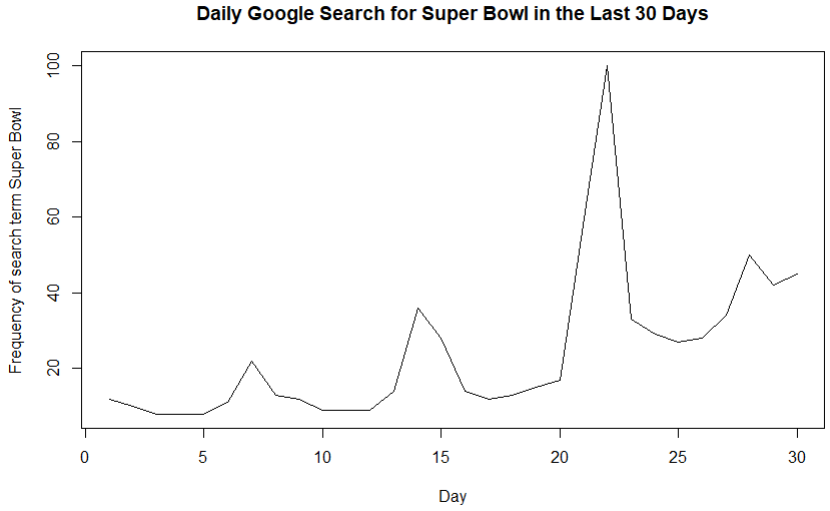
MA585
Time Series Analysis
Homework 2

1. Google Trends (https://www.google.com/trends/) provides information about popularity of search phrases. You can enter your phrase in the search box and from the drop down menu your can choose your time frame. For less than 90 days, you can view the daily data, otherwise you will see the weekly or monthly data. You can download the data in csv format by clicking on the [⇓] symbol from the top right corner (you need to be logged into your Google account). Here is the daily time series plot for the frequency of the search term "Super Bowl" in the last 30 days.

>library(TSA)

> plot.ts(superbowl, ylab="Frequency of search term Super Bowl",xlab="Day")

> title("Daily Google Search for Super Bowl in the Last 30 Days")



Your job is to think of three search terms each with the following characteristics:

1. a time series with a clear trend component
2. a time series with a clear seasonal component.
3. a time series with no clear trend or seasonal pattern and is poorly described by a combination of trend and seasonal components.

In each case download and plot the data. Explain why the data satisfies the requirements as stated above.

2. National Oceanic and Atmospheric Administration (NOAA) is a great resource for atmospheric time series data. Explore the website https://www.ncdc.noaa.gov/cdo-web/. You will find a wealth of time series data on climate. I have downloaded data related to recent climate patterns in Boston (2000 onwards) and it is available as a csv file "BostonClimateData.csv". Load the data in R. You will see that the data was collected from multiple weather stations in Massachusetts. Let's focus on the data collected at the Boston Logan Airport.

> BosData=subset(BosClimateData,BosClimateData$STATION_NAME=="BOSTON LOGAN INTERNATIONAL AIRPORT MA US")

The data has four variables.

TPCP - Total precipitation amount for the month (tens of mm)

TSNW - Total snowfall amount for the month (mm)

MMXT - Monthly mean maximum temperature (tenths of degrees of Celsius)

MNTM – Monthly mean temperature (tenths of degrees of Celsius)

Plot the time series for each variable and write a short description of the key features of each series. What proportion of months had total snowfall greater than a foot? What proportion of months mean maximum temperature exceeded $80^0 F$?

R Help: R function length gives you the length of a series.

```
>x=c(2,1,3,2,4,6.0,0,2)
> length(x)
[1] 8
>length(subset(x,x>=3))      # number of values greater than or equal to 3
[1] 3
```

3. a. Simulate a completely random process of length 100 with independent standard normal values. Plot the time series. Does it look "random"? Repeat this exercise a few times with a new simulation each time.

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

b. Repeat the problem in (a), except this time use random samples from a t-distribution with 4 df (replace rnorm(100) by rt(n=100, df=4)). If you didn't know that the data was generated from independent t-distribution, what feature of the data might have made you think that the data generating process may not be Gaussian?