

Intro to R

Welcome to class!

1. Introductions
2. Class overview
3. Getting R up and running



Photo by [Belinda Fewings](#) on Unsplash

About Us

Carrie Wright

Assistant Scientist, Department of Biostatistics, JHSPH

PhD in Biomedical Sciences

Email: cwright60@jhu.edu

Website: carriewright11.github.io



About Us

Ava Hoffman

Research Associate, Department of Biostatistics, JHSPH

PhD in Ecology

Email: ava.hoffman@jhu.edu

Website: avahoffman.com



About Us

Marta Karas

PhD Candidate, Department of Biostatistics, JHSPH

MS in Mathematics

Email: mkaras2@jhmi.edu

Website: martakarass.github.io



What is R?

- R is a language and environment for statistical computing and graphics
- R is the open source implementation of the S language, which was developed by Bell laboratories in the 70s.
- The aim of the S language, as expressed by John Chambers, is “to turn ideas into software, quickly and faithfully”



(source: <http://www.r-project.org/>,
[https://en.wikipedia.org/wiki/S_\(programming_language\)](https://en.wikipedia.org/wiki/S_(programming_language)),
https://en.wikipedia.org/wiki/Bell_Labs)

What is R?

- In 1991 Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand began developing R
- R is named partly after the first names of the first two authors and a play on the name of S.
- R is both open source and open development



(source: <http://www.r-project.org/>,
[https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language)))

Why R?

- High level language designed for statistical computing
- Powerful and flexible - especially for data wrangling and visualization
- Free (open source)
- Extensive add-on software (packages)
- Strong community



(source: <https://rladies-baltimore.github.io/>)

Why not R?

- Fairly steep learning curve
 - “Programming” oriented
 - Minimal interface
 - Little centralized support, relies on online community and package developers
 - Annoying to update
 - Slower, and more memory intensive, than the more traditional programming languages (C, Java, Perl, Python)



Introductions

What do you hope to get out of the class?

Why do you want to use R?



Photo by [Nick Fewings](#) on [Unsplash](#)

Course Website

http://jhudatascience.org/intro_to_r

Materials will be uploaded the night before class



Learning Objectives

- Reading data into R
- Recoding and manipulating data
- Writing R functions and using add-on packages
- Making exploratory plots
- Understanding basic programming syntax
- Performing basic statistical tests

Installing R

- Install the latest version from: <http://cran.r-project.org/>
- [Install RStudio](#)

Collection of R packages

We have an R package called jhur that will make sure all the packages are installed.

You can just copy and paste the below code into your console - we'll explain what it all means in the next day or two

```
install.packages("remotes")
remotes::install_github("muscchelli2/jhur")
```

Note it may take ~5-10 minutes to run.

Useful (+Free) Resources

- R for Data Science: <http://r4ds.had.co.nz/>
- Various “Cheat Sheets”: <https://www.rstudio.com/resources/cheatsheets/>
- Dataquest: <https://www.dataquest.io/>
- R reference card: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
- UCLA Institute for Digital Research and Education:
<http://www.ats.ucla.edu/stat/r/>
- Quick R: <http://statmethods.net/>
- Open Case Studies: <https://www.opencasestudies.org/>
- Tidyverse Skills for Data Science Book:
<https://jhubdatascience.org/tidyversecourse/>
- Tidyverse Skills for Data Science Course:
<https://www.coursera.org/specializations/tidyverse-data-science-r>
- Roger Peng or Jeff Leek Courses on Coursera: <https://www.coursera.org/>
- Course Website: http://jhubdatascience.org/intro_to_r/index.html