

Install R and RStudio for the programming language R and the environment

R is an open source language / tool for

1. Data analysis
2. Stat computing
3. visualizations

Originally used academicians and being adopted in industry also

R consists of

1. R language
2. Run time environment

Different tasks that can be carried out with R

1. Data analysis using both structured and unstructured data
  - a. Dplyr package for data manipulations & cleaning
  - b. Stringr package for unstructured text data
  - c. multiDplyr for big datasets
2. Stat modelling- different statistiscal measures are
  - a. Central tendencies like mean, median, mode
  - b. Measures of dispersion like range, variance, standard deviation
  - c. Measures of position like percentiles, quartiles, z-score
  - d. Measures of association like co-relation, regression coefficient
  - e. Shape of distribution like skewness and kurtosis
3. Visualization packages- ggplot2, plotly, grid, lattice, shiny for interactive visualizations
4. Interfaces for databases- access data from databases
  - a. Roracle for oracle database
  - b. RMySQL for interfacing with MySQL
  - c. Spreadsheets, NoSQL databases
5. R is cross platform compatible- you R version for windows, Linux, MacOS
6. ML packages: mlr, caret, randomforest for supervised and unsupervised ML models
7. Report generation- integrate visualization with summary to create report

Number of popular companies use R

1. Google
2. Facebook
3. IBM
4. uber