

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

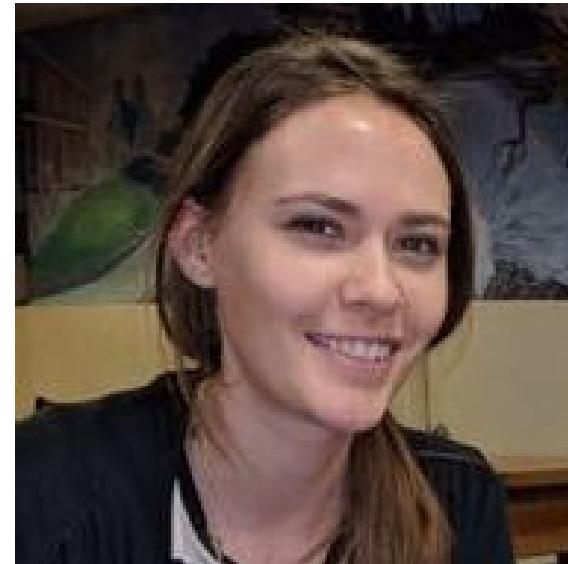
Carrie Wright

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

Candace Savonen

Research Associate, Department of Biostatistics, JHSPH

Masters in Neuroscience

Former Data Analyst for Childhood Cancer Data Lab

Email: csavone1@jhu.edu

Website: <https://www.cansavvy.com/>



About Us - TAs

Grant Schumock

PhD Candidate, Department of Biostatistics, JHSPH

BS in Nuclear Engineering

Email: gschumo1@jhmi.edu



About Us - TAs

Qier Meng

ScM Student, Department of Biostatistics, JHSPH

Bachelor's Degree in Mathematics

Bachelor's Degree in Neuroscience

Email: qmeng11@jhmi.edu



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
- Using add-on packages
- Making exploratory plots
- Understanding basic programming syntax
- Performing basic statistical tests
- Writing R functions

Course Format

- Lecture with slides (possibly “Interactive”)
- Lab/Practical experience
- Two 10 min breaks each day - timing may vary
- Jan 10-21, 2022, 8:30AM-11:50AM on Zoom
- No class on Jan 17th for Martin Luther King Jr. Day

CoursePlus

CoursePlus:

<https://courseplus.jhu.edu/core/index.cfm/go/syl:syl.public.view/coid/16733/>

Grading

1. Attendance/Participation: 20% - this can be asynchronous - just some sort of interaction with the instructors/TAs (turning in assignments, emailing etc.)
2. Homework: 3 x 15%
3. Final "Project": 35%

Homeworks and Final Project due by *Wednesday, Jan 26, 2022 at 11:59pm EST.*

If you turn homework in earlier this can allow us to potentially give you feedback earlier.

Note: Only people taking the course for credit must turn in the assignments. However, we will evaluate all submitted assignments in case others would like feedback on their work.

Installing R

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

RStudio is an integrated development environment (IDE) that makes it easier to work with R.

More on that soon!

Getting files from downloads

This course will involve moving files around on your computer and downloading files.

If you are new to this - check out these videos

If you have a PC: <https://youtu.be/we6vwB7DsNU>

If you have a Mac: <https://www.youtube.com/watch?v=Ao9e0cDzMrE>

Basic terms

- **Package** - a package in R is a bundle or “package” of code (and or possibly data) that can be loaded together for easy repeated use or for **sharing** with others.

Packages are sort of analogous to a software application like Microsoft Word on your computer. Your operating system allows you to use it, just like having R installed (and other required packages) allows you to use packages.

- **Function** - a function is a particular piece of code that allows you to do something in R. You can write your own, use functions that come directly from installing R, or use functions from additional packages.

A function might help you add numbers together, create a plot, or organize your data. More on that soon!

Tidyverse and Base R

We will mostly show you how to use tidyverse packages and functions.

This is a newer set of packages that can make your code more intuitive or readable.

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

- Course Website: http://jhudatascience.org/intro_to_r/index.html

Want more?

- Tidyverse Skills for Data Science Book:
<https://jhudatascience.org/tidyversecourse/>
- Tidyverse Skills for Data Science Course (can get certificate):
<https://www.coursera.org/specializations/tidyverse-data-science-r>
- R for Data Science: <http://r4ds.had.co.nz/> - Open Case Studies:
<https://www.opencasestudies.org/> - Dataquest: <https://www.dataquest.io/>

Need help?

- Various "Cheat Sheets": <https://www.rstudio.com/resources/cheatsheets/>
- R reference card: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
- R terminology: <https://cran.r-project.org/doc/manuals/r-release/R-lang.pdf>

Interested in Reproducibility? Check out Candace's courses at:

- https://jhudatascience.org/Reproducibility_in_Cancer_Informatics/ and
- https://jhudatascience.org/Adv_Reproducibility_in_Cancer_Informatics/