

STA130 Winter 2020

(Materials used in this presentation are provided by the U of T Statistical Sciences Department.

This presentation was prepared by Vivian Ngo.)

[Github.com/vivianngo97/STA130-Winter-2020](https://github.com/vivianngo97/STA130-Winter-2020)

viv.ngo@mail.utoronto.ca

Agenda

- Vocabulary
- R Tutorial
- Group discussion
- Writing Activity
- Mentorship Program

Vocabulary

- Cleaning data
- Tidy data
- Removing a column
- Extracting a subset of variables
- Filtering the data frame based on a condition (e.g. based on the values in one or more of the variables/columns)
- Sorting data based on the values of a variable
- Renaming the variables
- Grouping categories
- Defining new variables
- Producing new data frames
- Handling missing values (NAs)
- Creating summary tables

Group Discussion

- Why is data visualization so important?
- Why does your audience matter? (Think about what message you want to portray and the types of data/ visualizations that you'll use)
- Why is it important for data visualizations to be intuitive? How can you ensure your figures are intuitive to the intended audience?
- What might happen to a data visualization project if you failed to clean the data?

Writing Activity

- Q1. What questions, if any, do you have so far regarding the course materials?
- Q2. What is one of your favorite things about tutorial?
- Q3. What is one of your least favorite things about tutorial? (Other than it being on a Friday!)
- Q4. What is the thing you are most looking forward to for the mentorship program?