BIOS-584 Python Programming (Non-Bios Student)

Week 09

Instructor: Tianwen Ma, Ph.D.

Department of Biostatistics and Bioinformatics,

Rollins School of Public Health,

Emory University

Lecture Overview

- Pandas Tutorial 9 (week-09-pandas.ipynb)
- Debugging in PyCharm

Pandas

- A data manipulation package in Python for tabular data
  - In the form of rows and columns, known as DataFrames
- Functionality includes
  - Data transformation
  - Summary Statistics
  - Data merging
  - Integration with other data science packages, including NumPy, Matplotlib, Seaborn, Plotly, and Scikit-learn

What is Pandas used for?

- Import datasets from databases, spreadsheets, comma-separated values (CSV) files, etc.

- Clean datasets, i.e., handling missing values

- Tidy datasets by reshaping the structure into a suitable format prior to analysis.

- Aggregate data by calculating summary statistics.

- Visualize datasets and uncover hidden patterns.

Pandas Tutorial

- I will go over the Jupyter notebook today.

Debugging in PyCharm

- https://www.jetbrains.com/help/pycharm/debugging-your-first-python-application.html#summary
- Find Debug_example.py