

Experience

Memorial Sloan Kettering Cancer Center

March 2021 to Present

Data Analyst/Engineer with Andrew Vickers – (R, Python, Docker, Git, Bash/Zsh)

New York, NY

- Developed dcurves – Python package implementing Decision Curve Analysis or DCA
- Designed robust, R data pipelines for processing data pulled using SQL from institutional database, enabling surgeons to view their patient rates in comparison to their peers on a web app
- Develop AI radiology data pipeline implementing fine-tuned Llama3.1 and image segmentation model with RShiny dashboard to assist radiologists
- Built instructional website with forum for DCA (decisioncurveanalysis.org) and answer technical questions on DCA
- Authored R scripts for data cleaning and statistical analyses, transforming raw data to publication-ready reports for surgeons; applied statistical techniques like GAM, multivariable logistic regression, log-rank tests, Kaplan-Meier and CoxPH models, DCA; currently working with collaborators on seminal paper evaluating MSKCC and external hospital policies around performing post-chemotherapy retroperitoneal lymph node dissection in testicular cancer patients
- Completed large patient comorbidity analysis using survey data from my team’s electronic patient take-home survey rolled out throughout the hospital; used Posit’s ‘pointblank’ package for data validation; used analysis to demonstrate value of survey in assignment of accurate billing levels to patient visits

Sema4

February 2020 to February 2021

Bioinformatics Intern – (Python, R, AWS, Git, Docker, Bash/Zsh)

Stamford, CT

- Utilized high-performance computing resources, including AWS EC2 instances, to process and analyze large-scale genomic datasets, handling terabytes of data stored in S3 buckets
- Created automation pipelines for structural variant (SV) detection using Bash, Python, WDL, AWS, Git on over 100 large-scale genomic datasets (100GB+ BAM files each) using Delly, SvABA, Manta, and DRAGEN by Illumina
- Designed and executed data processing pipelines for cleaning, organizing, and comparing results from multiple SV detection algorithms, enhancing the efficiency and accuracy of genomic analysis workflows
- Employed genomic R libraries and ggplot2 to create sophisticated data visualizations, including Circos plots, stacked bar plots, heat maps, and scatter plots, to effectively communicate complex genomic findings

Education

Icahn School of Medicine at Mount Sinai

March 2021

Master's in Biomedical Data Science (MS)

Rutgers University – New Brunswick

May 2019

Bachelor's in Biomedical Engineering, Chinese (BE)

Languages

Mandarin Chinese	Hindi	Japanese	Spanish
●●●●●	●●●●●	●●●□□	●●●□□

Publications

<https://scholar.google.com/citations?hl=en&user=eR7hro0AAAAJ>