YIRU GONG

100 Haven Avenue, New York, NY 10032 | (917) 742-0036 | yg2832@cumc.columbia.edu

EDUCATION

Columbia University Mailman School of Public Health *New York, NY*
Master of Science (MS), Biostatistics 09/2021 – 05/2023 (Expected)

- Public Health Data Science Track; Relevant Courses: *Data Science, Biostatistical Methods*

University of Edinburgh - Zhejiang University Joint Institute *Edinburgh, UK*
Bachelor of Science with Honors in Biomedical Science, Dual degree program of UoE & ZJU 09/2017 – 06/2021

- GPA: 3.82
- Relevant courses: Applied Biomedical Science (Statistics); Introductory Data Science with Python and Tableau, Data Analytics for Customer Insights (NUS summer program)

SKILLS AND CERTIFICATES

- Proficient in R (tidyverse, Shiny), Python (Numpy, Pandas), and MATLAB; familiar with Linux, SQL, C language, and VBA
- Good command of MS Office software; Familiar with data visualization in Tableau
- Languages: Mandarin (Native), English (Proficient)
- Coursera Specialization Certificate: *Mathematics in Machine Learning* (ICL)
- EdX Certificate: *Introduction to Probability* (HarvardX), *Advanced Linear Algebra* (UT)

RELEVANT EXPERIENCE

GlaxoSmithKline (GSK), *Digital Analyst Intern, R&D Tech* Shanghai, China, 04/2021-07/2021

- Managed a project on “Establishment of a Database to Track Differences in Medication Treatment Response by Ethnicity”; designed drug data collection and analysis methods, established computational workflow using R and Python to support drug pharmacokinetics analysis.
  - Reduced manual workload from 2 months to 2 days by using both R and Python
  - Tidied semi-structured JSON data of > 400k FDA clinical trials by Elasticsearch in Python
  - Expanded information on the use of 56 chemical drugs and 10 bio-tech drugs by ethnicity
  - Identified three ethnic difference-related genes and signaling pathways
- Facilitated project on “Natural Language Processing (NLP)-based Medical Knowledge Graph Establishment.”
  - Presented researchers the principle of Natural Language Processing and AI to extract disease-related information from > 10 million research articles and official documents
  - Communicated with 6-8 Physicians and Clinicians for technical issues and designed big data-based solutions
  - Applied NLP to clinical trials to help colleagues identify potential competitor drugs and trials
  - Applied Deep Learning (NER, RE) to carry out auto-revision of medical writing and reduced time and effort on medical document translation and revision by 50%

DuPont Danisco Nutrition & Bioscience, *Sales/Marketing Intern, Intellifresh™* Shanghai, China, 03/2021-07/2021

- Initialized the Wechat Digital Marketing Platform for Danisco, localized and digitalized the Global marketing sheets into style favored by Chinese customers
- Expanded potential target customers from 200 to 2,000 by effectively improving the product exposure and customer management method, significantly increasing customer orders
- Established an automatic-filling labeling application document using VBA in Word and Excel. Significantly reduced the time for format editing and content filling

OTHER PROJECTS

Graduate Course: Data Science 09/2021-12/2021
Final Project: Establishing a Website of World Happiness Score and Related Factors using R and Shiny App

- Built a GitHub Website to visualize the happiness score in different countries and correlated social-economic factors
- Analyzed and visualized the yearly worldwide happiness score and GDP, COVID-19 Status, Life Pressure, and other factors by using interactive plots in ggplot, Plotly, and shiny app in R
- Established multivariable linear regression model to explore the effect of social factors on happiness level and applied five-fold cross-validation to verify the model
- Top rank (A) received

Prof. Roger Foo’s lab, Cardiovascular Research Institute, National University of Singapore 01/2021-5/2021
Graduate Thesis: Investigating the role of ADAR-mediated RNA-editing in Cardiomyopathy

- Established a computational pipeline in Linux and R for RNA-editing site identification from 324 large RNAseq data
- Performed multithread processing and core distribution by using R in the remote Linux Server

- Wrote dissertation and made thesis presentation to Department Dean and professors, top ranking received

Prof. Hangjin Jiang's lab, Center of Data Science, Zhejiang University 09/2019-12/2020

Project: Estimating the contribution of Gene-Environment Interactions in DNA methylation in cancer

- Reproduced the code and results of a paper on Gene-Environment Interactions of GWAS data in obesity using R
- Improved the statistical method and linear regression by applying Bayesian estimation. Model Accuracy improved from 50% to 90%; applied the improved model to a new field in DNA methylation

Project: Analyzing brain fMRI image between smoker and non-smoker by building statistical models

- Applied high dimension regression models (Tensor regression), sliding windows methods and neural networks to identify functional altered brain regions in smokers

Prof. Robert Young's lab, Usher Institute, University of Edinburgh 07/2020-08/2020

Project: Exploring the genetic and epigenetic regulation of promoter birth and death in human brain evolution

- Observed human brain's regulatory gene evolution by comparing genetic sequences between human and macaque's brain in R and Linux server; wrote report and analyzed the result and odds ratio in an R markdown file

Prof. Gedi Luksys's lab, Department of Neuroscience, University of Edinburgh 01/2020

Project: Building computational reinforcement model for Morris Water Maze

- Created poster which was presented at the 2020 Federation of European Neuroscience Societies (FENS) Forum
- Awarded by ZJE Student Overseas-exchange Scholarship
- Applied the principle of reinforcement learning to build a computational model to simulate the performance of mice in the Morris Water Maze behavior test, and estimate parameters for learning and memory ability
- Improved the model fitness from 20% to 80% by adding wall zone behavior simulation and refining parameter interval in MATLAB

Global Science Summer Program: Data Analysis, National University of Singapore Singapore

Course final project: Building customer comments-based recommendation system for hotels in Milan 08/2019

- Applied natural language processing to extract keywords from customer comments of hotels in Milan and correlate them with customer rating scores using Python packages
- Built a recommendation system based on the scores and keywords with R package "recommenderlab" and visualized the plots in Tableau. Both projects were marked as excellent (90+).

EXTRACURRICULAR ACTIVITIES

Head of Activity Department, CUMC CSSA, Columbia University 9/2021-now

- Made the organization's yearly budget plan and activity schedule
- Launched campus-level events such as Autumn hiking and practiced leadership by allocating works to members

Outstanding Participants, 17th Qiangyin Plan for entrepreneurship, Zhejiang University 11/2018-04/2019

- Wrote a business proposal for Anji Environment Protective Center to promote a sustainable business and public welfare model. The proposal was selected from 40 competitors to be presented on a public roadshow

President, Residential College Student Committee, ZJU International Campus 09/2018-06/2019

- Organized school-level events such as High table dinner (350 participants) and allocated related works to 20 staff
- Represented university to participate in the 5th Cross-Strait Forum on Education of Modern Colleges in Hong Kong
- Established connection with Residential College in more than 4 universities (Oxford University, HKU, University of Macao, SUSTC, Fudan University, etc) and launched exchange events, improved the reputation of Residential College globally.