

# Pinyi Wu

• Email: [pw2551@cumc.columbia.edu](mailto:pw2551@cumc.columbia.edu) • Cell Phone: (585) 210-5186 • LinkedIn: [paula-pinyiwu](#) • GitHub: [paulawucu](#)

## EDUCATION

---

### Mailman School of Public Health, Columbia University

New York, NY

*Master of Science in Biostatistics*

*Expected May 2023*

- **GPA:** 4.0/4.0 | **Courses:** Theoretical Neuroscience, Survival Analysis, Longitudinal Data Analysis, Numerical Methods, Biostatistical Methods, Data Science, Statistical Inference, Probability

### University of Rochester

Rochester, NY

*Bachelor of Science in Brain & Cognitive Science, Bachelor of Arts in Data Science*

*May 2021*

*Minor in Psychology*

- **GPA:** 3.99/4.0 | **Honors:** Summa Cum Laude, Highest Distinction
- **Courses:** Neurobiology, Auditory Perception, Sensory & Motor Neuroscience, Computational Model of Cognition, Data Mining, PHP, SQL, Artificial Intelligence

## RESEARCH INTEREST

---

- Brain Imaging, Neuropsychiatry Disorders, Sensory & Motor Neuroscience, Data Mining, Statistical Modeling

## SKILLS

---

- **Programming language:** Proficient in Python, R, SQL, Excel, LaTeX, Command Line; experience in Bash, Docker, Java, MATLAB, SAS
- **Laboratory skills:** EEG, monkey chaireing, brain perfusion fixation, mice brain dissection, animal behavior training
- **Languages:** English, native speaker of Mandarin, Japanese JLPT N2, intermediate learner of Spanish

## RESEARCH EXPERIENCE

---

### Brain Imaging Lab, Columbia University

New York, NY

*Research Assistant | Supervisor: Dr. Spiro Pantazatos*

*Jun. 2022 – present*

- Super-resolved clinical MRI scans to 1 mm-isotropic research-quality images using the “SynthSR” package in Python and Command Line.
- Dockerized procedures using Dockerfile to facilitate code distribution. Currently working to connect the procedure to XNAT to further facilitate imaging management and productivity, coded in JavaScript.

### Research at Department of Psychiatry, Columbia University

New York, NY

*Project Biostatistician | Supervisor: Dr. Richard Sloan*

*Feb. 2022 – present*

- Analyzed Midlife in the United States (MIDUS) dataset, to investigate the effect of childhood trauma on the changes in adult psychosomatic conditions across different time points using R.
- Conducted longitudinal analysis, cleaned and wrangled raw data, built Linear Mixed Effect Models. Also conducted modifier analysis on the effect of psychological factors on attenuating cognitive functioning.
- Found higher childhood trauma led to a greater decline in Episodic Memory during one’s adulthood. Preparing manuscripts for future publication.

### Adam Snyder Lab, University of Rochester

Rochester, NY

*Research Assistant*

*Feb. 2020 – May 2021*

- Assisted in research focusing on decoding EEG signals of novel detections of different kinds of visual components in 8 macaques, chaired macaques, prepared experiments setups.
- Analyzed EEG data using Python and MATLAB in collaboration with graduate students, built a logistic regression model to classify and predict motion direction from EEG signals, ran behavioral training using MATLAB.

### Xiongjie Yu Lab, Zhejiang University

Hangzhou, China

*Research Assistant*

*Jul. 2019 – Aug. 2019*

- Designed experimental protocol of discrimination task of time intervals between a pair of sound stimulus, trained one animal subject (rat) to complete experiment, coded in Spike 2 for training protocols and MATLAB for data analysis.
- Assisted in novel detection of sound frequency in nonhuman primates, examining neural activity using electrophysiology (EP).

- Trained C57 mice using the Tube Test protocols, used optogenetic stimulation (on mPFC) to examine changes in social hierarchy on 20 mice (5 cages), recorded and marked mice behaviors using BORIS. Presented the final product to the whole lab during the meeting.
- Used perfusion fixation to prepare brain tissue for microscopy observation, dissected brains and observed under fluorescence microscopy, assisted in observations and recording of calcium signals.

## **COURSE PROJECTS**

---

### ***Columbia University, Mailman School of Public Health***

#### **Breast Cancer Diagnosis and Optimizations**

**Spring 2022**

- Built a predictive model to classify malignant and benign images of breast cancer tissues in R.
- Coded from scratch, implemented a logistic-LASSO model to predict diagnosis. Developed a modified Newton-Raphson algorithm and a Pathwise Coordinate Optimization. Also coded for 5-fold cross validation to find optimal tuning parameters for the model. Reached AUC of 0.9994.
- Collaborated with a team of 5 to write and edit a 16-page summary report. Presented methods and results to the class.

#### **Hurricane Trajectory Prediction**

**Spring 2022**

- Built a predictive model to forecast hurricane trajectory and to explore seasonal differences of storms.
- Proposed a hierarchical Bayesian model. Coded for the Metropolis-Hastings algorithm to estimate posterior parameter distributions. Explored hurricane-related characteristics across seasons and years.
- Collaborated with a team of 5 to write and edit a 15-page summary report. Presented methods and results to the class.

#### **Data Analysis of Drug Overdose Across the US Jurisdictions**

**Fall 2021**

- Explored drug overdose death situation in the 51 US jurisdictions using R. Version controlled by GitHub.
- Built [website](#) and [R Shiny dashboard](#). Wrangled raw data and analyzed county-level death in Florida. Plotted county-level choropleth maps and “death by states” and “death by age and race” visualization on R Shiny.
- Collaborated with a team of 5 to write and edit a 20-page summary report. Wrote scripts, recorded, and narrated for a 2-min website presentation video.

### ***University of Rochester, Goergen Institute for Data Science***

#### **Capstone Project: Predicting Filter Failure in Trucks**

**Spring 2021**

- Built a comprehensive model to predict the probability of Diesel Particulate Filter failure of trucks based on various features using Python, in collaboration with the transportation company Vnomics.
- Spearheaded completing data preprocessing, data visualization, time series analysis, and final pipeline models. Independently completed building supervised learning models using scikit-learn.
- Achieved 85% model recall rate. Presented weekly accomplishments, midterm, and final product to the Vnomics team. Collaborated with a team of 5 to write and edit a 15-page report.

#### **Web Interface for Bookkeeping in a Bookstore**

**Fall 2020**

- Designed and implemented the database and web-interface of a virtual bookstore using HTML, PHP, and MySQL.
- Self-designed a comprehensive web-interface with various functions, including customers’ sign up, log ins, book search, purchases, and employees’ log ins and management of the databases.

#### **Improvement of Movie Recommendation Algorithm: A Hybrid Approach**

**Spring 2020**

- Designed and refined movie recommendation algorithm using integrated modeling, instructed by Professor Jiebo Luo.
- Coded in Python, combined content-based filtering, and user-based collaborative filtering approach to obtain an improved movie recommendation system, with more variability and relatively higher precision than the user-based collaborative filtering approach alone. Wrote a 5-page research paper.

## **TEACHING EXPERIENCE**

---

### ***Columbia University, Mailman School of Public Health***

#### **P8105: Data Science I**

*Teaching Assistant***New York, NY***Sept. 2022 – present*

- Held weekly office hours, answered R and GitHub related questions, offered coding help and instructions regarding homework. Graded homework, midterms, final projects.
- Provided instructions to 5 groups on their final projects. Held meetings to address programming related questions.

## Computing Club at CUIMC

Leader and Lecturer of the Python Workshop

New York, NY

Oct. 19<sup>th</sup>, 2021 & Nov. 1<sup>st</sup>, 2022

- Led the demonstration of the coding section using Google Colab notebooks, covered topics including Python basics data structures, data preprocessing, data frame, graphing, and machine learning basics. Utilized various data processing and machine learning libraries.

## Skills for Health and Research Professionals Training Program

Teaching Assistant for Machine Learning Bootcamp

Online

Jun. 16<sup>th</sup>-17<sup>th</sup>, 2022

- Assisted in teaching modules (Penalized Regression, Tree Based Methods, Dimension Reduction) in Machine Learning in R to healthcare professionals. Answered questions synchronously regarding concepts and coding tasks on Slack during lectures and lab sessions.

## University of Rochester

### Center for Excellence in Teaching and Learning

One-on-one Subjects Tutor

Rochester, NY

Feb. 2019 – May 2021

- Tutored peer students on a one-on-one basis in various subjects, including Computer Science, Brain and Cognitive Science, and Biology.
- Answered questions regarding course materials, offered instructions to homework assignments, made test preparation plans, and advised overall study skills in specific subjects upon the requests of each tutee.

### CSC 161: Intro to Programming

Lab Teaching Assistant

Rochester, NY

Sept. 2020 – May 2021

- Held bi-weekly lab sessions, answered Python-related questions, offered coding help and instructions regarding weekly assignments and a 3-staged project on various Python topics. Graded weekly lab assignments, projects, midterms, and final exams.

Workshop Leader

Sept. 2019 – May. 2020

- Led weekly workshop sessions to a group of 16 peer students, promoted peer discussion upon Python-based questions, explained concepts of syntax and programming logics, graded weekly quiz.

### BCS 204: Lab in Cognitive Neuroscience

Teaching Assistant

Rochester, NY

Sept. 2020 – Dec. 2020

- Held weekly office hour, offered instructions on analysis of behavioral, EEG, fMRI data using Excel, EEGLAB, and SPM12, assisted in recording EEG signals in the actual laboratory setting.

### BCS 110: Neural Foundations of Behaviors

Teaching Assistant

Rochester, NY

Sept. 2019 – Dec. 2019

- Held weekly recitations to a group of 15 students, reviewed lecture materials, answered questions.
- Additionally held four 2-hour review sessions with three other teaching assistants to help students prepare for unit exams. Graded 4 unit exams and 1 optional final.

## PAPER

---

- Mangold M.T, **Wu P.**, Sloan R.P., *The Lack of a Moderating Role of Psychological Wellbeing in the Relationship Between Childhood Trauma and Change in Cognition in Midlife in the United States* (2023) in preparation.

## OUTREACH

---

### Data Science for All – Women

Team Member

Online

Jul. 2022 – Aug. 2022

- In a group of five, investigated the effect of abortion bans on foster care entries.
- Scraped data from national services, cleaned raw data, fit random forest and XG-boost models, found more restrictive abortion laws is associated with higher rates of infants entering foster care systems.
- Presented the final product to more than 250 program attendees, TAs and instructors.

### China Thinks Big

Academic Coaching Team Member

Online

Oct. 2018 – Dec. 2018

- Reviewed public health related research papers written by Chinese high school students.
- Made academic suggestions and participated in grading.

## **Harvard Summit for Young Leaders in China**

**Hangzhou, China**

*Teaching Fellow*

*Aug. 12-22, 2018*

- Co-taught 3 seminars in Morphology to 30 students in total, answered questions, made class notes daily, graded homework and final exam.
- Taught a Capstone research project class to 9 students, provided detailed guidance on research methodology, including topic selection, data collection, data analysis, and presentation preparation.
- With two other fellows, led a house of 54 students, won the House Cup.

## **HONORS**

---

- **Dean's Scholarship** (\$28,000), University of Rochester, Sept. 2017 – Jan. 2021
- **Dean's List**, University of Rochester, Sept. 2017 – May 2021