

Useful open science tools to do research in psychiatry



Thomas Gargot
thomas_gargot@hotmail.com
EPA Congress,
Tuesday March 6th

<https://thecakereportblog.wordpress.com/> <https://github.com/Ouphix/TCR>



Useful open science tools to do research in psychiatry



Randomizer.org
Analyse data



Collect data
Limesurvey



Share data
OSF and GitHub



Write an article



Make the knowledge more visible

Goals of the workshop

- Understand the challenges of open science
- Learn the concepts of some tools
- Consider your own research/projects
- Rate cakes

<p>Who I am</p> <ul style="list-style-type: none">• Psychiatrist in Paris• PhD Student on Computer Science : use of IT in school• AETCC• Interest in IT and psychotherapy (CBT)• Master in Cognitive science in ENS• Former IT secretary of EFPT• Psychotherapy WG chairman <p> La Pitié-Salpêtrière</p> <p> EUROPEAN FEDERATION OF PSYCHIATRIC TRAINEES</p>	<hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
---	---

<p>Disclosure</p> <ul style="list-style-type: none">• None	<hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
---	---

<p>Ethical issues</p>	<hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
------------------------------	---

Problems

- **Cost:** How much will it cost to run this study?
- **Method:** What do they mean when they did this analysis?
- **Licence:** Does someone has the last version of this software ?
- **Memory:** Where is my data, how to re-use, how share it?
- **Paywall:** I can't get this article
- **Visibility:** Nobody cares about our work

Publication lucrative vicious circle





<https://www.youtube.com/watch?v=WnxqoP-cOZE&list=PLKipY1cRhemK1Fn6PbVe5zOfxLlI0YY&index=6>

Goals

- Decrease costs
- Improve transparency
- Improve accessibility to softwares, articles, and re-use of data
- Improve visibility for public

➔ Is it really possible ?!?

Why is there a need for open science?

**nature
human behaviour** PERSPECTIVE
PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021
OPEN

A manifesto for reproducible science

Marcus R. Munafò^{1,2*}, Brian A. Nosek^{3,4}, Dorothy V. M. Bishop⁵, Katherine S. Button⁶, Christopher D. Chambers⁷, Nathalie Pericie du Sert⁸, Uri Simonsohn⁹, Eric-Jan Wagenmakers¹⁰, Jennifer J. Ware¹¹ and John P. A. Ioannidis^{12,13,14}

Data from many fields suggests reproducibility is lower than is desirable^{8–14}; one analysis estimates that 85% of biomedical research efforts are wasted¹⁴, while 90% of respondents to a recent survey in *Nature* agreed that there is a ‘reproducibility crisis’¹⁵.

<https://www.nature.com/articles/s41562-016-0021>

Information world



- Share information don't decrease the ownership of the sharer
- Rather, better accessibility improve possibilities of not anticipated use (hacking) and feedback
- Open softwares are more flexible and have fewer bugs than closed/proprietary ones

Free / Open?



Open software: Free use, study, modification,
duplication and sharing
≠ Freemium strategy (demonstration)
≠ centralisation and user data collection

<https://www.open.com/~ysp/2016/02/07/product/>

An example of Open science pathway

Hands-on

Research protocol

- Make a data base of cookies tasting
- Rate them in an international context
- Select the best cakes in the world

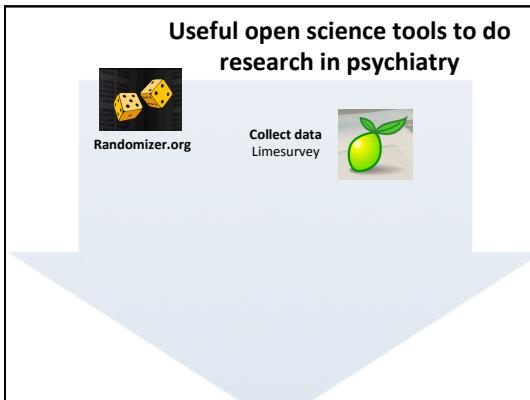


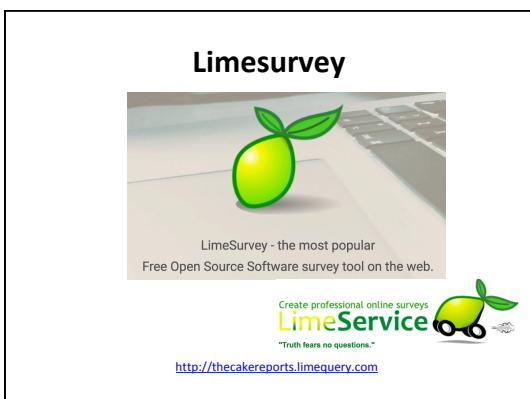
Sampling process during EPA congress, Madrid, 2016











How does it look like?

1 2 3 4 5 No answer

How does it taste?

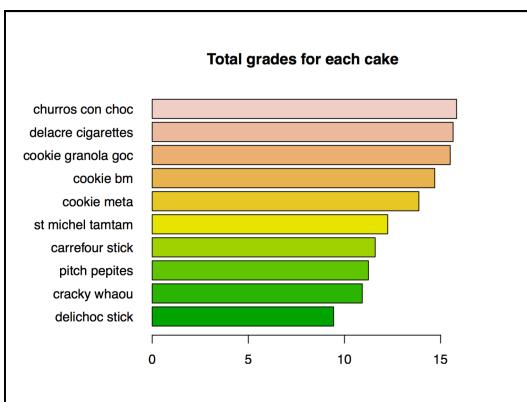
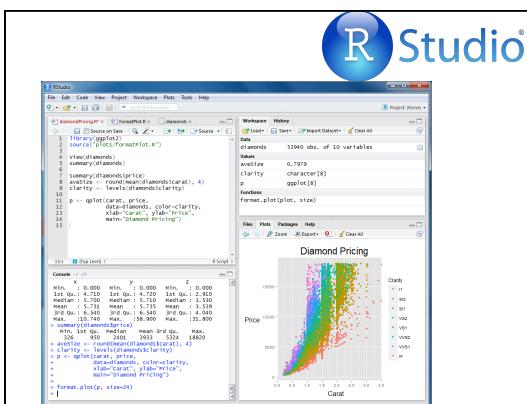
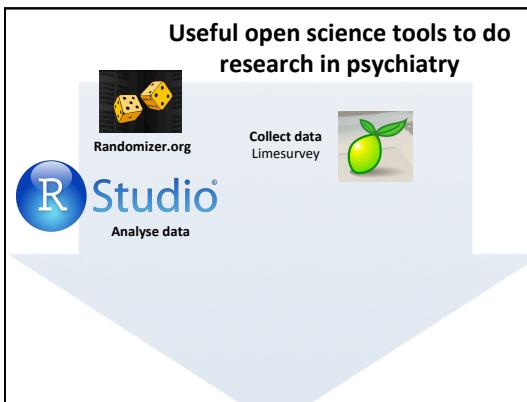
1 2 3 4 5 No answer

How is the texture?

1 2 3 4 5 No answer

How is the packaging?

1 2 3 4 5 No answer



Connect Limesurvey with R ? : Shinyapps

Assertiveness Rathus Scale

What is your pseudo ?
thog1

Number of observations to view:
10

Score Range	Frequency
-30 to -20	~25
20 to 30	~25

Your pseudo is thog1. Your score is -13. You are passive.

A score toward -30 is a sign that the subject has a lack of assertivity (passive style). A score around 0 is found in normal assertive behaviour.

<https://ouphix.shinyapps.io/affiDeSoi/>

The Python logo consists of a blue icon resembling a snake or a double helix, followed by the word "python" in a lowercase sans-serif font.

Useful open science tools to do research in psychiatry

- R Studio**
Analyse data
- Randomizer.org**
- Limesurvey**
- Share data**
OSF and GitHub
- GitHub**

20/12/17

Git Hub

- Share data

This repository Search Pull requests Issues Git

Ouphix / TCR

Code Issues Pull requests Projects Wiki Pulse Graphs Settings

The Cake Report <https://thecakereportblog.wordpress.com/> — Edit

0 commits 1 branch 0 releases All 2 contributors

Branch: master • New pull request Create new file Upload files Find file Clone or download

AnneCha join your secret via vie

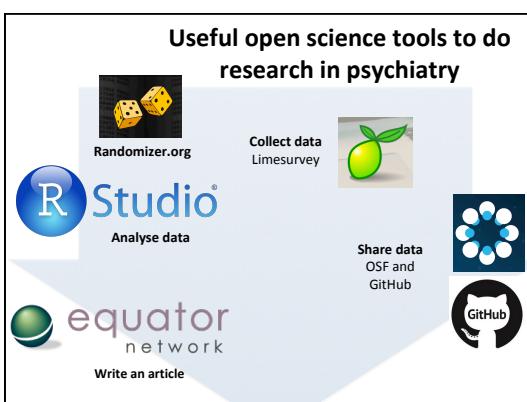
README.md README.rmd TheCakeReport.Rmd TheCakeReport.Rproj TheCakeReport.pdf carefur stick.png

Rmarkdown v1 Rmarkdown v1

7 months ago 7 months ago 7 months ago 6 months ago 6 months ago 6 months ago 6 months ago

Latest commit 1 week ago on 9 Apr







Equator Network

Enhancing the QUAlity and Transparency Of health Research

EUQUATOR resources in English | Spanish

[Home](#) [Library](#) [Toolkits](#) [Courses & events](#) [News](#) [Blog](#) [Librarian Network](#) [About us](#) [Contact](#)

Essential resources for writing and publishing health research

Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to reporting.

- Search for reporting guidelines
- Not sure which reporting guideline to use?
- Reporting guidelines under development
- Visit the library for

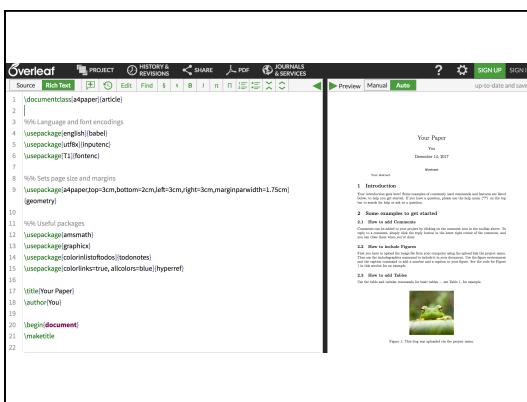
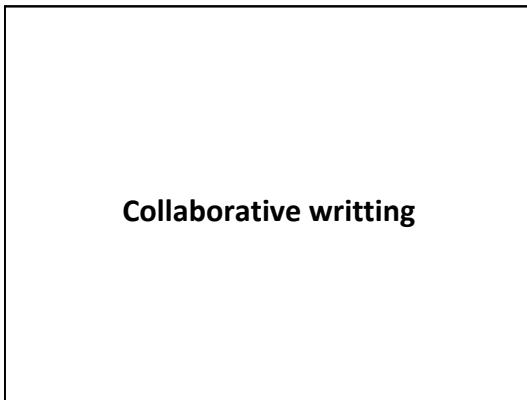
Reporting guidelines for main study types

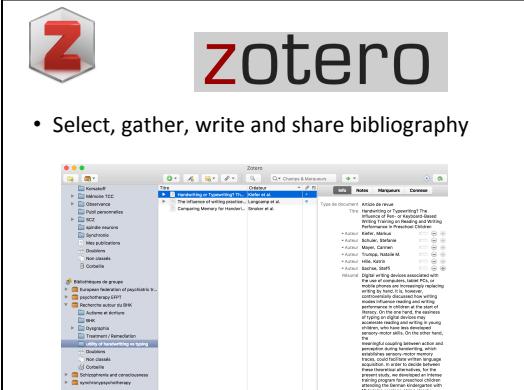
Randomized trials	CONSORT	Extensions	Other
Observational studies	STROBE	Extensions	Other
Evaluative studies	PRISMA	Extensions	Other
Care reports	CARE	Other	
Qualitative research	SQRG	COREQ	Other
Diagnostic / prognostic studies	STARD	TREND	Other
Quality improvement studies	SQUIRE	Other	
Economic evaluations	CHEERS	Other	
Animal pre-clinical studies	APPMV	Other	
Study protocols	SPIRIT	PRISMA-P	Other

Publication checklist for trial registration
The checklist is intended to help trialists and journal editors to ensure that all essential information is included in the trial registration record.



20/12/17





The screenshot shows the Zotero application window. At the top left is the Zotero logo, which is a stylized red 'Z' inside a hexagonal frame. The main title 'zotero' is centered above a list of items. Below the title, there is a bullet point: '• Select, gather, write and share bibliography'. The central part of the window displays a bibliography list with several entries. On the left side, there is a sidebar with various categories and sub-categories, such as 'Bibliothèque de groupes', 'Fonctionnalités de recherche', 'Recherches actives du GRK', 'Archives', 'Dossiers', 'Synthèses', and 'Traitements / Remarques'. Each entry in the bibliography list includes the author's name, the title of the work, and a small preview icon.

New ways to publish ?

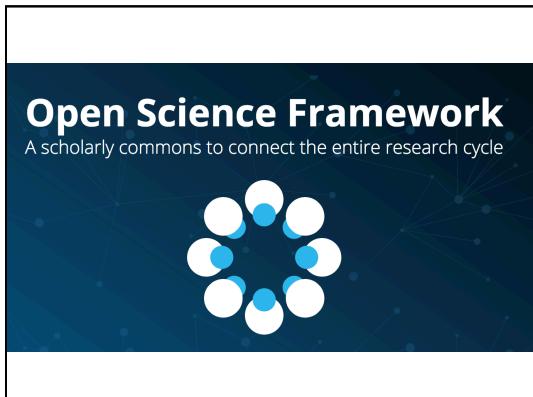


The screenshot shows the homepage of ArXiv : Preprints. The header features the text 'ArXiv : Preprints' and 'The Archives of the International Association for Child and Adolescent Psychiatry and Allied Professions'. Below the header, there is a navigation bar with links for 'Articles', 'Full Member Organizations', 'About IACAPAP', 'Media Gallery', 'Useful Links', 'Contact Us', and 'REGISTER'. A 'SUBMIT A MANUSCRIPT' button is also visible. The main content area contains a section titled 'About IACAPAP ArXiv' with a brief description and a photo of a person using a laptop. To the right, there are three logos: 'bioRxiv beta', 'PsyArXiv', and 'SOC ARXIV'.

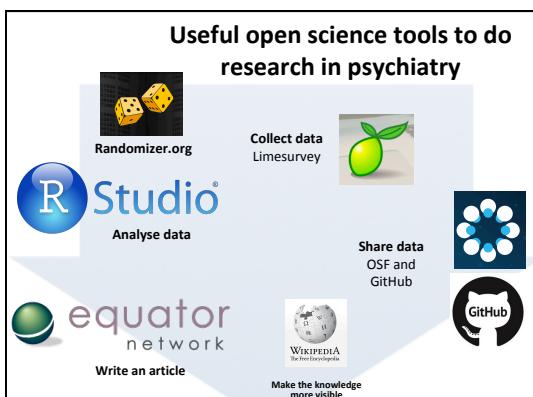
A screenshot of a website titled "EFPT Psychotherapy Guidebook". The page features a large image of a human brain in blue and white. At the top right, there is a small box showing "1 Following" and "Edit Journal". Below the title, a subtitle reads: "An evolving informal guide on psychotherapeutic therapies created from training experience by psychiatry trainees from across Europe for psychiatry trainees and other curious people." A "Follow" button with a person icon is located on the left. The navigation bar includes "Featured", "Submitted", "Pages", and "People". Below the main content, there is a section for "Rational Emotive Behaviour Therapy" featuring a butterfly logo, the text "EFPT Psychotherapy Guidebook chapter on Rational Emotive Behaviour Therapy", and the date "Featured on June 18, 2017". A "Psychoeducation" section follows, also featuring a butterfly logo, the text "EFPT Psychotherapy Guidebook chapter on Psychoeducation", and the date "Featured on June 18, 2017". On the far right, there is a "Sort" dropdown menu with the option "Authorea".

- Several way of practicing Open Access
- Not only Open Access Gold : journal where you need to pay to publish (new publisher's strategy)
- Use Open Repository as HAL in France, Zenodo, etc.
- Don't forget the licences (Creative Commons, etc)

LICENSES	TERMS
  BY	 Attribution Others can copy, distribute, display, perform and remix your work if they credit your name as requested by you
  BY SA	 No Derivative Works Others can only copy, distribute, display or perform verbatim copies of your work
  BY NC	 Share Alike Others can distribute your work only under a license identical to the one you have chosen for your work
  BY NC SA	 Non-Commercial Others can copy, distribute, display, perform or remix your work but for non-commercial purposes only.







20/12/17

WIKIPEDIA
The free encyclopedia

Article Talk

Biscuit

From Wikipedia, the free encyclopedia

See also: [Biscuit \(bread\)](#) and [Cookie](#)

For other uses, see [Biscuit \(disambiguation\)](#).

Biscuit is a term used for a diverse variety of baked, commonly flour-based food products. The term is applied to two distinct products in North America and the Commonwealth of Nations and Europe. The North American biscuit is typically a soft, leavened quick bread, and is covered in the article [Biscuit \(bread\)](#). This article covers the other type of biscuit, which is typically hard, flat and unleavened.

Contents [edit]

- 1 Varieties in making
- 2 Etymology
- 3 History
- 3.1 Biscuits for travel
- 3.2 Confectionery biscuits
- 4 Biscuit today
- 4.1 Commonwealth of Nations and Europe

American biscuit (left) and one variety of British biscuit (right) – the American biscuit is typically hard, flat and unleavened.

[Edit](#) [View history](#) [Search](#)

<https://en.wikipedia.org/>

WIKIPEDIA
The free encyclopedia

Article Talk

Biscuit

From Wikipedia, the free encyclopedia

See also: [Biscuit \(bread\)](#) and [Cookie](#)

For other uses, see [Biscuit \(disambiguation\)](#).

Biscuit is a term used for a diverse variety of baked, commonly flour-based food products. The term is applied to two distinct products in North America and the Commonwealth of Nations and Europe. The North American biscuit is typically a soft, leavened quick bread, and is covered in the article [Biscuit \(bread\)](#). This article covers the other type of biscuit, which is typically hard, flat and unleavened.

Contents [edit]

- 1 Varieties in making
- 2 Etymology
- 3 History
- 3.1 Biscuits for travel
- 3.2 Confectionery biscuits
- 4 Biscuit today
- 4.1 Commonwealth of Nations and Europe

American biscuit (left) and one variety of British biscuit (right) – the American biscuit is typically hard, flat and unleavened.

[Edit](#) [View history](#) [Search](#)

<https://en.wikipedia.org/>

WORDPRESS.ORG

Showcase Themes Plugins Mobile Support Get Involved About Blog Hosting Download WordPress

Search WordPress.org

Meet WordPress

WordPress is open source software you can use to create a beautiful website, blog, or app.

European Federation of Psychiatric Trainees

Dashboard Posts Media Pages Comments Events Feedback

Add a New Post

Wordfence could not get an API key from the Wordfence scanning servers when it activated. You can try to fix this by going to the Wordfence settings page and clicking the 'Check for API key' button. If you keep seeing this error it usually means your WordPress server can't connect to our scanning servers. You can try adding your WordPress host to allow your WordPress server to connect to it.

Enter title here

Publish

Save Draft Publish

Status: Draft Edit

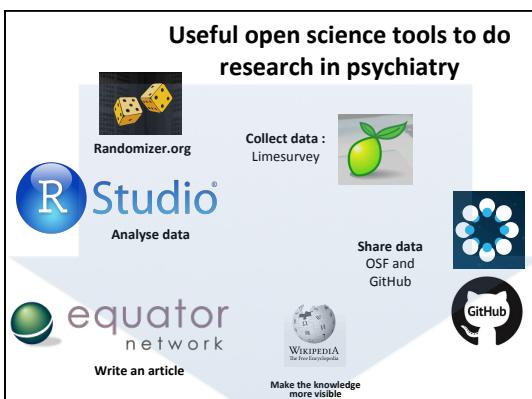
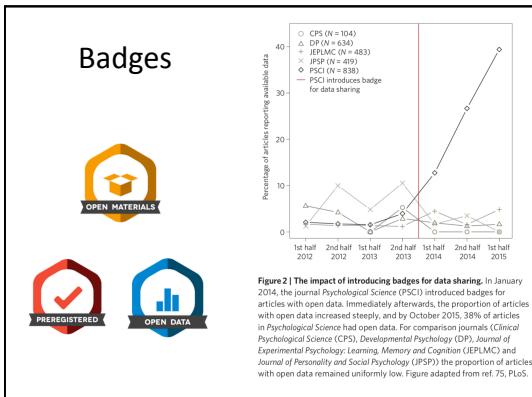
Validity: Public

Publish immediately

Postbody: Not Connected

close tags contact form published

[https://en.wikipedia.org/](#)



Base R

Cheat Sheet

Vectors			Programming				
Creating Vectors			For Loop				
<code>c(2, 4, 6)</code>	<code>2 4 6</code>	Join elements into a vector	<code>for (variable in sequence){</code> Do something	<code>while (condition){</code> Do something	<code>while (i < 5){</code> <code> print(i)</code> <code> i <- i + 1</code>		
<code>2:6</code>	<code>2 3 4 5 6</code>	An integer sequence	<code>}</code>	<code>}</code>	<code>}</code>		
Getting Help			While Loop				
Accessing the help files			Example				
<code>?mean</code>	Get help of a particular function.		<code>for (i in 1:4){</code> <code> j <- i + 10</code> <code> print(j)</code>	<code>while (i < 5){</code> <code> print(i)</code> <code> i <- i + 1</code>	<code>square <- function(x){</code> <code> squared <- x*x</code> <code> return(squared)</code>		
<code>help.search('weighted mean')</code>	Search the help files for a word or phrase.		<code>}</code>	<code>}</code>	<code>}</code>		
<code>help(package = 'dplyr')</code>	Find help for a package.						
More about an object			Vector Functions				
str(iris)			sort(x)				
Get a summary of an object's structure.			Return x sorted.				
class(iris)			rev(x)				
Find the class an object belongs to.			Return x reversed.				
Using Packages			unique(x)				
install.packages('dplyr')			See counts of values.				
Download and install a package from CRAN.			See unique values.				
library(dplyr)			Selecting Vector Elements				
Load the package into the session, making all its functions available to use.			By Position				
dplyr::select			<code>x[4]</code>				
Use a particular function from a package.			The fourth element.				
data(iris)			<code>x[-4]</code>				
Load a built-in dataset into the environment.			All but the fourth.				
Working Directory			<code>x[2:4]</code>				
<code>x[-(2:4)]</code>			Elements two to four.				
<code>x[c(1, 5)]</code>			<code>x[2:4]</code>				
By Value			All elements one and five.				
getwd()			<code>df <- read.table('file.txt')</code>				
Find the current working directory (where inputs are found and outputs are sent).			Input				
setwd('C://file/path')			Output				
Change the current working directory.			<code>write.table(df, 'file.csv')</code>				
Use projects in RStudio to set the working directory to the folder you are working in.			Description				
Named Vectors			Read and write a comma separated value file. This is a special case of read.table/write.table.				
Conditions			<code>load('file.RData')</code>				
<code>x['apple']</code>			<code>save(df, file = 'file.Rdata')</code>				
Element with name 'apple'.			Read and write an R data file, a file type special for R.				

Types

Converting between common data types in R. Can always go from a higher value in the table to a lower value.

<code>as.logical</code>	TRUE, FALSE, TRUE	Boolean values (TRUE or FALSE).
<code>as.numeric</code>	1, 0, 1	Integers or floating point numbers.
<code>as.character</code>	'1', '0', '1'	Character strings. Generally preferred to factors.
<code>as.factor</code>	'1', '0', '1' levels: 1, 0	Character strings with preset levels. Needed for some statistical models.

Maths Functions

<code>log(x)</code>	Natural log.	<code>sum(x)</code>	Sum.
<code>exp(x)</code>	Exponential.	<code>mean(x)</code>	Mean.
<code>max(x)</code>	Largest element.	<code>median(x)</code>	Median.
<code>min(x)</code>	Smallest element.	<code>quantile(x)</code>	Percentage quantiles.
<code>round(x, n)</code>	Round to n decimal places.	<code>rank(x)</code>	Rank of elements.
<code>signif(x, n)</code>	Round to n significant figures.	<code>var(x)</code>	The variance.
<code>cor(x, y)</code>	Correlation.	<code>sd(x)</code>	The standard deviation.

Variable Assignment

```
> a <- 'apple'
> a
[1] 'apple'
```

Matrices

`m <- matrix(x, nrow = 3, ncol = 3)`
Create a matrix from x.

<code>t(m)</code>	Transpose
<code>m %*% n</code>	Matrix Multiplication
<code>solve(m, n)</code>	Find x in: $m^{-1}x = n$
<code>tolower(x)</code>	Convert to lowercase.
<code>nchar(x)</code>	Number of characters in a string.

Lists

`l <- list(x = 1:5, y = c('a', 'b'))`
A list is a collection of elements which can be of different types.

`l[[2]]` `l[[1]]` `l$x` `l['y']`

<code>df</code> <- <code>data.frame(x = 1:3, y = c('a', 'b', 'c'))</code>	Second element of l.	New list with only the first element.	Element named x.	New list with only element named y.
---	----------------------	---------------------------------------	------------------	-------------------------------------

Also see the `dplyr` package.
A special case of a list where all elements are the same length.

List subsetting

1	x	y
2	a	<code>df\$x</code>
3	b	<code>df[[2]]</code>

Matrix subsetting

<code>df[, 2]</code>	<code>df\$y</code>	<code>head(df)</code>	<code>View(df)</code>	<code>summary(y ~ x, data=df)</code>
<code>nrow(df)</code>	Number of rows.	See the first 6 rows.	See the full data frame.	Generalised linear model.

List all variables in the environment.

Remove x from the environment.

Remove all variables from the environment.

Strings

Also see the `stringr` package.

`paste(x, y, sep = ' ')`
Join multiple vectors together.

`paste(x, collapse = ' ')`
Join elements of a vector together.

`grep(pattern, x)`
Find regular expression matches in x.

`gsub(pattern, replace, x)`
Replace matches in x with a string.

`toupper(x)`
Convert to uppercase.

`tolower(x)`
Convert to lowercase.

Factors

`factor(x)`
Turn a vector into a factor. Can set the levels of the factor and the order.

`cut(x, breaks = 4)`
Turn a numeric vector into a factor by 'cutting' into sections.

Statistics

<code>lm(y ~ x, data=df)</code>	Linear model.	<code>t.test(x, y)</code>	Test for a difference between means.
<code>glm(y ~ x, data=df)</code>	Generalised linear model.	<code>pairwise.t.test</code>	Perform a t-test for paired data.

`prop.test`
Test for a difference between proportions.

`anova`
Analysis of variance.

Distributions

	Random Variates	Density Function	Cumulative Distribution	Quantile
Normal	<code>rnorm</code>	<code>dnorm</code>	<code>pnorm</code>	<code>qnorm</code>
Poisson	<code>rpois</code>	<code>dpois</code>	<code>ppois</code>	<code>qpois</code>
Binomial	<code>rbinom</code>	<code>dbinom</code>	<code>pbinom</code>	<code>qbinom</code>
Uniform	<code>runif</code>	<code>dunif</code>	<code>punif</code>	<code>qunif</code>

Plotting

Also see the `ggplot2` package.

<code>plot(x, y)</code>	Values of x against y.
<code>hist(x)</code>	Histogram of x.

Dates

You can use the environment panel in RStudio to browse variables in your environment.

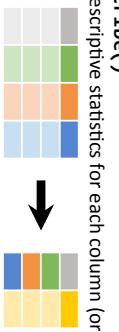
Summarize Data

`df['w'].value_counts()`
Count number of rows with each unique value of variable

`len(df)`
of rows in DataFrame.

`df['w'].nunique()`
of distinct values in a column.

`df.describe()`
Basic descriptive statistics for each column (or GroupBy)



pandas provides a large set of **summary functions** that operate on different kinds of pandas objects (DataFrame columns, Series, GroupBy, Expanding and Rolling (see below)) and produce single values for each of the groups. When applied to a DataFrame, the result is returned as a pandas Series for each column. Examples:

`sum()`
Sum values of each object.
`count()`
Count non-NA/null values of each object.
`median()`
Median value of each object.
`quantile([0.25,0.75])`
Quantiles of each object.
`apply(function)`
Apply function to each object.

`df.sum()`
Sum values of each object.
`df.count()`
Count non-NA/null values of each object.
`df.mean()`
Mean value of each object.
`df.var()`
Variance of each object.
`df.std()`
Standard deviation of each object.



Group Data

The examples below can also be applied to groups. In this case, the function is applied on a per-group basis, and the returned vectors are of the length of the original DataFrame.

`df.groupby(by="col")`
Return a GroupBy object, grouped by values in column named "col".

`df.groupby(level="ind")`
Return a GroupBy object, grouped by values in index level named "ind".

`df.rank(pct=True)`
Ranks rescaled to interval [0, 1].
`df.rank(method='min')`
Ranks. Ties get min rank.
`df.rank(method='dense')`
Ranks with no gaps.
`df.rank(method='max')`
Cumulative max.

`df.shift(1)`
Copy with values shifted by 1.
`df.shift(-1)`
Copy with values lagged by 1.
`df.cumsum()`
Cumulative sum.
`df.cummax()`
Cumulative max.

All of the summary functions listed above can be applied to a group.

Additional GroupBy functions:
`size()`
Size of each group.

Windows

`df.expanding()`
Return an Expanding object allowing summary functions to be applied cumulatively.
`df.rolling(n)`
Return a Rolling object allowing summary functions to be applied to windows of length n.

`df.plot.hist()`
Histogram for each column



Plotting

`df.plot.scatter(x='w',y='h')`
Scatter chart using pairs of points



Handling Missing Data

`df.dropna()`
Drop rows with any column having NA/null data.

`df.fillna(value)`
Replace all NA/null data with value.

`pd.qcut(df.col, n, labels=False)`
Bin column into n buckets.

`df.assign(Area=lambda df: df.Length*df.Height)`
Compute and append one or more new columns.

`df['Volume'] = df.Length*df.Height*df.Depth`
Add single column.

`pd.qcut(df.col, n, labels=False)`
Bin column into n buckets.

`pd.merge(adf, bdf, how='left', on='x1')`
Join matching rows from bdf to adf.

`x1 x2 x3`
`A 1 T`
`B 2 F`
`C 3 NaN`
`D NaN T`

`pd.merge(adf, bdf, how='inner', on='x1')`
Join data. Retain only rows in both sets.

`x1 x2 x3`
`A 1 T`
`B 2 F`
`C 3 NaN`
`D NaN T`

`pd.merge(adf, bdf, how='right', on='x1')`
Join data. Retain all values, all rows.

`x1 x2 x3`
`A 1 T`
`B 2 F`
`C 3 NaN`
`D NaN T`

`pd.merge(~adf.x1.isin(bdf.x1)]`
All rows in adf that do not have a match in bdf.

Filtering Joints

`x1 x2`
`A 1`
`B 2`

`adf[~adf.x1.isin(bdf.x1)]`
All rows in adf that do not have a match in bdf.

`ydf zdf`
`x1 x2`
`A 1`
`B 2`
`C 3`
`D 4`

`pd.merge(ydf, zdf, how='outer')`
Rows that appear in either or both ydf and zdf (Intersection).

`x1 x2`
`A 1`
`B 2`
`C 3`
`D 4`

`pd.merge(ydf, zdf, how='outer', indicator=True)`
• `query('_merge == "left_only"')`
• `drop(['_merge'], axis=1)`
Rows that appear in ydf but not zdf (Setdiff).

Combine Data Sets

`adfdiff`
`bdf`
`x1 x2`
`x1 x3`
`A 1`
`T`
`B 2`
`F`
`D 4`
`T`

`adfdiff`
`bdf`
`x1 x3`
`A 1`
`T`
`B 2`
`F`
`D 4`
`T`

`pd.merge(ydf, zdf, how='outer')`
Rows that appear in either or both ydf and zdf (Union).

`pd.merge(ydf, zdf, how='left_only')`
Rows that appear in ydf but not zdf (Setdiff).