BIOS-584 Python Programming (Non-Bios Student)

Week 01 – Introduction and Course Overview

Instructor: Tianwen Ma, Ph.D.

Department of Biostatistics and Bioinformatics,

Rollins School of Public Health,

Emory University

Lecture Overview

- Introduction
- Motivation
- Class Logistics
- Computing Setup

Introduction

Introduction

✉ ma3tian1wen2@emory.edu

- Research Assistant Professor in BIOS
- Originally from Nanjing, China
- PhD, MS, and BS all from the University of Michigan Ann Arbor (Go Blue!)
- Currently serve as a leading biostatistician at Emory Brain Health Center (BHC)

Introduction

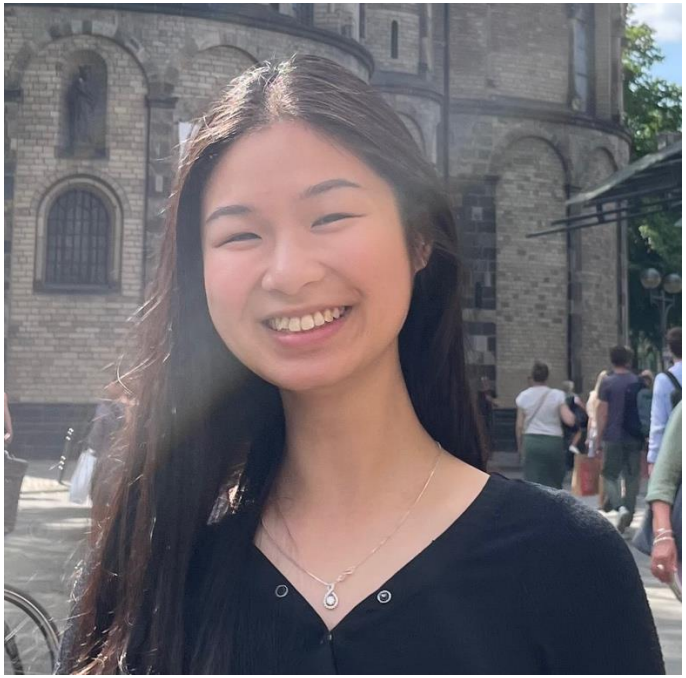
✉ ma3tian1wen2@emory.edu

- I bring Coco to campus every day (Don’t tell Rollins staff. Don’t worry and the department knows.).
- She is sweet and cute.
- She will be with you in class (sleeping somewhere in the back).
- You can pet her after class or during OH. She likes people very much!

My Teaching Style

- This is my first time to teach coding class targeting non-BIOS students, so fresh experience for both you and me.
- Hands-on activities help you learn better.
- Questions and discussions are encouraged.
- Your feedback helps me improve my teaching.

Teaching Assistant(s)



- Madeline Nguyen
- quynh.anh.nguyen@emory.edu
- Office Hours: 3-4 PM on Tuesday, GCR-359

Teaching Assistant(s)

- TA(s) will hold office hours and answer questions
- TA(s) will also be involved with grading assignment and quizzes.
- Feel free to ask questions during class, office hours, on Canvas, or via email.

Lab Sessions

- We have a two-hour lecture and one-hour online lab session each week.
  - Currently, 4-5 PM on Thursday
- I will assign additional problem sets to you on Datacamp (free with edu email address).

Lab Sessions

- I will go over common mistakes from HW and additional practices.
  - Send me questions ahead of time.
- Sometimes, I will review necessary mathematical and statistical concepts before we enter certain modules.
- Otherwise, it will become a regular office hour for me. Feel free to stop by my office or jump on the zoom link.

Office Hours

- What office hours are meant for:
  - Apply tools in practice
  - Discuss relevant issues in HW assignments
  - Improve your knowledge in data science

- What office hours are NOT meant for:
  - Solve the problem for you

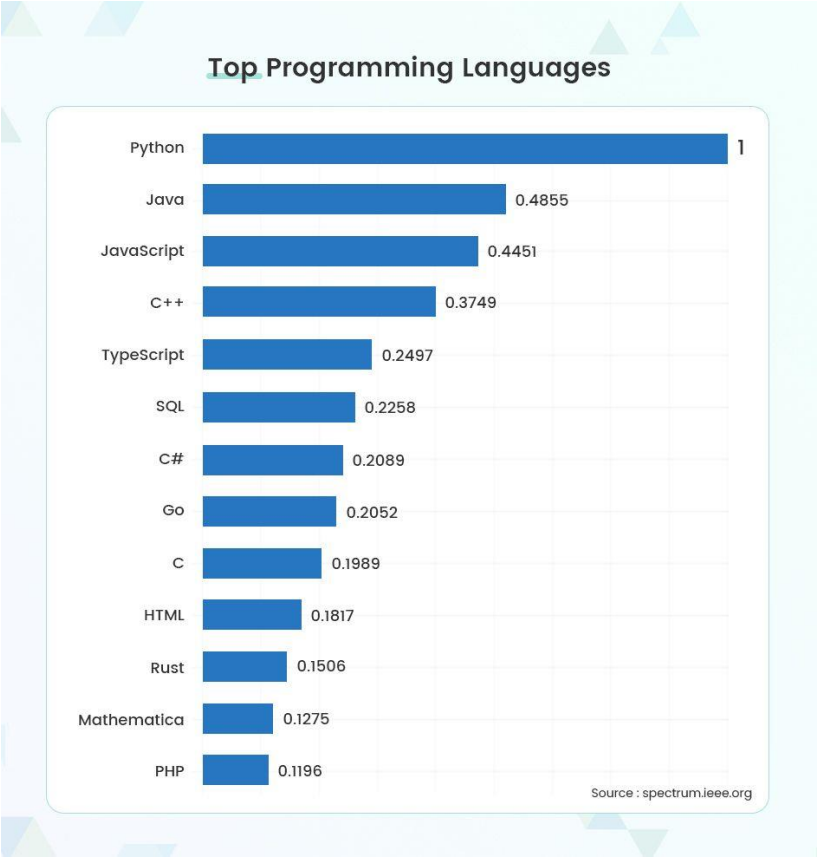
Class Etiquette

- Coding can be tough and make you feel frustrated.
- If the pace is too fast, let us know.
- Everyone comes from different backgrounds. Some sessions might be more engaging than others. If you already knew, helps others or explore new data science areas.
- Always be respectful to each other

Motivation

Why Python?

- The world's most popular programming language



Why Python?

- Great community and easy to learn
- Comprehensive packages and tutorials on machine learning and deep learning
- Higher efficiency than R even for statistical computing

- Java:

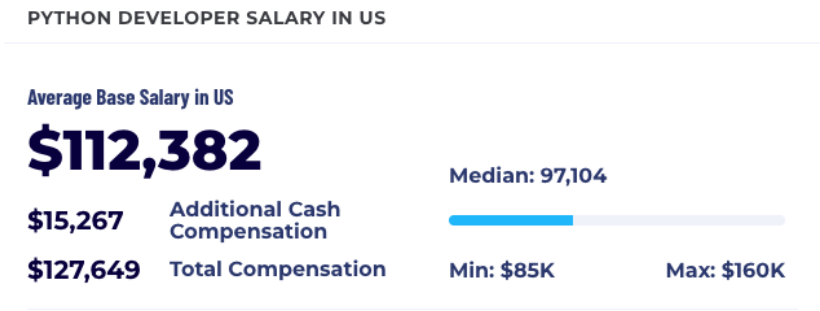
```java
public class Welcome {
    public static void main(String[] args) {
        System.out.println("Welcome to QTM151!");
    }
}
```

- Python:

```python
print("Welcome to QTM151!")
```

Why Python?

- Rewarding
- Despite current atmosphere, many AI startups require proficient Python skills



Class Logistics

Course Objectives

- Create Python Project using PyCharm
- Use Jupyter Notebook for reproducible workflows.
- Perform basic operations and write functions in Python.
- Familiarize themselves with NumPy, Pandas, Matplotlib for numerical operations, data wrangling and manipulation, and visualization, respectively.
- Implement linear models and understand principles of time series analysis.
- Learn basic machine learning via Scikit-learn.

Grades and Late Policy

- Homework (50%)
- Quizzes (10%)
  - Multiple choices or small data analysis
- Final Project (40%)
  - Four-stage development
  - Each section consists of 10%.
  - Final output due at the end of the semester

- Late assignments will automatically be graded for half-credit.
- You can use additional lab sessions to earn bonus point (counting towards the total score)
  - Only completed lab sessions before the deadline count.

Computing Setup

Software Installation

- Jupyter script file
- File with code in the Python programming language.
- Instructions for the program to follow

- PyCharm IDE
- Integrated development environment
- Virtual environment creation

- Python
- In the background

Python and PyCharm

- Make sure you have Python.
- If not, PyCharm offers features to simplify the process of installing and configuring Python for your projects automatically.
- If there is no existing Python installation, it provides the option to download and install a Python interpreter automatically, along with creating a virtual environment for the project.
- Read my tutorial on installing PyCharm.

GitHub

- A file management system in the cloud (also compatible with desktop app)
- Has version control
- Great for collaborative programming and reproducible research workflows

Jupyter Notebooks

- We will use Jupyter notebooks for our classes.
- I will prepare another tutorial on how to use them, too.
- It is like R markdown, a combination of code, texts, and visualizations.
- Bring your laptops to class.
- And you will try coding yourself in class

Next

- We will go over
  - GitHub.
  - PyCharm
  - Jupyter Notebook