

## Data visualization

It is the graphical representation of information and data.

By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

### 8 R Data Visualization tools:

- |                    |                 |
|--------------------|-----------------|
| 1. Bar Chart       | 2. Histogram    |
| 3. Box Plot        | 4. Scatter Plot |
| 5. Correlogram     | 6. Heat Map     |
| 7. Hexagon Binning | 8. Mosaic Plot  |

#### 1. Bar Chart

It is to show a comparison between two or more variables. Bar charts depict the comparison between the cumulative total across various groups.

Types of bar charts:

horizontal bar charts (standard format)

vertical bar charts (standard formats)

R can create both horizontal and vertical bars in a chart.

Besides, R also offers a stacked bar chart that lets you introduce different variables to each category.

In R, the `barplot()` is used to create bar charts.

## 2. Histogram

This representation breaks the data into bins (breaks) and depicts the frequency distribution of these bins.

Histograms provide a probability estimate of a variable, that is, the time period before the completion of a project

## 3. Box Plot

Box plot depicts five statistically significant numbers:

the minimum	the 75th percentile
the 25th percentile	the maximum.
the median	

## 4. Scatter Plot

Scatter plots depict numerous points in the Cartesian plane, wherein each point represents the values of two variables.

You can choose one variable in the horizontal axis and the second one in the vertical axis.

The function of a scatter plot is to track two continuous variables over time

## 5. Correlogram

correlogram, or correlation matrix, analyzes the relationship between each pair of numeric variables in a dataset.

It provides a quick overview of the complete dataset.

Correlograms can highlight the correlation amount between datasets at various points in time.

## 6. Heat Map

Heat maps are graphical representations of data in which individual values contained in a matrix are represented via different colors.

Heat maps allow you to perform EDA

## 7. Hexagon Binning

Hexagon binning is a type of bivariate histogram best suited for visualizing the structure in datasets with large n

## 8. Mosaic Plot

In R programming, the mosaic plot comes in handy while visualizing data from the contingency table or two-way frequency table.

It is a graphical representation of a two-way contingency table that represents the relationship between two or more categorical variables.

## Data Pre-processing

Many real world data is dirty and needs to be cleaned before used in code.

The process of cleaning a dataset is called Data Preprocessing.

Eg: Time series data

### Transforms for Time Series Data

Given a univariate time series dataset, there are four transforms that are popular when using machine learning methods to model and make predictions.

- Power Transform
- Difference Transform
- Standardization
- Normalization

## Autoregressive (AR) model

Autoregressive (AR) models are a subset of time series models, which can be used to predict future values based on previous observations.

AR models use regression techniques and rely on autocorrelation in order to make accurate predictions.

Time-series forecasting is the process of using a model to predict future values based on previously observed values.

Autoregressive models are powerful tools in the data scientist's toolbox for understanding how one variable may predict another.

Autoregressive (AR) modeling is one of the techniques used for time-series analysis. An autoregressive model is a time-series model that describes how a particular variable's past values influence its current value. In other words, an AR model attempts to predict the next value in a series by incorporating the most recent past values and using them as input data

Showing ACP

Showing PACP

Here are some of the alternative time-series forecasting methods to the AR modeling technique:

- MA (Moving average)
- ARMA (Autoregressive moving average)
- ARIMA (Autoregressive integrated moving average)
- SARIMA (Seasonal autoregressive integrated moving average)
- VAR (Vector autoregression)
- VARMA (Vector autoregression moving average)
- SES (Simple exponential smoothing)

The following are some of the key steps which needed to be done for training the AR model:

- Plot the time-series
- Check the stationarity
- Determine the parameter  $p$  or order of the AR model
- Train the model

A stationary time series a variable at time 't' is a linear function of prior observations or residual errors. Hence it is time for us to combine the two and have an Auto regressive moving average (ARMA) model.