

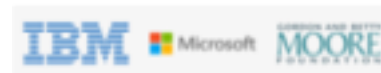
R과 RStudio

*R Programming*을 위한 환경 설정

여정을 위한 준비
: R과 RStudio



= 통계적 컴퓨팅과 그래픽을 위한
자유 소프트웨어 환경



R Project (Language)

: 다운로드 및 설치

<https://www.r-project.org/>

The screenshot shows the R Project website. A red arrow points from the URL above to the 'CRAN' link in the left sidebar. Another red arrow points from the 'Download and Install R' section to the 'Download R for Linux' link in the list of precompiled binaries.

The R Project for Statistical Computing

[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred mirror.

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

News

- [R version 3.5.1 \(Feather Spray\)](#) prerelease versions will appear starting Friday 2018-07-02. release is scheduled for Monday 2018-07-02.
- [R version 3.5.0 \(Joy in Playing\)](#) has been released on 2018-04-23.
- [R version 3.4.4 \(Someone to Lean On\)](#) has been released on 2018-03-15.
- [useR! 2018](#) (July 10 - 13 in Brisbane) is open for registration at <https://user2018.r-project.org>
- [The R Journal Volume 9/2](#) is available.
- [useR! 2017](#) took place July 4 - 7 in Brussels <https://user2017.brussels>
- The [R Logo](#) is available for download in high-resolution PNG or SVG formats.

Korea

- <https://cran.seoul.go.kr/>
- <http://healthstat.snu.ac.kr/CRAN/>
- <https://cran.biodisk.org/>
- <http://cran.biodisk.org/>

Download and Install R

Precompiled binary distributions of the base system. You likely want one of these versions of R:

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

Download Mirror 사이트에서 자신의 운영체제에 맞는 버전 다운로드

RStudio (IDE)

: 다운로드 및 설치

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

The screenshot shows the RStudio website. A red arrow points from the 'Download' link in the RStudio product section to the 'Download' button on the right. Another red arrow points from the 'Download' button to the 'Installers' section, which lists various operating systems and their corresponding RStudio versions. A purple text overlay '자신의 운영체제에 맞는 버전 다운로드' (Download the version that matches your operating system) is placed over the list of installers.

RStudio Desktop
Open Source License

FREE

DOWNLOAD

Learn more

Installers for Supported Platforms

| Installers |
|--|
| RStudio 1.1.453 - Windows Vista/7/8/10 |
| RStudio 1.1.453 - Mac OS X 10.6+ (64-bit) |
| RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (32-bit) |
| RStudio 1.1.453 - Ubuntu 12.04-15.10/Debian 8 (64-bit) |
| RStudio 1.1.453 - Ubuntu 16.04+/Debian 9+ (64-bit) |
| RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit) |
| RStudio 1.1.453 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit) |

자신의 운영체제에 맞는 버전 다운로드

Trouble Shooting

- ▶ *R* 혹은 *RStudio*의 프로젝트가 실행되는 경로상에는 한글이 있어서는 안됩니다.
 - ▶ 주로 윈도 환경에서 사용자의 계정명이 한글로 되어 있을 경우 이 문제가 일어납니다
- ▶ 윈도 환경의 영우, 패키지 인스톨 등 일부 기능은 관리자 권한으로 실행되어야 합니다.
 - ▶ 만약 이 문제가 있을 경우, 다음과 같이 해결합니다.
 - ▶ *R* 스튜디오 바로가기 아이콘 우클릭 > **[속성]** > 호환성 > 관리자 권한으로 이 프로그램 실행

RStudio

: 간단히 살펴보기

환경창: 분석 과정에서 생성한 데이터를 보여주는 창

소스창: 명령이나 메모를 자유롭게 기록

히스토리: 지금까지 입력, 실행한 내용들이 모여 있는 창

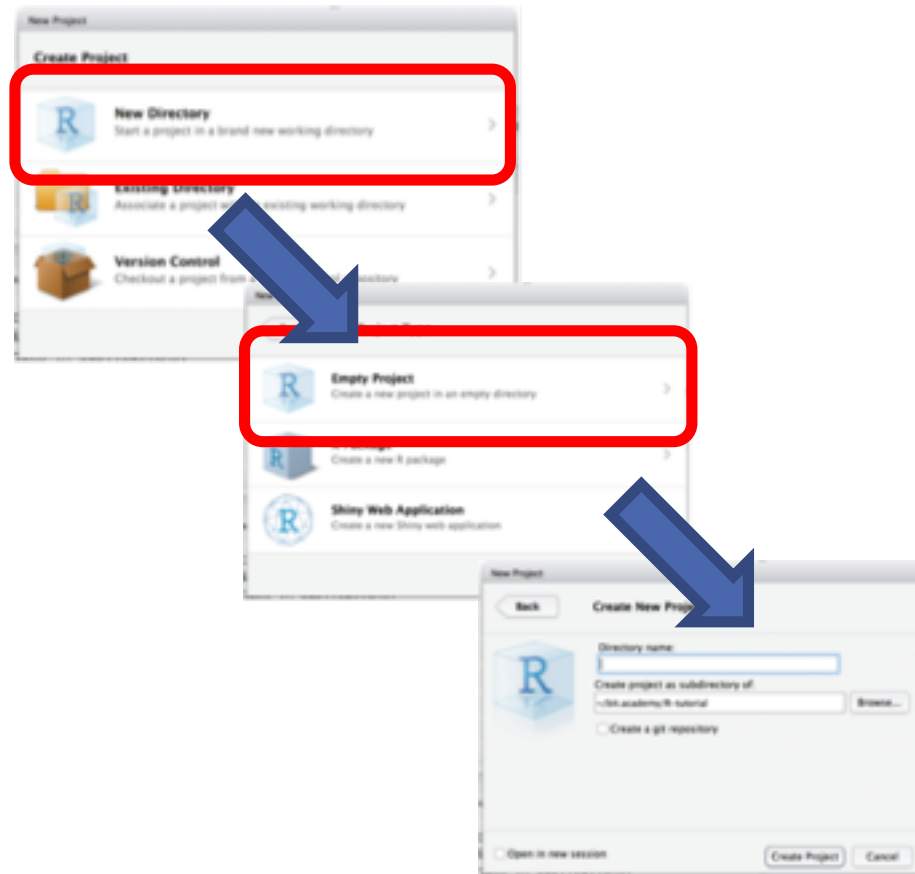
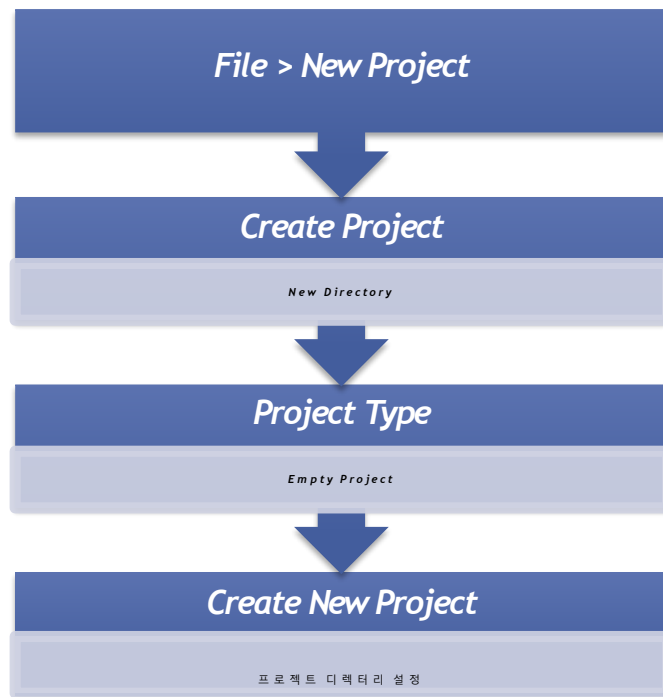


콘솔: 프롬프트에 명령어를 입력하고 실행하면 결과 출력

| 탭 | 기능 |
|-----------------|-----------------|
| Files | 워킹 디렉토리 내 파일 목록 |
| Plots | 그래프를 출력 |
| Packages | 설치된 패키지 목록을 보여줌 |
| Help | 도움말을 보여줌 |
| Viewer | 분석 결과를 웹문서로 출력 |

RStudio

: 프로젝트 생성

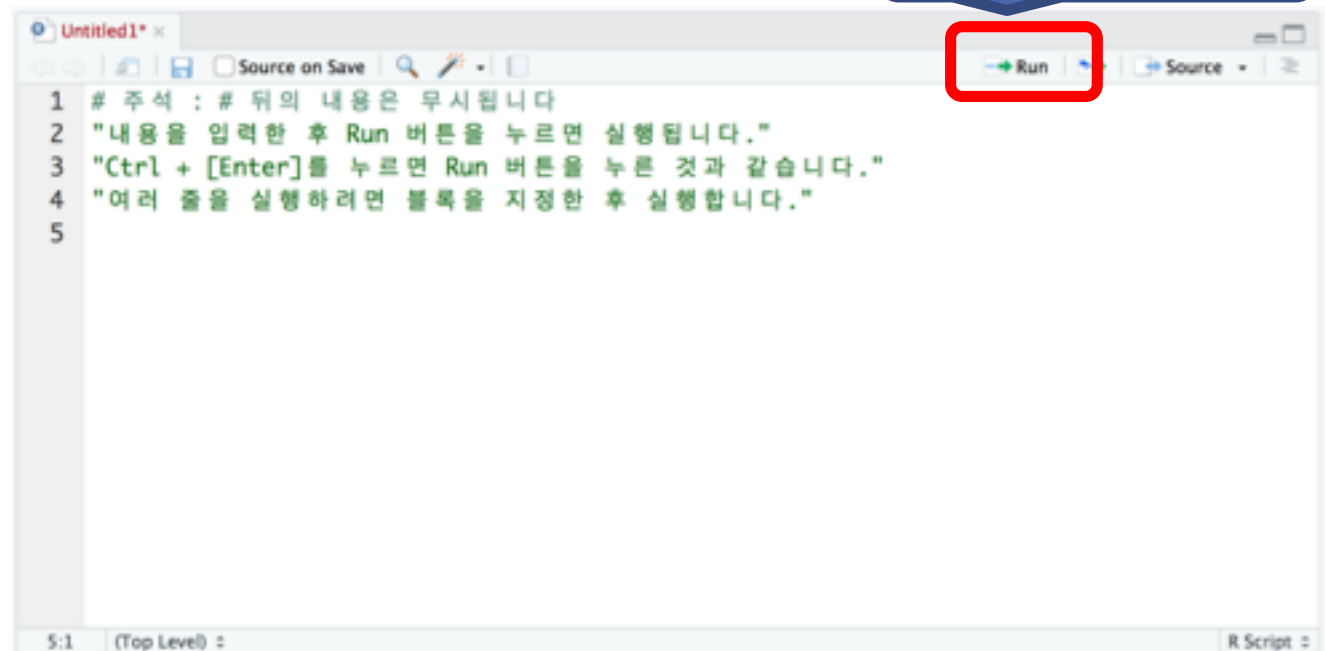


주의: 프로젝트 경로에 한글이 들어가면 오류

RStudio에 익숙해지기

- ▶ R과 RStudio를 설치하고 각 부분에 대해 간략히 살펴봅시다
- ▶ 콘솔창에 내용을 입력해 보고 출력값을 살펴봅시다
- ▶ 새 스크립트를 만들어 새 이름으로 저장하고 스크립트를 입력, 실행해봅니다.

Run = CTRL + [Enter]



R 이해하기

R Language

: *R* 프로그래밍 언어란?

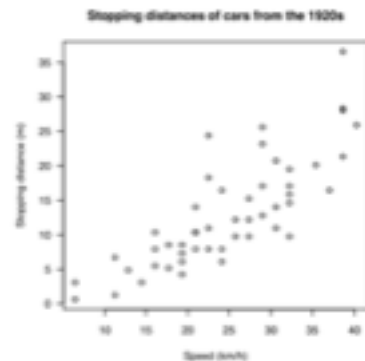
- ▶ 통계 계산과 그래픽을 위한 프로그래밍 언어이자 소프트웨어 환경
- ▶ 뉴질랜드 오클랜드 대학 로버트 젠틀맨(*Robert Gentleman*)과 로스 이하카(*Ross Ihaka*)에 의해 시작
- ▶ 현재는 *R* 코어 팀이 개발
- ▶ *S* 프로그래밍 언어(벨 연구소)의 구현체. *GPL* 하에 개발되는 *GNU S*라고도 한다
- ▶ 통계 소프트웨어의 개발과 자료 분석에 널리 사용
- ▶ 패키지 개발에 용이하여 통계학자들 사이에서 널리 사용

R = 어떤 형태의 데이터든 자유롭게 분석할 수 있는 범용 분석 툴

R의 주요 용도들

▶ 통계 분석

- ▶ 기초 통계 분석부터 고급 통계 분석 기법까지
- ▶ 최근 빅데이터의 추세에 따라 새로 개발되는 다양한 분석 기법들을 빠르게 적용 가능



▶ 머신러닝 모델링

- ▶ 다양한 데이터를 이용, 특정 변수를 예측할 수 있는 모델을 만드는 기법



R의 주요 용도들

▶ 텍스트 마이닝

- ▶ 문자로 이루어진 데이터를 분석
- ▶ 비정형 데이터의 분류와 감성 분석 등을 통해 유용한 정보와 지식, 패턴이나 트렌드를 찾아낼 수 있음
- ▶ 불필요한 데이터를 걸러내고 그 안에서 인사이트를 찾아내는데 도움

▶ 소셜 네트워크 분석

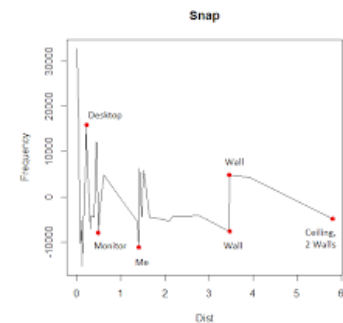
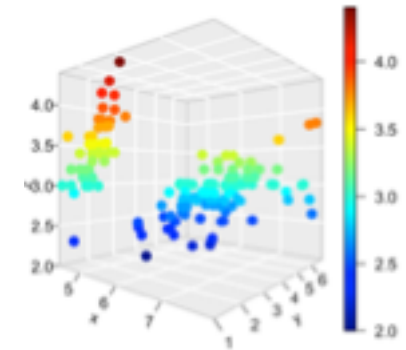
- ▶ 소셜 네트워크 사용자 관계 분석, 관심도 분석, 메시지 전파 경로 분석 등

▶ 시각화(Visualization)

- ▶ 데이터를 직관적으로 한눈에 알아볼 수 있도록 풍부한 시각화 기능 제공

▶ 비정형 데이터 분석

- ▶ 이미지, 사운드 등 그동안 데이터 분석에서 다루어지지 않았던 다양한 형태에 데이터를 분석, 처리



누가 **R**을 사용하는가?

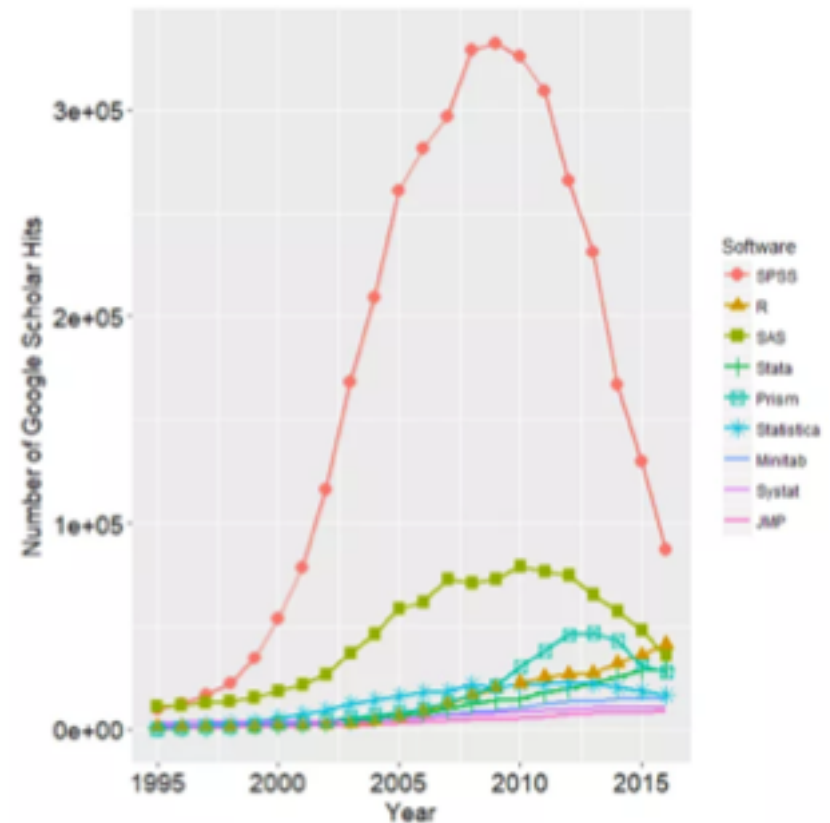
일반 용도는 물론 학술논문에도 널리 활용중

What self-service analytic tool are you currently using?



The Popularity of Data Science

<http://r4stats.com/articles/popularity/>

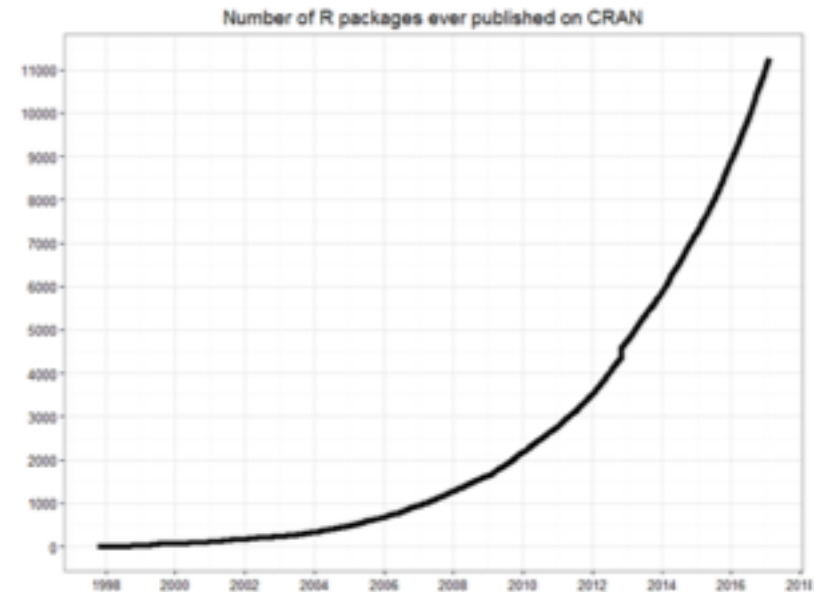


Tiobe Index : June 2018

| Jun 2018 | Jun 2017 | Change | Programming Language | Ratings | Change |
|----------|----------|--------|----------------------|---------|--------|
| 1 | 1 | | Java | 15.368% | +0.88% |
| 2 | 2 | | C | 14.936% | +8.09% |
| 3 | 3 | | C++ | 8.337% | +2.61% |
| 4 | 4 | | Python | 5.761% | +1.43% |
| 5 | 5 | | C# | 4.314% | +0.78% |
| 6 | 6 | | Visual Basic .NET | 3.762% | +0.65% |
| 7 | 8 | ▲ | PHP | 2.681% | +0.11% |
| 8 | 7 | ▼ | JavaScript | 2.495% | -0.53% |
| 9 | | ▲ | SC | 2.336% | +3.36% |
| 10 | 14 | ▲ | R | 1.452% | -0.70% |
| 11 | 11 | | Perl | 1.252% | -0.91% |
| 12 | 18 | ▲ | Objective-C | 1.181% | -0.78% |
| 13 | 16 | ▲ | Visual Basic | 1.154% | -0.66% |
| 14 | 9 | ▼ | Perl | 1.147% | -1.16% |
| 15 | 12 | ▼ | Swift | 1.145% | -1.06% |
| 16 | 10 | ▼ | Assembly language | 0.915% | -1.34% |
| 17 | 17 | | MATLAB | 0.894% | -1.10% |
| 18 | 15 | ▼ | Go | 0.879% | -1.17% |
| 19 | 13 | ▼ | Delphi/Object Pascal | 0.875% | -1.28% |
| 20 | 20 | | PL/SQL | 0.848% | -0.72% |

R의 장점 : Open Source & Free to Use

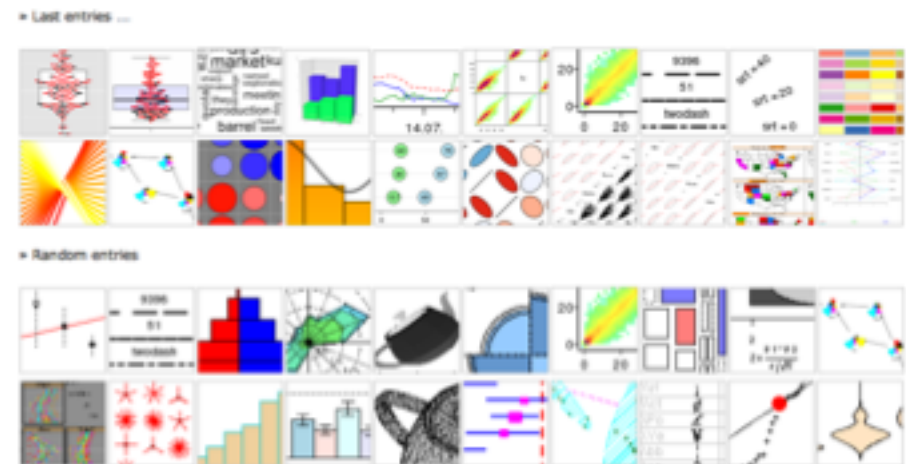
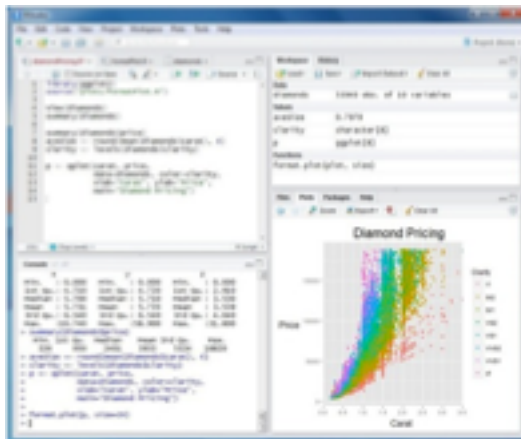
- ▶ **오픈소스 & FREE**
 - ▶ R은 무료이며 상용툴이 제공하는 기능을 대부분 갖추고 있다
 - ▶ 다양한 패키지와 최신 분석 기법을 활용
 - ▶ 새로운 분석 기법이 나오면 빠른 시간 내에 패키지가 업로드
 - ▶ 최근의 추세는 비정형 데이터가 늘어남에 따라 새로운 분석기법들이 다수 등장
- ▶ **CRAN : Over 10,000 packages (2017. 1 기준)**



<http://blog.revolutionanalytics.com/2017/01/cran-10000.html>

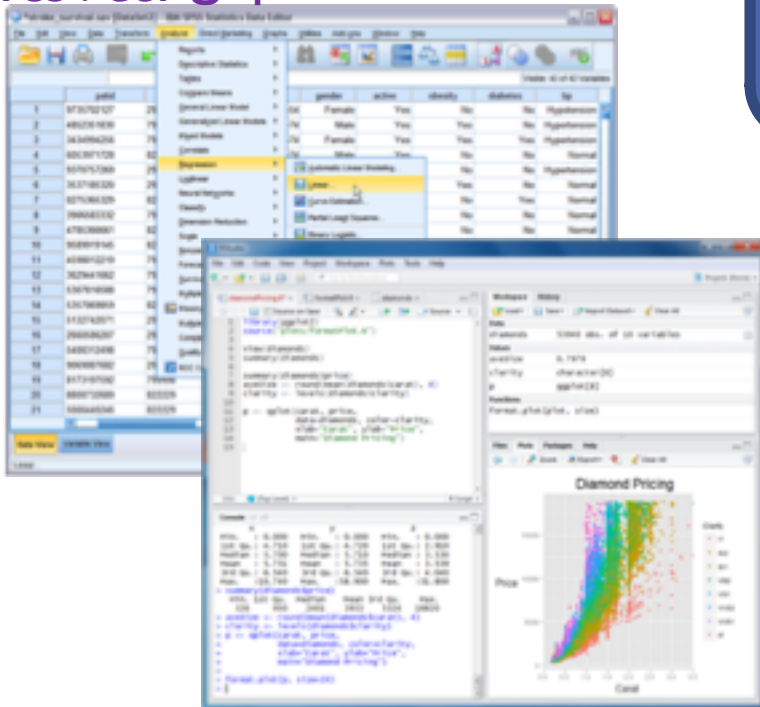
R의 장점 : *Data Visualization*

- ▶ 데이터 분석 결과를 이해하기 쉽도록 그래프로 작성
- ▶ 고품질의 그래프와 데이터 시각화를 손쉽게 작성할 수 있다
- ▶ 데이터 분석부터 시각화 작업까지 하나의 도구로 완성할 수 있어 효율적



R의 장점 : 프로그래밍 방식

SPSS : GUI 방식



프로그래밍 방식: RStudio

“재현 가능한 고품질 분석을 생성하기 위한 최선의 수단”

“어떤 통계 기법을 사용하던 그에 맞는 R 패키지가 나와 있다”

맷 애덤스(데이터 사이언티스트)

▶ 재현성의 확보

- ▶ 분석의 전 과정을 코드로 남길 수 있다

▶ 오류를 쉽게 파악, 수정할 수 있다

- ▶ 분석 결과에 이상이 발견되어도 오류를 쉽게 파악하여 수정할 수 있다

▶ 공동 작업을 할 수 있다

- ▶ 모든 분석 과정이 코드로 남기 때문에 코드를 공유하면서 공동으로 작업을 할 수 있다

R과 Python

- ▶ **Python**도 R과 함께 데이터 분석 도구로 주목
 - ▶ 범용 프로그래밍 언어
 - ▶ **Pandas** : 통계, 데이터 처리 관련 모듈
 - ▶ **TensorFlow** : 딥러닝 라이브러리
 - ▶ 전문 프로그래밍 언어이므로 더 많은 공학적 지식이 필요
- ▶ 주 목적이 데이터를 다루는 것이라면 **R**,
소프트웨어 개발에 데이터 분석을 필요로 한다면 파이썬을 추천

