

HARRISON WANG

PH.D. STUDENT

✉ harrison.c.wang@gmail.com

🌐 <https://harrisonized.github.io/>

in <https://www.linkedin.com/in/harrisonized/>

🐙 <https://github.com/harrisonized>

PROGRAMMING: Python, R, SQL, Matlab, Datalog

DATABASES & WEB FRAMEWORKS: Postgresql, Datomic, MongoDB, Flask, RShiny

CLOUD COMPUTING & DASHBOARDING: AWS, EC2, S3, Sagemaker, Redshift, Athena, Quicksight, Tableau, Power BI

MACHINE LEARNING: Regression, Classification, Clustering, Time Series Analysis, Natural Language Processing

WET LAB: qPCR, Flow Cytometry, Cloning, Mouse Studies, RNA FISH, ELISA, Tissue Culture, Protein Purification, Assay Development

EXPERIENCE

UNIVERSITY OF PENNSYLVANIA

Ph.D. Student

Aug. 2022 to Current

Laboratory of Dr. Roxane Tussiwand

Ph.D. Student in the NIH-Penn Immunology Partnership Program

Rotation work:

- Fourth rotation in Roxane Tussiwand's lab at the NIH, studying the role of DNase enzymes in autoimmunity.
Wet lab: PCR, flow cytometry, mouse take down, Dry lab: R, Flowjo
- Third rotation in Mariana Kaplan's lab at the NIH, studying sex differences in neutrophils.
Wet lab: human neutrophil isolation, RNA extraction, qPCR, Dry lab: Python
- Second rotation in Montserrat Anguera's lab, studying how X-inactivation contributes to female bias in autoimmune diseases.
Wet lab: RNA FISH and H3K27me3 immunofluorescence. Dry lab: R, image processing using Python
- First rotation in Michela Locci's lab, exploring the role Activin A in rheumatoid arthritis.
Wet lab: tissue culture, mouse IP and tail dermal injections, anesthesia, euthanasia, flow cytometry. Dry lab: Flowjo, Python

THERMO FISHER SCIENTIFIC

Research and Development Data Scientist

South San Francisco

Feb. 2021 to Aug. 2022

- Built a **scalable, end-to-end ETL process** to achieve **real-time monitoring** of SARS-CoV-2 RT-PCR test data from labs and universities using **Amazon AWS** and **Power BI**. Listed as third author on a **symposium abstract** based on this work
- **Lead the development** of the **data pipeline**. Heavily **refactored** and **optimized** the existing data pipeline for **scalability** through clever use of **parallelization** and **batch processing**, cutting the time it takes to transform 2000 csv files from 2 hours to 3 minutes. Add new **KPIs** and support **backward compatibility**
- **Analyzed** performance of sample retests using **Python** and **SQL**, enabling stakeholders to understand the limit-of-detection of our **SARS-CoV-2 RT-PCR** tests
- Designed dashboards in **Amazon Quicksight** and **Power BI** to monitor data ingestion and filtering, eliminating up-front time in scoping projects for data projects
- Created new **Plotly interactive visualization** for 96- and 384-well plates, enabling the Data Science team to rapidly **identify anomalies** in new data
- Saved a legacy **Matlab** codebase from disrepair through heavy **refactoring**

INVITAE

Dry Lab Operations Data Scientist

San Francisco

Jan. 2020 to Sept. 2020

- Discovered the **root cause** for a high-visibility, time-sensitive issue on a new assay that resulted in batch failures impacting 500+ samples
- Analyzed **trends** using **Python** and **SQL** to **monitor next-generation sequencing (NGS)** production-line data for anomalies
- Efficiently diagnosed **customer issues** by writing a **Python script** to **automate** generation of standardized **visualizations** and **statistics**
- Wrote maintainable **Datalog** and **SQL** queries for **Flask** app used to migrate data from **Datomic** to **Amazon Redshift** for consumption by **Tableau**
- Built an 8-page **Tableau dashboard** to enable our operations team to carefully monitor a new product launch

SANOFI, FORMERLY TRUE NORTH THERAPEUTICS

Research Associate 2

South San Francisco

Jan. 2017 to Mar. 2019

- Led **protein-engineering project** on of lead antibody drug (BIVV009). Handled all **cloning and expression** of antibody variants and most of protein purification for the lab. **Collected binding and efficacy (KD and IC50) data** and performed **regression analysis**, revealing a log-linear relationship between KD and IC50. This enabled us to select which variants to use in downstream experiments.
- Led **research project** to make a protein complex of BIVV009 and its target (C1s of the complement immune system) for crystallography study.
- Wrote **Python script** to automate design of short DNA oligos, reducing a task that normally took 2 hours/week of manual labor to a 3 minute script.
- Wet lab: tissue culture, protein expression and purification, ELISAs, affinity and kinetics characterization, crystallography

GENE YEO LAB, UCSD

Staff Research Associate 1

La Jolla

May 2013 to Nov. 2016

- Developed **standardized protocols** to create tagged cell lines for **next-generation sequencing** (ENCODE)
- **Co-authored a Neuron paper** that included my experiments on elucidating the mechanism for how a mutation in an RNA-binding protein can cause ALS
- **Co-authored a Cell paper** that included my experiments on using a new genome-editing technology (CRISPR/Cas9) to track RNA in live cells
- Wet lab: **fluorescent cell imaging**, **next-generation sequencing**, high-throughput sequencing, CRISPR-Cas9, **qPCR**

EDUCATION

Metis Data Science Bootcamp

Apr. 2019 to June 2019

University of California, San Diego

2015

Double Major: B.S. Physics, B.S. Physiology & Neuroscience