

# HANA LEE

---

[hanalee07@gmail.com](mailto:hanalee07@gmail.com) • [hanalee.info](http://hanalee.info) • [github hnlee](https://github.com/hnlee) • [linkedin hanalee07](https://www.linkedin.com/in/hanalee07)

## SUMMARY

- Scientist interested in exploring and analyzing big data to solve meaningful problems.
- Former career as a genetics and genomics researcher, with expertise in analysis of next-generation sequencing data sets and scientific computing for the life sciences.

## ROLES AND EXPERIENCES

### *Data analysis and visualization*

- Built data cleaning and processing pipelines in Python for terabyte-scale data sets
- Used R for exploratory data analysis and statistical testing, including nonparametric methods, generalized linear models, and multiple testing correction
- Wrote Python script to semiautomate colony counting on images
- Rewrote R code developed by collaborator in Python with data storage in SQLite for faster implementation

### *Teaching and communication*

- Introduced undergraduate and doctoral students to programming through Python and R
- Co-taught a workshop for scientists on analyzing next-generation sequencing data in R
- Expert at communicating complex information to nonspecialist audiences through data visualization and oral presentations

### *Machine learning projects*

- **San Francisco crime classification**
  - Developed model averaging predictions from logistic regression and random forests using Python as project for Chicago Python User Group (ChiPy) Fall 2015 Mentorship Program
  - Submitted to Kaggle competition and placed in top 25% of the leaderboards
  - Presented work at the ChiPy January 2016 meeting (see [slides](#))
- **Gas mileage prediction:** Developed generalized linear model trained on fuel economy data set using R, and created webapp using shiny to deliver mileage predictions and display plots in response to user input (see [webapp](#))
- **Bicep curl classification:** Developed random forests model trained on accelerometer data set using R, and wrote report in knitr (see [report](#))

### *Web development*

- Maintained a donations website written in web2py for alumni organization
- Created backend in django with SQLite database for a website to manage event invitations
- Built web interface in django with MySQL database to crowdsource and display media metadata

## EDUCATION

- 2012 **Ph.D.** Molecular Biology, University of California, Berkeley
- 2007 **A.B.** Biochemical Sciences, Harvard University