Tidyverse R demo

Brown bag

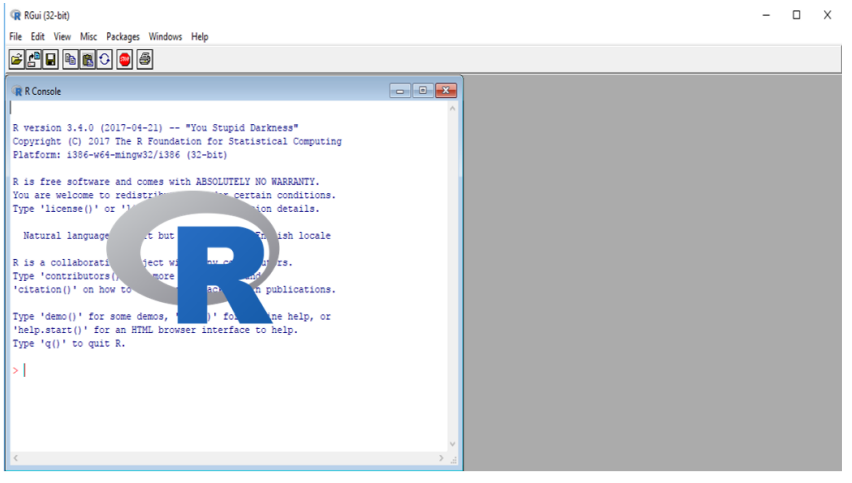
Abdou Khouakhi & Roel Tersteeg

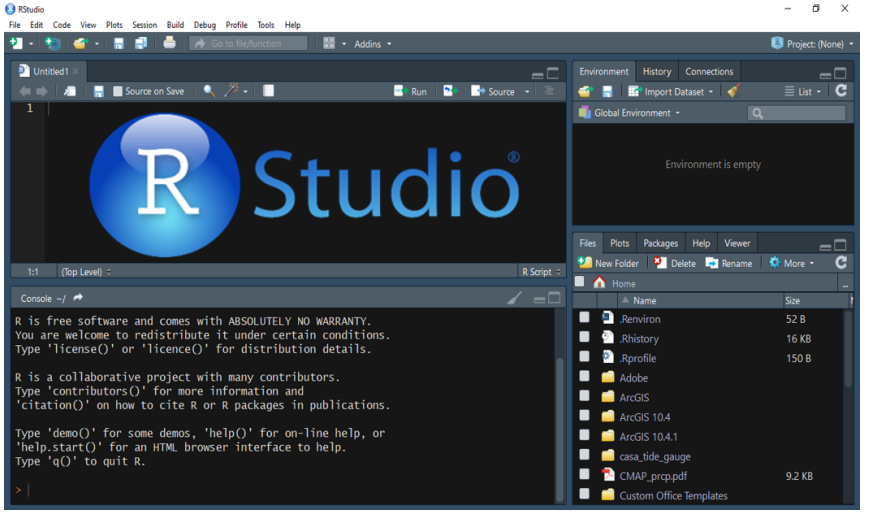
05 April, 2018, Loughborough University, ABCE

What is R?

- Language + environment for data analysis, statistical computing and graphics
- Free and open source
- Written by Ross Ahaka & Robert Gentleman in 1996 and extended by others
- An implantation of S language written by John Chambers and others
- *"The Most powerful statistical computing language in the planet"* according to the Developer of SPSS.

Difference between R and Rstudio





Difference between R and Rstudio

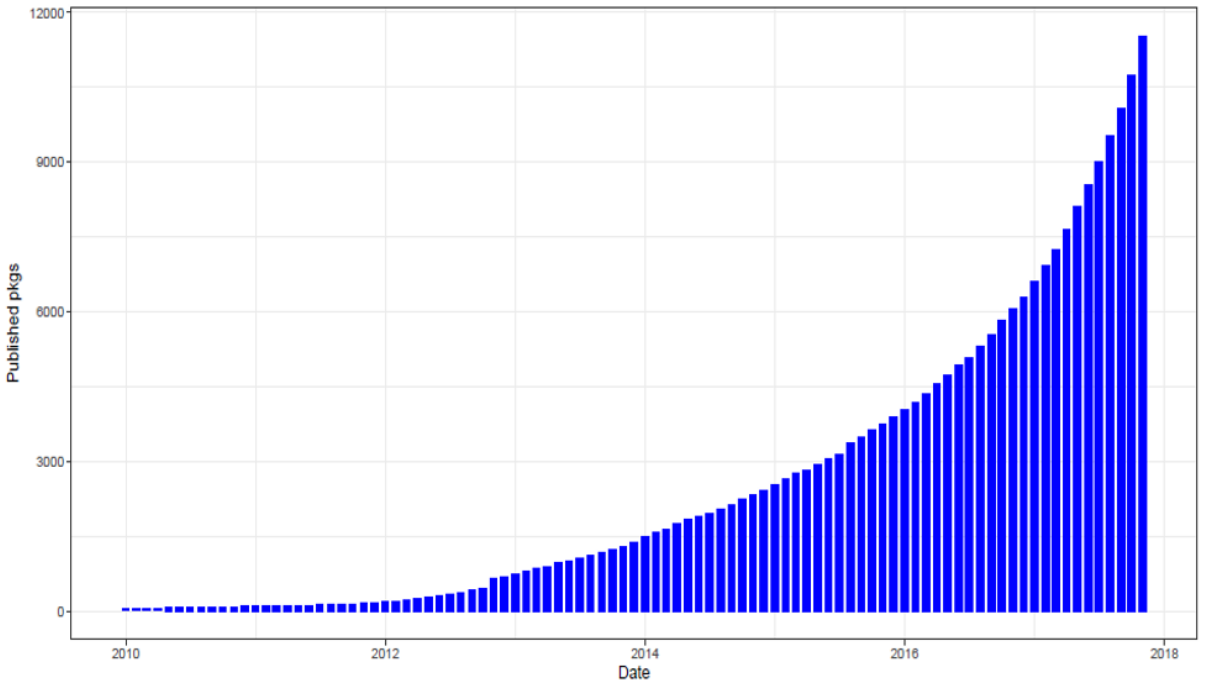
What makes R different than other statistical software (SPSS, STATA, SAS, Excel...)?

- **R** is free and open source!
- **R** comprises all of the capabilities of the above, and more
- **R** is powerful in graphics
- **R** has a large community
- **R** is good for developing algorithms
- **R** is great for reproducible research
- **R** can handle much larger datasets than Excel

Packages

https://cran.r-project.org/

A collection of functions, data, and documentation that extends the capabilities of R



Data types in R

R works with numerous data types. Some of the most basic types are:

- Decimals values are called numerics.
- Natural numbers called integers. Integers are also numerics.
- Boolean values (TRUE or FALSE) are called logical.
- Text (or string) values are called characters.

Data types in R

R works with numerous data types. Some of the most basic types are:

- Decimals values are called numerics.
- Natural numbers called integers. Integers are also numerics.
- Boolean values (TRUE or FALSE) are called logical.
- Text (or string) values are called characters.

R distinguishes between several types of objects:

- Vector
- Matrix
- Time series
- Data frames
- Lists
- Functions
- Graphics...

What can you do with R ?

- Programming
- Analytics (data science and machine learning)
- Graphics and visualizations (e.g. ggplot)
- Build apps/interactive graphics (e.g. shiny)
- Reporting/dynamic documents (e.g Rmarkdown)
- R can interact with APIs
- Interface with other programming languages (e.g. rpython)

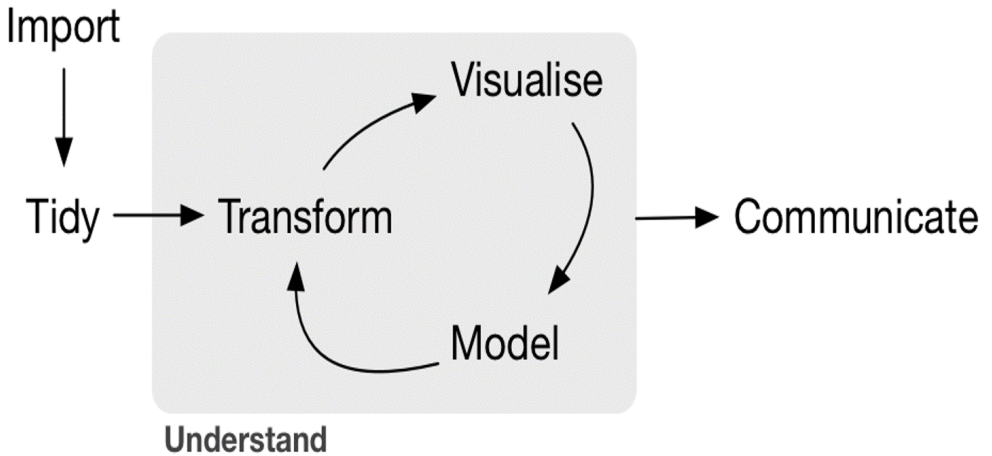
To learn more..

- There are entire books devoted to R (e.g. R cookbook, R for Everyone: Advanced Analytics and Graphics)
- Numerous (free) Web-based tutorials and user manuals (e.g. R for data science)
- The best way to learn R is through trial-and-error
- Embedded help, commands help()

Examples of online learning sites:

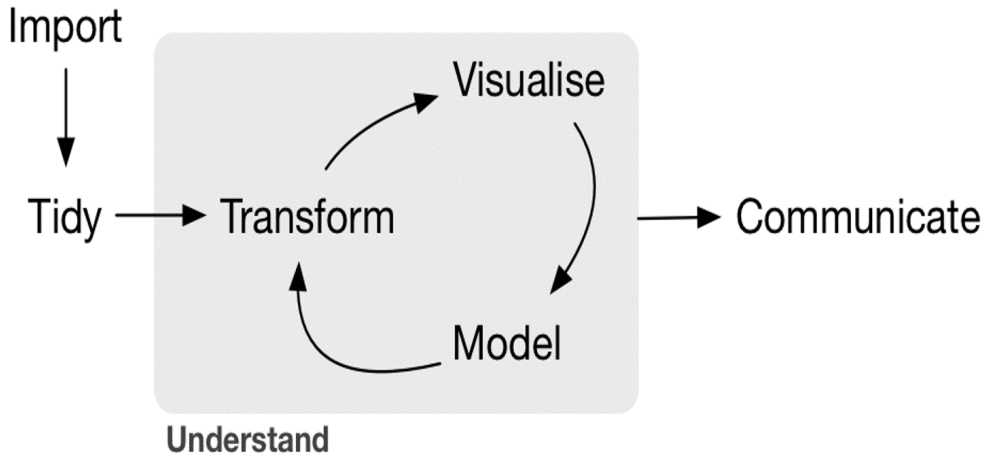
- **https://www.rstudio.com/resources/training/**
- **https://www.datacamp.com/courses/free-introduction-to-r**
- **http://swirlstats.com/**
- **http://www.statmethods.net/**
- **https://stackoverflow.com/**

R demo



Data Science Pipeline Grolemund and Wickham

R demo



Data Science Pipeline Grolemund and Wickham

Data sets: Time series of tropical cyclones

- Import the data
- Check the data structure
- Summary statistics
- Data wrangling (select, filter, arrange…)
- Data visualization

Data Wrangling

dplyr package functions/ verbs

```
- mutate() : Add new columns (or overwrite old one)
- filter() : subset rows
- select() : subset columns
- arrange() : order rows
- summarise() : summarise rows
- group_by() : grouping elements used with summarise()
```

GGplot2

- Concept of ‘Grammar of graphics’
- Build-in statistics (e.g. for regression lines and histograms)
- Variety of ways of building a graph
- For quick graphs and yet highly adaptable

GGplot2

- Concept of ‘Grammar of graphics’
- Build-in statistics (e.g. for regression lines and histograms)
- Variety of ways of building a graph
- For quick graphs and yet highly adaptable

Basic graph

```
ggplot(data = , aes(x = , y = )) +
        geom_point()

p1 <- ggplot(data = , aes(x = , y = ))
        p1 + geom_point()

ggplot() +
        geom_point(data = , aes(x = , y = ))
```

Adding layers

```
ggplot() +
        geom_point(data = , aes(x = , y = )) +
```

- Add layer with different data
- Add theme objects, e.g. labels
- Change the scale and scale type
- Add/change legend
- Change coordinate system