HANA LEE

hanalee07@gmail.com • hanalee.info • **github** hnlee • **linkedin** 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.

SKILLS

- Scientific computing and visualization

  - **R:** `caret`, `ggplot2`, `shiny`, `knitr`
  - **Python:** `numpy`, `pandas`, `scikit-learn`, `seaborn`, `jupyter`

- Relational databases: **SQLite**, **MySQL**

- Web frameworks: **Django**, **web2py**

ROLES AND EXPERIENCES

- **Data analysis and visualization:** I built data cleaning and processing pipelines in Python for terabyte-scale data sets and used R for exploratory data analysis and statistical testing, with nonparametric methods, generalized linear models, and multiple testing correction.

- **Teaching and communication:** I introduced undergraduate and doctoral students to programming through Python and R. I also co-taught a workshop for scientists on analyzing next-generation sequencing data in R. I am expert at communicating complex information to nonspecialist audiences through data visualization and oral presentations.

- **Self-directed learning:** As side projects, I built a classification model to predict exercise movements, created a webapp to deliver predictions on gas mileage, developed a web interface to crowdsource media metadata, wrote the backend for a website to manage invitations, and maintained a donations website. I am currently working on the Yelp Dataset Challenge and Kaggle competitions.

EDUCATION

- 2012 **Ph.D.** Molecular Biology, University of California, Berkeley

- 2007 **A.B.** Biochemical Sciences, Harvard University