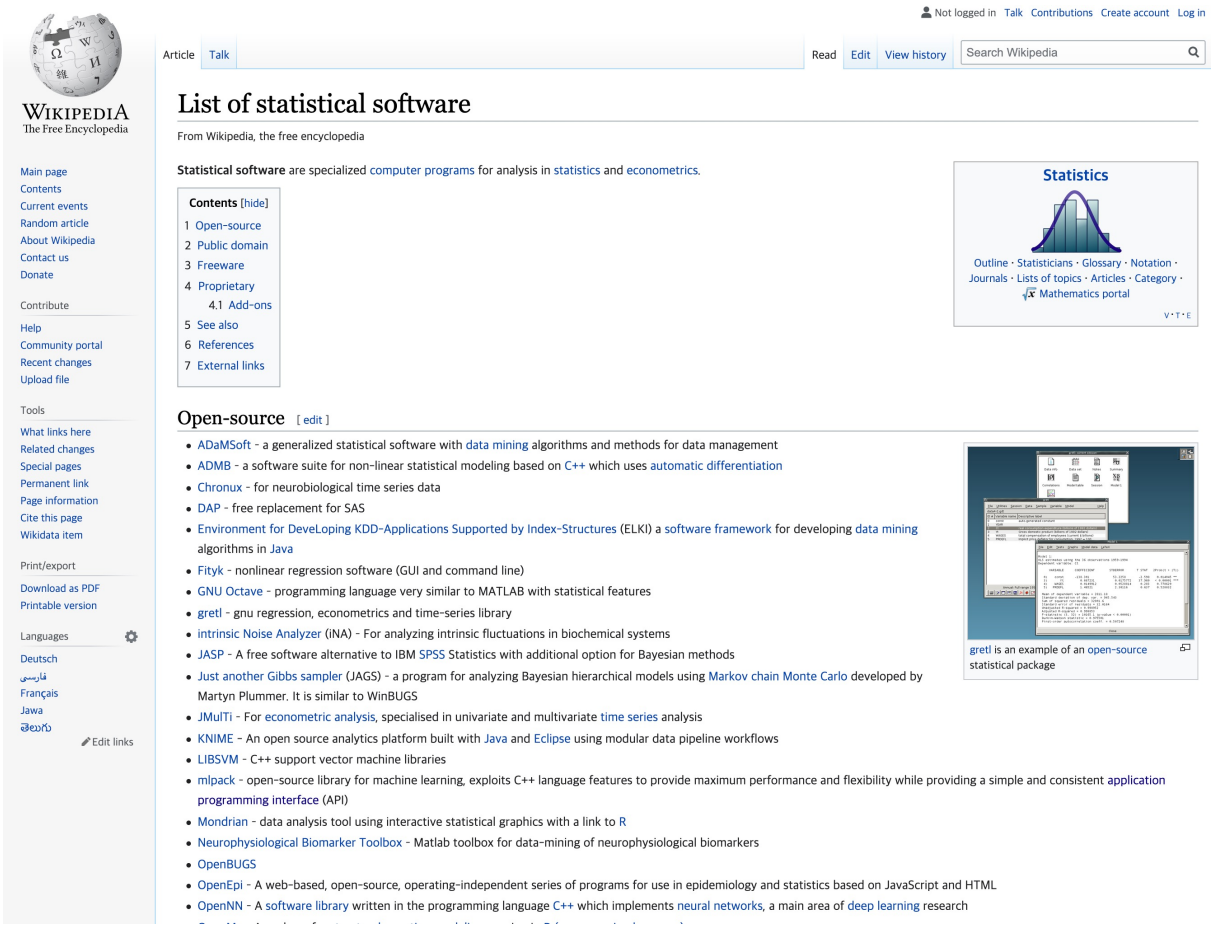
# Lecture 02
# Softwares and programming languages

# Software packages for statistics

- There are so many software packages for statistics

https://en.wikipedia.org/wiki/List_of_statistical_software



Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

Languages
Deutsch
فارسی
Français
Jawa
తెలుగు
Edit links

## List of statistical software

From Wikipedia, the free encyclopedia

**Statistical software** are specialized computer programs for analysis in statistics and econometrics.

Contents [hide]
1 Open-source
2 Public domain
3 Freeware
4 Proprietary
4.1 Add-ons
5 See also
6 References
7 External links

Statistics
Outline · Statisticians · Glossary · Notation · Journals · Lists of topics · Articles · Category · Mathematics portal

### Open-source [edit]

- ADaMSoft - a generalized statistical software with data mining algorithms and methods for data management
- ADMB - a software suite for non-linear statistical modeling based on C++ which uses automatic differentiation
- Chronux - for neurobiological time series data
- DAP - free replacement for SAS
- Environment for DeveLoping KDD-Applications Supported by Index-Structures (ELKI) a software framework for developing data mining algorithms in Java
- Fityk - nonlinear regression software (GUI and command line)
- GNU Octave - programming language very similar to MATLAB with statistical features
- gretl - gnu regression, econometrics and time-series library
- intrinsic Noise Analyzer (iNA) - For analyzing intrinsic fluctuations in biochemical systems
- JASP - A free software alternative to IBM SPSS Statistics with additional option for Bayesian methods
- Just another Gibbs sampler (JAGS) - a program for analyzing Bayesian hierarchical models using Markov chain Monte Carlo developed by Martyn Plummer. It is similar to WinBUGS
- JMulTi - For econometric analysis, specialised in univariate and multivariate time series analysis
- KNIME - An open source analytics platform built with Java and Eclipse using modular data pipeline workflows
- LIBSVM - C++ support vector machine libraries
- mlpack - open-source library for machine learning, exploits C++ language features to provide maximum performance and flexibility while providing a simple and consistent application programming interface (API)
- Mondrian - data analysis tool using interactive statistical graphics with a link to R
- Neurophysiological Biomarker Toolbox - Matlab toolbox for data-mining of neurophysiological biomarkers
- OpenBUGS
- OpenEpi - A web-based, open-source, operating-independent series of programs for use in epidemiology and statistics based on JavaScript and HTML
- OpenNN - A software library written in the programming language C++ which implements neural networks, a main area of deep learning research

gretl is an example of an open-source statistical package

# The ones that I have used before… and now…

OLDIES (that I am not using any more)

Basic statistics

- SAS (during my undergrad)
- SPSS (during my masters)
- STATA (for my wife)

Multivariate

- LISREL
- Mplus
- AMOS

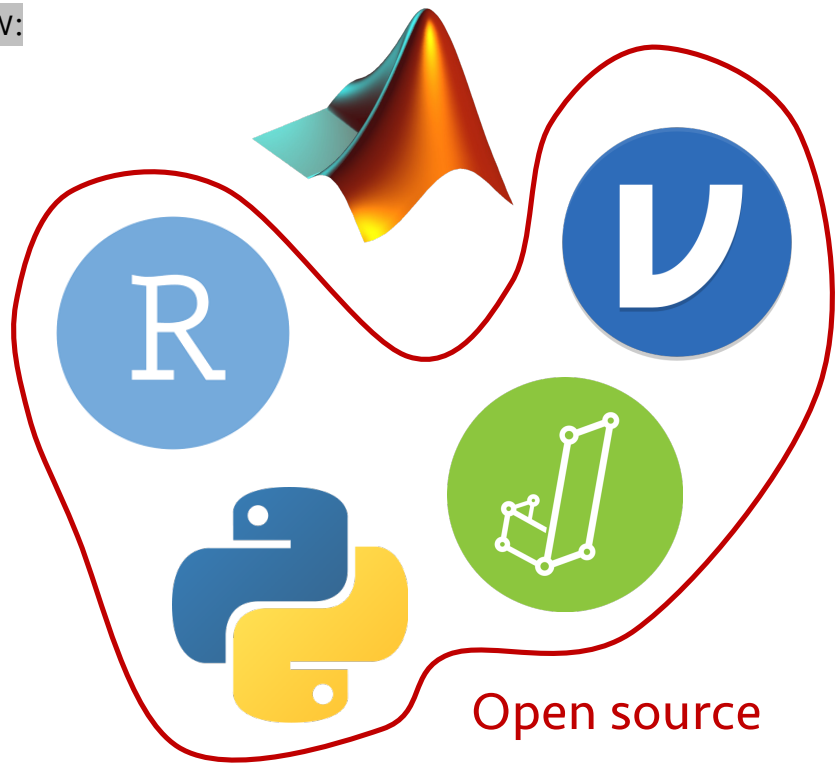
What I am using now:

Basic statistics

- Matlab
- R
- JASP
- JAMOVI
- Python

Multivariate

- Matlab
- Python
- R




Cultural changes for the last 20 years

# The ones that I have used before… and now…

성균관대학교 SUNGKYUNKWAN UNIVERSITY(SKKU)

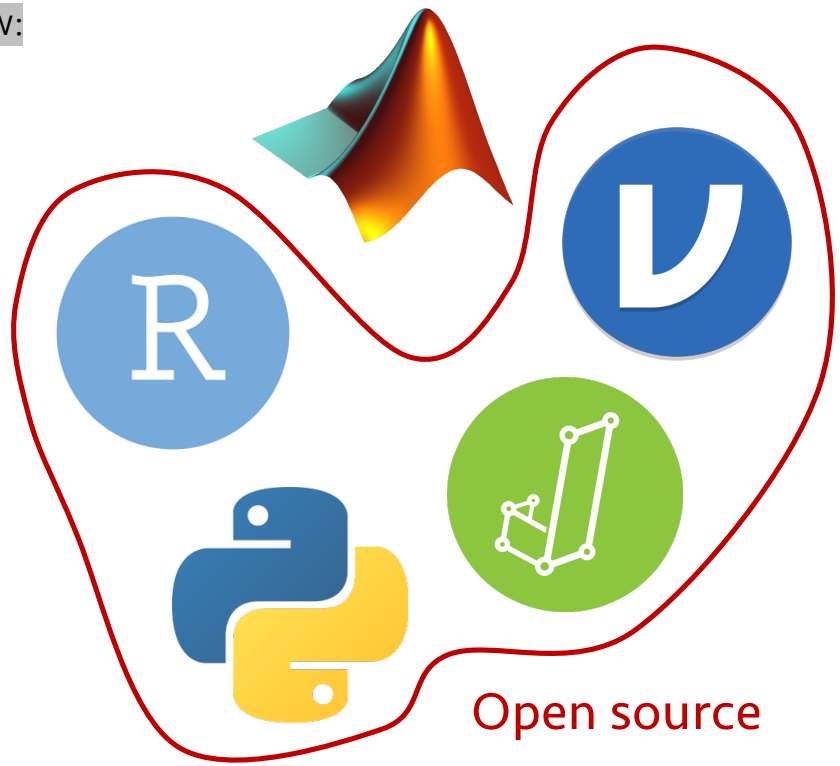
학생지원
편의 / 복지
학생활동
IT 서비스
IT Call center
모바일 앱 이용안내
노트북/SW 대여
무선랜 이용신청
전산망 이용 신청 / 서버계정신청
킹고포털
홈빌더 / 구글 드라이브
영상회의시스템 / 전화
증명발급
대표홈페이지 수정 요청

* 학내 공용 SW 목록

at SKKU

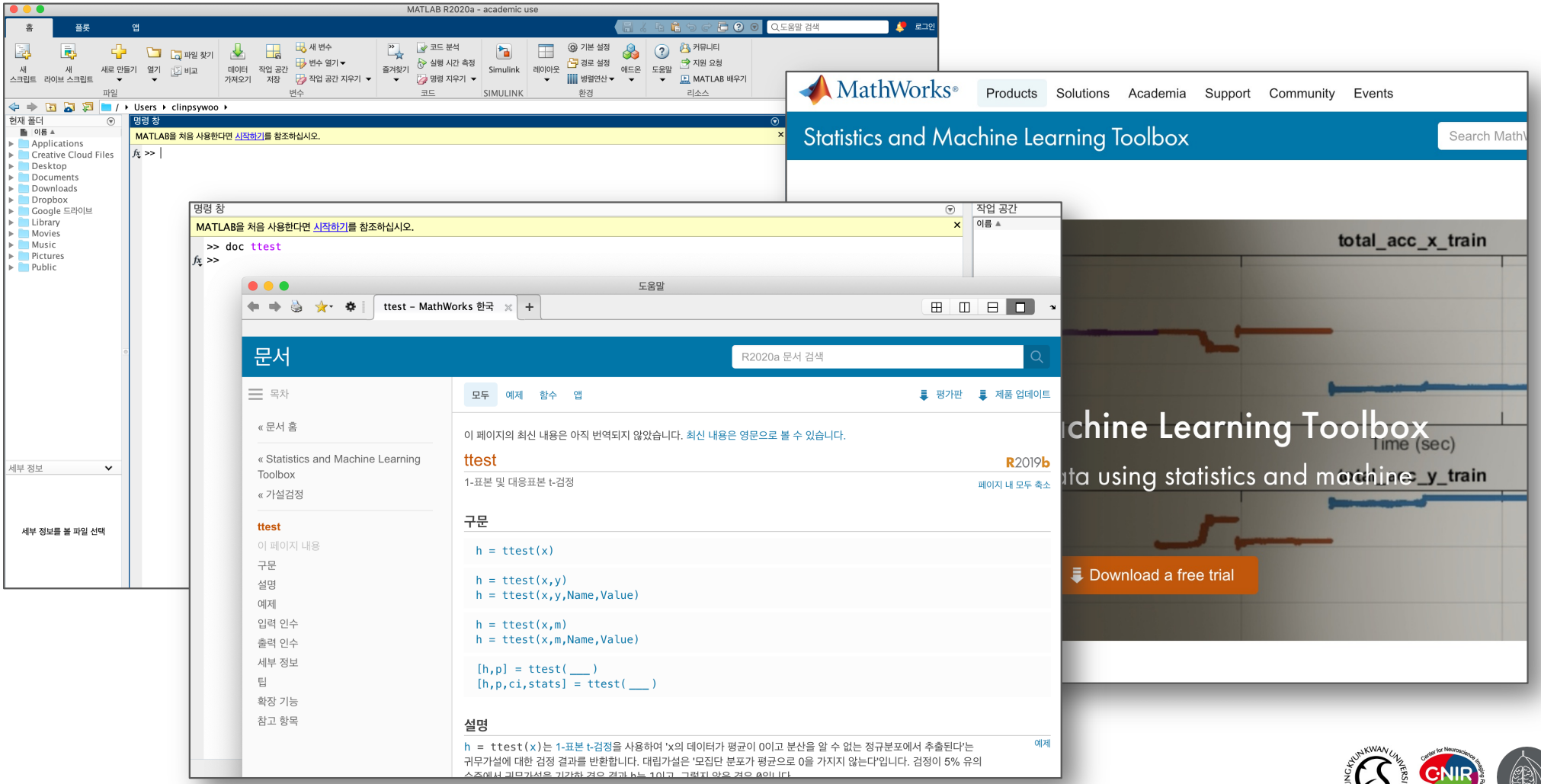
| SW 종류 | 사용 지역 | | 사용자 구분 | | |
|---|---|---|---|---|---|
| | 인사캠 | 자과캠 | 교원 | 직원 | 학생 |
| MS Window OS | ○ | ○ | ○ | ○ | × |
| MS Office | ○ | ○ | ○ | ○ | × |
| MS O365 Service | ○ | ○ | × | × | ○ |
| Adobe | ○ | ○ | ○ | ○ | × |
| 아래한글 | ○ | ○ | ○ | ○ | × |
| MATLAB | × | ○ | ○ | ○ | ○ |
| SAS | ○ | ○ | ○ | ○ | ○ |
| SPSS | ○ | ○ | ○ | ○ | × |
| STATA | ○ | ○ | ○ | ○ | × |
| V3 | ○ | ○ | ○ | ○ | × |
| ALTOOLS | ○ | ○ | ○ | ○ | × |
| IT(OA)강좌 | ○ | ○ | ○ | ○ | ○ |

now:




Cultural changes for the last 20 years

# Matlab


MATLAB R2020a - academic use
MATLAB을 처음 사용한다면 시작하기를 참조하십시오.
>> doc ttest
>>
MathWorks
Products
Solutions
Academia
Support
Community
Events
Statistics and Machine Learning Toolbox
total_acc_x_train
Download a free trial
도움말
ttest - MathWorks 한국
문서
R2020a 문서 검색
ttest
R2019b
1-표본 및 대응표본 t-검정
구문
h = ttest(x)
h = ttest(x,y)
h = ttest(x,y,Name,Value)
h = ttest(x,m)
h = ttest(x,m,Name,Value)
[h,p] = ttest(___)
[h,p,ci,stats] = ttest(___)
설명

# Matlab

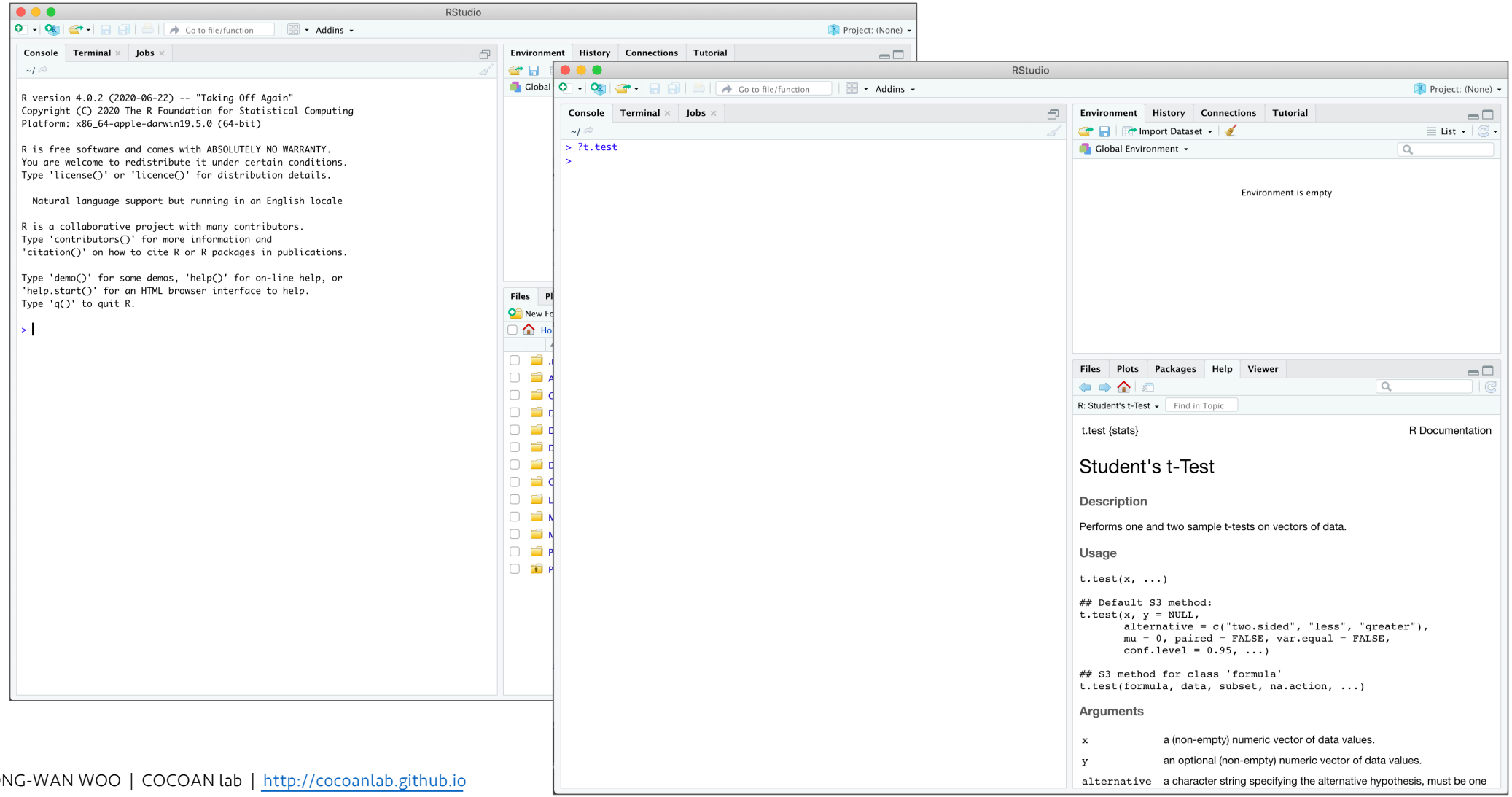
**Strengths:**

- Command line coding
- Simple and fast
- Well-developed toolboxes (e.g., statistics and machine learning toolbox)
- Open source tools for neuroimaging and others (thick user base)

**Weaknesses:**

- Expensive (what if you graduate the school?)
- The popularity is decreasing (the user base is getting weaker)
- Deep learning?

# R


RStudio
Console
Terminal
Jobs
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin19.5.0 (64-bit)
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.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> ?t.test
Environment
History
Connections
Tutorial
Environment is empty
Files
Plots
Packages
Help
Viewer
t.test {stats}
R Documentation
Student's t-Test
Description
Performs one and two sample t-tests on vectors of data.
Usage
t.test(x, ...)
## Default S3 method:
t.test(x, y = NULL,
alternative = c("two.sided", "less", "greater"),
mu = 0, paired = FALSE, var.equal = FALSE,
conf.level = 0.95, ...)
## S3 method for class 'formula'
t.test(formula, data, subset, na.action, ...)
Arguments
x a (non-empty) numeric vector of data values.
y an optional (non-empty) numeric vector of data values.
alternative a character string specifying the alternative hypothesis, must be one

# R

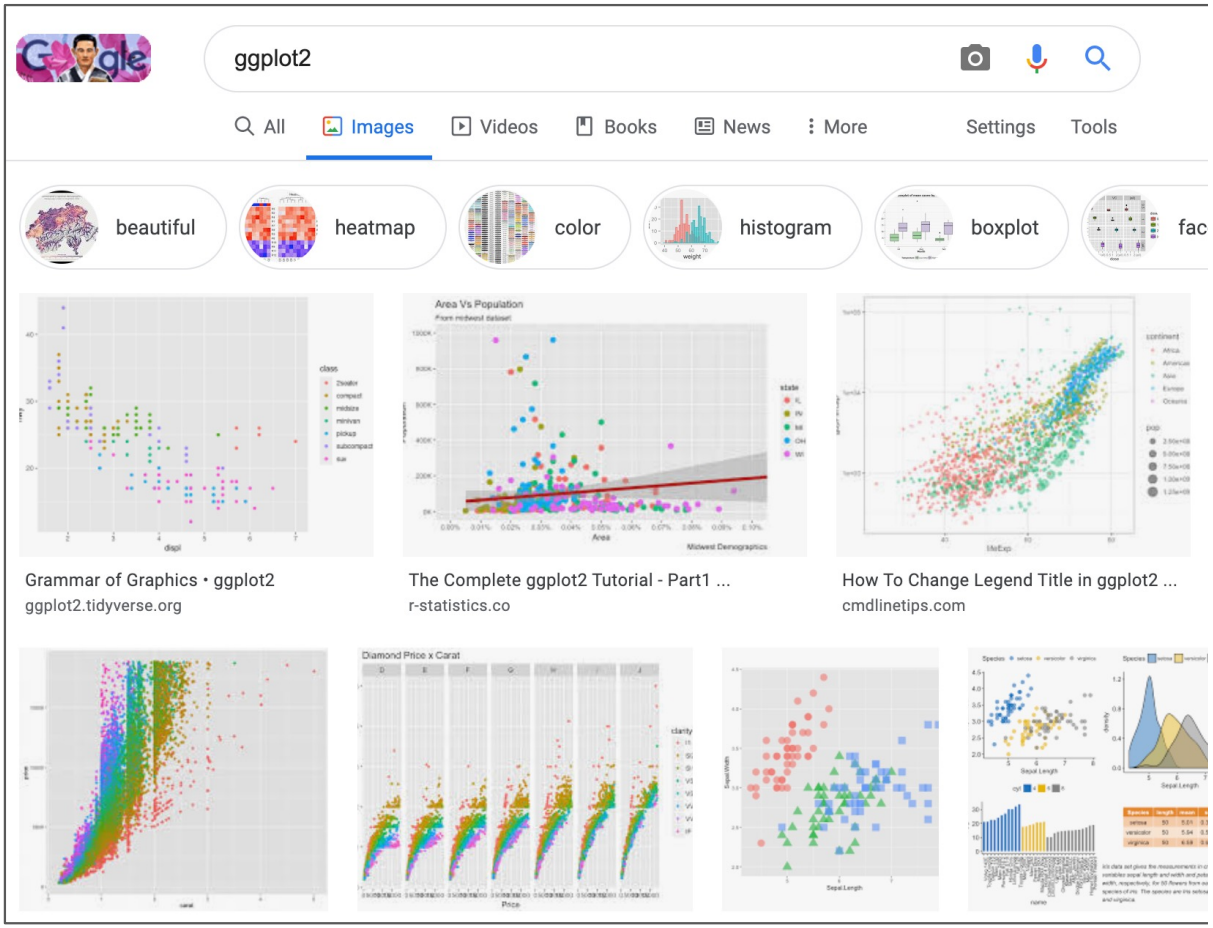
**Strengths:**

- Command line coding
- Simple and fast
- Thick user base!!! Esp. in statistics;

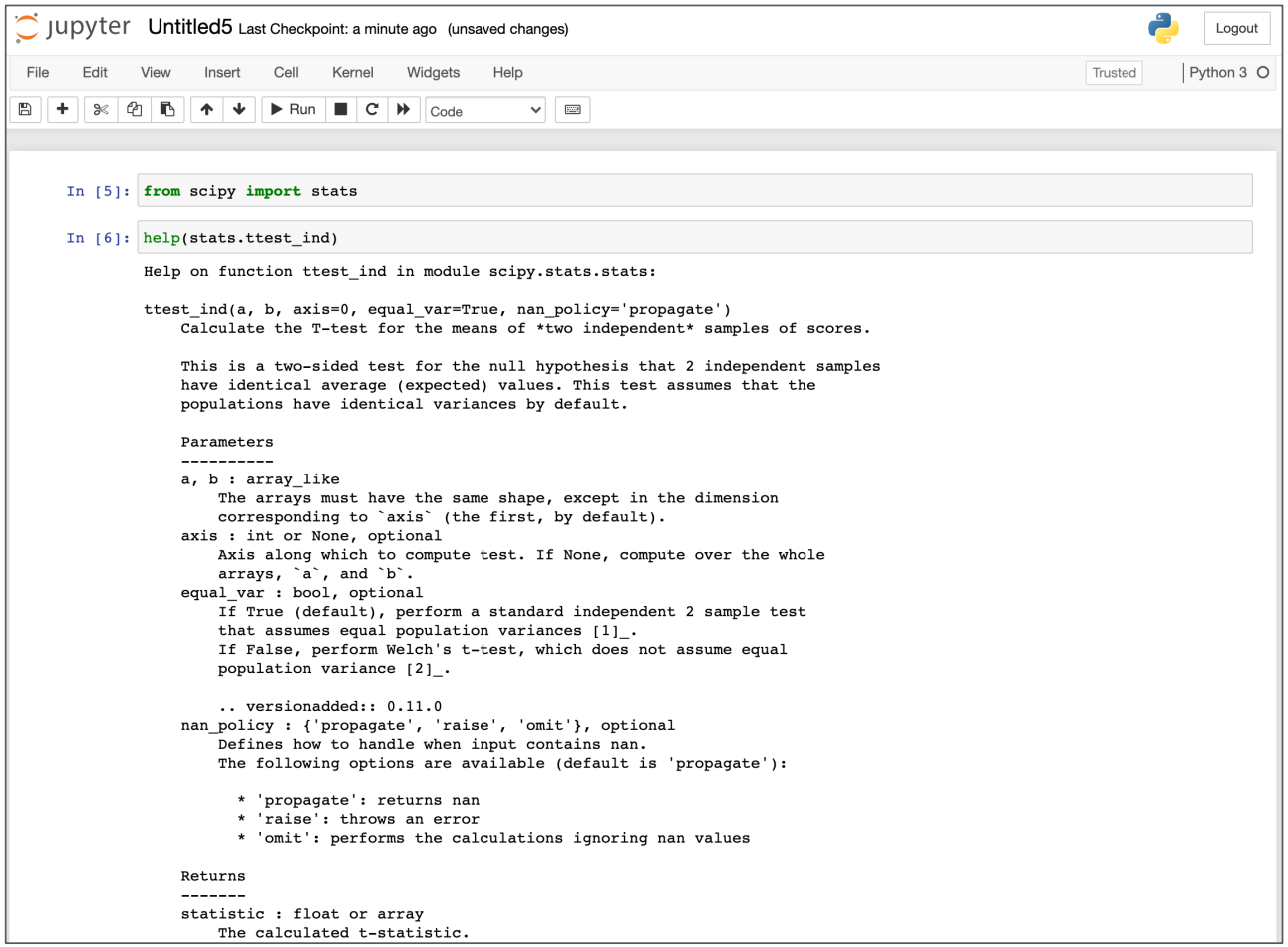
  so many useful packages for advanced stats
- Good visualization tools (e.g., ggplot2)
- The popularity is increasing

**Weaknesses:**

- No good connection with neuroimaging
- The popularity is lower than python
- Deep learning?



# Python


jupyter Untitled5 Last Checkpoint: a minute ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [5]: from scipy import stats
In [6]: help(stats.ttest_ind)
Help on function ttest_ind in module scipy.stats.stats:
ttest_ind(a, b, axis=0, equal_var=True, nan_policy='propagate')
Calculate the T-test for the means of *two independent* samples of scores.
This is a two-sided test for the null hypothesis that 2 independent samples
have identical average (expected) values. This test assumes that the
populations have identical variances by default.
Parameters
a, b : array_like
The arrays must have the same shape, except in the dimension
corresponding to `axis` (the first, by default).
axis : int or None, optional
Axis along which to compute test. If None, compute over the whole
arrays, `a`, and `b`.
equal_var : bool, optional
If True (default), perform a standard independent 2 sample test
that assumes equal population variances [1]_.
If False, perform Welch's t-test, which does not assume equal
population variance [2]_.
.. versionadded:: 0.11.0
nan_policy : {'propagate', 'raise', 'omit'}, optional
Defines how to handle when input contains nan.
The following options are available (default is 'propagate'):
* 'propagate': returns nan
* 'raise': throws an error
* 'omit': performs the calculations ignoring nan values
Returns
statistic : float or array
The calculated t-statistic.

# Python

**Strengths:**

- Command line coding
- Simple and fast
- Thick thick user base! Esp. in machine learning and others;

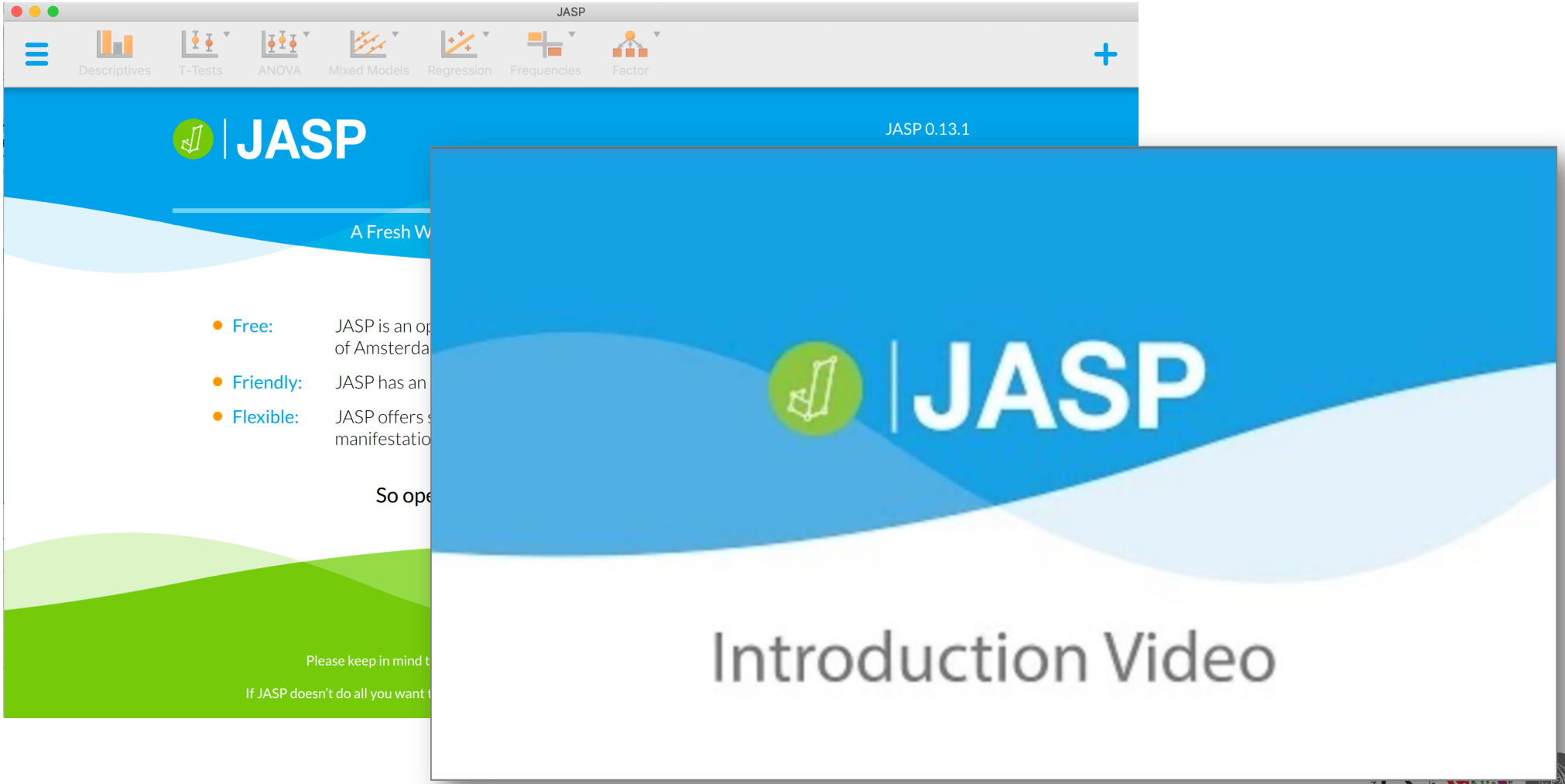
  so many useful packages for everything
- Deep learning!
- The popularity is rapidly increasing

**Weaknesses:**

- Not really good for statistics (maybe compared to R or matlab), but still okay

# JASP

https://jasp-stats.org/2017/06/27/new-5-minute-video-introducing-jasp/



# JASP

**Strengths:**

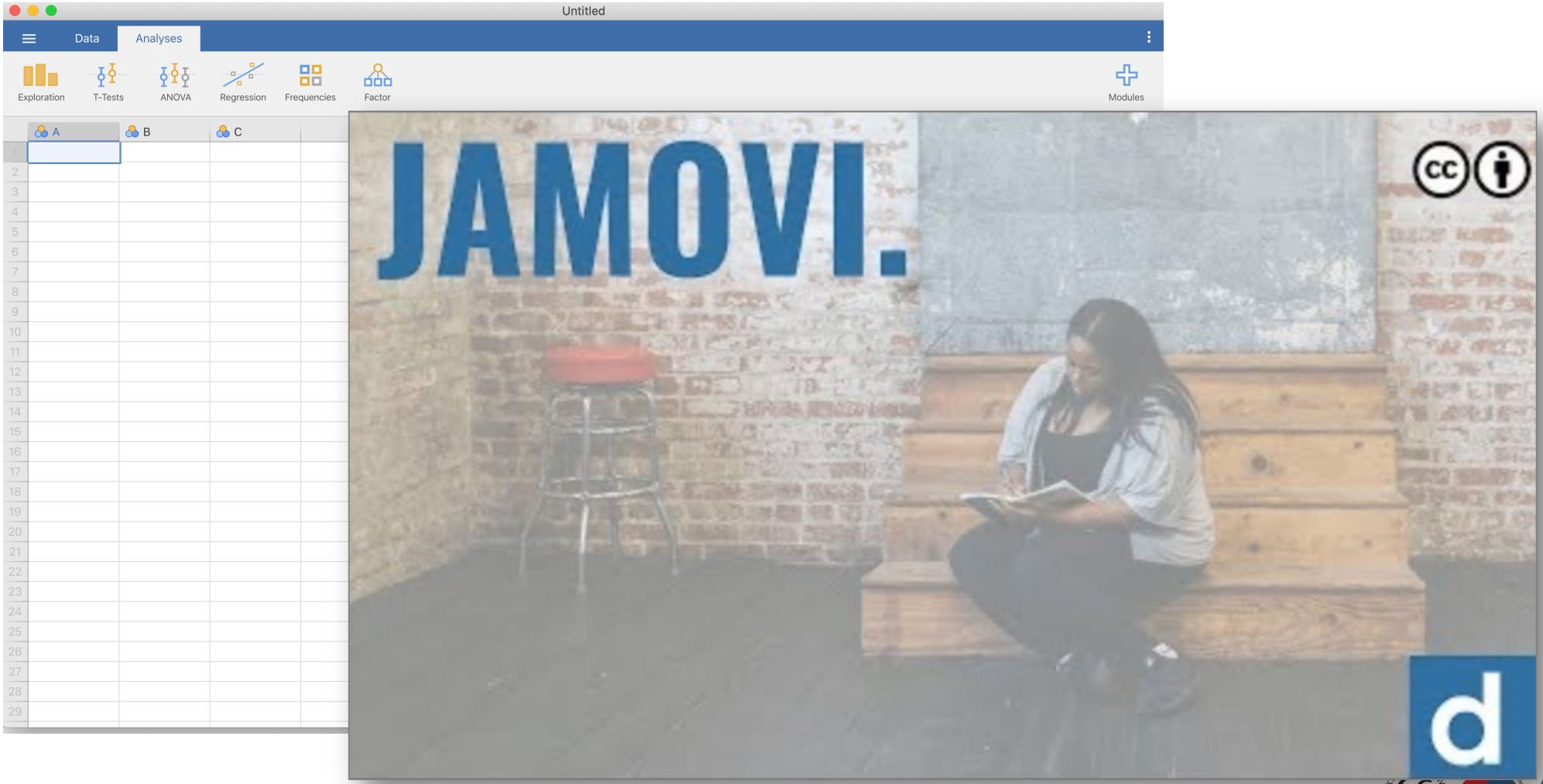
- Free!
- Graphical user interface
- Simple and easy
- Bayesian analysis

**Weaknesses:**

- Not command line
- No flexibility
- No development

# JAMOVI

https://www.youtube.com/watch?v=Ej9e8lzaeDE&feature=emb_logo



# JAMOVI (similar to JASP)

**Strengths:**

- Free!
- Graphical user interface
- Simple and easy
- Bayesian analysis

**Weaknesses:**

- Not command line
- No flexibility
- No development

Interestingly, JAMOVI was developed by the previously JASP developers.

http://blog.efpsa.org/2017/03/23/introducing-jamovi-free-and-open-statistical-software-combining-ease-of-use-with-the-power-of-r/

# Thus, our choice for the class is…

JAMOVI

(JASP)

Matlab

R