Introduction to R: Review

Julia Gallucci

Data Science Institute, University of Toronto

2023-12-16

Review

- We will be working in groups and using the tools we've learned to explore an open access Toronto data set!

- Join a breakout room! Jessie and I will be there to help with any errors or complications you may have.

- Once complete, we will rejoin the main room and each group will present what they've done!

Review

Choose a dataset from Open Data Toronto
(https://open.toronto.ca/)

Your job is to do the following,

▶ **Load in the data** use the appropriate read_*() function

▶ **Explore the data** take a look at your data!

(recall: glimpse() or head() function) what types are
present in the dataset ?

what are the summary statistics of your data? (recall:
summary() function)

Review

▶ **Clean your data**

convert variables to the appropriate type if applicable (i.e., if your dataset has 1/2 representing Male/Female explicitly coerce from numeric to factor)

remove any NA values from your data

▶ **Manipulate your data**

use functions such as `filter()`, `arrange()`, `sort()`, `group()`, `mutate()`

recall the %>% operator to combine multiple commands!

▶ **Wrangle your data**

try out `pivot_*()` functions where applicable to tidy your data!

Review

▶ **Loop over your data**

create for loop to loop over an aspect of the columns in your dataset (i.e., mean of each column) (recall `map()` function)

▶ **Save your work**

in an RMarkdown file to present to the class. Be sure to follow best coding practices (i.e., commenting your code, good use of variable names, lines/spaces etc.)