



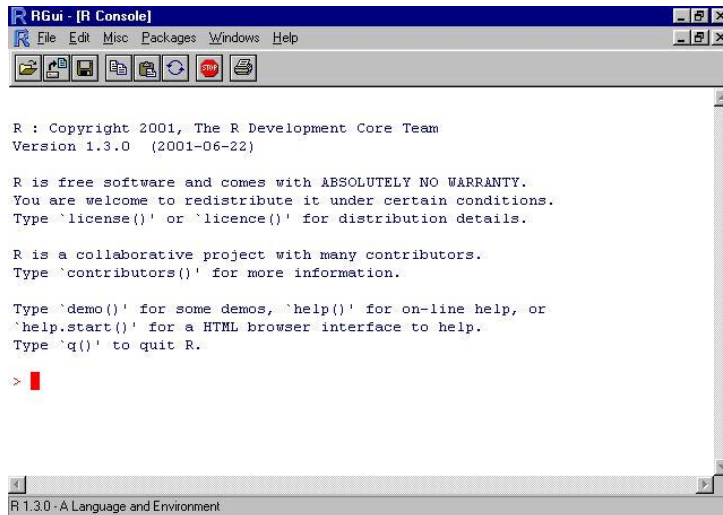
# R Overview

Pilsung Kang

School of Industrial Management Engineering

Korea University

# History of



```
RGui - [R Console]
R File Edit Misc Packages Windows Help

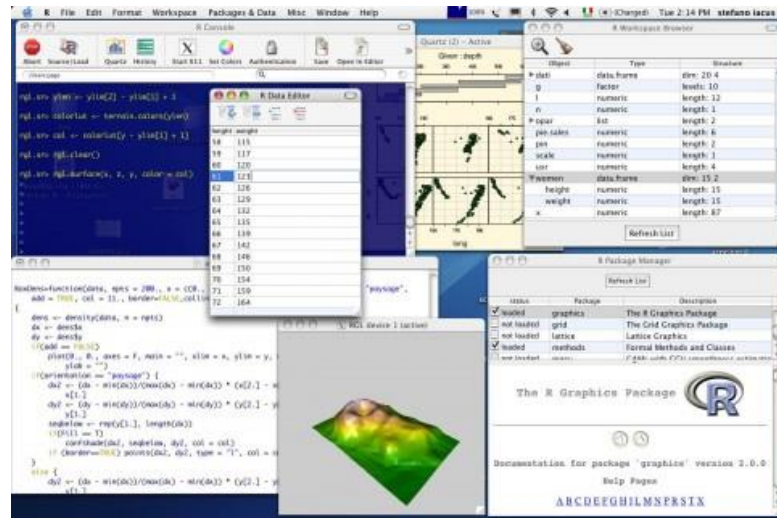
R : Copyright 2001, The R Development Core Team
Version 1.3.0 (2001-06-22)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit 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 extensible
  - ✓ 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

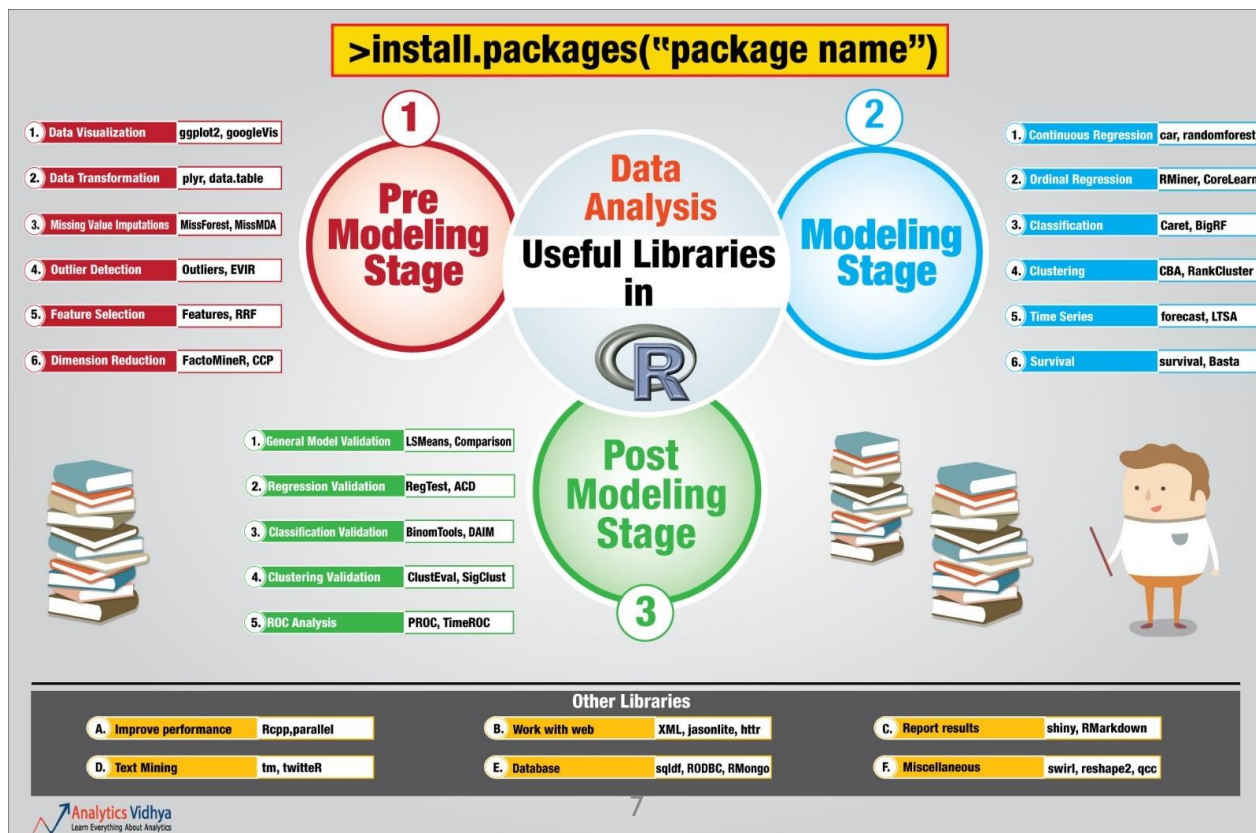
- With packages like Shiny and Markdown, reporting the results of an analysis is extremely easy with R. You can make reports with the data, plots and R scripts embedded in them. You can even make interactive web apps that allow the user to play with the results and the data.

# 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



## 20 BEST LIBRARIES FOR DATA SCIENCE IN R

COMMITTS CONTRIBUTORS FEATURES

DATA MANIPULATION	dplyr	4 354	136	<ul style="list-style-type: none"> <li>powerful library for data wrangling</li> <li>works with local data frames and remote database tables</li> <li>precise and simple command syntax</li> </ul>
	data.table	3 211	43	<ul style="list-style-type: none"> <li>quick aggregation of large data</li> <li>laconic flexible syntax and a wide suite of useful functions</li> <li>friendly file reader and parallel file writer</li> </ul>
	lubridate	1 427	45	<ul style="list-style-type: none"> <li>a set of functions to work with date and time format</li> <li>easy and fast parsing of date-time data</li> <li>expanded mathematical operations on time data</li> </ul>
	jsonlite	908	11	<ul style="list-style-type: none"> <li>robust and quick parsing JSON objects in R</li> <li>great tool for interacting with web APIs and building pipelines</li> <li>functions to stream, validate, and prettify JSON data</li> </ul>
GRAPHIC DISPLAYS	ggplot2	3 903	133	<ul style="list-style-type: none"> <li>powerful implementation of the grammar of graphics visualization</li> <li>developed static graphics system</li> <li>takes care of plot specifications</li> </ul>
	Corrplot	299	8	<ul style="list-style-type: none"> <li>abilities to visualize correlation matrices and confidence intervals</li> <li>contains algorithms to do matrix reordering</li> <li>flexible appearance details settings</li> </ul>
	lattice	132	0	<ul style="list-style-type: none"> <li>high-level visualization system</li> <li>emphasis on multivariate data</li> <li>efficiently copes with nonstandard requirements</li> </ul>
HTML WIDGETS	plotly	2 989	26	<ul style="list-style-type: none"> <li>rich features and plenty of available charts</li> <li>web-based toolbox for building visualizations</li> <li>abilities to make ggplot2 graphics interactive</li> </ul>
	ggvis	2 159	21	<ul style="list-style-type: none"> <li>implementation of an interactive grammar of graphic</li> <li>incorporates shiny reactive programming model and dplyr grammar of data transformation</li> </ul>
	DT DataTables	1 919	21	<ul style="list-style-type: none"> <li>displays R matrices and data frames as interactive HTML tables</li> <li>creates sortable tables with a minimum of code</li> <li>many useful features and styling options for tables</li> </ul>
	rCharts	638	11	<ul style="list-style-type: none"> <li>interactive JS charts from R</li> <li>tools for creation, customization, and sharing</li> </ul>

REPRODUCIBLE RESEARCH	R Markdown	5 467	96	<ul style="list-style-type: none"> <li>transparent tool for easy dynamic report generation in R</li> <li>enables integration of R code into LaTeX, LyX, HTML, Markdown, AsciiDoc, and reStructuredText documents</li> </ul>
	rmarkdown	2 297	56	<ul style="list-style-type: none"> <li>next generation implementation of R Markdown based on pandoc</li> <li>many static and dynamic output formats</li> <li>abilities to define new formats for custom publishing requirements</li> </ul>
	slidify	302	7	<ul style="list-style-type: none"> <li>generates reproducible html5 slides from r markdown</li> <li>allows embedded code chunks and mathematical formulas</li> <li>rich sharing and customizing opportunities</li> </ul>
MACHINE LEARNING	mlr	3 915	55	<ul style="list-style-type: none"> <li>extensible framework for classification, regression, survival analysis, and clustering</li> <li>easy extension mechanism through S3 inheritance</li> </ul>
	dmlc XGBoost	3 188	259	<ul style="list-style-type: none"> <li>implementation of the Gradient Boosted Decision Trees algorithm</li> <li>reach tools for regression, classification, and ranking problems</li> <li>high speed and performance</li> </ul>
	caret	1 659	59	<ul style="list-style-type: none"> <li>many models for classification and regression</li> <li>powerful tools and algorithms for creating predictive models</li> </ul>
	gbm	731	26	<ul style="list-style-type: none"> <li>represents Generalized Boosted Regression Models</li> <li>includes plenty of regression methods</li> <li>tools variable selection and final stage precision modeling</li> </ul>
	Prophet	190	20	<ul style="list-style-type: none"> <li>high-quality forecasts for time series data</li> <li>manages data that has multiple seasonality with linear or non-linear growth</li> <li>robust to missing data, shifts in the trend, and large outliers</li> </ul>
	randomforest	56	0	<ul style="list-style-type: none"> <li>implements Breiman's random forest algorithm for classification and regression</li> <li>builds multiple decision trees and gives back the mean prediction of the individual trees</li> </ul>

Updated: December 2017

Created by ActiveWizards

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



