**Module 5 – Time Series in R Assignment**

**Name: Sunil Raj Thota**

**NUID: 001099670**

**Date: 02/17/2021**

**Title: Time Series in R Assignment**

**ALY 6015 – Intermediate Analytics**

**Prof. Roseanna Hopper**

**Introduction**

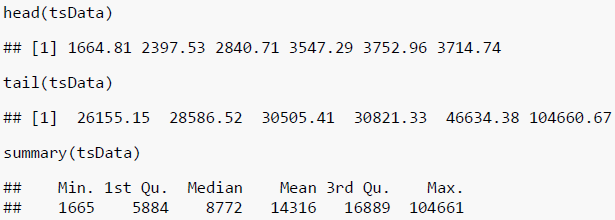
In this task, I used R and R studio to perform Descriptive Statistics and Time Series Analysis on the sales of souvenirs and volcanoes dataset from the Time Series Data Library, by Rob J Hyndman Professor. We can notice a list of all data sets in R using the function called data().

The datasets used for the analysis are the “sales” and “volcano” to discover the time series analysis, seasonality, smoothing, and components, ARIMA on the variables. Also, used additional packages such as forecast, TTR, and SMA libraries.

**Analysis**

**Part A**

**Problem 1**

****

Let's perform some Exploratory Data Analysis and Time series analysis using "sales" data set. To get this data set we need to search on Google as Monthly Sales for a Souvenir Shop in Australia from Jan 1987 - Dec 1993. After that, i have loaded the data set using scan() function. I also installed all the necessary package from the packages tab which is right side to the work space in R Studio.

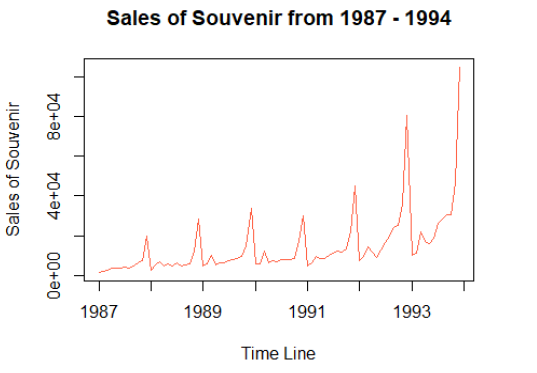
Or we can also install the packages by using install.packages("package name") command. Once it is loaded, we can use it in the code for further analysis and calculations. Loaded the necessary library into the work space. Loaded the sales Data set into the Environment.

I have also installed TTR package enables us to use SMA function which is required to smooth time data series by using Simple Moving Average. I have also installed 'TTR', 'SMA', to perform Smooth Time Data series by simply moving the averages. Let's install all the above packages.

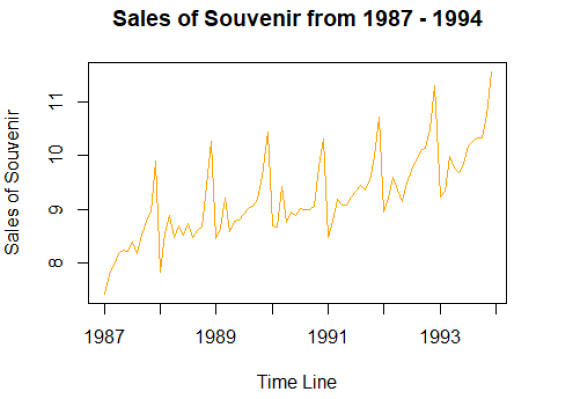
To View the diabetes Data set we use View() command, To observe the structure of the Data set we use str() command, and head () and tail() shows first and last few rows in the Data set. Summary() Provides the Descriptive Stats of the sales columns. We noticed 5 variables from the statistics given in the summary.

**Problem 2**

I have utilized the "ts" function to time series data. I took frequency as 12 to format the data into 12 levels as the months are 12 from 1987 - 1993. I used "ts.plot" function to plot the time series data. In this plot, X Axis depicts the time line and Y Axis plots the number of sales of souvenir. The plot is of Additive type which is not applicable to utilize it for the time series analysis.

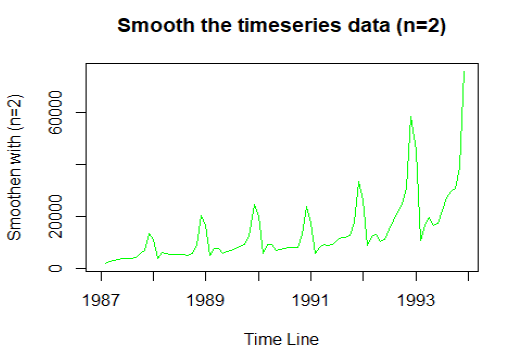


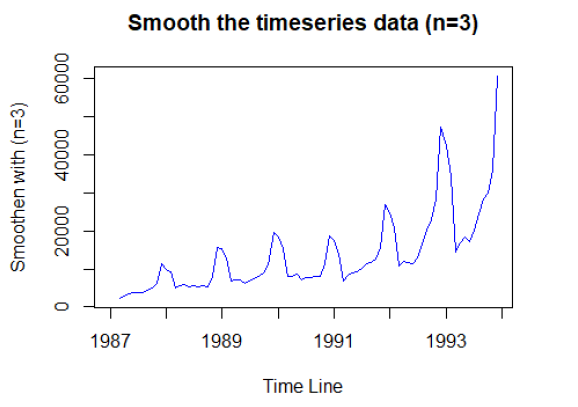
Because the difference between consecutive time series data is same. To analyze time series data, I need to convert it to multiplicative types by taking "Log" natural for the above dataset. Let's plot the data as shown above. I had used ts.plot' but it did not provide any differentiation and produced the same result.

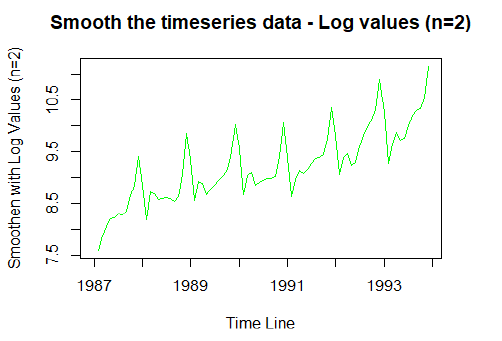


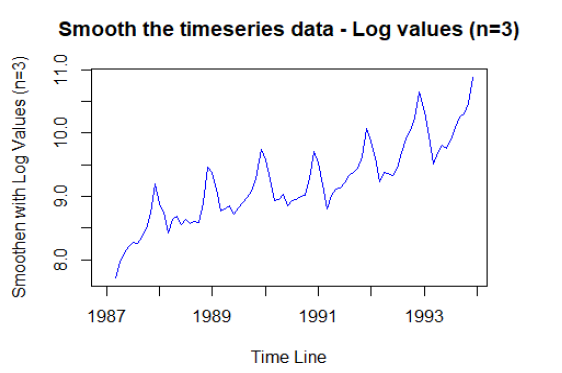
**Problem 3**

Let's smoothen the above four plots by using simple moving averages "SMA()" function. This function usually requires 2, 3 consecutive numbers and avg. them and take the next consecutive by averaging the data set. Let's use various values of "n" to alter the smoothing level.

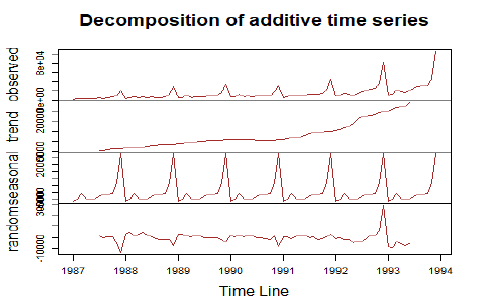


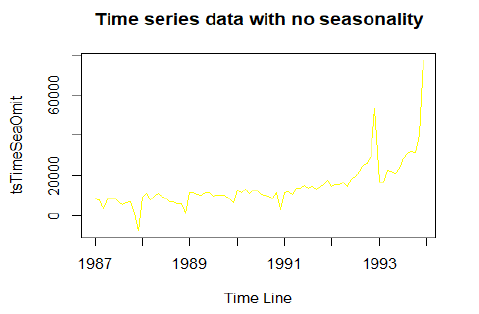






The peaks in the time series analysis is determined by SMA. Let's now analyze the components of a time series by using the "decompose()" function to segregate various components.



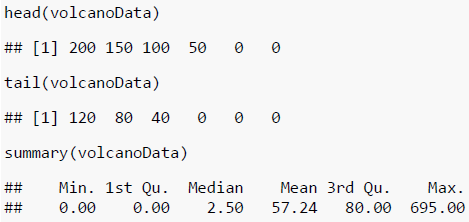


After that, I have plotted the graph to analyze these components. In this, we already saw that there are four various components in this time series analysis as observed, trend, random, and seasonal. The seasonal attribute is recurring over the time line and is capable. To get a sure shot on this data we need to eradicate the seasonal aspect which does not give exact analysis of the trends.

Once that component is eradicated from the analysis we can see more precise information on the time series data behavior to observe the rise and fall of the data. From the plot we can depict that the unseasoned hike in the end

**Part B**

**Problem 1**

****

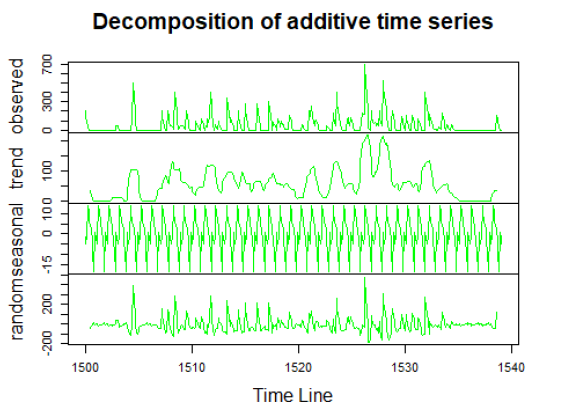
Let's perform some Exploratory Data Analysis and Time series analysis using "volcano" data set. After that, I have loaded the data set using scan() function. I also installed all the necessary package from the packages tab which is right side to the work space in R Studio.

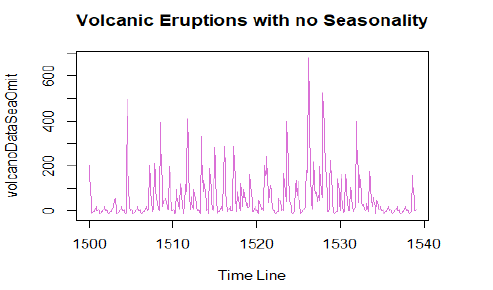
Or we can also install the packages by using install.packages("package name") command. Once it is loaded, we can use it in the code for further analysis and calculations. Loaded the necessary library into the work space. Loaded the sales Data set into the Environment. I have also installed forecast package enables us to forecast. Let's install all the above packages.

To View the diabetes Data set we use View() command, To observe the structure of the Data set we use str() command, and head () and tail() shows first and last few rows in the Data set. Summary() Provides the Descriptive Stats of the volcano columns. We noticed 5 variables from the statistics given in the summary. Let's use the ARIMA to discover the correlations and its problems

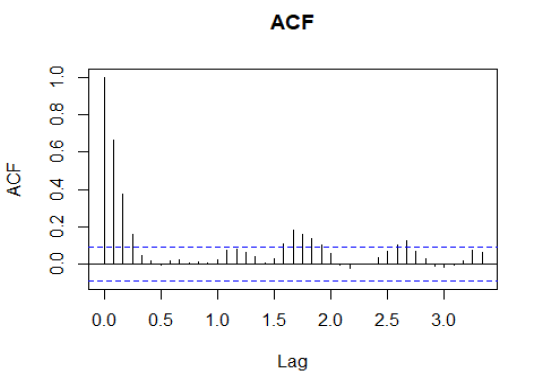
**Problem 2**

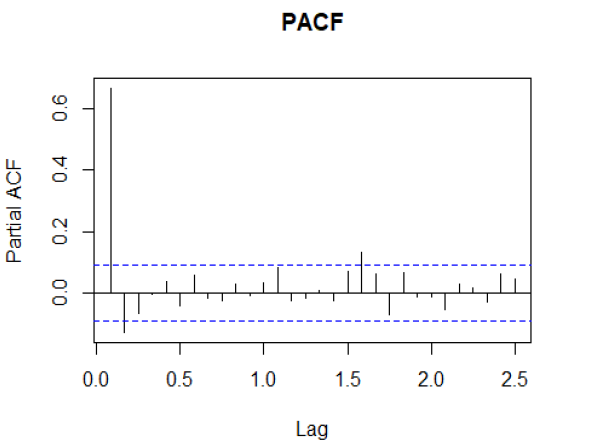
Let's check the time series data for Volcano and took frequency as 12 because it has 12 levels for 12 months. To check for seasonality let's use decompose() and eradicate this component from the data and observe the volcanic eruptions trends.



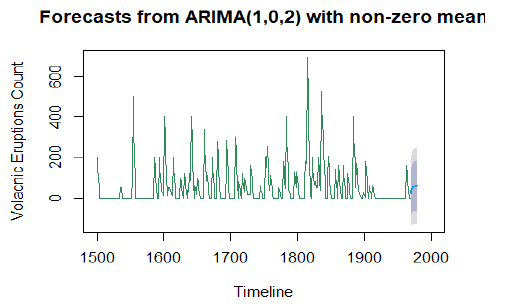


The plot depicts the total number of volcanic eruptions from 1500 - 1540. And, also let's use acf() and pacf() functions to check correlation.





Where acf() is auto correlation and pacf() is partial correlation. acf() performs auto-correlation on the time series data with lagged attributes. Where pacf() is partial auto-correlation function that is used to observe the residuals correlation.



Let's create time series with ARIMA function for the volcanoes data set. auto.arima() function checks the best value of q, p, and automatically. forecast() function is used to predict the volcanic eruptions trends. We can see that a prediction at the timeline end.

**Conclusion**

To perform multiple analytical operations, we use R on a wider range of data like various datasets. R is a very powerful tool to perform analysis which is mainly built by researchers, statisticians, and developers. In the beginning, we did Descriptive Statistics Analysis, Exploratory Data Analysis, and continued with the time series analysis, seasonality, smoothing, and ARIMA to get understandings from the provided data sets. Plotted relevant parameters, to helps us to understand the data in a clear way. Used forecast and TTR libraries.

**References**

[1] Prof. Rob Hyndman Using R for Time Series Analysis was retrieved from https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html#decomposing-time-series

[2] Time Series and Forecasting was retrieved from https://www.statmethods.net/advstats/timeseries.html

[3] Ruslana Dalinina (January 10, 2017) Introduction to Forecasting with ARIMA in R was retrieved from https://blogs.oracle.com/datascience/introduction-to-forecasting-with-arima-in-r