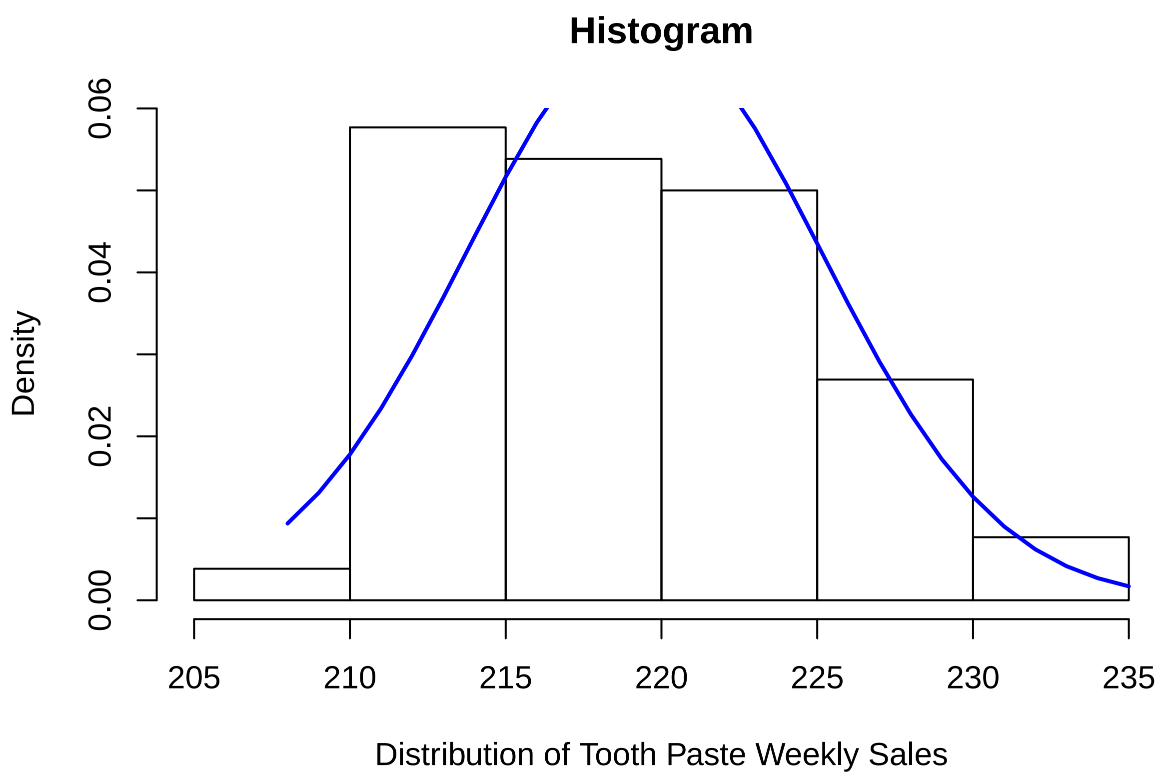
```
groc = read.table("groceries.csv", header = T, sep = ',')
groc
```

```
##          Date ToothPaste PeanutButter Biscuits
## 1    6-Jan-08        224          462      381
## 2   13-Jan-08        235          488      398
## 3   20-Jan-08        226          431      349
## 4   27-Jan-08        226          495      397
## 5    3-Feb-08        222          439      367
## 6   10-Feb-08        215          452      366
## 7   17-Feb-08        221          475      397
## 8   24-Feb-08        213          413      347
## 9    2-Mar-08        213          470      387
## 10   9-Mar-08        220          455      388
## 11 16-Mar-08        223          449      367
## 12 23-Mar-08        214          433      360
## 13 30-Mar-08        216          471      394
## 14  6-Apr-08        223          458      383
## 15 13-Apr-08        221          439      359
## 16 20-Apr-08        219          462      382
```

```
## 17 27-Apr-08        215            443       372
## 18  4-May-08        216            456       377
## 19 11-May-08        220            454       380
## 20 18-May-08        217            441       366
## 21 25-May-08        227            490       403
## 22  1-Jun-08        228            442       365
## 23  8-Jun-08        224            464       372
## 24 15-Jun-08        212            431       363
## 25 22-Jun-08        208            453       380
## 26 29-Jun-08        213            446       380
## 27  6-Jul-08        211            428       361
## 28 13-Jul-08        217            468       392
## 29 20-Jul-08        216            430       364
## 30 27-Jul-08        215            453       374
## 31  3-Aug-08        212            437       372
## 32 10-Aug-08        216            461       385
## 33 17-Aug-08        215            430       363
## 34 24-Aug-08        223            480       396
## 35 31-Aug-08        222            437       366
## 36  7-Sep-08        217            452       369
## 37 14-Sep-08        212            442       374
## 38 21-Sep-08        217            464       388
## 39 28-Sep-08        226            462       384
## 40  5-Oct-08        224            445       364
## 41 12-Oct-08        219            457       376
## 42 19-Oct-08        229            492       406
## 43 26-Oct-08        231            447       364
## 44  2-Nov-08        223            457       367
## 45  9-Nov-08        215            449       375
## 46 16-Nov-08        215            457       381
## 47 23-Nov-08        215            441       370
## 48 30-Nov-08        223            474       393
## 49  7-Dec-08        226            448       371
## 50 14-Dec-08        218            443       362
## 51 21-Dec-08        220            476       395
## 52 28-Dec-08        223            450       374
```

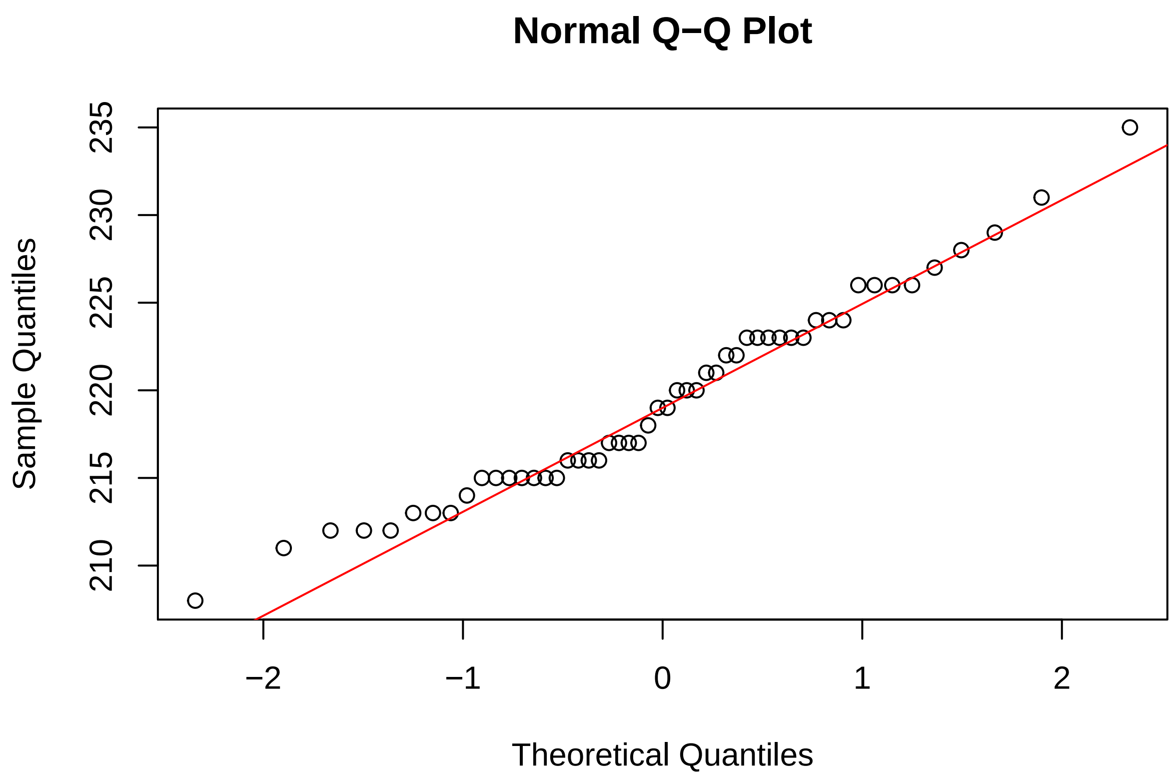
a) Analyze the distribution of ToothPaste weekly sales using a histogram and normal quantile plot.

```r
# histogram
par(mfcol = c(1,1))
hist(groc$ToothPaste, xlab = "Distribution of Tooth Paste Weekly Sales", prob = TRUE, main = "Histogram
xfit <- seq(min(groc$ToothPaste), max(groc$ToothPaste))
yfit <- dnorm(xfit, mean = mean(groc$ToothPaste), sd = sd(groc$ToothPaste))
lines(xfit, yfit, col = "blue", lwd = 2)
```



Based on the result, it is more likely skewed to the right rather than symmetric due to the fact that the tail at the right side is longer.

```
# quantile plot
qqnorm(groc$ToothPaste)
qqline(groc$ToothPaste, col = 2)
```



it is not a perfect normal distribution, but it is very close to the normal distribution.

b) Test the hypothesis of normality for the distribution of ToothPaste weekly sales using the Jarque Bera test at 5% significance level.

```
normalTest(groc$ToothPaste, method = c("jb"))
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 1.6157
##   P VALUE:
##     Asymptotic p Value: 0.4458
##
## Description:
##  Wed Sep 27 14:12:30 2017 by user:
```

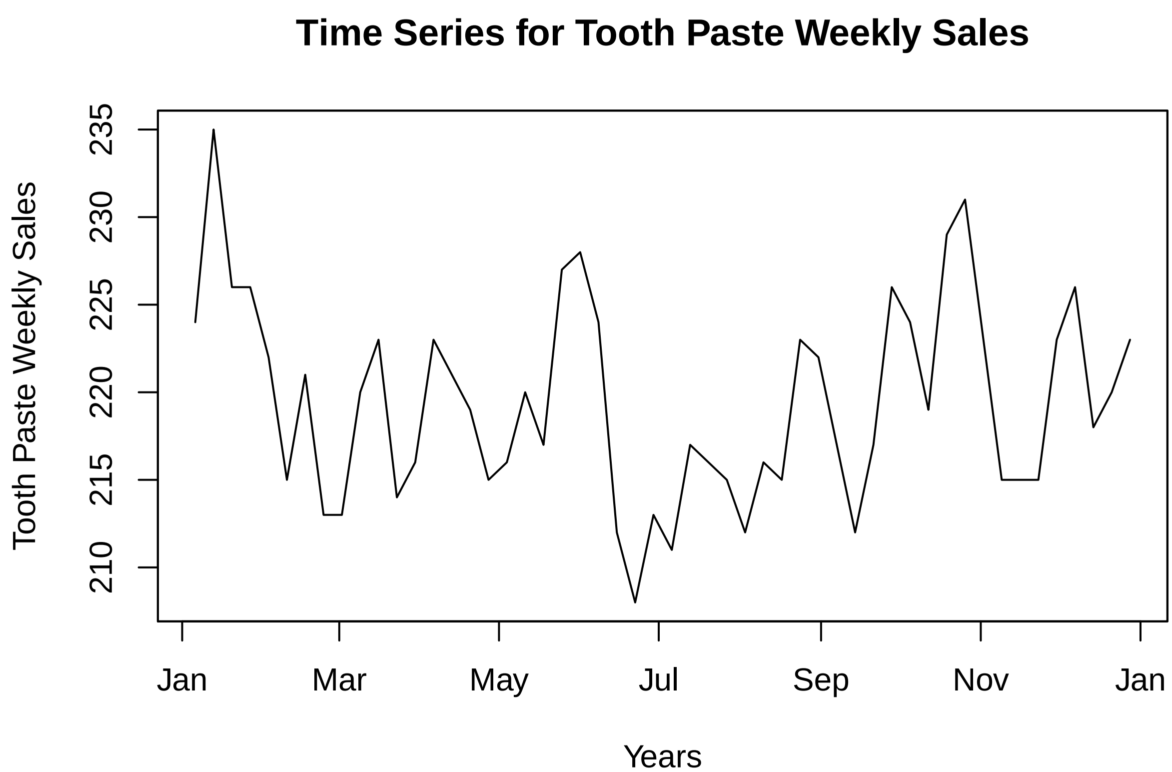
Based on the normal test, p-value is 0.4458 which is much larger than 0.05 so that we fail to reject the hypothesis.

c) Create a time plot for the time series of ToothPaswte weekly sales.

```r
#Time Series
TStooth = zoo(groc[,2], as.Date(as.character(groc[,1]), format = "%d-%b-%y"))
TStooth
```

```
## 2008-01-06 2008-01-13 2008-01-20 2008-01-27 2008-02-03 2008-02-10
##        224        235        226        226        222        215
## 2008-02-17 2008-02-24 2008-03-02 2008-03-09 2008-03-16 2008-03-23
##        221        213        213        220        223        214
## 2008-03-30 2008-04-06 2008-04-13 2008-04-20 2008-04-27 2008-05-04
##        216        223        221        219        215        216
## 2008-05-11 2008-05-18 2008-05-25 2008-06-01 2008-06-08 2008-06-15
##        220        217        227        228        224        212
## 2008-06-22 2008-06-29 2008-07-06 2008-07-13 2008-07-20 2008-07-27
##        208        213        211        217        216        215
## 2008-08-03 2008-08-10 2008-08-17 2008-08-24 2008-08-31 2008-09-07
##        212        216        215        223        222        217
## 2008-09-14 2008-09-21 2008-09-28 2008-10-05 2008-10-12 2008-10-19
##        212        217        226        224        219        229
## 2008-10-26 2008-11-02 2008-11-09 2008-11-16 2008-11-23 2008-11-30
##        231        223        215        215        215        223
## 2008-12-07 2008-12-14 2008-12-21 2008-12-28
##        226        218        220        223
```

```r
#plot
plot(TStooth, xlab = "Years", ylab = "Tooth Paste Weekly Sales", main = "Time Series for Tooth Paste We
```



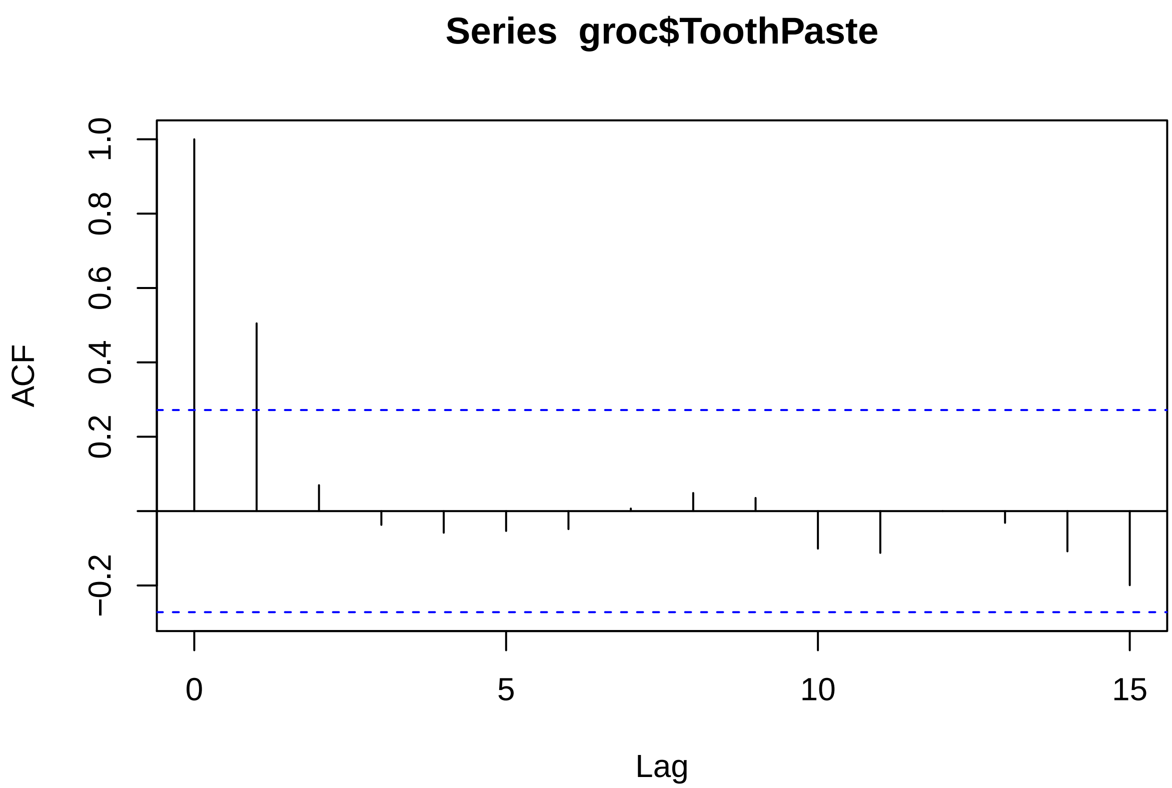
This data shwos that it does not have any patterns. However, I see that biggest sales of toothpaste occured in January, June, and November.

d) Compute and plot the first 15 lags of the autocorrelation function for ToothPaste weekly saels and discuss if the series shows evidence of serial correlation.

```
acf(groc$ToothPaste, lag = 15, plot = F)
```

```
##
## Autocorrelations of series 'groc$ToothPaste', by lag
##
##      0      1      2      3      4      5      6      7      8      9
##  1.000  0.505  0.070 -0.037 -0.058 -0.053 -0.048  0.007  0.049  0.035
##     10     11     12     13     14     15
## -0.101 -0.112  0.000 -0.032 -0.108 -0.199
```

```
acf(groc$ToothPaste, lag = 15, plot = T)
```



The graph shows autocorrelations and all bars falls in the boundaries except for bar at 1 (k=1). Therefore, I can conclude with this is a serieal correlation due to k=1

e) Use the Ljung Box test to hypothesis that ToothPaste weekly sales have significant serial correlation.

```
# To lag 6
Box.test(groc$ToothPaste, lag = 6, type = 'Ljung')
```

```
##
##  Box-Ljung test
##
## data:  groc$ToothPaste
## X-squared = 14.902, df = 6, p-value = 0.02103
```

since the p-value is 0.02103 when lag is 6, so we can conclude that we can reject the hypothesis

```r
# To lag 12
Box.test(groc$ToothPaste, lag = 12, type = 'Ljung')
```

```
##
##  Box-Ljung test
##
## data:  groc$ToothPaste
## X-squared = 16.685, df = 12, p-value = 0.1619
```

The p-value is 0.16 which is larger than 0.05. Thus, fail to reject the hypothesis when lag is 12

f) Discuss in general the importance of weak stationarity for time series analysis, and explain the methode that are used to analyze whether a TS is stationary

The most important thing for weak stationarity for time series is a constant mean and acf (variance). The method that you can detect if TS is starionary is to use acf. ACF depends on lag and does not change with time. The easiest way to check is to plot the time series. The time series is stationary if they do not have trends and seasonality.