HSS 611 - Week 1: Course Overview

2023-08-28

Welcome to Programming for DHCSS

- Instructor: Taegyoon Kim, Ph.D. in Political Science and Social Data Analytics
- Teaching assistant: Namgoong Minsang, Ph.D. student in Culture Technology

Welcome to Programming for DHCSS (cont'd)

- Language? Python
- Who is this for?
  - Broadly applicable techniques as long as it concerns data work
  - The needs of DHCSS research

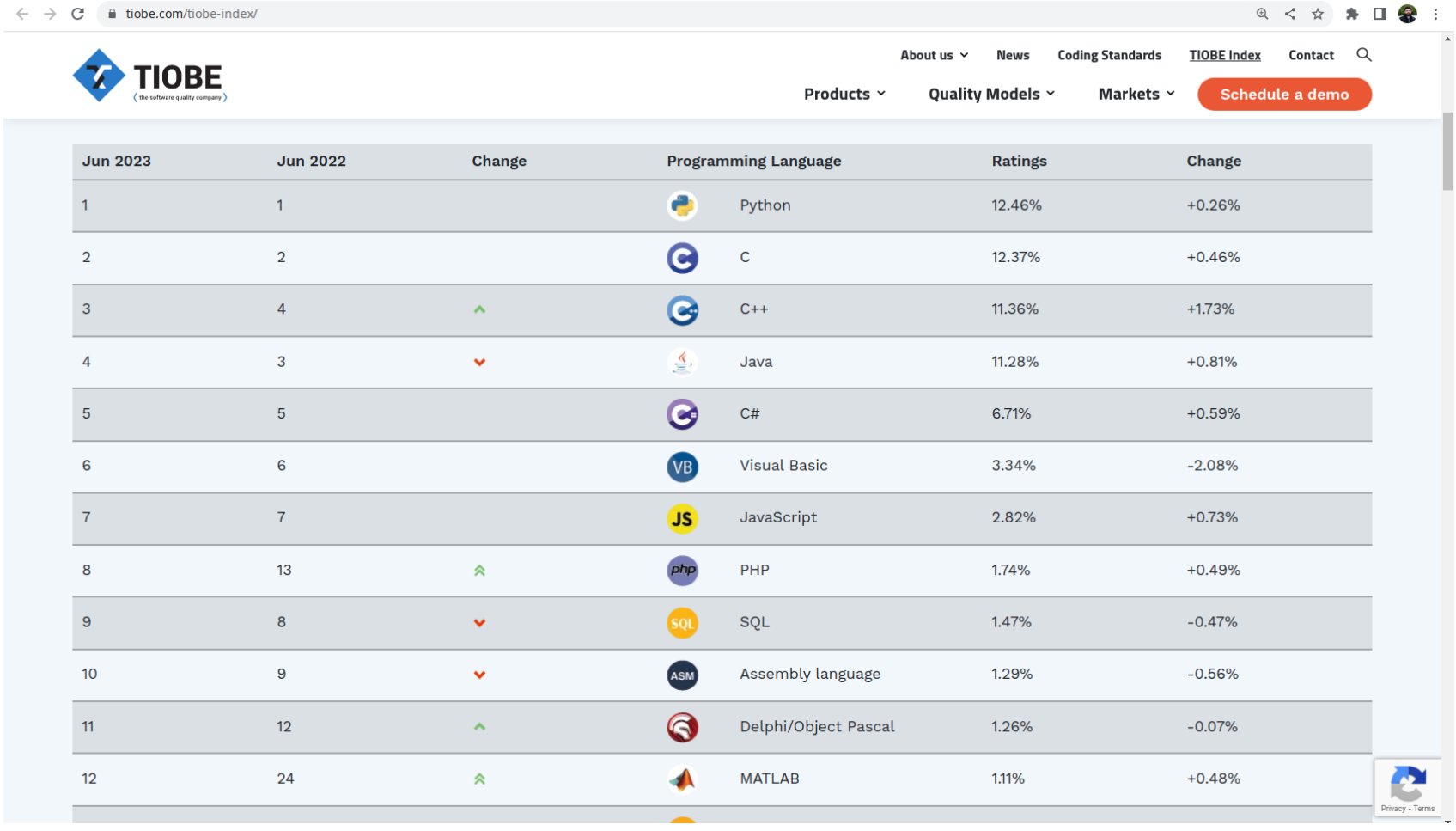
Welcome to Programming for DHCSS (cont'd)

- How challenging will this course be?
  - It depends
  - On one hand, lectures assume little prior knowledge
  - On the other hand, in-class activities and assignments will accommodate more advanced learners

Python

- A general-purpose programming language
- Created by Guido van Rossum
- First released in 1991

Python's Popularity



Jun 2023	Jun 2022	Change	Programming Language	Ratings	Change
1	1		Python	12.46%	+0.26%
2	2		C	12.37%	+0.46%
3	4	↑	C++	11.36%	+1.73%
4	3	↓	Java	11.28%	+0.81%
5	5		C#	6.71%	+0.59%
6	6		Visual Basic	3.34%	-2.08%
7	7		JavaScript	2.82%	+0.73%
8	13	⇈	PHP	1.74%	+0.49%
9	8	↓	SQL	1.47%	-0.47%
10	9	↓	Assembly language	1.29%	-0.56%
11	12	↑	Delphi/Object Pascal	1.26%	-0.07%
12	24	⇈	MATLAB	1.11%	+0.48%

Figure 1: Python Tiobe

Python's Popularity (cont'd)

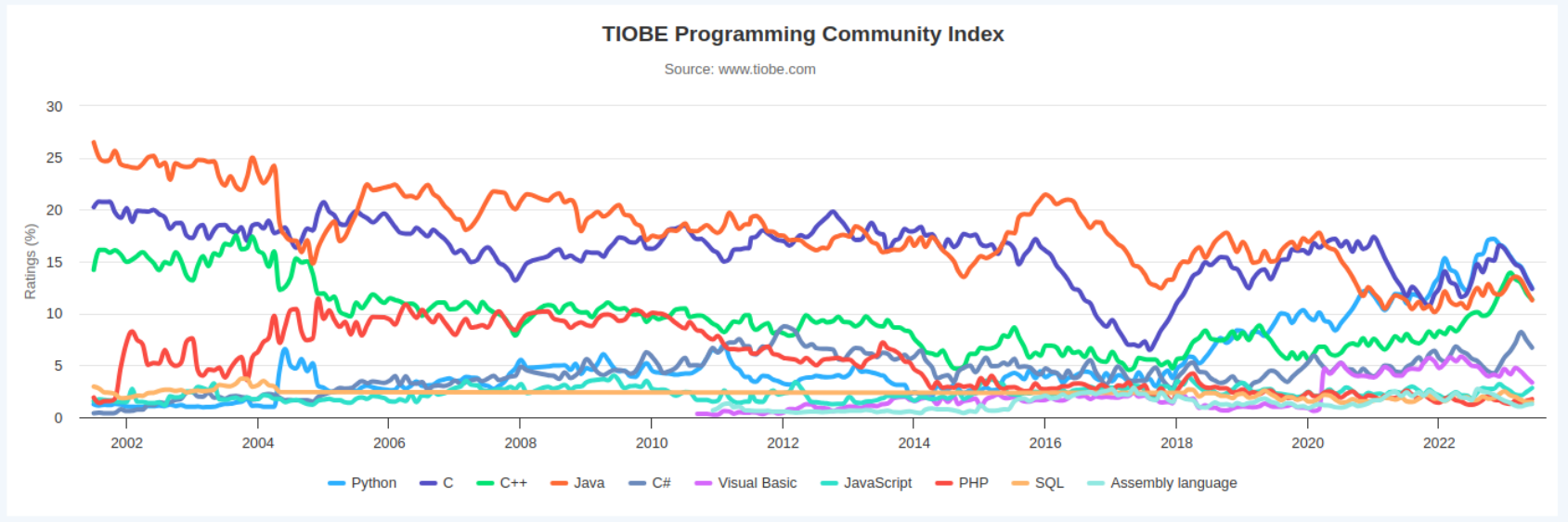


Figure 2: Tiobe over Time

Python's Popularity (cont'd)

- Open source & free
- English-like, readable syntax
- Very popular across domains

Python's Popularity (cont'd)

- Used in web development, game development, data science, academic research, etc.
- Booming in machine learning & artificial intelligence

Python for DHCSS: Data Collection

- A lot of data live on the web
- Need to be collected, properly formatted, and stored
- Python ideal for all of them

Python for DHCSS: Data Collection (cont'd)

- APIs (Application Programming Interface)
- Web scraping

Python for DHCSS: Data Analysis and Visualization

- Pandas
- Numpy
- Matplotlib

Python for DHCSS: NLP/Text-as-data, Machine Learning, Network Analyis, etc.

- NLTK, spaCy, Gensim, Transformers, etc.
- Scikit-learn for traditional ML, PyTorch / Tensorflow for deep learning, etc.
- NetworkX, igraph, etc.

Statistics? Other Languages?

- R
- Stata, Matlab, SPSS, etc.

Course Structure

- First half: fundamentals
- Second half: advanced topics
- Let's check the syllabus

Course Structure (cont'd)

- Monday: lecture (with live coding)
- Wednesday: tutorial presentations + lab sessions
  - Lab sessions will involve exercises with varying degrees of difficulty
  - For lab sessins, proceed as far as you can before wrapping up

Major Tasks: Attendance

- Attend
- Stay engaged in class

Major Tasks: Tutorial Presentation

- Prove your mastery of the techniques covered in that week
- Deliver two tutorial presentations
  - One topic from the first half, another from the second
- Help other students perfect the techniques by demonstrating in detail how the techniques work in practice

Major Tasks: Tutorial Presentation (cont'd)

- Encourage to use data of your own interest
- Encourage to go deeper than what is covered in class
- Feel free to incorporate what you've learned in previous weeks

Major Tasks: Tutorial Presentation (cont'd)

- Minimum 20 minutes on Wednesdays
- Starting in the 2nd week
- Sign up *here*
- See examples from sites like Real Python

Major Tasks: Final Paper

- Write a research paper that involves a significant amount of data work
- Choose a topic of your own interest
  - Ideally, select a topic and data you would actually use for a paper
- Construct a comprehensive data-driven research pipeline, covering data collection, pre-processing, analysis, and visualization
- Focus on detailed accounts of data collection, cleaning, manipulation, analysis, and visualization
- Minimize discussions of theories or prior literature

Major Tasks: Final Paper (cont'd)

- Use this as an opportunity for data collection/analysis/visualization for your (MA) paper
- Deliver a 15-minute presentation in the final week
- Email the paper to the instructor, along with scripts and data, by the end of the semester

Major Tasks: Problem Set

- Several problem sets will be assigned throughout the course