

R for Data Science

R for Data Science workshop

2019-05-01 (updated: 2019-07-15)

R for Data Science

Everyone please do the following now if you haven't already:

1. Sign up for a free account at rstudio.cloud
2. Visit <https://tinyurl.com/kcdc2019-r-workshop> to join our shared workspace
3. Open the r-for-data-science project
4. Save a Permanent Copy of the project

TEMPORARY



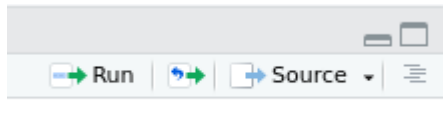
Save a Permanent Copy



Your turn

In the `r-for-data-science` project open the following file:

`eat-cake-first.R`



Start with the cursor on the first line, and then run the code by repeatedly pressing the "Run" button at the top right of the editor pane.

Note: In RStudio, "Run" executes the current line or selection, while "Source" executes the contents of the whole file.

You should see a graph after pressing "Run" 4 times.

Try changing one or both of the countries to any of the following:

Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Poland, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States

Welcome



Email: davidbody@bigcreek.com

Twitter: [@david_body](https://twitter.com/david_body)

Welcome

Workshop goals

- Introduce R, RStudio, and the Tidyverse
- Learn enough to begin using R for real work
- Overview of what is available
- Have fun

Welcome

Introductions

- Your name
- Why are you here?
- One other thing about you

Welcome

Outline

- Introduction to R, RStudio, RStudio Cloud
- Visualizing data with ggplot2
- Data wrangling with dplyr and tidyr
- [lunch?]
- Exploratory data analysis
- R programming
- R Markdown
- Shiny

Logistics

- Breaks
- Lunch
- Rest rooms

Sponsors

- Visit their booths
- Thank them

<https://r4ds.had.co.nz/>

O'REILLY®



R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Golemund

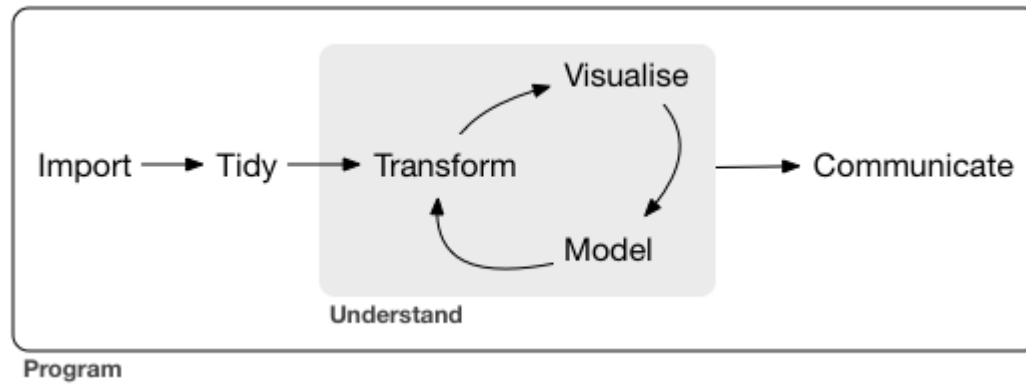


Image source: [R for Data Science](#) by Hadley Wickham & Garrett Golemund.

R, RStudio, & RStudio Cloud

R is a language

RStudio is an IDE (also a company)

RStudio Cloud is RStudio (the IDE) hosted by RStudio (the company)

What is R?

From the official R website:

R is a **free software environment** for **statistical computing** and **graphics**. It compiles and runs on a wide variety of **UNIX** platforms, **Windows** and **MacOS**.

- general purpose programming language
- large standard library with strong support for
 - data wrangling
 - data analysis & statistical computations
 - plotting and data visualization
- large number of contributed libraries
- written in R, C++, and Fortran
- conceived in 1992 as an open source replacement for the S language. Initial release in 1995, version 1.0 in 2000, version 3.6 in 2019

RStudio IDE

- Free and open source integrated development environment for R
- RStudio Desktop (Windows, MacOS, Linux)
- RStudio Server (via browser)
- RStudio Cloud (hosted by RStudio, the company)

RStudio (the company)

- Founded by J.J. Allaire (creator of Cold Fusion)
- Chief Scientist is Hadley Wickham
- RStudio has contributed to many packages, including
 - Tidyverse - R packages for data science
 - Shiny - web framework for R
 - RMarkdown - combine R code with markdown
 - knitr - generate dynamic reports
 - packrat - dependency management
 - devtools - package development tools
- rstudio::conf January 27-30, 2020, San Francisco

Session videos for past conferences are available at rstudio.com

Tidyverse

www.tidyverse.org

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Core tidyverse packages:

- ggplot2
- dplyr
- tidyr
- readr
- purrr
- tibble
- stringr
- forcats

05 : 00

Your turn

RStudio (Cloud)

Project overview

How to get help

R Markdown notebook

[your-turn/00-introduction.Rmd](#)