2. 1D Data Visualization

Prepared by Prajwal Singh

Data Acquisition and Inspection

- Read only the Gender and Height data columns from the attached Excel file.
    - *df = pd.read_excel(<filename>, usecols=<list of columns>)*
- You must install openpyxl package before using Pandas read_excel method.
    - (ex: pip install openpyxl)
- Check the exact name of the Height data column.
- You may use Pandas function to change the column names to shorten them for convenience.

Data Cleaning

- The gender column data may have some non-uniformity on the name.
    - For example: "Male" may appear as "male" in some rows. Use Pandas function to resolve that problem.

Data Visualization

- Use either of Plotly Express, Seaborn to create the following distribution plots
    - Histogram plot
    - Box plot
    - Violin plot

Note that the plots must show distributions of the Male and Female heights separately (may overlap) on the same plot.

(This dataset contains the height, weight and 4 fingerprint measurements (length, width, area and circumference) collected from 200 participants. This data was collected to perform a regression analysis to assess whether a significant relationship exists between fingerprint size and physical stature. The dataset was downloaded from Here)