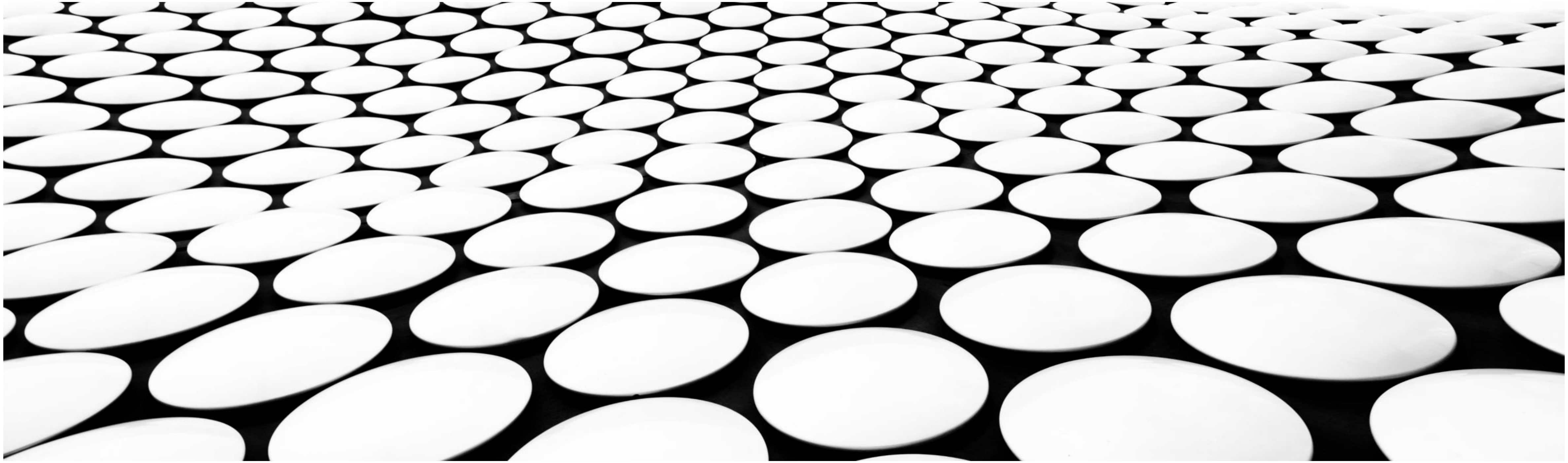
TIME SERIES

INTRODUCTION

Prof. Dr. Klaus Fabian Côco
Prof. Dr. Patrick Marques Ciarelli

UNIVERSIDADE FEDERAL
DO ESPÍRITO SANTO

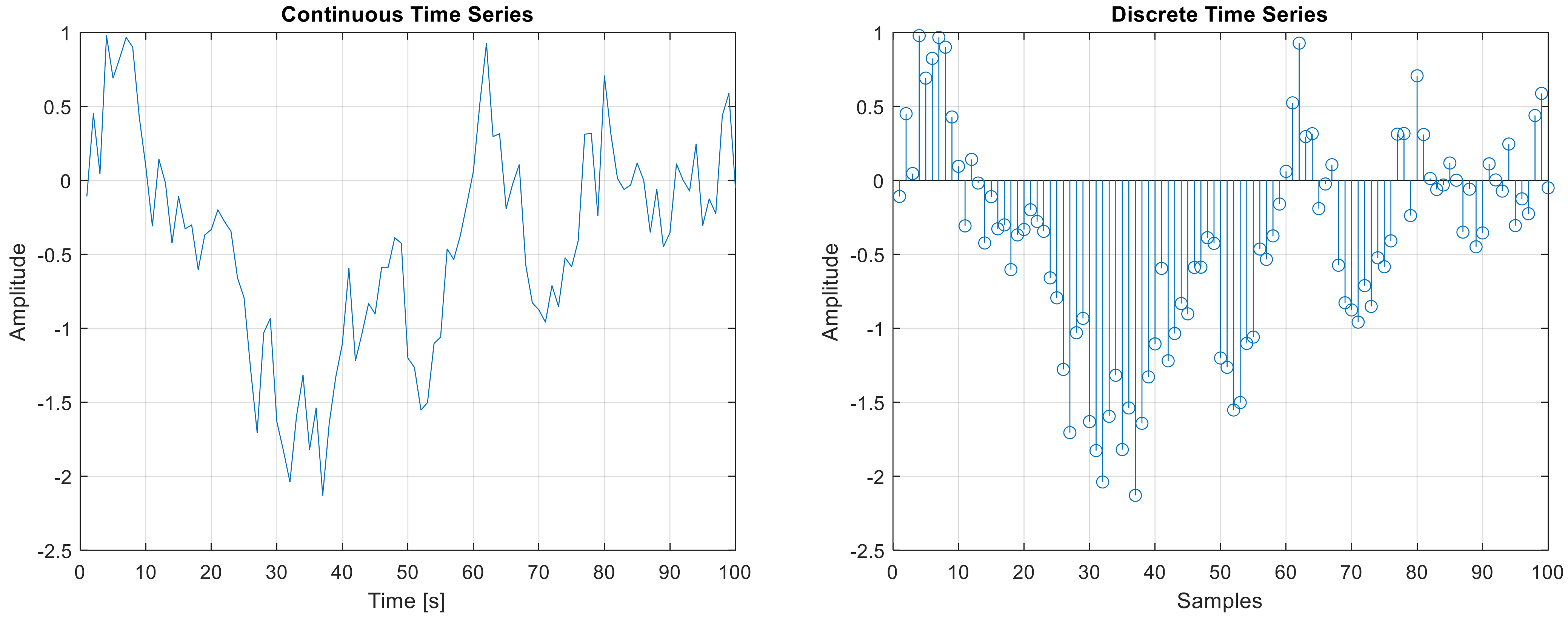


DEFINITIONS

INTRODUCTION TO TIME SERIES

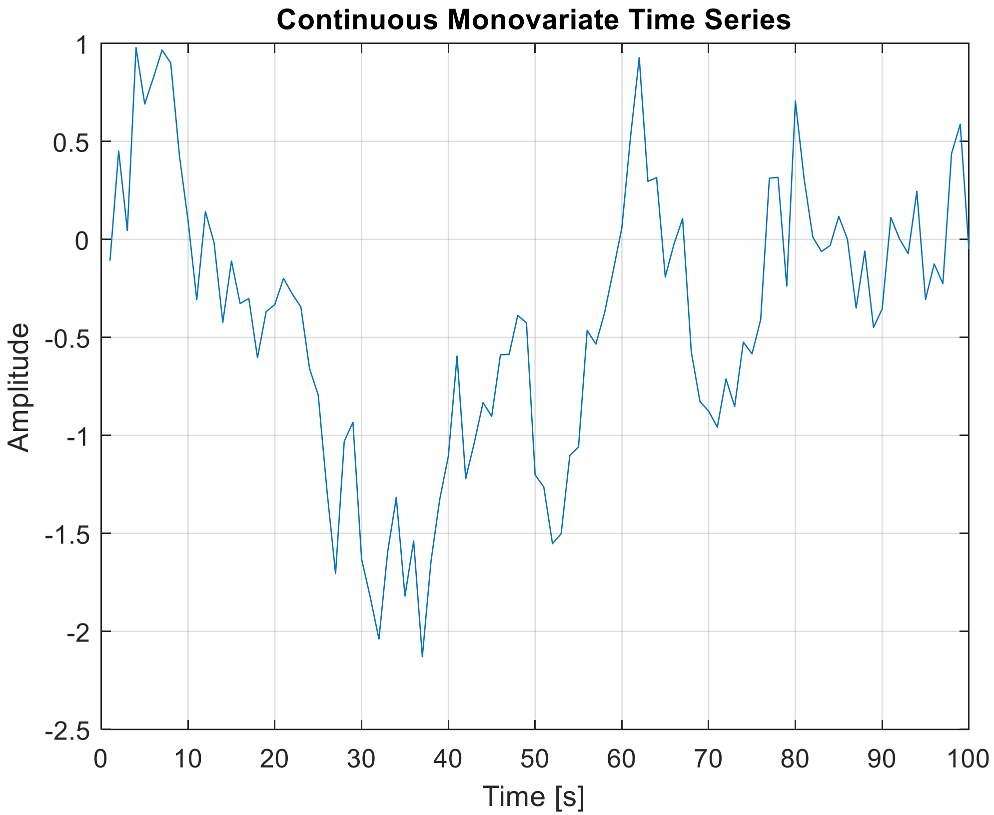
DEFINITIONS

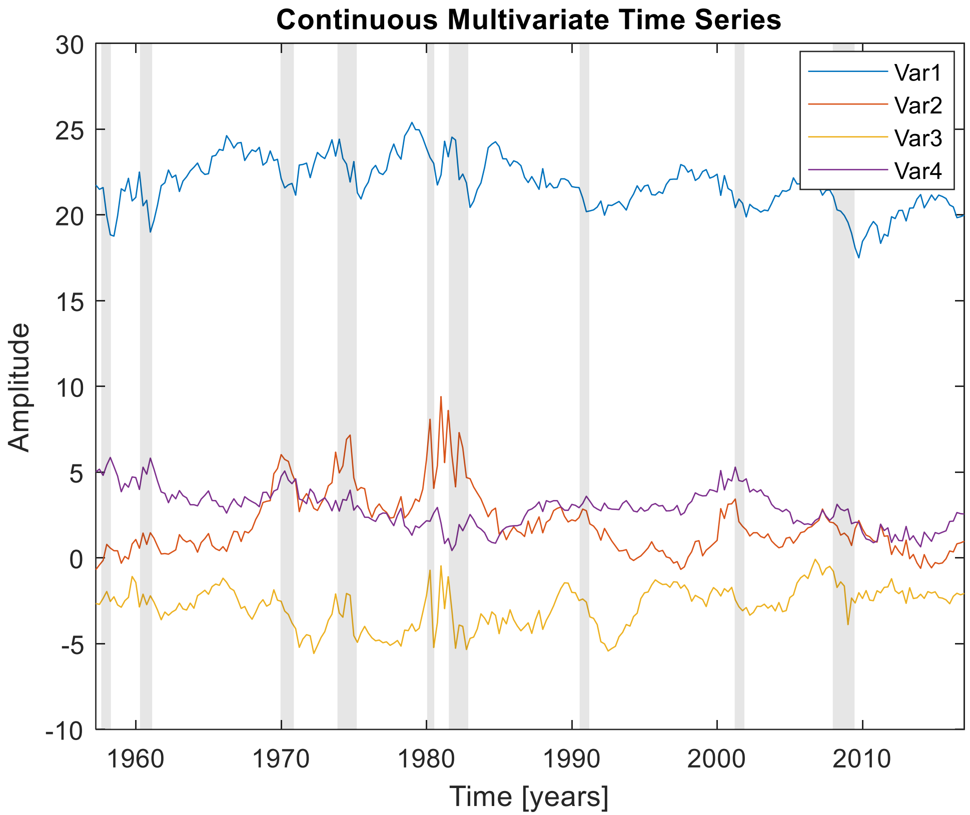
A Time Series is a set of observations taken sequentially in time.
An intrinsic feature of a time series is that, typically, adjacent observations are dependent.
Time Series Analysis is concerned with techniques for the analysis of this dependence.



INTRODUCTION TO TIME SERIES
DEFINITIONS

A time series is <u>monovariate</u> when it has only a single variable.
A time series is <u>multivariate</u> when it has more than one variable.





OBJECTIVES OF TIME SERIES ANALYSIS

INTRODUCTION TO TIME SERIES

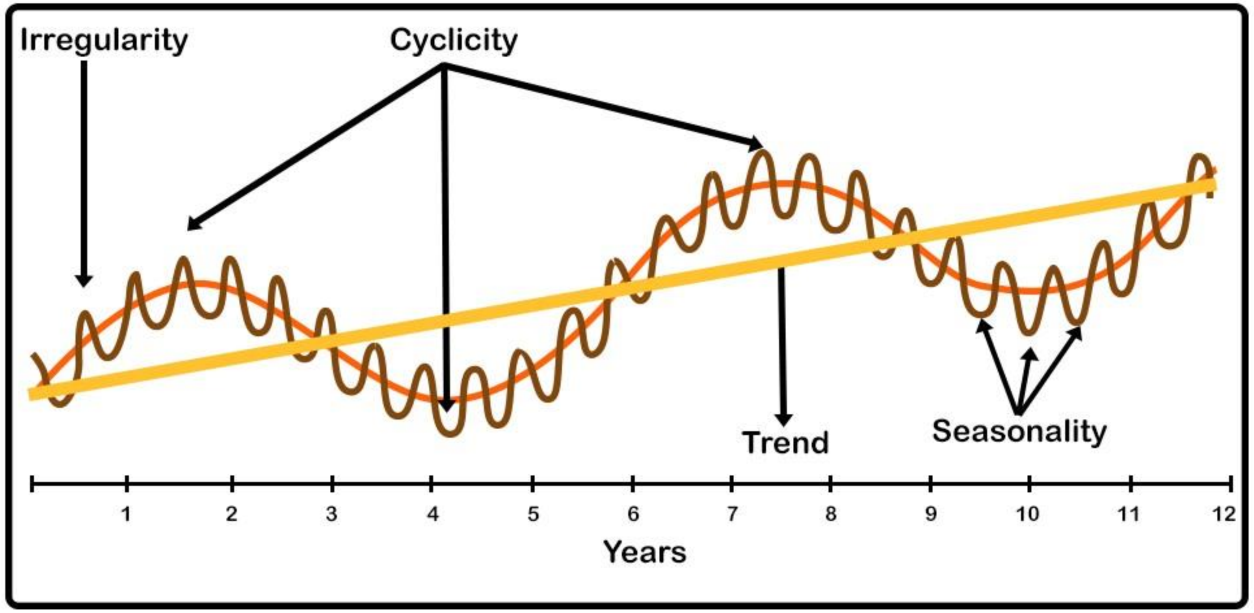
OBJECTIVES OF TIME SERIES ANALYSIS

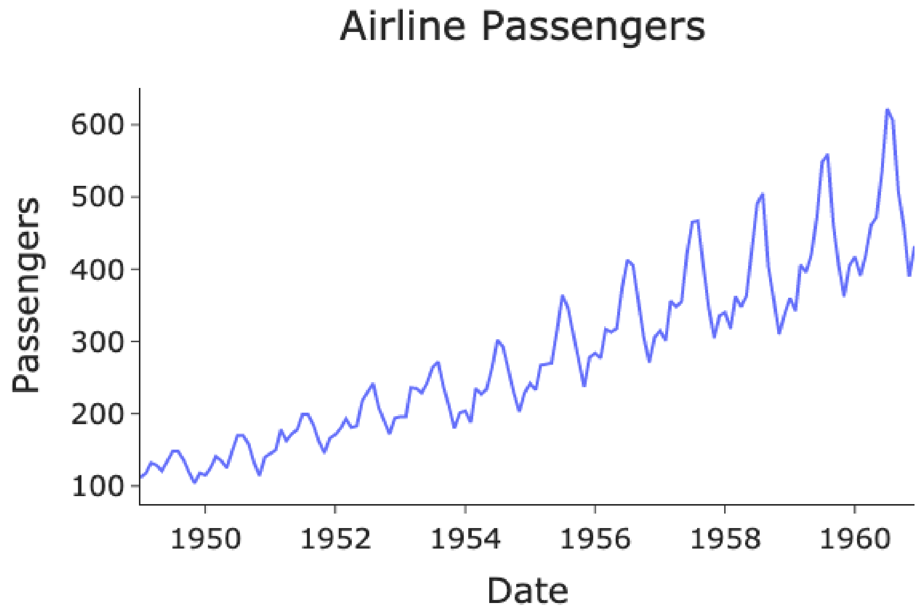
- There are four main objectives of time series analysis:
    - Description;
    - Explanation;
    - Prediction (Forecasting);
    - Control.

INTRODUCTION TO TIME SERIES

OBJECTIVES OF TIME SERIES ANALYSIS

- Description: A descriptive analysis is typically done by plotting the time series data. Plotting provides a ”high level” overview of the time series and its main components: the trend, seasonality, cycle, and random variations.
- Statistical measures, such as mean, standard deviation, and median, computed from a time series or from a time period of a time series can also be used to give a summary of the characteristics of a time series.

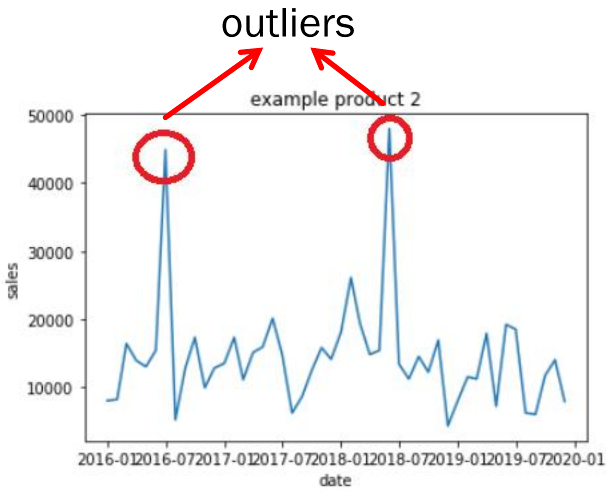


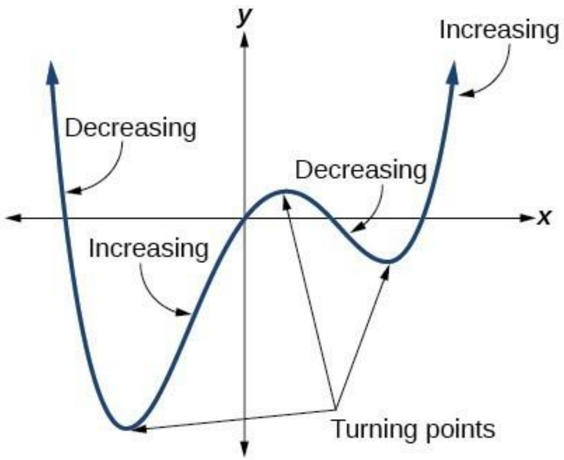


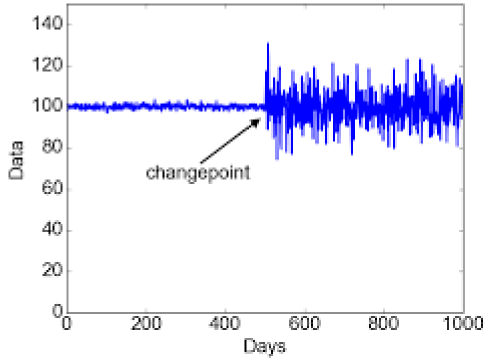
INTRODUCTION TO TIME SERIES

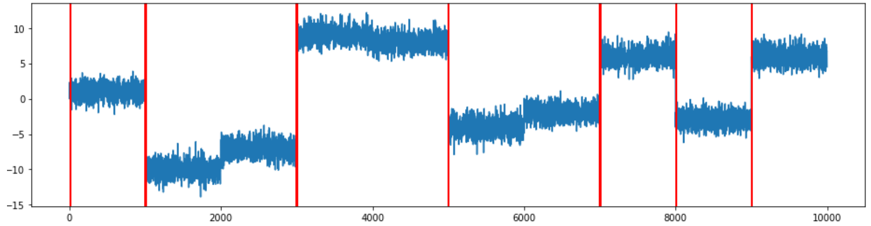
OBJECTIVES OF TIME SERIES ANALYSIS

- <u>Description</u>: Plotting the time series data can also reveal relevant points:
  - <u>Outliers</u>, which are points that appear inconsistent with the data pattern;
  - <u>Turning points</u> (TP), which are points in time when a series which had been increasing (decreasing) reverses and, for a time, decreases (increases); and
  - <u>Change points</u> (CP), which are points in time when a series undergoes a significant change in its statistical properties. Herein, a CP is distinguished from a TP, i.e., a TP is always a CP but a CP may not be necessarily a TP.







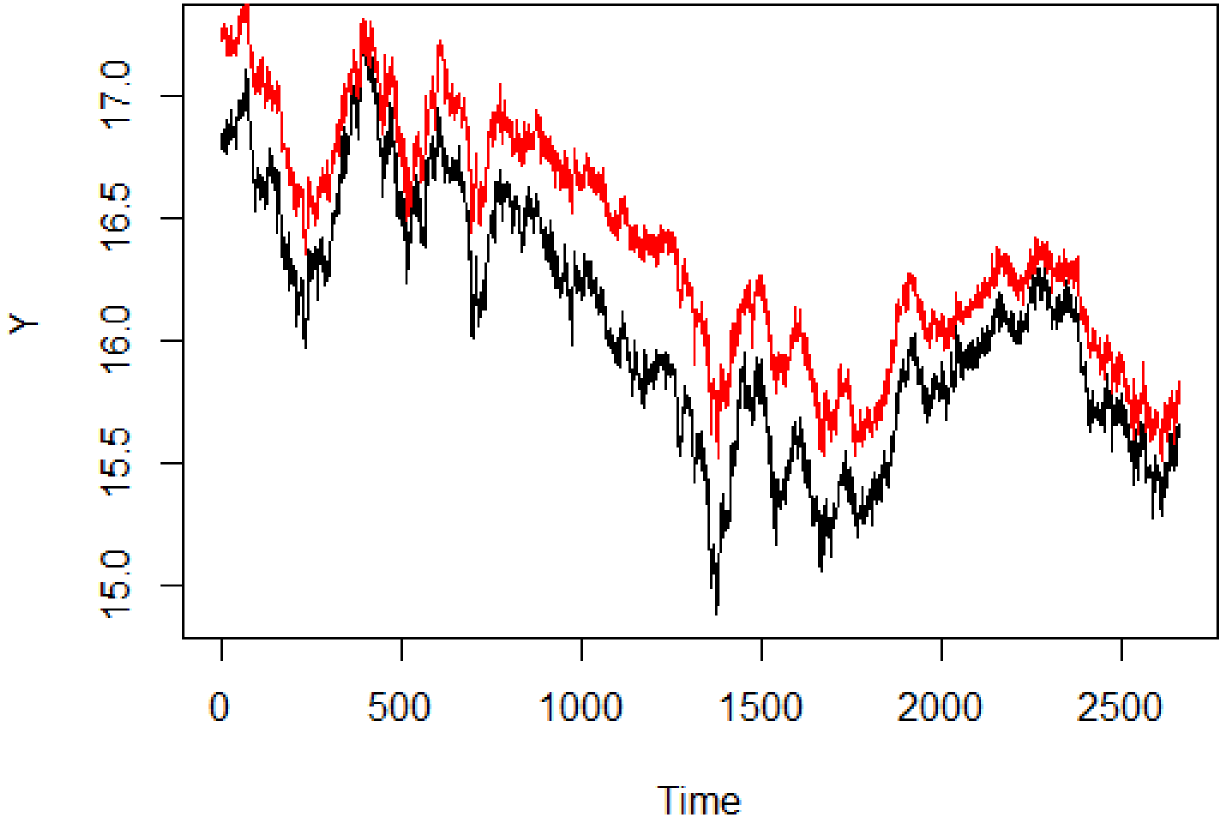


chance points

INTRODUCTION TO TIME SERIES

OBJECTIVES OF TIME SERIES ANALYSIS

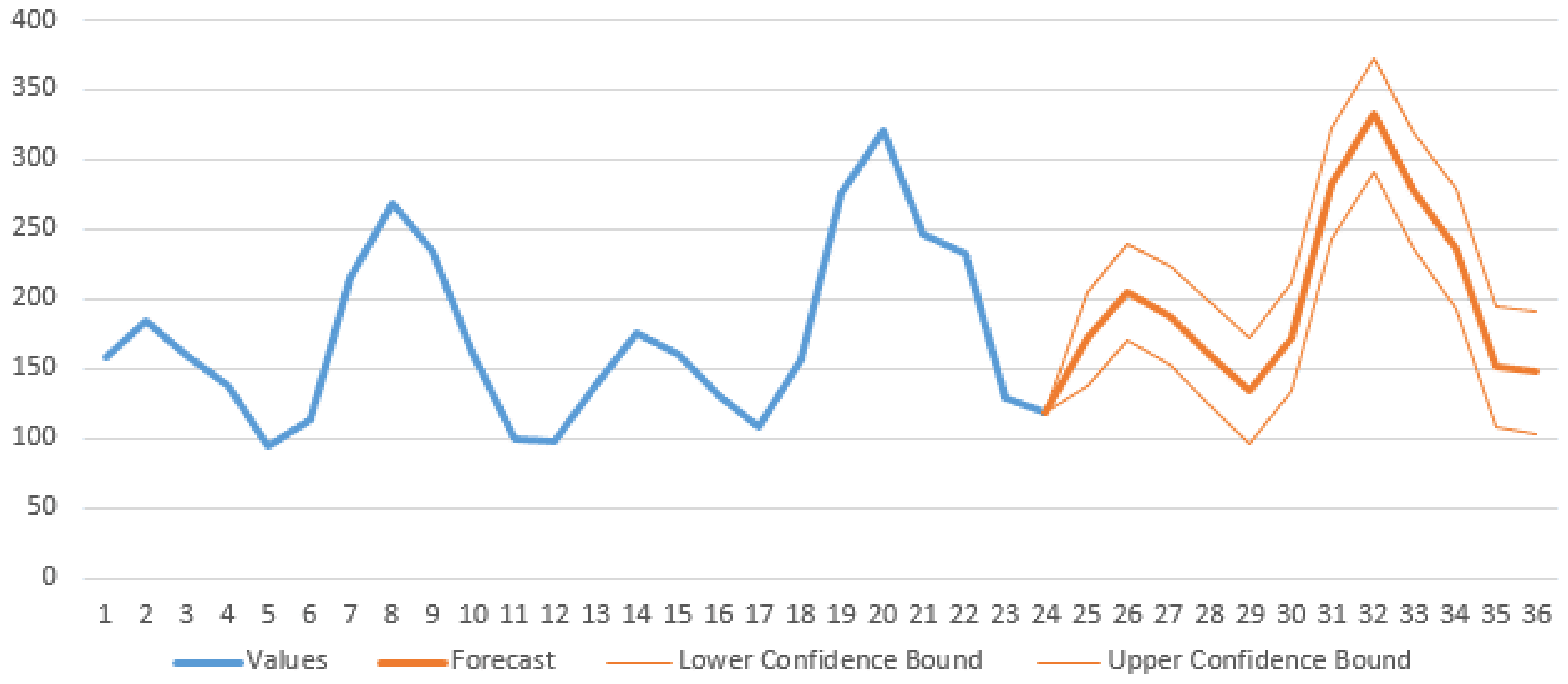
- Explanation: Understand the generator mechanism of the time series. We can use the changes in one variable to explain another variable, and thereby understand how the variables are related. Additionally, we can explain the behavior of a time series by training models (modeling).



INTRODUCTION TO TIME SERIES

OBJECTIVES OF TIME SERIES ANALYSIS

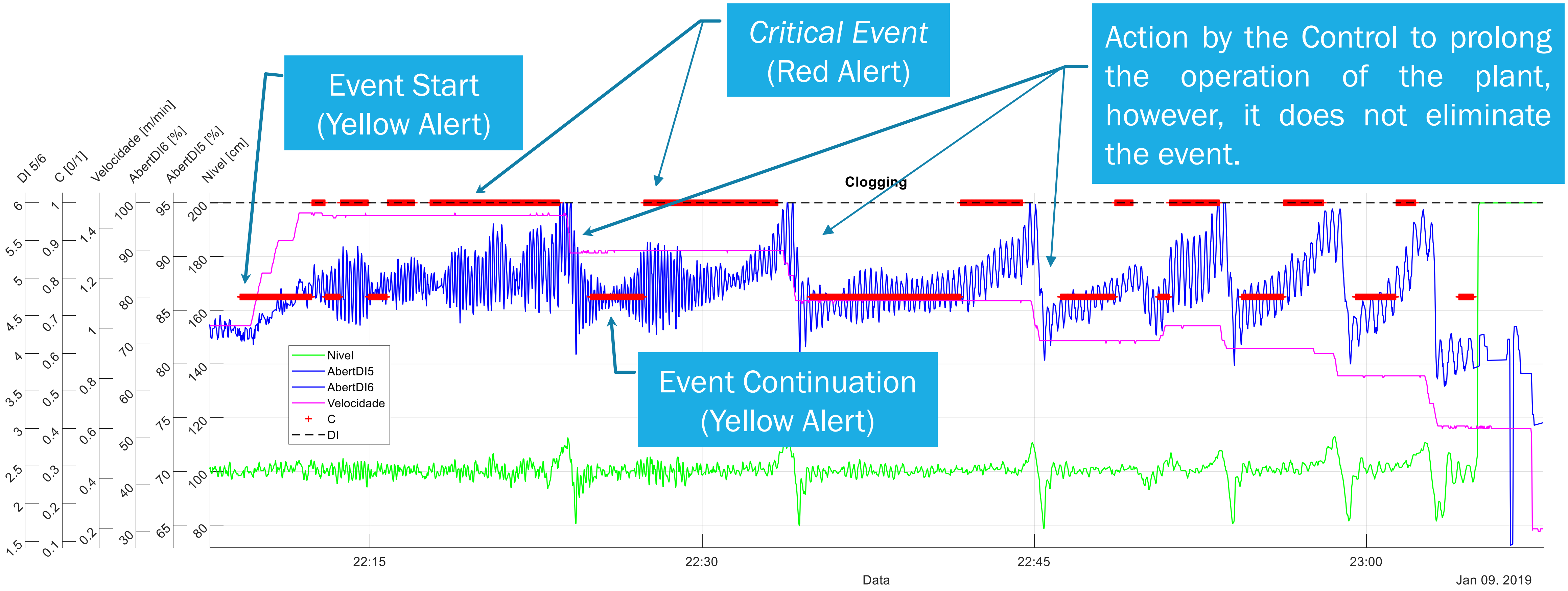
- <u>Forecasting</u>: Given a time series, one may want to predict the future values of the series. It is an important task in many areas, such as economic analysis and industrial process. Prediction and forecasting are <u>normally</u> used interchangeably.



INTRODUCTION TO TIME SERIES

OBJECTIVES OF TIME SERIES ANALYSIS

- <u>Control</u>: When time series is generated to measure the quality of a process to control it (taking of actions to keep the process on target). Control procedures are of several different kinds. Forecasting future values and classifying time series segments help identify the moments to interfere in the process.



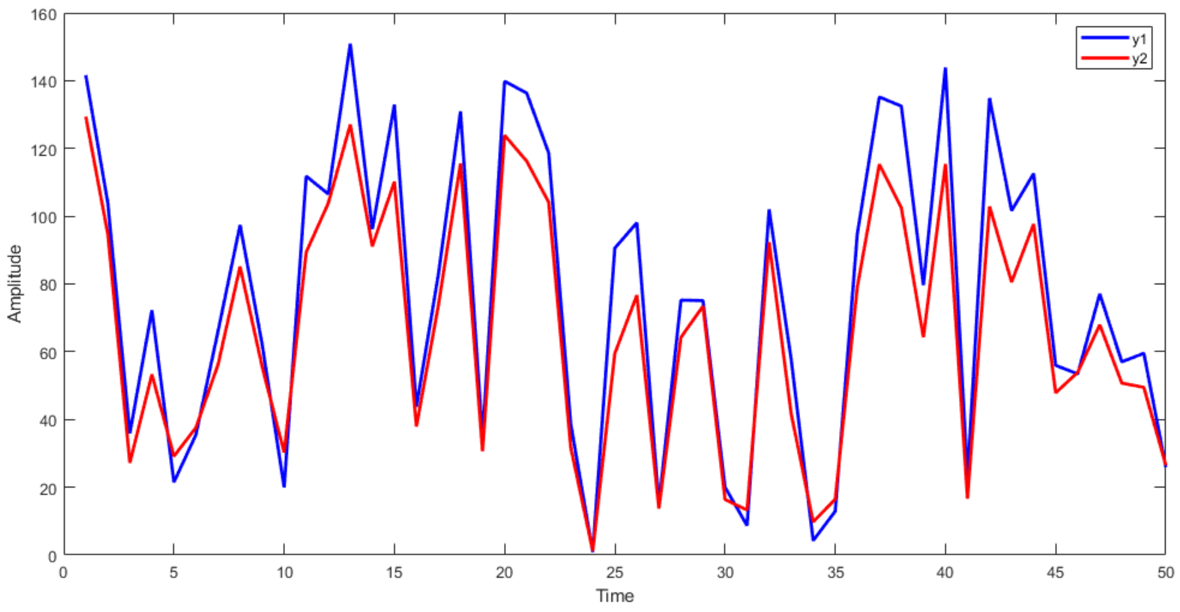
CONCEPTS

INTRODUCTION TO TIME SERIES

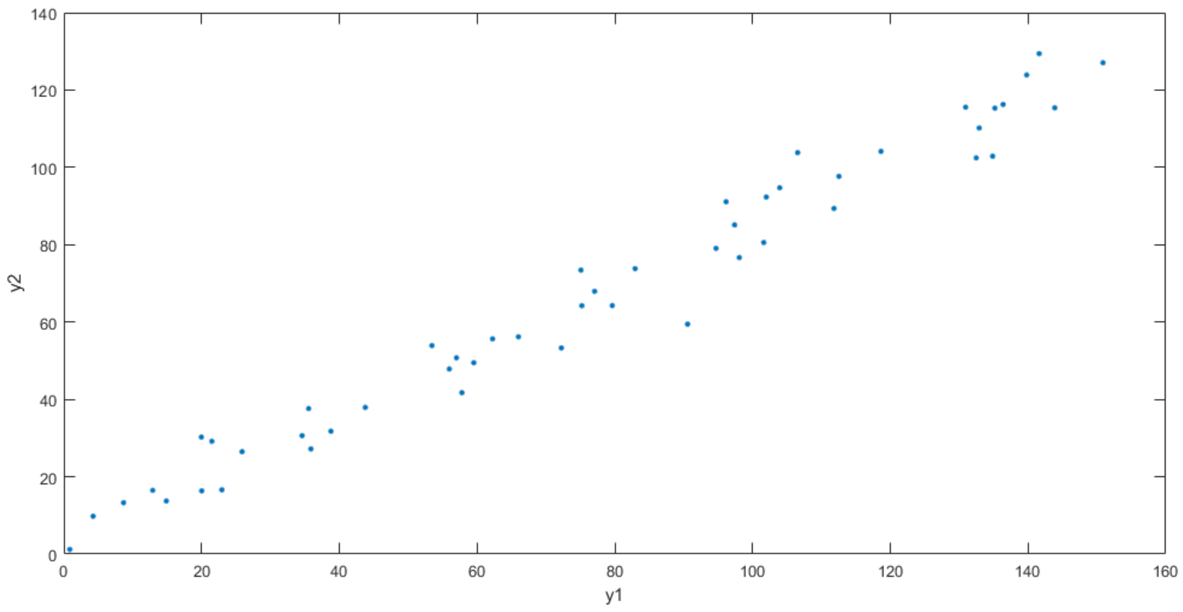
CONCEPTS

Correlation is any statistical dependence between two or more variables.
Collinearity is when a strong linear correlation exists between two or more (multicollinearity) variables.

signals y1 and y2 in time



y1 vs y2

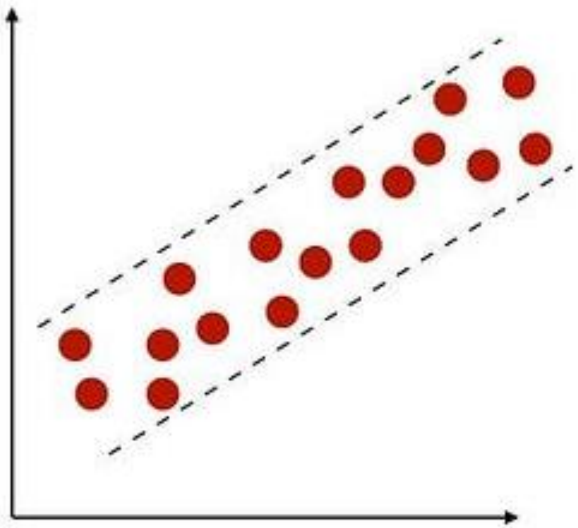


INTRODUCTION TO TIME SERIES

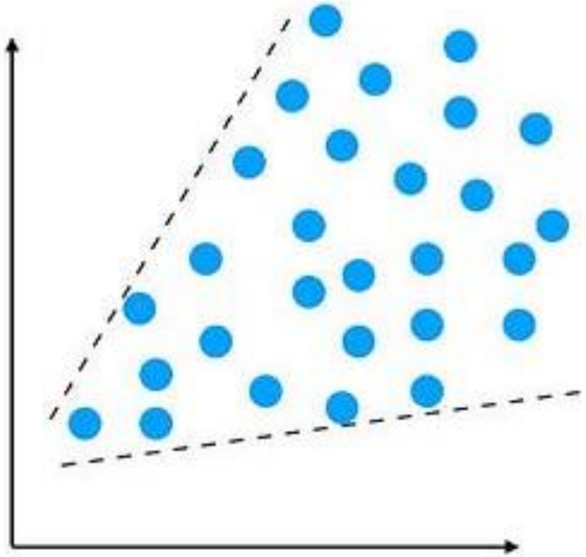
CONCEPTS

<u>Homoscedasticity</u> is when the variable has the same variance across the range of values of a second variable that predicts it.
<u>Heteroscedasticity</u> is when the variance is unequal across the range of values of a second variable that predicts it (the variance is not constant).

Heteroskedasticity is seen as a problem because some regression techniques, such as ordinary least squares (OLS), assume that the variables have constant variance.



homoscedasticity

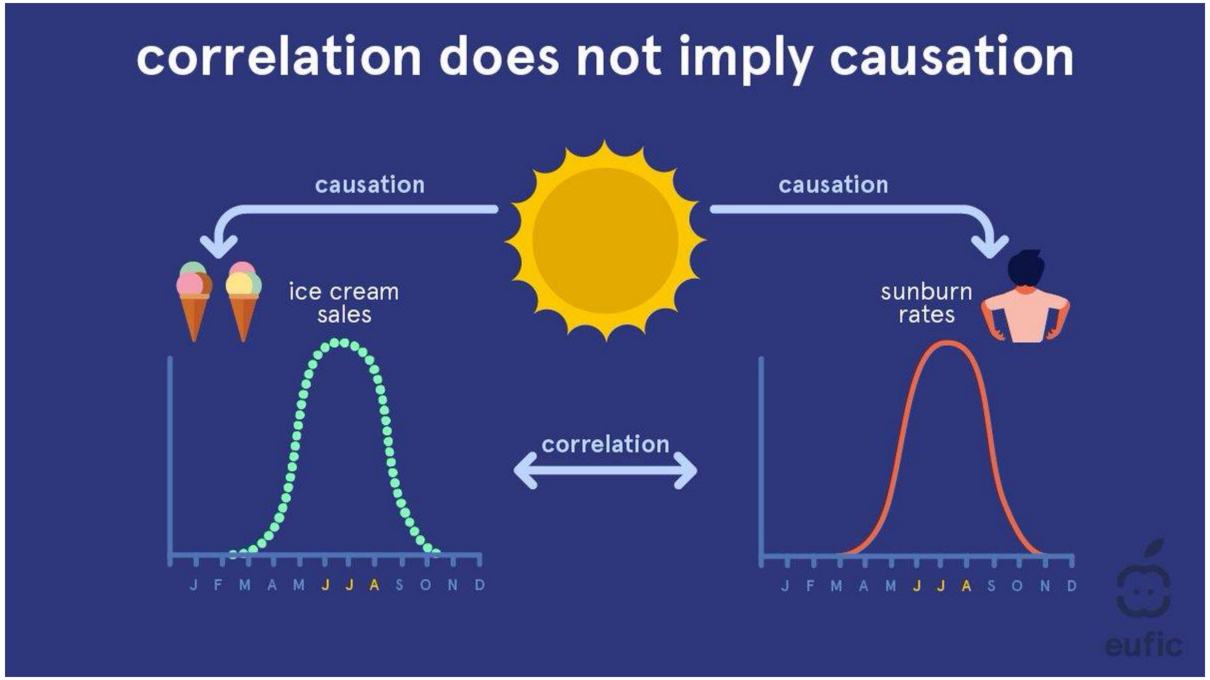


heteroscedasticity

INTRODUCTION TO TIME SERIES

CONCEPTS

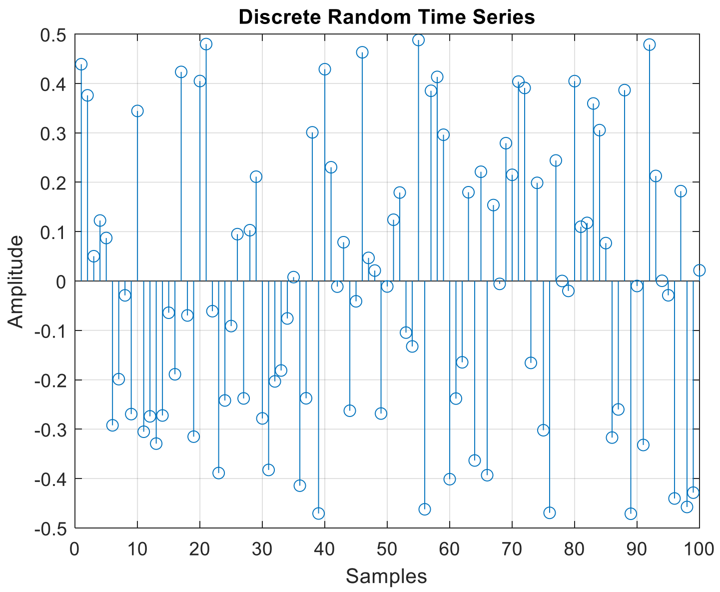
Causality is a characteristic of a system in which its output at any time depends only on input values at the present and past instants, that is, in a subordination of cause and effect.

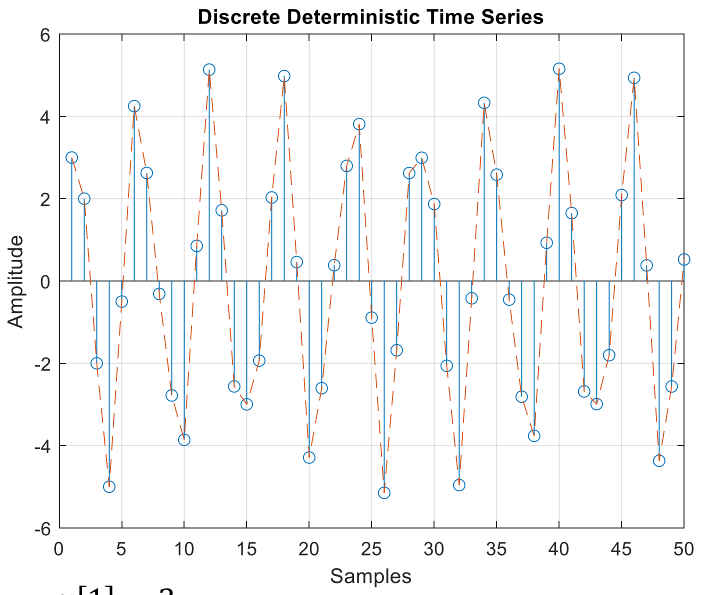


INTRODUCTION TO TIME SERIES

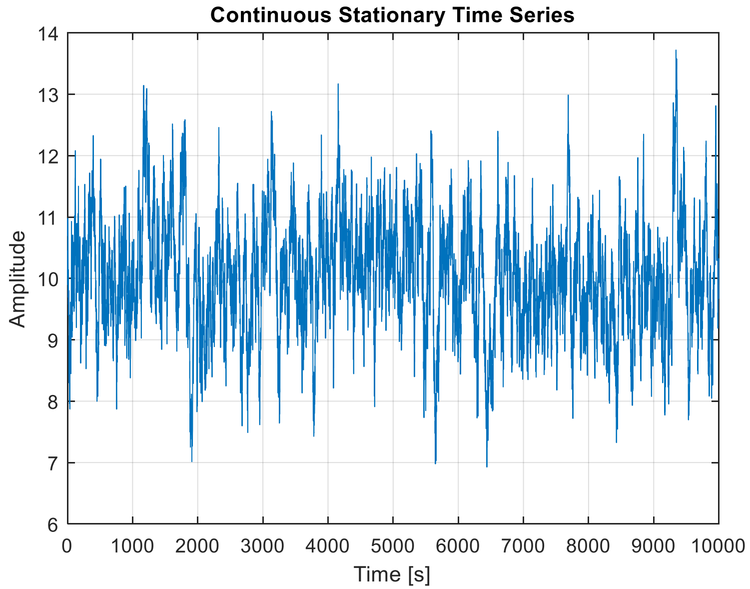
CONCEPTS

A time series is <u>random</u> when adjacent observations are <u>independent</u> of each other.
A time series is <u>deterministic</u> when adjacent observations are <u>dependent</u> of each other.
A time series is <u>stationary</u> when whose properties do not depend on the time at which the series is observed.





$$y[1] = 3$$
$$y[2] = 2$$
$$y[n] = y[n-1] + 0{,}5(-1)^n y[n-1] - y[n-2]$$



(Mean = 10.04, Variance = 0.90)

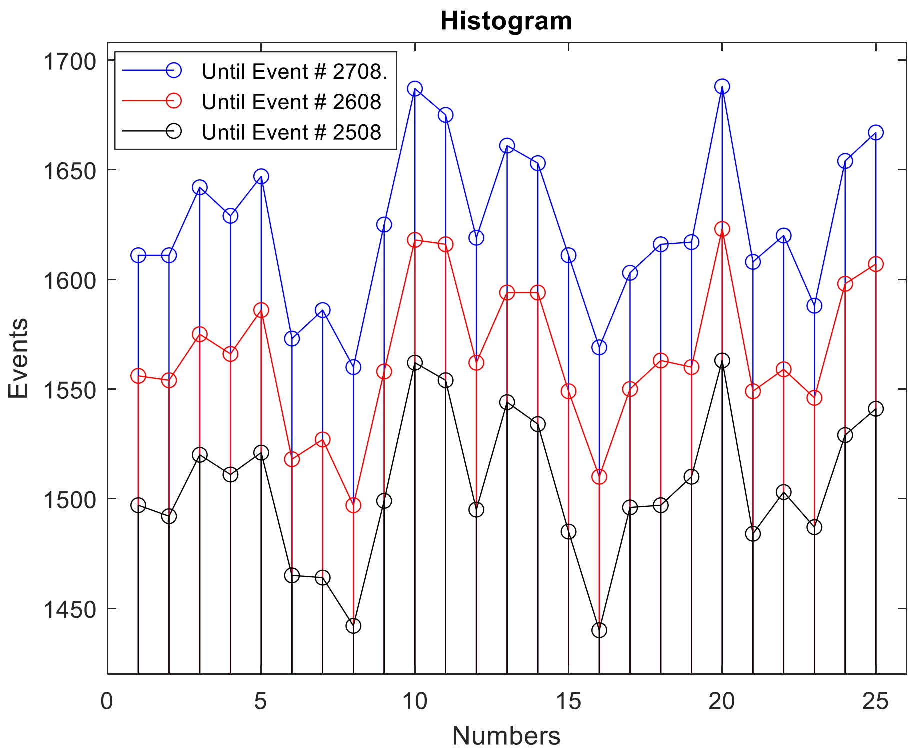
INTRODUCTION TO TIME SERIES

CONCEPTS

Stationarity is associated with the invariance (or constancy) of the statistical properties of a stochastic process.

Let $y(t)$ be a time series. A time series is strictly stationary (or strongly stationary) when its finite-dimensional probability distributions are invariant by time translations.

$$F(y, t+\tau) = F(y, t)$$



INTRODUCTION TO TIME SERIES

CONCEPTS

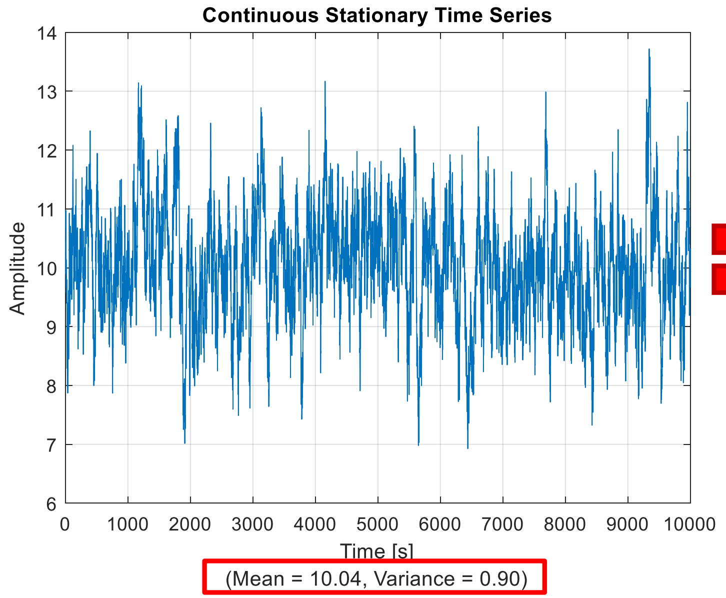
Strongly stationary is quite difficult to verify or test in practice. A less demanding definition of stationary is weakly stationary.

Let $y(t)$ be a time series. A time series is <u>**largely stationary**</u> (or <u>**weakly stationary**</u>) if its mean, variance and autocorrelation are:

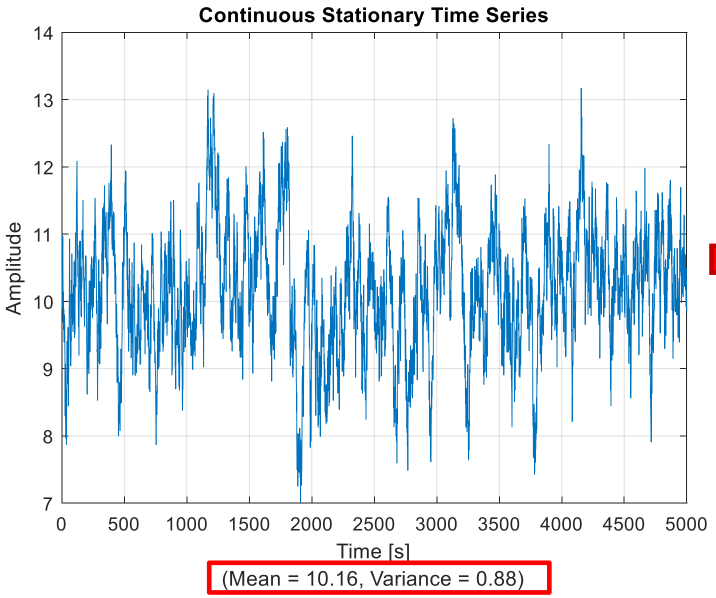
$$E\{y(t)\} \cong \mu \quad \text{(constant)}$$

$$V\{y(t)\} \cong \sigma^2 \quad \text{(constant)}$$

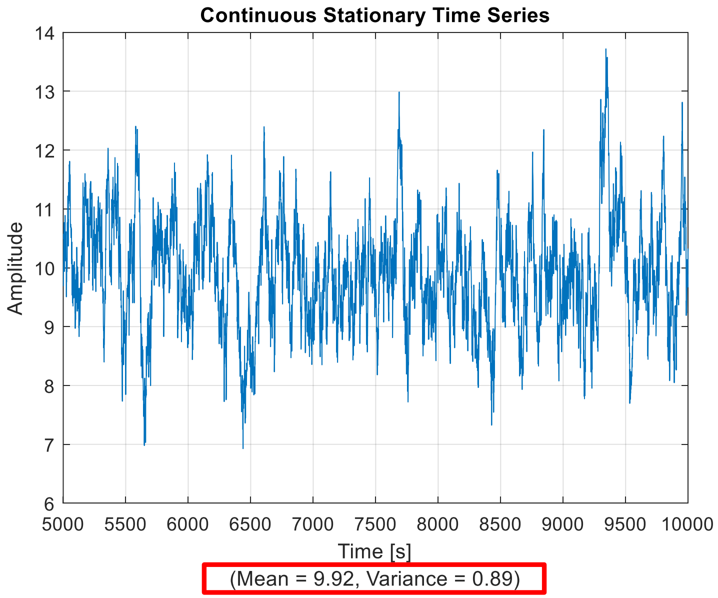
$$\gamma(t_1, t_2) \cong \gamma(t_2 - t_1)$$ (it depends only on the time difference and not on the instant at which it is calculated)



Signal from 1 to 10000



Signal from 1 to 5000



Signal from 5001 to 10000

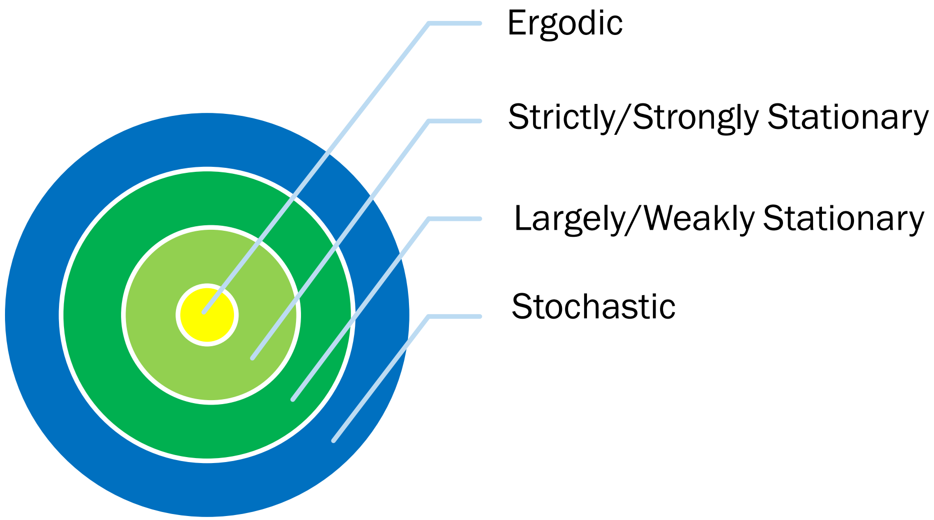
INTRODUCTION TO TIME SERIES

CONCEPTS

Ergodicity is the condition in which only one execution of the process is enough to obtain the statistics of the entire process.

A weakly stationary process is ergodic to the mean, that is, the time average converges as time tends to infinity.

Stationarity does not imply ergodicity. Ergodicity is a restrictive form of stationarity.



INTRODUCTION TO TIME SERIES

CONCEPTS

Example of stationary process that is not ergodic:

Let $X(t) = A$ for all t, where A is a zero-mean, unit-variance random variable, In other words, $X(t)$ can assume any value of the random variable A, however, its value does not change over time ($X(t)$ is not time-dependent).

Note that the process is stationary, because its properties do not depend on the time at which the values are observed.

The mean of many executions of the process is:

$$\mu\big(X(t)\big) = \mu(A) = 0$$

However, the time average of a unique execution of the process is A. This value does not always converge to zero as time tends to infinity (except when $X(t) = 0$).

Therefore, this process is not ergodic because it is not possible to obtain the statistics of the entire process with only a unique execution.

INTRODUCTION TO TIME SERIES

CONCEPTS

Examples of ergodic processes:

Example 1: Let $X(t) = A$, so that for each time t, $X(t)$ can assume a value of A, being A a zero-mean, unit-variance random variable.

This process is also stationary, because its properties do not depend on the time at which the values are observed.

In this case, with a single and sufficiently long execution of the process, we have its statistical properties. Therefore, this process is ergodic.

Example 2: Let $X(t)$ be a process that represents repeated rolls of an unbiased six-sided die, such that at each instant t the die is rolled.

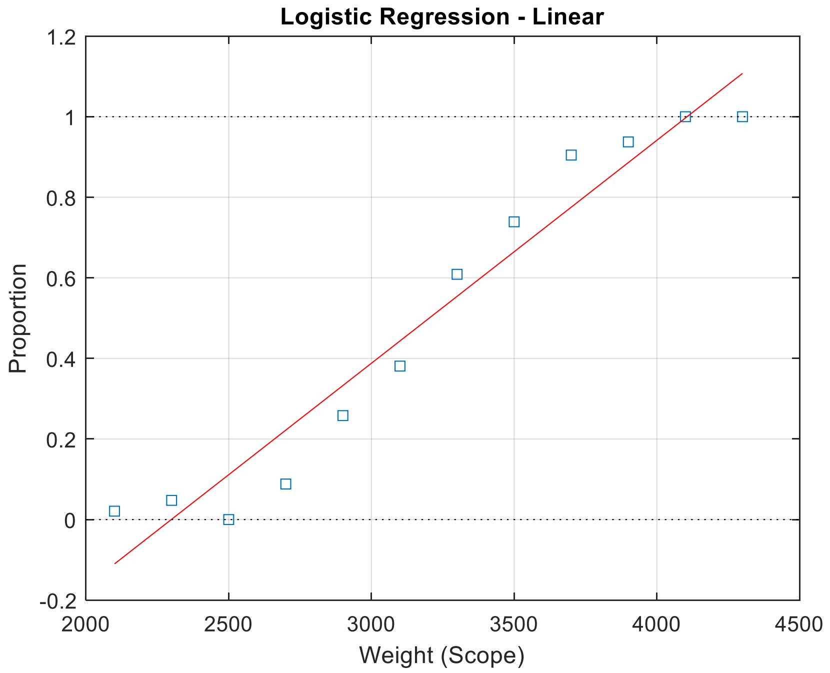
In this case, a single and sufficiently long execution of the process is enough to obtain the statistical properties of the process. Therefore, the process is ergodic.

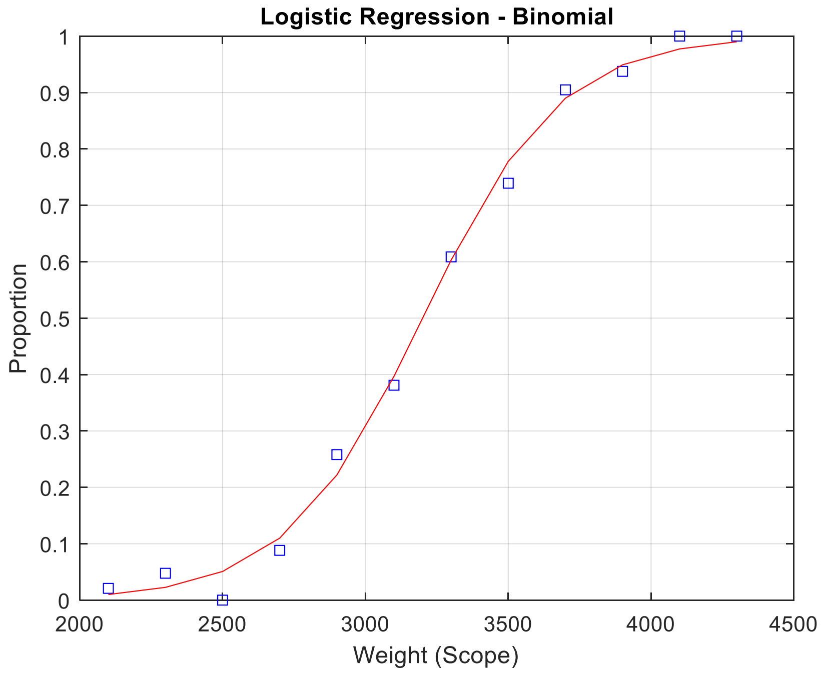
INTRODUCTION TO TIME SERIES

CONCEPTS

Regression analysis is a set of methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the relationship between them.

It is normally useful only for estimating values surrounding the known values of the target variable (within the scope of the variable). There need not be a temporal relationship between the samples of the variable (i.e., it need not be a time series).





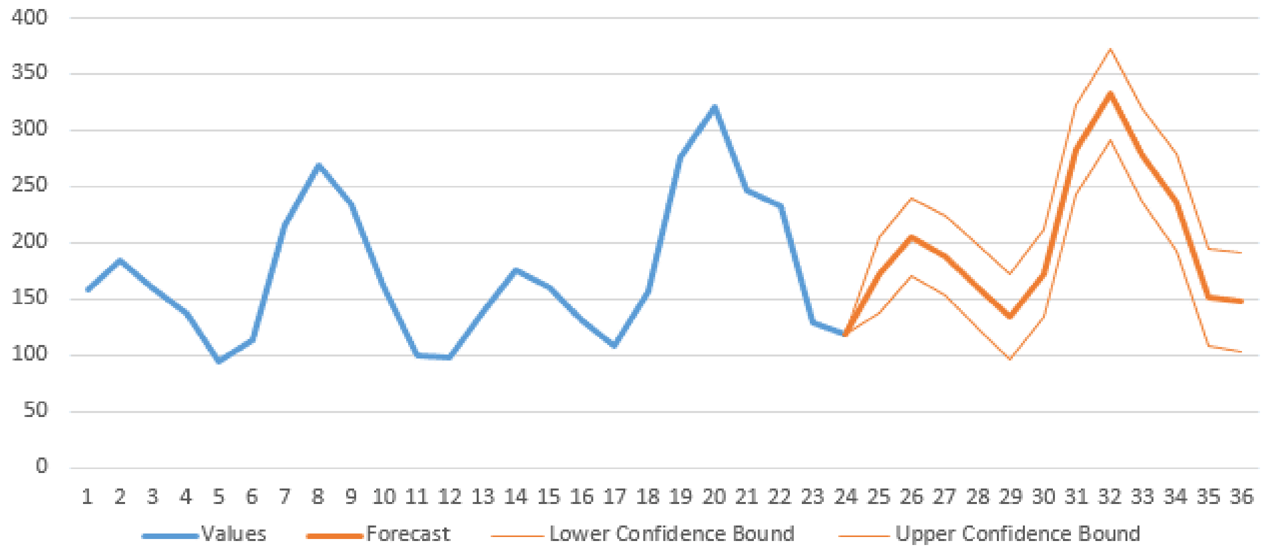
INTRODUCTION TO TIME SERIES

CONCEPTS

Forecasting is a method that consists of estimating future values of a variable based on present and past values of the same variable.

It is an extrapolation problem and seeks to predict what is outside the scope of the variable in a temporal relationship (i.e., applied only to time series).

Other variables can also be used to forecast the values of the target variable. They are called exogenous variables. Their future values must be known in order to include them in the prediction process. The inclusion of exogenous variables can enhance the accuracy of forecasts.

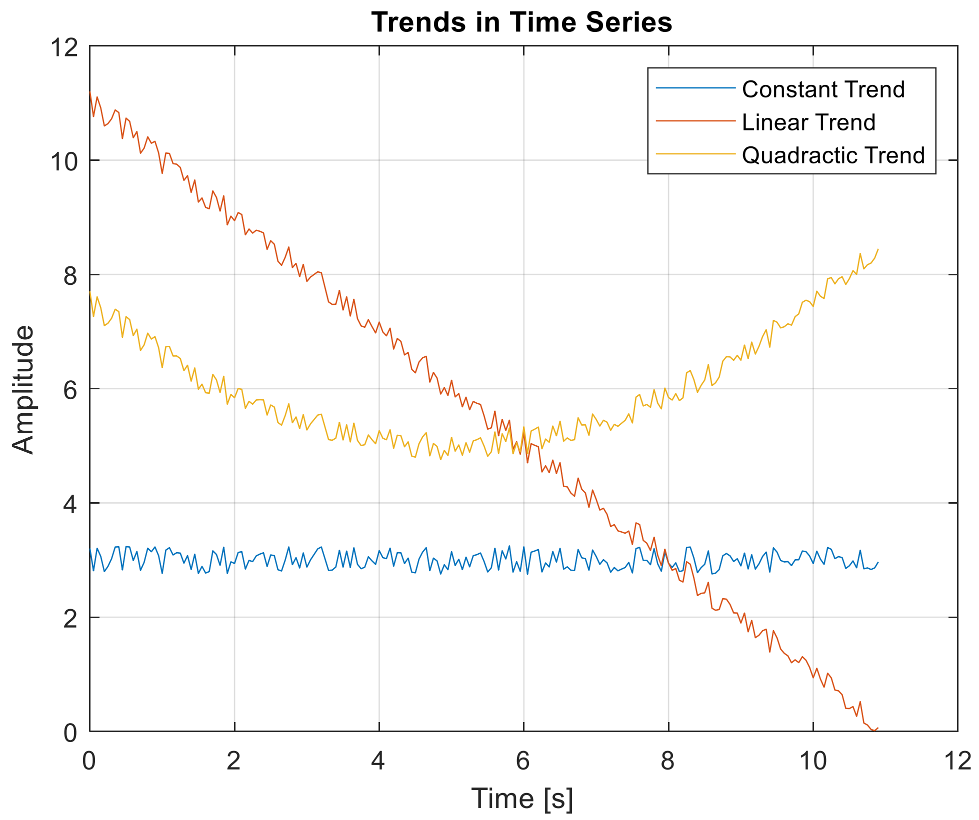


TYPICAL PHENOMENA IN TIME SERIES

INTRODUCTION TO TIME SERIES

TYPICAL PHENOMENA IN TIME SERIES

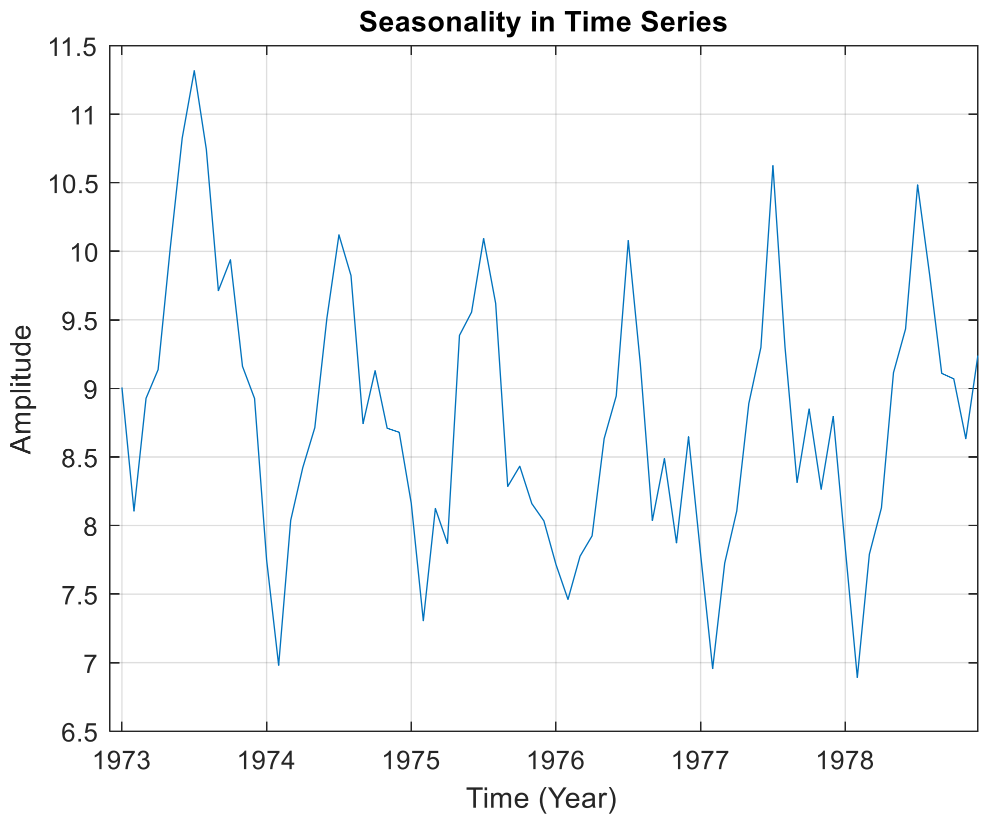
The <u>trend</u> of a series indicates its “long-term” behavior, that is, whether it increases, decreases, or remains stable, and how fast are these changes. In the most common cases, the trend is a constant, linear or quadratic.



INTRODUCTION TO TIME SERIES

TYPICAL PHENOMENA IN TIME SERIES

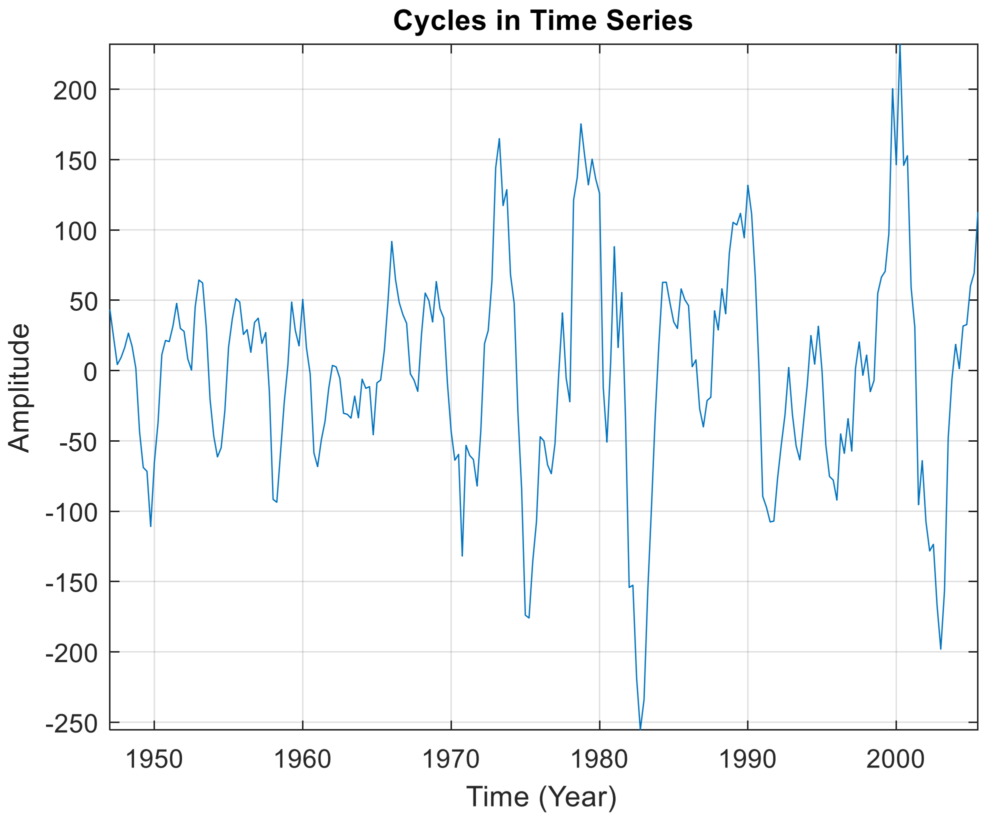
The <u>seasonality</u> is a characteristic of a time series in which the data experience regular and predictable changes that are repeated at each time interval. Seasonality is always of a fixed and known frequency, such as every year or every week.



INTRODUCTION TO TIME SERIES

TYPICAL PHENOMENA IN TIME SERIES

Cycles are characterized by rising and falling oscillations in the series that are not of a fixed frequency.

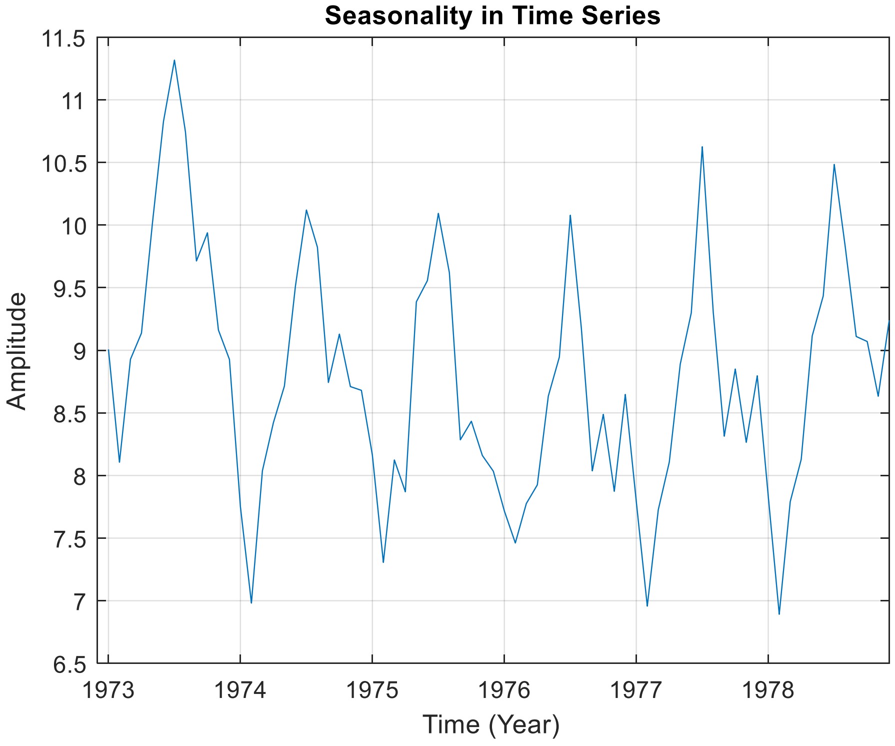


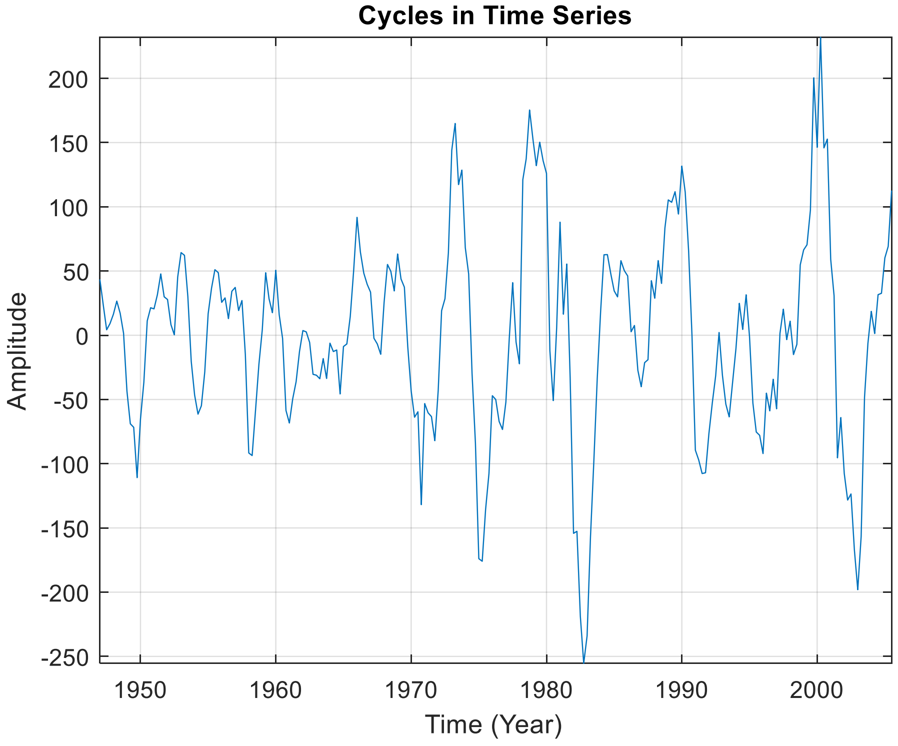
INTRODUCTION TO TIME SERIES

TYPICAL PHENOMENA IN TIME SERIES

Cycles vs Seasonality.

The essential difference between the seasonal and cyclical components is that the former has easily predictable movements, occurring at intervals regular movements of time, whereas cyclic movements tend to be irregular (no fixed frequency).



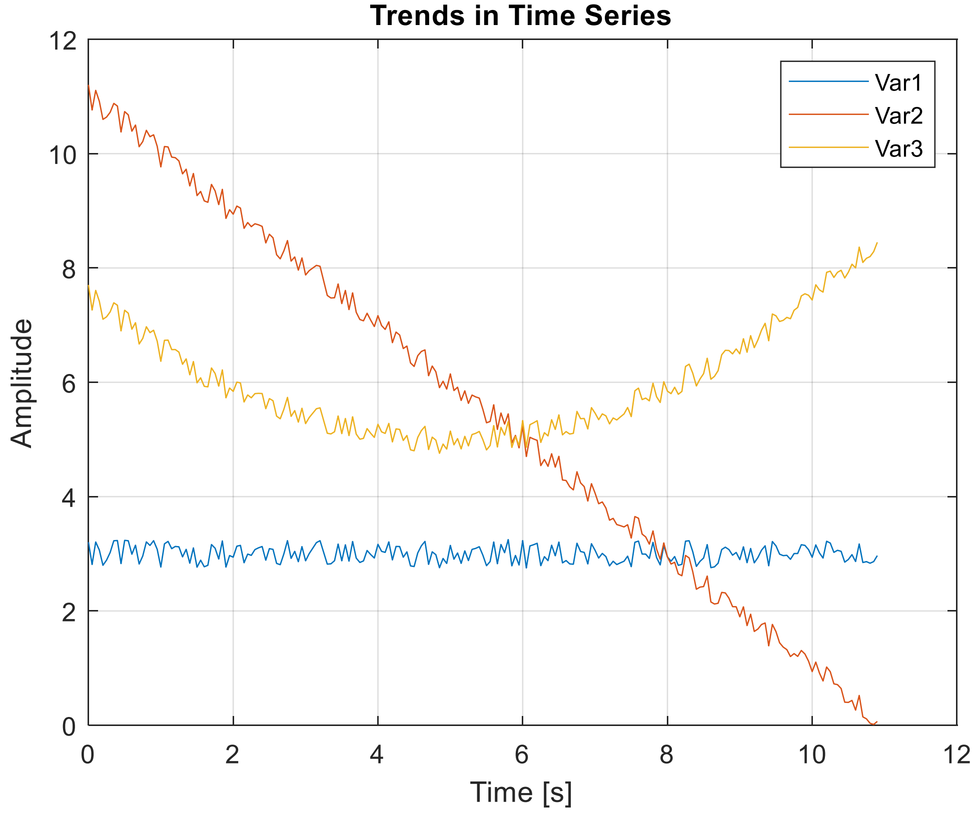


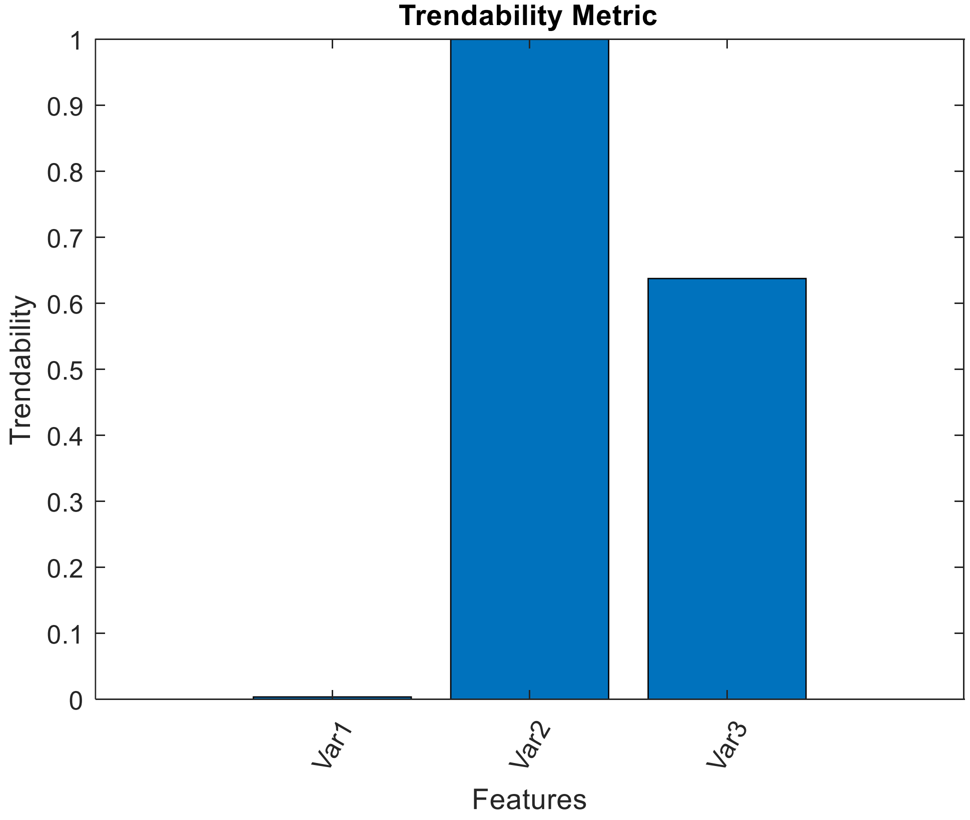
MATHEMATICAL MEASUREMENTS IN TIME SERIES

INTRODUCTION TO TIME SERIES

MATHEMATICAL MEASUREMENTS IN TIME SERIES

<u>Trendability</u> is the measure of similarity between the trajectories of a feature measured in several run-to-failure experiments. A more trendable feature has trajectories with the same underlying shape. Conversely, any feature that is non-trendable is a less suitable condition indicator. The values of trendability range from 0 to 1: trendability is 1 if Var is perfectly trendable. Trendability is 0 if Var is perfectly non-trendable.

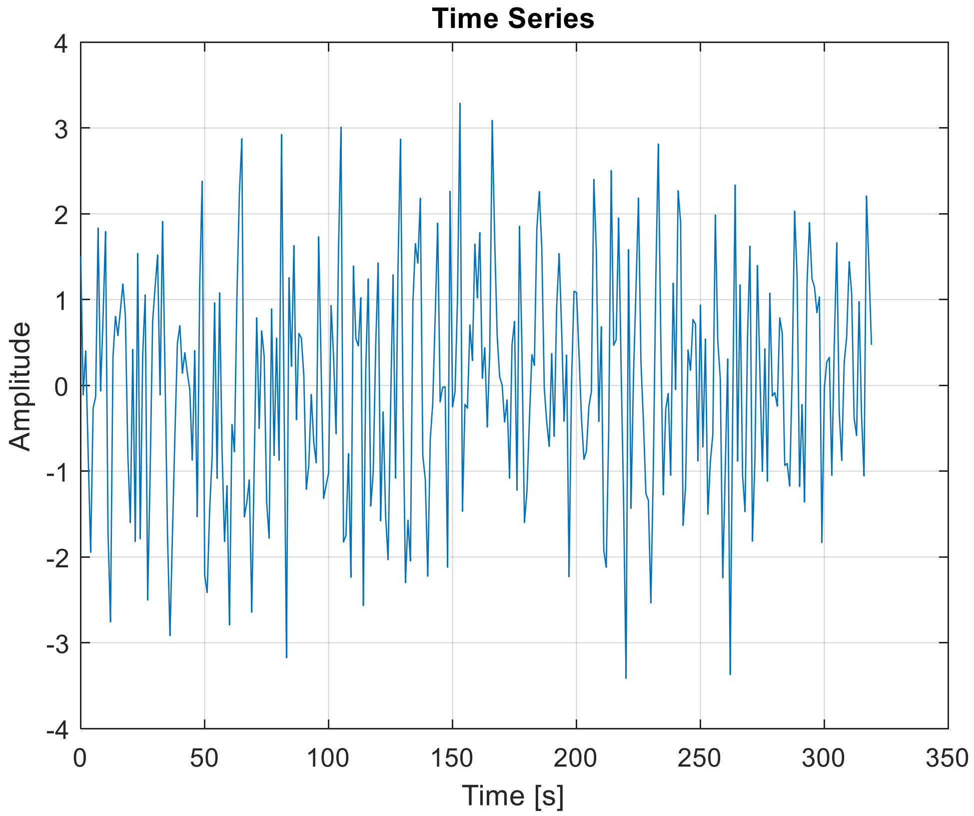


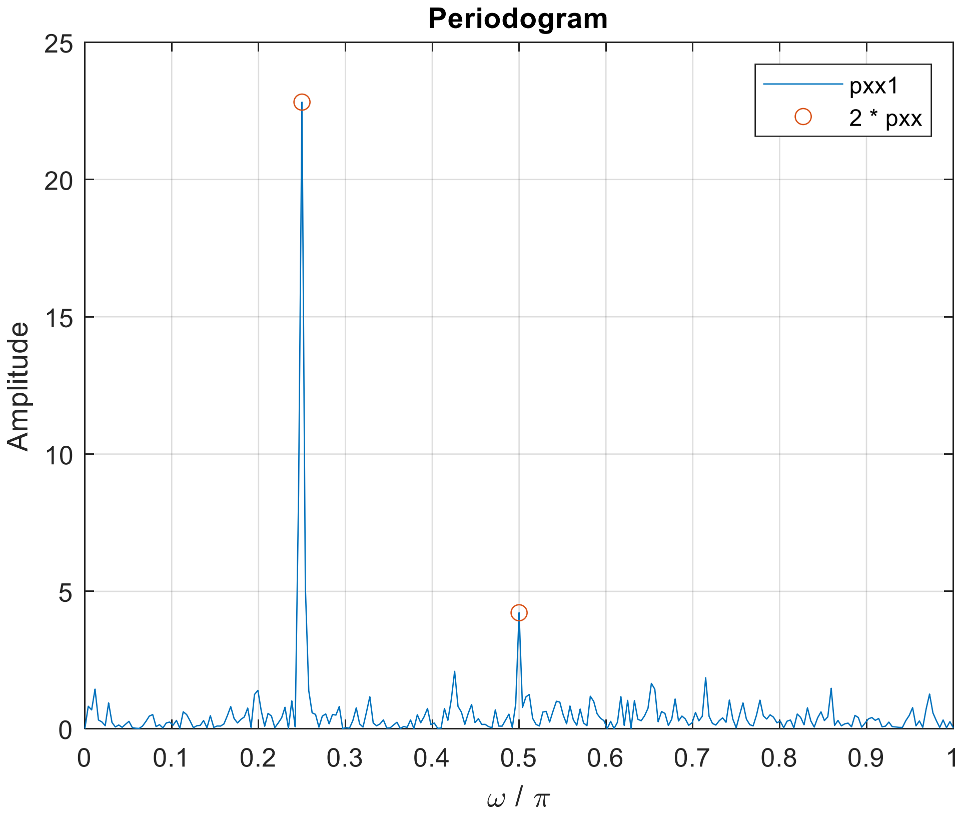


INTRODUCTION TO TIME SERIES

MATHEMATICAL MEASUREMENTS IN TIME SERIES

<u>Periodogram</u>: is an estimate of the spectral density of a signal. A periodogram is used to identify the dominant periods (or frequencies) of a time series.



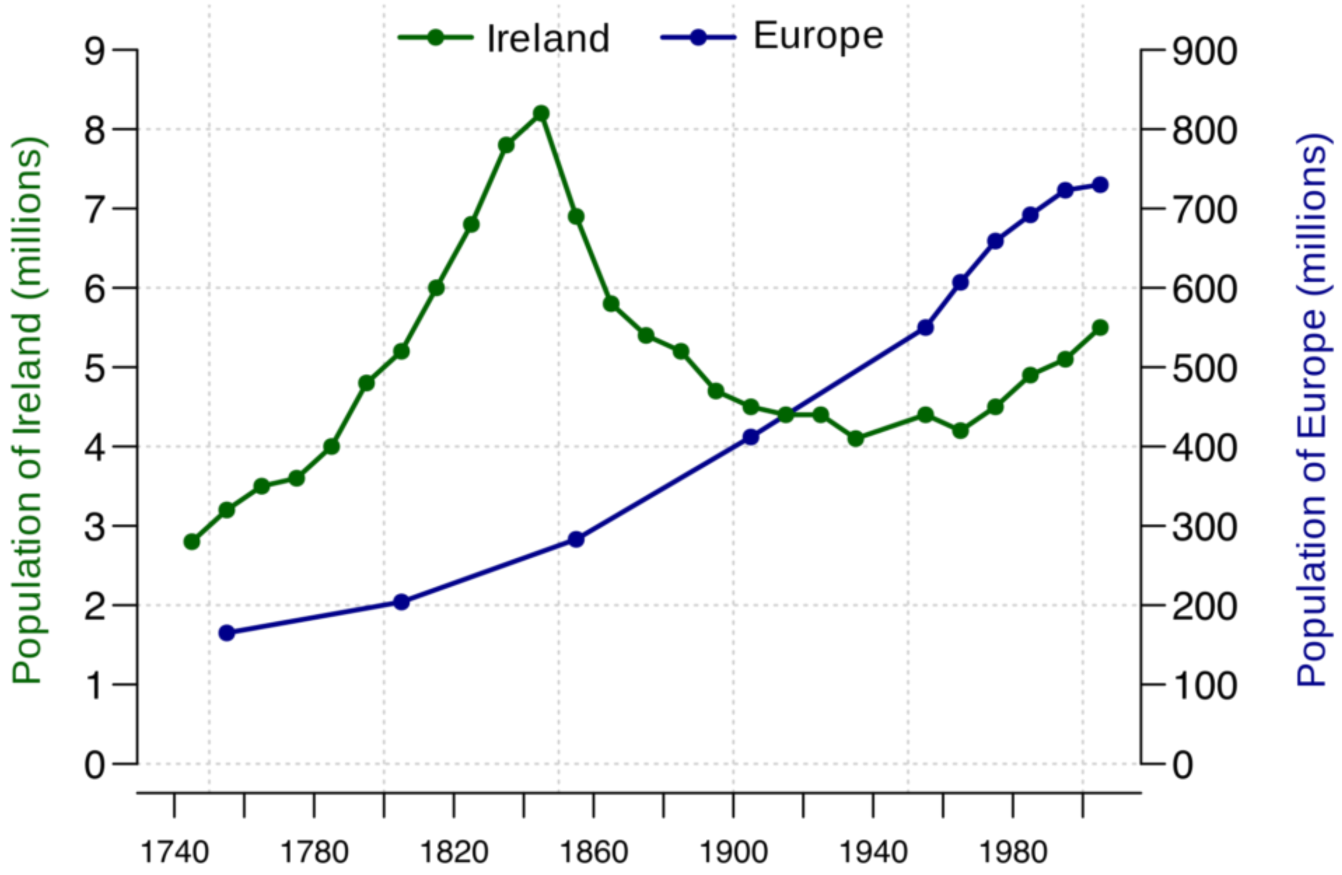


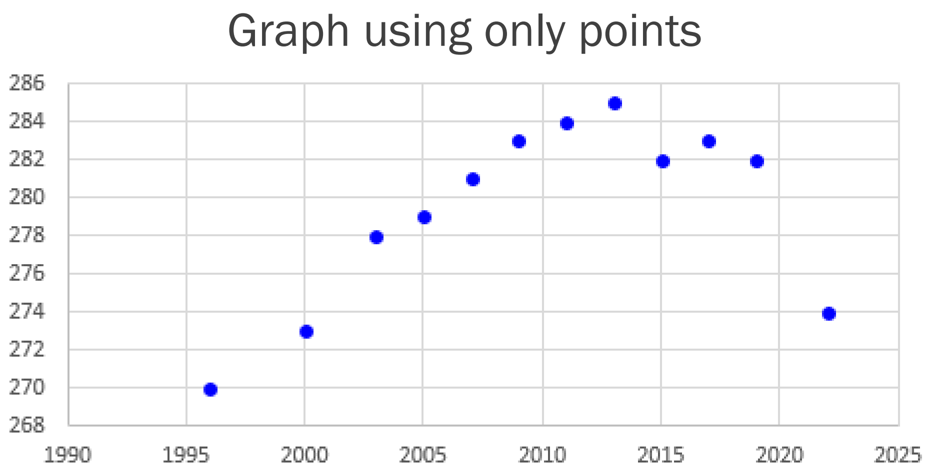
VISUALIZATION OF TIME SERIES DATA

INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

<u>Line Graph</u>: is the simplest way to represent time series data. It is intuitive, easy to create, and helps the viewer get a quick sense of how something has changed over time. It uses points connected by lines to show how a dependent variable and independent variable changed.

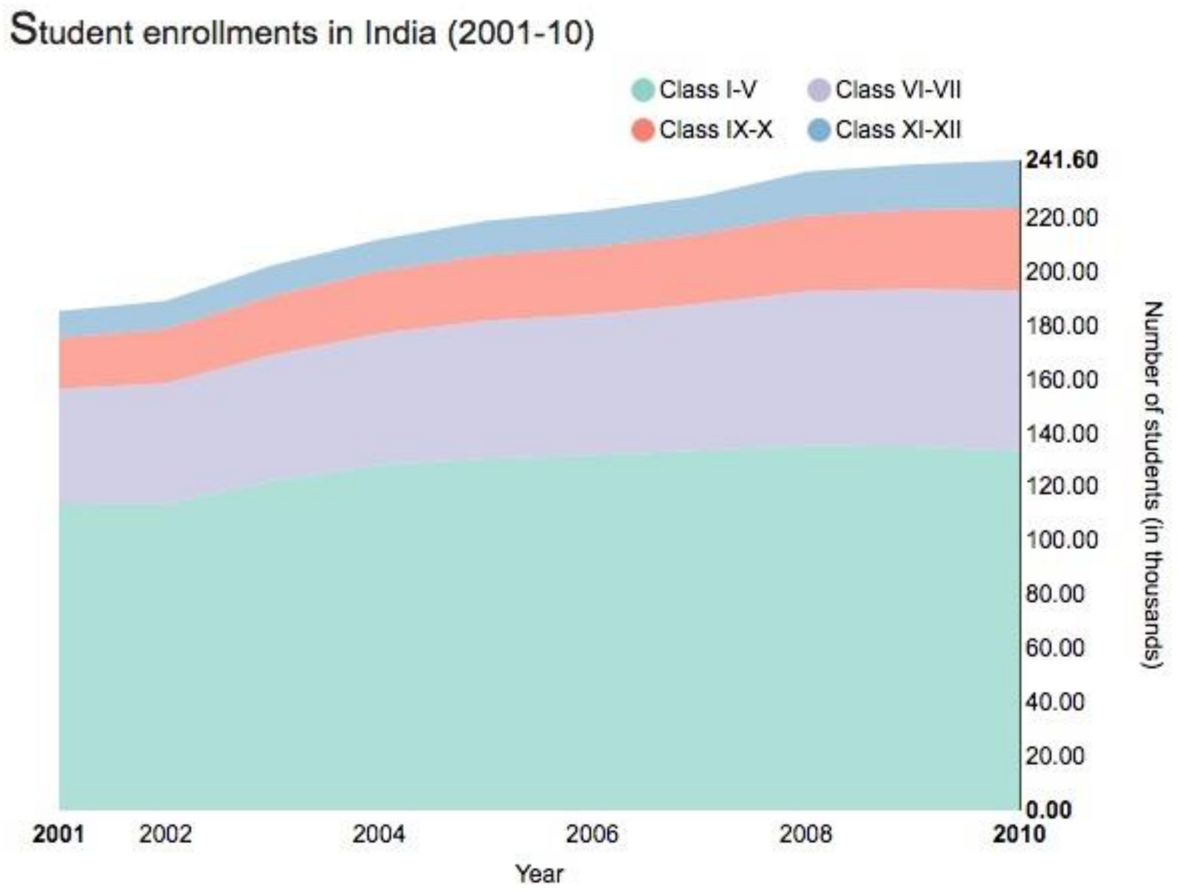




INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

Stacked Area Chart: is similar to a line graph in that it has points connected by straight lines on a two-dimensional chart. However, in an area chart, multiple variables are “stacked” on top of each other, and the area below each line is colored to represent each variable.



INTRODUCTION TO TIME SERIES

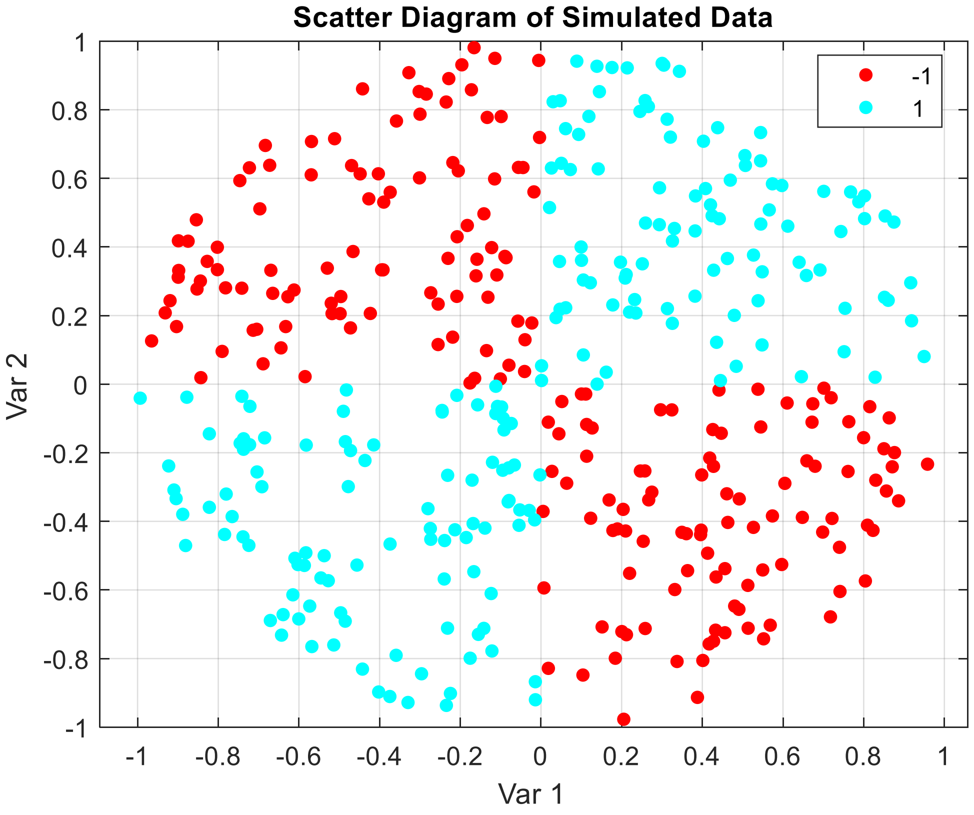
VISUALIZATION OF TIME SERIES DATA

Waterfall Chart: shows how an initial value is increased or decreased by a series of intermediate values. They are often used in accounting to visualize the different components of a total income or expense. If you need to visualize how an initial value changes as a result of intermediate values, a waterfall chart is the way to go.

INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

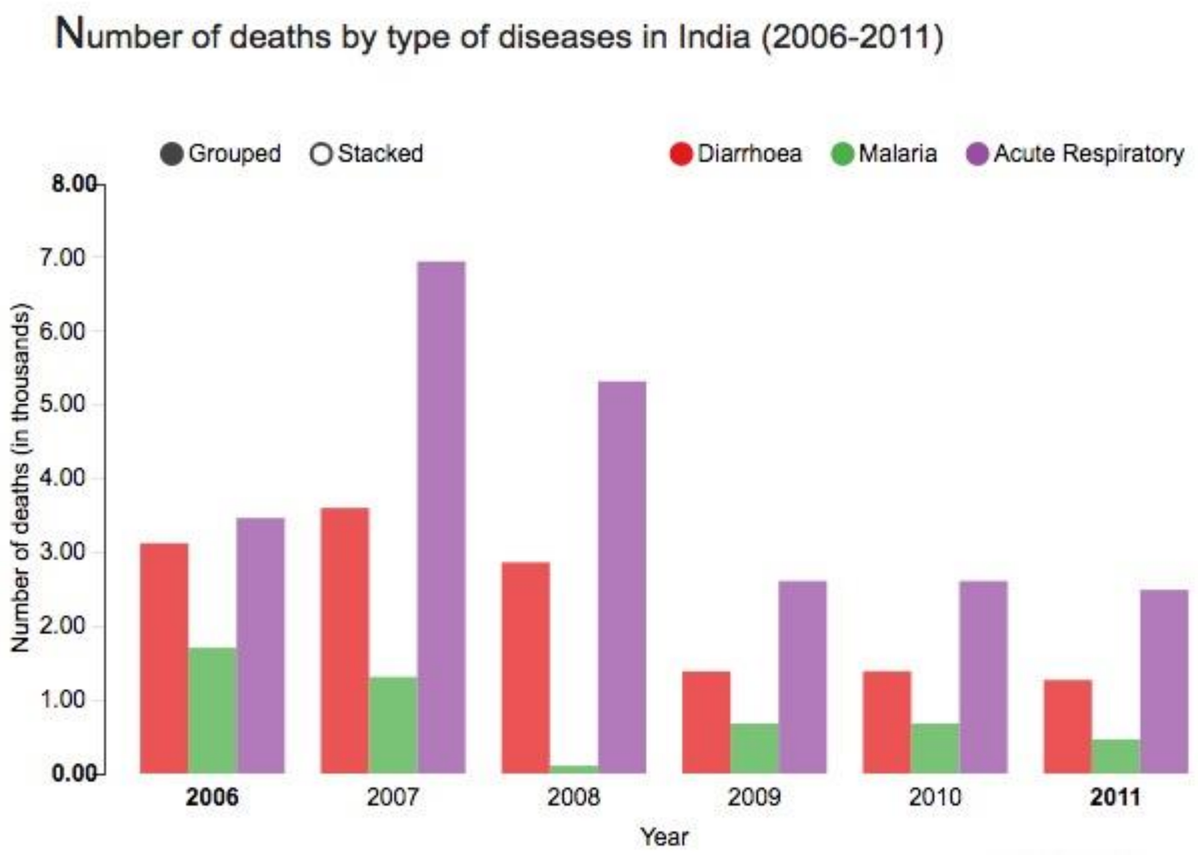
<u>Scatter Diagram</u>: is a graphical representation of the relationship between two numeric variables. It is useful to identify correlation between variables.



INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

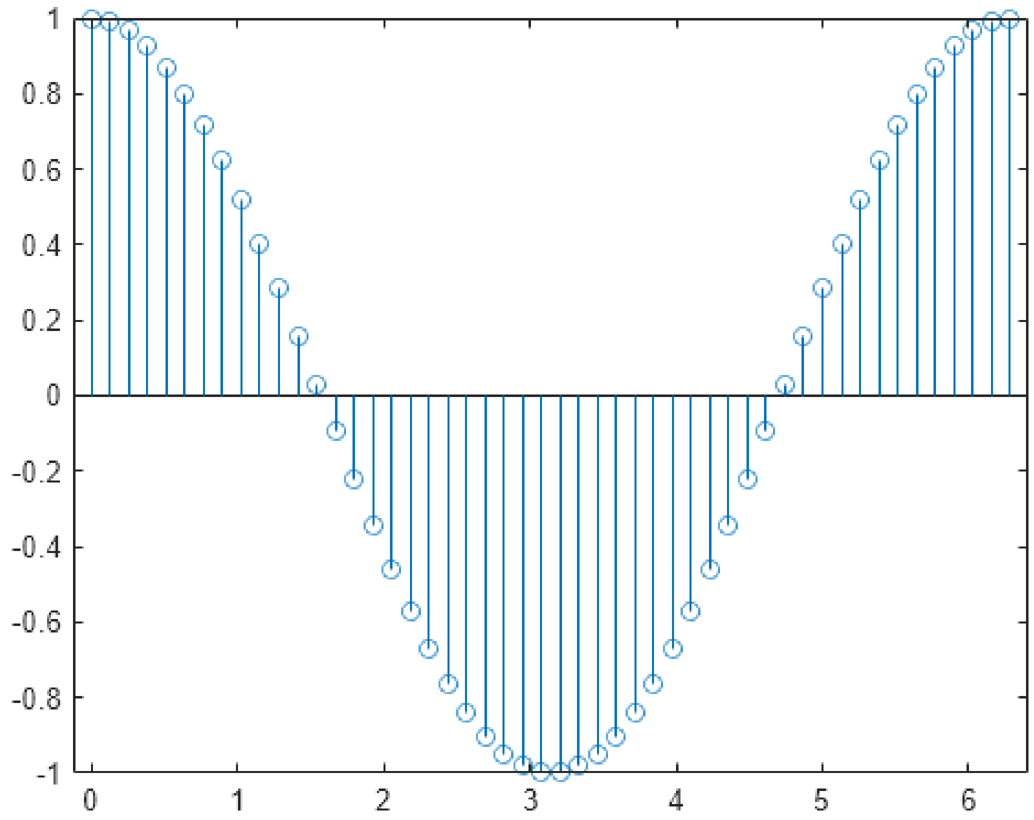
Bar Chart: represents data as horizontal or vertical bars. The length of each bar is proportional to the value of the variable at that point in time. A bar chart is the right choice for you when you wish to look at how the variable moved over time or when you wish to compare variables versus each other.



INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

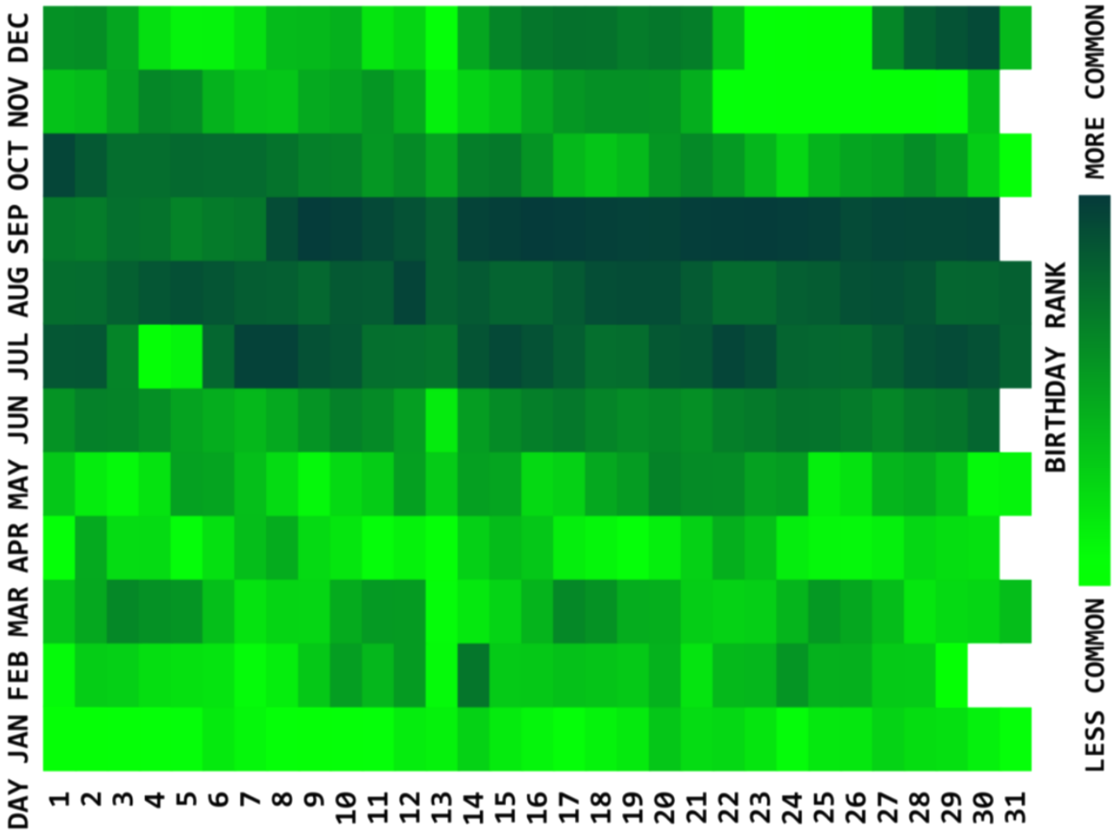
<u>Stem Plot</u>: is quite similar to a bar chart. The difference is that, instead of using a bar for each data point, a stem plot uses a vertical line with a marker at the top. Stem plots can be useful if there are a lot of data points. Since the line and marker occupy less area than the bar, the plot looks less cluttered.



INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

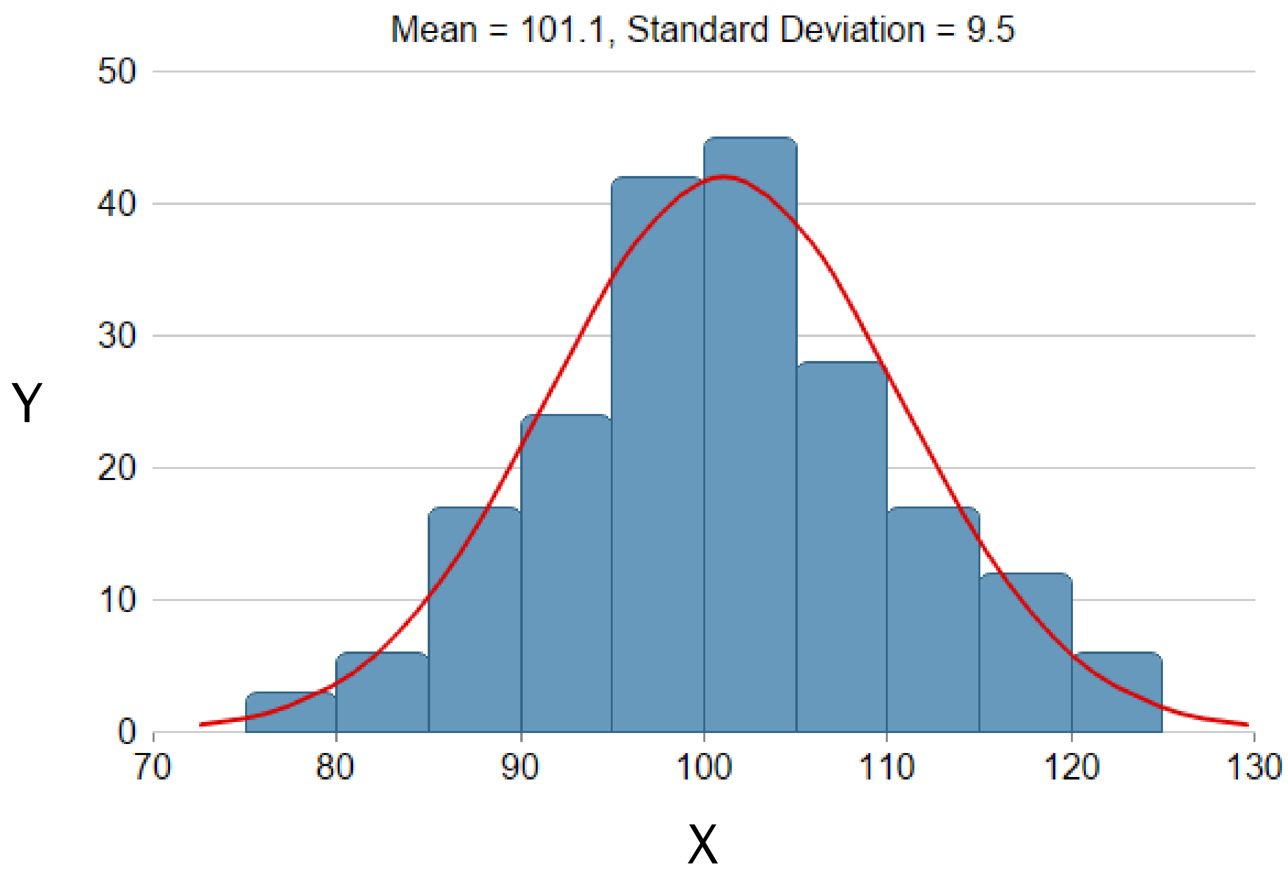
<u>Heat Map:</u> helps identify quickly “hot spots” or regions of high concentrations of a given variable. When adapted to temporal visualizations, heat maps can help us explore two levels of time in a 2D array.



INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

<u>Histogram:</u> is a way to view the distribution of data. The y-axis is dedicated to count, and the x-axis is divided into bins. The Histogram visualization is a bar graph that displays the number of data points that fall within "bins".



INTRODUCTION TO TIME SERIES

VISUALIZATION OF TIME SERIES DATA

<u>Boxplot:</u> displays a lot of information about the data. Among other things, the median, the interquartile range (IQR) and the outliers can be read in a boxplot.