

Title: Download the Iris flower dataset, compute and summarize statistics for features of dutaset Perform data visualization and plot histogram as well as boxplot for the same.

problem statement: Data Visualization III

Download the Iris flower doctaset or any other dataset into a dataframe.

- 1. List down the featurer and their types (e-g numeric, nominal) available in the dataset.
- 2. Create a histogram for each features in the dutaset to illustrate the features the distribution
- 3. create a boxplot for each feature in the dataset.

Objectives: learn classification techniques how to plot histogram and boxplot for given dataset.

Pre-requisites: R- Programming Retudio

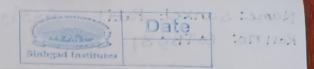
Theory: Let's design a basic data analysis program in Rusing R studio by utilizing the featurer of R studio by utilizing to create some visual representation of their data

How to Install Rstudio?

we have to follow three basic steps in same order to run

Rand R studio on your system.

- · Install R (Download the binary setup file for R from the following link)
- · Install Rstudio
- · Install R Packages.



following steps will be performed to achieve your goal.

- · Downloading/importing data in R
- · Transforming Data / Running queries on data.
- · Basic data analysis using startistical averages.
- · Plotting data distribution

Typical data analysis process

Data analysis deals with collecting, inspecting, cleansing, transforming and moderling data to glean valuable, insights and support better decision in an organization.

Data Exploration: - Having Indentified business problem, a data analyst has to go through the data provided by client to analyse the root cause of the problem.

Data preparation: This is the most crucial step of the data analysis process wherein any data anomalies. with the data have to be modered in the right direction.

+ Data Modelling: -

modelling step begins once the data how been prepared modelling is an iterative process wherein the model is not repeatedly for improvement.

2. Validation-

In this step, the model provided by the client and the model developed by the data analyst are validated against each other to find out if the developed model mutthe business requirement.

3. Implementation of the model and tracking.

model is implemented in production and is tested for accurrent and efficient.



1. Importing Data in R studio:-

data set Acs. There are two ways to import this duta in R. one way is to import the data programatically by executing the following command in the data console window of Rstudie

First download it is into you local computer and use the imports dataset featurer of R studio.

I click on the import dataset button in the top-tight section under the environment tab. select the file you want to import and then click open.

2. After setting up the preferences of seperator, name and other parameters click on the import button.

2. Transforming Data.

you can use various transformation features of R to manipulate the data.

· To Access a particular column, Exage husband inourcare.
· To run some querier on data, you can use the subset
function of R. Let's say I want those nows from the
dataset in which the age husband is greater than
age-wife.

The first parameters to the subset function is the dotaframe you want to apply that function to and the second parameter is the boolean condition that needs to be checked for each row to be included or not.



Getting stabistical Averager from data:

Following functions can be used to conculate the averages

- · For mean of any column, run: mean (acspage husband)
- · Median, run: median (acstage husband)
- · Quntite, run: quantile (acstage-wife)
- · variance, tun: var (acs dage wife)

4. Plotting Data

for R. Any data set imported in R can visualized using the plot and several other functions of R.

where S is the subset of the original dataset and typeop's set the plot type as point you can also choose the and other change variables to Lebe.

To draw a Histogram of dutaset, you can run the command nist (acs & number_children)

Similarly for bar plots, tunthe following Set of command.

Conclusion:

thus we have learned how to use features of R studio when learning data visualization. How to create histogram for each features of dato set, boxplot etc