INTRODUCTION TO TIME SERIES ANALYSIS -II

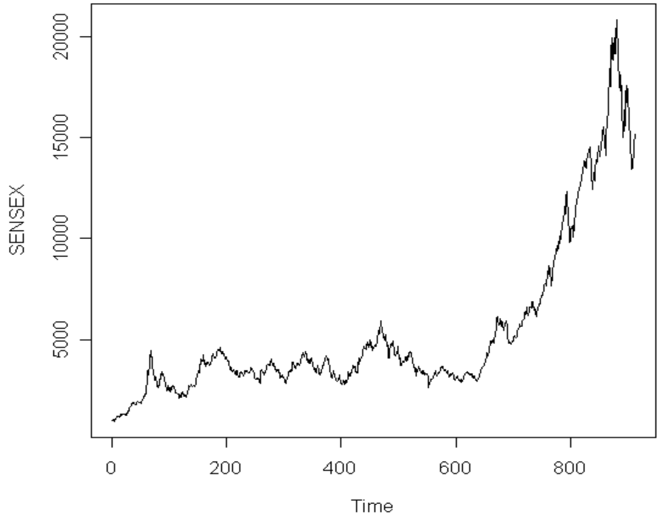
DATA SCIENCE INSTITUTE

Recap: What is Time Series?

- Time Series is a sequence of values measured over time.

- Time series can be

- ❑ Annual- (GDP, Company Turnover)
- ❑ Quarterly-(GDP, Company Turnover)
- ❑ Monthly-(Inflation Rates)
- ❑ Daily-(Stock prices, Gold Prices)



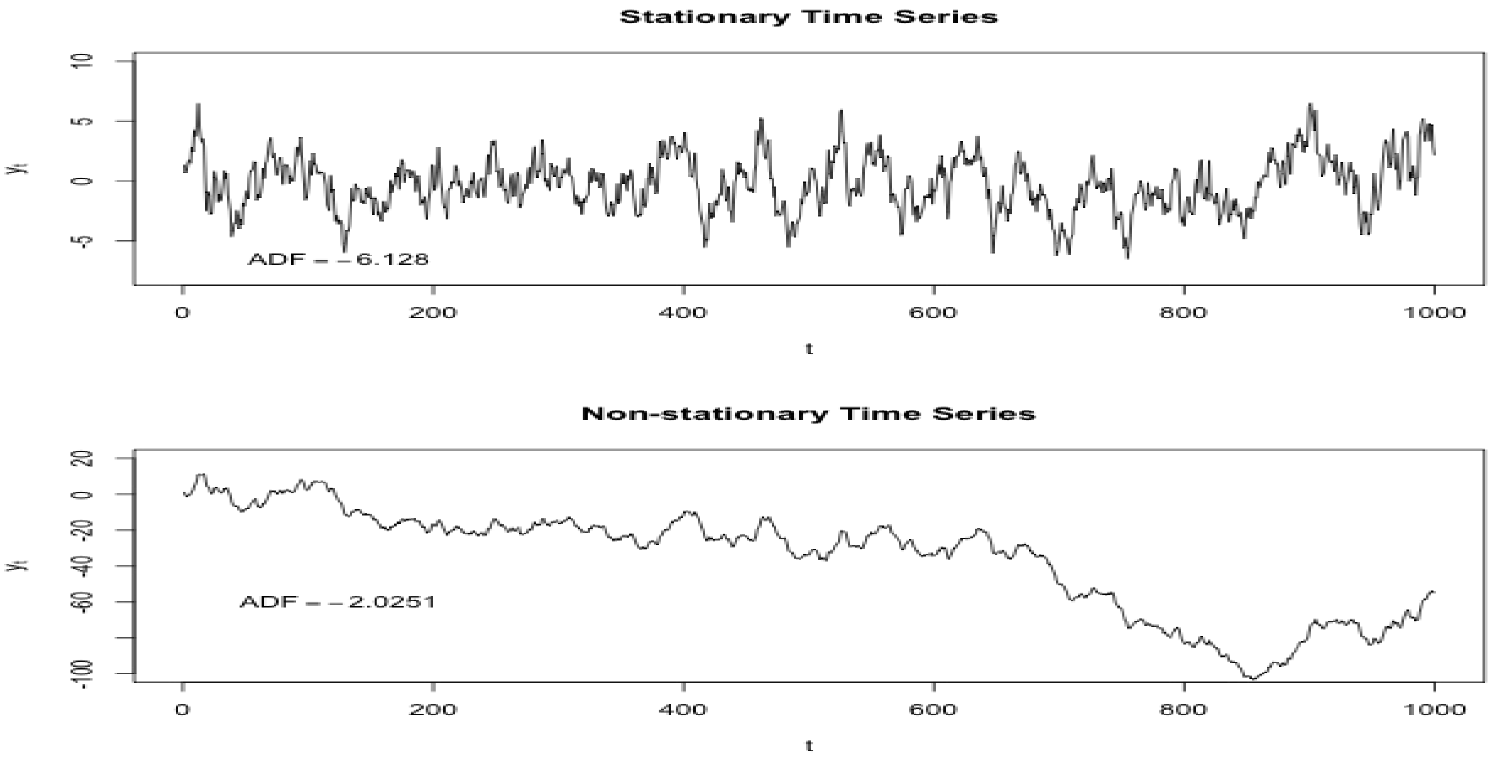
DATA SCIENCE INSTITUTE

Recap: Time Series Data Analysis

- **Analyze trend and seasonality** present in the data.
- **Decompose** time series into its components.
- Forecast future values of the time series.

DATA SCIENCE INSTITUTE

Recap: Stationary Time Series



- **Time series process is called stationary if the statistical properties of the process remain unchanged over time.**

Recap:How to Make a Non Stationary Time Series Stationary?

- There are 2 methods to make a non stationary time series stationary.
  - Differencing
  - De-trending

- Differencing: A non stationary time series can be made stationary by differencing. Consider the following non- stationary process

$$Y_t=Y_{t-1}+U_t \quad t=1,2,3.....$$

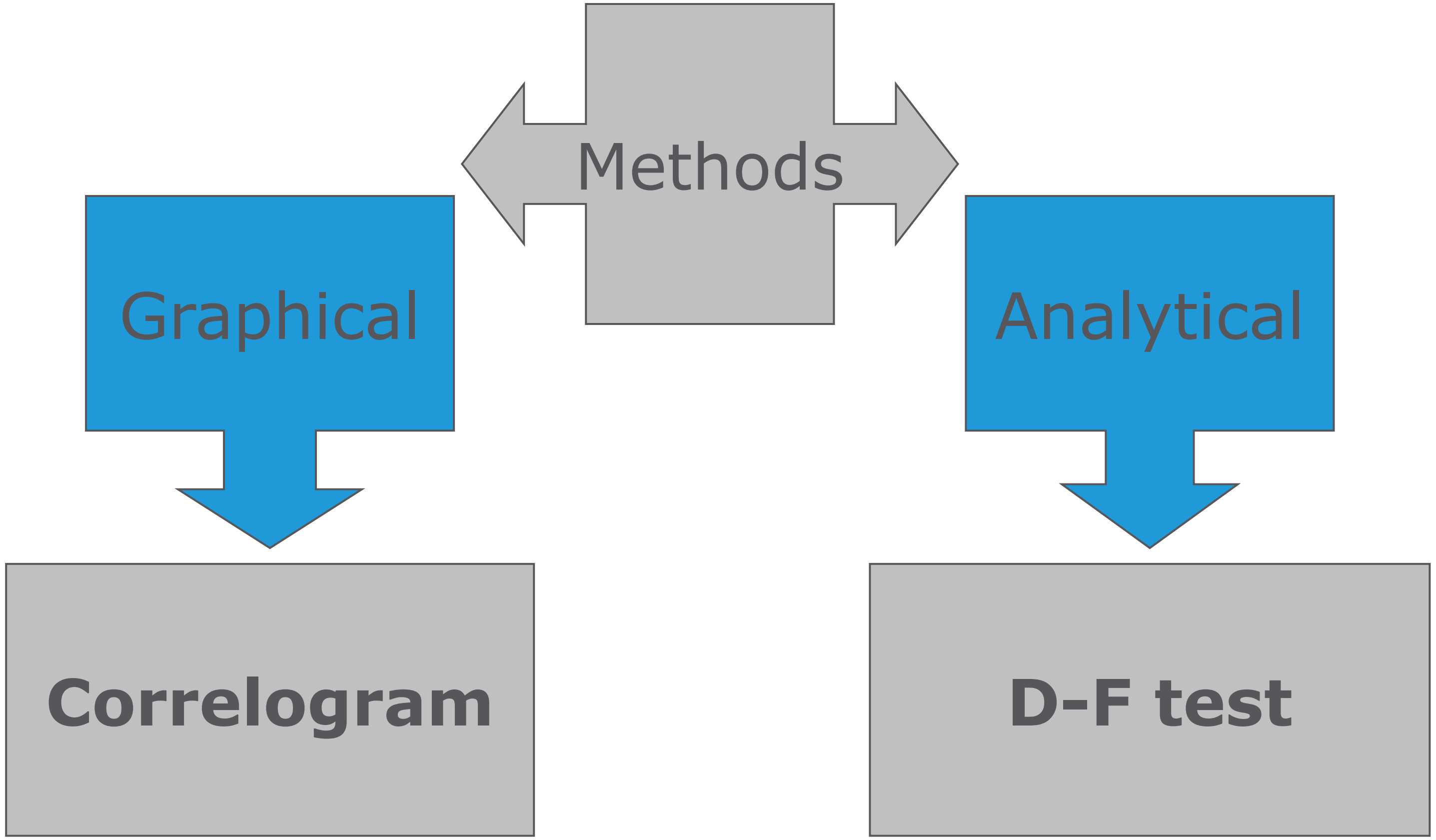
We assume that U_t is a random series with constant mean μ and constant variance σ^2 also it is serially uncorrelated i.e (U_t is stationary).

- Hence $Y_t-Y_{t-1}=\Delta Y_t= U_t$, which is a stationary time series.

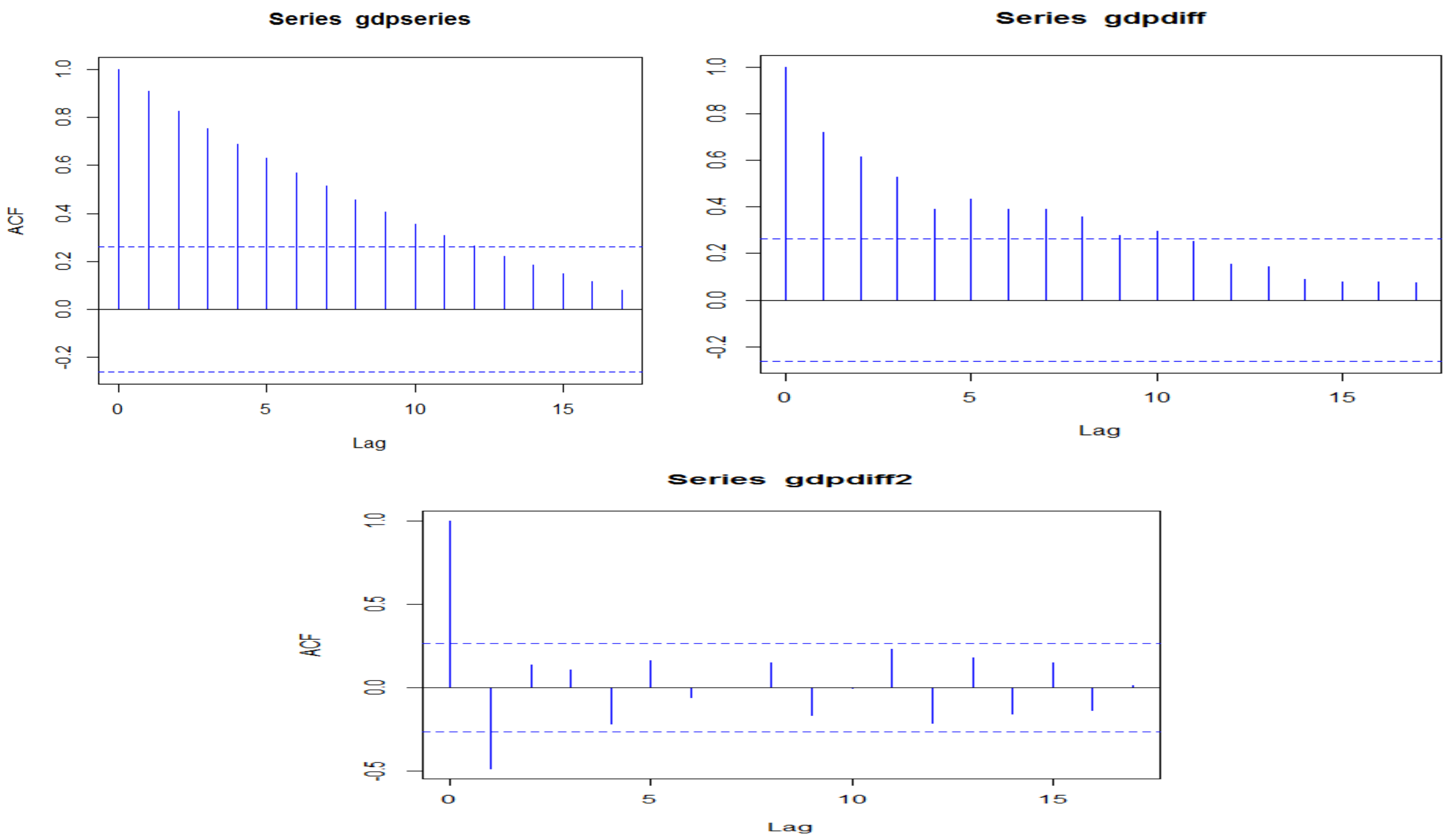
- Differencing can be well applied in case of stochastic time series.

DATA SCIENCE INSTITUTE

Recap: Identifying Stationary Time Series



Recap: Time Series Analysis in R
Correlograms



Stationarity is achieved with second order difference

DATA SCIENCE INSTITUTE

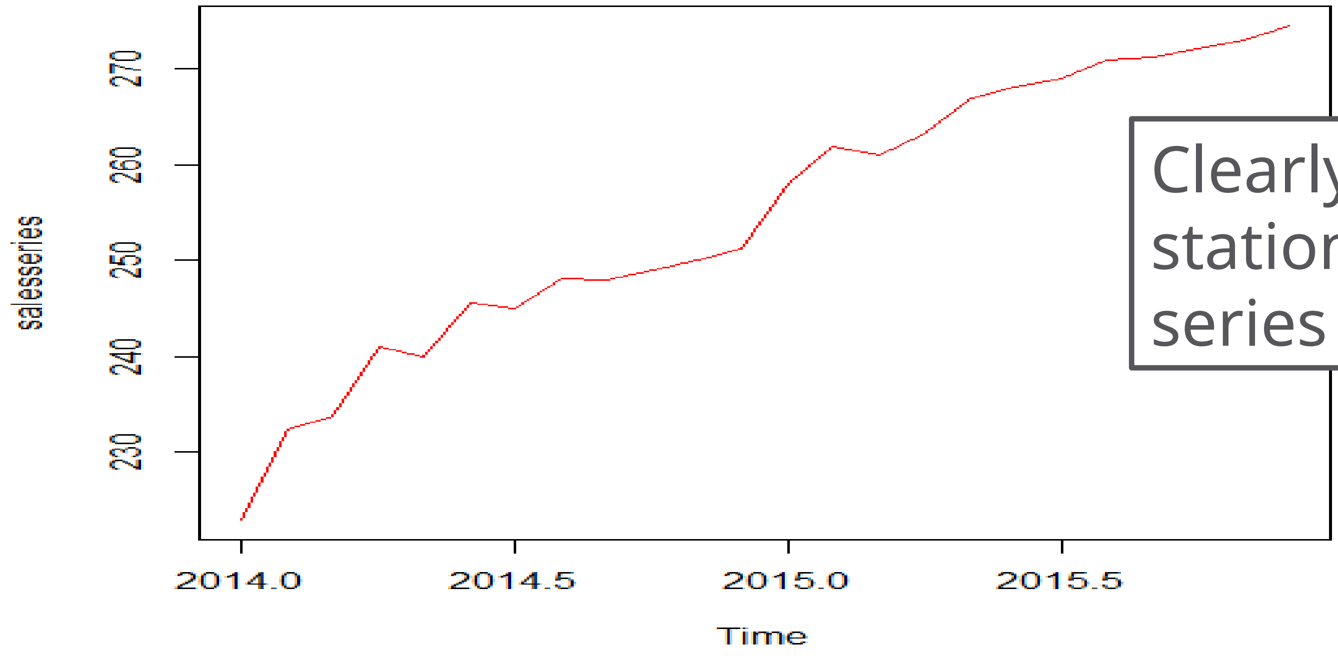
Monthly Sales Data
2014-2015

Year	Month	Sales
2014	Jan	223
2014	Feb	232.5
2014	Mar	233.8
2014	Apr	241.1
2014	May	240
2014	Jun	245.6
2014	Jul	245
2014	Aug	248.1
2014	Sep	248
2014	Oct	249.1
2014	Nov	250
2014	Dec	251.2
2015	Jan	258
2015	Feb	261.9
2015	Mar	261
2015	Apr	263.3
2015	May	267
2015	Jun	268.1
2015	Jul	269
2015	Aug	271
2015	Sep	271.2
2015	Oct	272.1
2015	Nov	273
2015	Dec	274.5

Sales figures (in Rs. Millions)
for 2 years.

Time Series Analysis in R
Plot Time Series…

```
salesdata<-read.csv(file.choose(),header=T)
#Create time series object and plot. Frequency=12 indicates monthly data
salesseries<-ts(salesdata$Sales,start=c(2014,1),end=c(2015,12),frequency=12)
plot(salesseries,col="red")
```



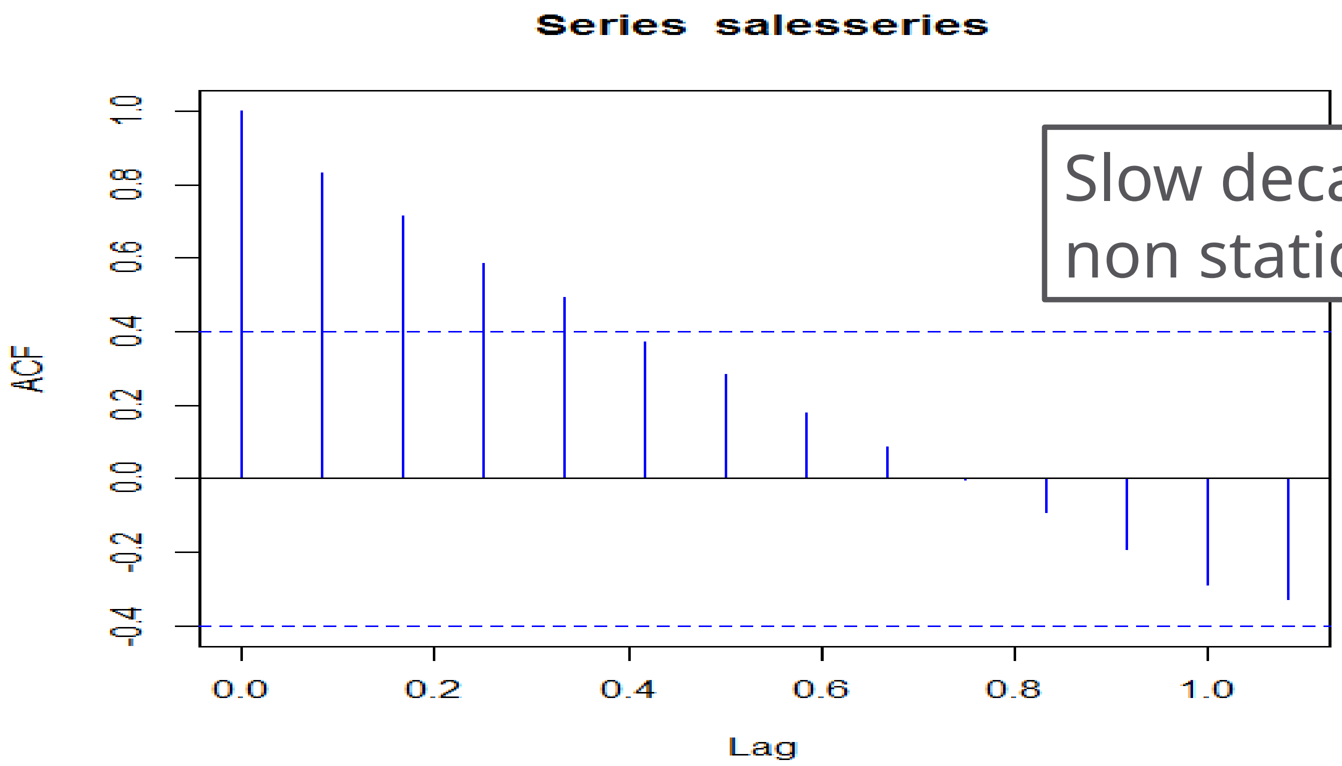
Clearly a non stationary time series

DATA SCIENCE INSTITUTE

Time Series Analysis in R
Correlogram

#Obtain correlogram

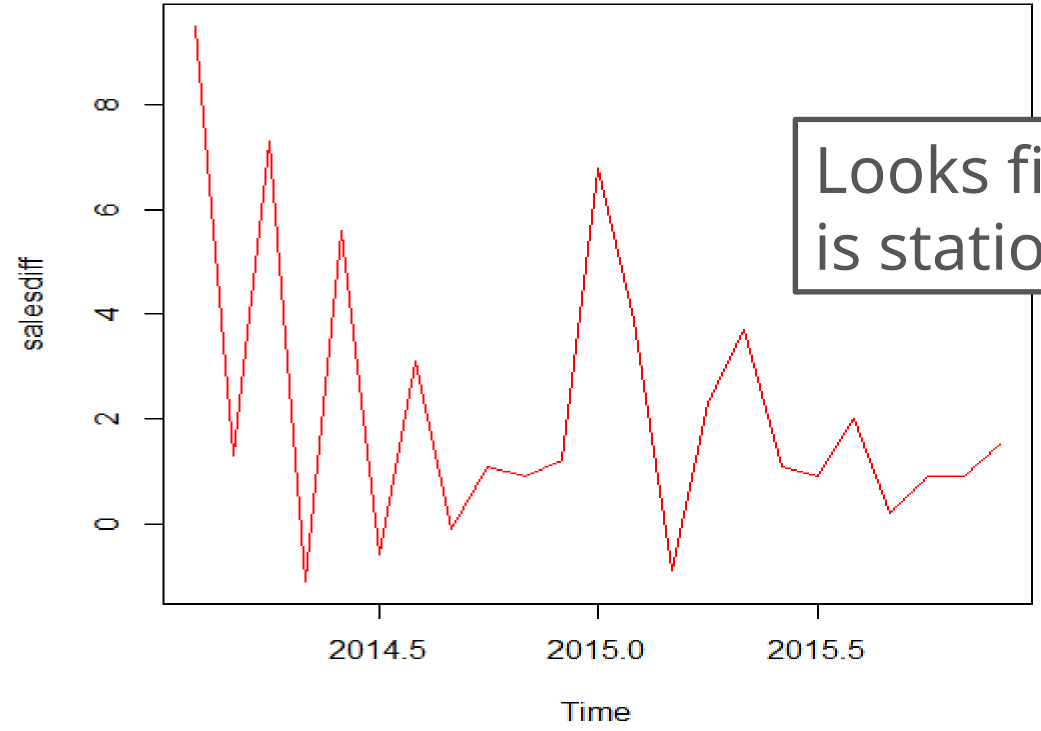
```
acf(salesseries,col="blue")
```



Slow decay confirms non stationarity

Time Series Analysis in R
Plot Difference Time Series

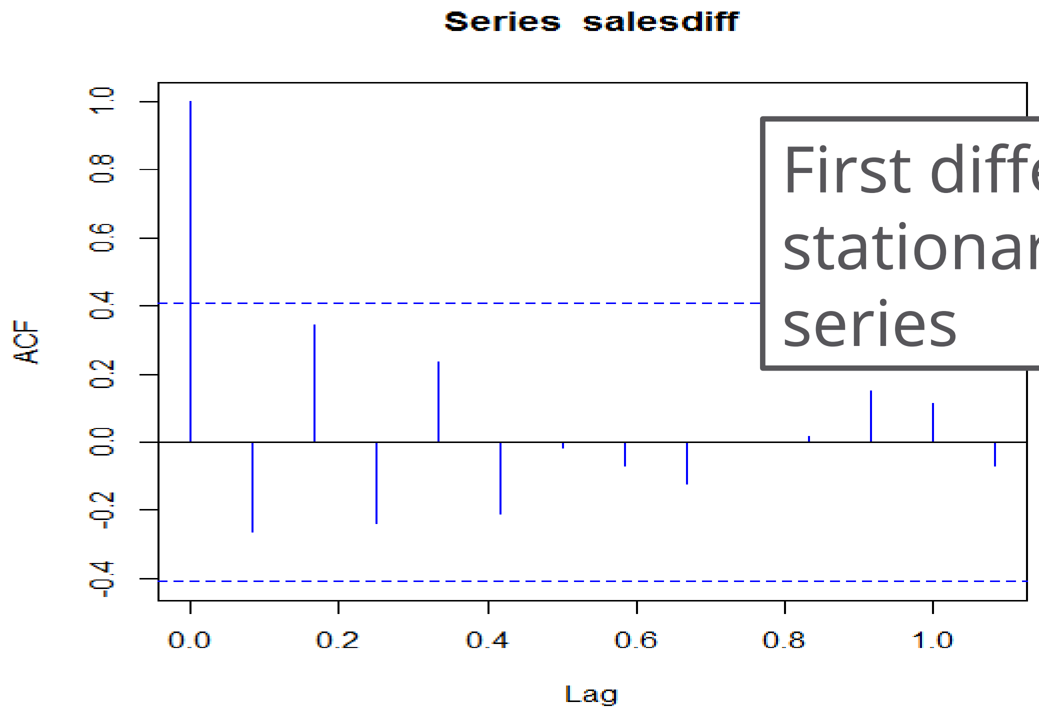
```
salesdiff<-diff(salesseries,differences=1)
plot(salesdiff,col="red")
```



Looks first difference is stationary

Time Series Analysis in R
Correlogram for Difference Time Series

acf(salesdiff,col="blue")



First difference is stationary time series

How Many Times Should Time Series Be Differenced to Make Stationary?

```
library(forecast)
ndiffs(salesseries)
> ndiffs(salesseries)
[1] 1
```

```
library(urca)
summary(ur.df(salesseries,lag=0))

Value of test-statistic is: 3.6856
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.66 -1.95  -1.6
```

H0: Time Series in Nonstationary

Do not reject H0

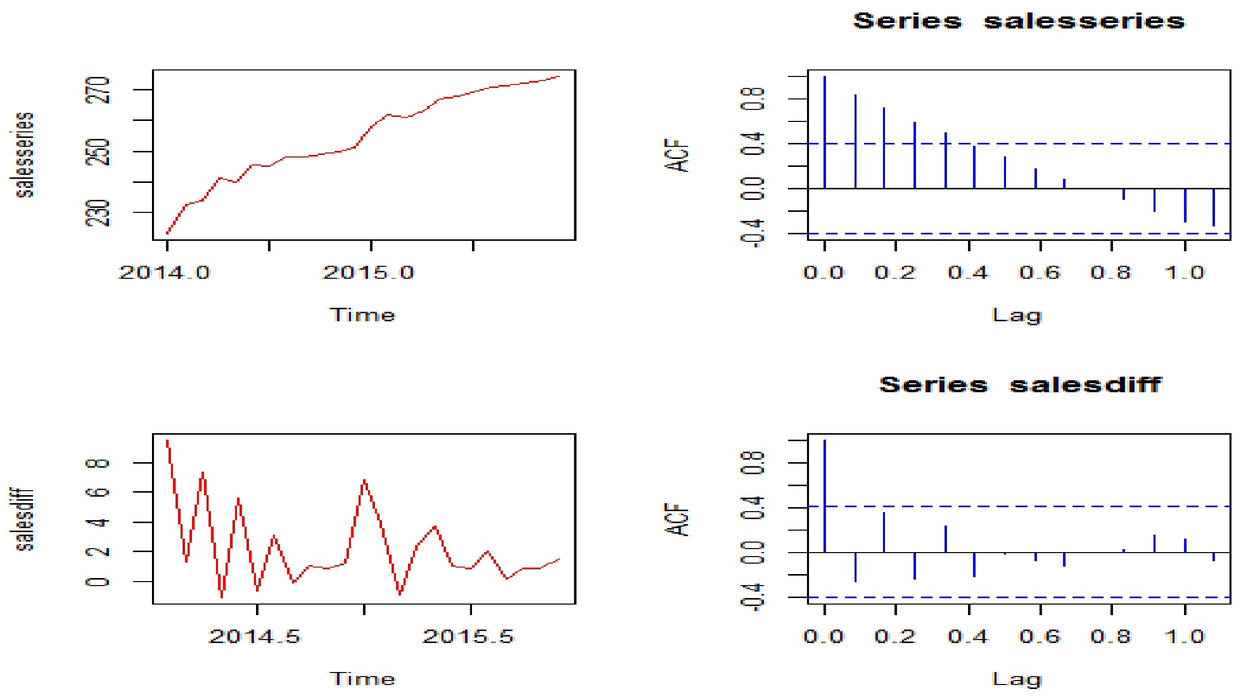
```
summary(ur.df(salesdiff,lag=0))

Value of test-statistic is: -4.6349
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.66 -1.95  -1.6
```

Reject H0 as value of test statistic is less than 5 pct critical value(-1.95)

DATA SCIENCE INSTITUTE

Showing Graphs in One Panel

```
par(mfrow=c(2,2))
plot(salesseries,col="red")
acf(salesseries,col="blue")
plot(salesdiff,col="red")
acf(salesdiff,col="blue")
```



DATA SCIENCE INSTITUTE

Components of Time Series? Trend, Seasonality and Cyclic Pattern

- **Trend** refers to the long-term increase or decrease in the time series.
- **Seasonality** in a time series refers to predictable and recurring trends and patterns over a period of time, normally a year. An example of a seasonal time series is retail data, which sees spikes in sales during holiday seasons like Christmas or Diwali.
- A **cyclic pattern** exists when data exhibits rises and falls that are not of fixed period. The duration of these fluctuations is usually of at least 2 years.

Time Series Decomposition

Time series can be considered as having three components:

a) Trend
b) Seasonal Component
c) Random Component

For example, if we assume an additive model, then we can write

$$Y_t = S_t + T_t + R_t$$

where Y_t is the data at period t, S_t is the seasonal component at period t, T_t is the trend component at period t and R_t is the remainder (or irregular or error) component at period t.

Time Series Decomposition

Alternatively, a multiplicative model would be written as

$$Y_t = S_t * T_t * R_t$$

The additive model is most appropriate if the magnitude of the seasonal fluctuations or the variation around the trend-cycle does not vary with the level of the time series.

DATA SCIENCE INSTITUTE

Time Series Decomposition

Step 1: Obtain moving averages covering one season. For example: In case on monthly data obtain average of 13 values (6 previous,6 post and given month value). This provides trend component of the time series.

Step 2: Eliminate trend component from original time series. Calculate Yt –Tt

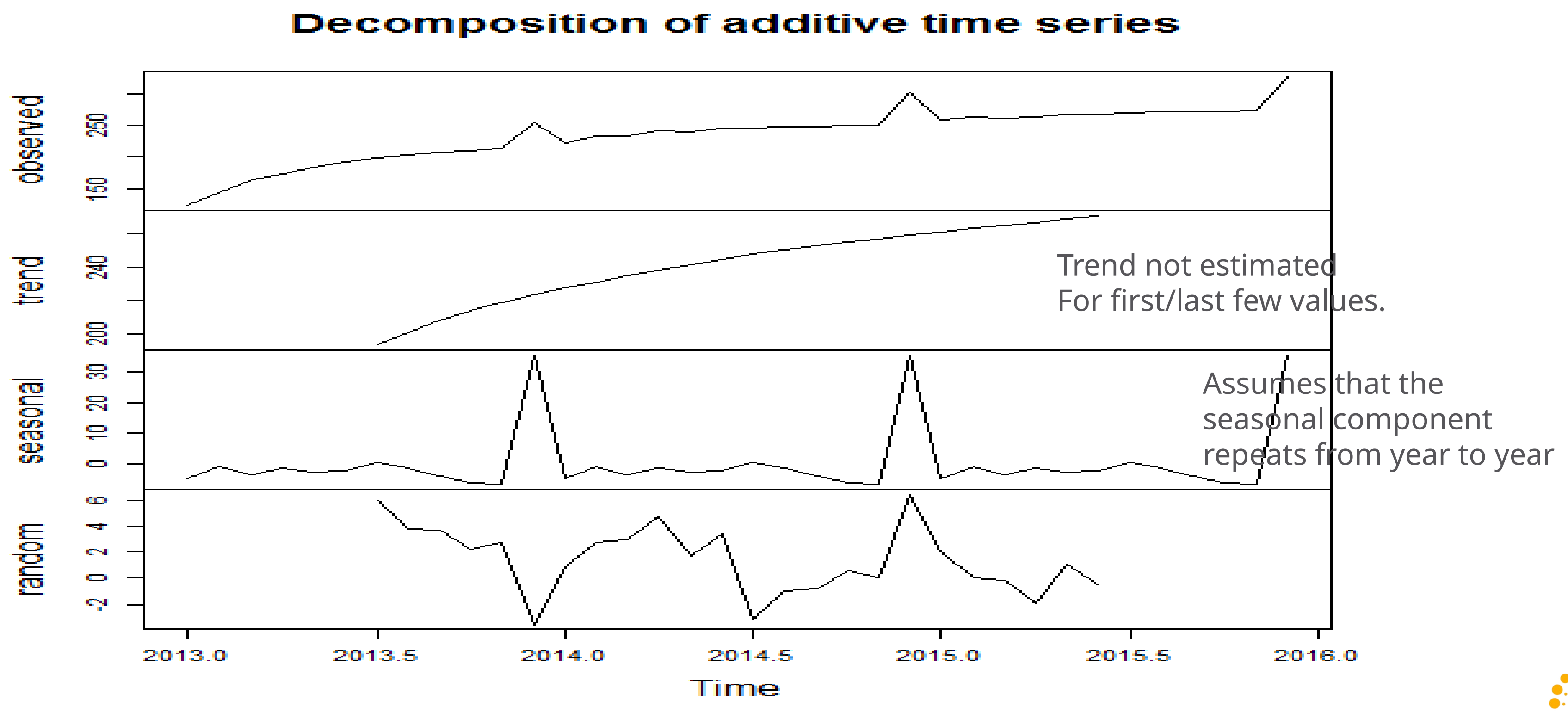
Step 3: To estimate the seasonal component for each month, simply average the detrended values for that month. For example, the seasonal index for March is the average of all the detrended March values in the data. These seasonal indexes are then adjusted to ensure that they add to zero.

The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components.

DATA SCIENCE INSTITUTE

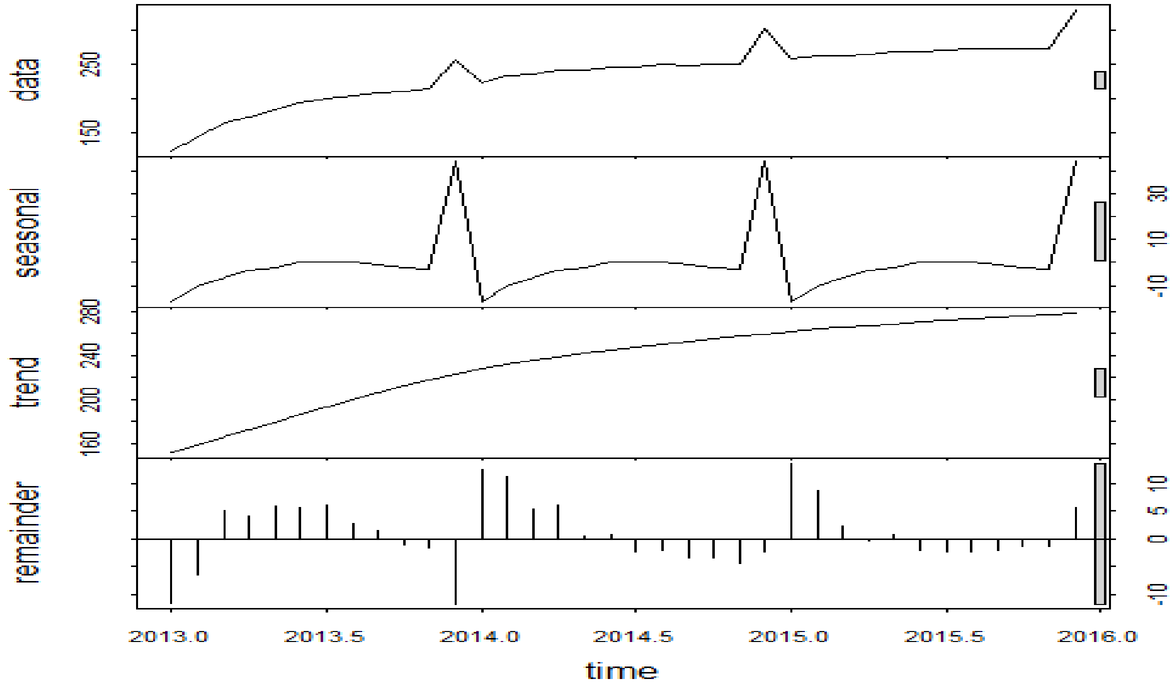
Time Series Decomposition in R

```
salesdata<-read.csv(file.choose(),header=T)
salesseries<-ts(salesdata$Sales,start=c(2013,1),end=c(2015,12),frequency=12)
decomp<-decompose(salesseries)  # can specify type="multiplicative"
plot(decomp)
```



Time Series Decomposition in R
Local Regression Method (LOESS)

```
salesdata<-read.csv(file.choose(),header=T)
salesseries<-ts(salesdata$Sales,start=c(2013,1),end=c(2015,12),frequency=12)
decomp<-stl(salesseries,s.window="periodic")
plot(decomp)
```



STL is an acronym for “Seasonal and Trend decomposition using Loess”,
while Loess is a method for estimating nonlinear relationships.

THANK YOU!!

DATA SCIENCE
INSTITUTE