Welcome to COMP 204
Computer Programming for Life Sciences!

Introduction

Yue Li

What's computer programming?

- Implementing a set of instructions in computer to do things
- Implementation involves (1) **Designing** and (2) **Executing** such instructions on the computer
- Many (if not all) applications in life sciences involve computer programming:
  - Identifying transcription factor binding sites based on sequence motifs
  - Predicting regulatory genomic sequences that alter gene expression changes
  - Identifying brain subregions from neuroimaging data corresponding to depression symptoms
  - Identifying genetic predisposition that cause complex phenotypes (e.g., height, BMI, breast cancer)
  - Predicting future disease onsets and treatment effects based on patients' health record histories

How to make computer to do complex tasks?

- The problem is that computer is only good at *basic and repetitive tasks* like multiplying/adding/dividing numbers
- How do we go from adding numbers to identifying disease mutations sounds non-trivial at all!
- Imagine we have set of a laboratory toolkit (e.g., beakers, test tubes, pipettes) and some basic reagents. They do not mean anything unless we design and implement an chemical experiment using these basic tools.
- Programming is about putting basic instructions together in specific ways.
- Programming is very empowering skill to have (arguably the closest skill to superpower)

Key Course Information

Introduction to key concepts in computer science:

- Key elements in computer programming: variables, expressions, types, functions, conditionals, loops, objects and classes, data structures (lists, strings), modular software design, libraries, file input/output, debugging.
- Building blocks of computer algorithm: conditionals, loops, logic relations, arithmetics (+,-,*,/), and that's all!

Emphasis on applications in the life sciences:

- (DNA $\rightarrow$ RNA $\rightarrow$ Protein) * environments $\rightarrow$ phenotypes,
- Some key questions: how mutation affects gene expression and phenotypes or diseases? How specific gene expression are turn on and off during development and adulthood? How distinct species differ and share in common in their genomes?

Prerequisites: BIOL 112 and a CEGEP level mathematics course

Restrictions: Only one of COMP 204, COMP 202, COMP 208, or COMP 364 can be taken for credit. COMP 204 cannot be taken for credit with or after COMP 250, COMP 206, or COMP 364.

Key Course Information

Objectives: By the end of this course, students will be able to:

1. Design and describe precise, unambiguous instructions that a computer can use to solve a problem/perform task(s)
2. Translate these instructions into a language that a computer can understand (Python)
3. Write Python scripts that solve complex problems by decomposing them into simpler subproblems
4. Apply programming-style and structure conventions to make your programs easy to understand, debug and modify
5. Learn independently about new programming-language features/libraries by reading documentation and experimenting

Key Course Information

Instructors:

- Yue Li: yue.yl.li@mcgill.ca
- Office: McConnell 324 (January) and Trottier 3105
- Office hours: Wednesday 11:30-12:30

Schedule: (MW) Rutherford (Physics) 112; (F) Leacock 26
Web page: `https://www.cs.mcgill.ca/~yueli/teaching/COMP204_Winter2019/home.html`
MyCourse: Every lecture will be recorded
Teaching assistants:

- Ayrin Ahia-Tabibi: airin.ahia-tabibi@mail.mcgill.ca
  Office hours (Trottier 3090): Tuesday 2:00-3:00
- Pouriya Alikhani: pouriya.alikhani@mail.mcgill.ca
  Office hours (Trottier 3090): Thursday 10:00-11:00

CSUS Helpdesk

HOURS: 12pm - 5pm (mon-fri)
LOCATION: Trottier 3090

WHO ARE WE? WHAT DO WE DO?

- U2 and U3 students who have taken this course and want to help you!
- We are a **FREE** drop-in tutoring service, perfect for study help, and guidance on assignments.
- We provide review sessions for midterms and finals for intro courses!

Key Course Information

Schedule of topics covered + All lecture notes:

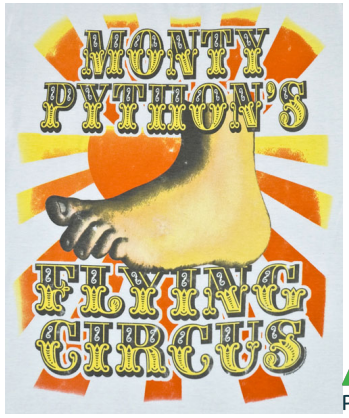
- `https://www.cs.mcgill.ca/~yueli/teaching/COMP204_Winter2019/schedule.html`
- Lectures will be recorded and made available on MyCourses

Online material: How to Think Like a Computer Scientist: Interactive Edition (Python)

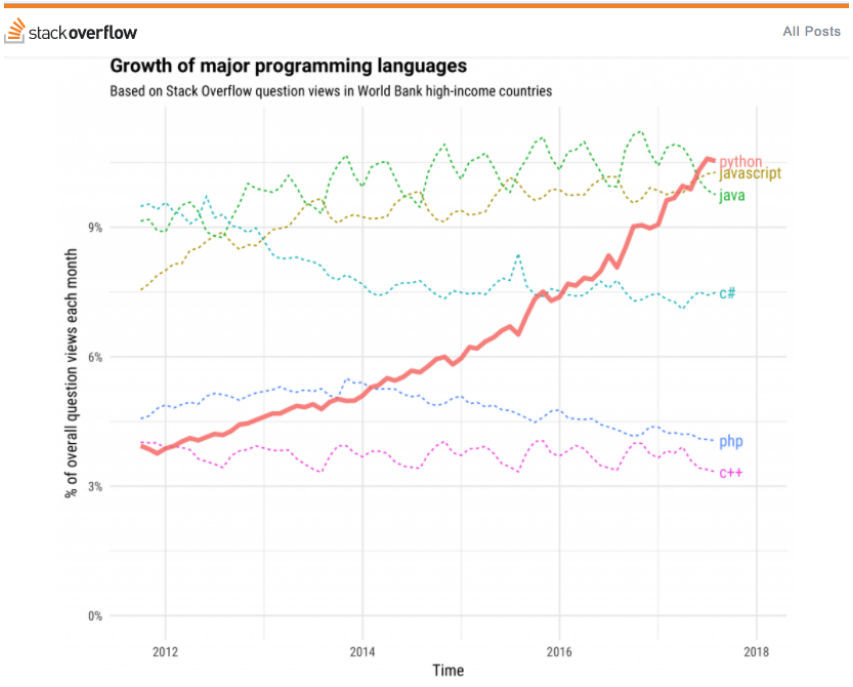
- `http://interactivepython.org/courselib/static/thinkcspy/index.html`

Python programming language

- Created by Guido Van Rossum (early 90s)
  - Named after 'Mounty Python's Flying Circus'
- Python is one of the most popular programming language in data science and machine learning (used heavily by Google, e.g., Tensorflow)
- Version 3.7 (latest to date) will be used in this course
- Install Anaconda (Python API manager): `https://docs.continuum.io/anaconda/install/`



ANACONDA
Powered by Continuum Analytics



Course evaluation

In-class Quiz: 10%, active at the start of each lecture, password access announced in class, the best 25 quizes (i.e., 0.4% per quiz)
Assignments: 35% (5 assignments worth 7% each)
Midterm exam: 20%

- Tentative Time: Tuesday, February 19th from 18:30-20:00 pm; Location: ENGMC 304
- Mix of multiple choice and short/long answer written questions (mini final exam)

Final exam: 35%

- 3-hour final exam, time and place TBD

Total grade = 0.1*Quiz + 0.35*Assignments + 0.2*Midterm + 0.35*Final

- There is no 100% final option: all assignments will count, except for medical reasons.
- In exceptional situations, students may write a supplemental examination: contact your Student Affairs Office.
- Students who receive unsatisfactory final grades will NOT have the option to submit additional work.

Assignments

- 5 Python programming assignments, each aiming at addressing a specific biological question using programming techniques introduced in class.
- Solutions must be submitted electronically on MyCourses. Every student is responsible for verifying that their submissions are successful.
- Due at 11:59:59 PM on the following dates (roughly two weeks to complete each assignment):
  - Assignment 1 Posted: January 17; Due: February 1
  - Assignment 2 Posted: February 5; Due: February 13
  - Assignment 3 Posted: February 20; Due: March 7
  - Assignment 4 Posted: March 15; Due: March 29
  - Assignment 5 Posted: April 1; Due: April 16
- At the time the assignment is released, you will have all of the background to complete the assignment.
- Working hard on your assignments will improve your score on exams.

Late Policy

- Late assignments will be deducted 20% each day or fraction thereof for which they are late, including weekend days and holidays:
  - 0-24 hours late = 20% deduction
  - 24-48 hours late = 40% deduction.
  - >48 hours late = not be accepted (grade of 0%).
- Programming assignments are notoriously time-consuming. DO NOT leave it to the last minute!
- If you have only partially finished an assignment, document the parts that do not work, and submit what you managed to complete for partial credit.
- Individual exceptions to the lateness policy will not be granted without appropriate justification submitted in writing and supported by documentary evidence.

Assignment grades

- Assignment marks will be posted on myCourses. It is your responsibility to check that the marks are correct and to notify your section instructor of any errors or missing marks.
- If you believe that your assignment was graded incorrectly, you should first email the TA who marked your assignment. Their email address should be in the feedback left on your assignment. If you and the TA cannot resolve the discussion, then you should contact your instructor.
- Complaints about grading must be formulated within two weeks of the release of the grade.

Getting help

- The instructor and teaching assistants will not answer questions by email.
- Post your questions on myCourses or ask them at office hours
- Answer each other's question on MyCourses BUT do NOT provide solution code.
- Only email the instructors or TAs for private matters, and do not count on a quick response.

Plagiarism Policy

McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism, and other academic offenses under the Code of Student Conduct and Disciplinary Procedures (see www.mcgill.ca/integrity/ for more information).

- Include your name and McGill ID number at the top of each source code file that you submit. By doing so, you are certifying that the program or module is entirely your own.
- Work submitted for this course must represent your own efforts. Assignments must be done individually. Do not rely on friends or tutors to do your work for you. You must not copy any other person's work in any manner (electronically or otherwise), even if this work is in the public domain or you have permission from its author to use it and/or modify it in your own work.
- Do not give a copy of your work to any other person, or post your solutions on any publicly accessible repository.

Collaboration Policy

The plagiarism policy is not meant to discourage interaction or discussion among students.

- Discuss assignment questions with instructors, TAs, and your fellow students.
- However, there is a difference between discussing ideas and working in groups or copying someone else's solution. A good rule of thumb is that when you discuss assignments with your fellow students, you should not leave the discussion with written notes.
- When you write your solution to an assignment, you should do it on your own.

McGill Français

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Brief introduction about me

- A new Assistant Professor at the School of Computer Science
- Research in applied machine learning in computational biology
- Machine learning: hierarchical Bayesian inference, deep generative model, matrix/tensor factorization
- Computational biology: mining electronic health records, inferring risk mutations from genome-wide association studies and functional genomic data, gene regulatory
- Homepage: `https://www.cs.mcgill.ca/~yueli`
- You will hear more about my research later in the course

Overview of topics and schedule

https://www.cs.mcgill.ca/~yueli/teaching/COMP204_Winter2019/schedule.html

In-class quiz

Go to myCourse COMP204 to complete the quiz. I will let you the password in the class.