**Predictive Modeling**

**Midterm exam**

**Mutya Sowmya**

**001303248**

1- Decomposition Graph

Features: BIRTH_YR and CNT

#loading the data

data <- read.csv(file = 'C:/Users/sowmy/Downloads/data.csv')

#decomposition graph for BRTH_YR

timeseries <- ts(data$BRTH_YR, frequency=12)

tscomponents <- decompose(timeseries)

plot(tscomponents)

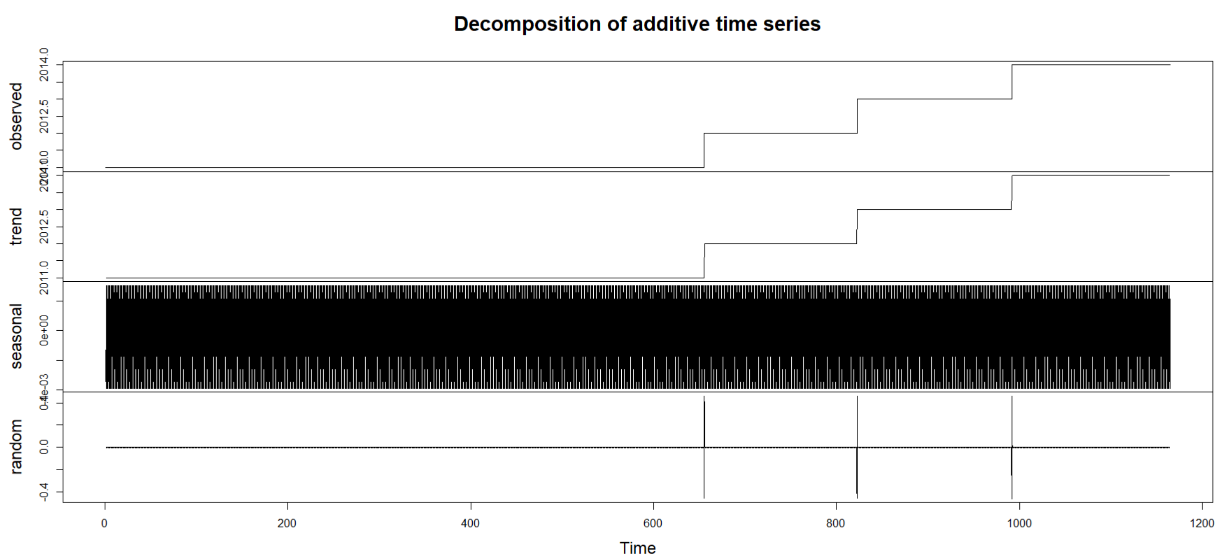
#decomposition graph for CNT

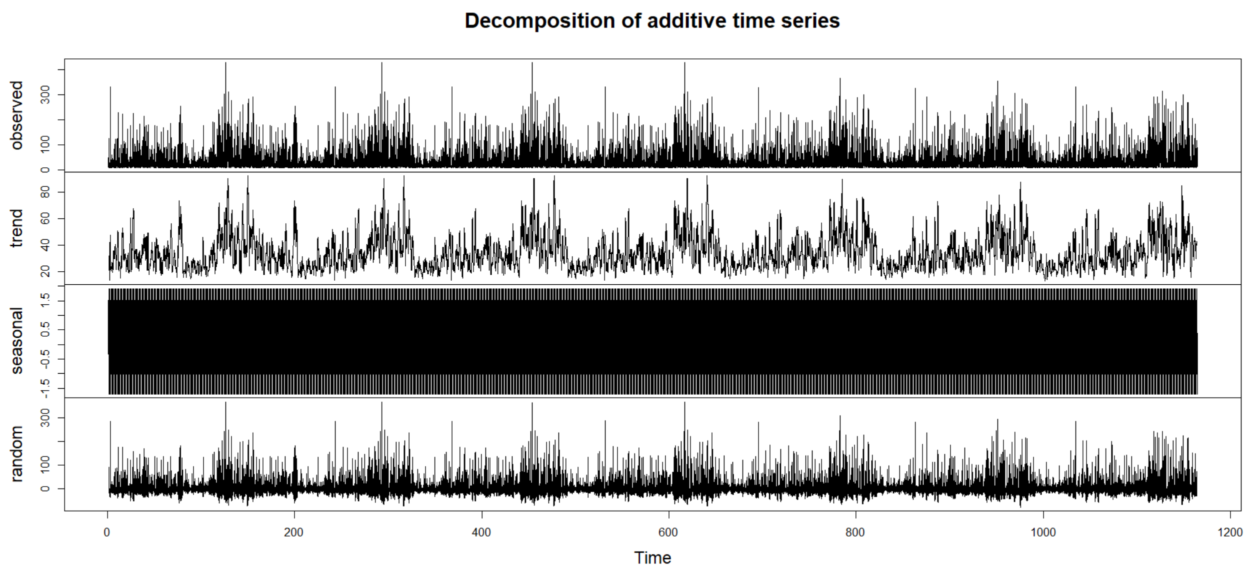
timeseries <- ts(data$CNT, frequency=12)

tscomponents <- decompose(timeseries)

plot(tscomponents)

a) Make the decomposition graph.






b) Is there any seasonality in baby's names in NYC?

Seasonality in BRTH_YR:

Yes, there exists seasonality in BRTH_YR as the pattern repeat is depicted regular intervels.

Seasonality in CNT:

Yes, the decomposed time series of the CNT clearly shows a seasonality trend. The area in the middle of the plot that is colored black represents the seasonal component. A distinctive sign of seasonality in the data is this pattern's regular cyclical fluctuation, which repeats over predictable time intervals.

c) Extract information from graph.

The long-term increasing or declining tendency in the data is represented by the trend component. It appears as a smooth rising curve in the graph.

Monthly or annual changes are examples of the recurrent patterns in the data that are captured by the seasonal component. It appears as a wave-like pattern on the graph.

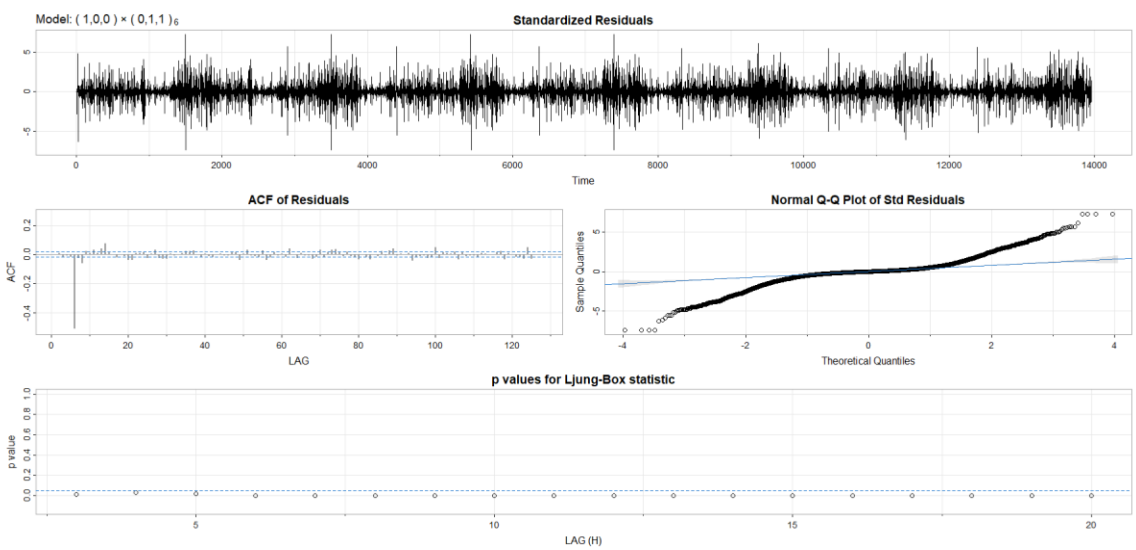
The data's erratic swings that are not represented by the trend or seasonal components are symbolized by the random component. It can be seen as the haphazard spots surrounding the seasonal components and trend in the graph.

2- ARIMA Graph

Features: BIRTH_YR and CNT

```
#Arima graph for CNT
library(astsa)
diff6 = diff(data$CNT, 6)
diff1and6 = diff(diff6,1)
acf2(diff1and6,12)
dtb=residuals(lm (diff6~time(diff6)))
acf2(dtb)
sarima(dtb, 1,0,0,0,1,1,6)
```

a) Make the ARIMA(1,0,0)(0,1,1)6.




b) What is the lag based on ARIMA? Why?

Based on the information provided in the image, the ARIMA model being evaluated appears to be (1, 0, 0) × (0, 1, 1)6, which can be interpreted as follows:

The lag for the seasonal component is 6, as indicated by the (0, 1, 1)6 part of the model specification. This suggests that the seasonal pattern, if present, has a periodicity of 6 time periods.

Like example it can be months quarters and so on…

c) Interpretation of the ARIMA graph:

Standardized Residuals Plot: The residuals are randomly distributed around zero, indicating that the model has effectively captured the data dynamics.

ACF of Residuals: There are no notable spikes beyond the confidence intervals in the autocorrelation function (ACF) plot, suggesting that the residuals do not exhibit significant autocorrelation patterns.

Standard Q-Q Plot: The points on the standard quantile-quantile (Q-Q) plot align well with the theoretical quantile line, indicating that the assumption of normality for the residuals is reasonable.

Ljung-Box Test: The p-values from the Ljung-Box statistic indicate that the null hypothesis of no autocorrelation in the residuals cannot be rejected at common significance levels for the provided lags, further supporting the randomness of the.