

Assignment on

Time Series Analysis & Forecasting

WM-ASDS 10

(Chapter 1 to 5)

**Ms. Nasrin Khatun**

Department of Statistics & Data Science

Submitted by

**Shourav Majumder**

**20231046**

10th batch, Section: B

Professional Masters in Applied Statistics & Data Science

Department of Statistics

**Jahangirnagar University**

**CHAPTER 1 (Time Series Data)**

**Definition:**

When data is collection over a period of time then it is called time series data.

* Daily IBM stock price
* Annual Google profit

**Types of Time Series Data:**

* Cross sectional data
* Panel data
* Longitudinal data

**Cross sectional data:**

Cross sectional data refers to a type of data collected by observing many subjects, individuals at the same point of time. Each observation represents a distinct entity.

Single point of time, multiple variables and no time dimension.

**Panel data:**

The combination of time series and cross-sectional data is called panel data. Like Inflation rate of Bangladesh between 1970 to 2022.

**Longitudinal data:**

This data refers to a type of data collected by following the same subjects, individuals or entities over a period

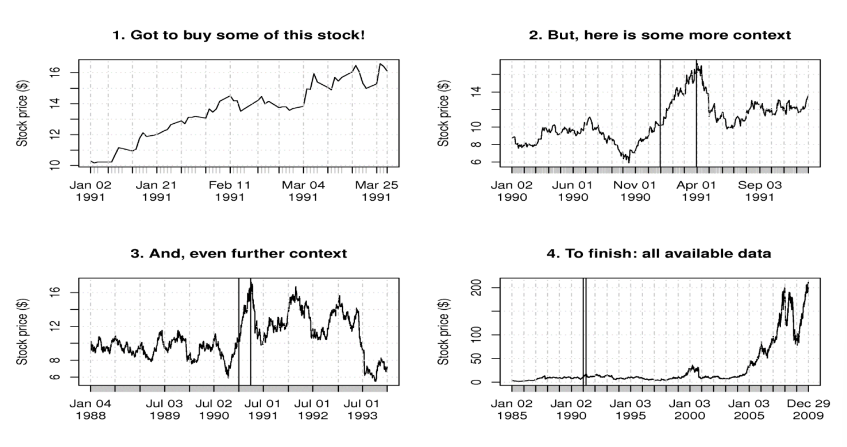
. 

Fig: Graphical Plot of Time Series Data

**Time Series Definition:**

* It is a sequence of observations on a variable measured at successive points in time or over a period of time.
* It is a series of data points indexed in Time order.
* It is a sequenced taken at a equally spaced points in time.

**Time Series Analysis:**

* Time series analysis involves developing models that best capture in order to understand the underlying causes.
* Primary objectives to develop mathematical models that can describe those datasets.
* Analyze data in order to extract meaningful statistics & others characteristic.

**Different types of Time Series**:

* Discrete Time Series: when observations are taken only at a specific time, usually equally spaced.
* Continuous Time Series: When data points are measured or recorded continuously over a specified time interval.

**Components of Time Series:**

* Secular trend
* Cyclic Variation
* Seasonal Variation
* Irregular Variation

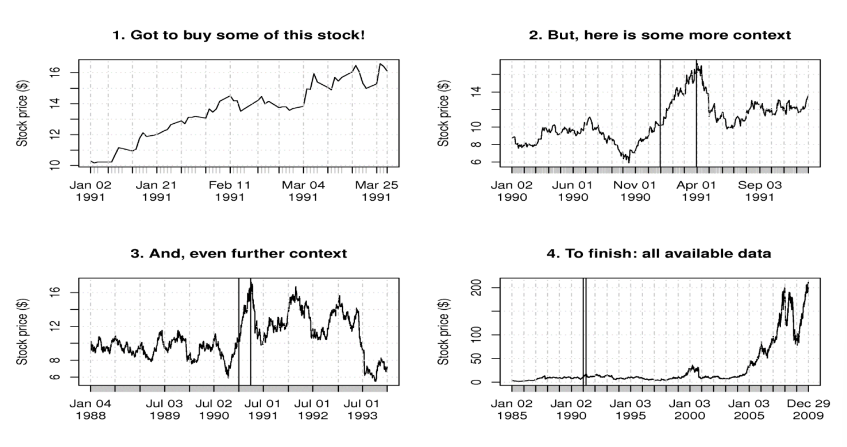


Fig: Decomposition of a Time Series Data

**Forecasting:**

Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecast.

It is a common statistical task in business, where it helps to inform decision about scheduling of production, transportation, long term strategic planning.

There is three types of forecast, short term, medium term and long term forecast.

**Software we use for forecasting:**

R programming, Python Programming, Minitab ETC

**CHAPTER 2 (Time Series Graphics)**

**Time Plot in R:**

R code for autoplot.

autoplot(a10)+

ggtitle("ANTIDIABETIC DRUG SALES")+

ylab('million')+

xlab("year")

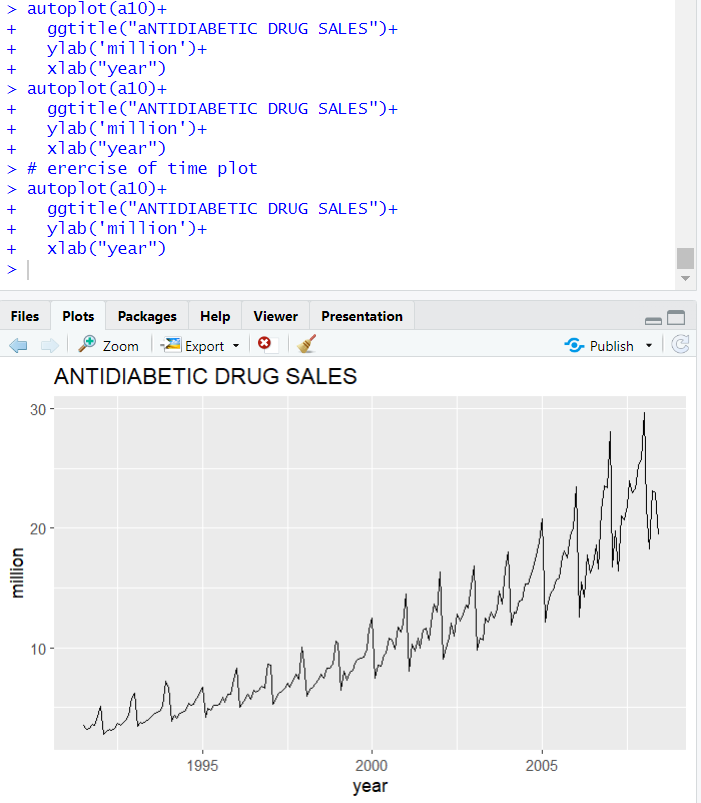


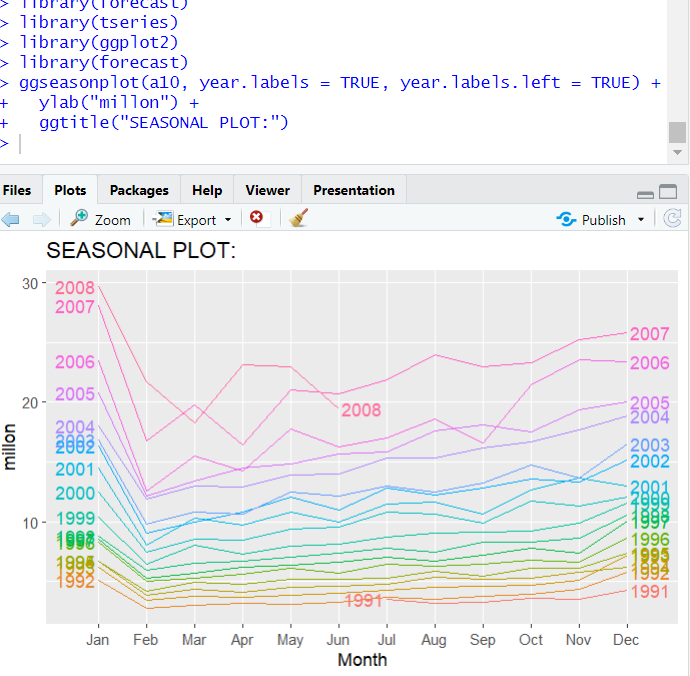
Fig: Time Series Plots

**Seasonal Plots:**

ggseasonplot(a10, year.labels = TRUE, year.labels.left = TRUE) +

ylab("millon") +

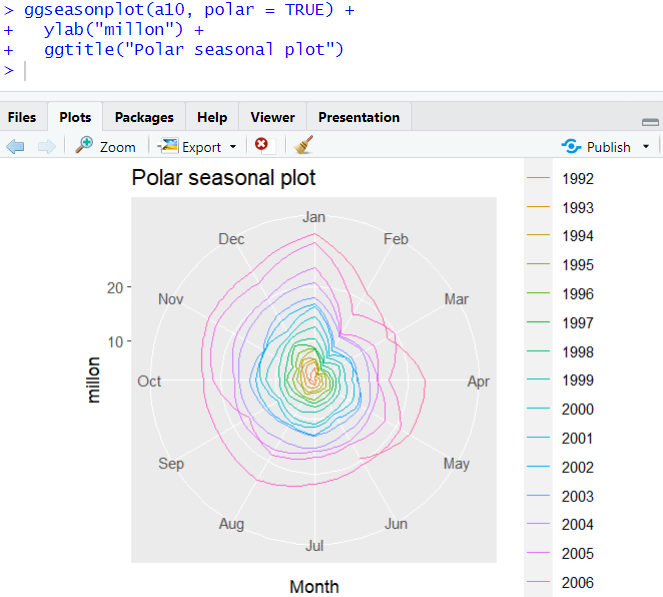
ggtitle("SEASONAL PLOT:")



ggseasonplot(a10, polar = TRUE) +

ylab("millon") +

ggtitle("Polar seasonal plot")

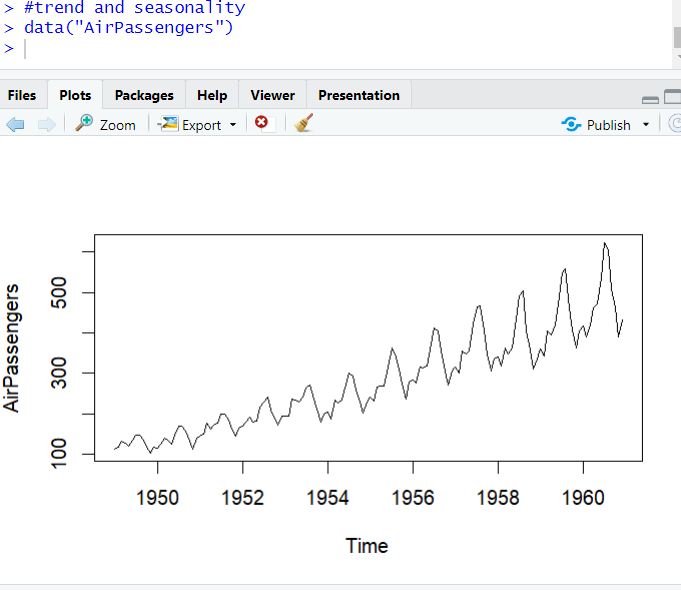


**Scatter Plot:**

#trend and seasonality

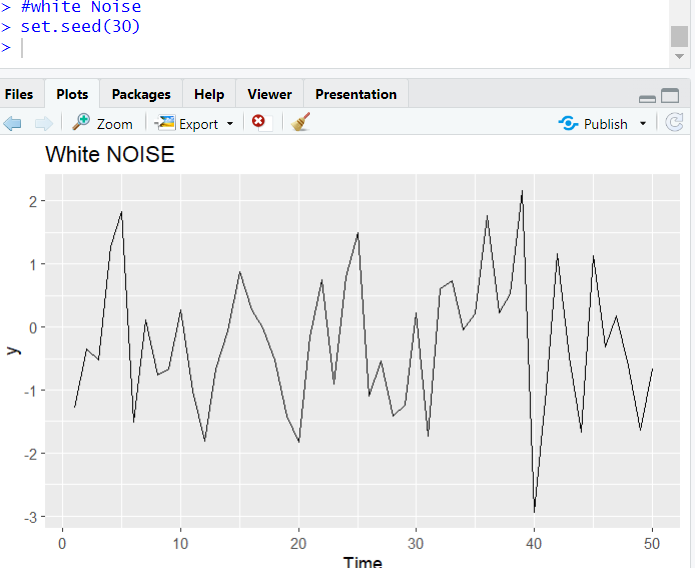
data("AirPassengers")

plot.ts(AirPassengers)



**White Noise:**

Time series that shows no autocorrelation.



Chapter 2 all exercise is solved below:

library(fpp2)

1. Use the help menu to explore what the series gold, woolyrnq and gas represent. These are available in the forecast package.

# See the structures of datas

str(gold)

str(woolyrnq)

str(gas)

# a. Use autoplot to plot each of these in separate plots.

autoplot(gold)

autoplot(woolyrnq)

autoplot(gas)

writeLines("")

# b. What is the frequency of each commodity series? Hint: apply the frequency() function.

print("Frequency")

print("gold")

frequency(gold)

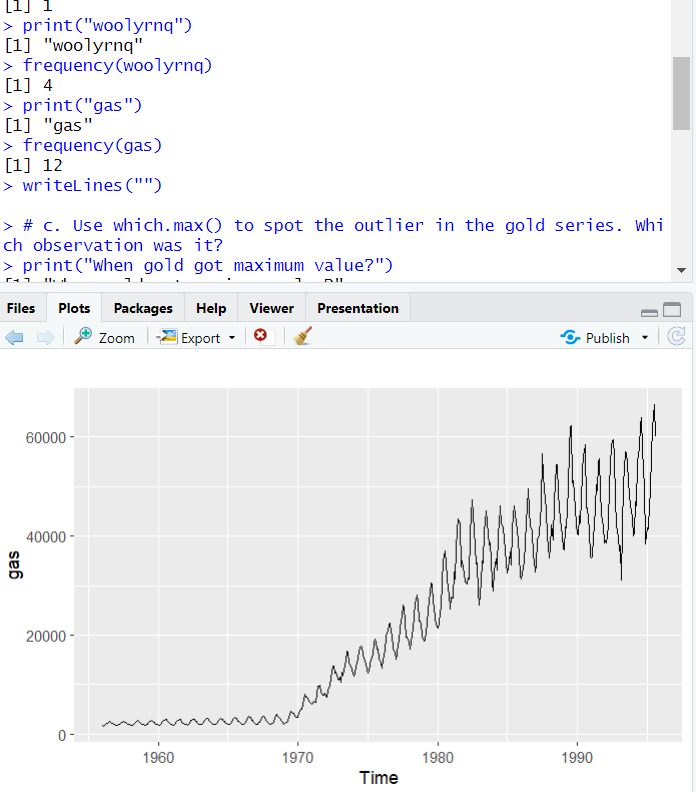
print("woolyrnq")

frequency(woolyrnq)

print("gas")

frequency(gas)

writeLines("")



For more to get please check my github account. All solve as well as others PMASDS, JU projects are uploaded there.

Github Link : <https://github.com/shrveel/PMASDS_JU_time_series_analysis/>

It is being requested that, check all the R files uploaded on the link according to the chapters.

Thank You

(please turn over)

**CHAPTER 3 (Trend Estimation)**

R for time series and decomposition

#time series decomposition

data("AirPassengers")

print(AirPassengers)

class(AirPassengers)

start(AirPassengers)

end(AirPassengers)

frequency(AirPassengers)

#checking missing values

sum(is.na(AirPassengers))

data = ts(AirPassengers, frequency = 12)

data

plot(data)

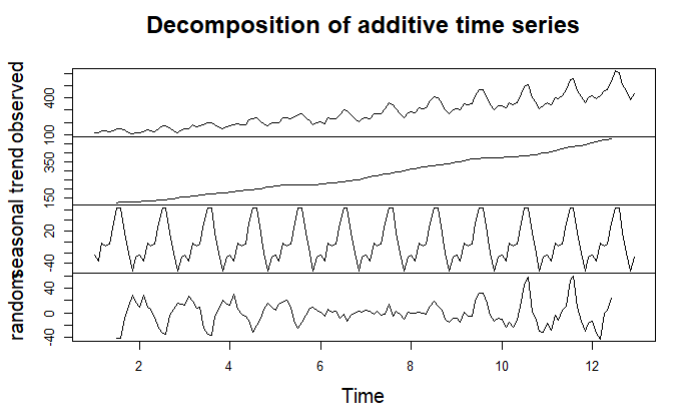
#decomposition

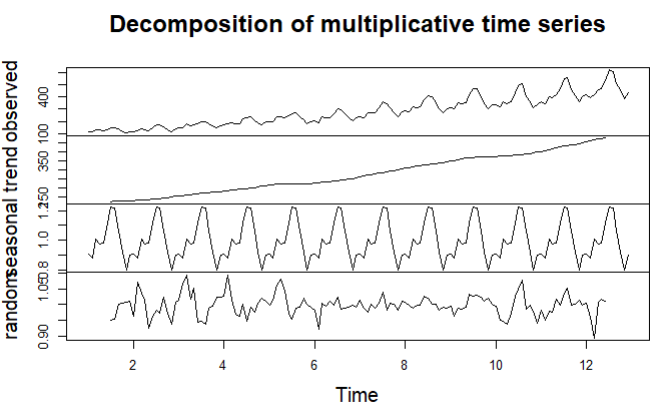
decom = decompose(data)

plot(decom)

decom = decompose(data, "multiplicative")

plot(decom)





**Types of Trend:**

* Deterministic: which are consistently increase and decrease at time series.
* Stocastic: which are not consistently increase or decrease at time series.

**Identifying a trend:**

• you can plot time series data to see if a trend is obvious or not

• create line plots of your data and inspect the plots for obvious trends

• add linear and nonlinear trend lines to your plots and see if a trend is

obvious

**Removing a trend:**

▪ a time series with a trend is called non-stationary

▪ an identified trend can be modelled

▪ once modelled, it can be removed from the time series dataset, this is called

detrending the time series

▪ if a dataset does not have a trend or we successfully remove the trend, the

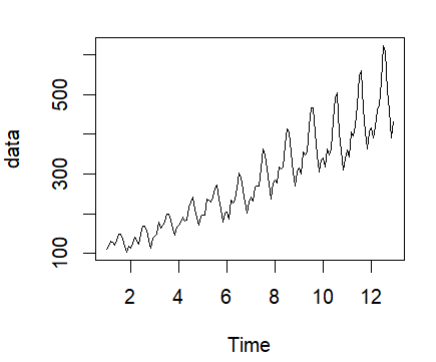
dataset is said to be trend stationary.

**Elimination of Trend/ Detrend by Differencing method:**

# Detrend by Diffencing

data

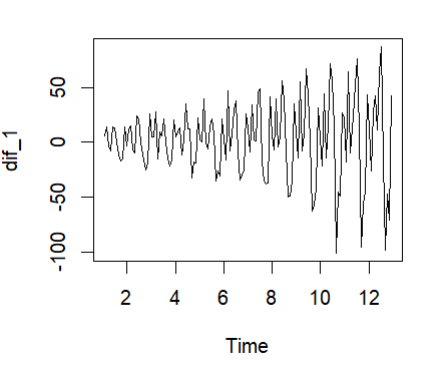
plot(data)



dif\_1 = diff(data)

dif\_1

plot(dif\_1)



Detrend by Model Fitting:

# Detrend by model fitting

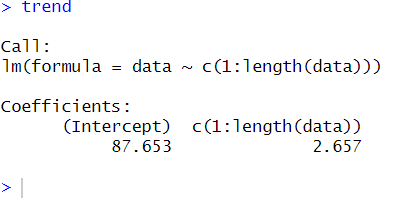
library(lmtest)

data = ts(AirPassengers, frequency = 12)

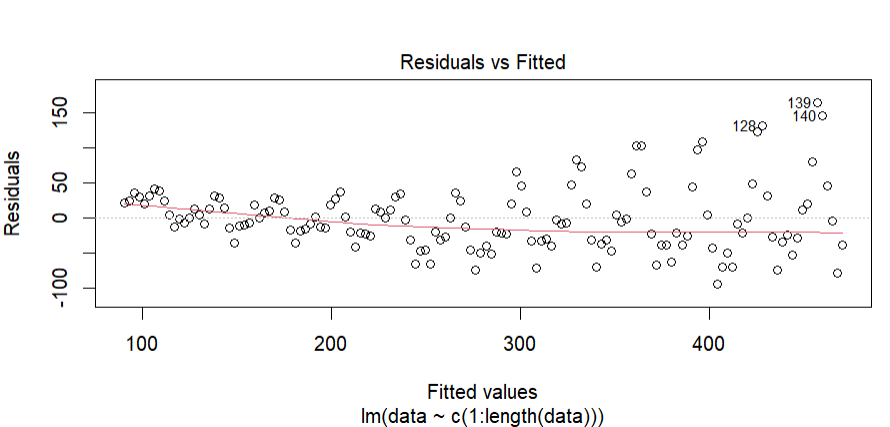
plot(data)

trend = lm(data~c(1:length(data)))

trend

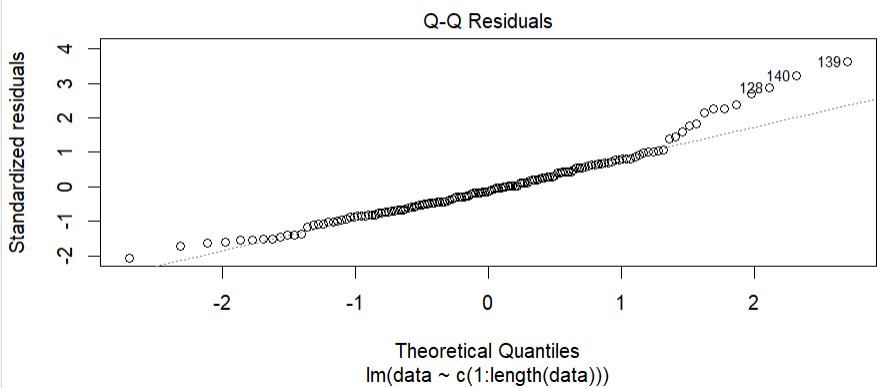


plot(trend)



detrend = residuals(trend)

plot.ts(detrend)



**CHAPTER 4 (Seasonality Estimation & Elimination)**

**Seasonality:**

* Seasonal variations are fluctuations that coincide with certain seasons and are repeated year after year.
* Seasonal fluctuations in a time series can be contrasted with cyclical patterns and the latter occur when the data exhibits rise and falls that are not of a fixed period.

**Method for finding seasonality**

* Method of simple average
* Ratio to trend method
* Ration-to-moving average method
* Method of link relatives

**Method of simple average**

* Method of simple average is easy and simple to execute.
* This model is based on additive model of the time series.
* Data do not contain any trend and cyclic components.

**Ratio-to-trend method:**

* Use the ration to trend method to determine the quarterly indices for the data.
* Yearly increment in the trend value = 12
* Quarterly increment in the trend value = 12/4 = 3
* This method is an improvement over the simple average method.
* This method is certainly more logical procedure for measuring seasonal variations.
* The main defect is, is there are cyclical swings in the series, the trend whether a straight line or a curve can never follow the actual data.

**Ratio to moving average:**

* Determine the four-quarter moving total.
* Locate the specific seasonal for the corresponding quarters.
* Calculate adjusted seasonal index using arithmetic mean and correction factor.
* It is the most satisfactory, edible and widely used method.

**How seasonal values are interpreted:**

* The original time series data is divided by adjusted seasonal index, resulting in de-seasonalized data.
* Reason for Depersonalizing the time series is to remove the seasonal fluctuations so that the trend and cycle can be studied.
* Forecast the quarter data with linear trend equation.

**Link Relative Method:**

* It is calculated by dividing the figure of each season by the figure of immediately preceeding season and multiplying it by 100.

R code for seasonality & De-seasonality:

data("AirPassengers")

print(AirPassengers)

plot.ts(AirPassengers)

decompose= decompose(AirPassengers)

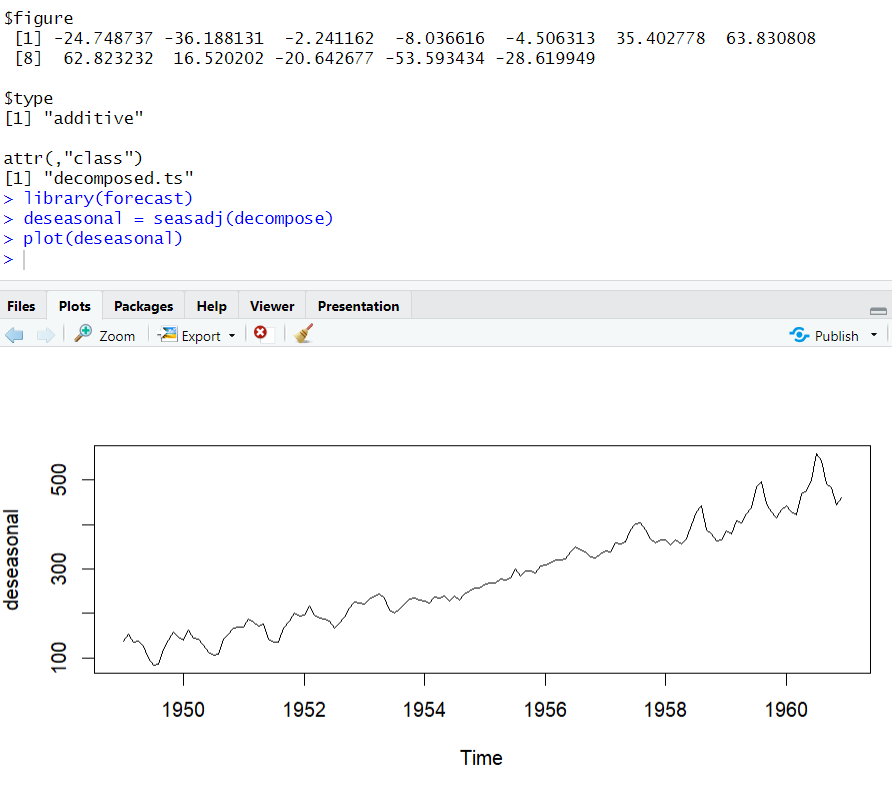
plot(decompose)

print(decompose)

library(forecast)

deseasonal = seasadj(decompose)

plot(deseasonal)



**Chapter 5 (Stationary Vs Non-stationary)**

**Stationary:**

A time series is called stationary time series if means and variance are constant over time.

**Types:**

* Weakly Stationary
* Strongly Stationary

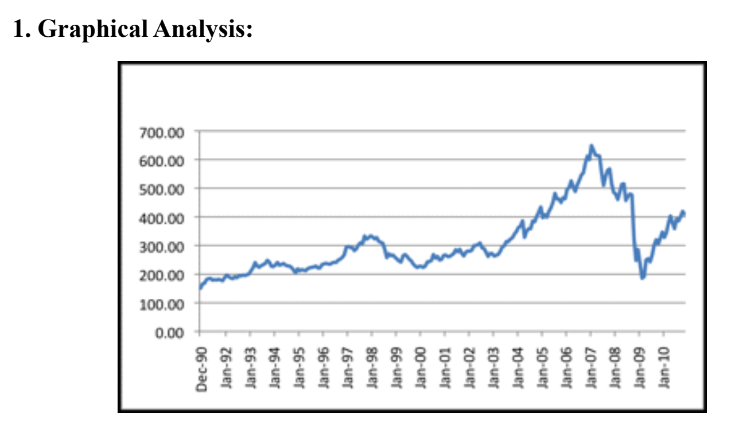
**Non-stationary time series:**

A time series is called non-stationary time series if mean and variance are not constant over time.

**For checking whether a data is stationary or not by those methods:**

* Graphical Analysis
* Autocorrelation function (ACF) and correlogram.
* Unit root test

**Graphical Analysis:**



If this ACF value close to zero it’s indicating that stationary otherwise, it’s a non-stationary time series data.

We will see ACF & correlogram when checking stationary or not in R code.

R code for Autocorrelation Function

library(tseries)

library(forecast)

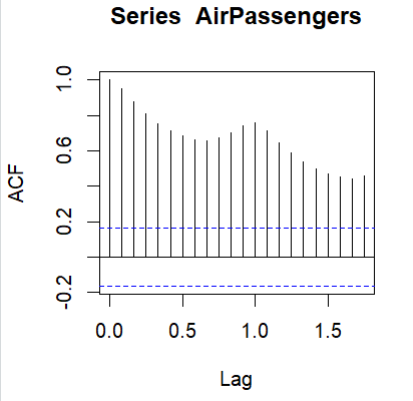
data("AirPassengers")

plot(AirPassengers)

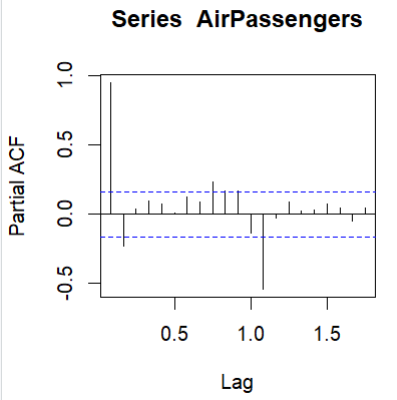
# acf and pacf plot

plot.ts(AirPassengers)

acf(AirPassengers)



pacf(AirPassengers) #partial ACF



**White Noise in Time Series Analysis:**

If time has no autocorrelation, then time series exist white noise. It is an important concept in time series analysis and forecasting.

**Important for two main reasons:**

* Predictability: If your time series is white noise, then, by definition, it is random.
* Model Diagnostics: the series of errors from a time series forecast model should ideally be white noise

If a time series data is white noise that means that each observation is uncorrelated with other observation that’s why there is no specific pattern or we can say that the data is random. If data is random, we cannot make any prediction for future value.

**Checking white noise method:**

* Create a line plot
* Calculate summary statistics
* Create an autocorrelation plot

**Checking White Noise process in R:**

#generate some random data

set.seed(1)

size = 200

#create an empty array to hold ts data

x = numeric(size)

#create the first value

x[1] = rnorm(n = 1, mean = 0, sd = 1)

#create loop to have enough variables

for (t in 2:size) {

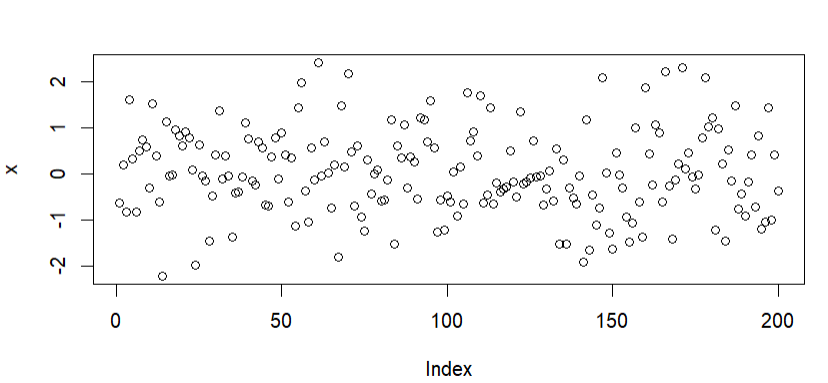
w.error = rnorm(n=1, mean = 0, sd=1)

x[t] = w.error

}

#white noise has been created now we plot

plot(x)



* White noise time series is defined by a zero mean, constant variance, and zero correlation.
* If our time series is white noise, it cannot be predicted.
* For testing white noise, Q-statistics is applied, which is developed by BOX-Pierce.

**THANK YOU**