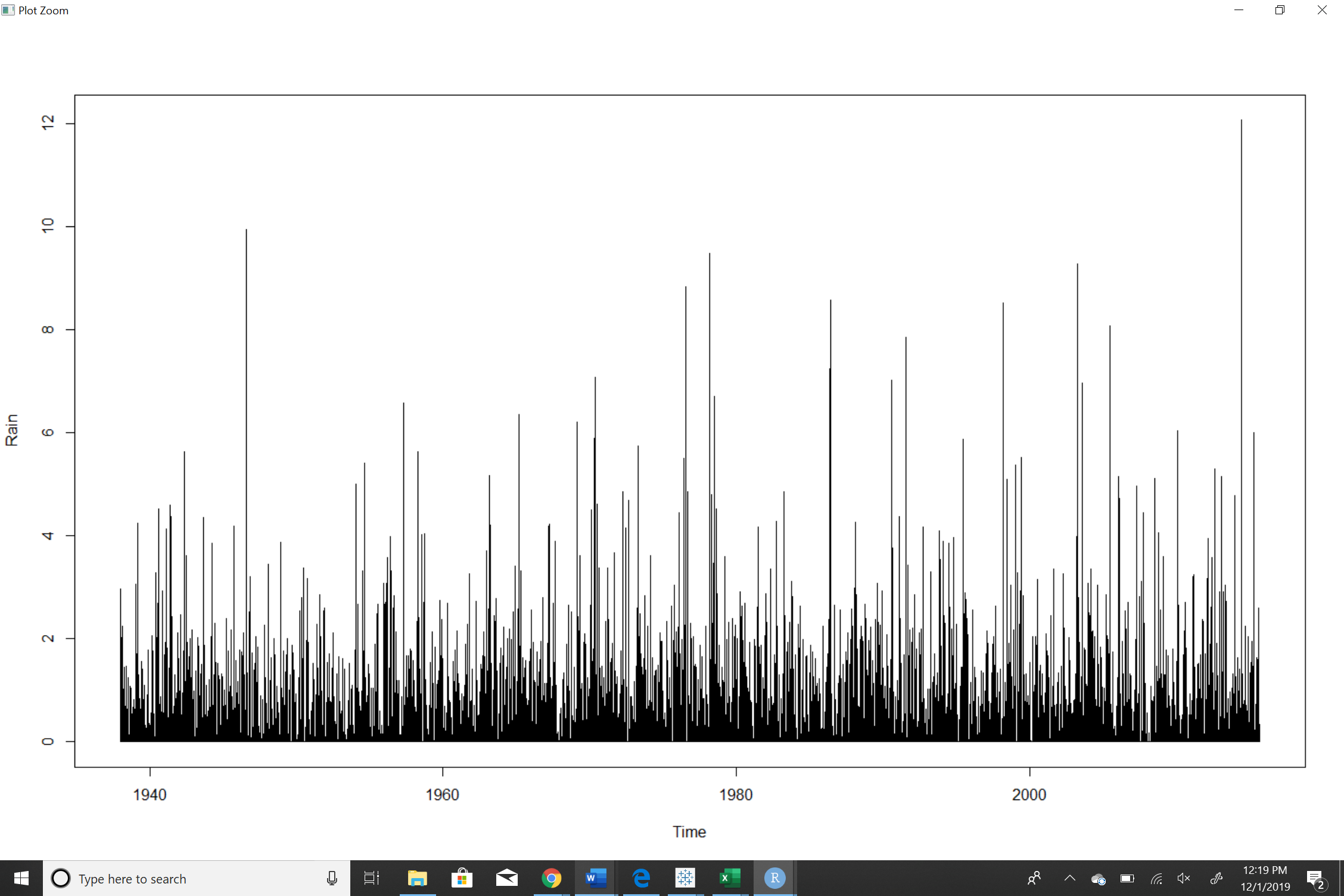
**Time Series Lab 2 Assignment**

1. **Create plots for time series (line plot)**

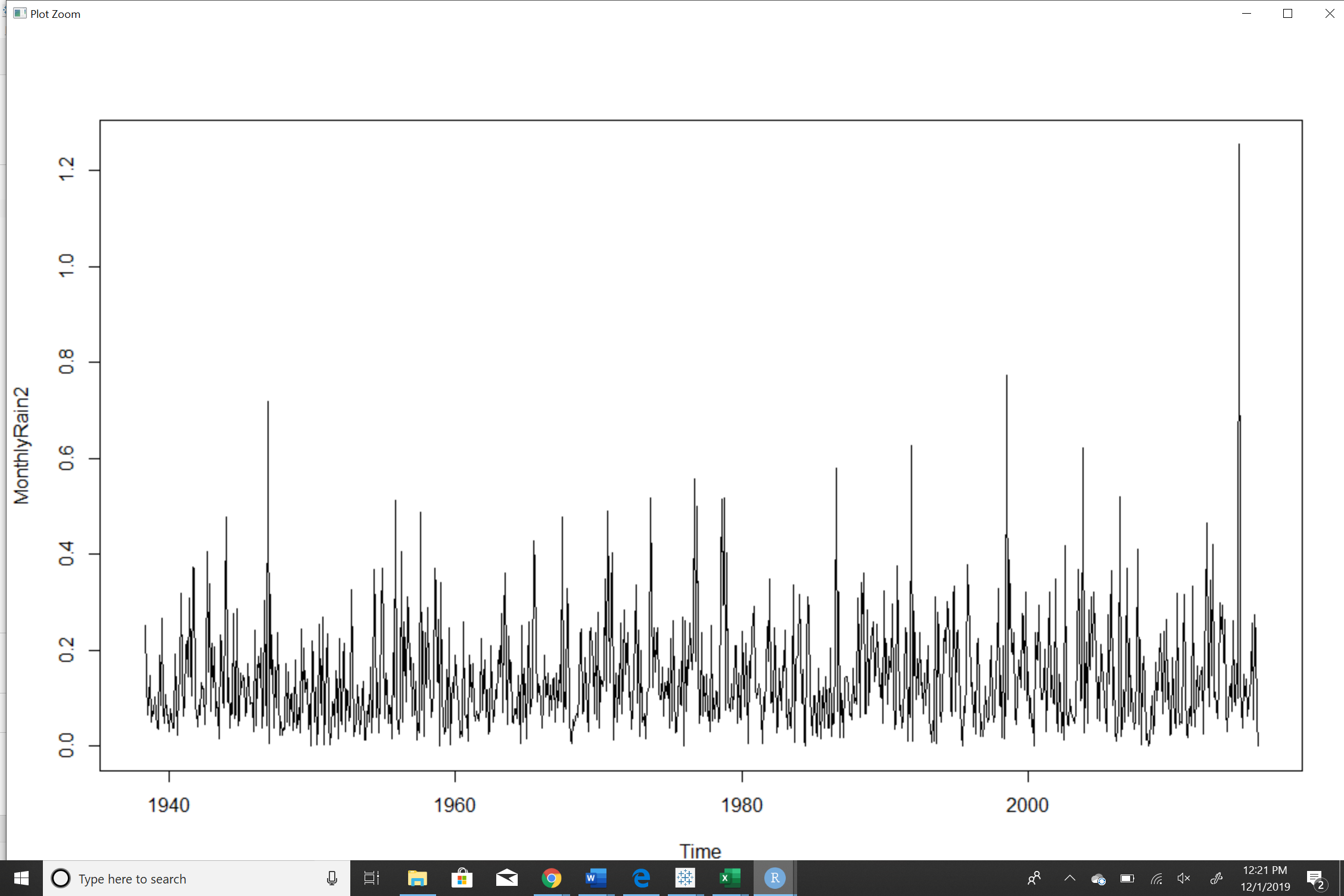
***Full Data:***



**The above graph is plotted for the entire rain data for Houston collected from 1938 to 2016. There exists a seasonality but not a prominent one, in order to adjust seasonality, we can collapse data into monthly average.**

1. **Adjust for the seasonality issues by collapsing the data into monthly averages**

***Monthly average:***



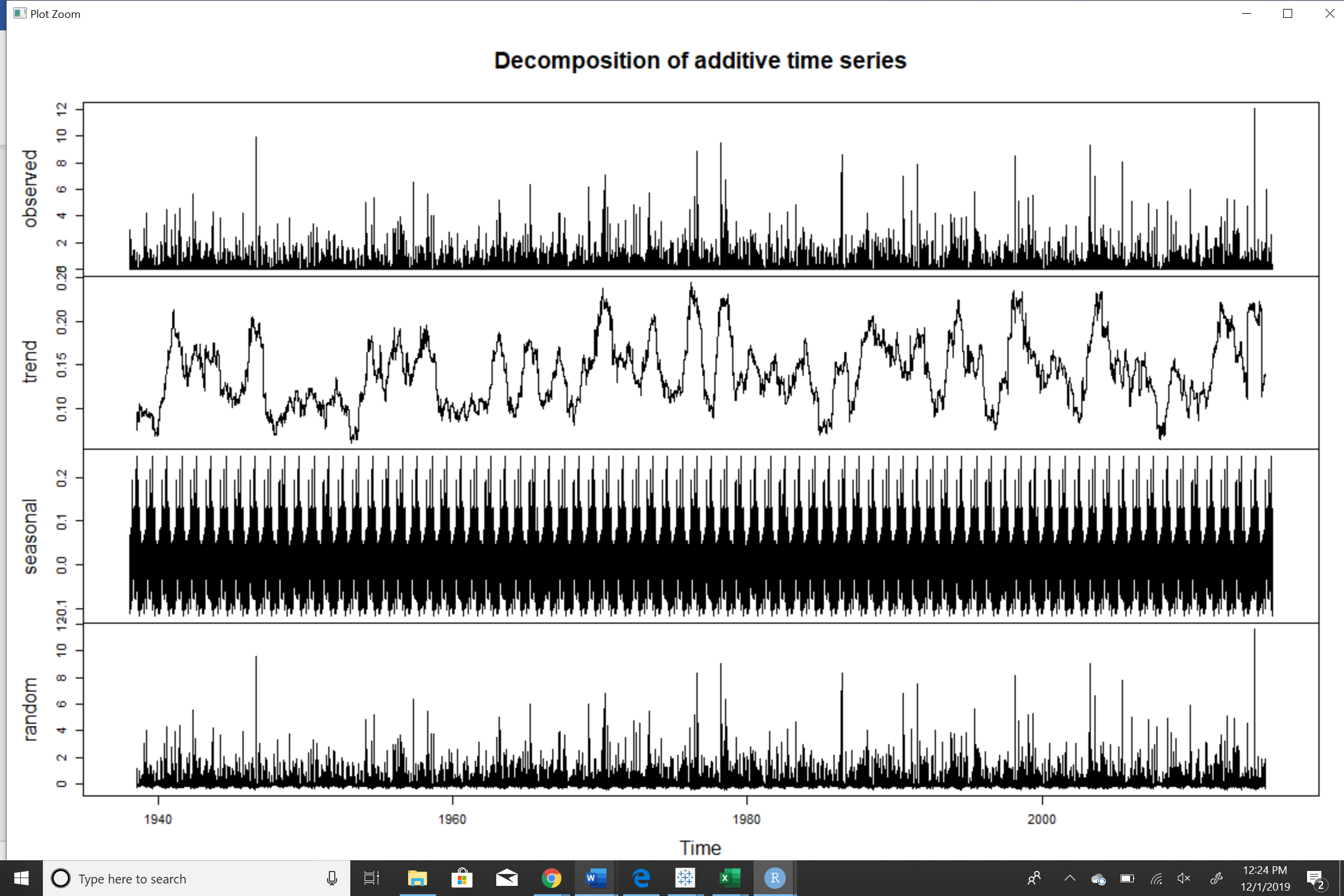
**The above graph is slightly better version of the dataset, it is formed based on monthly average and shows prominent seasonality.**

1. **Decompose the time series into its component parts (decompose)**

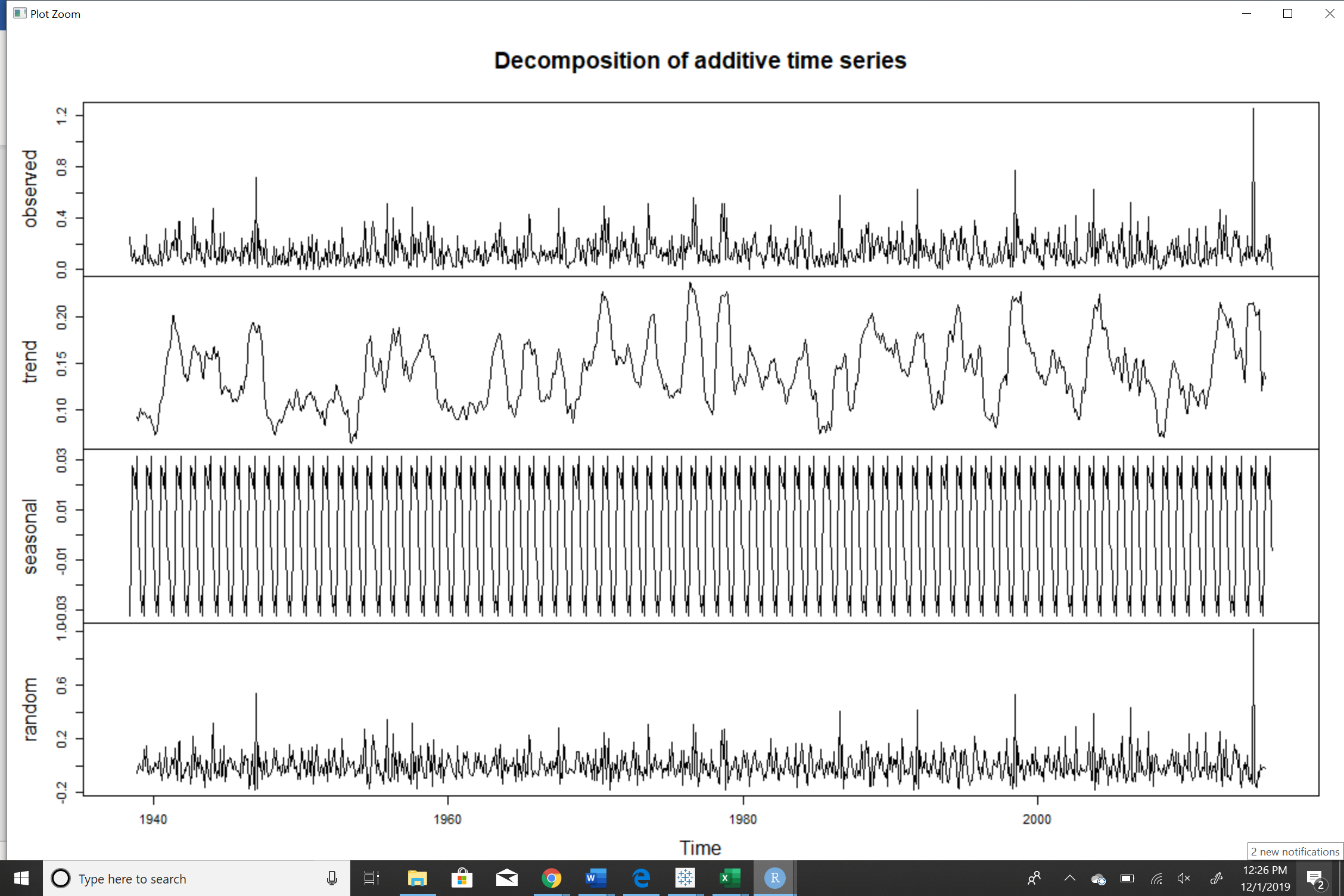
***Observed: actual value***

***Random: removal of trend and seasonality from actual gives random***

**Full data:**

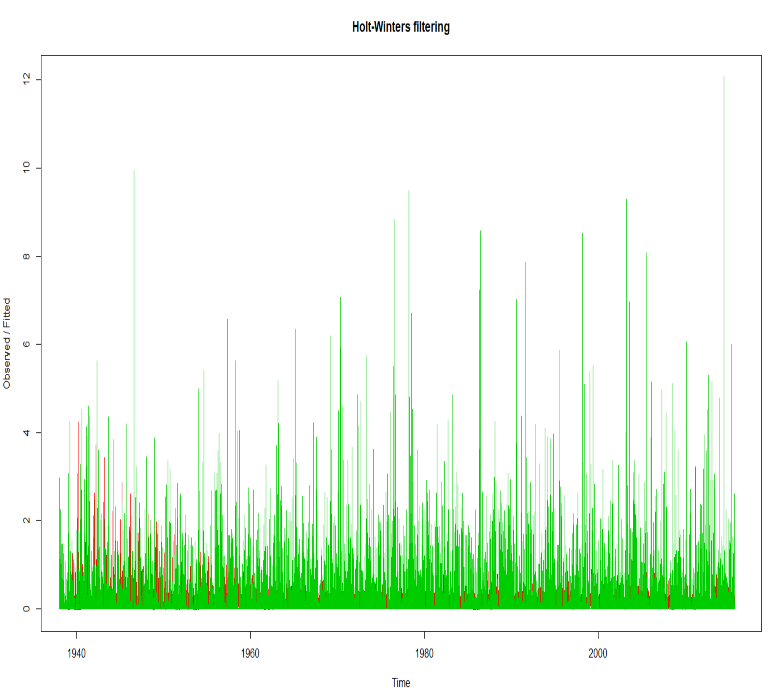
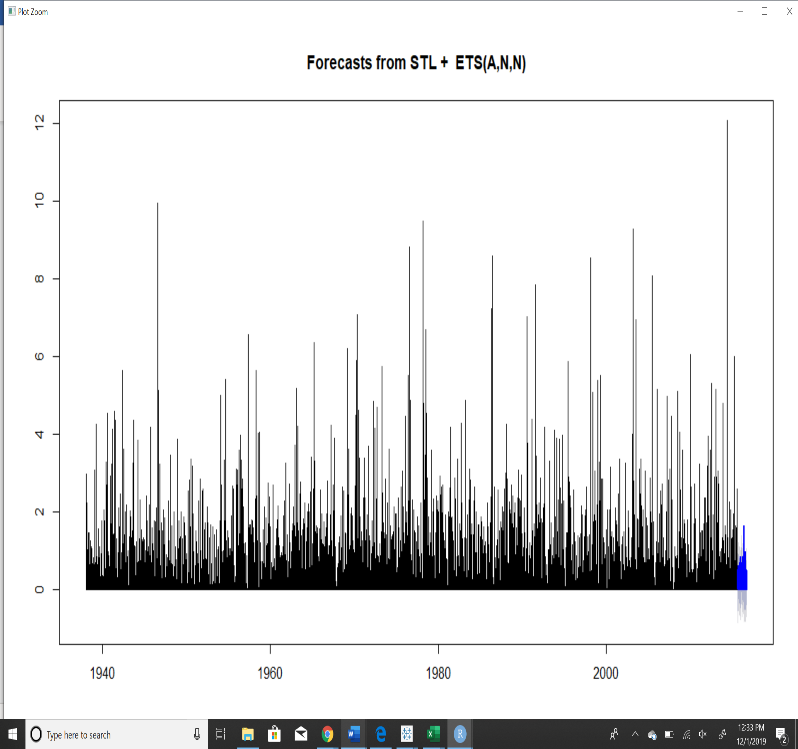


**Monthly Data:**



1. **Model and forecast using exponential smoothing**

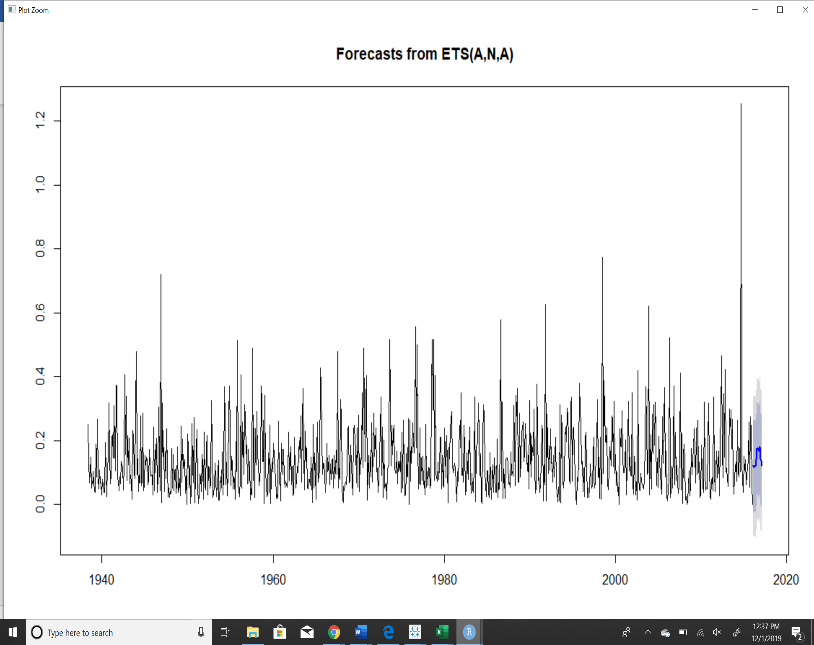
Full Data1$SSE

[1] 7437.527

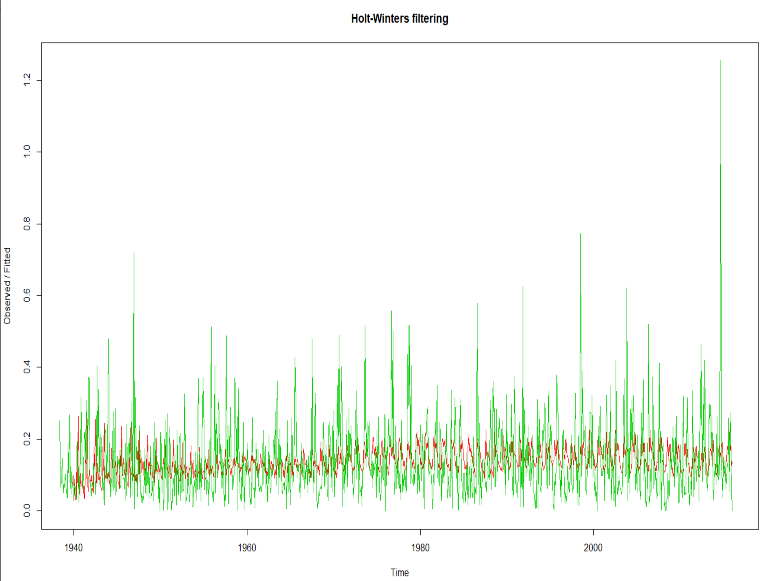
**From the above graph it can be understood that the prediction is not accurate.**

**From the graph, it is clear that at the beginning it was a better predictor than at the ends. This can be inferred from the red lines spiking up in the beginning and decomposing at the end**

Monthly Data$SSE

[1] 12.46868

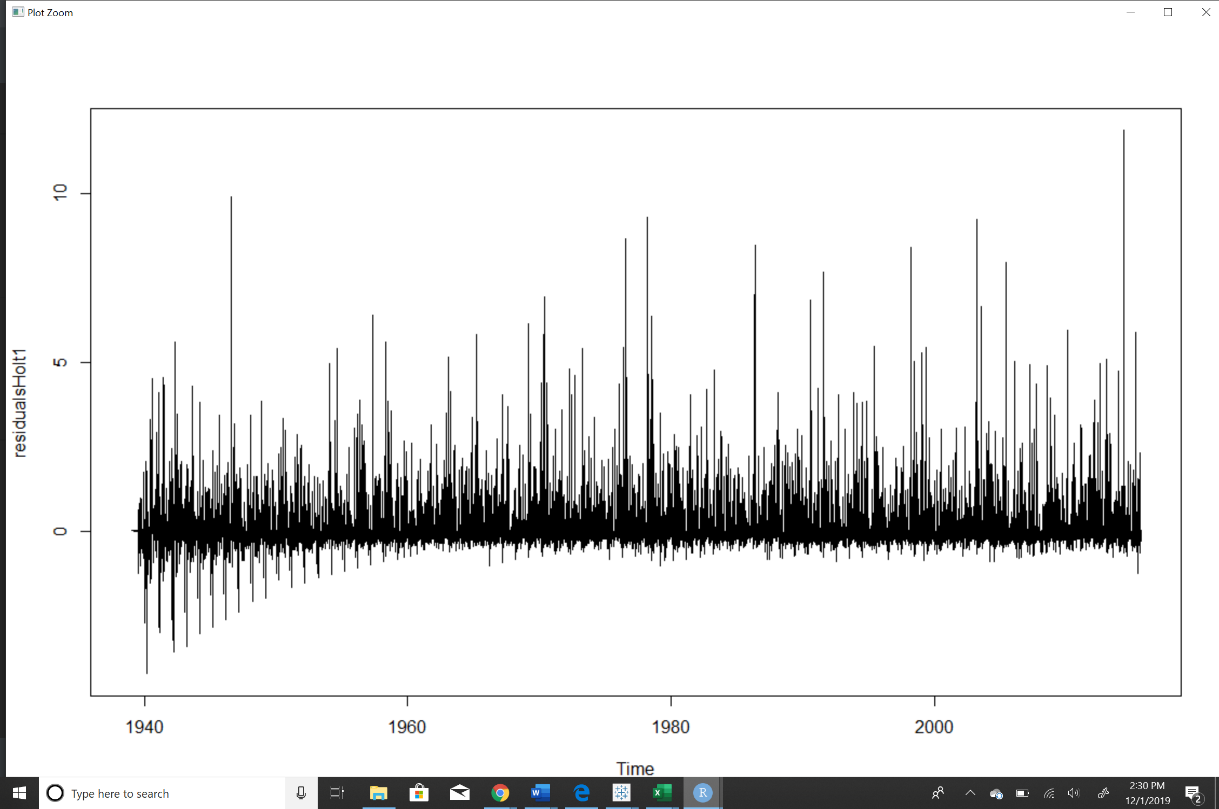
**From the above graph it can be understood that the prediction is slightly promising and shows a cyclical pattern.**

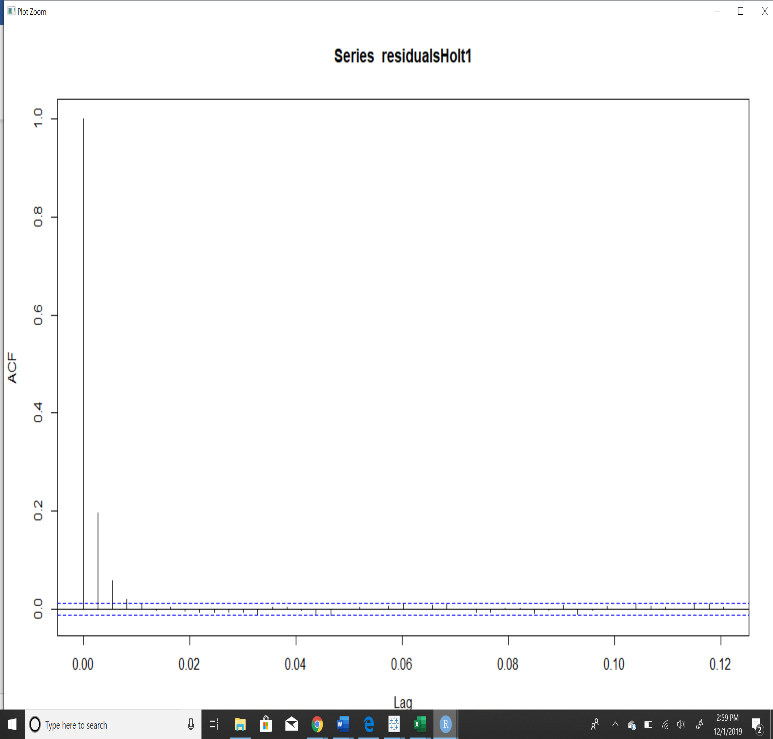


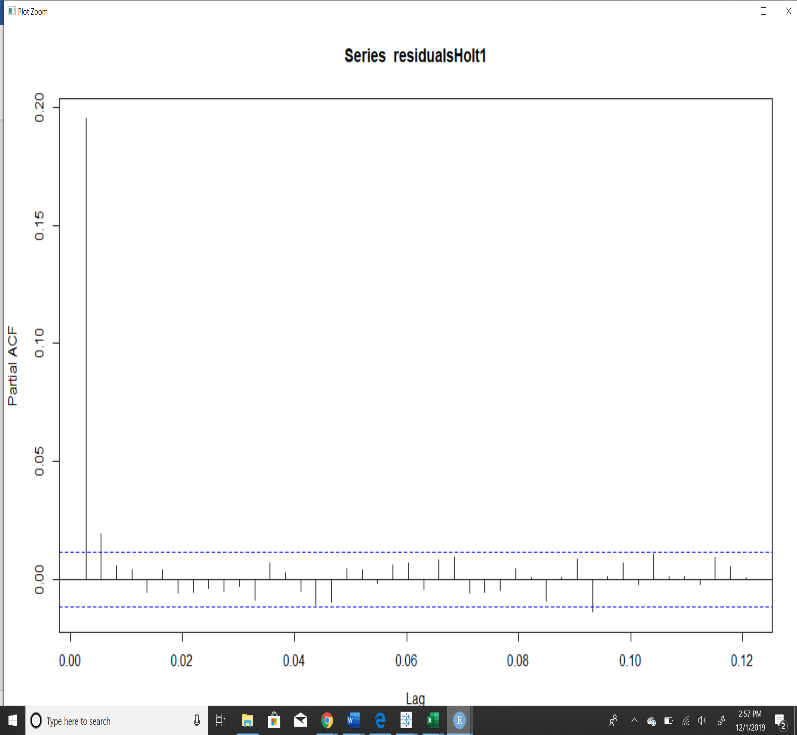
**From the graph, Monthly data shows slightly better prediction than the Full data. The red lines are not spiked, but are evenly spread across, offering better prediction trend.**

1. **Check the model through residual analysis**

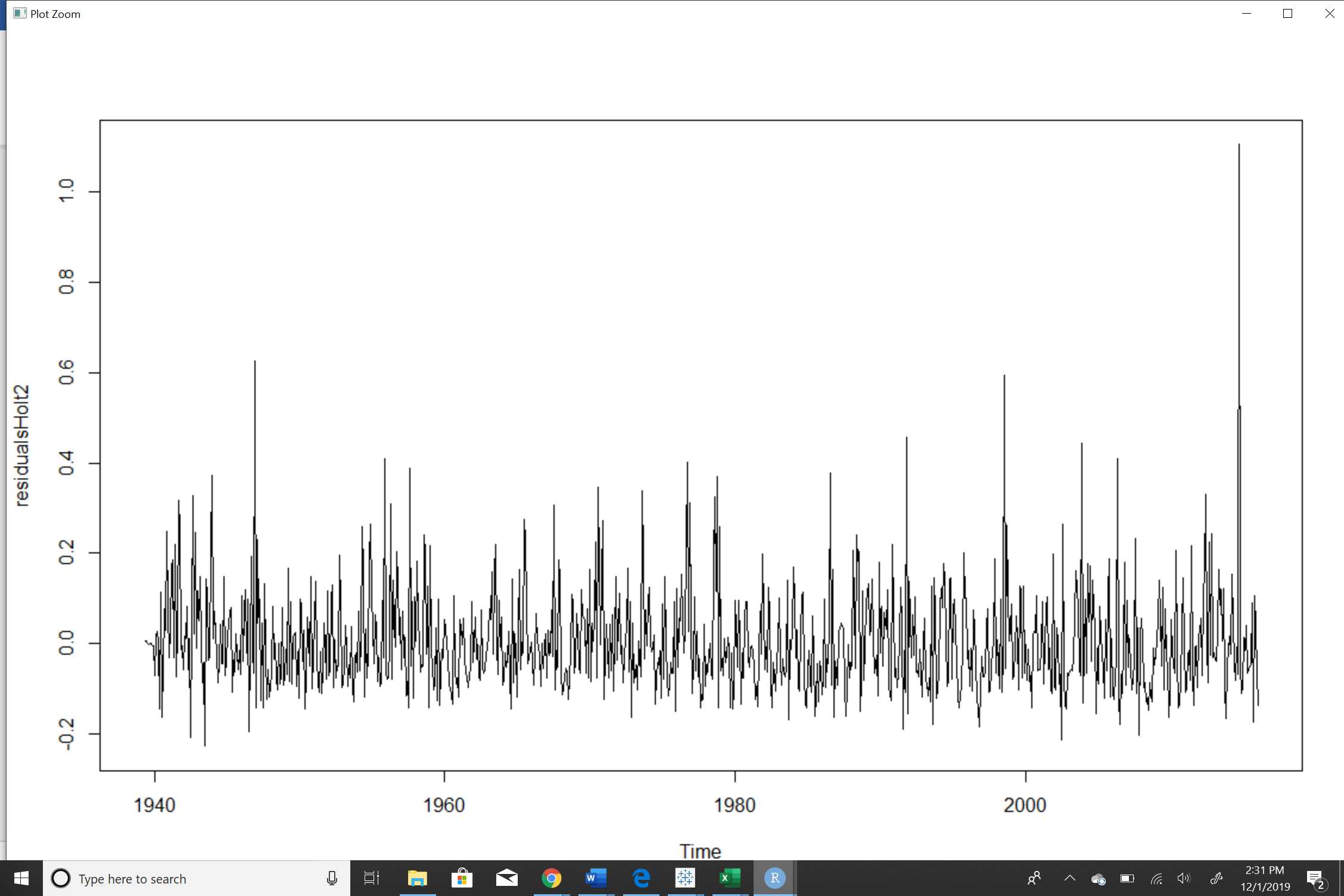
***Full Data:***

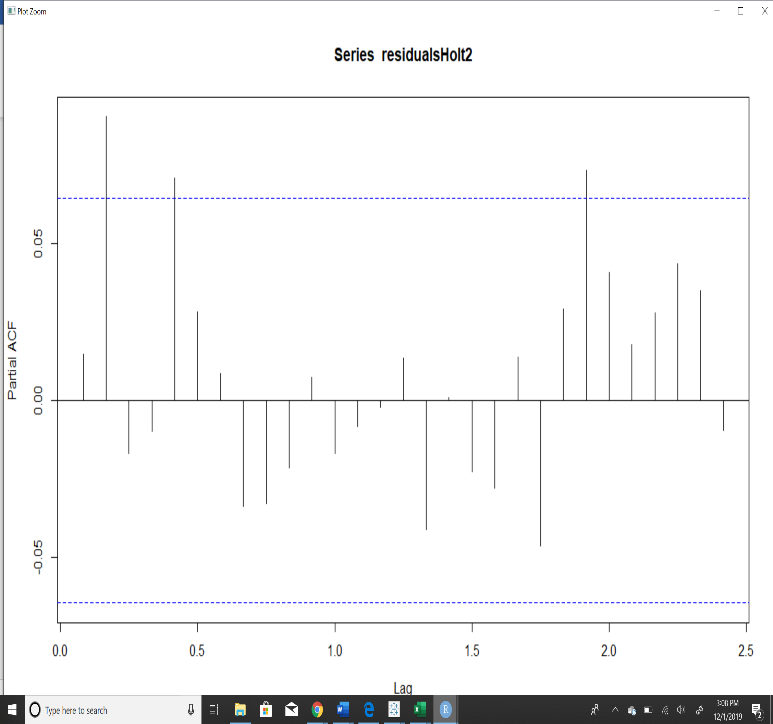
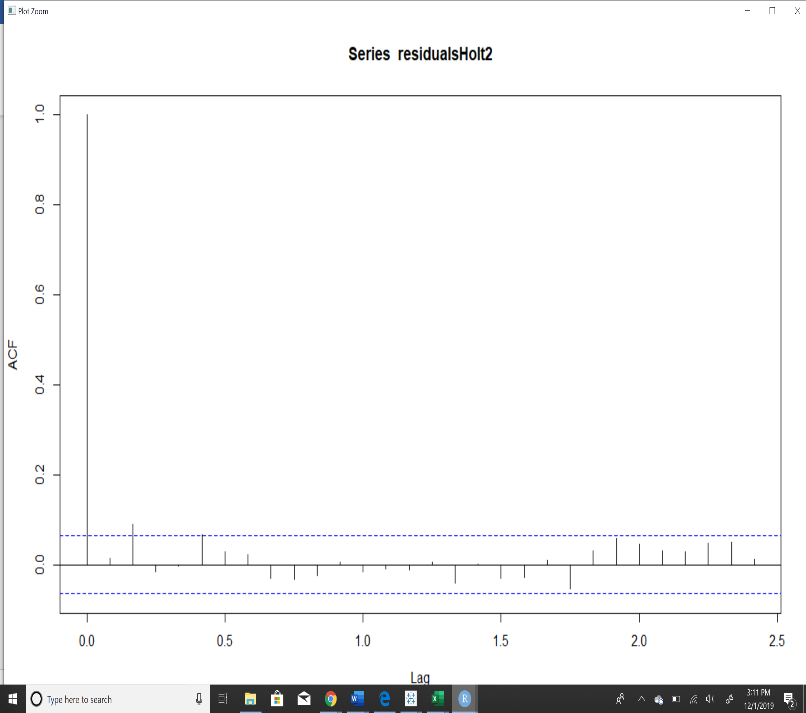


**It can be inferred from the above residual plot that it is not evenly spread for full rain data. They are evenly spread in the beginning and towards the end they are skewed on one side. The model is underpredicting.**



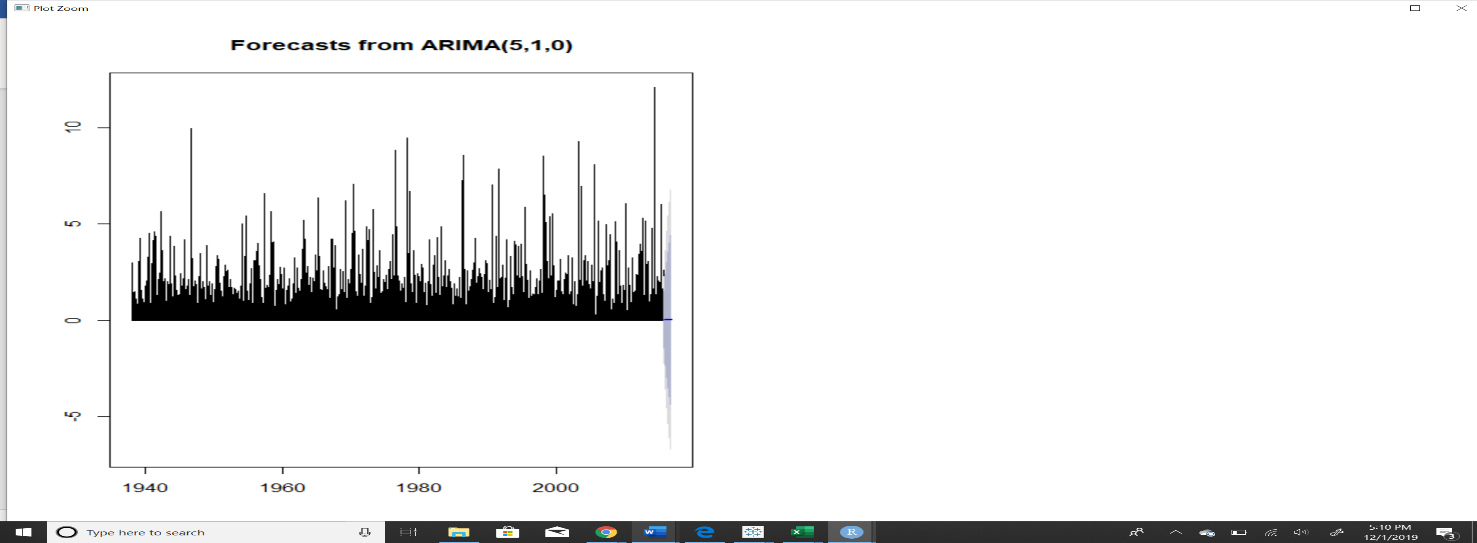
**From the ACF and PACF, it is clear that model with full rain data is not a good predicting model, it under predicts the rain trend in future (The number of lines above the mean blue line is more in ACF and need find a better model).**

***Monthly Data:*** 

**It can be inferred from the above residual plot that it is almost evenly spread out top and bottom along zero. But there are more lines above than below, so that model is still underpredicting but slightly better than 365 data.**

**From the ACF and PACF, it is clear that model with monthly average rain data is a good predicting model, it slightly shows better prediction of the rain trend in future (The number of lines above the mean blue line is less in ACF and makes a better model).**

1. **Model and forecast using an auto arima model**

* ***Full Rain Data:***

**Best model: ARIMA(5,1,0)**

Series: Rain

ARIMA(5,1,0)

Coefficients:

ar1 ar2 ar3 ar4 ar5

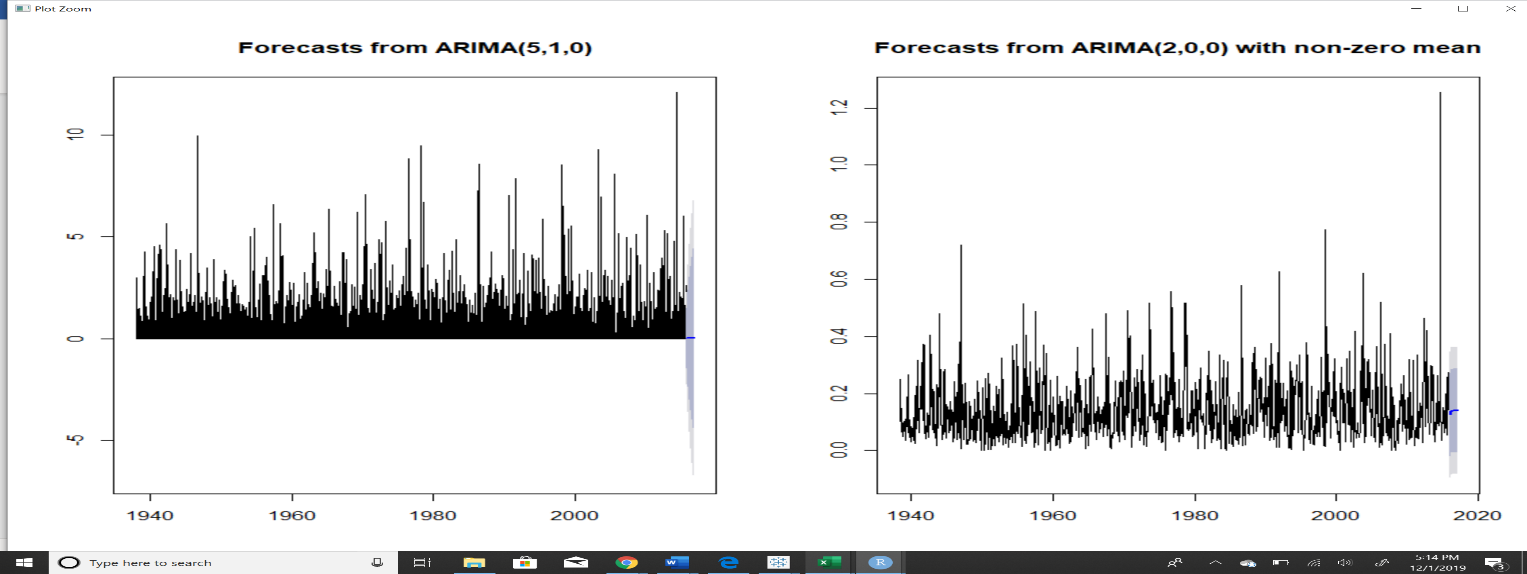
-0.6603 -0.5161 -0.3945 -0.2674 -0.1542

s.e. 0.0059 0.0069 0.0072 0.0069 0.0059

sigma^2 estimated as 0.2694: log likelihood=-21640.35

**AIC=43292.69** AICc=43292.7 BIC=43342.21

* ***Monthly Rain Data:***



**Best model: ARIMA(2,0,0**)

Series: MonthlyRain2

ARIMA(2,0,0)

Coefficients:

ar1 ar2 mean

0.0506 0.0970 0.1404

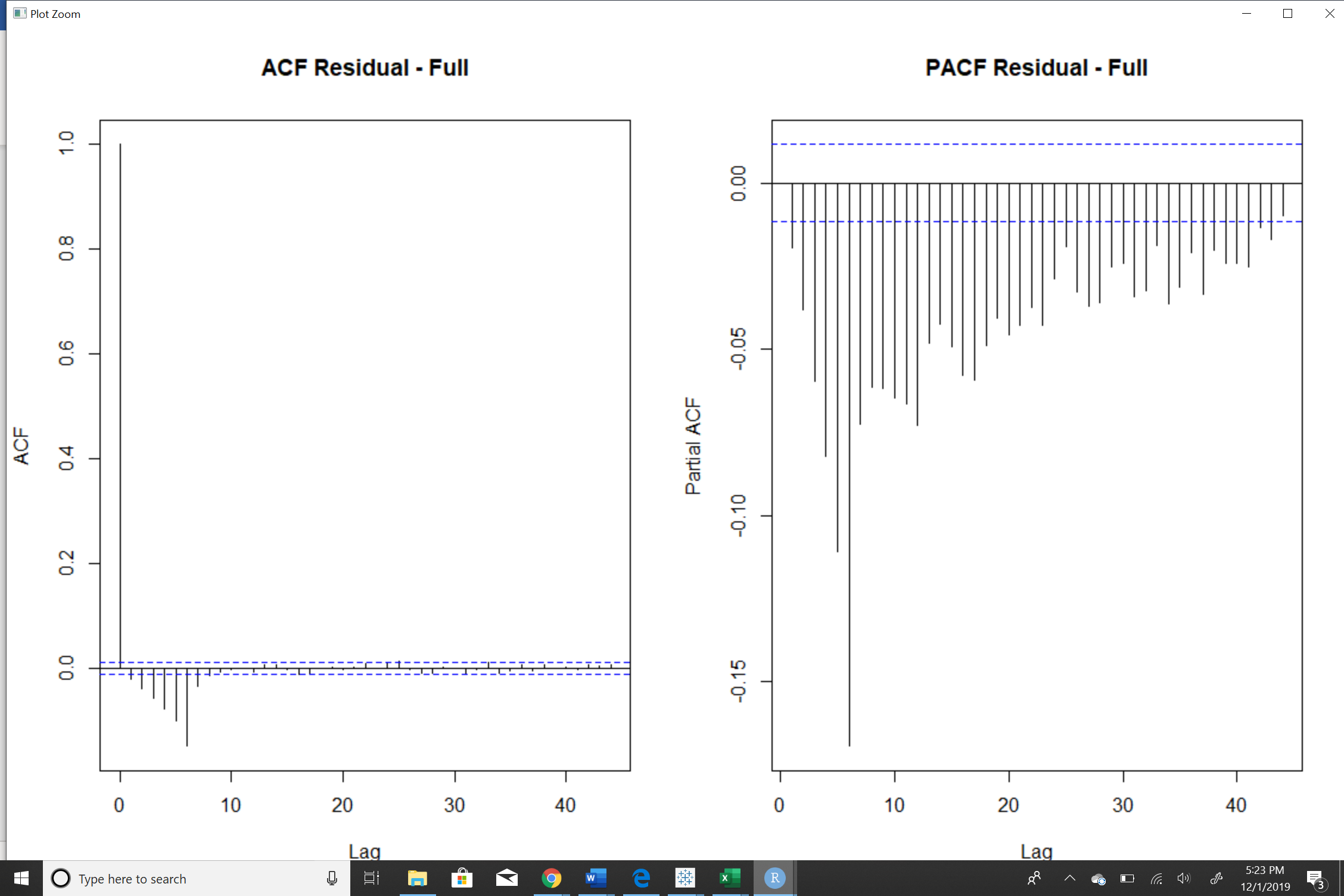
s.e. 0.0326 0.0326 0.0043

sigma^2 estimated as 0.01278: log likelihood=711.53

**AIC=-1415.06** AICc=-1415.02 BIC=-1395.71

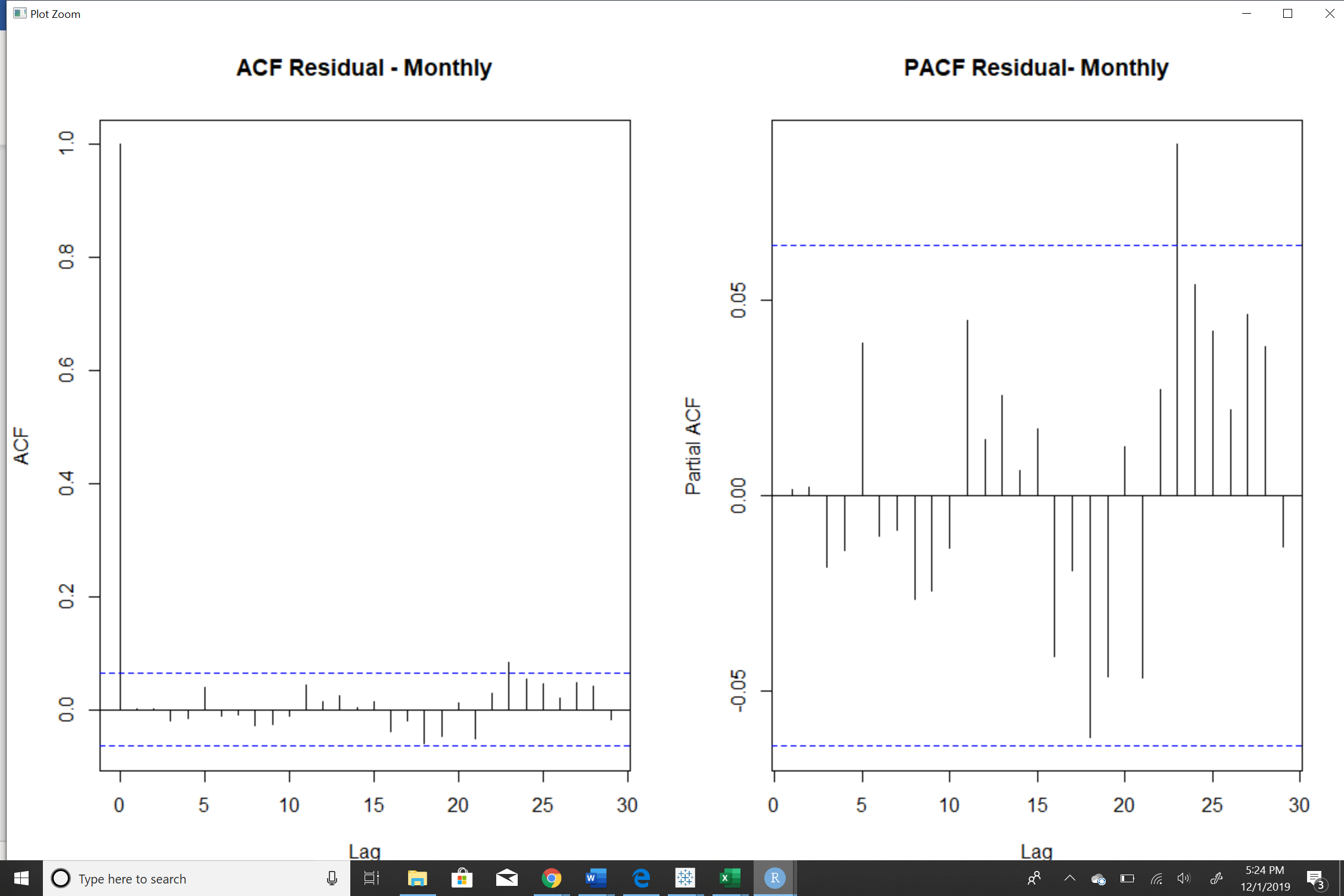
1. **Check the model through residual analysis**

* ***Full Rain data***



**From the ACF and PACF, it is clear that model with full rain data using auto arima is a bad predicting model, in fact exponential smoothening model gave a better result than this.**

* ***Monthly Rain data***



**From the ACF and PACF, it is clear that model with monthly average rain data is a good predicting model, it slightly shows better prediction of the rain trend in future (The number of lines above the mean blue line is less in ACF and makes a better model).**