"Introduction to R" series

Aug.30: Part 1 – Data Wrangling

Sep.06: Part 2 – Data Visualization

Sep.13: Part 3 – Data Analysis

Sep.20: Part 4 – Real-world Data Analysis Using R

National Institute of Allergy and Infectious Diseases

Yuyan Yi, PhD **Mina Peyton, PhD**

Bioinformatics and Computational Biosciences Branch

Bioinformatics and Computational Biosciences Branch (BCBB)

- **My background:** Biomedical Statistical Specialist in the Bioinformatics and Computational Biosciences Branch in the Office of Cyber Infrastructure and Computational Biology

- Check us out: https://www.niaid.nih.gov/research/bioinformatics-computational-biosciences-branch
- Do you conduct research involving genomics (clinical, meta and microbial), imaging, data science and biostatistics, or structural biology?
  - We offer collaboration opportunities to address research questions at no direct cost to the NIAID research community and collaborators

What is R?

- Created by Ross Ihaka and Robert Gentleman – the name derived from the first letter of each of their names
- First released in March of 1995
- More than 2 million people use the programming language worldwide
- Over 18,000 packages in CRAN repository that do various tasks and analyses
- FOSS (free and open-source) software

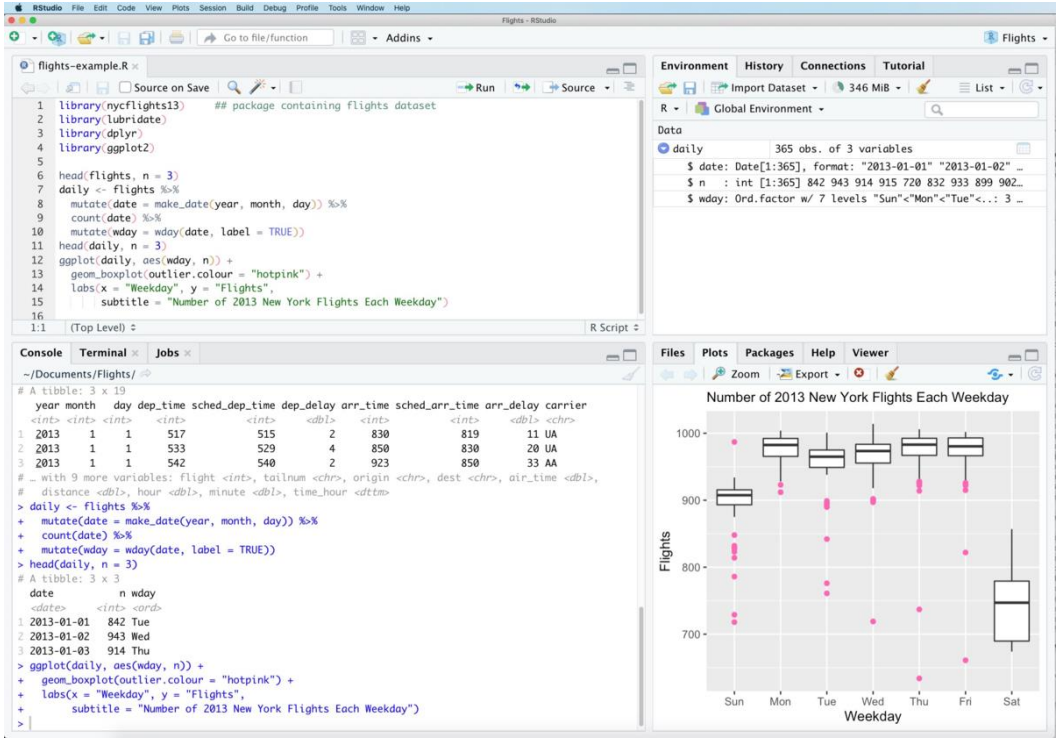
Why learn R?

- Top 20 programming language according to TIOBE index and top 10 language in PYPL list (Aug 2024)
- Significant usage within biomedical research community as well as a high demand skill from employers
- It is specially suited for data analysis, statistics, visualization, machine learning, and other tasks within data science
- Useful tool especially with complement of another language like python if focused on data science as a career path

How does one use R?

- Can be used on the command line
- Better used within an IDE (integrated development environment) such as Rstudio
- Load data, install R packages, and then interactively explore, visualize, analyze, and report results

RStudio



National Institute of Allergy and Infectious Diseases

RStudio - Wikipedia. (2024). Retrieved 26 August 2024, from https://en.wikipedia.org/wiki/RStudio

NIAID

Where to find R?

- Within NIAID, you can use self-service:
  - Windows Software Center: https://inside.niaid.nih.gov/it-equipment/installing-scientific-software-using-software-center
  - Mac self-service: https://inside.niaid.nih.gov/it-equipment/installing-scientific-software-mac-self-service
- Within NIH, you can use the Biowulf HPC cluster: https://hpc.nih.gov/
- Download and install R and Rstudio:
  - R: https://www.r-project.org/
  - Rstudio: https://www.rstudio.com/

Tutorial overview

- I will provide a tutorial of using R in the next part of the seminar
- Next slides will cover the pre-requisites and the set-up if you would like to follow along
- I will keep advancing during the tutorial to ensure we complete it and will try to keep a reasonable pace
- If you simply want to listen and complete the tutorial later or if you get stuck, no worries…
  - You can go back to the material and complete later
  - Goal today is to introduce the material and demonstrate its use

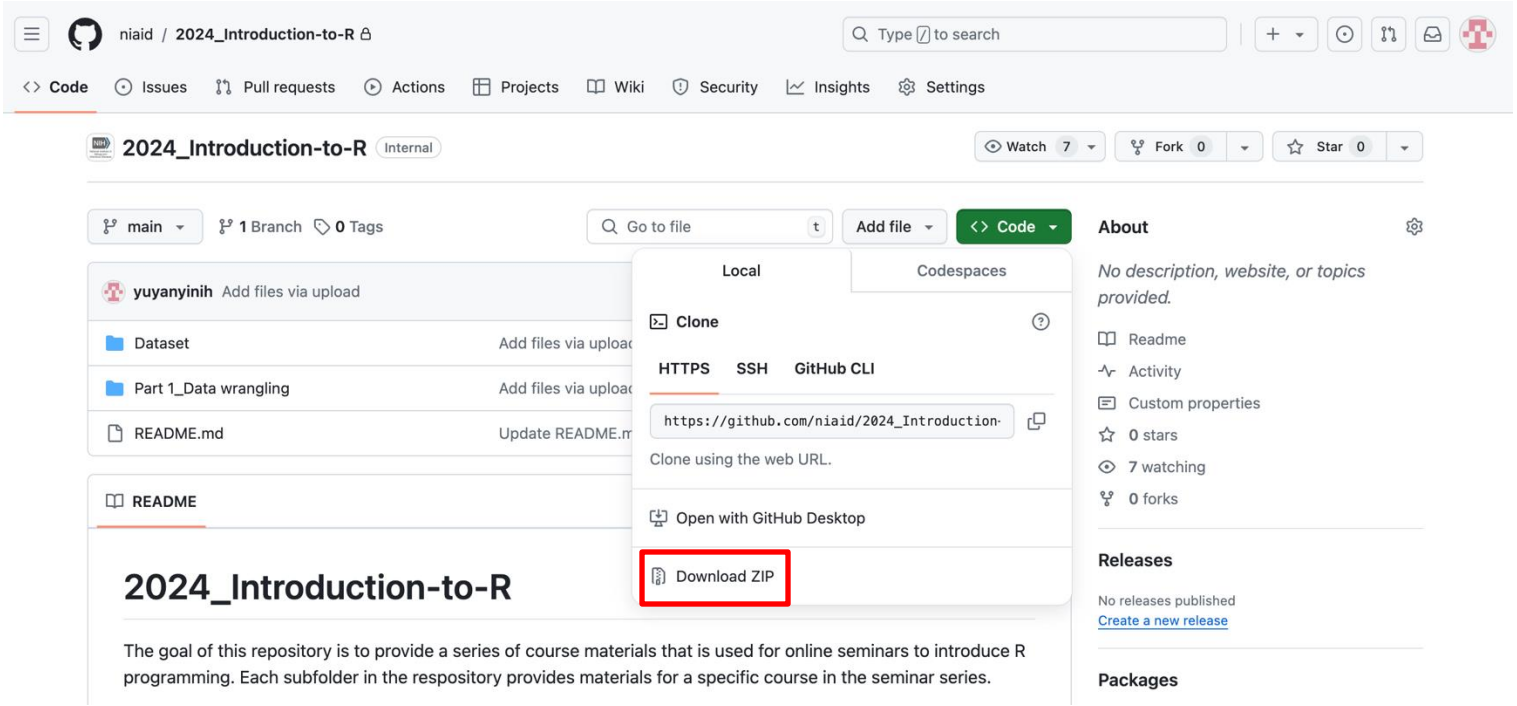
Pre-requisites for tutorial

- You will need to have R and RStudio installed, and the course materials downloaded from GitHub
  - R version 4.2.1+ (https://cran.r-project.org/)
  - Rstudio (https://www.rstudio.com/products/rstudio/)
  - Location of course materials: https://github.com/niaid/2024_Introduction-to-R

- If within NIAID, consider self-service for R and Rstudio install:
  - Windows self-service: https://inside.niaid.nih.gov/it-equipment/installing-scientific-software-using-software-center
  - Mac self-service: https://inside.niaid.nih.gov/it-equipment/installing-scientific-software-mac-self-service

- We will be using the R script file (2024_Part 1_Intro to R_Data wrangling.R)

Tutorial Set Up



National Institute of Allergy and Infectious Diseases

https://github.com/niaid/2024_Introduction-to-R.git

NIAID

Tutorial Set Up

- Unzip the downloaded file in your downloads folder
- Open the subfolder “Part 1_Data wrangling” and then the “2024_Part 1_Intro to R_Data wrangling.R” file
- If Rstudio is installed, then this should open the project in RStudio