# Lan Huong Nguyen

PhD Candidate · ICME, Stanford University

2101 California St, Mountain View, CA 94040 · (+1) 626-200-7792 · lanhuong@stanford.edu

⌂ http://nlhuong.github.io | github.com/nlhuong | in nlhuong

## Education

**Stanford University**                                                                                                   Stanford, CA
PhD in Computational and Mathematical Engineering (GPA 3.9)                          Sep. 2013 - Exp. Jun. 2019
- Robust non-linear dimensionality reduction and latent structure recovery. Advised by Prof. Susan Holmes.
- Relevant coursework: numerical linear algebra, convex optimization, stochastic processes, probabilistic graphical models, distributed algorithms, statistical learning theory, Bayesian statistics, algorithms in advanced machine learning, theories of deep learning, statistical and machine learning methods for genomics

**California Institute of Technology**                                                                              Pasadena, CA
B.S. in Applied Mathematics and Economics (GPA 3.8)                                           Sep. 2009 - Jun. 2013
- Graduated with honors from the Department of Computing & Mathematical Sciences.

## Experience

**Google for Clinicians**                                                                                              Palo Alto, CA
Senior Data Scientist                                                                                            Oct. 2021 - Present
- Oversee efforts to ensure high rigor, consistency and scalability across evaluation work, as Tech Lead for a DS subteam.
- Establish strategies to accurately assess performance of clinical tools under development by defining success and quality metrics, designing evaluation templates, implementing robust statistical methodologies, and conducting data analysis.
- Provide analytical leadership for cross-functional multi-year projects that have culminated in successful launches of at least two key product features: Conditions tab - a patient problem list automatically extracted from text, and Second Level Insights - a collection of medical information related to specific conditions algorithmically retrieved from EHR.

**Google Health**                                                                                                       Palo Alto, CA
Data Scientist                                                                                                      Oct. 2019 - Sep. 2021
- Conducted quality evaluation of search and browse tools for clinicians built on top of Electronic Health Records data.
- Performed statistical analysis and data exploration to guide product development and feature enhancements.
- Designed and developed frameworks to assess performance and robustness of prediction models.

**High-dimensional Data Statistics Graduate Research**                                          Stanford, CA
Researcher in Holmes Lab                                                                                    Sep. 2014 - Sep. 2019
- Designed robust dimensionality reduction and manifold learning methods for inferring latent structures and associated uncertainties in high-dimensional data.
- Built statistical models for single cell RNA-seq and microbiome data.
- Developed `R` and `Shiny` software packages implementing new data analysis and visualization techniques.

**R for Data Science Introductory Course**                                                             Stanford, CA
Course Instructor for CME/STATS 195                                                                    Fall '16, '17, '18
- Created a short course on R application to data science; topics covered: data wrangling, exploration, visualization and modeling, `http://cme195.github.io`.
- Taught a class of 60+ graduate and undergraduate students each year.

**Genentech**                                                                                                   South San Francisco, CA
Intern in Department of Bioinformatics                                                                   Jun. 2017 - Oct. 2017
- Studied the role of COP1 in activation and repression kinetics of LPS-induced genes.
- Developed an open-source Bioconductor package for analyzing and visualizing short time-course RNA-seq data.

**AOL Advertising**                                                                                                   Palo Alto, CA
Intern in Large-Scale Analytics Team                                                                      Jun. 2014 - Oct. 2014
- Analyzed DoubleClick ad exchange auction bid data consisting of ∼4-5 million entries per day.
- Estimated market trends and winning price distribution. Designed Real Time Bidding algorithms.

**Talentoday Startup**                                                              Palo Alto, CA
Consultanting Project                                                   Sep. 2012 - Feb. 2013
- Worked with 30k+ user data to find the most predictive survey questions using QR factorization, and sparse CCA.
- Developed a matrix completion answer-imputation technique reducing the number of required questions by 30%.

## Skills

**Programming**    R, Stan, Python, Matlab, C++, Spark, Scala, SQL
**Languages**         English, Polish, Vietnamese

## Publications and Packages

A. Feder, I. Laish, S. Agarwal, U. Lerner, A. Atias, C. Cheung, P. Clardy, A. Peled-Cohen, R. Fellinger, H. Liu, L. H. Nguyen, et al.. **Building a Clinically-Focused Problem List From Medical Notes.** Proceedings of the 13th International Workshop on Health Text Mining and Info Analysis (2022) `https://aclanthology.org/2022.louhi-1.8`

M. Roodgar, F. P. Suchy, L. H. Nguyen, et al.. **Chimpanzee and pig-tailed macaque iPSCs Improved culture and generation of primate cross-species embryos.** Cell Reports (2022) `https://doi.org/10.1016/j.celrep.2022.111264`

M. Roodgar, A. Babveyh, L. H. Nguyen, et al.. **Chromosome-level de novo assembly of the pig-tailed macaque genome using linked-read sequencing and HiC proximity scaffolding.** GigaScience (2020) `https://doi.org/10.1093/gigascience/giaa069`

J.A. Grembi, L.H. Nguyen, et al.. **Gut microbiota plasticity correlated with sustained weight loss after a low-carb or low-fat dietary intervention.** Scientific Reports (2020) `https://doi.org/10.1038/s41598-020-58000-y`

L.H. Nguyen, S. Holmes. **Diffusion t-SNE for manifold learning.** Machine Learning in Computational Biology (2019) `https://mlcb.github.io/mlcb2019_proceedings/papers/paper_45.pdf`

L.H. Nguyen, S. Holmes. **Ten quick tips for effective dimensionality reduction.** PLOS Comp. Biology (2019) `https://doi.org/10.1371/journal.pcbi.1006907`

L.H. Nguyen., S. Holmes. **Bayesian Unidimensional Scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations.** BMC Bioinformatics (2017) `https://doi.org/10.1186/s12859-017-1790-x`

L.H. Nguyen. **'TimeSeriesExperiment' package for analysis and visualization of short time course data.** Bioconductor (2018) `https://doi.org/doi:10.18129/B9.bioc.TimeSeriesExperiment`

P.H.T. Kamga, B. Li, M. McKerns, L.H. Nguyen, M. Ortiz, H. Owhadi, and T.J. Sullivan. **Optimal uncertainty quantification with model uncertainty and legacy data.** Journal of the Mechanics and Physics of Solids, 72 (2014) `http://dx.doi.org/10.1016/j.jmps.2014.07.007`