"Introduction to R" series

Aug.30 1-3pm: Part 1 – Data Wrangling
Sep.06 1-3pm : Part 2 – Data Visualization
Sep.13 1-3pm : Part 3 – Data Analysis
Sep.20 1-3pm : 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

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/

Part 2: Data Visualization Review

- Materials: https://github.com/niaid/2024_Introduction-to-R/tree/main/Part%202_Data%20visualization
- Recording: https://bioinformatics.niaid.nih.gov/programming/55.2.1

- Base R plotting system
  - https://r-charts.com/base-r/
  - Base R graphics functions can be customized adding legends, texts, grids, modifying the axes, among other functions.
- Package ggplot2
  - https://ggplot2.tidyverse.org/
- Three good places to start with different level and purpose:
  - The Data Visualization and Communication chapters in R for Data Science.
  - Make common graphics as quickly as possible: The R Graphics Cookbook by Winston Chang.
  - If you've mastered the basics and want to learn more, read ggplot2: Elegant Graphics for Data Analysis.

Part 3: Data Analysis Overview

- Correlation
- Principal component analysis
- Statistical Test
  - One/Two sample t-test, Paired t-test,
  - One-way ANOVA, Two-way ANOVA, Pairwise t-test
  - Fisher's test, Chi-square test
- Statistical Model Analysis:
  - Linear regression model
  - Logistic regression model
- Other reference sources:
  - R Tutorials for Applied Statistics

Tutorial Set Up

- Download file from Github link or attachment:

https://github.com/niaid/2024_Introduction-to-R/tree/main/Part%203_Data%20Analysis

- Open the subfolder “Part 3_Data analysis” and then the “2024_Part 3_Intro to R_Data analysis.Rmd” file

- If Rstudio is installed, then this should open the project in RStudio