# Hw2 Stat153

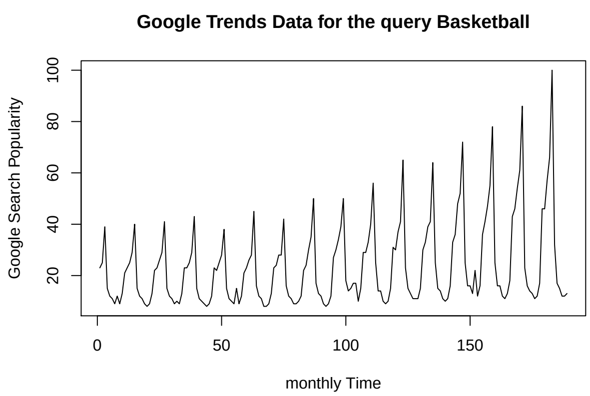
Ziyuan Li

9/22/2021

## Computer Exercises

### 1

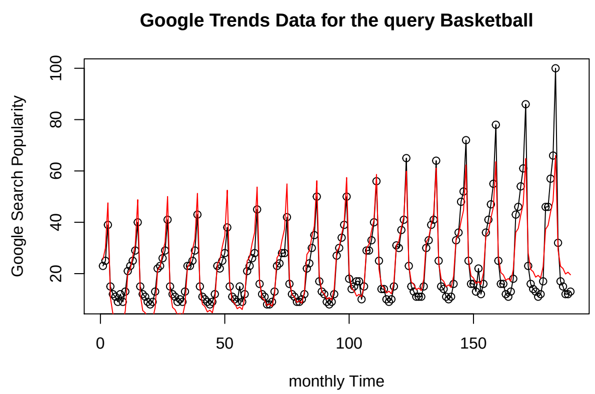
```
#1a
Basketball = read.csv("/Users/li/Desktop/STAT153/HW2/Basketball.csv", header = T)
basket=Basketball$Basketball
plot(basket, type = "l", ylab = "Google Search Popularity", xlab = "monthly Time", main = "Google Trends
```



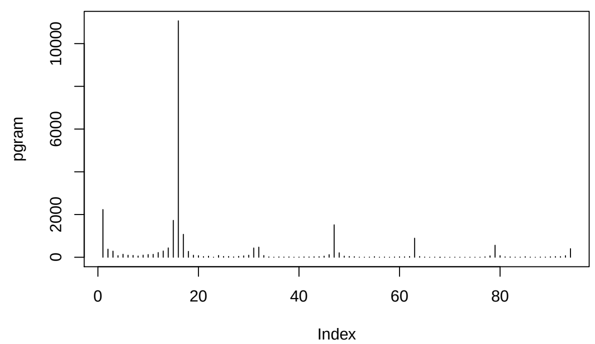

There is increasing trend and seasonality.

```
#1b
#We fitted a seasonal function to the data:
t = 1: length(basket)
f1 = 1
f2 = 2
f3 = 3
f4 = 4
f5 = 5
f6 = 6
d = 12
v1 = cos(2*pi*f1*t/d)
v2 = sin(2*pi*f1*t/d)
v3 = cos(2*pi*f2*t/d)
v4 = sin(2*pi*f2*t/d)
v5 = cos(2*pi*f3*t/d)
```

```
v6 = sin(2*pi*f3*t/d)
v7 = cos(2*pi*f4*t/d)
v8 = sin(2*pi*f4*t/d)
v9 = cos(2*pi*f5*t/d)
v10 = sin(2*pi*f5*t/d)
v11 = cos(2*pi*f6*t/d)

lin.mod = lm(basket ~ 1 + t  + v1 + v2 + v3 + v4 + v5 + v6 + v7 + v8 + v9 + v10 + v11 )
plot(t, basket, type = "o", xlab = "monthly Time", ylab = "Google Search Popularity", main = "Google Tr
points(t, lin.mod$fitted, type = "l", col = "red")
```




```
#1c
m = floor(length(basket)/2)
pgram = abs(fft(basket)[2:(m+1)])^2/length(basket)
plot(pgram, type = "h")
```



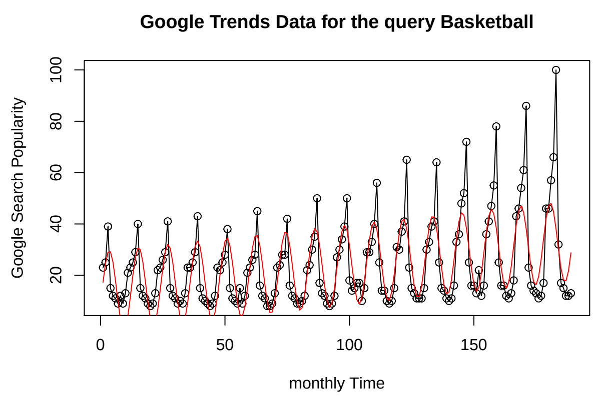

When j = 16, the periodogram achieve the largest value.

length(basket)=189

this implies sinusoid of frequency = 16/189 = 0.0847 and period = 189/16 = 11.8125.

Because basketball query is a monthly data, so this implies there may be a 12-month period.And because 12-month is a non-Fourier frequency, so there is leakage.

```
#1d
t = 1: length(basket)
f1 = 16
d=length(basket)
v1 = cos(2*pi*f1*t/d)
v2 = sin(2*pi*f1*t/d)
dft.mod = lm(basket ~ 1 + t + v1 + v2)
plot(t, basket, type = "o", xlab = "monthly Time", ylab = "Google Search Popularity", main = "Google Tr
points(t, dft.mod$fitted, type = "l", col = "red")
```

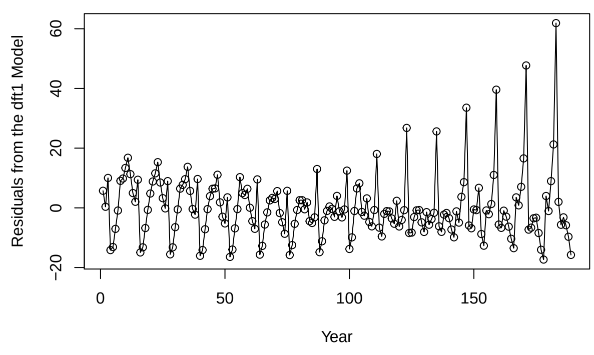



1)These new fitted values don't oscillate at the right frequency with the data.

2)This model doesn't fit well.

3)Because there is still other component of Fourier sinusoid which combine together with the single dominant sinusoid to form the final data. And as shown in periodogram, there are other components with strong signals like j = 1,15...

```
#1e
plot(t, dft.mod$residuals, type = "o", xlab = "Year", ylab = "Residuals from the dft1 Model")
```

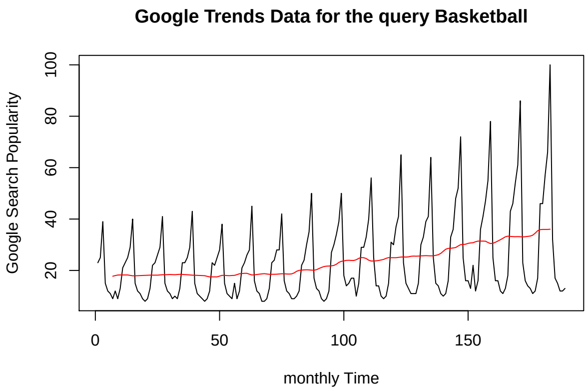



There still important sinusoids present even after removing the first dominant sinusoid.

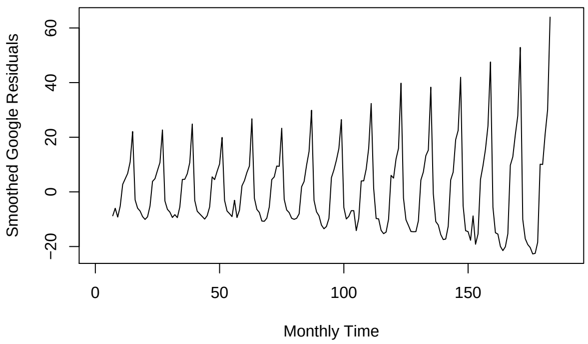
### 1f

Because DFT implies there may be an annual period, so use d = 12 to take a local average of 1 year data.

```
#1f
ts.dat = basket
tme = 1:length(ts.dat)
q=3
sm.par = c(0.5,1,1,1,1,1,1,1,1,1,1,1,0.5)
mt = filter(ts.dat, sm.par/sum(sm.par))


plot(ts.dat, type = "l", ylab = "Google Search Popularity", xlab = "monthly Time", main = "Google Trends
lines(tme, mt, type = "l",col="red")
```



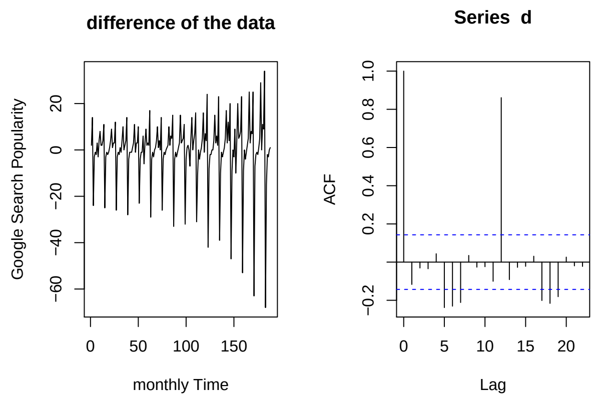

```
res=ts.dat-mt

plot(res, type = "l", ylab = "Smoothed Google Residuals", xlab = "Monthly Time", )
```



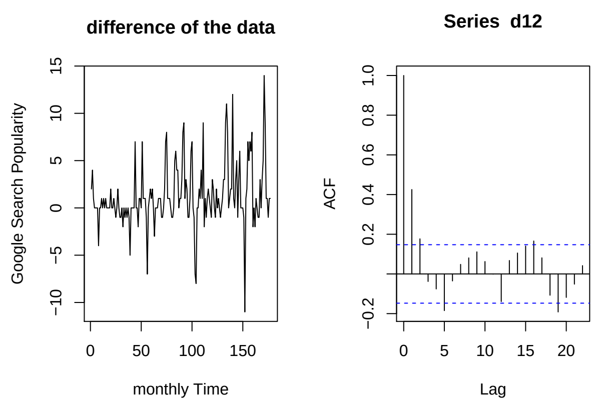

The residuals's variance is increasing, which means it is non-stationary.

```
#1g
ts.dat = basket
d = diff(basket,lag = 1)
d12 = diff(basket,lag = 12)
dd12 = diff(d12)

par(mfrow = c(1,2))
```

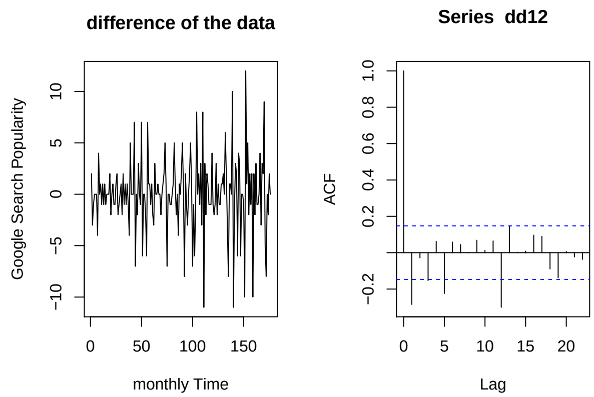
```r
plot(d, type = "l", ylab = "Google Search Popularity", xlab = "monthly Time", main = "difference of the
acf(d)
```




```r
plot(d12, type = "l", ylab = "Google Search Popularity", xlab = "monthly Time", main = "difference of th
acf(d12)
```




```r
plot(dd12, type = "l", ylab = "Google Search Popularity", xlab = "monthly Time", main = "difference of
acf(dd12)
```



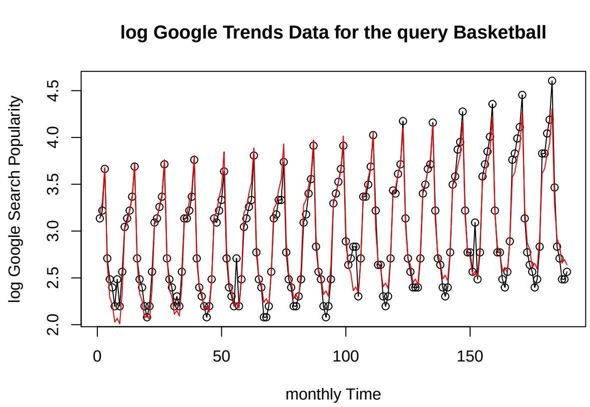

The first difference of the seasonal difference of the data (with 12 months being a season) seems most stationary.

### 1h

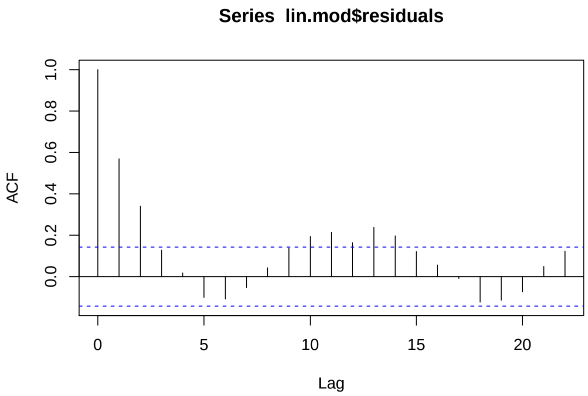
I use following regression model to eliminate heteroskedasticity.

$$log(y_t) = \beta_0 + \beta_1 t + \sum_{j=1}^{6} \left[ \beta_{2j} \cos\left(\frac{2\pi jt}{12}\right) + \beta_{2j+1} \sin\left(\frac{2\pi jt}{12}\right) \right] + w_t$$

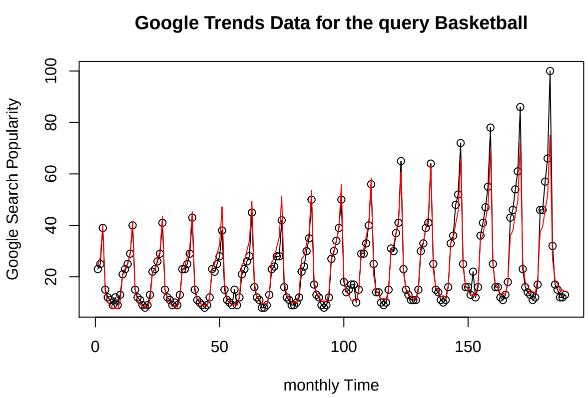
```
#1h
ts.dat = log(basket)
t = 1: length(ts.dat)
f1 = 1
f2 = 2
f3 = 3
f4 = 4
f5 = 5
f6 = 6
d = 12
v1 = cos(2*pi*f1*t/d)
v2 = sin(2*pi*f1*t/d)
v3 = cos(2*pi*f2*t/d)
v4 = sin(2*pi*f2*t/d)
v5 = cos(2*pi*f3*t/d)
v6 = sin(2*pi*f3*t/d)
v7 = cos(2*pi*f4*t/d)
v8 = sin(2*pi*f4*t/d)
v9 = cos(2*pi*f5*t/d)
v10 = sin(2*pi*f5*t/d)
v11 = cos(2*pi*f6*t/d)

lin.mod = lm(ts.dat ~ 1 + t + v1 + v2 + v3 + v4 + v5 + v6 + v7 + v8 + v9 + v10 + v11 )
sd=mean((lin.mod$residuals)^2)

plot(t,ts.dat, type = "o", xlab = "monthly Time", ylab = "log Google Search Popularity", main = "log Goo
points(t, lin.mod$fitted, type = "l", col = "red")
```



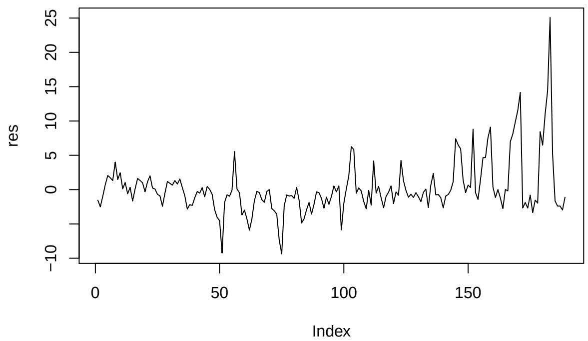

```
acf(lin.mod$residuals)
```




```
plot(t,basket, type = "o", xlab = "monthly Time", ylab = "Google Search Popularity", main = "Google Tre
points(t, exp(lin.mod$fitted+sd/2), type = "l", col = "red")
```




```
res= basket-exp(lin.mod$fitted+sd/2)

plot(res,type="l")
```




The residual of fitting trend and seasonality of log(basket) is almost stationary, but log(basket) may not still be perfectly homoscedasticity ,especially when t > 170.