Shaun Porwal

New York, NY   (732) 318-7592   shaun.porwal@gmail.com   https://shaunporwal.com   U.S. Citizen

in Shaun Porwal
shaunporwal

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

- Created **dcurves**, a Python package implementing Decision Curve Analysis (DCA), empowering researchers with an advanced statistical tool for clinical decision-making

- Engineered robust R data pipelines integrating SQL-extracted institutional data, enabling surgeons to benchmark patient outcomes through a web-based analytics app

- Developed an AI-driven radiology pipeline, fine-tuning Llama3.1 and image segmentation models, with results visualized via an interactive RShiny dashboard

- Built and maintain **decisioncurveanalysis.org**, an instructional website with a technical forum for DCA users, providing support and answering advanced user questions

- Led statistical analyses for a landmark study on post-chemotherapy RPLND policies in testicular cancer, applying cutting-edge methods (GAM, logistic regression, Kaplan-Meier, CoxPH, DCA) to produce actionable insights

- Conducted a hospital-wide patient comorbidity analysis using survey data, validating with Posit's 'pointblank,' and demonstrated the survey's effectiveness in improving billing accuracy for patient visits

Sema4   February 2020 to February 2021

Bioinformatics Intern - (Python, R, AWS, Git, Docker, Bash/Zsh)   Stamford, CT

- Processed large-scale genomic datasets (terabytes) using AWS EC2 and S3, leveraging high-performance computing (HPC) for efficient analysis

- Automated structural variant (SV) detection pipelines for 100GB+ BAM files with Bash, Python, WDL, and AWS, integrating tools like Delly, SvABA, Manta, and DRAGEN

- Streamlined genomic workflows by designing pipelines to clean, organize, and compare results from multiple SV detection algorithms, improving accuracy and efficiency

- Created data visualizations with genomic R libraries and ggplot2, including Circos plots, stacked bar plots, heat maps, and scatter plots, to present complex genomic findings

Education

Icahn School of Medicine at Mount Sinai   March 2021   **Rutgers University - New Brunswick**   May 2019

Master's in Biomedical Data Science (MS)   Bachelor's in Biomedical Engineering, Chinese (BE)

Languages

Mandarin Chinese	Hindi	Japanese	Spanish
■■■■■	■■■■■	■■■□□	■■■■□

Publications

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