Lab 1 - R-Basics ; PSY 503, Fall 2025

Please upload to canvas/ e-mail me the link to your project. You can work together in pairs if you wish. In this case, please let me know who you worked with. Use comments such as #1, #2, #3, and #4 to indicate the beginning of each task. Use comments to add verbal answers where needed as well.

---

(0) Create a new project. Create a new file Lab_1.r *(or a .rmd/.qmd file if you are familiar with R-markdown/Quarto)* in which you write the code to do the following.

(1) Install and load the following packages

- coursekata
- tidyverse
- datasauRus (among other things loading this package makes the dataframe *datasaurus_dozen* available in its namespace)

(2) Display the overall structure of this dataframe, using str(). Examine the data, and explain what you believe this dataframe is representing. What are the types of variables represented here (i.e. Nominal, Ordinal, Interval, Ratio)?

    - **Hint:** *the data is tidy which means that*
        - *(i) Each column is a variable*
        - *(ii) Each row is an observation (also called a case, research unit, or an object; all these words mean the same thing)*

- print all the values of the variable/column titled *dataset*

- next, run the following command: print (unique(lab1_data$dataset)). What do you believe is being printed?

(3) Select a subset of lab1_data which has the value of *"dataset"* equal to "circle", and save it in a new variable temp_data
(refer to section 2.9 of courseKata)

- draw a scatterplot of it (y against x), using ggplot
- ***or***
- gf_point which is an abstraction over ggplot*(refer to section 4.2 of CourseKata)*

Attempt this for a few other values of the variable *"dataset"*. What do you observe?