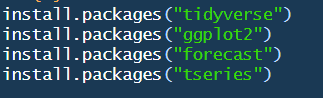
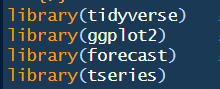
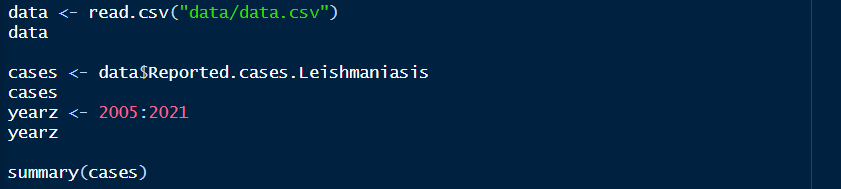
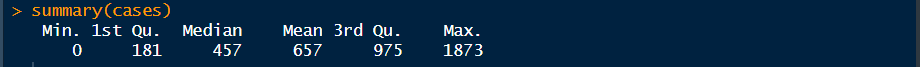
**TIME SERIES PROJECT**

In this section we I am downloading the required packages for the project then later library them to have them accessible in the environment.



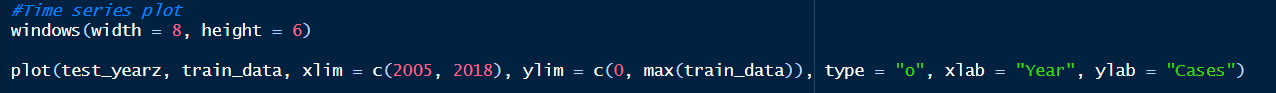


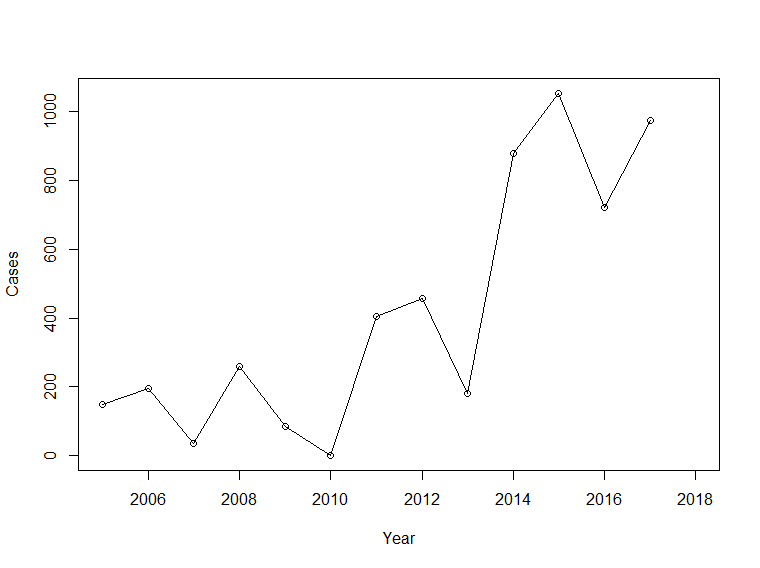
**Tidyverse** : Used for data manipulation  
**ggplot2**: Used for data visualization  
**forecast**: Used for time series forecasting and ARIMA  
**tseries**: Used for time series  
  
   
  
In this section I am reading data from a csv file then singling out the number of cases and the years since this will be very important in our time series model development.  
I am also getting the summary statistics for the cases  
  
  
  
The output of the summary statistics  


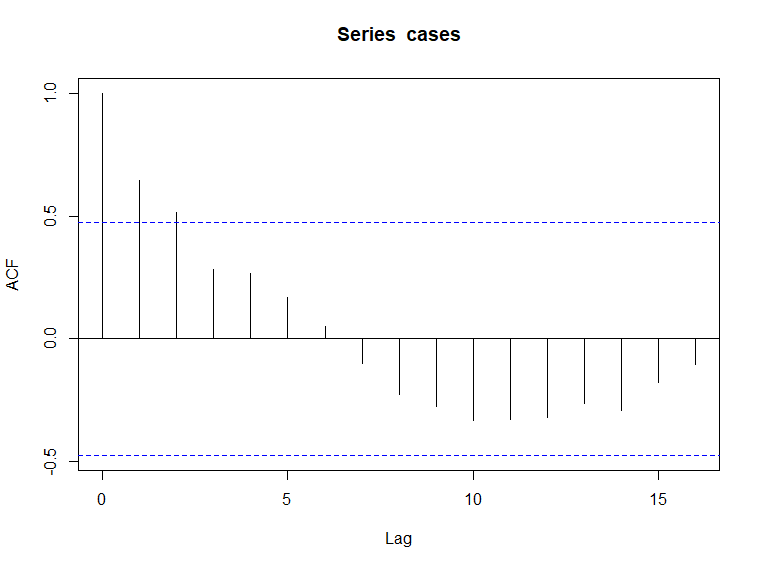
The mean of the cases reported is 657  
The range of the reported cases is 1873

At this point, am splitting the data so that I can train the model with one split and predict with the other one. The Training dataset is 17% while the Test one is 25%

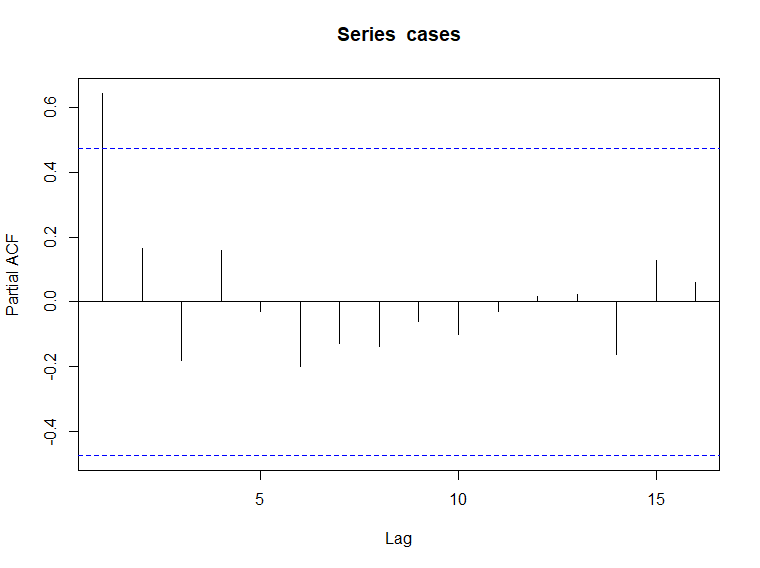


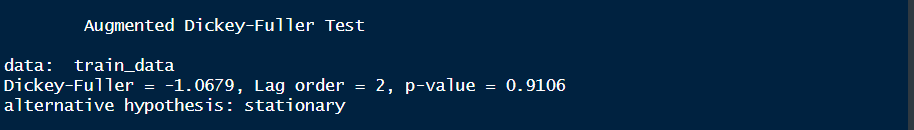
Plotting the train data  
  




This plot show how cases are distributed along the years. It shows the data is not stationary.  
  
inspecting further we check on ACF and PACT at this level  
  
ACF  


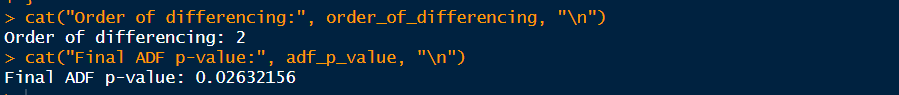
PACF



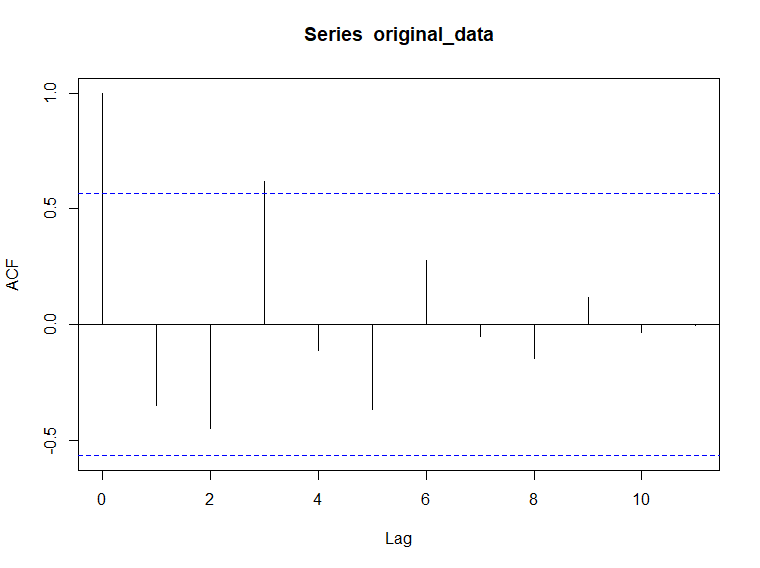
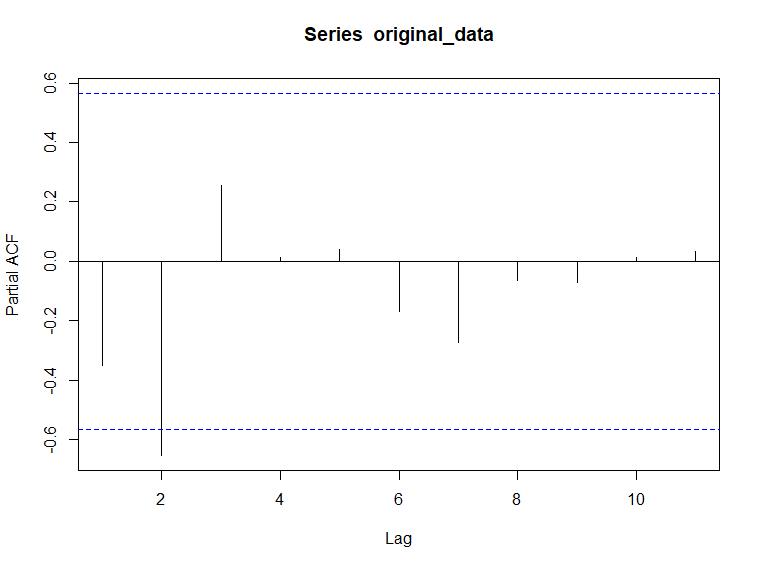
ACF and PACF clearly indicates that our data is not stationary.  
  
Augmented Dickey-Fuller Test  
  


H0 : Time series is non-sationary

H1 : Time series is stationary  
  
From our ADF test results, p-value is 0.9106,

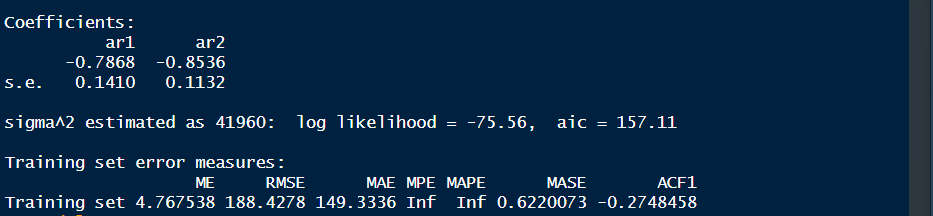
Since 0.9106 < 0.05, we have no enough evidence to reject the null hypothesis thus the time series data is non-stationary  
  
  
How should we make our time series data stationary? We carryout differencing  


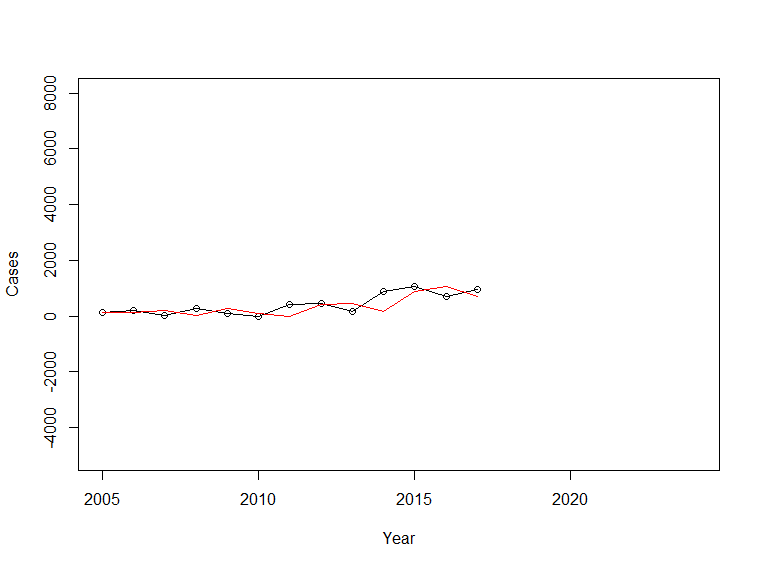
After carrying out differencing we notice that differencing of order 2 makes our data stationary.

ACF and PACF of differenced data respectively  
  
  


From the ACF and PACF above, we can see come up with ARIMA model of the following order  
  
1.ARIMA(2,2,3)  
We can play around with the AR and MA values to get an optimum model.  
  
2.ARIMA(0,2,3)

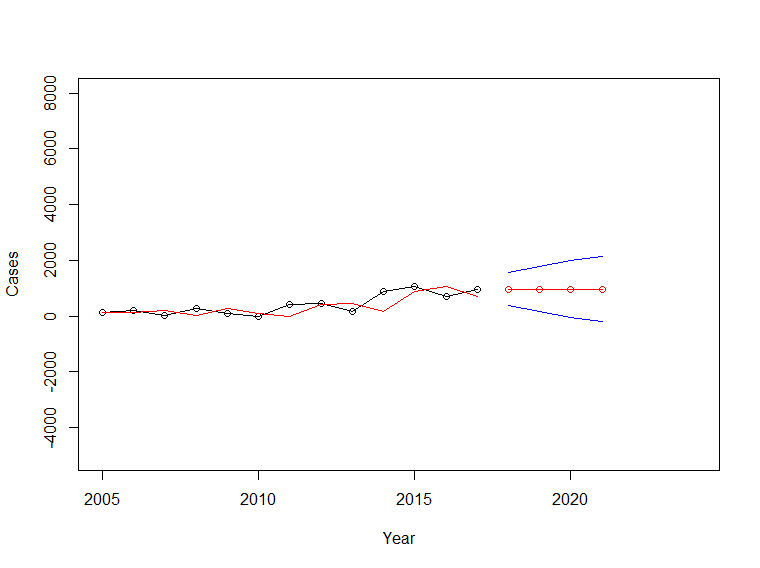
3.ARIMA(2,2,0)

4.Automatic generated model  
  
From the 3 models, ARIMA(2,2,0) seems to have the smallest AIC value of 157.11 as shown  
  


Plotting the residuals to show how far the predicted values fall away from the actual values  
  


The above diagram shows the actual values and how far the predicted values fall away from the actual values. The red line shows the distance between predicted an actual values.   
  
Forecasting the test data, we get the following values

The below plot shows predicted values in the differenced state



Recovered differenced predicted values in the normal state are as shown below

1950 2925 3900 4875