

EDUCATION

University of Toronto (St. George Campus)

Honours Bachelor of Science - Statistical Science & Bioinformatics

2021-09 – 2025-06

ASIP Co-op Program

GPA: 3.96/4.00; Average: A+

Undergraduate Courses: Multivariable Calculus, Linear Algebra, Probability & Statistics, Python Programming, Human Genetics, Molecular & Cell Biology, Organic Chemistry

EXPERIENCE

Lunenfeld-Tanenbaum Research Institute @Sinai Health

Toronto, ON

Research Assistant with Prof. Frederick Roth and Dr. Jochen Weile

2023-04 – Present

- Re-processed the raw data underlying existing variant effect maps (atlas of the functional effects of all possible genetic variants on proteins) with the latest versions of the TileSeqMave, a bioinformatics pipeline for the calculation of fitness scores from sequencing reads and provides a suite of quality control visualizations.
- Compiled benchmark sets of variants with known pathogenicity from online databases and literature for each map.
- Evaluated the predictions made by different versions of variant effect maps using the precision-recall curve and use them to infer evidence strength for clinical interpretation by calculating the log-likelihood ratio for pathogenicity.
- Gave recommendations for optimizing the implementation of TileseqMave pipelines based on analysis results.
- Delivered presentation on the methodologies and key finding of this project to the entire lab and wrote a report. (Link to my report)

Dalla Lana School of Public Health

Toronto, ON

Research Trainee with Prof. Kuan Liu and Prof. Kevin Thorpe

2023-05 – Present

- Conducted a series of simulations in R statistical software to simulate Randomized Controlled Trials (RCTs) with different outcomes (continuous, binary, repeatedly measured) under different missing mechanisms.
- Applied common missing data handling methods such as complete case analysis, imputations and inverse probability weighting to each simulated data set, and assessed their performance by employing a range of evaluation metrics.
- Provided recommendations on how to deal with missing data in RCTs based on evaluation results.
- Acquired in-depth knowledge and expertise in missing data methodologies, as well as the design and analysis of RCTs.
- Gave a poster presentation on research showcase Day. (Link to my poster)

SKILLS

Languages & Tools: Python, R, Unix Shell, LaTeX, PyMOL

Libraries: Numpy, Pandas, Scikit, Matplotlib, SpaCy

PROJECTS

Treemap Applications

2023-03 – 2023-04

- Modelled different kinds of real-world hierarchical data using tree data structure in Python, such as modelling the file system, and modelling moves in chess games.
- Implemented recursive operations on trees, and developed an algorithm to generate a geometric treemap visualization.
- Leveraged the os module in Python to create a program that seamlessly interacts with computer's file system.
- Used inheritance to design classes that share a common interface, allowing for code reusability and promoting a structured approach to object-oriented programming.

Optimal Grouping Design

2023-01 – 2023-02

- Implemented a Python program to put students into optimal groups with respect to different criteria and survey data.
- Compared different grouping algorithms including AlphaGrouper (groups student alphabetically), GreedyGrouper (forms groups using greedy algorithm), and SimulatedAnnealingGrouper (forms groups using simulated annealing algorithm).
- Visualized the scores of groups created by different grouping algorithms, as well as some simple statistics about how well they do using side-by-side bar plots.
- performed thorough unit testing on a complex program comprising multiple interacting classes.

Analysis of Hypertension & Low-Income Data in Toronto

2022-11 - 2022-12

- Extracted effective hypertension & low-income data from the CSV file using Python.
- Performed data cleaning & standardization and selected suitable statistical methods for analysis.
- Utilized the object-oriented programming paradigm to implement functions. Additionally, designed comprehensive test cases for each function to ensure thorough testing and validate program functionality.
- Drew conclusions regarding the correlation between hypertension rates and low-income rates on the neighborhood-level.
- Practiced skills in using Python and relevant libraries to conduct data analysis.

HONORS AND AWARDS

Summer Undergraduate Data Science (SUDS) Research Program Award

2023

Awarded to fund a summer research project for undergraduate students in data science field. (\$7200 CAD)

Dean's List Scholar

2022, 2023

Awarded to students who had excellent GPA in the University