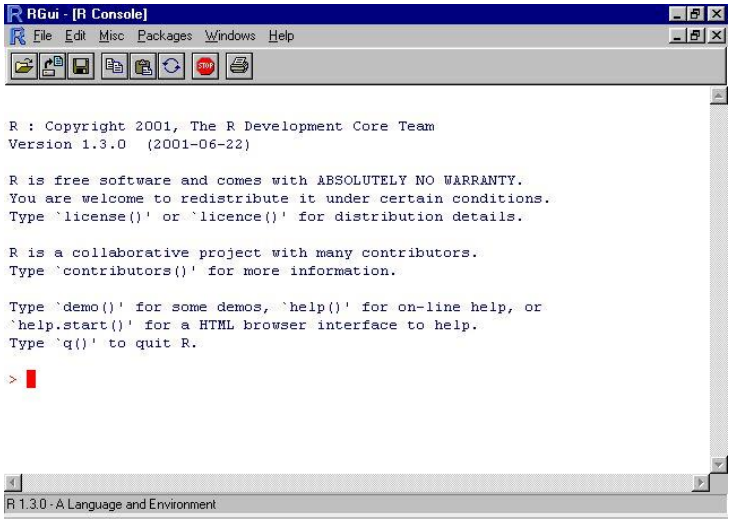
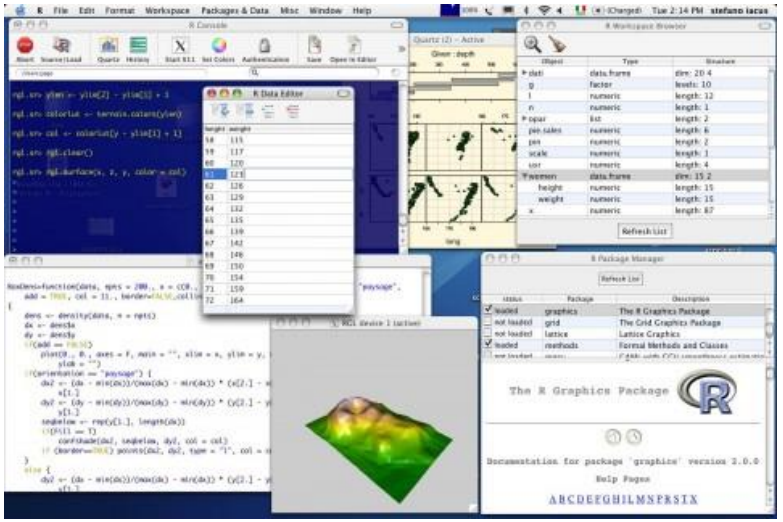
R Overview

Pilsung Kang
School of Industrial Management Engineering
Korea University

History of R





- R is an open source programming language and software environment for statistical computing and graphics
  - ✓ Created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team
  - ✓ Highly extensible through the use of user-submitted packages for specific functions or specific areas of study

The R Language

- R is an expression-based language
  - ✓ User type language expressions at the R prompt
  - ✓ These expressions are evaluated by the R interpreter
  - ✓ The computed values of the expressions are printed
- R is extensile
  - ✓ Users can implement new functionality in the form of functions
  - ✓ Developers can implement new packages of functionality that extends the base system

Why R

- Advantage of R Programming (https://data-flair.training/blogs/pros-and-cons-of-r-programming-language/)

  ✓ Open Source

  - R is an open-source programming language. This means that anyone can work with R without any need for a license or a fee. Furthermore, you can contribute towards the development of R by *customizing its packages, developing new ones and resolving issues*.

  ✓ Exemplary Support for Data Wrangling

  - R provides exemplary support for data wrangling. The packages like *dplyr, readr* are capable of transforming messy data into a structured form.

  ✓ The Array of Packages

  - R has a vast array of packages. With over 10,000 packages in the CRAN repository, the number is constantly growing. These packages appeal to all the areas of industry.

  ✓ Quality Plotting and Graphing

  - R facilitates quality plotting and graphing. The popular libraries like *ggplot2* and *plotly* advocate for aesthetic and visually appealing graphs that set R apart from other programming languages.

Why R

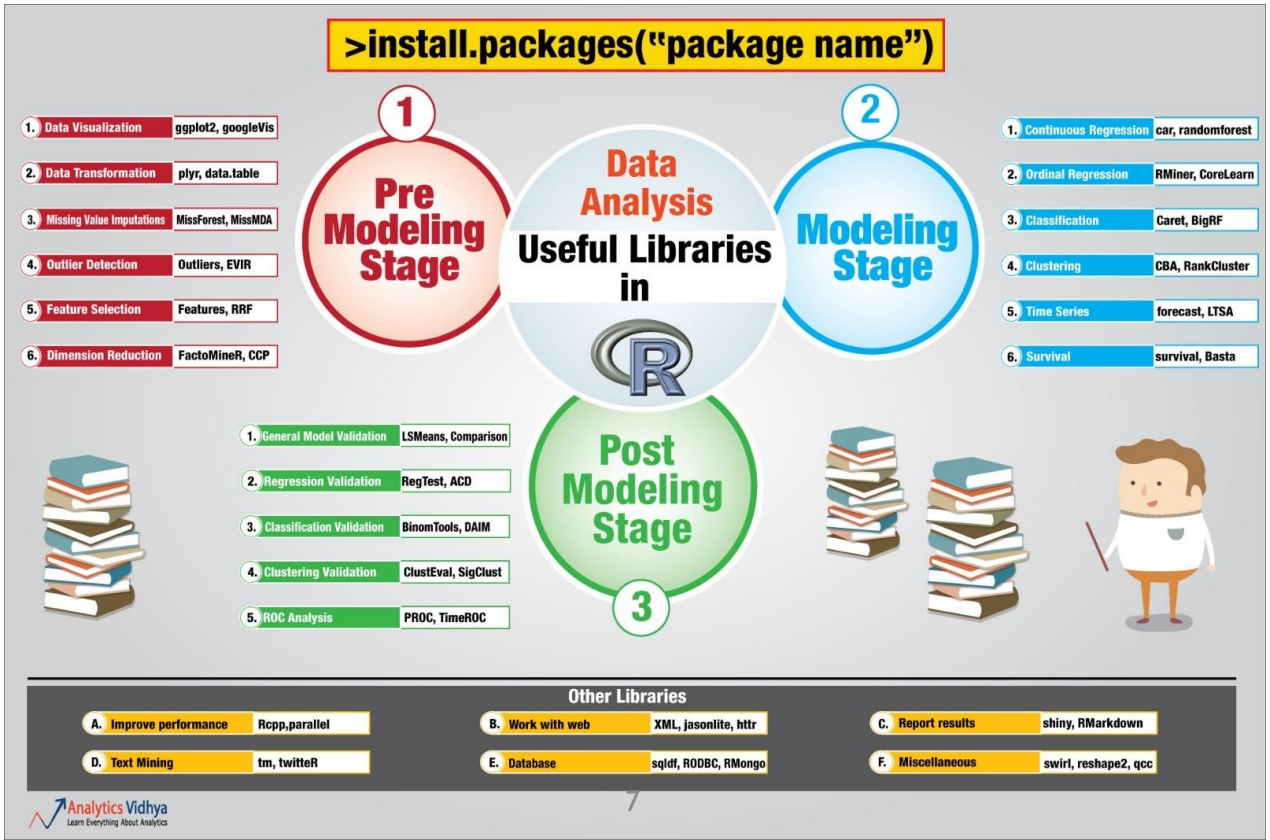
- Advantage of R Programming (https://data-flair.training/blogs/pros-and-cons-of-r-programming-language/)
  - ✓ Highly Compatible
    - R is highly compatible and can be paired with many other programming languages like C, C++, Java, and Python. It can also be integrated with technologies like Hadoop and various other database management systems as well.
  - ✓ Platform Independent
    - R is a platform-independent language. It is a cross-platform programming language, meaning that it can be run quite easily on Windows, Linux, and Mac.
  - ✓ Eye-Catching Reports
    - With packages like Shiny and Markdown, reporting the results of an analysis is extremely easy with R. You can make reports with the data, plots and R scripts embedded in them. You can even make interactive web apps that allow the user to play with the results and the data.

Why R

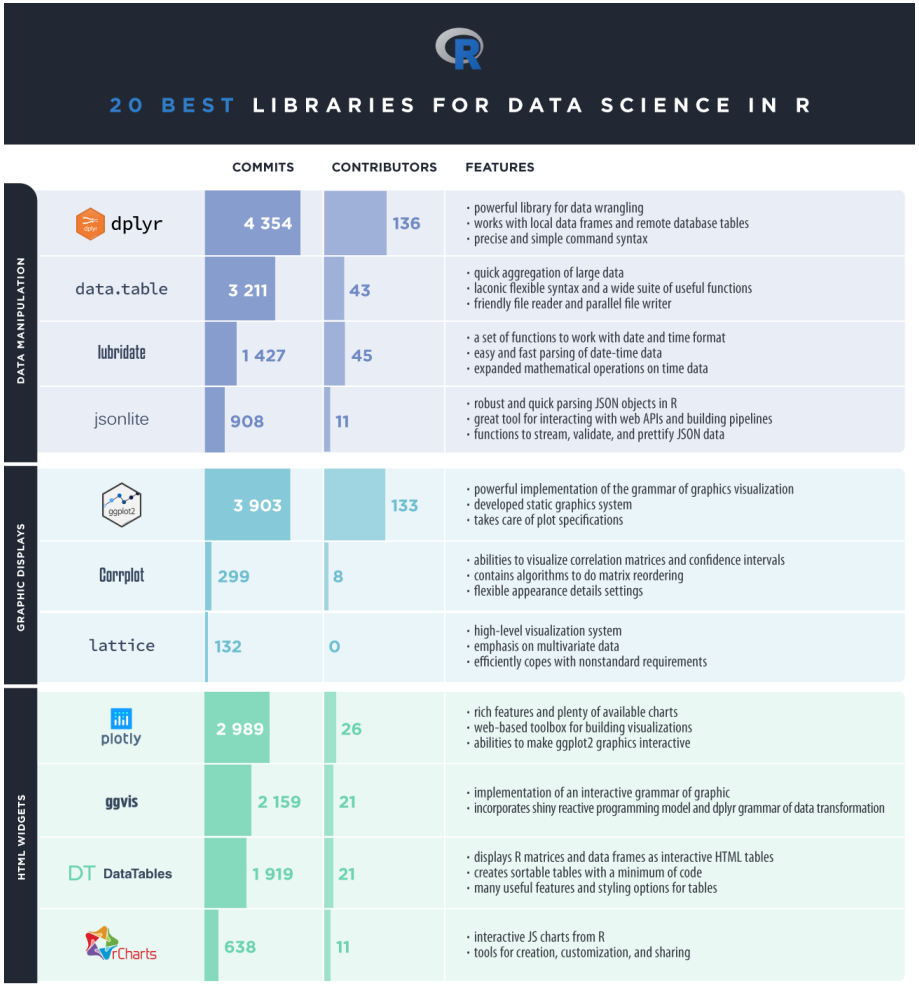
- Advantage of R Programming (https://data-flair.training/blogs/pros-and-cons-of-r-programming-language/)
  - ✓ Machine Learning Operations
    - R provides various facilities for carrying out machine learning operations like *classification, regression and also provides features for developing artificial neural networks.*
  - ✓ Statistics
    - R is prominently known as the lingua franca of statistics. This is the main reason as to why R is dominant among other programming languages for developing statistical tools.
  - ✓ Continuously Growing
    - R is a constantly evolving programming language. It is a state-of-the-art technology that provides updates whenever any new feature is added.

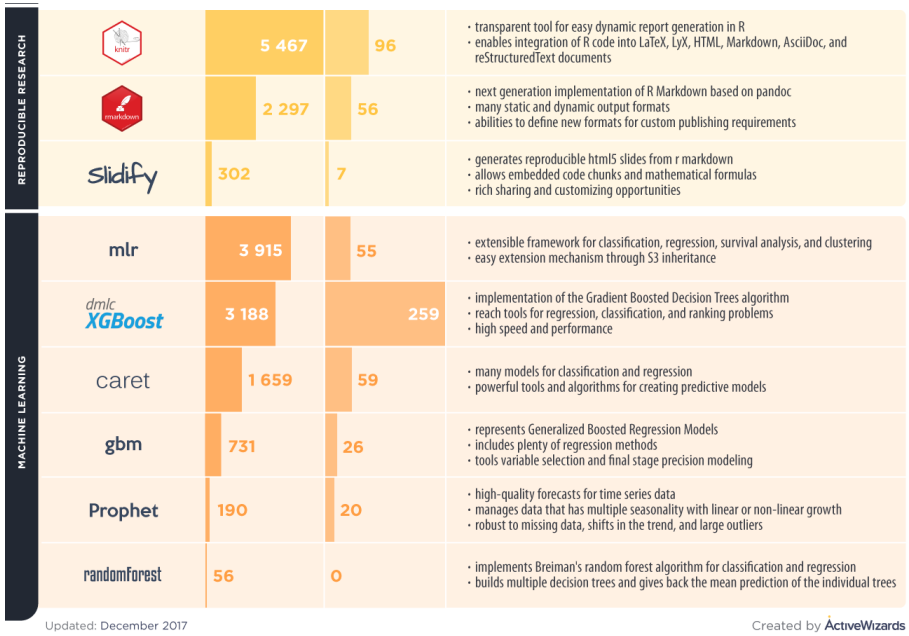
Why R?

- R packages (https://www.analyticsvidhya.com/blog/2017/03/create-packages-r-cran-github/)
  - ✓ Everyone can make, distribute, and use R packages
  - ✓ Almost every statistical analysis can be possible
  - ✓ 14,791 packages are available (2019. 09. 02) → 16,159 packages in 2020. 08. 24



Popular Packages for Data Science





https://medium.com/activewizards-machine-learning-company/top-20-r-libraries-for-data-science-in-2018-infographic-956f8419f883


ANY
questions ?