585 HW2

Aditya Thakur

11/02/2021

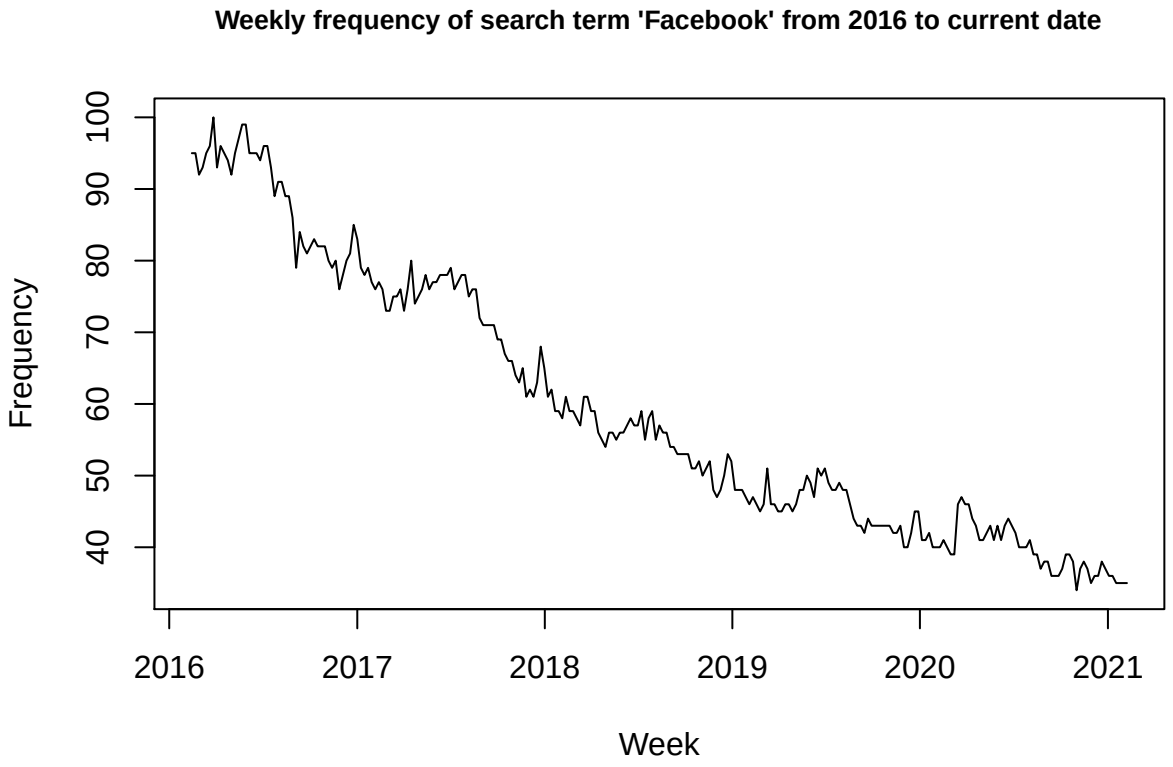
Question 1

```
library(readr)
```

```
##
## Attaching package: 'readr'
```

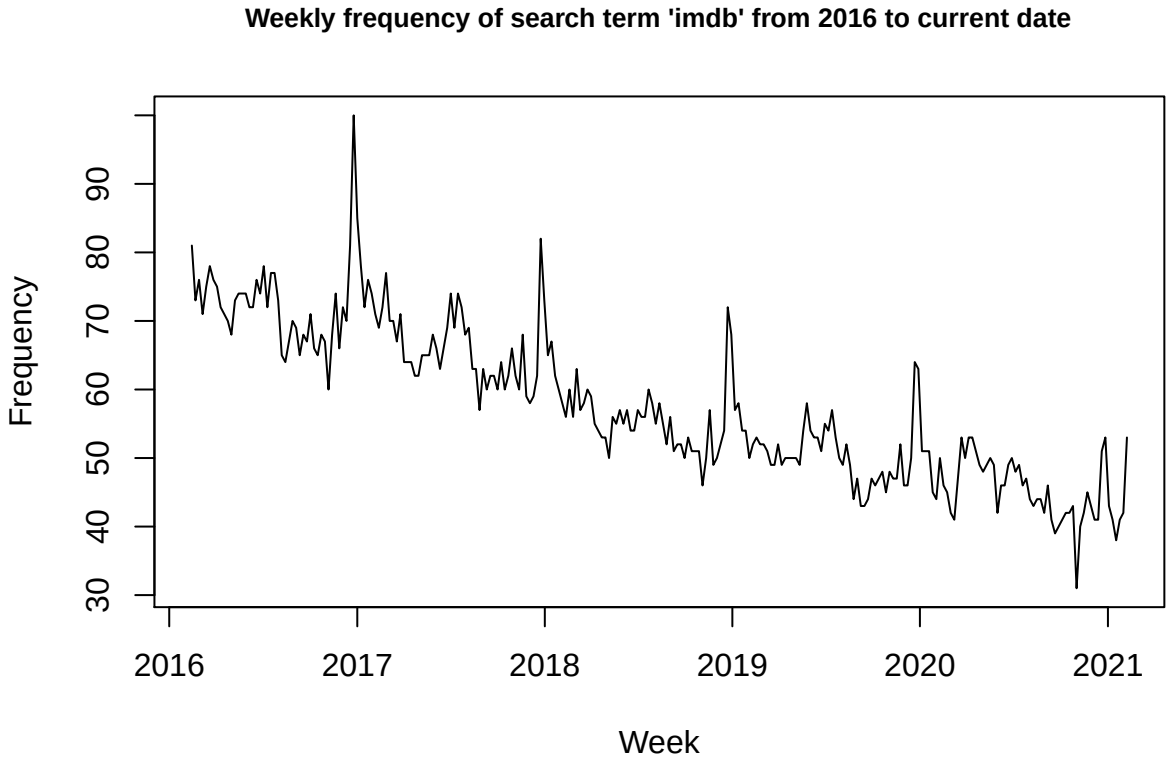
```
## The following object is masked from 'package:TSA':
##
##     spec
```

```
fbts <- read.csv("~/Downloads/FBTS-Trend.csv")
fbts$Week<-as.Date(fbts$Week,format="%Y-%m-%d")

plot(fbts,type='l',
     main="Weekly frequency of search term 'Facebook' from 2016 to current date"
     ,cex.main=0.8)
```



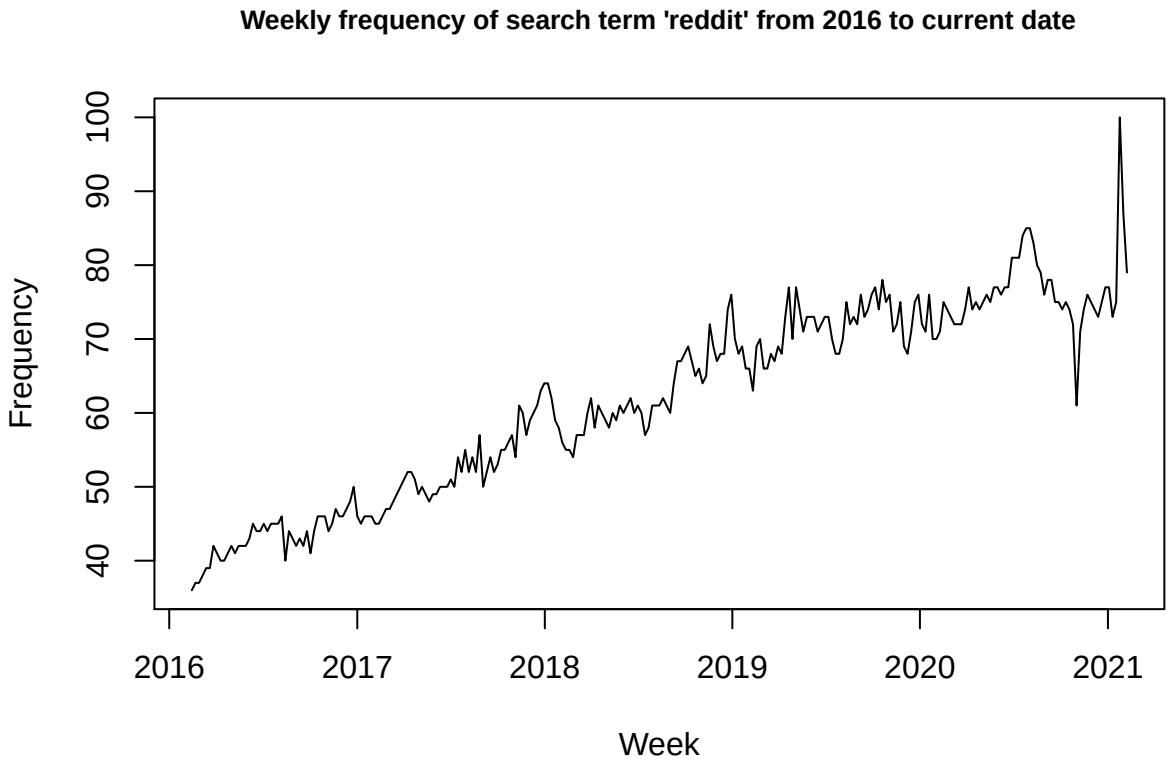
The search term 'Facebook' has been declining in popularity ever since 2016, as visible through the plot. There is a general negative trend component existent in this data, as search queries for this term have been declining in a linear fashion from 2016, perhaps due to the large privacy issues plaguing the company.

```
imdb <- read.csv("~/Downloads/imdb- Trend.csv")
imdb$Week<-as.Date(imdb$Week,format="%Y-%m-%d")

plot(imdb,type='l',
     main="Weekly frequency of search term 'imdb' from 2016 to current date",
     cex.main=0.8)
```



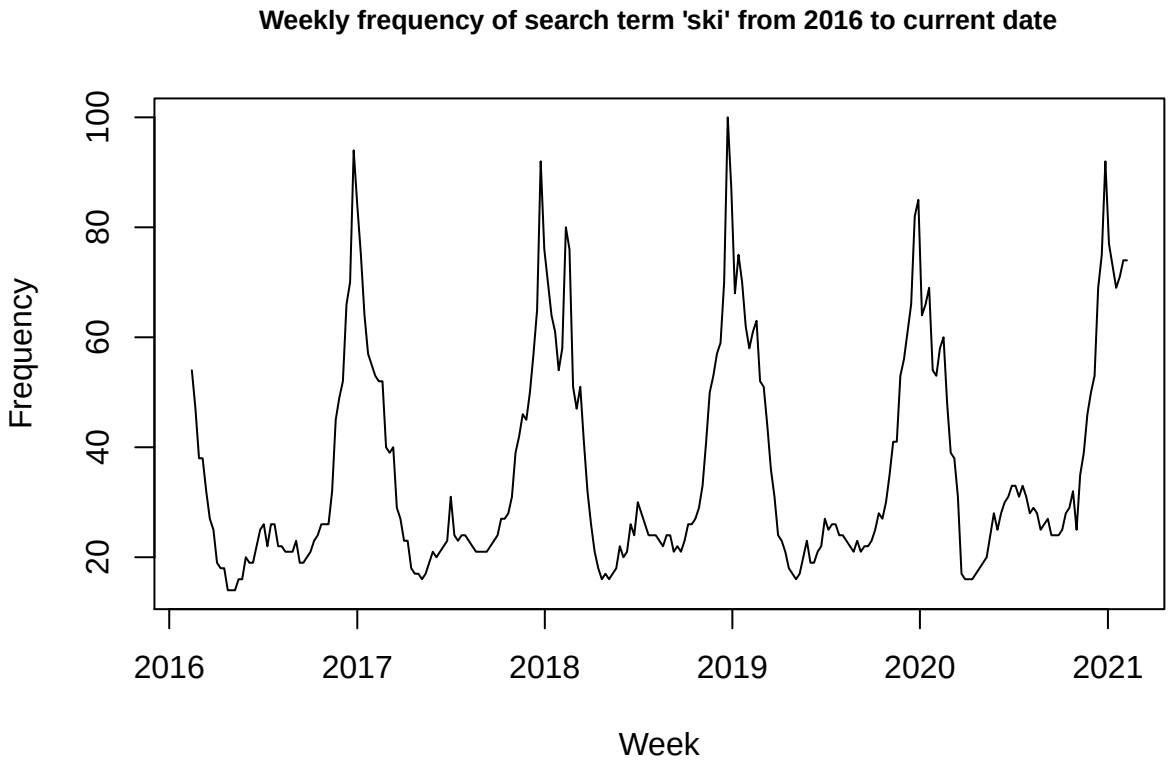
This plot too has a trend component, but it is slightly less negative than the previous term (Facebook). Still here a negative trend can be seen, indicating that searches for 'imdb', a popular internet movie data base, has been declining at a steady rate since 2016.

```
reddit <- read.csv("~/Downloads/Reddit-Trend.csv")
reddit$Week<-as.Date(reddit$Week,format="%Y-%m-%d")

plot(reddit,type='l',
     main="Weekly frequency of search term 'reddit' from 2016 to current date",
     cex.main=0.8)
```



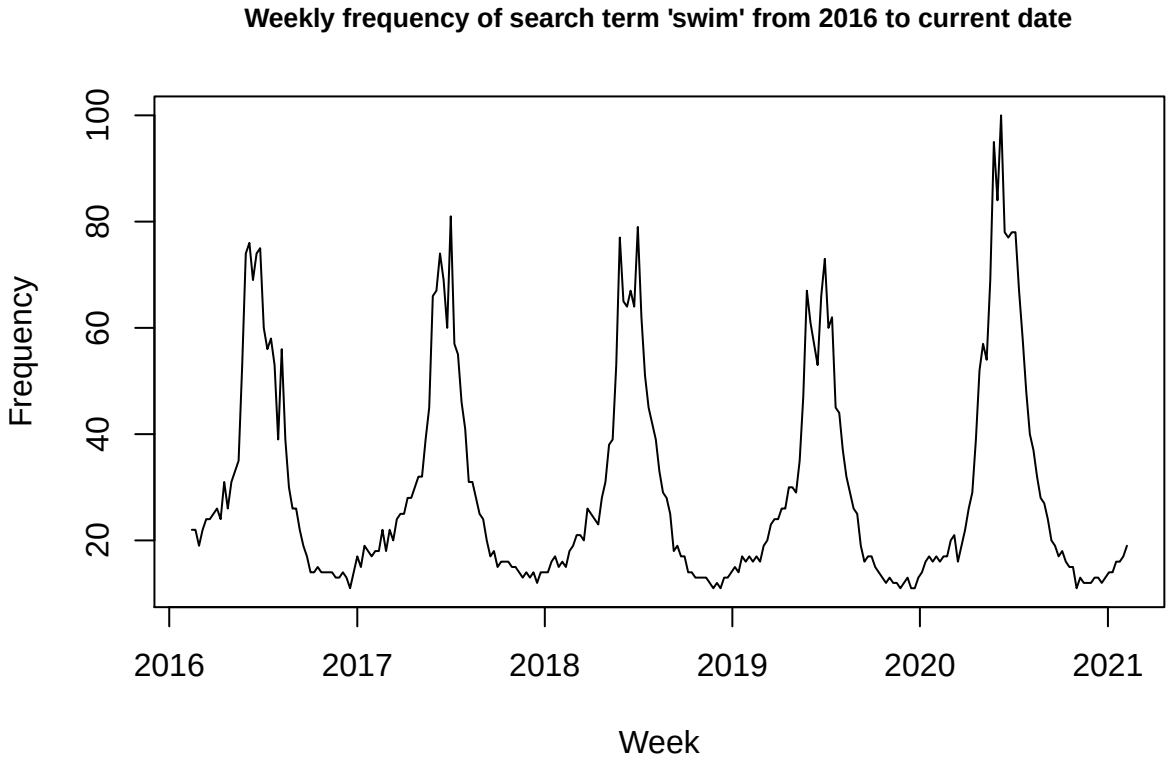
As opposed to the trends in the previous two graphs, the trend for the search term 'reddit' has been increasing steadily since 2016. There is a sharp jump seen near the beginning of 2021 (perhaps due to the Gamestop incident) however there is still a general upward trend.

```
ski <- read.csv("~/Downloads/ski- seasonal.csv")
ski$Week<-as.Date(ski$Week,format="%Y-%m-%d")

plot(ski,type='l',
     main="Weekly frequency of search term 'ski' from 2016 to current date",
     cex.main=0.8)
```



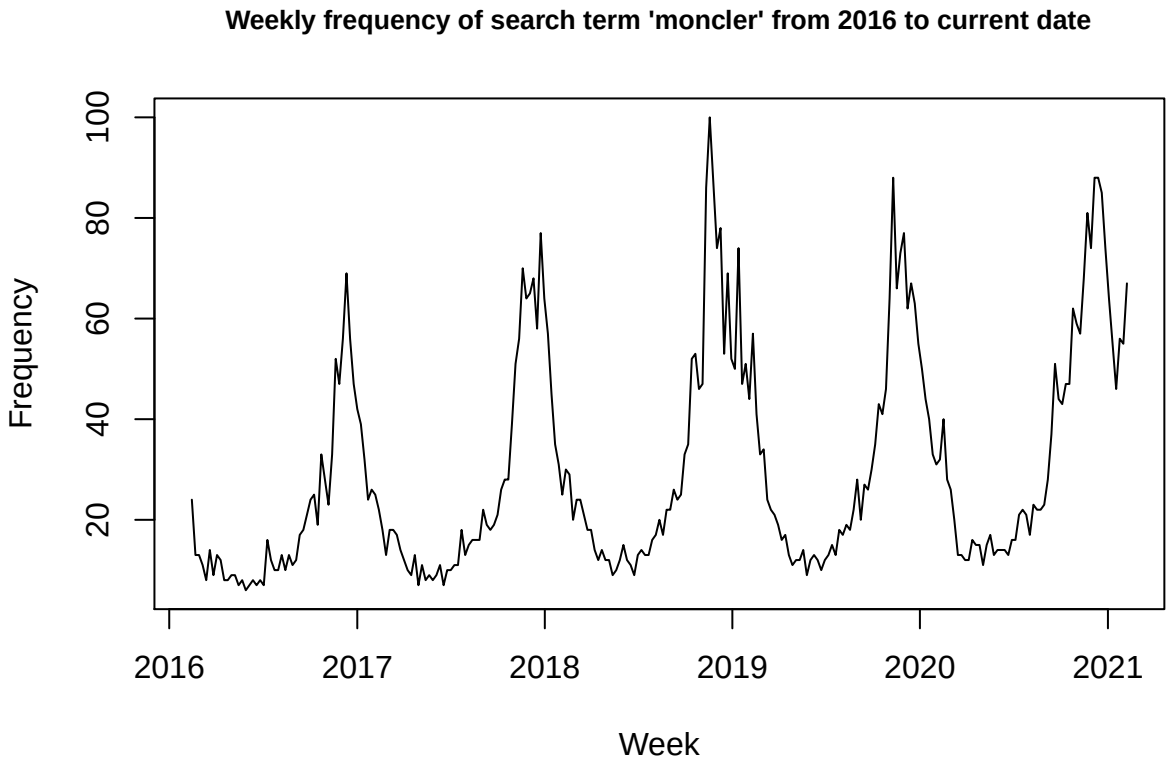
Here we see our first seasonal data set for the search term 'ski'. This term is the most popular in the concluding months of the year, which makes sense as these months are during the winter in most places, where a lot of individuals opt to ski.

```
swim <- read.csv("~/Downloads/swim-season.csv")
swim$Week<-as.Date(swim$Week,format="%Y-%m-%d")

plot(swim,type='l',
     main="Weekly frequency of search term 'swim' from 2016 to current date",
     cex.main=0.8)
```



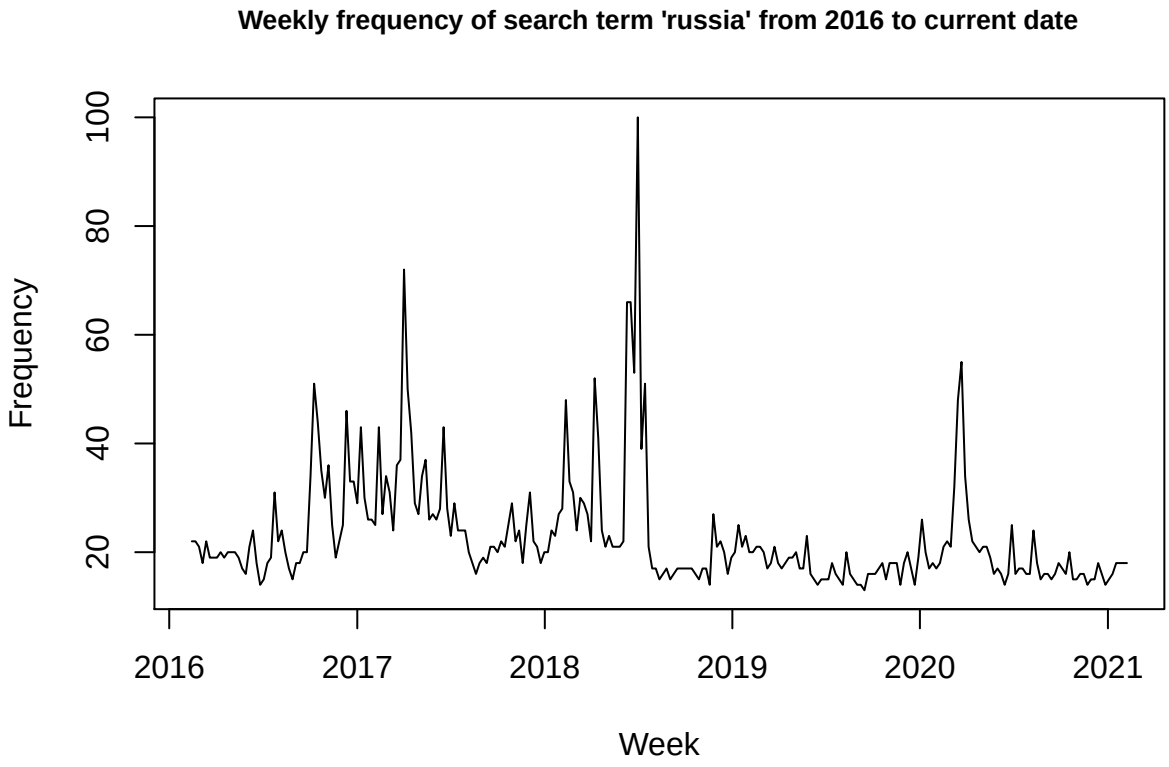
Another seasonal data set, this term too is highly influenced by the seasons. As swimming is a popular activity during the summer period, we see spikes in the middle of each year (Jun-Aug) indicating the effect of the weather on the popularity of this activity.

```
moncler <- read.csv("~/Downloads/moncler-season.csv")
moncler$Week<-as.Date(moncler$Week,format="%Y-%m-%d")

plot(moncler,type='l',
     main="Weekly frequency of search term 'moncler' from 2016 to current date",
     cex.main=0.8)
```



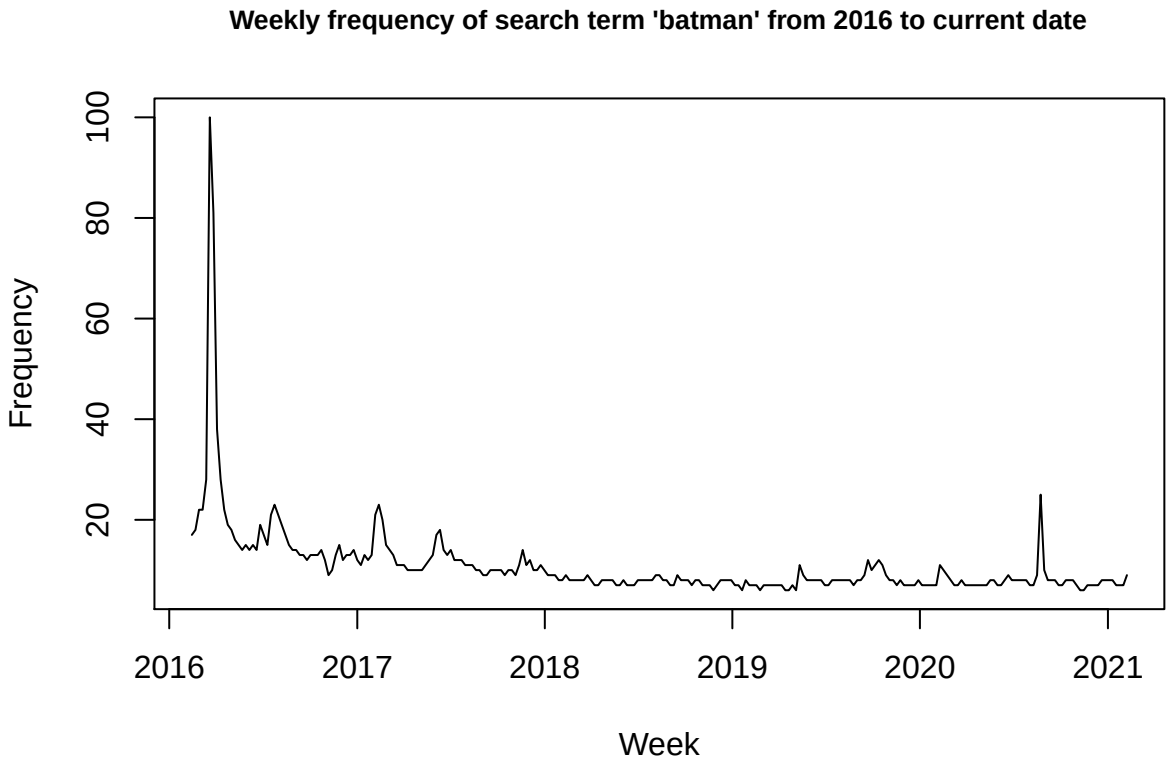
Moncler is an Italian luxury fashion company, predominantly selling winter wear. It therefore makes sense that searches for their products peak during the winter months of each year, resembling a seasonal component in the data set.

```
russia <- read.csv("~/Downloads/russia-noth.csv")
russia$Week<-as.Date(russia$Week,format="%Y-%m-%d")

plot(russia,type='l',
     main="Weekly frequency of search term 'russia' from 2016 to current date",
     cex.main=0.8)
```



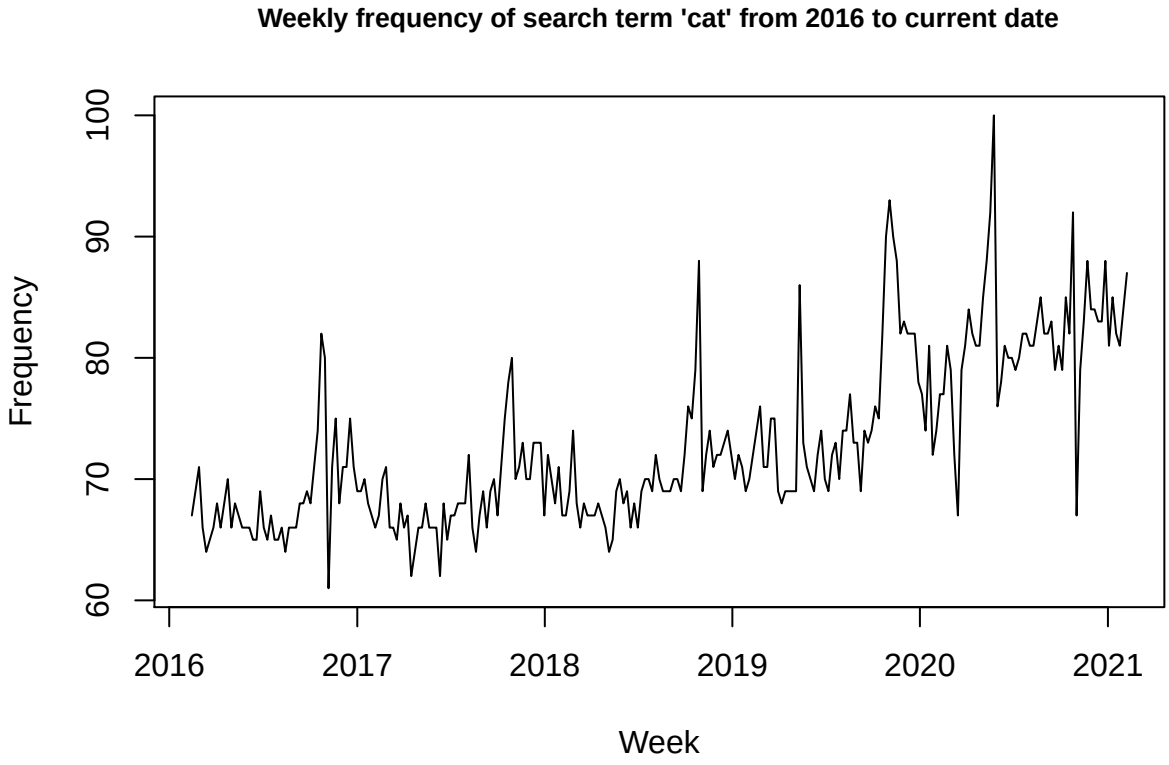
Here we see that for the search term 'Russia', there is no clear seasonal or trend component. We see spikes in frequency of searches in 2017 and again in 2019, however, there is nothing to suggest that these spikes are seasonal, as they are not repeated at a fixed time in every year. Also, there is no general upward or downward trend in the searches.

```
batman <- read.csv("~/Downloads/batmants-noth.csv")
batman$Week<-as.Date(batman$Week,format="%Y-%m-%d")

plot(batman,type='l',
     main="Weekly frequency of search term 'batman' from 2016 to current date",
     cex.main=0.8)
```



The search frequency for the term 'Batman' rised sharply in 2016. However, ever since, the term as maintained a low search rate, with no seasonal changes occurring in 2017-2021 and no general trend either.

```
cat <- read.csv("~/Downloads/cat-noth.csv")
cat$Week<-as.Date(cat$Week,format="%Y-%m-%d")

plot(cat,type='l',
     main="Weekly frequency of search term 'cat' from 2016 to current date",
     cex.main=0.8)
```



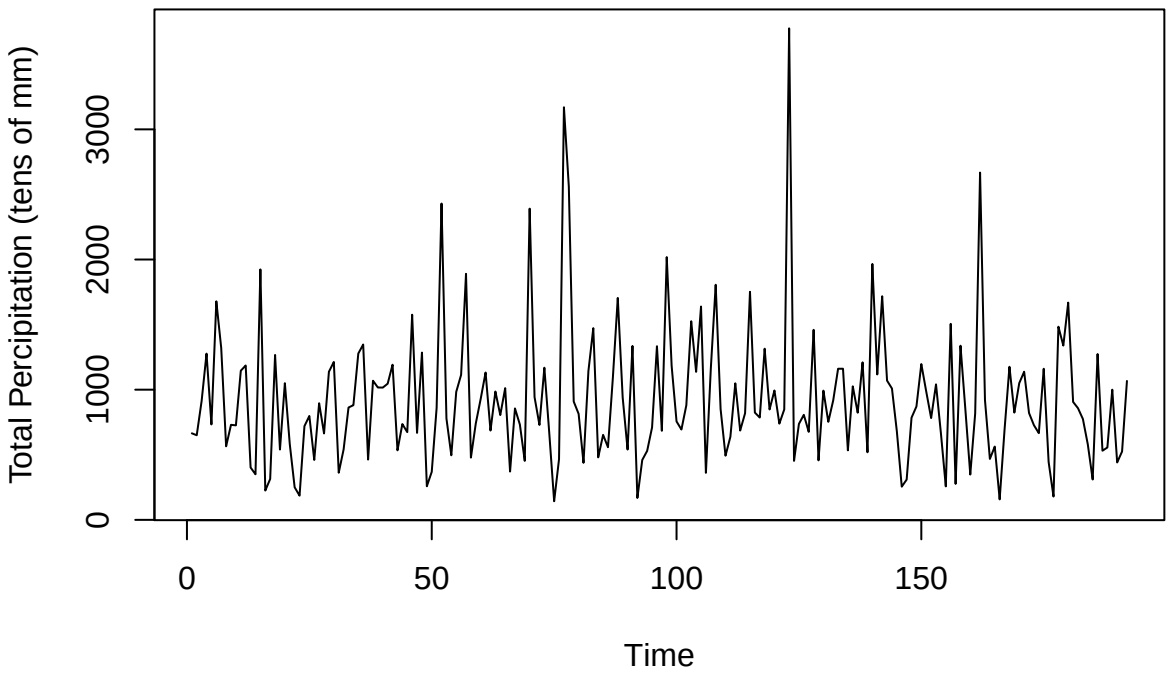
Although there is a slight upward trend forming from 2019, there is nothing to suggest any seasonal effects on the dataset testing the frequency of searches for the term 'cat'.

Question 2

```
BosClimateData<-read_csv("~/Downloads/BosClimateData.csv")
```

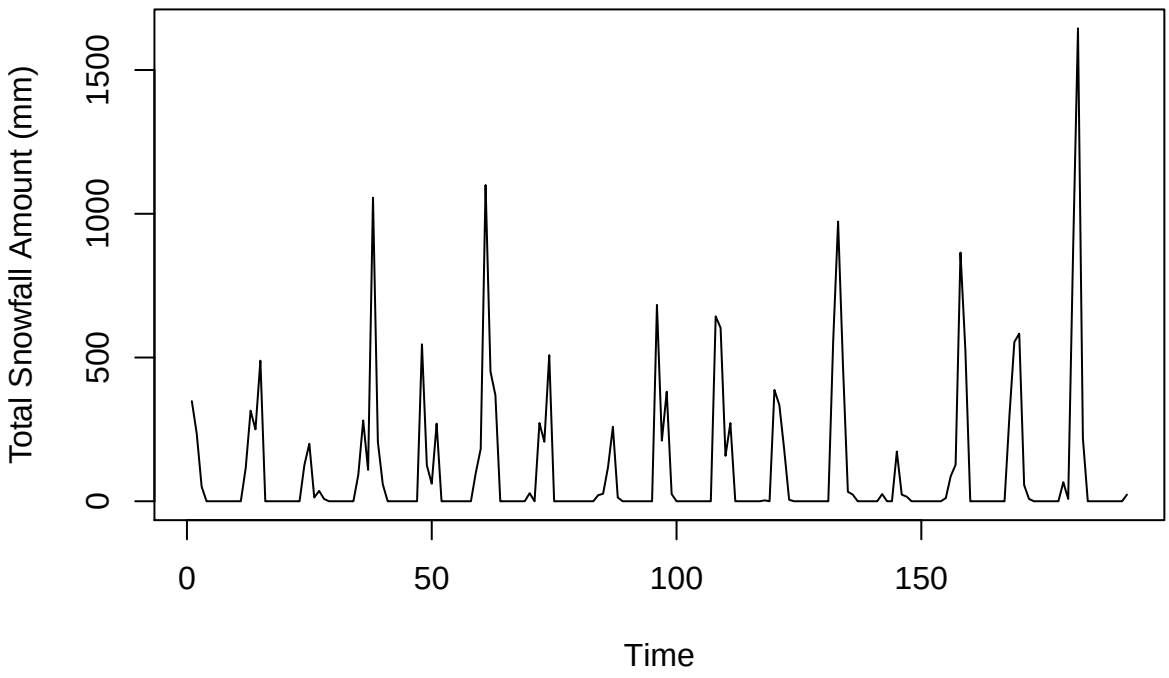
```
## Parsed with column specification:
## cols(
##   STATION = col_character(),
##   STATION_NAME = col_character(),
##   DATE = col_double(),
##   TPCP = col_double(),
##   TSNW = col_double(),
##   MMXT = col_double(),
##   MMNT = col_double()
## )
```

```
BosData=subset(BosClimateData,
               BosClimateData$STATION_NAME==
                 "BOSTON LOGAN INTERNATIONAL AIRPORT MA US")
plot.ts(BosData$TPCP,ylab="Total Percipitation (tens of mm)")
```



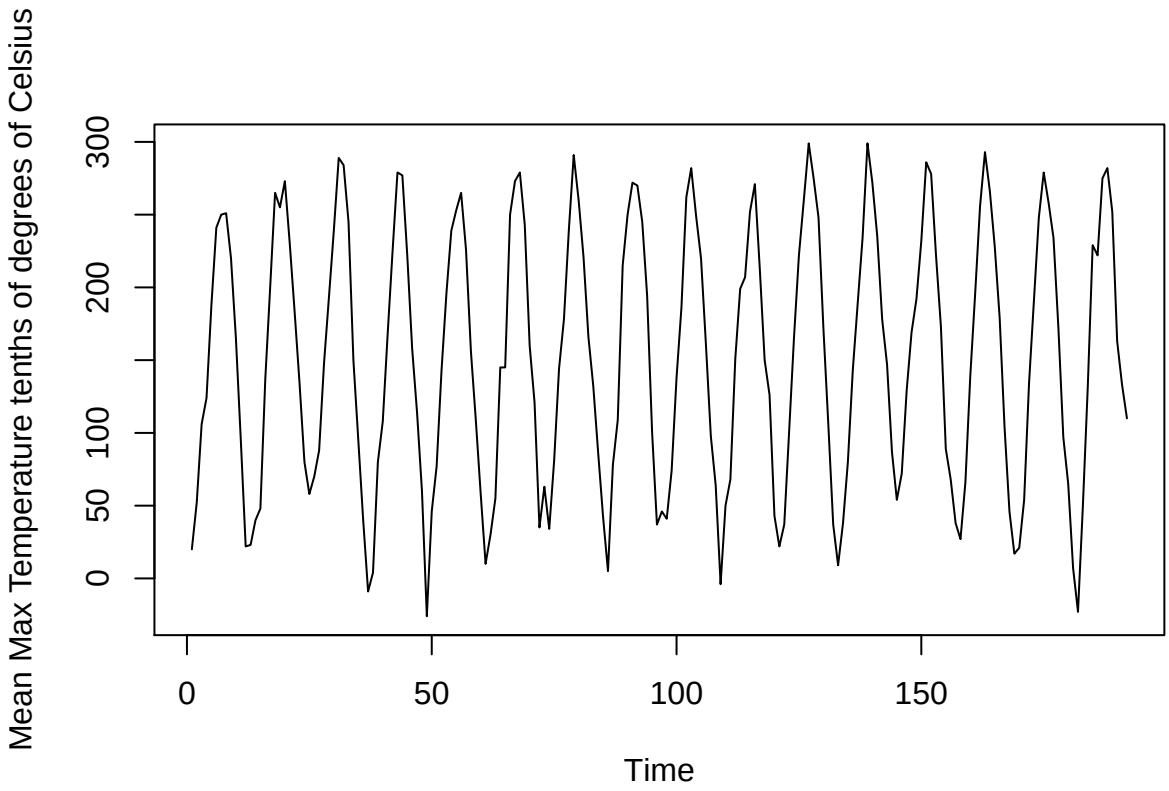
In this series, we see a large spike occurring between time periods 100 and 150, indicating a larger event (that resulted in larger amounts of precipitation). It is somewhat surprising to not find a seasonal component in this time series, as one would expect there to be periods where the rainfall is fairly consistent.

```
plot.ts(BosData$TSNW,ylab="Total Snowfall Amount (mm)")
```



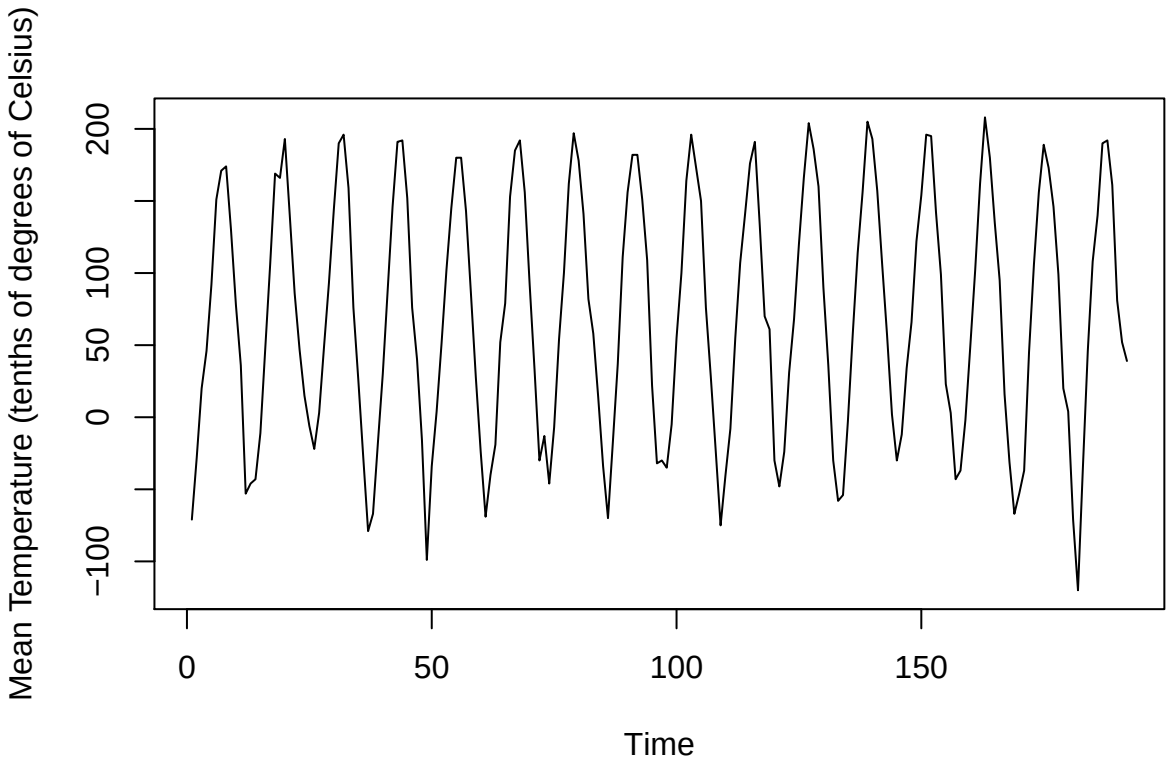
This data makes sense logically. There are only a few months in every year that it does snow and for the rest of the months, there is no snow. Therefore, we see these sharp inclines and decreases in our series.

```
plot.ts(BosData$MMXT,ylab="Mean Max Temperature tenths of degrees of Celsius")
```



This series and this plot is very useful to explain the fluctuations in the temperature that we witness across seasons. Thee points where the mean max temperature is high can be identified as the summer months, while there is a sharp decline to the winter months represented by the decrease in temperature.

```
plot.ts(BosData$MMNT,ylab="Mean Temperature (tenths of degrees of Celsius)")
```



Similar to the plot above, we see the fluctuations of temperature across time. From this, we can make out the 'summer' and 'winter' months.

```
snow_prop=length(subset(BosData$TSNW,BosData$TSNW>=304.8))/length(BosData$TSNW)
snow_prop
```

```
## [1] 0.125
```

12.5% of the months had snowfall greater than a foot. I used 304.8 as the minimum subset value as 1ft=304.8mm

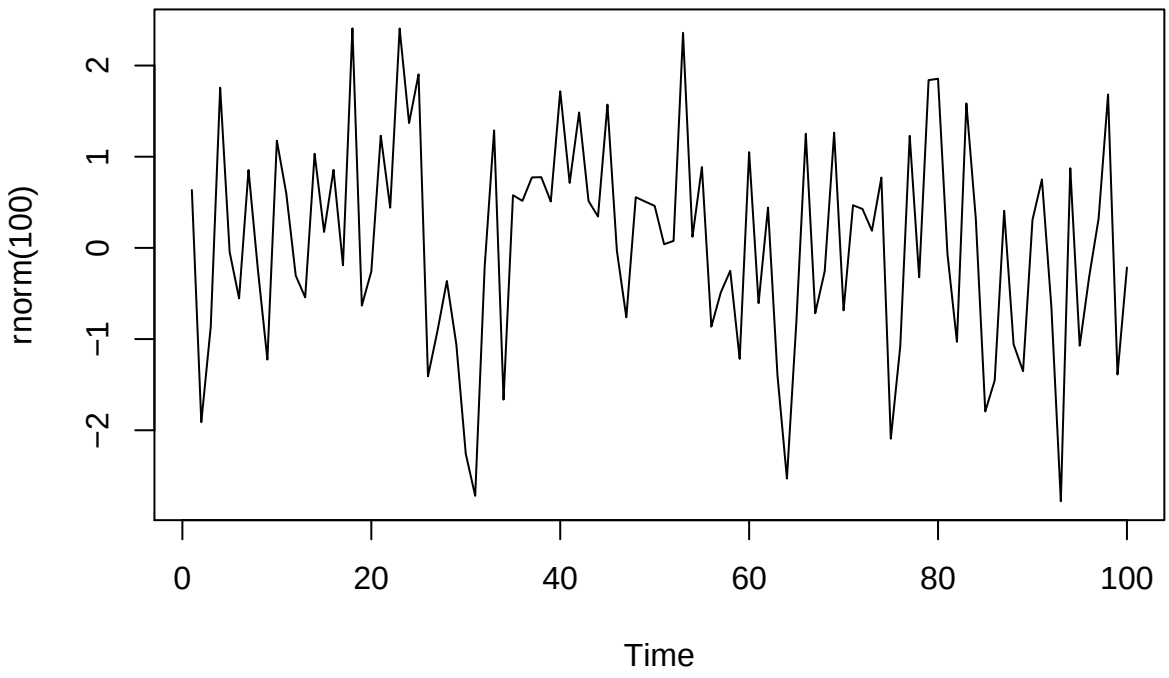
```
BosData$MMXT <- ((BosData$MMXT/10)*(9/5))+32
temp_prop=length(subset(BosData$MMXT,BosData$MMXT>80))/length(BosData$MMXT)
temp_prop
```

```
## [1] 0.1145833
```

The mean maximum temperature exceeded 80 degrees F for 11.45% of the months.

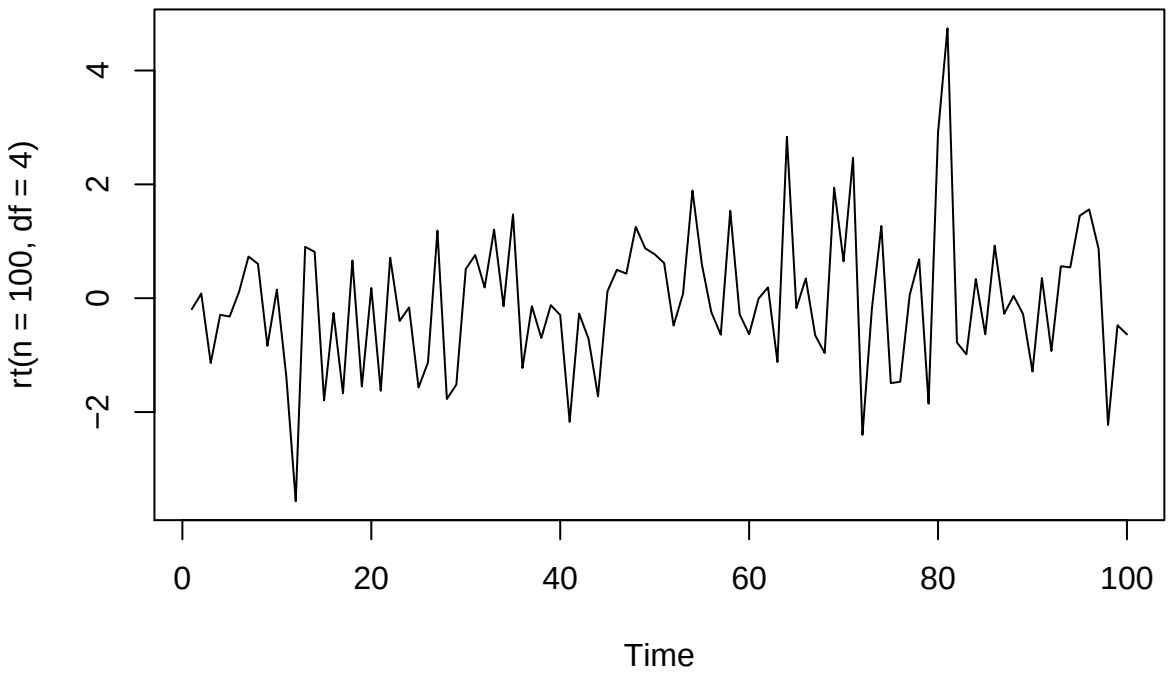
Question 3

```
plot.ts(rnorm(100))
```



Yes, the plot above shows that there is no general trend along the data. It does appear to be random, as there are oscillations up and down at a constant rate throughout the time period, with the lengths of these oscillations changing randomly throughout. As there are these oscillations around 0, we can infer that the mean of the data is 0. Additionally, at any given time point, the variance appears to be finite.

```
plot.ts(rt(n=100,df=4))
```



Firstly, as visible, a lot of the data is concentrated above 0. Therefore, the average of the data does not appear to be 0. There also seems to be additional volatility clustering at different time points. This indicates that the data has higher kurtosis than a Gaussian distribution. All these elements indicate that this is not a Gaussian distribution.