THE UNIVERSITY OF CHICAGO

Time Series Analysis and Forecasting: Assignment 1

Submitted by: Akshay Chhabra, James Foster and Susan Parker
Professor: Vadim Sokolov
TA: Danny Ng
Submitted on: July 13, 2015

Problem 1:
A classic example of a non-stationary series is the daily closing stock prices. Download a OHLC data of a stock from Google Finance. You can change a ticker to your favorite stock. Plot the daily closing prices for the stock and the ACF.Explain how each plot shows the series is non-stationary and should be differenced.
Answer:

Get the data- We will read the data for two stocks- Google and Walmart

```
getSymbols("GOOGL",src='google')
```

```
##      As of 0.4-0, 'getSymbols' uses env=parent.frame() and
##   auto.assign=TRUE by default.
##
##   This  behavior  will be  phased out in 0.5-0  when the call  will
##   default to use auto.assign=FALSE. getOption("getSymbols.env") and
##   getOptions("getSymbols.auto.assign") are now checked for alternate defaults
##
##   This message is shown once per session and may be disabled by setting
##   options("getSymbols.warning4.0"=FALSE). See ?getSymbols for more details.
```

```
## [1] "GOOGL"
```

```
getSymbols("WMT",src='google')
```

```
## [1] "WMT"
```

Next let's find the daily close price of the stocks and the Autocorrelation function using the ACF R function

```
A = Cl(GOOGL)
B=Cl(WMT)
acf.Google<-acf(A ,type="correlation",lag.max = 60, plot = F)
acf.Walmart<-acf(B ,type="correlation",lag.max = 60, plot = F)
```

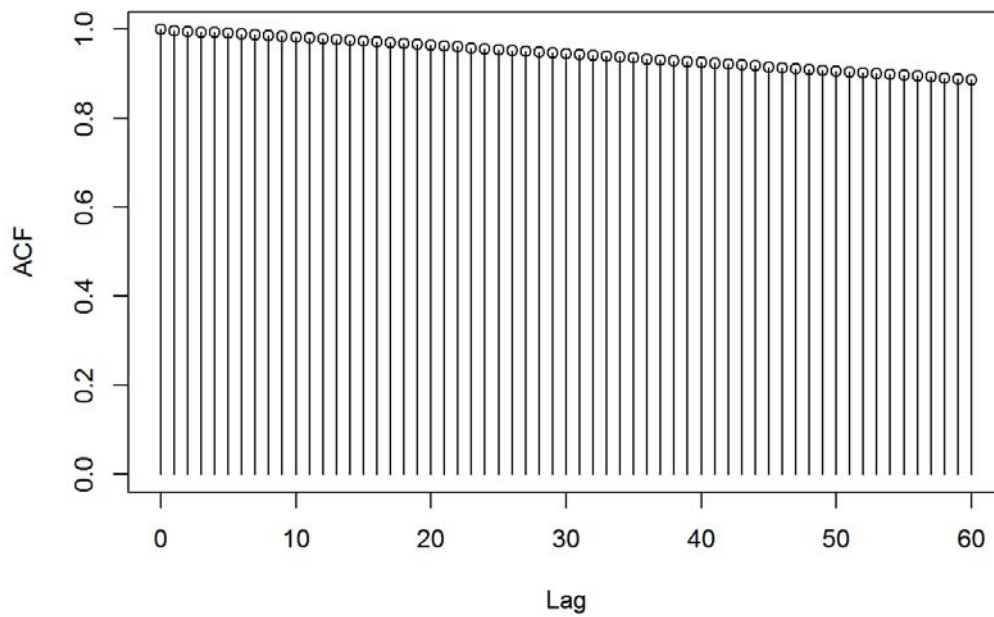### Finally, lets plot the daily close prices and the ACF

```
plot(A,main="Google")
```

Google


```
plot(x=0:60,y=acf.Google$acf,type='p',main = "Google",xlab="Lag", ylab="ACF",lwd=1, ylim=c(0,1))
lines(acf.Google$lag,acf.Google$acf,type='h')
```
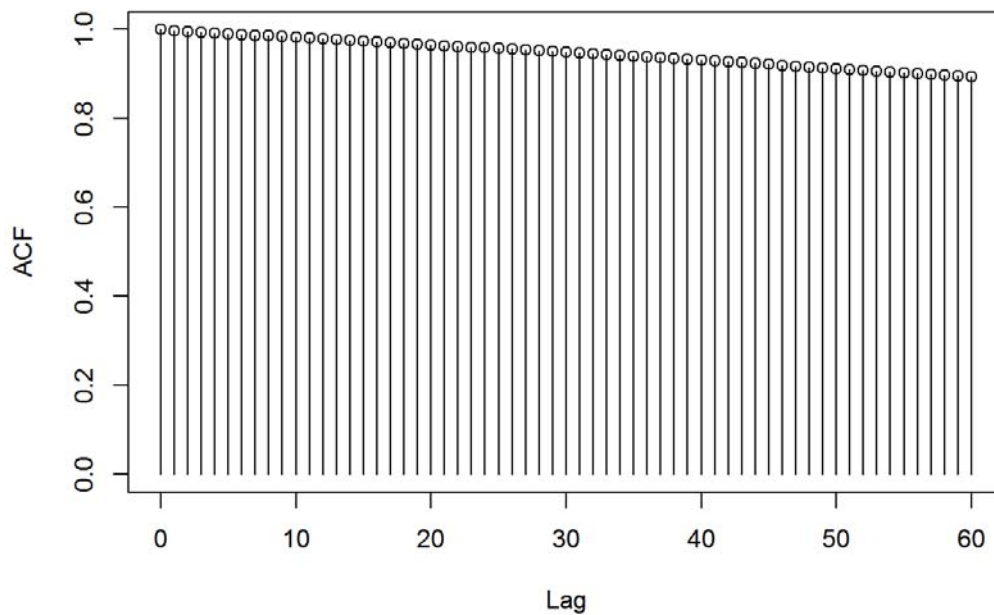
Google


```
plot(B,main="Walmart")
```

Walmart


```
plot(x=0:60,y=acf.Walmart$acf,type='p',main="Walmart", xlab="Lag", ylab="ACF",lwd=1, ylim=c(0,1))
lines(acf.Google$lag,acf.Walmart$acf,type='h')
```

Walmart

Analysis:

- **Closing price plot-** In the plots of the closing prices for both Google and Walmart, there seems to be visible trend and a seasonality elements. The stocks prices for both these companies have gone up in the past few years, which explain the upward trend. The seasonality could be beacuse of the quarterly reports. Due to the existence of the trend and seasonality components, the first two moments are not constant and thus both the series are not stationary.
- **ACF plot-** The ACF plots of both the stocks show that there is correlation between the lags. Even lags till 60 are affecting the current stock price.Thus, the series of the closing prices is not stationary


Problem 2:

Plot the ACFs for 36 random numbers, 360 random numbers and for 1,000 random numbers.Explain the differences among these figures.

Do they all indicate the data are white noise? Why are the critical values at different distances from the mean of zero? Why are the autocorrelations different in each figure when they each refer to white noise?
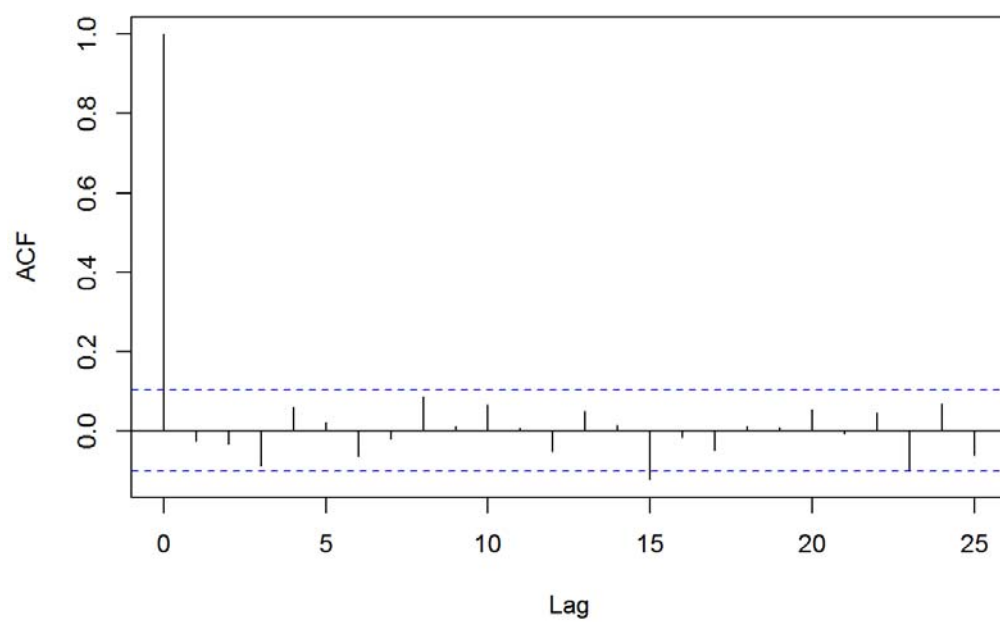**Answer:**

The blue lines in the below plots are the bounds for statistical significance.
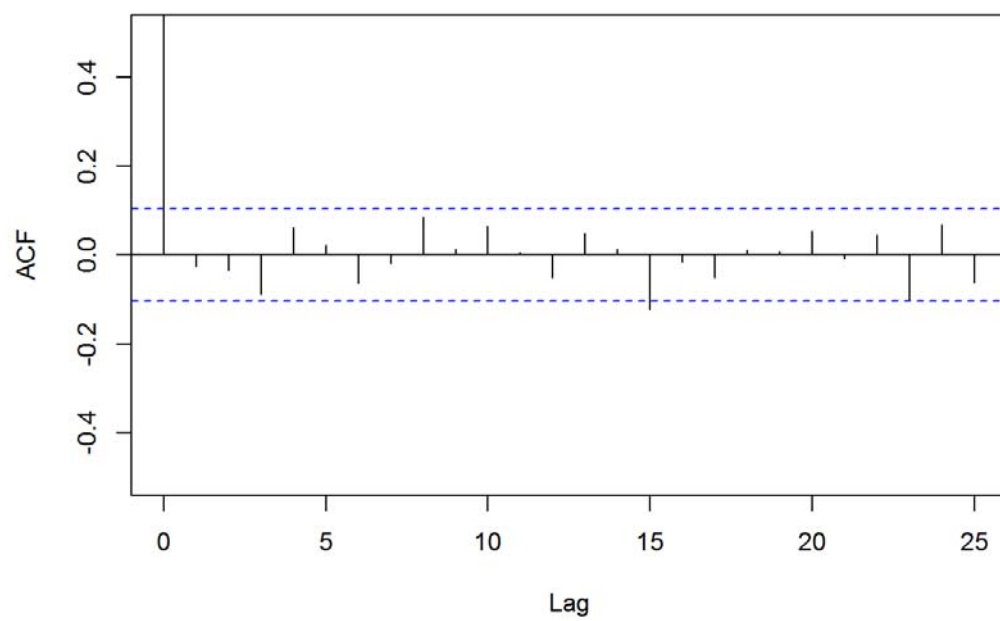
```
plot(acf(rnorm(36)),ylim=c(-0.5,0.5))
```

Series rnorm(36)
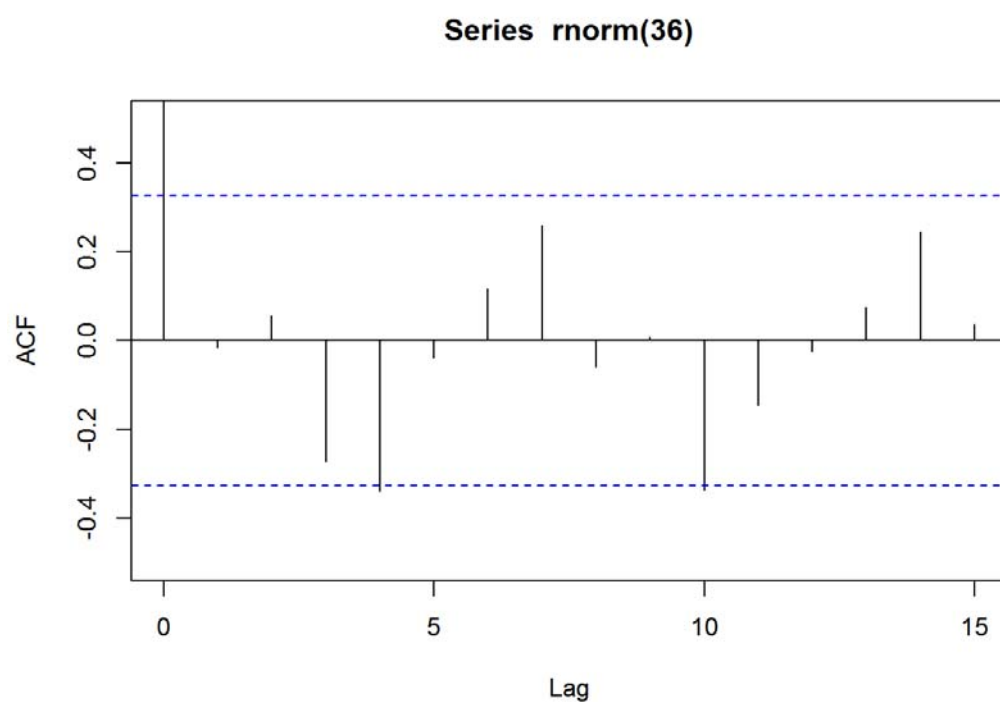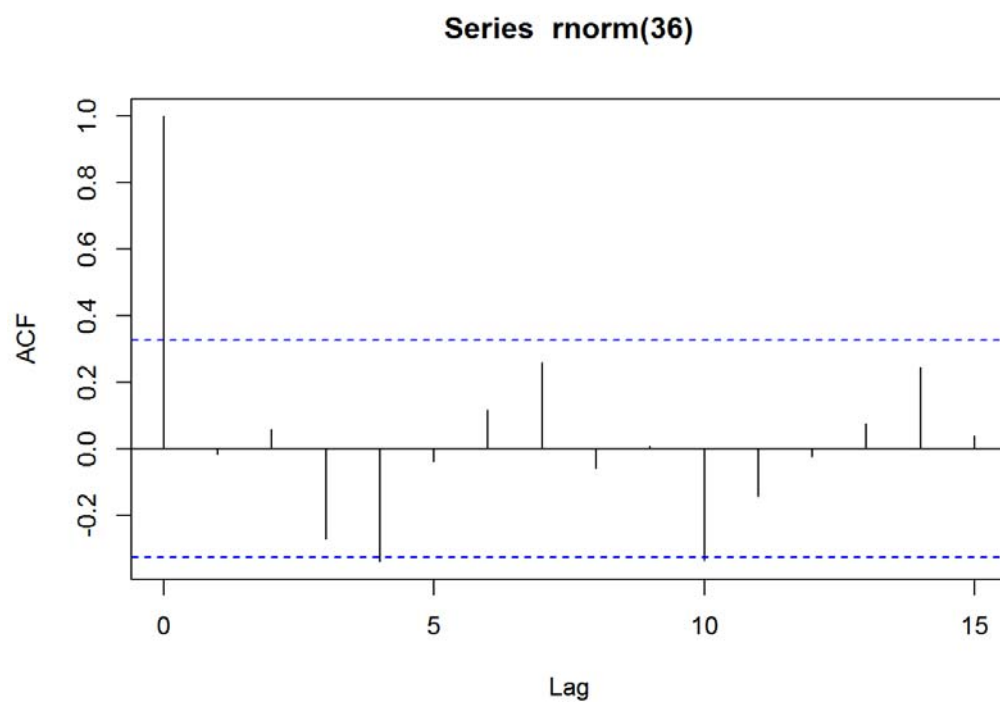
Series rnorm(36)


```
plot(acf(rnorm(360)),ylim=c(-0.5,0.5))
```
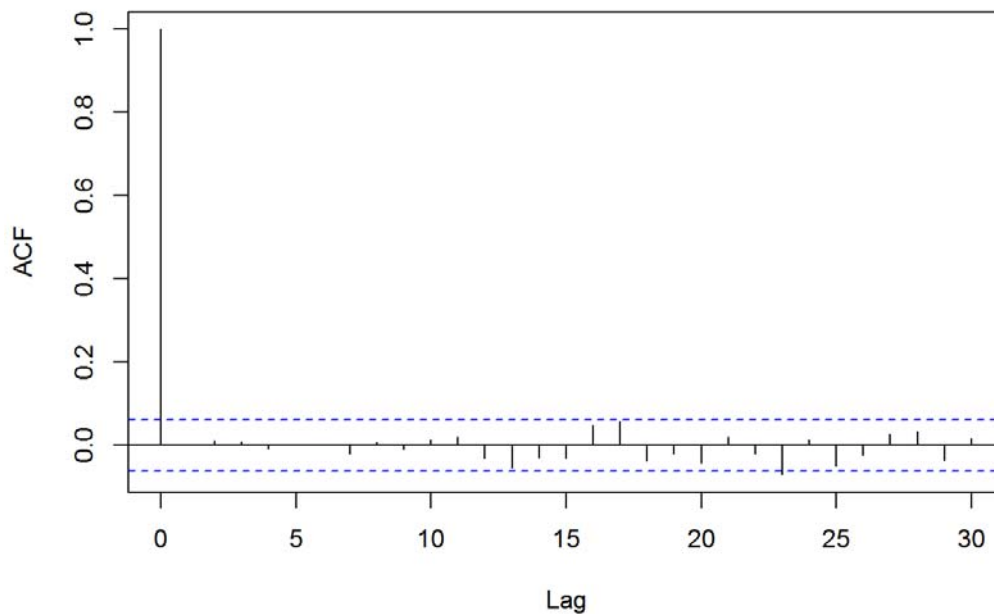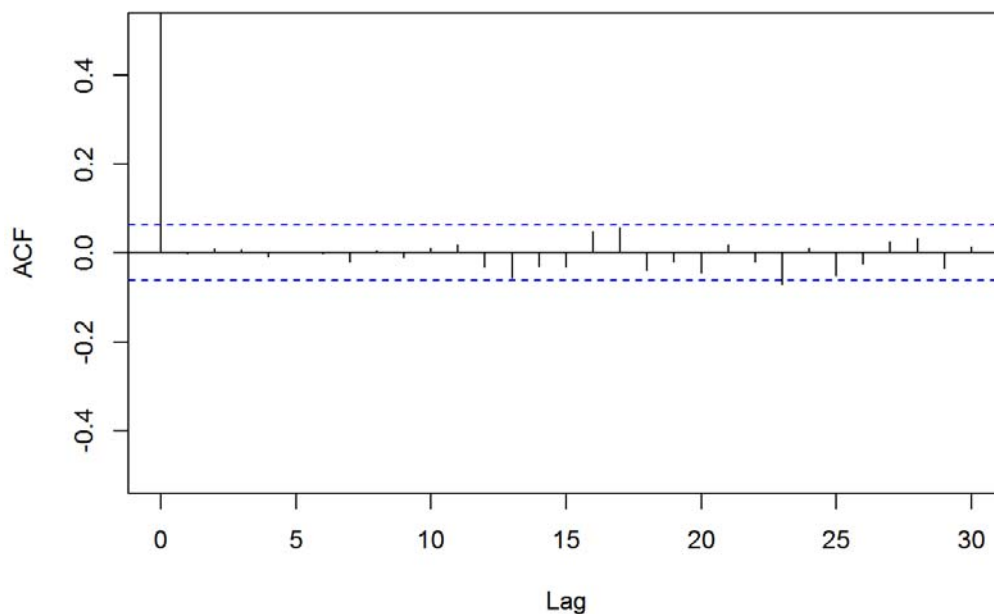
Series rnorm(360)

Series rnorm(360)


```
plot(acf(rnorm(1000)),ylim=c(-0.5,0.5))
```

Series rnorm(1000)

Series rnorm(1000)

Analysis:

With the exception of lag 0, which is always 1 by definition, almost all of the autocorrelations fall within the confidence limits. In addition, there is no apparent pattern. This is the absence of a pattern we expect to see if the data are in fact random. A few lags slightly outside the confidence limits do not neccessarily indicate non-randomness. Thus the data are white noise. With a smaller sample size, it is easy to misunderstand the plot given that the correlations seem high. As we increase the sample size, the correlation and thus the confidence intervals decrease. It becomes clearer that the correlations are due to randomness. Critical values are based up the number of records accessed with more records corresponding to tighter limits. For each data set, the autocorrelations are different because each data set is an independent draw of random numbers.

Problem 3:

Google maintains records of the popularity of search terms over time at http://www.google.com/trends (http://www.google.com/trends). You can examine daily data by selecting a time period of less than 90 days, otherwise you'll see weekly or monthly data. The numbers reported are relative search volume (you can read more about it at https://support.google.com/trends/answer/4355164?hl=en&rd=1 (https://support.google.com/trends/answer/4355164?hl=en&rd=1). Your task is to find:

1. one search term dominated by a weekly or annual seasonal pattern
2. one search term dominated by an increasing trend

3. one search term that would be poorly described by a combination of trend and seasonality

In each case, you should download the relevant data as a .csv file (click on the settings in the top right corner on the trends page), read the data into R and produce a plot of the data.

Answer:

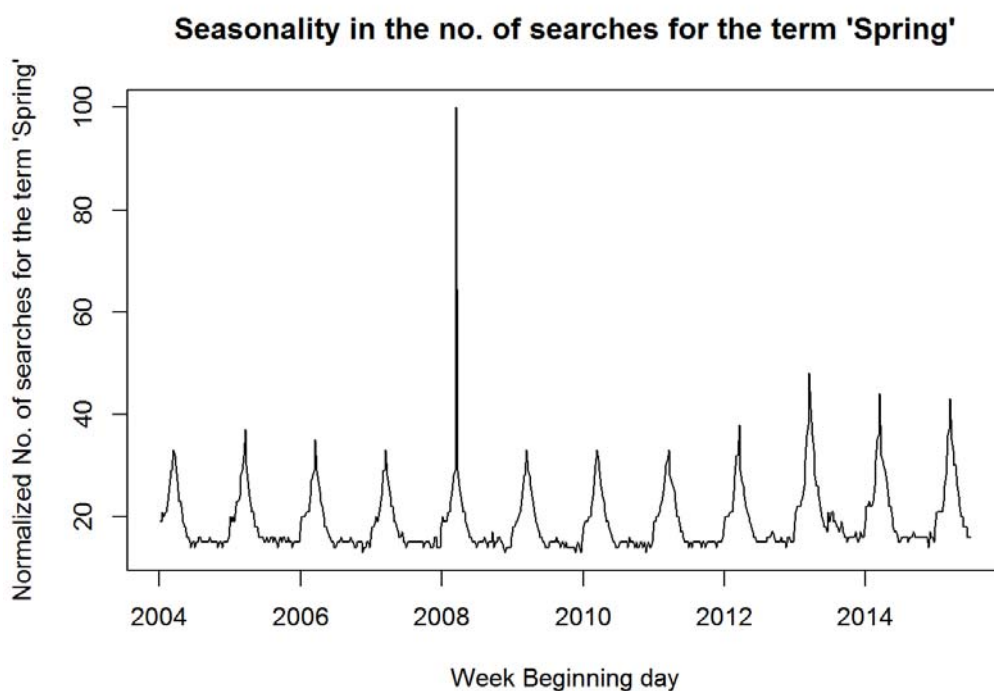1. **One search term dominated by a weekly or annual seasonal pattern**
   Read the data

```
seasonal<-read.csv("Seasonality- spring- report.csv",sep=",",skip = 4,stringsAsFactors = F,nrows = 600)
head(seasonal)
```

```
##                     Week spring
## 1 2004-01-04 - 2004-01-10     19
## 2 2004-01-11 - 2004-01-17     19
## 3 2004-01-18 - 2004-01-24     21
## 4 2004-01-25 - 2004-01-31     20
## 5 2004-02-01 - 2004-02-07     21
## 6 2004-02-08 - 2004-02-14     21
```

Extract the week beginning day from the data and then plot the normalized number of searches for the term "Spring"

```
seasonal$week.begin=as.vector(substr(seasonal$Week,1,10))
plot(parse_date_time(seasonal$week.begin,"%y %m %d"),seasonal$spring,data=seasonal,type="l",xlab="Week Beginning d
ay",ylab="Normalized No. of searches for the term 'Spring'",main="Seasonality in the no. of searches for the term
'Spring'")
```


The seasonality in the data is clearly visible.

2. **One search term dominated by an increasing trend**
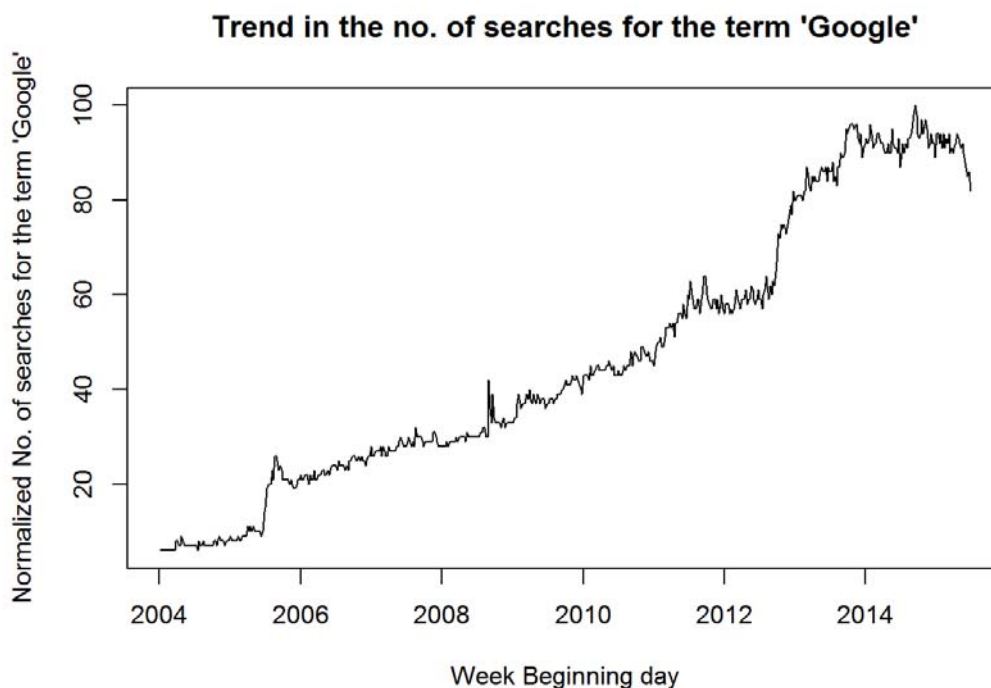   **Read the data**

```
Trend<-read.csv("Trend- Google- report.csv",sep=",",skip = 4,stringsAsFactors = F,nrows = 600)
head(Trend)
```

```
##                     Week google
## 1 2004-01-04 - 2004-01-10      6
## 2 2004-01-11 - 2004-01-17      6
## 3 2004-01-18 - 2004-01-24      6
## 4 2004-01-25 - 2004-01-31      6
## 5 2004-02-01 - 2004-02-07      6
## 6 2004-02-08 - 2004-02-14      6
```

Extract the week beginning day from the data and then plot the normalized number of searches for the term "Google"

```
Trend$week.begin=as.vector(substr(Trend$Week,1,10))
plot(parse_date_time(Trend$week.begin,"%y %m %d"),Trend$google,type="l",xlab="Week Beginning day",ylab="Normalized
No. of searches for the term 'Google'",main="Trend in the no. of searches for the term 'Google'")
```


The increasing trend in the data is clearly visible.

3. **One search term that would be poorly described by a combination of trend and seasonality**
   This question seems to be ambiguous to me. So I have plotted two graphs
   1. The plot of a term that can be explained by both a trend and seasonality
   2. The plot of a term that cannot be explained by a trend or seasonality

Read the data

```
Trend.Seasonal<-read.csv("NBA- report.csv",sep=",",skip = 4,stringsAsFactors = F,nrows = 600)
head(Trend.Seasonal)
```

```
##                         Week nba
## 1 2004-01-04 - 2004-01-10   25
## 2 2004-01-11 - 2004-01-17   27
## 3 2004-01-18 - 2004-01-24   27
## 4 2004-01-25 - 2004-01-31   26
## 5 2004-02-01 - 2004-02-07   25
## 6 2004-02-08 - 2004-02-14   30
```
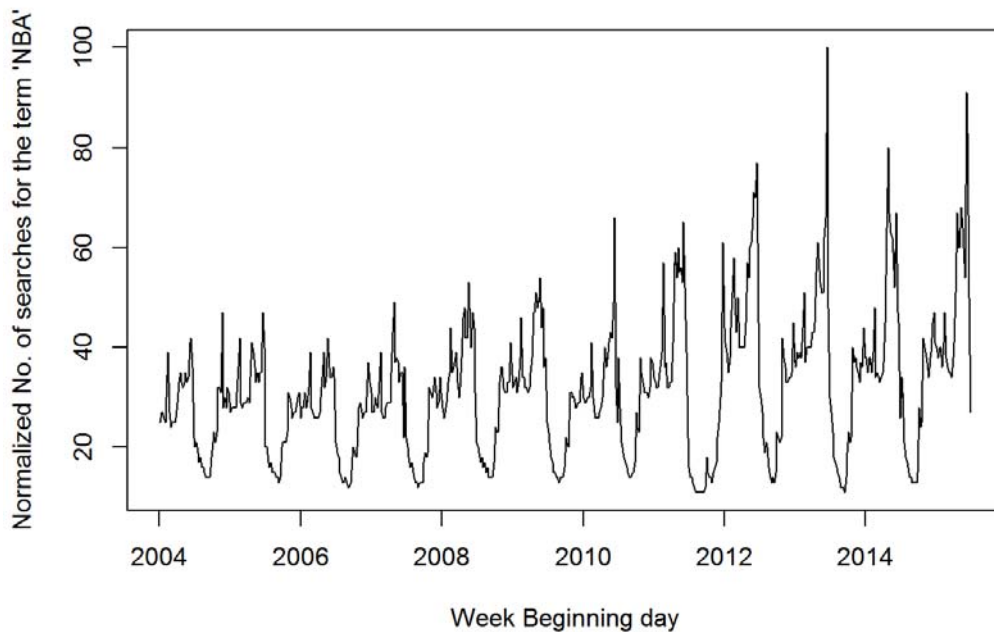
Extract the week beginning day from the data and then plot the normalized number of searches for the term "NBA"

```
Trend.Seasonal$week.begin=as.vector(substr(Trend.Seasonal$Week,1,10))
plot(parse_date_time(Trend.Seasonal$week.begin,"%y %m %d"),Trend.Seasonal$nba,type="l",xlab="Week Beginning day",y
lab="Normalized No. of searches for the term 'NBA'",main="Trend in the no. of searches for the term 'NBA'")
```

Trend in the no. of searches for the term 'NBA'

The increasing trend and seasonality during NBA season in the data is clearly visible.

Now lets look at a plot that doesn't seem to have seasonality or trend- This is a plot of the search term "Obama"

Read the data

```
Trend.Seasonal<-read.csv("Obama.csv",sep=",",skip = 4,stringsAsFactors = F,nrows = 600)
head(Trend.Seasonal)
```

```
##                        Week obama
## 1 2004-01-04 - 2004-01-10     0
## 2 2004-01-11 - 2004-01-17     0
## 3 2004-01-18 - 2004-01-24     0
## 4 2004-01-25 - 2004-01-31     0
## 5 2004-02-01 - 2004-02-07     0
## 6 2004-02-08 - 2004-02-14     0
```

Extract the week beginning day from the data and then plot the normalized number of searches for the term "Obama"

```
Trend.Seasonal$week.begin=as.vector(substr(Trend.Seasonal$Week,1,10))
plot(parse_date_time(Trend.Seasonal$week.begin,"%y %m %d"),Trend.Seasonal$obama,type="l",xlab="Week Beginning da
y",ylab="Normalized No. of searches for the term 'Obama'",main="Trend in the no. of searches for the term 'Obam
a'")
```

Trend in the no. of searches for the term 'Obama'

Week Beginning day
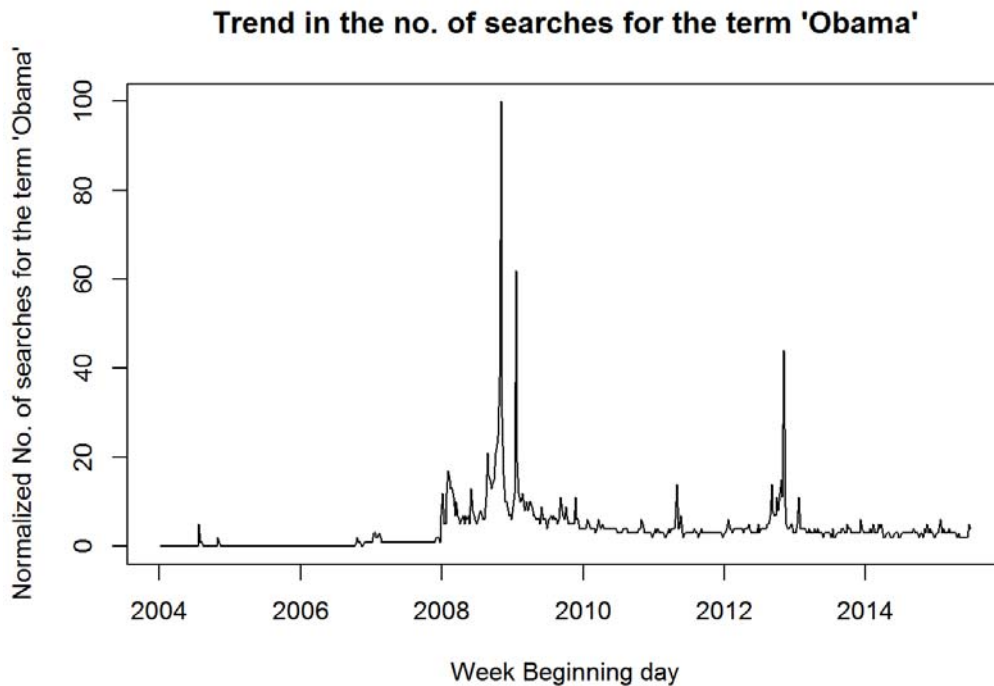
There doesn't seem to be any seasonality or trend component.

Problem 4:

(Shumway 1.6) Consider the time series $x_t = \beta_1 + \beta_2 t + w_t$ where $\beta_1 \, and \, \beta_2$ are known constants and w_t is the white noise proces with variance σ_w^2

1. Determine whether x_t is stationary.
2. Show that the process $y_t = x_t - x_{t-1}$ is stationary

Answer:

1. **Determine whether x_t is stationary.**

   In order for x_t to be stationary, its statistical properties such as its mean should be constant over time.

   $E(x_t) = E(\beta_1 + \beta_2 t + w_t)$

   Since $E(w_t) = 0$, we get

   $E(x_t) = \beta_1 + \beta_2 t$

   Since the mean depends on time t, x_t is not stationary

2. **Show that the process $y_t = x_t - x_{t-1}$ is stationary**

   First lets check for the mean

   $E(y_t) = E(x_t) - E(x_{t-1})$

   $= \beta_1 + \beta_2 t- = \beta_1 - \beta_2(t-1) = \beta_2$

   The mean is indpendent of time

   Next lets check for the autocovariance

   $cov(y_{t+h}, y_t)$

   $= E(y_{t+h} y_t) - E(y_{t+h})E(y_t)$

   $= E((x_{t+h} - x_{t+h-1})(x_t - x_{t-1})) - E(x_{t+h} - x_{t+h-1})E(x_t - x_{t-1})$

   $= E((\beta_2(t+h) + w_{t+h} - \beta_2(t+h-1) - w_{t+h-1})(\beta_2(t) + w_t - \beta_2(t-1) - w_{t-1})) - $
   $\quad E(\beta_2(t+h) + w_{t+h} - \beta_2(t+h-1) - w_{t+h-1})E(\beta_2(t) + w_t - \beta_2(t-1) - w_{t-1})$

   $= E((\beta_2 + w_{t+h} - w_{t+h-1})(\beta_2 + w_t - w_{t-1})) - E(\beta_2 + w_{t+h} - w_{t+h-1})E(\beta_2 + w_t - w_{t-1})$

   We know that $E(w_t) = 0 \, and \, E(w_t^2 = \sigma^2)$. Thus, we get

   $= \beta_2^2 + E(w_{t+h}w_t) - E(w_{t+h}w_{t-1}) - E(w_{t+h-1}w_t) + E(w_{t+h-1}w_{t-1}) - \beta_2^2$

   $= E(w_{t+h}w_t) - E(w_{t+h}w_{t-1}) - E(w_{t+h-1}w_t) + E(w_{t+h-1}w_{t-1})$

   Thus, $cov(y_{t+h}, y_t)$

   $= \begin{cases} 2\sigma^2 & if \ h = 0 \\ -\sigma^2 & if \ h = 1 \\ 0 & Otherwise \end{cases}$

   This shows that the autocovariance is independent of time.

   Now, since the mean and the autocovariance are independent of time, y_t is stationary.

Problem 5:

Let x_t be a stationary time series, and define

$$y_t = \begin{cases} x_t & for\ t\ odd \\ x_t + 3 & for\ t\ even \end{cases}$$

1. Show that $cov(y_t, y_{t+h})$ does not depend on t
2. Is y_t stationary?

Answer:

Since x_t is stationary, we know that the $cov(x_{t+h}, x_t)$ is independent of time t (This is a requirement of stationarity)

1. $cov(y_{t+h}, y_t) = E(y_{t+h}y_t) - E(y_{t+h})E(y_t)$

    1. When t is odd
       $cov(y_{t+h}, y_t)$
       $= E(y_{t+h}y_t) - E(y_{t+h})E(y_t)$
       $= E(x_{t+h}x_t) - E(x_{t+h})E(x_t)$
       $= cov(x_{t+h}, x_t)$
       Which we know is independent of time t
    2. When t is even
       $cov(y_{t+h}, y_t)$
       $= E(y_{t+h}y_t) - E(y_{t+h})E(y_t)$
       $= E((x_{t+h} + 3)(x_t + 3)) - E(x_{t+h} + 3)E(x_t + 3)$
       $= E(x_{t+h}x_t + 3x_{t+h} + 3x_t + 9) - (E(x_{t+h}) + 3)(E(x_t) + 3)$
       $= E(x_{t+h}x_t) + 3E(x_{t+h}) + 3E(x_t) + 9 - E(x_{t+h})E(x_t) - 3E(x_t) - 3E(x_{t+h}) - 9$
       $= E(x_{t+h}x_t) - E(x_{t+h})E(x_t)$
       $= cov(x_{t+h}, x_t)$
       Which we know is independent of time t
       Thus, $cov(y_t, y_{t+h})$ does not depend on time

2. In order for y_t to be stationary, its statistical properties should be constant over time. We already know that the autocovariance of y_t is independent of time. Lets look at mean.
   $$E(y_t) = \begin{cases} E(x_t) & for\ t\ odd \\ E(x_t) + 3 & for\ t\ even \end{cases}$$
   We know that $E(x_t)$ is constant since x_t is stationary. Thus,
   $$E(y_t) = \begin{cases} Constant & for\ t\ odd \\ Constant + 3 & for\ t\ even \end{cases}$$
   Since the mean of y_t is dependent on time t, it is not stationary.

Problem 6:

Consider the following time series model: $x_t = w_{t-1} + 2w_t + w_{t+1}$

where the w_t are a white noise process with variance σ^2. Determine the autocovariance and autocorrelation functions of x_t.

Answer:

Autocovariance
$$= \gamma x(t+h, t)$$
$$= E(x_{t+h}x_t) - E(x_{t+h})E(x_t)$$

Since $E(x_t) = E(x_{t+h}) = 0$, Autocovariance

$$= E(x_{t+h}x_t)$$
$$= E((w_{t-1+h} + 2w_{t+h} + w_{t+1+h})(w_{t-1} + 2w_t + w_{t+1}))$$

$$= \begin{cases} 6\sigma^2 & if\ h = 0 \\ 4\sigma^2 & if\ h = \pm 1 \\ 0 & Otherwise \end{cases}$$

where we used that $E(w_{t1}w_{t2}) = 0, E(w_t) = 0\ and\ E(w_t^2) = \sigma^2$

Autocorrelation function

$$= \frac{\gamma(h)}{\gamma(0)}$$

$$= \frac{cov(x_{t+h}, x_t)}{cov(x_t, x_t)}$$

Since $E(x_t) = E(x_{t+h}) = 0$, Autocorrelation

$$= \frac{E((x_{t+h})(x_t))}{E((x_t)(x_t))}$$

$$= \frac{E((w_{t-1+h} + 2w_{t+h} + w_{t+1+h})(w_{t-1} + 2w_t + w_{t+1}))}{6\sigma^2}$$

$$= \begin{cases} 1 & if \ h = 0 \\ \frac{2}{3} & if \ h = \pm 1 \\ 0 & Otherwise \end{cases}$$

where we used that $E(w_{t1} w_{t2}) = 0, E(w_t) = 0 \ and \ E(w_t^2) = \sigma^2$

─────────────**End of Document**─────────────