

2. R 이해하기

I R

Why R ?

- 오픈 소스로 무료 (Open-Source) : FREE!
- 수많은 다양한 패키지(Packages) : R 공식 배포인 CRAN(open-source repository)을 통해 약14000개 (2019)
- Data Scientists가 가장 많이 사용하는 programming languages 중에 Python 과 더불어 최상위
- 학계나 과학자들을 통해 개발 Statistical 방법론에 특화 (예: 패널데이터 분석, 시계열 분석 등에 다양한 패키지)
- 다양한 커뮤니티를 통해 지원

R. Community

- R-bloggers : <https://www.r-bloggers.com>
- The R Consortium : <https://www.r-consortium.org/projects>
- #rstats hashtag : <https://twitter.com/search?q=%23rstats>
- R-Ladies : <https://rladies.org>
- Local R meetup groups : <https://jumpingrivers.github.io/meetingsR/r-user-groups.html>
- Rweekly : <https://rweekly.org>
- DataCarpentry(<http://www.datacarpentry.org>) and Software Carpentry(<https://software-carpentry.org>)
- R Conferences : <https://jumpingrivers.github.io/meetingsR/events.html>
- Github : <https://github.com/trending/developers/r?since=weekly>

etc

R 환경설정

다운로드 R 설치 파일 (R 설치 파일 배포 : CRAN)



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2019-07-05, Action of the Toes) [R-3.6.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

R for Windows

Subdirectories:

- [base](#) Binaries for base distribution. This is what you want to [install R for the first time](#).
- [contrib](#) Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
- [old contrib](#) Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
- [Rtools](#) Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.



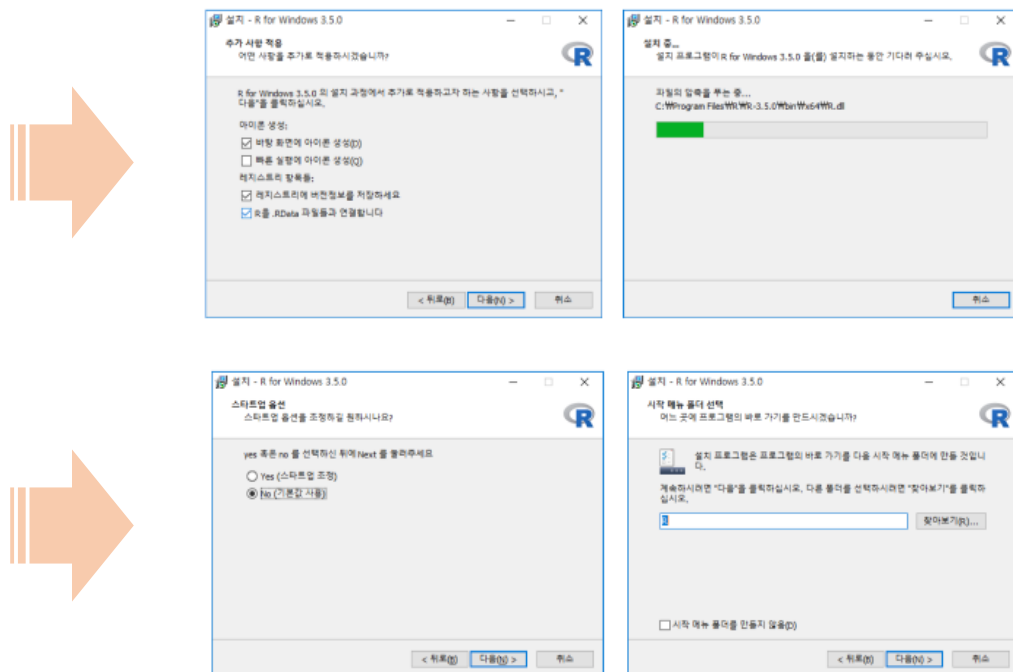
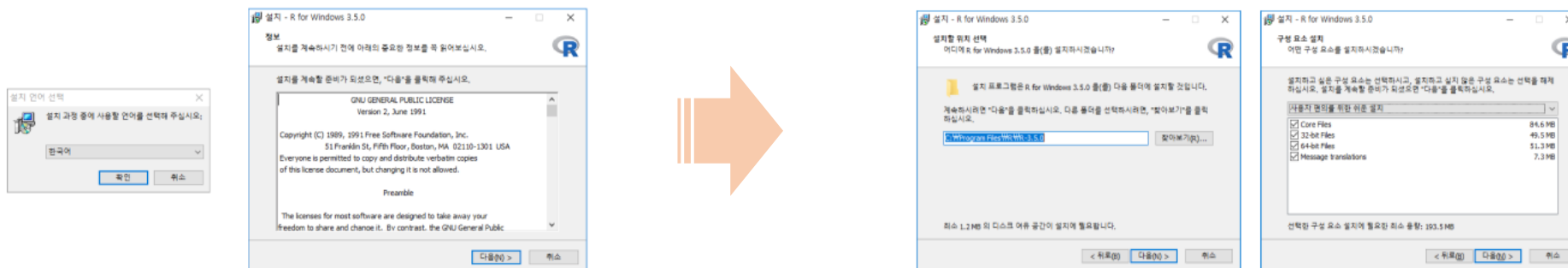
R-3.6.1 for Windows (32/64 bit)

[Download R 3.6.1 for Windows](#) (81 megabytes, 32/64 bit)

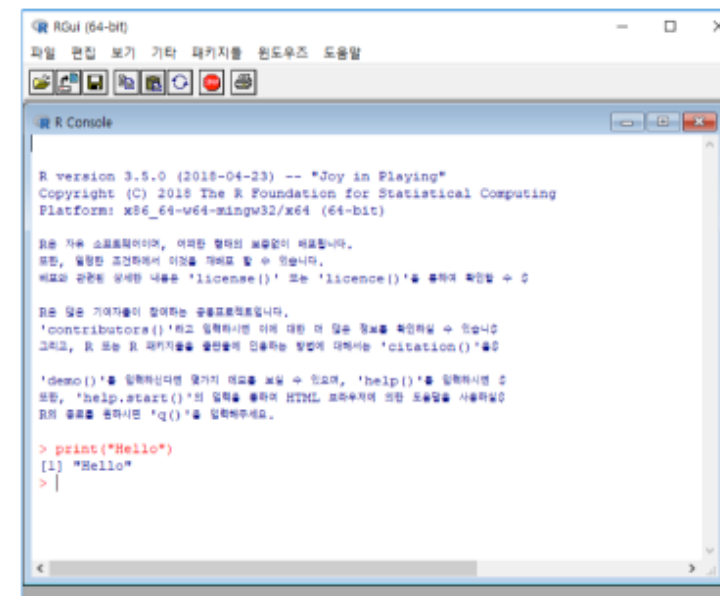
[Installation and other instructions](#)
[New features in this version](#)

R 환경설정

R 설치 (OS: window)



R 설치완료 (실행)



R 환경설정

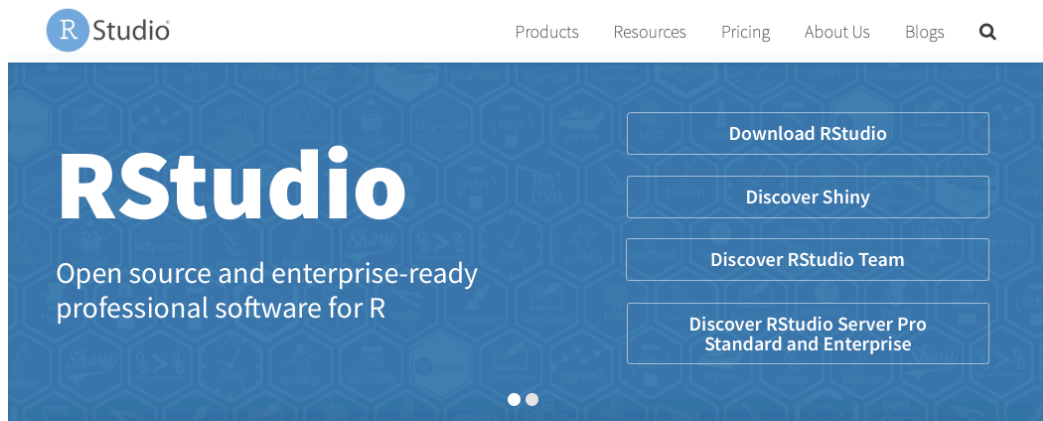
통합 개발 환경IDE(Integrated Development Environment)

RStudio : <https://www.rstudio.com/products/RStudio/>

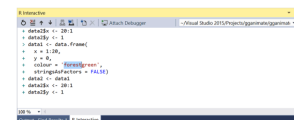
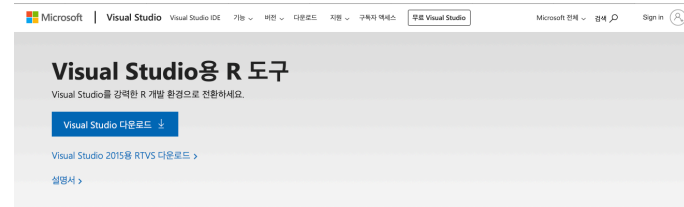
Visual Studio용 R : <https://docs.microsoft.com/ko-kr/visualstudio/rtps/installing-r-tools-for-visual-studio>

jupyter notebook R kernel : <https://irkernel.github.io/requirements/>

RStudio



R Tools for Visual Studio



R 대화형 창

생각하는 속도로 프로토타입

대화형 창에서 바로 코드를 실행하거나, 편집기에서 Ctrl+Enter를 누르거나, 기록 창에서 계산 결과 또는 그래프를 빠르게 볼 수 있습니다. 위쪽 확장을 키를 눌러서 편집기에서처럼 한 줄 또는 여러 줄의 코드를 편집할 수 있습니다.

Introducing RStudio Team

RStudio's recommended professional data science solutions for every team. RStudio Team includes:



RStudio Server Pro

Take control of your R code



RStudio Connect

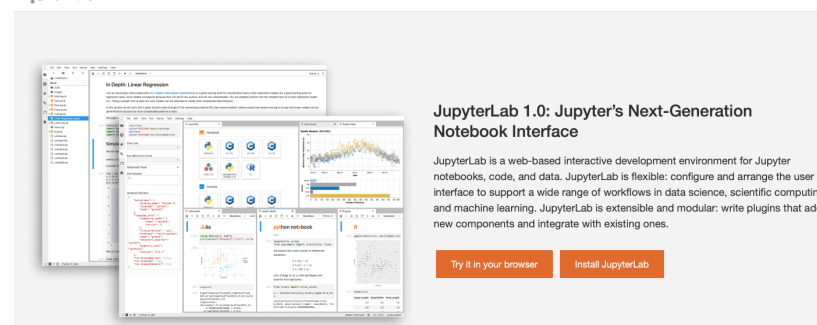
Connect data scientists with decision makers



RStudio Package Manager

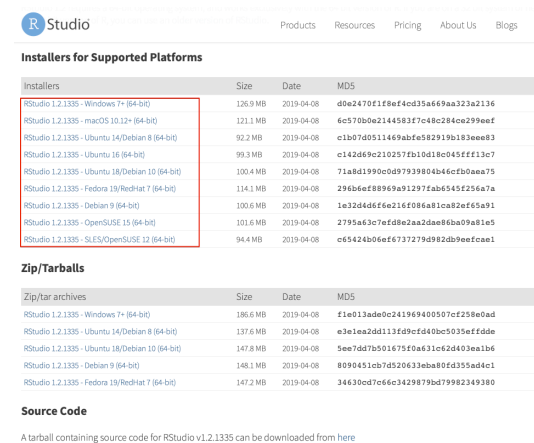
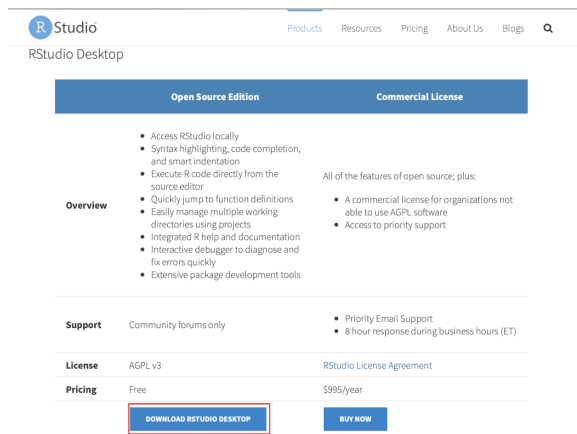
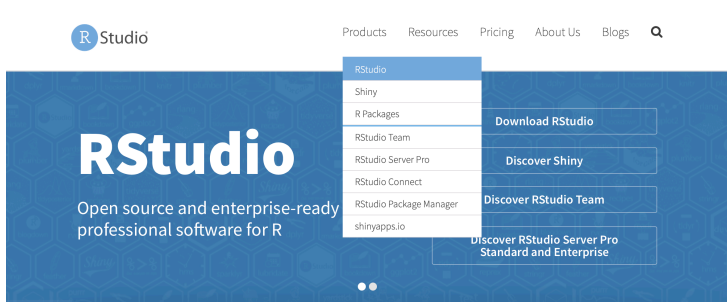
Share and manage R packages

R Kernel for Jupyter Notebook



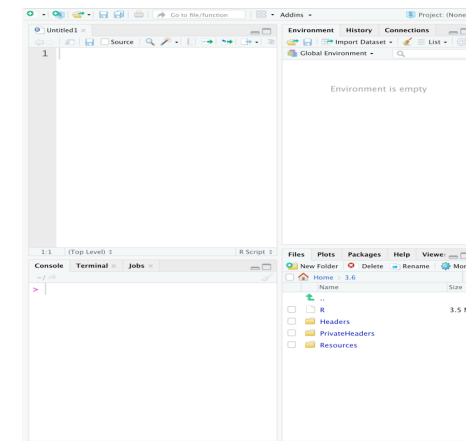
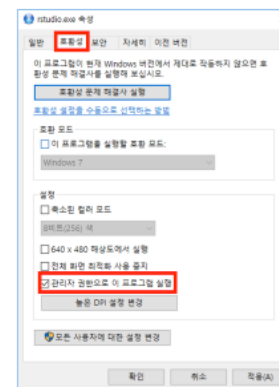
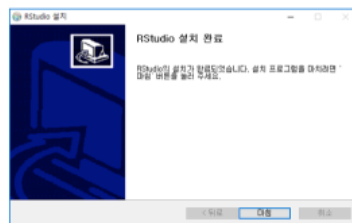
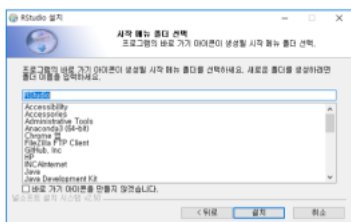
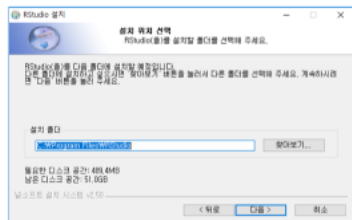
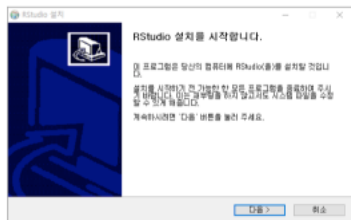
R 환경설정

RStudio 설치 (IDE)

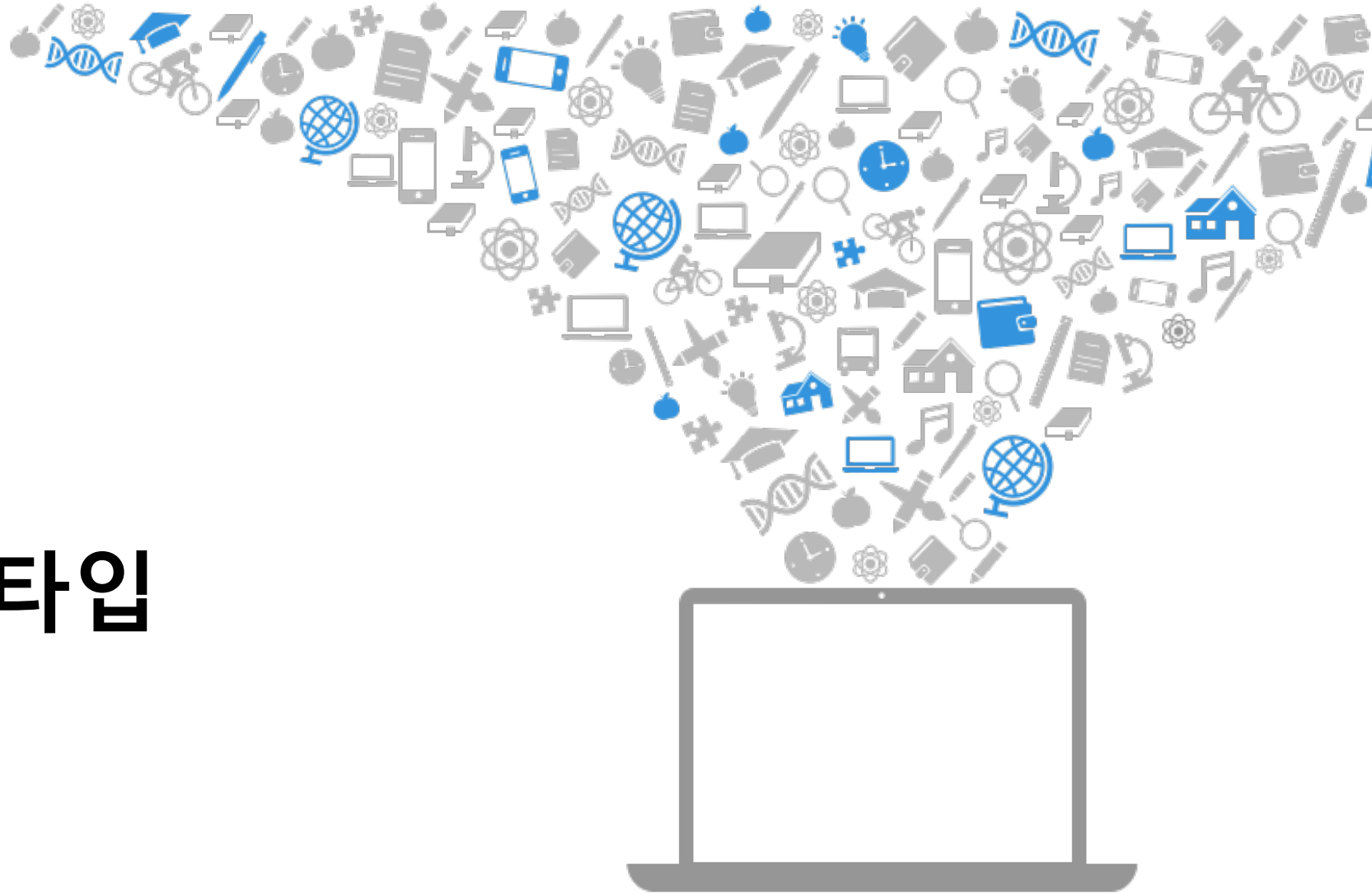


RStudio 관리자 실행으로 설정

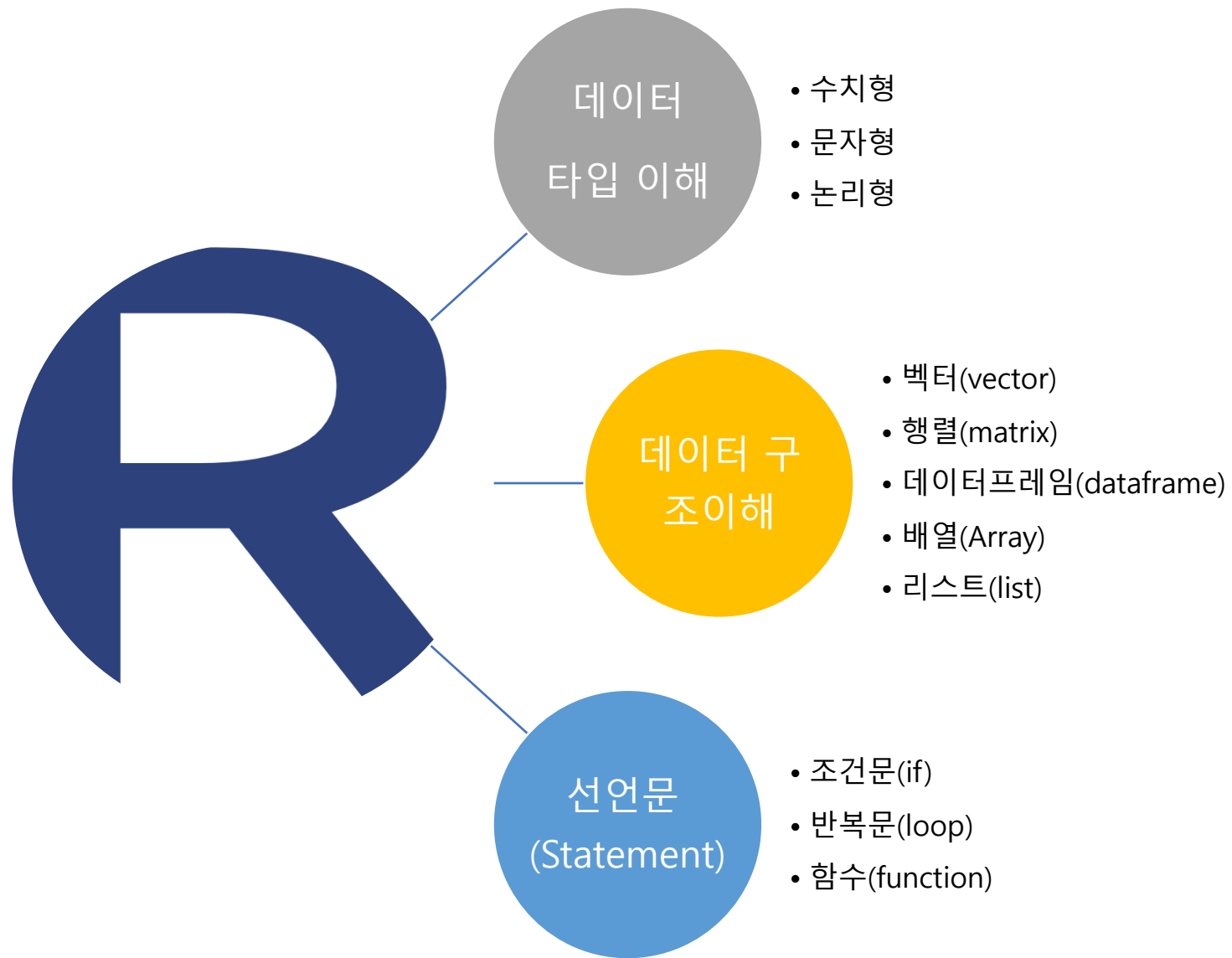
RStudio 실행



2.2 R 데이터 타입



I R 기본적 이해



I R 데이터 타입

변수, 객체(object) : 할당하여 생성

산술연산자 : + , -, *, /, %/%, %% , ^
논리연산자 : ==, !=, >=, <=, >, <

Rule of R language

- 첫 글자 반드시 문자
- 대소문자 구분
- 할당연산자 : <-, ->, =, 변수를 생성할 때 사용

자료구조

수치형

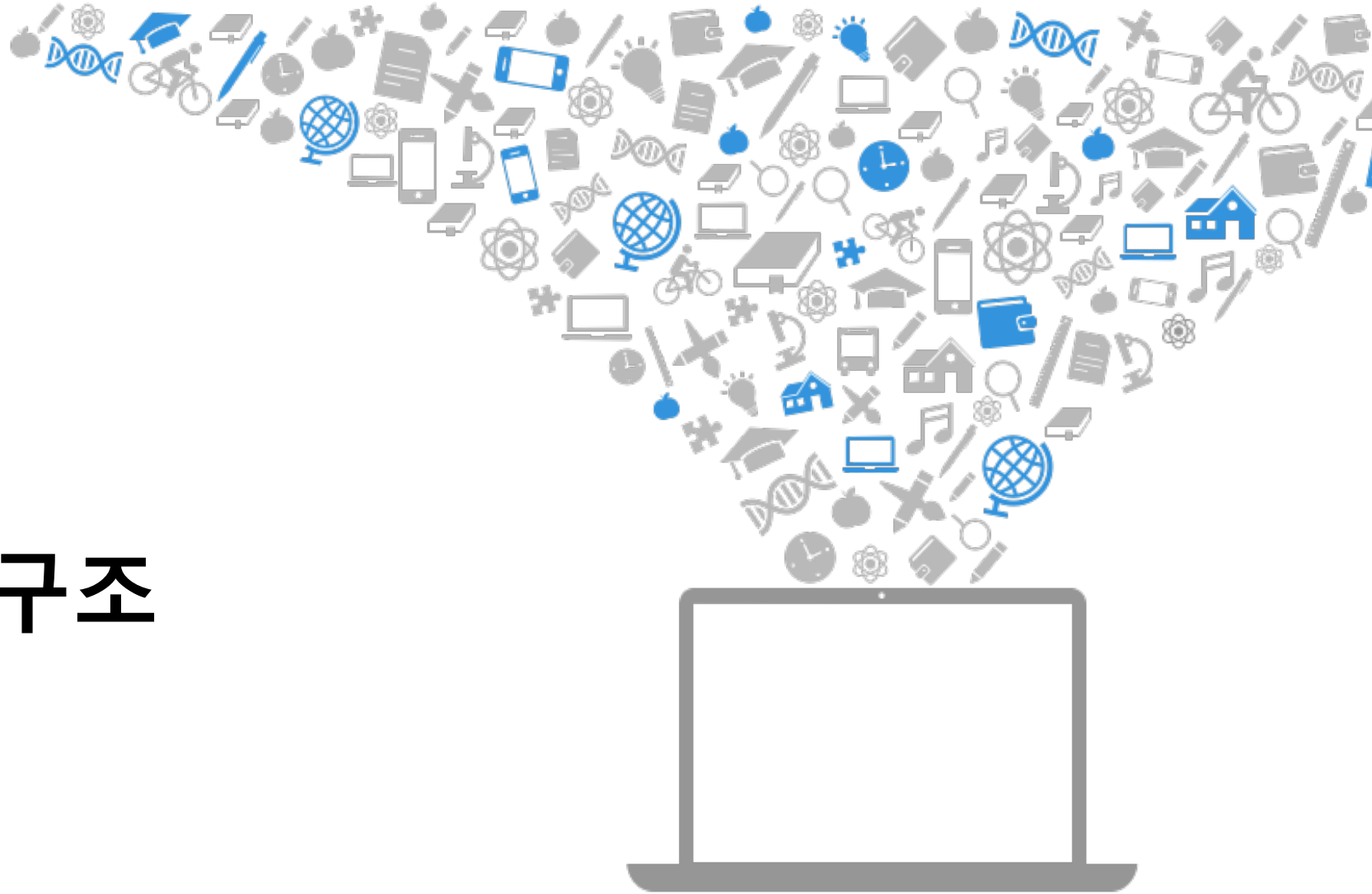
- 실수(numeric) : 1.3
- 정수(integer) : 3, 3L,
- 복소수(complex) : 2i

문자형

- 문자열(character) : "string"
- 범주형(factor) : levels(male, female)
- 날짜(date) : "2019-09-23", format ('y-m-d')
- 시간(time) : format = (h:m:s)

논리값 : TRUE, FALSE, T, F

2.3 R 데이터 구조



R 데이터 구조

스칼라(Scala)

벡터(vector)

- 생성
- indexing(slicing)
- 연산

행렬(Matrix)

- 생성
- indexing(slicing)
- 연산

배열(Array)

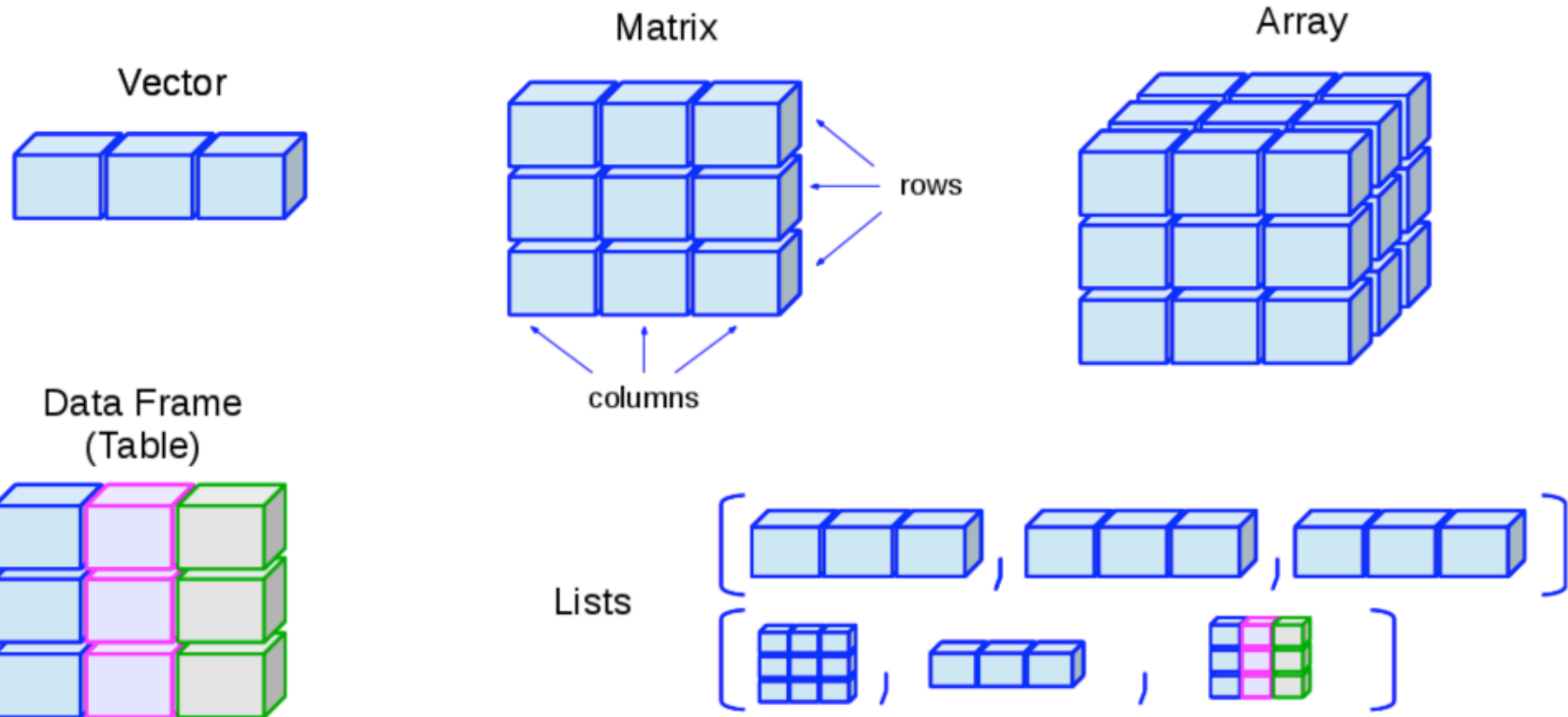
- 생성
- indexing(slicing)
- 연산

데이터 프레임(data frame)

- 생성
- indexing(slicing)
- 연산

리스트(list)

- 생성
- indexing(slicing)
- 연산



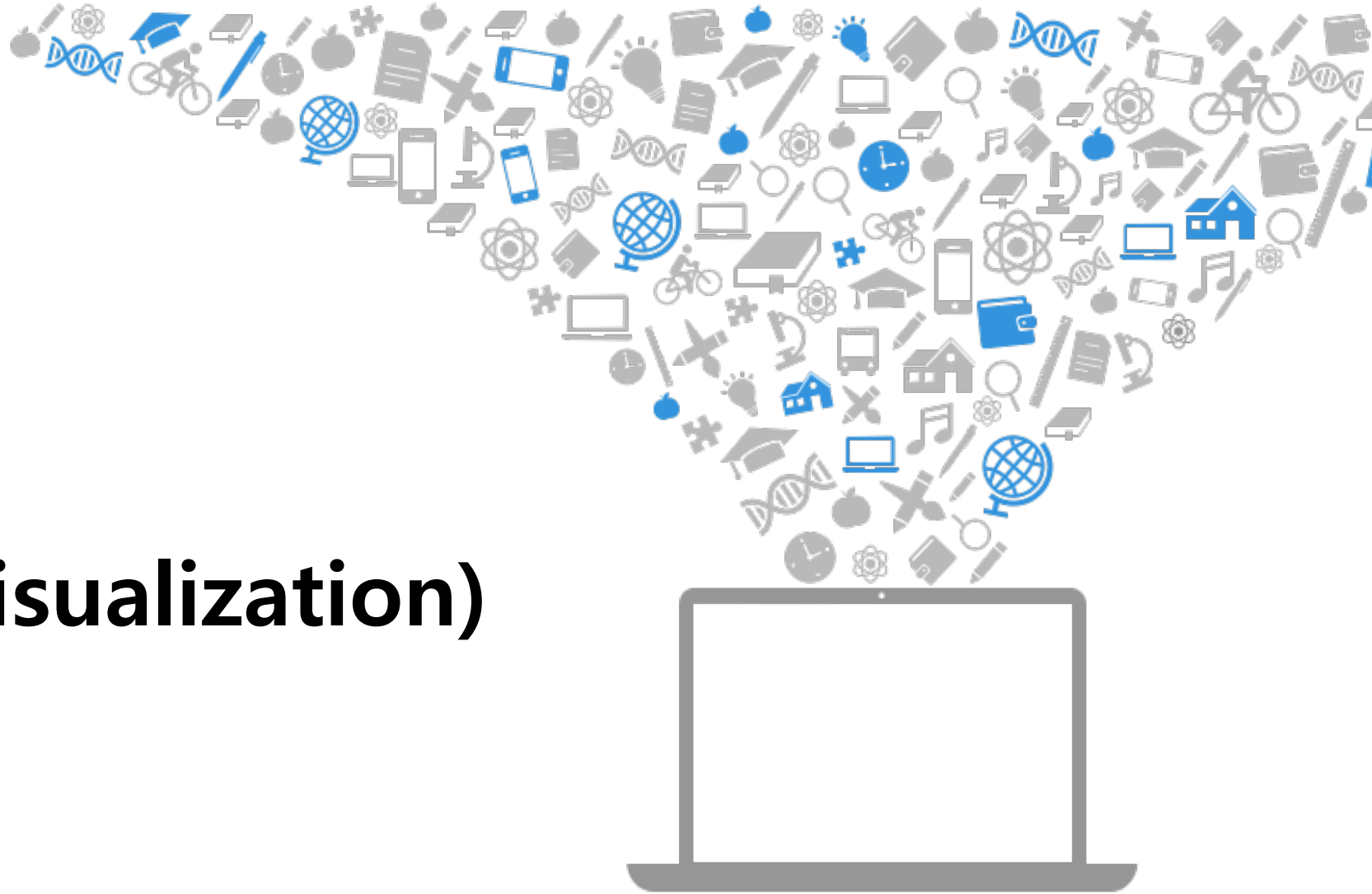
source: <http://venus.ifca.unican.es/Rintro/dataStruct.html>

2.4

**조건문(if),
반복문(loops),
함수(function)**







2.5 시각화 (Visualization)



R 주요 패키지 예시)






20 BEST LIBRARIES FOR DATA SCIENCE IN R

		COMMITTS	CONTRIBUTORS	FEATURES
DATA MANIPULATION	 dplyr	4 354	136	<ul style="list-style-type: none">powerful library for data wranglingworks with local data frames and remote database tablesprecise and simple command syntax
	data.table	3 211	43	<ul style="list-style-type: none">quick aggregation of large datalaconic flexible syntax and a wide suite of useful functionsfriendly file reader and parallel file writer
	lubridate	1 427	45	<ul style="list-style-type: none">a set of functions to work with date and time formateasy and fast parsing of date-time dataexpanded mathematical operations on time data
	jsonlite	908	11	<ul style="list-style-type: none">robust and quick parsing JSON objects in Rgreat tool for interacting with web APIs and building pipelinesfunctions to stream, validate, and prettify JSON data
GRAPHIC DISPLAYS	 ggplot2	3 903	133	<ul style="list-style-type: none">powerful implementation of the grammar of graphics visualizationdeveloped static graphics systemtakes care of plot specifications
	Corrplot	299	8	<ul style="list-style-type: none">abilities to visualize correlation matrices and confidence intervalscontains algorithms to do matrix reorderingflexible appearance details settings
	lattice	132	0	<ul style="list-style-type: none">high-level visualization systememphasis on multivariate dataefficiently copes with nonstandard requirements
HTML WIDGETS	 plotly	2 989	26	<ul style="list-style-type: none">rich features and plenty of available chartsweb-based toolbox for building visualizationsabilities to make ggplot2 graphics interactive
	ggvis	2 159	21	<ul style="list-style-type: none">implementation of an interactive grammar of graphicincorporates shiny reactive programming model and dplyr grammar of data transformation
	DT DataTables	1 919	21	<ul style="list-style-type: none">displays R matrices and data frames as interactive HTML tablescreates sortable tables with a minimum of codemany useful features and styling options for tables
	 rCharts	638	11	<ul style="list-style-type: none">interactive JS charts from Rtools for creation, customization, and sharing



20 BEST LIBRARIES FOR DATA SCIENCE IN R

		COMMITTS	CONTRIBUTORS	FEATURES
REPRODUCIBLE RESEARCH	 knitr	5 467	96	<ul style="list-style-type: none">transparent tool for easy dynamic report generation in Renables integration of R code into LaTeX, LyX, HTML, Markdown, AsciiDoc, and reStructuredText documents
	 markdown	2 297	56	<ul style="list-style-type: none">next generation implementation of R Markdown based on pandocmany static and dynamic output formatsabilities to define new formats for custom publishing requirements
	slidify	302	7	<ul style="list-style-type: none">generates reproducible html5 slides from r markdownallows embedded code chunks and mathematical formulasrich sharing and customizing opportunities
MACHINE LEARNING	mlr	3 915	55	<ul style="list-style-type: none">extensible framework for classification, regression, survival analysis, and clusteringeasy extension mechanism through S3 inheritance
	 dmlc XGBoost	3 188	259	<ul style="list-style-type: none">implementation of the Gradient Boosted Decision Trees algorithmreach tools for regression, classification, and ranking problemshigh speed and performance
	caret	1 659	59	<ul style="list-style-type: none">many models for classification and regressionpowerful tools and algorithms for creating predictive models
	gbm	731	26	<ul style="list-style-type: none">represents Generalized Boosted Regression Modelsincludes plenty of regression methodstools variable selection and final stage precision modeling
	Prophet	190	20	<ul style="list-style-type: none">high-quality forecasts for time series datamanages data that has multiple seasonality with linear or non-linear growthrobust to missing data, shifts in the trend, and large outliers
	randomForest	56	0	<ul style="list-style-type: none">implements Breiman's random forest algorithm for classification and regressionbuilds multiple decision trees and gives back the mean prediction of the individual trees

Updated: December 2017

Created by  ActiveWizards