HW1

Oluwafemi Shobowale

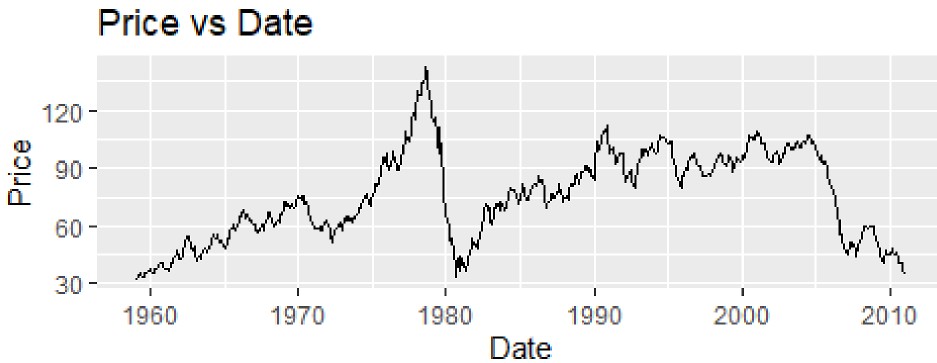
04/10/2022

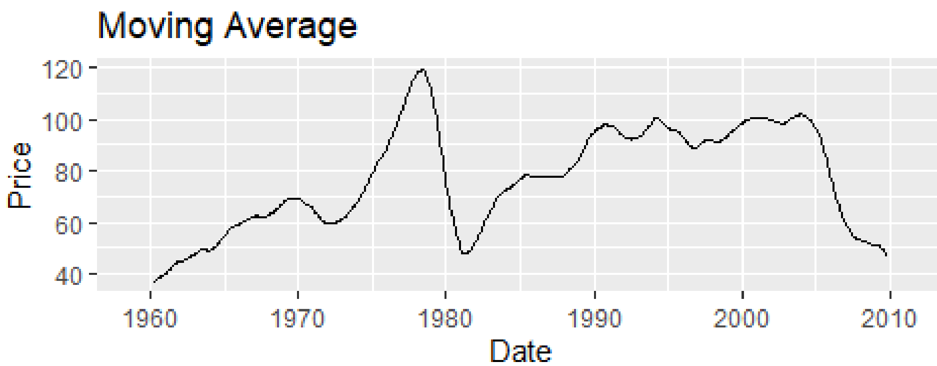
```
library(readxl)
df1<- read.csv("crudeoil.csv")
df1$date = as.Date(df1$date,format = "%d-%b-%y")
df1$date =as.character(df1$date)
b = ts(df1$price,c(2004,01),frequency=12)

g1<-ggplot(df1,aes(x=time(b),y=b,group=1))+geom_line()+labs(title = "Price vs
Date")+xlab("Date")+ylab("Price")
#30-Days  moving average over
b_smoth = ma(b, order = 30 ,centre = TRUE)
g2<-ggplot(df1,aes(x=time(b_smoth),y=b_smoth,group=1))+geom_line()+labs(title
= "Moving Average")+xlab("Date")+ylab("Price")
```

1a,

The plots shows no definite trend either upward or downward over time, there seems to be some sort of unclear trend. The range of random values is greater than the range of values in the seasonal plot, hence there is no seasonality.

```
# set the plotting area into a 1*2 array
library(gridExtra)
# Use grid.arrange to put plots in columns
grid.arrange(g1, g2,ncol=1)
```

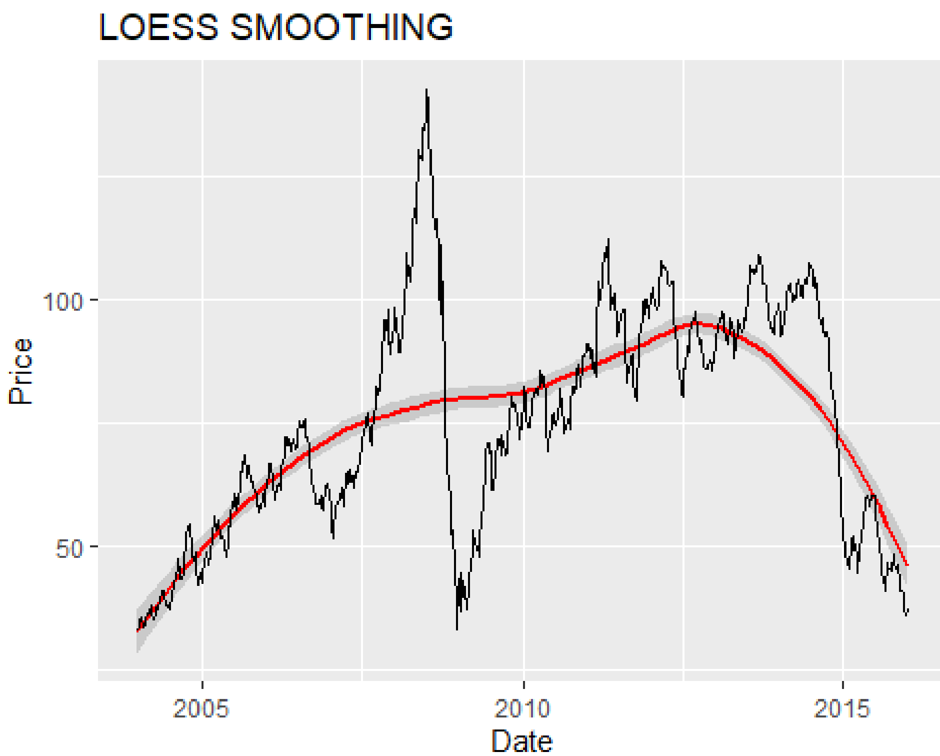




1b,

LOESS shows a more smooth curve than moving average. There seems to be some sort of upward and down trends parabolic in nature.

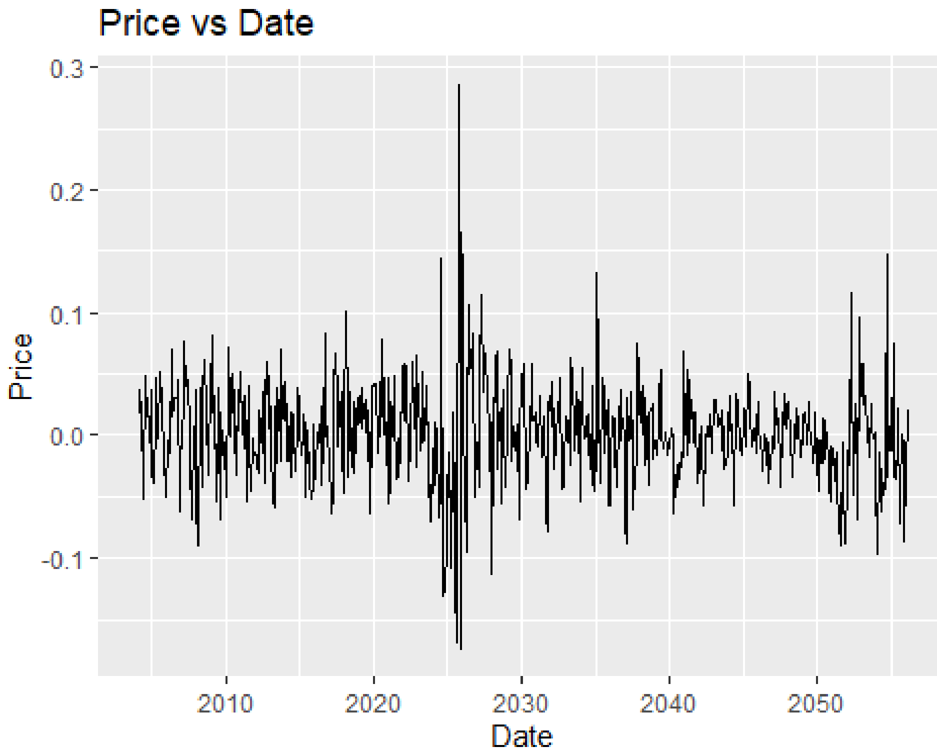
```
df1$date = as.Date(df1$date,'%Y-%m-%d')
ds = data.frame(df1$price,df1$date)
# Remember, it LOESS is the default smoothing for geom_smooth!
ggplot(data=ds, aes(x=df1$date, y=df1$price)) + geom_line() +
geom_smooth(col="red")+geom_line()+labs(title = "LOESS
SMOOTHING")+xlab("Date")+ylab("Price")
```



1c,

Here we can see that the change rate of price hovers around zero.

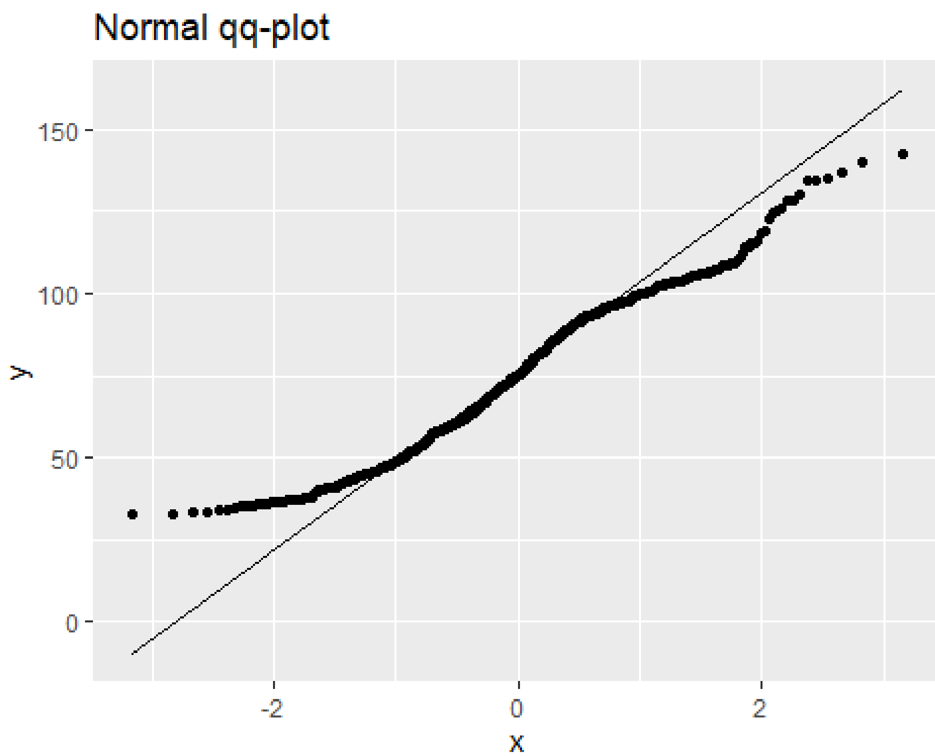
```
library(dplyr)
df1$lag<-lag(df1$price, shift=1)
df1$p_rate<-(df1$price-df1$lag)/df1$lag
ggplot(df1,aes(x=time(b),y=df1$p_rate,group=1))+geom_line()+labs(title =
"Price vs Date")+xlab("Date")+ylab("Price")
```



1d,

Quantile plot shows that the rate or normally distributed, with fat tails on both end. Kurtosis test is ‘excess’ hence normal=0. The jarque bera test since p <0.05 we can assume the rate is normally distributed.

```
ggplot(df1, aes(sample =df1$lag))+stat_qq() + stat_qq_line()+labs(title =
"Normal qq-plot")
```




```
df1=na.omit(df1)
# A fairly high peak and possibly fatter tails, let's look at the kurtosis
library(fBasics)
kurtosis(df1$p_rate)   # Note, this is "excess" kurtosis, so normal = 0

## [1] 4.627621
## attr(,"method")
## [1] "excess"

# If you want regular kurtosis where normal = 3, then set a method
kurtosis(df1$p_rate, method="moment")

## [1] 7.627621
## attr(,"method")
## [1] "moment"

skewness(df1$p_rate)  # Not much in the way of skewness!

## [1] 0.2586747
## attr(,"method")
## [1] "moment"

normalTest(df1$p_rate)

##
## Title:
##  Shapiro - Wilk Normality Test
##
```

```
## Test Results:
##   STATISTIC:
##     W: 0.9572
##   P VALUE:
##     1.635e-12
##
## Description:
##  Sun Apr 10 14:27:58 2022 by user: soboa
```

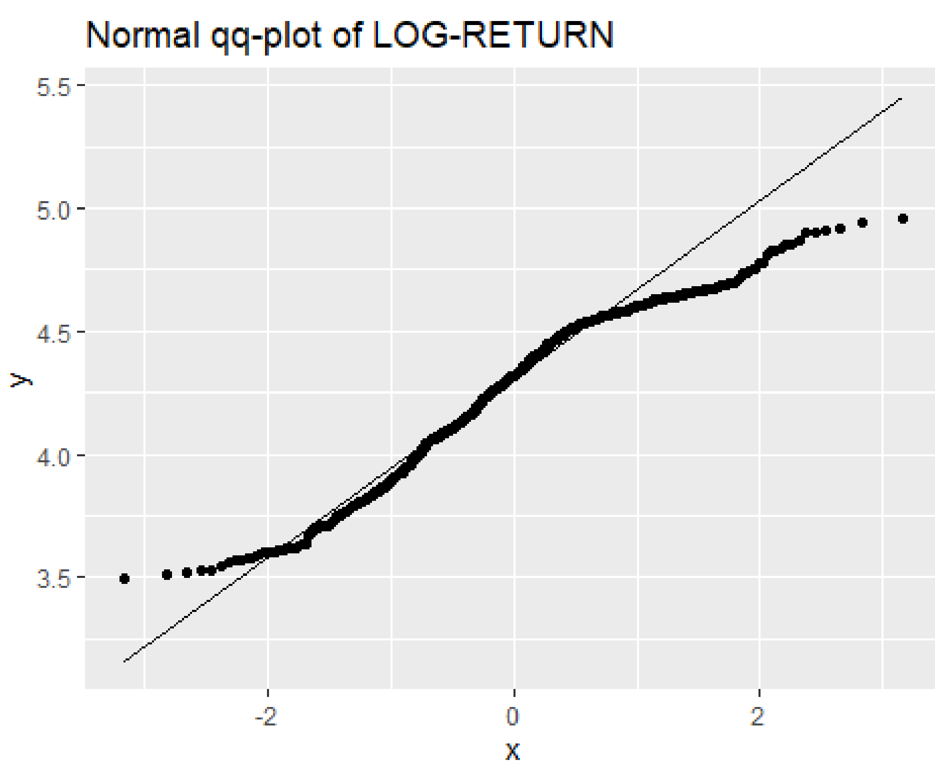
```
library(tseries)
df1=na.omit(df1)
jarque.bera.test(df1$p_rate)
```

```
##
##  Jarque Bera Test
##
## data:  df1$p_rate
## X-squared = 571.5, df = 2, p-value < 2.2e-16
```

1e,

Log transformation does not affect the normality,both the rate and the log-rate are normally distributed, passing the normality test.

```
df1$log_rate= log(df1$price)
ggplot(df1, aes(sample =df1$log_rate))+stat_qq() + stat_qq_line()+labs(title
= "Normal qq-plot of LOG-RETURN")
```



```r
#test normality of log price
library(fBasics)
kurtosis(df1$log_rate)   # Note, this is "excess" kurtosis, so normal = 0

## [1] -0.7395231
## attr(,"method")
## [1] "excess"

# If you want regular kurtosis where normal = 3, then set a method
kurtosis(df1$log_rate, method="moment")

## [1] 2.260477
## attr(,"method")
## [1] "moment"

skewness(df1$log_rate)  # Not much in the way of skewness!

## [1] -0.4314748
## attr(,"method")
## [1] "moment"

normalTest(df1$log_rate)

##
## Title:
##  Shapiro - Wilk Normality Test
##
## Test Results:
##   STATISTIC:
##     W: 0.9595
##   P VALUE:
##     4.078e-12
##
## Description:
##  Sun Apr 10 14:27:58 2022 by user: soboa

library(tseries)
df1=na.omit(df1)
jarque.bera.test(df1$log_rate)

##
##  Jarque Bera Test
##
## data:  df1$log_rate
## X-squared = 33.504, df = 2, p-value = 5.305e-08
```

2

2a,

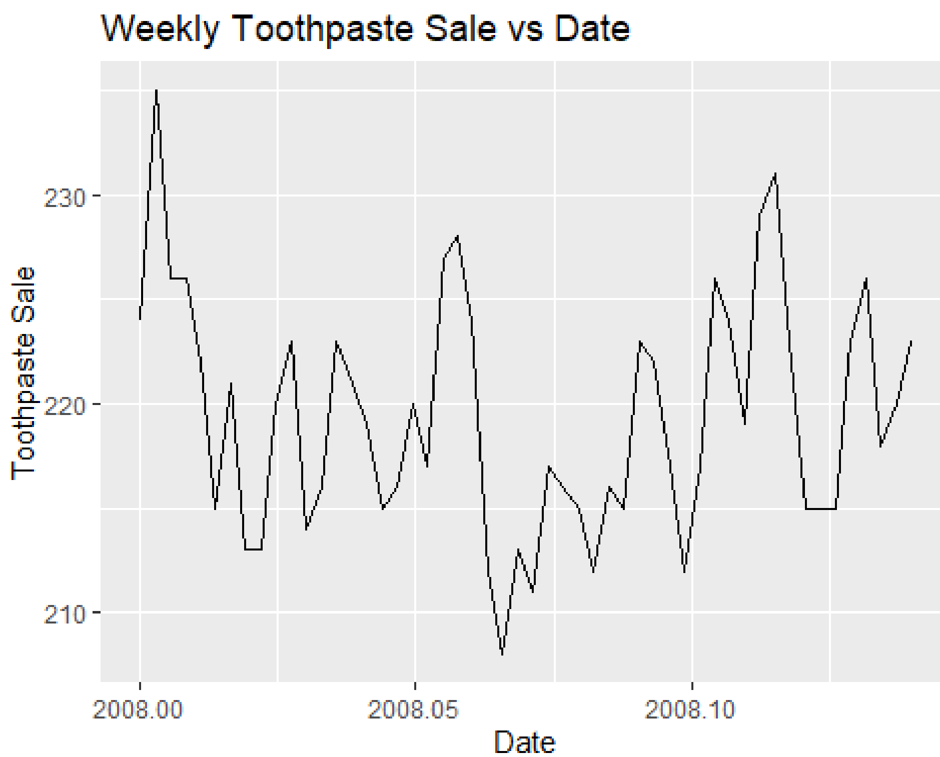
The frequency is set to 365 since the data only covers 1yr.

```r
library(readxl)
df2<- read.csv("groceries.csv")
df2$Date = as.Date(df2$Date,format = "%d-%b-%y")
df2$Date =as.character(df2$Date)
df2_ts = ts(df2$ToothPaste,c(2008,01),frequency=365)
```

2b,

The trend in the below graph shows a combination of downward and upward over time, there seem to be some sort of oscillating periodic effect in the series which suggest weekly seasonality

```r
ggplot(df2,aes(x=time(df2_ts),y=df2_ts,group=1))+geom_line()+labs(title =
"Weekly Toothpaste Sale vs Date")+xlab("Date")+ylab("Toothpaste Sale")
```



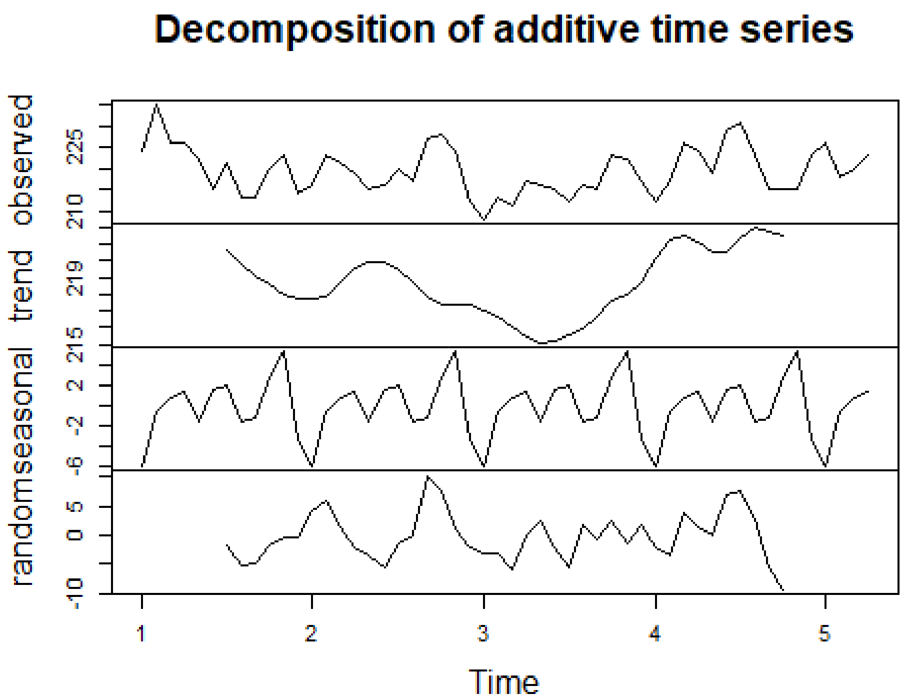
2c,

Series is additive because the series does not increase as the amplitude/value of toothpaste increase.

2d,

The trends shows some sort of oscillation in the series as suggested. There is no seasonality since the range of the random observation is larger than the range of seasonal observation

```r
df2_ts1 = ts(df2$ToothPaste,frequency = 12)
plot(decompose(df2_ts1))
```



3

3a,

ts frequency is 12, since the ts data is shown with 12 months.

```r
library(fpp2)
head(auscafe)
```

```
##         Apr    May    Jun    Jul    Aug    Sep
## 1982 0.3424 0.3421 0.3287 0.3385 0.3315 0.3419
```

```r
head(auscafe,235)
```

```
##         Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct
## 1982                      0.3424 0.3421 0.3287 0.3385 0.3315 0.3419 0.3584
## 1983 0.3686 0.3481 0.3658 0.3511 0.3605 0.3471 0.3645 0.3760 0.3776 0.3741
## 1984 0.3888 0.3771 0.3978 0.3833 0.4140 0.3815 0.3930 0.4090 0.3947 0.4245
## 1985 0.4256 0.3918 0.4161 0.4198 0.4462 0.4068 0.4485 0.4660 0.4552 0.4936
## 1986 0.5043 0.4529 0.4803 0.4972 0.5308 0.4848 0.5263 0.5384 0.5374 0.5704
## 1987 0.5716 0.5251 0.5436 0.5581 0.5651 0.5423 0.5986 0.5837 0.5927 0.6233
## 1988 0.6050 0.5861 0.6253 0.6119 0.6305 0.6355 0.6587 0.6564 0.6600 0.6702
## 1989 0.7328 0.6607 0.7132 0.6939 0.7100 0.7217 0.7413 0.7455 0.7669 0.7937
## 1990 0.8575 0.7644 0.8403 0.8046 0.8089 0.7990 0.8150 0.8280 0.8124 0.8330
## 1991 0.8617 0.7711 0.8133 0.7974 0.8206 0.8013 0.8294 0.8535 0.8820 0.9389
## 1992 0.9383 0.8615 0.9361 0.9320 0.9294 0.8685 0.8913 0.8755 0.9141 0.9357
```

```
## 1993 0.9178 0.8381 0.8698 0.8621 0.8521 0.8276 0.8823 0.8669 0.9049 0.9312
## 1994 0.9845 0.9016 1.0152 0.9391 0.9414 0.9348 1.0133 1.0175 1.0407 1.0618
## 1995 1.0763 0.9816 1.0989 1.0678 1.0833 1.0450 1.0944 1.1097 1.1261 1.1665
## 1996 1.2131 1.1280 1.1800 1.1693 1.1457 1.1090 1.1381 1.1457 1.1049 1.1422
## 1997 1.1801 1.0595 1.1480 1.1413 1.1703 1.1135 1.1646 1.1733 1.1541 1.2048
## 1998 1.1859 1.0500 1.1411 1.1067 1.1442 1.0882 1.1622 1.1448 1.1495 1.2586
## 1999 1.2440 1.1244 1.2454 1.2360 1.2714 1.2080 1.2192 1.2336 1.2613 1.3144
## 2000 1.2974 1.2067 1.3246 1.2525 1.2821 1.2747 1.3179 1.3288 1.4324 1.4608
## 2001 1.4708 1.3380 1.5212 1.4413 1.4461 1.3976 1.4774 1.5046 1.4372 1.4885
##         Nov    Dec
## 1982 0.3747 0.4331
## 1983 0.3906 0.4590
## 1984 0.4365 0.4881
## 1985 0.5056 0.5639
## 1986 0.5638 0.6632
## 1987 0.6219 0.7398
## 1988 0.6708 0.7739
## 1989 0.8075 0.9427
## 1990 0.8515 0.9457
## 1991 0.9538 1.0483
## 1992 0.9318 1.0458
## 1993 0.9630 1.0995
## 1994 1.0702 1.1883
## 1995 1.1943 1.3147
## 1996 1.1294 1.2308
## 1997 1.1837 1.3067
## 1998 1.2008 1.3102
## 1999 1.3183 1.4553
## 2000 1.4163 1.5386
## 2001
```

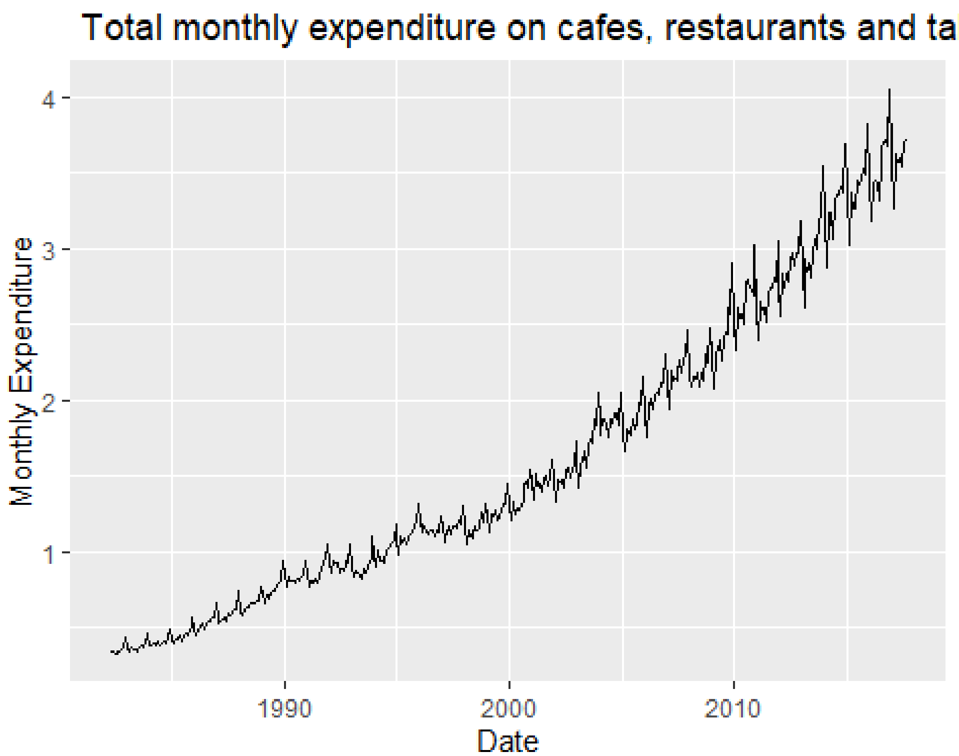
```r
frequency(auscafe)
```

```
## [1] 12
```

3b,

The seems to be a exponential trend in the series, seasonality also occurs in the series monthly.Series is multiplicative Since series tends to increase exponentially it exhibits.

```r
ggplot(auscafe,aes(x=time(auscafe),y=auscafe,group=1))+geom_line()+labs(title
= " Total monthly expenditure on cafes, restaurants and takeout food services
in Australia ")+xlab("Date")+ylab("Monthly Expenditure")
```

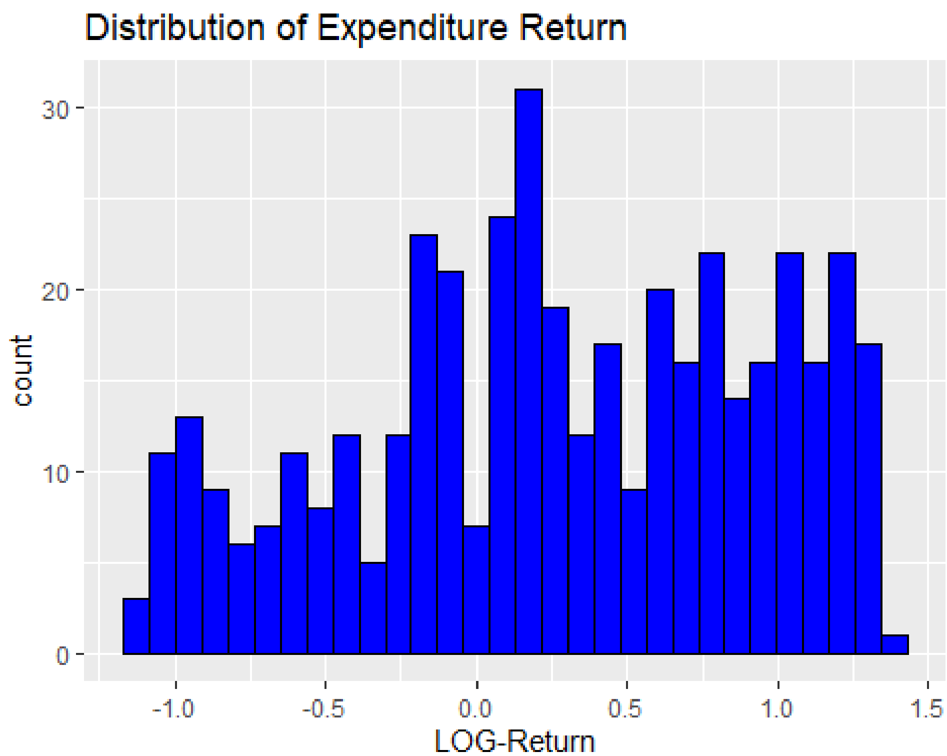


3c,

The plots shows that the log-return is not normally distributed. Distribution is non-symmetric. Kurtosis is -0.9 hence not bell shaped.

```
ggplot(auscafe, aes(sample =log(auscafe)))+stat_qq() +
stat_qq_line()+labs(title = "Normal qq-plot of LOG-RETURN")
```


```r
# And their histogram
ggplot(auscafe, aes(log(auscafe))) + geom_histogram(col="black", fill="blue")
+
  xlab("LOG-Return") + ggtitle("Distribution of Expenditure Return")
```




```r
# A fairly high peak and possibly fatter tails, let's look at the kurtosis
library(fBasics)
kurtosis(log(auscafe))

## [1] -0.9011837
## attr(,"method")
## [1] "excess"
```

3d,

The normal test and jarque bera rejects normality.

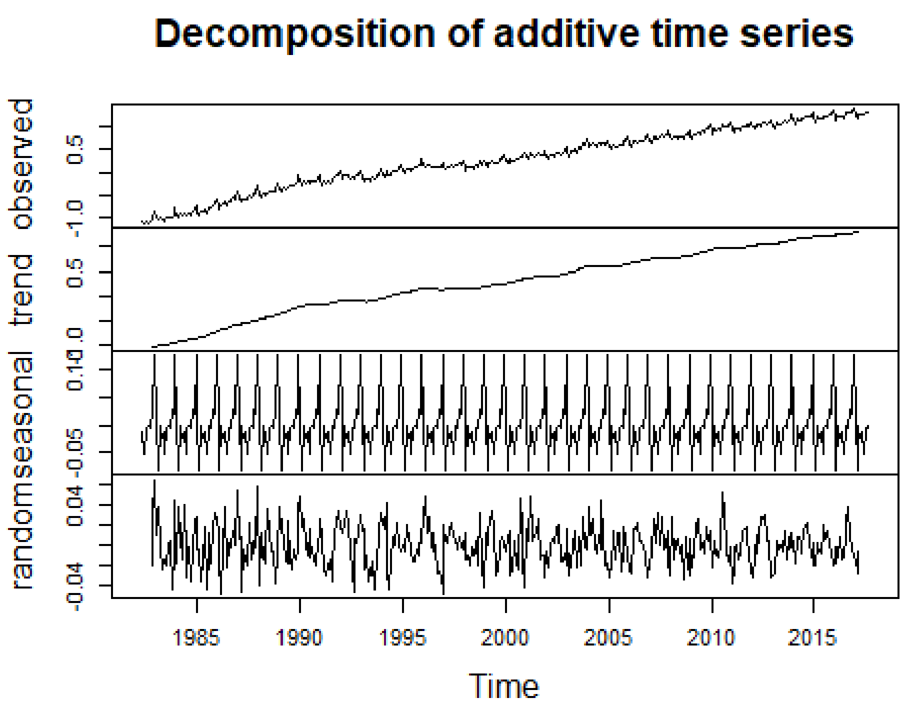
```r
normalTest(auscafe)

##
## Title:
##  Shapiro - Wilk Normality Test
##
## Test Results:
##   STATISTIC:
##     W: 0.9232
##   P VALUE:
##     5.984e-14
##
## Description:
##  Sun Apr 10 14:27:59 2022 by user: soboa
```

```r
library(tseries)

jarque.bera.test(auscafe)

##
##  Jarque Bera Test
##
## data:  auscafe
## X-squared = 37.505, df = 2, p-value = 7.175e-09
```

3e,

The seem to be some sort of weekly seasonality present in the series also there is a linear trend increasing upward over time.

```r
plot(decompose(log(auscafe)))
```



4

4a,

Seasonality is important to understand beacuse it indicated the fluctuation of demand or supply due to any variable/factor of interest.

4b,

The author breaks down the negative consequences of seasonal pattern into three categories: employment, investment and environment. Seasonality affects the employment market since more jobs are available during peak seasons and less jobs are available off-peak seasons. Investments are affected based on the return of capital, this means that seasonality causes a low annual return.Lastly, seasonality causes over crowing or overuse during peak season which affects the environment.

4c,

Seasonality in tourism can also be caused by social media. A spike or constant report of a location or attraction can cause seasonality. One vital reason to study this can be to boost the economy and also control the cause and effect of social media on seasonality in tourism.

4d,

Seasonality in a monitoring system like a heart beat monitor relays on seasonality. The heartbeat has a periodic rhythm over time. The most important reason why seasonality is useful to be studied in a monitoring system is to detect abnormality in breathing pattern. Since we expect a seasonal response from the sensor when picking signals from the chest wall we can monitor or investigate when an heartbeat is off which can early detect heard diseases.