INSIGHT DATA SCIENCE

INSIGHT

260 Sheridan Ave Suite 310
Palo Alto, CA 94306

T 650-279-8908
F 650-284-0456
info@insightdatascience.com

June 2014 Insight Fellows: Profiles and Projects

Adam Morgan
University of California at Berkeley
PhD, Astrophysics

CharityVerity.com
CharityVerity: Predicting charity ratings to guide effective altruism.

Ahna Girshick
University of California at Berkeley
PhD, Vision Science & Computer Science

Wikiscore.co
Wikiscore: The Wikipedia page scorer.

Aparupa Das Gupta
University of California at Los Angeles
PhD, Operations Research

PriceWizard.us
Price Wizard: Recommending the right price for selling items on Craigslist.

Aravind Natarajan
University of Pittsburgh
Postdoctoral Scholar, Physics

FunWithASL.net
Signs: Learn English through ASL!

Asif Imran

Wisconsin IceCube Particle Astrophysics Center
Postdoctoral Scholar, Astrophysics

Git-Discover.com
Git-discover: Unlock the potential of Github.

Benoit Dherin

University of California at Berkeley
PhD, Mathematics & Statistics

GradeExpectation.org
GradeExpectation: Piazza forum analyzer.

Carlos Cunha

Stanford University
Research Associate, Astrophysics

JauntApp.net
Janut: Recommendations along the way!

Christopher Klein

University of California at Berkeley
PhD, Astronomy

WhatsForDinner.us
What's for Dinner: Recommending dinner to fit your nutritional goals.

David Paulsen

University of Pittsburgh
PhD, Cognitive Neuroscience

RxFxFinder.com
RxFx: Prescription drug recommender system.

Dirk Neumann

California Institute of Technology
Postdoctoral Scholar, Computation & Neural Systems

nuanceq.com
Nuance: Got a Feeling?

Gaston Sanchez

University of California at Berkeley
Postdoctoral Scholar, Statistics

SkinCareProds.info
SkinCareProds: Scoring personal care products.

Heather Shapiro

University of California at Davis
Postdoctoral Scholar, Cognitive Neuroscience

Whenzi.me
Whenzi: Predicting when patients will be funded on Watsi.

Igor Bespamyatnov

University of Texas at Austin
Postdoctoral Scholar, Plasma Physics

News-Domains.net
NewsNet: Graph Recommender for news websites.

Jason French

Northwestern University
PhD, Differential Psychology

SunnySideUp.solar
Sunny Side Up: Removing the barrier to solar energy estimation.

Jesus Martinez-Manso
University of Florida
PhD, Astrophysics

EdXPredictor.com
EdXPredictor: Find out your chances of certification!

Kelty Allen
University of California at Berkeley
PhD, Mathematical Logic

AmazonExplorer.info
AmazonExplorer: A deeper look into Amazon product reviews.

Kevin Ford
University of California at San Francisco
Postdoctoral Scholar, Biophysics/Neuroscience

tapmapper.us
TapMapper: Find out what the locals are drinking/tweeting.

Matthew George
University of California at Berkeley
PhD, Astrophysics

StreetsAheadMaps.com
StreetsAhead: Image recognition with Google Street View.

Meredith Trotter
Stanford University
Postdoctoral Scholar, Biology: Evolutionary and Demographic Theory

Kick-Or-GoGo.com
Should I Kick or Should I Go? : Optimize your crowd funding campaign.

Michael Ramm
University of California at Berkeley
PhD, Physics

hoppr.me
Hoppr: Route planner and recommender.

Michael Woods
University of California at Davis
PhD, Physics

MelodyMapr.com
MelodyMapr: Your road trip radio companion.

Monica Hsu
University of California at San Francisco
PhD, Biophysics

truValuation.com
truValuation: Find the true value of homes in San Francisco.

Nicholas Peterson
Ohio State University
PhD, Mathematics

math.stackexpert.net
Math.StackExpert: Finding you high-quality questions on Math.StackExchange that you'll WANT to answer!

Nicolas Tilmans
Stanford University
PhD, Biochemistry

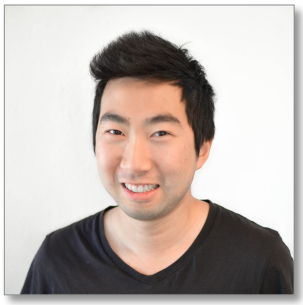
DrugScope.info
DrugScope: Scope out your medicine.

Nicole Romano
Stanford University
PhD, Materials Engineering (Neural Regeneration)

SetListEngine.com
{SetList]: Set your life to music.

Ravi Menon
University of California at San Diego
Postdoctoral Scholar, Earth & Ocean Sciences/Signal Processing

SearchPy.net
SearchPy: Find relevant Python code on Stack Overflow.



Ray Gao
University of Toronto
PhD, Physics

UpFindr.com
UpFindr: Find your activity partners.

Sara Salha
University of California at Los Angeles
PhD, Physics

CraigSublets.com
CraigsList++: Insightful search for room sublet.

Sumin Tang
Caltech / University of California at Santa Barbara
Postdoctoral Scholar, Astrophysics

Movie2Books.com
Movie2Books: Find books you may like based on your favorite movie.

Vanessa Heckman

California Institute of Technology
PhD, Civil Engineering & Geophysics

Lenderstanding.com *username: guest password: insight123*
Lenderstanding: Helping lenders help: Credit risk model for Zidisha.

Vickie Zhang

University of California at Berkeley
PhD, Bioengineering

VentureNetwork.us
VentureNetwork: Recommend high potential startups based on investor's portfolio.

Wafa Soofi

Georgia Institute of Technology / Emory University
PhD, Biomedical Engineering & Computational Neuroscience

CrowdSkippr.com
CrowdSkippr: Tell us where. We'll tell you when.

Woei Ling Leow

Lawrence Berkeley National Laboratory
Postdoctoral Scholar, Energy Efficiency

AnalytiCar.org
AnalytiCar: Pick a car, driven by data!

Zack Chadick

University of California at San Francisco
Postdoctoral Scholar, Systems Neuroscience

StealThisHouse.info
Steal This House: Crime doesn't pay…'cept for your bottom line!

linkedin.com/in/qmorgan
github.com/qmorgan

Adam N. Morgan

qmorgan@gmail.com
(510) 229-7683

Experience

Insight Data Science Fellow (Palo Alto, CA) *June 2014 - Present*

CharityVerity.com - Predicting charity ratings to guide effective altruism

- Created CharityVerity.com, a web application to predict charity rankings
- Scraped ~7500 charity rankings from evaluator CharityNavigator.org to build training set
- Built a prediction model using random forest regression on IRS data for 200K+ charities
- Deployed app and database (MySQL) to AWS, using Flask as the web framework

Independent Side Projects (Berkeley, CA) *July 2013 - Present*

Automated image analysis pipeline to quantify biological cell migration

- Collaborated with a bioengineer to automate the image analysis of hydrogel image stacks
- Quantified cell migration by wrapping a custom Python detection association code around software designed to catalog objects from astronomical images

Disaggregation of household energy consumption into individual appliance use

- Built a framework to predict appliance energy consumption based on a time series of electronic signatures for the Kaggle project "Belkin Energy Disaggregation Competition"
- Developed a solution with three collaborators that placed 7th out of 165 entries (top 5%)

Graduate Student Researcher at UC Berkeley (Berkeley, CA) *July 2008 - May 2014*

Projects focused on real-time classification and analysis of time-series and imaging data utilizing a variety of parametric and non-parametric statistical methods (including parameter estimation, classification/regression, and model selection). Built a framework of software pipelines for automated data reduction and analysis.

Decision engine for rapid, machine-learned resource allocation

- Established a collaboration (3 astronomers & 3 statisticians) to develop a process to make rapid telescope time allocation decisions to capture rare, high-priority astronomical events.
- Trained a classifier using Random Forests in *R* within a larger Python framework
- Pipeline autonomously triggers on new events to parse, classify, & send results to web/email
- Results published in Morgan et al. 2012, *The Astrophysical Journal*, 746, 170

Multivariate time-series modeling of a complex astrophysical event

- Led an international collaboration (20 people from 7 institutions) to model the behavior of Gamma-ray burst 120119A, which showed significant levels of early-time color change
- Collated noisy, asynchronous time-series data across several telescopes to model the physical properties of the explosion and the surrounding environment using SciPy
- Provided the strongest evidence thus far for dust destruction around a gamma-ray burst
- Results published in Morgan et al. 2014, *MNRAS,* 440, 1810

Teaching and Outreach

- Organizer and web developer (AWS EC2 + RDS) for PhD career education conference Beyond Academia (beyondacademia.org).
- Helped develop & run annual 3-day UC Berkeley Python Boot Camp for four years
- Two-time graduate student instructor for "Python Computing for Data Science" course at UC Berkeley.

Education

- **University of California, Berkeley** - Berkeley, CA - ***National Science Foundation (NSF) Fellow***
  Ph.D. in Astrophysics *July 2008 – May 2014*
- **The University of Cambridge** - Cambridge, United Kingdom - ***Marshall Scholar***
  M.Ast in Mathematics *Oct. 2007 – June 2008*
  M.Phil in Physics (Astronomy) *Oct. 2006 – Sept. 2007*
- **The Pennsylvania State University** - State College, PA - ***Goldwater Scholar***, ***Astronaut Scholar***
  B.S. in Astronomy and Astrophysics with Honors; B.S. in Physics *Aug. 2002 – May 2006*

Skills

- **Languages:** Python, MySQL, prior experience with R
- **Tools:** Git, SciPy, NumPy, pandas, matplotlib, scikit-learn, Flask, d3.js

Ahna Girshick

510.604.1567
San Francisco, CA
ahna.girshick@gmail.com

@AhnaGirshick
lightdark.org
github.com/ahna
linkedin.com/in/ahnagirshick

Skills Summary

Languages: Python, Matlab, SQL, C++, C, OpenGL, Mathematica
Tools: Unix, scipy, numpy, Flask, UIKit
Some exposure: R, NLTK, Lisp, HTML, CSS, d3.js, Cinder, Objective C, Cinder

Experience

Fellow, INSIGHT DATA SCIENCES – 2014-present

- Created Wikiscore.co, a web app that analyzes and rates Wikipedia pages.
- Scraped 100k+ Wikipedia pages using Python and stored results in a MySQL database on AWS.
- Trained a random forest classifier and analyzed Wikipedia's readability, using Scikit-learn and NLTK.
- Developed the web app using Python, Flask, Twitter Bootstrap, HTML, Jinja2, CSS.

Senior Producer, SNIBBESTUDIO.COM (now eyegroove.com) – 2012-2013

- Managed startup team of 6 engineers and designers to create interactive music and visualization apps for major musicians including Björk, Philip Glass, Metric, and Passion Pit.
- Produced 6 highly-acclaimed apps for iOS, Mac, Android, Windows, LEAP Motion, using iOS, C++, UIKit, OpenGL, Cinder, Objective C. Conceptualized apps, designed user experience designs, spearheaded social integration and real-time app usage analytics, contributed to production code.
- Received coverage from the App Store, *WIRED*, *Rolling Stone*, and 2 Webby Honoree Awards.

Postdoctoral Fellow, UNIVERSITY OF CALIFORNIA, Berkeley, CA. – 2011-2012

- Conducted web experiments to enhance data visualization design for optimal visual understanding, using d3.js, Matlab, Mechanical Turk.

Postdoctoral Fellow, NEW YORK UNIVERSITY, New York, NY – 2008-2010

- Received 3-yr NIH NRSA Fellowship to develop a model to predict biases in visual perception.
- Conducted visual psychophysics experiments in lab resulting in more than 1 million data points.
- Developed probabilistic machine learning model to reverse-engineer the brain's prior knowledge of edge orientation, using optimization, bootstrapping and model comparisons in Matlab.
- Measured orientation statistics at multiple spatial scales in a large set of photos and discovered close match to human prior knowledge.
- Simulated a neural network that embeds the prior knowledge from image statistics and whose behavior closely matches human perceptual biases, in Matlab.
- Published results in *Nature Neuroscience* and received press in *Science News, NPR, Nature.*

Graduate student researcher, UNIVERSITY OF CALIFORNIA, Berkeley, CA. – 2001-2007

- Received 4-yr DOE Computational Sciences Fellowship to research the mechanisms of visual perception of images and digital displays.
- Developed probabilistic models of visual perception of 2D and 3D displays, and of sensory integration of multiple sources of information (stereo, perspective, focus, haptics).
- Designed and conducted visual psychophysics experiments in the lab using C, C++, OpenGL.
- Analyzed results and conducted Monte Carlo simulations in Matlab to support hypotheses.
- Published results in *Nature Neuroscience, SIGGRAPH* and received press in *NPR*, *NY Times*.

Education

Vision Science, University of California, Berkeley, CA – PhD, 2007
Computer Science, University of Minnesota, Twin Cities, MN – MS, 1996
Computer Science, University of Minnesota, Twin Cities, MN – BS, 1993

Aparupa Das Gupta

Ph: 707-364-3348 ; dg.aparupa@gmail.com
https://github.com/aparupadg; https://www.linkedin.com/in/aparupadg

EDUCATION

- University of California, Los Angeles
  Ph.D. in Operations Management 2009-2014
  M.S. in Statistics 2014 (expected)
- Purdue University, Indiana
  M.S. in Industrial Engineering 2006-2008
- Indian Institute of Technology, Kharagpur, India
  Integrated B.S. and M.S. in Mathematics and Computing 2001-2006

EXPERIENCE

Fellow at Insight Data Science : June 2014-Present

- Developed pricewizard.us, a web application in Python that recommends appropriate selling-prices for any seller of used furniture items on Craigslist.
- Scraped (using Scrapy) Craigslist website for all furniture advertisements posted in San Francisco for three weeks, extracted structured information using text processing, and stored the data in MySQL.
- Built price prediction model using linear regression (with scikit-learn) on each category of items that was sold previously, and used it to recommend the price for a seller's item. Also computed histogram of price distribution and word-cloud of top adjectives/nouns (with NLTK) for similar items sold earlier.
- Designed an interactive front end using Flask and Javascript. Hosted the website on Amazon S3.

M.S. Statistics, Relevant Projects, UCLA : 2013-2014

- Used logistic regression and SVM models with high-dimensional feature vectors constructed from IMDb database to predict box office performance of movies. Implemented in Python (with scikit-learn)
- Studied competition between four brands of a consumer product category by using a multinomial logit choice model for a retailer's scanner panel data and revealed preference segments. Implemented in R.
- Analyzed Istanbul stock exchange (ISE) return data using kernel density estimation and applied kernel and linear regression models to find any significant effect of national holidays and other international stock exchange return indices on ISE. Implementation was done in R.

Ph.D. Research, Mathematical Modeling of Service Encounters, UCLA: 2009-2014

- Analyzed optimal scheduling of all activities of a service experience to maximize customer satisfaction.
- Proposed a non-linear programming model for customer satisfaction from service experience by employing evidence from behavioral studies. Proposed prescriptive recommendations and heuristics for scheduling activities. Implementation was done using GAMS and MATLAB.

Summer Intern, Reliance Infocomm, India: Summer 2005

- Developed an image morphing application for Reliance camera phones with the option of adding animation effects to them. Implementation was done in J2ME.

SKILLS

- Languages: Python, SQL, R, MATLAB, GAMS
- Relevant Courses: Machine Learning, Convex Optimization, Integer Programming, Statistical Programming, Monte Carlo Methods for Optimization

AWARDS AND HONORS

- Honorable mention (from top 3), INFORMS Behavioral Operations Best Working Paper Award (2013)
- Third place, INFORMS Service Science Best Student Paper Award (2013)
- Recipient of Doctoral Fellowship, UCLA Anderson School of Management (2009-2013)
- Recipient of Frederick N. Andrews Fellowship, Purdue University, (2006–2008)

Aravind Natarajan

Palo Alto, CA | anat01@me.com | (412) 352 9178
aravindnatarajan.com | github.com/aravindnatarajan | linkedin.com/in/anat01

Experience

FELLOW AT INSIGHT DATA SCIENCE, PALO ALTO **JUNE 2014 - PRESENT**

- Developed SIGNS, a tool to teach English to deaf children using American Sign Language.
- Obtained sign video data from HandSpeak.com and stored the data in a MySQL database. Corpus of words pertaining to children's literature was obtained from the Gutenberg project.
- Performed language modeling, part-of-speech tagging, phonetic analysis, and word clustering analysis using Python's Natural Language Processing and Machine Learning toolkits.
- Front end developed using Flask and Javascript, and hosted at funwithasl.net using AWS.

RESEARCH ASSOCIATE, PITTSBURGH PARTICLE PHYSICS, ASTROPHYSICS, AND COSMOLOGY CENTER **SEP 2013 - MAY 2014**
MCWILLIAMS FELLOW AT THE CENTER FOR COSMOLOGY, CARNEGIE MELLON UNIVERSITY **SEP 2009 - AUG 2013**

- Analyzed data from the Sloan Digital Sky Survey using MySQL and Python's scikit-learn package, and built a classifier to identify quasars using Support Vector Machines.
- Worked with over 100 hours of radio data obtained from the Green Bank Telescope using C and Python, to perform cross correlations of radio maps with optical galaxy catalogs, to quantify the abundance of neutral Hydrogen.
- Contributed as a full member of the SCI-HI collaboration to collect and analyze radio data obtained from Isla Guadalupe, to study the masses and luminosities of the first generation of stars.
- Served as co-Producer and scientific advisor of a planetarium show titled ``The Hydrogen Sky''. Presented a public talk titled ``The Dark Matter Puzzle'' at the Pittsburgh Allegheny Observatory.

POSTDOCTORAL FELLOW IN PARTICLE ASTROPHYSICS, BIELEFELD UNIVERSITY, GERMANY **OCT 2007 - JULY 2009**

- Worked on the cosmological implications of particle dark matter through a Markov Chain Montecarlo analysis of cosmic microwave background data, using codes written in Python and C.
- Studied the infall of dark matter on to the Milky Way Galaxy using computer simulation code written in C, to identify regions of high dark matter density in the solar neighborhood.

Education

- Ph.D (Physics), University of Florida at Gainesville — Aug 2002 - Aug 2007
- Bachelor of Engineering (Electronics and Communication), Bangalore University, India — Sep 2000

Skills

- Machine Learning and Natural Language Processing using scikit-learn and NLTK. Bayesian estimation of model parameters using Markov Chain Montecarlo simulations.
- Python, C, MySQL, Javascript.
- Published 26 papers in major journals, with over 300 citations.

Asif Imran

Palo Alto, CA
(515) 450-3117 • asif_imran@icloud.com • asifimran.org • @5sigma
www.linkedin.com/in/asifimranphd • Github:aimran

Experience

Insight Data Science — **Palo Alto, CA**
Fellow — *June 2014 – Present*

- Created git-discover.com, a web app designed to discover outstanding contributors on Github.
- Developed a Python framework to obtain 50,000 Github user profiles using the Github API and stored the data in a MySQL database.
- Implemented a custom recommendation engine based on page-rank algorithm to identify talented programmers by leveraging user profiles along with activity and popularity levels of Github repos.
- Deployed an interactive web app on Amazon S3 using Flask, JQuery and Twitter-Bootstrap.

Wisconsin IceCube Particle Astrophysics Center — **Madison, WI**
Research Associate — *2013 – 2014*

- Developed the main real-time gamma-ray flare monitoring system for the **HAWC Observatory**. The C++ application with a tightly integrated SQL-backend allowed HAWC to perform real-time searches for the first time.
- Designed Bayesian statistical algorithms to evaluate the performance of the monitoring system using Monte-Carlo simulations.
- Awarded a competitive *NASA* grant as the Co-Investigator to conduct joint observing campaigns with complementary observatories.
- For a complete listing of my publications, see: http://bit.ly/asif_imran_publication_list

Los Alamos National Laboratory — **Los Alamos, NM**
Postdoctoral Scientist — *2010 – 2013*

- Led data acquisition group to select and develop the principal data acquisition system for HAWC.
- Built and deployed the data acquisition system (C, zmq) from ground-up with a $> 99\%$ live time.
- Wrote C-based libraries for synchronized readout of an array of single board computers with a net throughput rate of 500 MBytes/sec – unprecedented in current generation of astrophysical experiments with comparable size and mission.

Iowa State University — **Ames, IA**
Graduate Student Researcher — *2004 – 2010*

- Developed a C++-based analysis pipeline to measure the density of diffuse radiation field that resulted in new limits on emissions from distant galaxies.
- Developed a Monte Carlo simulation package for the VERITAS collaboration that was central to characterizing the sensitivity of the telescope.
- Implemented an artificial neural network based energy estimation method for VERITAS that led to a 20% improvement in accuracy over previous measurements.

Skills

Languages: Python, C, C++, SQL

Tools: NumPy, SciPy, Matplotlib, Pandas, Scikit-learn, ROOT, Git/SVN, bash & regexp

Education

Iowa State University — **Ames, IA**
Ph.D., Astrophysics — *June 2010*

Grinnell College — **Grinnell, IA**
B.A., Physics — *May 2003*

Benoit Dherin

Phone (510) 220-2167 • **Github** BenoitDherin • **Email** benoit.dherin@gmail.com • **LinkedIn** linkd.in/1IpOmW
Teaching db.tt/XE0N3kMR • **Research** db.tt/yHuHYav5 • **Website** stat.berkeley.edu/~bdherin

Education

PhD, Mathematical Physics, Federal Institute of Technology, ETH Zürich **2004**

Master's of Science, Mathematics, Geneva University **1999**

Experience

Fellow, Insight Data Science **June 2014–present**

Created (www.gradeexpectations.org) a data-analysis and machine-learning web application and aimed at predicting and improving student performance by analyzing posting behavior in online class forums

- Stored student posts in MongoDB and extracted aggregated metrics on student-posting behavior using Python, PyMongo, NumPy, and pandas
- Trained a learning algorithm to predict student achievements from posting behavior using scikit-learn
- Created D3-generated force-directed graphs to convey insights about student predicted learning achievements from their posting behavior metrics
- Implemented the front end with Flask and Bootstrap, hosted on AWS

Lecturer, UC Berkeley Statistics Department **January 2013–June 2014**

Taught several core courses for the statistics-major curriculum, including:

- Game Theory (combinatorial and strategic multi-agent games)
- Concepts of Statistics (hypothesis testing, linear regression, parameter estimation)
- Concepts in Computing with Data (data science with Python and R)

Research Scientist **August 2006–December 2013**

(ETH Zürich, Geneva University, Utrecht University, UC Berkeley, and São Paulo University)

- Developed new differential geometric models for quantum mechanics
- Taught courses in linear algebra, ordinary differential equations, quantum mechanics, and quantum computing
- A list of publications can be found here: front.math.ucdavis.edu/search?a=dherin

Cofounder and Developer, Quadrature, Ltd **June 2000–August 2006**

- Built an innovative scheduling and accounting web application for the medical industry using genetic algorithms for scheduling appointments
- Developed a custom web framework in C++ generating dynamically HTML and JavaScript and serializing business objects into a MySQL database
- Prototyped the UI with Tcl/Tk and Java JFC/Swing

Skills

- Statistics, machine learning, data analysis, R, Python, SQL, and Unix
- Git, version control, and VirtualBox
- C/C++ and Java (prior experience)

Carlos E. Cunha

Menlo Park, CA - 734.757.0647 - carloscunha47@gmail.com - https://github.com/dcurl47

Technical Skills

◦ **Statistical techniques**: Markov chain Monte Carlo, neural networks, kd-trees, nearest-neighbors, deconvolution, spectral decomposition, optimization, principal component analysis, clustering, spectral connectivity analysis,numerical analysis of partial differential equations, bayesian and frequentist techniques, anomaly detection.
◦ **Programming languages:** C, Python, SQL, HTML, Octave, Gnuplot.

Professional Experience

Fellow at Insight Data Science June 2014 - Present

◦ Created Jaunt (www.jauntapp.net), a web app for finding Yelp recommendations along a driving route using Python/Flask, MySQL, jQuery and Javascript.
◦ Stored and filtered the Yelp Phoenix dataset in a MySQL database. Developed a ranking algorithm to select most relevant results. Expanded the scope of the project to the entire US by collaborating with data engineer Rick Seeger.
◦ Deployed app on Amazon Web Services using Google Maps, Mapquest and Yelp APIs.

Kavli Fellow at KIPAC - Stanford University September 2011 - Present

◦ Simulated performance of IFU spectrograph using Python and IDL, leading to addition of new instrument on board NASA planned billion-dollar satellite mission WFIRST.
◦ Improved nearest-neighbor searches using supervised linear and non-linear metric learning techniques developed in Python (with Scikit-Learn and Pybrain).
◦ Managed anomaly detection project using neural nets and SVM to search for satellite galaxies in large datasets.
◦ Developed simulations in C and Python to realistically assess, identify and correct the effects of incomplete and incorrect training data used for machine learning and model-based photometric redshift estimation.
◦ Pioneered a technique (and code in C) to use the clustering of galaxy clusters to constrain primordial non-Gaussianity and forecast the expected yields of upcoming surveys.
◦ Over 60 seminars and invited talks at international conferences and the most prestigious universities in ten countries.
◦ Over 20 scientific publications in the most prestigious journals in Astrophysics (ApJ, MNRAS, Physical Review).
◦ Elected builder of the DES survey, which grants lifetime authorship rights to all DES publications.

Postdoctoral Fellow at the University of Michigan September 2008 - August 2011

◦ Performed PCA analysis of time-series data from combined cosmological techniques using C.
◦ Led optimization of the survey strategy of the DES survey, a $>$ \$100 million survey involving 25 institutions in seven countries. Improved projected survey output figure-of-merit by 40% at no additional cost.
◦ Managed student project on spectral connectivity analysis to improve photo-z estimation.
◦ Implemented Fisher Information Matrix analyses of information content of cosmological surveys.
◦ Published a catalog of photometric redshifts of nearly 100 million galaxies from the Sloan Digital Sky Survey.
◦ Released a suite of C++ codes (github rep: probwts) that use nearest-neighbor techniques to identify and alleviate selection biases in training sets. Used by the most important ongoing surveys in the field (SDSS, DES, HSC).

Research Assistant at the University of Chicago October 2002 - August 2008

◦ Developed and tested machine learning (regression, neural nets and nearest-neighbors) as well as model-based techniques of photometric redshift estimation.
◦ Optimized filter shapes and exposure times of photometric surveys using an MCMC algorithm written in C.
◦ Evaluated system of around 30 coupled second-order partial differential equations to explore the effect of mass-varying neutrinos in the evolution of space-time and the energy-density fluctuations within it.
◦ Developed a cross-calibration technique to optimally combine information from imperfect experiments.

Education

The University of Chicago Chicago, IL
PhD and M.S in Astrophysics August 2008/June 2005

Brandeis University Waltham, MA
B.S in Physics with high honors and B.A.S in Mathematics May 2002
Physics Faculty Prize for 2002.
Summa Cum Laude.

Christopher R. Klein

Contact
ckleinastro@gmail.com
Berkeley, CA
(510) 859-7748

Links
http://github.com/ckleinastro
http://linkedin.com/in/crklein
http://twitter.com/ckleinastro

EDUCATION

University of California, Berkeley, Berkeley, CA
Ph.D., Astrophysics — May 2014
M.A., Astrophysics — December 2010

California Institute of Technology, Pasadena, CA
B.S., Astrophysics, *with honors* — June 2008

COMPUTER SKILLS

Languages: Python, SQL, matlab. Prior experience: R, C, Java.
Tools: NumPy, SciPy, matplotlib, PyMC, pandas, git, MySQL, Scikit-learn, Flask.

EXPERIENCE

Insight Data Science – *Fellow* — June 2014 – present
Palo Alto, CA

- Developed `http://whatsfordinner.us`, an online tool that recommends dinner compositions to meet nutritional goals and provides analysis and visualization of common food cuisines.
- Used Python, NumPy and MySQL to ingest and analyze 10 years of USDA food intake survey data (31,261 dinners) and the USDA Food and Nutrient Database (9,537 food listings).
- Leveraged word2vec to cluster food items and design themed dinners, as well as to learn about American eating habits and discover insights relating to cuisine and food pairings.
- Incorporated Flask, jQuery, d3, and Bootstrap to create a dynamic website front end.

University of California, Berkeley – *Graduate Student Researcher* — July 2008 – May 2014
Berkeley, CA

- Researched pulsating variable stars and measured astronomical distances with high precision.
- Formulated a novel Bayesian linear regression methodology to simultaneously fit period–luminosity relations using MCMC with Python and PyMC.
- Developed an automatic image reduction pipeline in Python leveraging MySQL telescope scheduling and observation databases, and built a webpage interface with Django.
- Conducted predictive simulations in Python for a NASA-proposed all-sky survey mission.
- Contributed to the software and hardware development of a new 6-band telescope camera.
- Published in *Mon. Not. Roy. Astron. Soc.* and *Astrophys. J.* as first author.
- Coauthored 20+ publications. Delivered 10+ oral and poster conference presentations.

University of California, Berkeley – *Graduate Student Instructor* — July 2008 – May 2014
Berkeley, CA

- Wrote & graded homework problems and contributed lectures for three semesters of undergraduate astronomy classes and two semesters of "Python Computing for Data Science."

California Institute of Technology – *Undergraduate Researcher* — June 2007 – June 2008
Pasadena, CA

- Developed Java observation planning program with GUI for Keck Telescope.
- Conducted undergraduate thesis research on high redshift galaxy stellar gas outflow rates.

David J. Paulsen

Palo Alto, California · paulsendavidjay@gmail.com · 541-510-1840

https://sites.google.com/site/paulsendavidjay · https://github.com/paulsendavidjay

www.linkedin.com/in/paulsendavidjay

EXPERIENCE

Insight Data Science **June 2014 – Present**
Fellow *Palo Alto, CA*

- Created and built RxFx (RxFxFinder.com), a web application that takes a user ranking of most commonly reported side effects to make personalized prescription drug recommendations.
- Utilized ~2.6 million records of openFDA data; reduced ~300k drug entries to ~90k using grep and Python pattern replacement, and n-gram similarity; stored and joined using MySQL.
- Generated algorithm for recommendations using side-effect ranking and linear optimization.
- Deployed on Amazon Web Services using Flask and Twitter Bootstrap.

University of Pittsburgh **September 2012 – June 2014**
Postdoctoral Research Scholar *Pittsburgh, PA*

- Designed and implemented analysis pipeline for a longitudinal neuroimaging study in Python and Bash.
- Deconstructed and analyzed 3-D imaging files using mixed-effects logistic regression, bootstrapping, false discovery rate correction, and parallel processing in R; reassembled results into 3-D format.
- Wrote program to simulate longitudinal data for estimating the effects of measurement reliability in mixed-effects models.
- Created and automatized MySQL database to monitor data acquisition and preprocessing.
- Published 1 paper (+1 in revision) in peer-reviewed journals; prepared 3 conference presentations.

Duke University **August 2007 – August 2012**
PhD Graduate Student *Durham, NC*

- Designed and conducted experiments on risky decision-making.
- Programmed gambling games in Matlab for neuroimaging, eye-tracking, and behavioral studies.
- Modeled choice behavior in R using logistic regression, eye-tracking, and neuroimaging data.
- Crafted custom eye-tracking data analysis pipeline to create quantified measures of emotion.
- Published 6 papers (4 as first author) in peer-reviewed journals; prepared 12 conference presentations.

EDUCATION

- Ph.D., Psychology & Neuroscience, Duke University, 2012
- M.S., Psychology, University of Oregon, 2006
- B.S., Psychology, Philosophy, University of Oregon, 2003

(see www.linkedin.com/in/paulsendavidjay for relevant graduate coursework and full publication list)

SKILLS & TOOLS

Languages: R, Python, SQL, Matlab, Bash
Tools: *ggplot*, linear mixed-effects models (*lmer*), MySQL, Pandas (some experience), Flask

SELECT AWARDS

- Mortimer D. Sackler Summer Institute for Developmental Neuroscience, Weill Cornell Medical College, 2012
- Honorarium, Scientific Research Network on Decision Neuroscience & Aging, Stanford Center on Longevity, 2011
- Predoctoral Ruth L. Kirschtein National Research Service Award, National Institutes of Health, 2010-2012

Dirk Neumann

dirk@caltech.edu . (626) 817 3475 . San Francisco, California

opani.com/dirkneumann . github.com/dirkneumann

Data Science Experience:

Insight Data Science, Palo Alto, CA — **Fellow** *Jun 2014—Jul 2014*

- Developed high-fidelity emotion recognition algorithm for natural language in Python: **nuanaceq.com**
- Deployed distributed analysis backend in Spark on Amazon EC2 cluster, using Google's Word2Vec model
- Served as an analytic consultant to find behavioral donation patterns for medical crowd-funding startup **watsi.org**

Opani.com, Pasadena, CA — **Founder / CEO, Full-Stack Hacker** *May 2010—Present*

- Developed the company from initial idea, product launch, led pivot into a profitable consulting business
- Contributed Python and racecondition-less Go production code powering **opani.com**
- Expert advisor on distributed load testing and behavioral analysis for mobile and web companies
- Authored Opani's data science blog that attracted 420,000 unique page views: **blog.opani.com**

University of Miami, Miami, FL — **Adjunct Assistant Professor** *May 2014—Present*

- Designed diagnostic classifier for autism using whole-brain scans from 1,000 participants in PySpark
- Researched brain-network interaction serving mental flexibility in autism and neurotypical controls

California Institute of Technology, Pasadena, CA — **Visiting Researcher** *Sep 2009—Present*

- Published papers and conducted research in disorders such as autism and Williams syndrome
- Developed high-throughput analytics pipeline for large medical data sets in Python
- Co-organized Bayesian seminar about linear mixed-effects models, Gaussian processes, Jags / RStan

APropell.com – Developed audience visualization and content recommendation *Jan 2014—Mar 2014*

- Ranked pages by engagement metrics from Google Analytics using a novel Page-rank-like algorithm

TheClosetScience.com – Created real-time recommender system for 1m+ items *Jun 2013—Dec 2013*

- Achieved CTR improvements between 2-15%, and developed iOS interactive analytics interface

University of Technology München, and **Tübingen**, Germany — **Research Assistant** *1998—2003*

- Developed self-localization vision software for mobile soccer robots using Bayesian and particle filters
- Researched brain waves decoder for locked-in patients using correlation dimension

FiANTEC, now Elaxy, Frankfurt, Germany — **Software Engineer** *1994—1996*

- Developed business software in SQL and Pascal for the financial service industry

Education:

California Institute of Technology, Pasadena, CA 2003—2009
Computation and Neural Systems, Ph.D., Minor in Genetics

University of Technology Munich, Munich, Germany 2001—2003
Computer Science, M.S.

Eberhard Karls University Tübingen, Tübingen, Germany 1998—2001
Psychology, M.S., Computer Science, B.S. / Vordiplom

Technical Skills:

Python, R, C, JavaScript, Node [5-10y] MongoDB, Spark, MySQL, HDFS [1-5y] Robotic soccers, distributed autonomous systems, particle filters, distributed machine learning, computational neuroscience [5-10y]
D3, jQuery, Bootstrap [5y] AWS, Open Stack Cloud, Ubuntu [5y] SciPy, Pandas, Scikit-Learn [1-5y]

Gaston Sanchez

(510) 697-0882
Berkeley, CA
gaston.stat@gmail.com

www.gastonsanchez.com
github.com/gastonstat
linkedin.com/in/sanchezgaston

EXPERIENCE

Fellow - *Insight Data Science.* (Jun 2014 - present)

- Created Web app "www.skincareprods.info" to score personal care products based on health risk.
- Scraped data from the Good Guide and EWG using R, stored database in MySQL.
- Applied Regression Trees with cross-validation for computing health scores of personal care products.
- Front end developed with Python (Flask). Visualization with D3js.

Postdoctoral Engineer - *Unit of Sensometrics and Chemometrics. ONIRIS-Nantes, France.* (Mar 2013 - Mar 2014)

- Built a prototype toolbox in R for multiblock data analysis methods.
- Lead seminars, training workshops, and tutoring on multivariate methods and data analysis with R.
- Created R packages tester, turner, and matrixkit for simplifying development of other packages.

Postdoctoral Researcher - *Nielsen Lab. University of California, Berkeley.* (Jun 2011 - Mar 2013)

- Performed research on statistical tests used in genetic association studies, and created the R package AssotesteR.
- Developed the R packages colortools, arcdiagram (arc diagrams) and DiscriMiner (discriminant analysis).
- Continued as guest researcher between March 2014 and May 2014.

Research Analyst II - *Biology Scholars Program (BSP). UC, Berkeley.* (Feb 2011 - Nov 2012)

- Established the roadmap for studying the scope of the Biology Scholars Program between 1994 - 2010.
- Carried out cluster analysis for segmenting program members in order to target specific audiences and needs.
- Conducted analysis to measure the impact of the program on its members. Applied statistical tests in R to compare the difference of the program members against students outside the program.

Research Analyst II - *Program for Reproductive Health & the Environment (PRHE). UC San Francisco.* (Aug 2011 - May 2012)

- Collaborated with PRHE's scientists on statistical analyses and bibliometric studies on Bisphenol A (BPA) datasets.
- Analyzed data from the biomonitoring study "Chemicals in Our Bodies". Analysis in R with statistical tests to identify threshold levels for different chemicals (metals and PBDA's).

Statistical Programmer - *LIAM. Universitat Politecnica de Catalunya, Spain.* (Jul 2005 - Jul 2009)

- Research on advanced methods for structural equation modeling and segmentation analysis.
- Developed the R package "plspm" for partial least squares path analysis (most popular PLS open source package).
- Carried out a customer satisfaction analysis for clients of Sabadell Bank. Analysis performed in R using structural equation models. Provided recommendations to increase Perceived Brand Image and Customer Loyalty.
- Built path models for the Barcelona School of Informatics Alumni Satisfaction Survey to evaluate perceptions between educational experience and job experience. Findings used to modify the syllabus of Computer Science degrees.

EDUCATION

PhD in Statistics - *Universitat Politecnica de Catalunya (UPC), Spain.*
(Sep 2004 - Jul 2009)

BA in Actuarial Science - *Instituto Tecnologico Autonomo de Mexico (ITAM), Mexico.*
(Jan 1997 - Jun 2002)

SKILLS

Data Analysis	Multiblock methods, dimension reduction techniques, partial least squares.
Software	R (expert), Python (basic), SQL (basic), HTML, CSS, Jekyll, LaTeX, Markdown.
Languages	Spanish (native), English (fluent), Catalan (fluent), French (conversational), Italian (conversational).

Heather Shapiro

San Francisco, CA
shapiro.heather@gmail.com
(908) 752-1032

www.heather-shapiro.com
www.linkedin.com/in/shapiroheather
www.github.com/shapiroheather

Experience

Fellow, Insight Data Science, Palo Alto CA, 2014–present
- Developed whenzi.me, a web app for predicting when patients on watsi.org will be funded
- Scraped watsi.org for data on more than 1800 patients in need of funding for medical treatments
- Applied Random Forest Regression to predict the time within which patients would attain funding
- Built and deployed app using Python, MySQL, Javascript, jQuery, Flask and Amazon Web Services

Co-Founder, NeuRecall Inc., San Francisco CA, 2013–present
- Co-Founded cognitive science-based technology startup built upon an adaptive learning platform
- Sourced investment leads and cultivated partnerships with educational organizations
- Conducted extensive market research and detailed reports on the competitive landscape
- Applied for accelerator programs (finalist at TechStars and LearnLaunchX)

Postdoctoral Scholar, UC Davis MIND Institute, Sacramento CA, 2013–2014
- Designed, developed, and managed studies of brain development in children with disorders
- Performed brain imaging (MRI and EEG) to examine brain structure and function in children
- Analyzed and visualized data using Matlab and R

Intern, Mission Bay Capital, UCSF Institute for Quantitative Bioscience, San Francisco CA, 2012–2013
- Attended entrepreneurial pitches and completed due diligence on potential investments
- Conducted market research and produced reports describing market segments and competitors
- Interviewed industry experts and potential customers

Business Development Fellow, UC Davis Graduate School of Management, Davis CA 2010–2011
- Completed MBA courses related to business development and technology ventures
- Developed and pitched two business plans to angel investors

PhD Researcher, UC Davis MIND Institute and Neuroscience Graduate Group, Davis CA, 2007–2013
- Programmed battery of child-friendly computerized behavior tests (disguised as games)
- Tested more than 100 children on cognitive psychology behavior paradigms
- Performed more than 100 MRI scans on children aged 7-14
- Conducted linear and nonlinear statistical analyses, including classification and machine learning
- Programmed data analyses and visualizations using Matlab and R
- Presented research at 10 invited talks and 20 poster presentations nationwide
- Published four peer-reviewed manuscripts in leading neuroscience journals

Education

PhD., University of California at Davis, Davis CA, 2007–2013
Major: Cognitive Neuroscience

Honors B.A. (*Cum Laude*), Hamilton College, Clinton NY, 2001–2005
Major: Neuroscience, Minor: Dance

Skills

Programming: Python, SQL, R, MATLAB, Javascript, Flask
Tools: machine learning, statistics, NLTK, NumPy, SciPy, pandas, Matplotlib
Software: FSL, SPM, FreeSurfer, Photoshop, Illustrator

Leadership

Mentored 3 PhD students and 20 undergraduates in brain imaging methods and analysis (2007-2013)
Managed research interns for the Cognitive Analysis and Brain Imaging Lab, UC Davis (2007-2013)
Served on the Admissions Committee for UC Davis Graduate Program in Neuroscience (2011-2012)
Organized the UC Davis student-run seminar series in neuroscience (2009-2011), neuroscience recruitment weekend (2009), and neuroscience retreat (2008-2009)
President of Hamilton College Chapter of Habitat for Humanity (2002-2004)

Personal

Dual U.S. and E.U. citizenship
Proficient in written and spoken French
Outdoor Enthusiast; Completed five triathlons, including “Escape from Alcatraz” (2011)

IGOR BESPAMYATNOV

(512) 744-8497 linkedin.com/in/bespam
igorbest@gmail.com Palo Alto, CA github.com/bespam

SKILLS SUMMARY

- **Languages:** Python, IDL, Matlab, HTML/JS/CSS, bash; some experience with Java, Ruby
- **Tools**: Pandas, NumPy, SciPy, Scikit-learn, NLTK, IPython, Flask, BeautifulSoup, Gephi, NetworkX
- **Analysis Toolkits and DBs**: RapidMiner, Orange, MySQL, MongoDB, CouchDB, Neo4j
- **Visualization**: D3.js, DataTables.js, HighCharts.js, Tableau, Matplotlib, bootstrap.js, sigma.js
- **Hadoop**: MapReduce, AWS, Cloudera, S3, Heroku, Pig, pydoop; some experience with Hive, Storm
- **Other skills**: Data acquisition, machine learning, spectral analysis, numerical simulations

EXPERIENCE

Insight Data Science Palo Alto, CA
Fellow June 2014 – present

- Designed and built a graph-based recommender (news-domains.net) for news-related websites.
- Scraped Alexa.com and listofnewspapers.com for list of news related domains.
- Processed 2TB Common Crawl page-to-page dataset (3.5B pages, 128B links) into aggregated news domain-to-domain hypergraph (6K domains, 40K links) using AWS EMR and Apache Pig.
- Processed graph with Gephi, build recommender engine with Python and stored data into MySQL.
- Deployed web app on AWS using Flask, JQuery, Bootstrap, Sigma.js and MySQL.

Data Science Web Projects 2013 – present

- **Stock Market Recommender:** Content-based recommender (www.findstock.us)
- **US opinion of other countries:** Interactive dashboard for US tweet sentiments (www.usopinion.info)
- **US Allergy Map:** US allergy map and history (http://bl.ocks.org/bespam/5691344)
- **Book Graph:** Graph representation of a given book (www.bookgraph.net)
- **Concept Riddle**: Association search within ConceptNet5 graph (conceptapp.herokuapp.com)
- **Fuzzy-SQL:** ConceptNet5 hypergraph enhanced SQL for fuzzy queries (fuzzysql.heroku.com)
- **Tableau Visualization**: Airplane collisions dataset (public.tableausoftware.com/profile/bespam)

Massachusetts Institute of Technology Cambridge, MA
Visiting Scientist Sep 2009 – May 2014

- Validated the collisional-radiative numerical model for ITER project.
- Organized and participated in national and international collaborations (US, Germany, China).

The University of Texas at Austin Austin, TX
Postdoctoral Fellow Sep 2008 – Nov 2012

- Implemented the CXRS diagnostic system for real-time data acquisition and data analysis.
- Developed a 3D neutral beam simulation package ALCBEAM used internationally.

EDUCATION

The University of Texas at Austin, PhD in Plasma Physics Austin, TX, 2004 - 2008
Novosibirsk State University, BS, MS in Physics Novosibirsk, Russia, 1999-2001

Jason A. French

Palo Alto, CA 94306
Mobile: (312) 970-9261

Github: github.com/frenchja
Email: frenchja@gmail.com
Homepage: jason-french.com

Employment

Fellow June, 2014 - Present
Insight Data Science Palo Alto, CA

- Designed and built http://www.sunnysideup.solar, a web app to calculate hypothetical solar output and recommend compatible solar energy systems using Python, Flask, multiple regression, and computer vision.
- Collected utility data from Open Energy Information and solar data from the National Renewable Energy Laboratory.
- Developed algorithm using Scikit-image to identify roofs using Google Maps API and calculate the sq. footage.
- Outlined frontend using Twitter Bootstrap 3 & JQuery.

Ph.D. Candidate September, 2009 - Present
Northwestern University Evanston, IL
https://sapa-project.org

- Co-developed https://sapa-project.org using PHP, R, and MySQL. SAPA collects data from 40,000+ participants/year, which are then analyzed using R, visualized, and published.
- Presented and published multiple projects assessing individual differences in personality traits, cognitive abilities, and vocational interests and their association with vocational and educational outcomes using SAPA data.
- Developed open-source surveys and items for assessing temperament and scientific interests using structural equation modeling and item response theory.
- Published analytical and methodological R tutorials for researchers on sapa-project.org/blog.

Education

Ph.D. Cognitive Psychology, Northwestern University, 2014 (Expected September).

Adviser: William Revelle, Ph.D.

Dissertation: The Relation of Scientific Attitudes to Traits, Abilities, and Interests over Time.

M.S. Cognitive Psychology, Northwestern University, 2011.

B.A. Psychology and Spanish, University of Michigan, Ann Arbor, 2007.

Released Software

French, J. A. & Condon, D. M. (2013). SAPATools: Software for analyzing the Synthetic Aperture for Personality Assessment. R package version 2013.12.26. URL https://github.com/frenchja/SAPATools

French, J. A., Condon, D. M., & Revelle, W. (2012). ICAR: International Cognitive Ability Resource. R package version 0.1. URL http://www.icar-project.org

Skills

Programming:

Languages: R (Very Proficient), LaTeX, Python, & PHP.

Statistical Analysis: R-Project, SPSS, SAS, SVM Machine Learning, Structural Equation Modeling, Item-Response Theory, Hierarchical Linear Modeling, Survey Methodology.

Software Frameworks: E-Prime, DirectRT, PsychoPY, SciPY, NumPY, ggplot2.

Databases: Filemaker Pro, MySQL, PostgreSQL.

Web Frameworks: PHP, NodeJS, Flask, Django.

Language:

Conversant in Spanish.

Jesus Martinez-Manso

Mountain View, CA *www.linkedin.com/in/jmartinezmanso*
j.martinez.manso@gmail.com *github.com/jmmanso*
(352) 256 3988

SKILLS

Languages: Python, Mathematica,SQL, Bash, past experience with C++
Tools: Numpy/Scipy, PyCUDA, Pandas, Scikit, MySQL, Flask, Bootstrap, Git

EXPERIENCE

Fellow - Insight Data Science, Palo Alto, CA (*June 2014 - present)*

- Built edXPredictor.com, a web application that assesses the performance of students in the online course platform edX and makes personalized course recommendations.
- Created a MySQL database with historical student information and applied a quadratic discriminant model to predict user academic success in a particular online course.
- Determined prediction accuracy from cross-validation, reaching 90% in 4 courses.
- Designed an interactive front-end with Flask, Bootstrap and hosted it on AWS.

Graduate Researcher - University of Florida, Gainesville, FL (*2009 - 2014)*

- Described with state-of-the-art accuracy the relation between the masses of galaxies and their host dark matter halos in the young Universe.
- Found the solution to the problem of the ultra-dense quiescent galaxies, finding that these were less massive than previously thought.
- Developed Python code to compute 2-point clustering statistics of a large number of galaxies using the GPU device, achieving speed gains by ~1000x.
- Detected and characterized galaxies from infrared imagery. Quantified levels of noise, incompleteness and false-positives by performing photometric simulations on noise-dominated data.
- Built generic data analysis tools, including Monte Carlo samplers, maximum-likelihood estimators, numerical integrators and matrix manipulation methods.
- Instructed undergraduates in 4 astrophysics courses.

Web designer - GEANSOLAR, Chiclana, Spain *(2007-2008)*

- Designed, built and maintained the website of a local renewable-energy company, using HTML, PHP and Javascript.

EDUCATION

- PhD in Astronomy, University of Florida (*2009-2014)*
- B.S. Theoretical Physics, University of Sevilla, Spain (*2003-2009)*

AWARDS

- University of Florida Doctoral Fellowship (*2009-2013)*
- Academic Achievement Scholarship, *(2004-2007)*
  Spanish Ministry of Education

Kelty Allen

· kelty.allen@gmail.com · (510) 545 3582 · Santa Clara, CA · linkedin.com/in/keltyallen · github.com/KeltyAllen ·

Skills

- **Languages** Python, SQL, some experience with MATLAB, C++, Java
- **Tools** NumPy, SciPy, MatPlotLib, Scikit Learn, LaTeX
- **Other Skills** Machine Learning (Classification, Clustering etc), Numerical Analysis, Linear Algebra

Experience

Fellow at Insight Data Science Palo Alto, CA — June 2014 - present

- Developed AmazonExplorer.info, a web-based tool to provide sellers and consumers of Amazon products with analysis of product quality and popularity based on review history, using data collected by SNAP.
- Analyzed product review history using natural language processing to determine the most representative reviews from a time period selected algorithmically as most interesting to sellers or as input by user.
- Collaborated with Insight Data Engineering Fellow Isura Edirisinghe in optimizing database design for analysis of 35GB dataset.
- Designed an interactive front end using Flask and Twitter Bootstrap. Hosted the website on AWS.

Side Project in Earthquake Aftershock Cluster Analysis — March 2014

- Discovered Earthquake aftershock patterns using cluster analysis (DBSCAN, K-Means). Analyzed USGS data using Python and Scikit Learn. Verified by comparing to Omori-Utsu and Gutenberg-Richter laws.

PhD Research at UC Berkeley Berkeley, CA — September 2007 - May 2014

- Applied techniques from the study of algorithmic randomness in computability theory to the study of Brownian motion in probability theory to prove new results that bridged both fields.
- Used techniques and results from mathematical logic and probability theory to provide a new proof that the solution to the Dirichlet problem (an important problem in physics) is computable .
- Invited to speak at the Joint Mathematics Meeting, AMS Sectional Meeting, and seminars at UC Berkeley, UCLA, Penn State, and University of Paris VII.
- Graduate Student Instructor for undergraduate mathematics courses in numerical analysis, multivariable calculus, single-variable calculus and precalculus.
- Relevant Coursework: Computability Theory, Quantum Computation, Probability Theory, Efficient Algorithms and Intractable Problems.

Research Assistant, Oberlin College Physics Oberlin, OH

Quantum Dynamics of Adsorbed Hydrogen — June-December 2006

- Used Diffuse Reflection Infrared Spectroscopy to investigate the quantum dynamics of hydrogen adsorption in novel metal-organic frameworks.
- Wrote and maintained MATLAB code used in quantum behavior modeling.

Astronomy - Study of Interstellar Medium Using Pulsars — November 2004 - April 2005

- Simulated scintillation of radio waves from a compact source through the interstellar medium using Fortran.
- Observed Interstellar Medium by scintillation of radio waves from pulsars; observations transmitted from Arecibo, Puerto Rico.

Physics REU Internship at AFIT Dayton, OH — Summer 2005

- Wrote HELEEOS Lite, a faster, Java-based version of HELEEOS - High-Energy Laser End to End Operational Simulation - an existing MATLAB-based laser simulation used by the Air Force.

Math REU Internship at Clemson University Clemson, SC — Summer 2004

- Developed and implemented computer algorithms to improve known lower bounds for Ramsey Numbers using C++.
- Developed and implemented computer algorithms to explore existence of four-saturated graphs and three-saturated graphs of degree two using C++.

Education

PhD Logic and Methodology of Science, UC Berkeley — May 2014
BA Physics and Mathematics with Honors, Oberlin College — May 2007

Honors and Awards

University of California Dissertation Year Fellowship — 2013-2014
John Stern Merit Scholarship, Natural Sciences, Oberlin College — 2003-2007
National Merit Scholar — 2003-2007

Kevin J Ford

CONTACT kjford@gmail.com
858 945-1523
LinkedIn: http://linkd.in/1n2neOu
Github: https://github.com/kjford

SKILLS

Languages: Python, MATLAB, SQL, BASH scripting, Javascript (some experience)
Tools: scikit-learn, pandas, numpy, scipy, pylab, BeautifulSoup, networkx, Gephi, d3

EXPERIENCE

Fellow - *Insight Data Science* (Palo Alto, CA) 2014-Present

- Developed Tapmapper.us, a social network analysis of regional beer preferences.
- Scraped beer-related information from BeerAdvocate.com using BeautifulSoup and mined continuous stream of Twitter data for beer related references. Constructed a MySQL database to store and aggregate data.
- Built a cleaning and normalization pipeline for data to group tweets by beer and region using geolocation.
- Designed an interactive user interface using Flask and Bootstrap, hosted on AWS.

Independent Projects 2013-Present

- Competed in 4 Kaggle competitions including a top 25% finish. Developed exploratory analysis and prediction tools in Python using scikit-learn Random Forrest, Gradient Boosting, Support Vector Machines, and home-built tools for neural networks.
- Created statistical analysis tools for genomics data using Python.
- Developed flyGeneRefNet, a network analysis of ~20,000 *Drosophila* genes based on references in ~100,000 research publications. Built Python based tools to query Postgres database, construct interaction network, cluster genes using Spectral Clustering, and visualize network using Gephi.

Postdoctoral Fellow - *UC San Francisco Biopysics* (San Francisco, CA) 2012-Present

- Designed and carried out a large-scale electrophysiological screen of ~300 genes that allow the nervous system to adapt to changes. Developed MATLAB code to analyze electrophysiologyical output, visualize potential targets, and perform statistical analysis to detect candidate genes.
- Pioneered novel fluorescence imaging technique to visualize neuronal voltage signals. Developed MATLAB code to extract and analyze data.

Graduate Researcher - *UC Berkeley Neurobiology* (Berkeley, CA) 2005-2011

- Developed MATLAB code base to organize, analyze, and visualize imaging and electrophysiological data from neuroscience experiments and numerical simulations.
- Worked collaboratively to develop code to analyze movies of neuronal activity.
- Authored 6 publications in high profile journals. Awarded 2 competitive fellowships.
- Presented research at neuroscience and applied mathematics meetings. Taught undergraduate neuroscience and biochemistry courses.

Assistant/Sys Admin - *Moores DNA Sequencing Core* (La Jolla, CA) 2003-2005

- Administered staff computers and servers for DNA sequencing facility.
- Created computer-programming solutions to fit client and staff needs using web development, Javascript, and MySQL/PHP.
- Designed and implemented an online customer satisfaction survey database used to enhance customer satisfaction.

EDUCATION

University of California, Berkeley (Berkeley, CA) December, 2011
Ph.D. in Molecular and Cell Biology

University of California, San Diego (La Jolla, CA) June, 2005
B.S. in Biology (*Summa Cum Laude*)

Matt George

510-778-3674 | mattgeorge1@gmail.com | Berkeley, CA
github.com/mrgeorge | linkedin.com/in/mattgeorge0

Skills

- Python
- SQL
- IDL
- C (some)
- Data analysis
- Bayesian modeling
- Markov Chain Monte Carlo
- Project management
- Machine learning
- Classification, clustering
- Experiment design
- Robotics enthusiast
- UNIX, OS X
- bash, awk
- git, svn
- Flask

Experience

Fellow, Insight Data Science *(Palo Alto, CA)* *June 2014 - present*

- Built StreetsAhead, a tool to identify text and objects in Google Street View images using a computer vision and crowdsourcing API (*streetsaheadmaps.com*).
- Explored deep neural networks implemented in C++ and Python for feature learning in images, trained on a GPU with 75,000 labeled images from machine learning repositories.
- Sped up queries by caching results in MySQL database.
- Designed web front-end with Flask, jQuery, and Bootstrap, and deployed to AWS.

Graduate Student Researcher, UC Berkeley *(Berkeley, CA)* *Aug. 2008 - present*

- Invented new techniques to hunt for dark matter using galaxy shapes and motions extracted from large imaging and spectroscopy surveys.
- Developed Bayesian models of covariance between properties of 100,000 galaxies for cosmological inference with Python (Numpy, Scipy, Pandas), IDL and C.
- Presented novel visualizations using Matplotlib and Scikit-Image to large audiences at international conferences and seminars.
- Successfully proposed for observing time at world's largest optical telescopes.
- Constructed a web bookmarking tool in PHP and JavaScript for journal discussion groups that improved user participation with seamless voting for papers on arxiv.org.
- Taught and developed curriculum for technical and non-technical astronomy classes.
- Mentored 30 third-graders, 2 undergraduates, and 3 graduate students in science outreach.

Side Project: BusTripper (*github.com/mrgeorge/BusTripper*) *Feb. - May 2014*

- Applied machine learning algorithms to GIS transit data to predict bus arrival times and prevent bunching, using Scikit-Learn and Pandas, with Berkeley startup VIA Analytics.
- Increased accuracy, speed, and scalability of predictions over existing heuristic approach.

Research Assistant, Cambridge University *(Cambridge, UK)* *Sep. 2007 - Aug. 2008*

- Detected faint signal from X-ray satellite data delivering first-of-its-kind measurement of hot gas surrounding galaxy clusters using astronomical image processing tools and IDL.

Education

UC Berkeley	Ph.D. in Astrophysics	*anticipated Aug. 2014*
Cambridge University	M.Phil. in Physics	*Aug. 2008*
Harvard University	B.A. in Physics & Astronomy, *magna cum laude*	*June 2007*

Awards and Recognition

National Science Foundation Graduate Research Fellowship	*2009 - 2012*
Herschel Smith Scholarship from Harvard for research at Cambridge University	*2007 - 2008*
Goldberg Prize for best thesis from Harvard Astronomy Department	*June 2007*
Authored over 20 peer-reviewed publications (5 as lead) with 900+ citations	
Research highlighted in popular media: MSNBC, Physics Today, NASA website	

Meredith Trotter

http://people.stanford.edu/mtrotter mer.trotter@gmail.com
https://github.com/mtrotter/ San Jose, CA
http://www.kick-or-gogo.com 520 223 7747

DATA SCIENCE EXPERIENCE

Fellow 2014 - Present
Insight Data Science *Palo Alto, CA*

- Created Kick-or-gogo.com a crowdfunding site recommender app that predicts the probability of success of a planned campaign on the Kickstarter and Indiegogo platforms.
- Utilized Python with Selenium and BeautifulSoup to scrape a training set of >25,000 closed campaigns from each site.
- Applied random forest classification and regression models to build predictors of funding success for both sites using Python Scikit-learn with data stored in MySQL
- Deployed as interactive web application using Flask, hosted on AWS.

Postdoctoral Researcher, Population Studies 2010 - Present
Stanford University *Stanford, CA*

Created novel analytic and modeling tools to derive insights from large datasets of genetic, ecological and demographic data, including:

- Derived new sensitivity metrics for population growth and extinction risk in changing environments that are now in use by ecologists and demographers (MATLAB).
- Developed analysis methods for whole-lifecycle time series microarray analysis that identified new gene sets correlated with aging and other life transitions (MATLAB).
- Analyzed historical data from Human Mortality Database to show how increasing lifespan inequality in the US is driven by mortality increases in the young (R).

Built graduate-level online course: Essential Mathematics for Research in Life and Social Sciences

Postdoctoral Researcher, Evolutionary Theory 2009 - 2010
University of Arizona *Tucson, AZ*

- Demonstrated how failures in genetic robustness can actually help evolution produce complex adaptations, using stochastic simulation models (C++, MATLAB).

Graduate Researcher 2004 - 2008
University of Otago *Dunedin, New Zealand*

- Built stochastic mutation-selection models that illustrated how natural selection under pairwise contests can maintain high levels of genetic variation (C++, MATLAB).
- Employed as Zoology Department Statistics Consultant, providing advice on experimental design and statistical analysis to graduate students and staff.

EDUCATION

PhD in Theoretical Population Genetics Sept 2004 - 2008
University of Otago *Dunedin, NZ*

BSc. Hons in Marine Biology, minor in Math 2000 - 2004
Dalhousie University *Halifax, NS*

LANGUAGES: Python, MATLAB, SQL, C++ (prior experience), R (prior experience)
TOOLS: BeautifulSoup, Flask, MongoDB, MySQL, NumPy, Pandas, Scikit-Learn, Scrapy

Michael Ramm

micramm@gmail.com | www.michael-ramm.com | Redwood Shores, CA | 646-498-9995
@Michael_Ramm | www.linkedin.com/in/mramm | www.github.com/micramm

EXPERIENCE

Insight Data Science – Fellow Palo Alto, CA 6/2014 – Current

- Developed Hopper, www.hoppr.me, a Python-based web app to optimize route planning given multiple destinations and business rating preferences.
- Utilized Yelp API and Foursquare API to obtain 60,000 business listings in the Bay Area, and MapQuest API for geocoding and distance information.
- Cleaned and visualized data with Pandas, stored records in MySQL database.
- Implemented a traveling salesman problem solver in Python using itertools.
- Employed a logistic regression classifier to recommend additional locations by using Yelp features with Foursquare labels.
- Deployed web app using AWS, RDS, Flask, Javascript, Bootstrap, and Google Maps API.

UC Berkeley – Graduate Student Researcher Berkeley, CA 3/2009 – 5/2014

- Designed and constructed an ultra-high vacuum chamber, trap, and electronics for confining ions.
- Programmed experimental control software in Python for complex ion trapping experiments using PyQt and asynchronous framework Twisted.
- Performed numerical Monte Carlo simulations of interacting quantum systems with Cython.
- Analyzed collected data in NumPy and Scipy to determine the quantum state of ion motion by fitting to models of atom-laser interaction.
- Published results in peer-reviewed journals including Nature Physics and Physical Review Letters.
- Recipient of Department of Energy Science Graduate School Fellowship.

General Atomics – National Undergraduate Fellow San Diego, CA 6/2007 – 9/2007

- Developed an efficient numerical integration technique in Matlab to study magnetic plasma confinement.
- Analyzed data from DIII-D tokamak to improve suppression of plasma instabilities.
- Earned award at American Physical Society conference for best undergraduate poster.

EDUCATION

UC Berkeley Berkeley, CA 9/2008 – 5/2014

- Ph.D. in Physics.
  Thesis title: Energy Transport and Quantum Correlations in Trapped Ions.

Stanford University Stanford, CA 9/2004 – 6/2008

- B.S. in Physics and Mathematics.
  Additional coursework in Computer Science and Economics.

SKILLS

Languages: Python, SQL, Mathematica, Matlab, Java (prior experience), C++ (prior experience).
Tools: SciPy, NumPy, Pandas, Matplotlib, Twisted, PyQt, Scikit-learn, Flask, Git, AWS, Linux, LaTeX.
Other skills: Analog and digital electronics, mechanical design, vacuum systems, photolithography.

Michael A. Woods

925.321.3560
physicsmichael@gmail.com
Daly City, CA

github.com/vgm64
linkedin.com/in/michaelwoodsphd
mike.strollthroughthewoods.com

Skills

Languages: Python, C++/ROOT, MySQL, HTML, PHP, Matlab (prior experience)

Tools: NumPy, SciPy, pandas, scikit-learn, BeautifulSoup, git, Flask, AWS

Miscellaneous: Rocks cluster computing, SGE, sysadmin, NFS, Puppet, Raspberry Pi, Arduino, HDF5

Experience

- **Insight Data Science** — Palo Alto, CA
  Fellow — *June 2014 – Present*
  - Designed and built melodymapr.com, a web app that plans a radio station playlist for road trips.
  - Joined FCC radio antenna engineering data with scraped program information from multiple sources.
  - Interfaced GoogleMaps API to MySQL to overlay available radio stations by genre for a given route.
  - Deployed stack on AWS using Flask/Bootstrap/JavaScript.

- **Python software development: python-fit** — Davis, CA
  2012 – Present
  - Authored the python-fit module, available on PyPI.
  - Focused on fast and effortless curve fitting in Python for rapid data exploration.
  - Wrapped powerful ODR techniques from SciPy in an easy to use interface.

- **Large Underground Xenon (LUX) Dark Matter Collaboration** — Davis, CA — Lead, SD
  Graduate Student Researcher — *2008 – 2014*
  - Coordinated cross-country analysis pipelines, data processing, and cluster computing for a custom Python/C++/Matlab/MySQL analysis framework.
  - Developed machine learning solutions for position reconstruction and event characterization in Python and C++.
  - Invented a technique to include variable size input vectors in machine learning algorithms.
  - Managed storage and processing of 120 TB data, remotely and locally.
  - Deployed two computing clusters for real-time processing, and projected expansion needs.

Education

- **University of California** — Davis, CA
  Ph.D. in Physics — *March 2014*

- **University of California** — Davis, CA
  B.S. in Physics — *June 2007*

Awards

- **2011 Summer Graduate Student Researcher Award** — Davis, CA
  June 2011
  - Awarded to support promising graduate research in engineering-related applications and methods.

- **2009 UC Davis Graduate Block Grant Fellowship** — Davis, CA
  March 2009
  - Recognition of outstanding research.

Monica Tremont Hsu

monica.tremont@gmail.com • 314.341.8994 • Burlingame, CA • github.com/MonicaHsu • linkedin.com/in/monicatremonthsu

EXPERIENCE

Insight Data Science—Fellow
Palo Alto, CA *June 2014 – present*

- Developed TruValuation.com, a web app that predicts real estate prices in San Francisco using data collected from Redfin.com, WalkScore.com and data.sfgov.org.
- Scraped real estate and neighborhood information using Selenium. Merged data sources using geolocation tags. Cleaned and parsed relevant data using Python. Performed random forest regression using Scikit Learn.
- Predicted home prices for a blind test set with a median error of 1.5%, an improvement over predictions from list prices (12%) and Zillow.com's Zestimates (6.9%).
- Produced a web app using mySQL, AWS, Flask, Javascript, JQuery and Google Maps API.

University of San Francisco—Adjunct Faculty
San Francisco, CA *September 2012 – January 2014*

- Expanded a biological cell model to accurately predict protein diffusion. Contributed to the model using Fortran and analyzed the results using Matlab.
- Taught ~100 students per semester in Physiology, Biology and Kinesiology.
- Created a new MCAT prep course for undergraduate students.

UCSF—Graduate Student Researcher
San Francisco, CA *August 2007 – August 2013*

- Constructed a model for microtubule assembly kinetics using Fortran and Matlab. Validated the model with published experimental data.
- Scraped an online database of published protein structures and developed an algorithm for simplifying the structures to create a model for cell transport.
- Instructed graduate level students in Statistical Mechanics and Cell Dynamics.

William and Mary—Undergraduate Student Researcher
Williamsburg, VA *August 2005 – May 2007*

- Created and modeled systems of differential equations in Matlab to represent food webs with complex regulatory environmental feedbacks.

EDUCATION

University of California, San Francisco *San Francisco, CA*
Doctorate of Philosophy in Biophysics *August 2013*

College of William and Mary *Williamsburg, VA*
Bachelor of Science in Biology and in Chemistry *May 2007*

SKILLS

- **Languages**: Python, SQL, MATLAB, html, Fortran (prior experience)
- **Tools**: Pandas, SkLearn, Matplotlib, NumPy, SciPy, git, mySQL, Flask, AWS

NICHOLAS R. PETERSON

Sunnyvale, CA • 614-477-8412 • nrpeterson@gmail.com
nrpeterson.info • linkedin.com/in/nrpeterson • github.com/nrpeterson

Relevant Experience

Fellow at Insight Data Science, Palo Alto, CA — June 2014-Present

- Developed Math.StackExpert (http://math.stackexpert.net), a web app to help expert users of Math.StackExchange find high-quality, interesting questions to answer.
- Fetched questions from the StackExchange API using Python; stored them in a MySQL database.
- Performed text analysis on questions in scikit-learn to train predictors of question quality, and ran a clustering algorithm on 1000+ content tags to reduce to 28 broad content categories.
- Built a front-end using Flask, D3, and Bootstrap, and deployed it to AWS.

The Ohio State University, Columbus, OH

Graduate Mathematics Researcher — 2007-2013

- Explored a new random "k-out" digraph process which exhibits a rich-get-richer scheme.
- Applied careful approximations/asymptotics and probabilistic methods to find the likely structure of the digraph given a large vertex set, and to compare/contrast to a uniform model.
- Performed numerical simulations/experiments in Python and Mathematica to guide research.

Lecturer of Mathematics — 2013-2014

- Lectured precalc (150 students), multivariable calc (x4, 60 each), and combinatorics (10 students).

Web App Developer for Michigan Tech Mathematical Sciences, Houghton, MI — 2005-2007

- Developed web apps (PHP/MySQL), including: a dept. directory which users could update, a scheduling/time clock app, and an admin system for Mathematica labs using barcodes.

Relevant Skills

Languages:	Python, Mathematica, SQL, Javascript. Previous experience with PHP, C/C++, Java, Octave, and R.
Tools:	Numpy, Scipy, Scikit-Learn, Flask, BeautifulSoup, D3, jQuery, Bootstrap.
Other Skills:	Graph/Algorithm Analysis, Applied Probability, Asymptotics.
Coursera:	Machine Learning (Andrew Ng, Stanford University), March-June 2014.

Education

Ph.D. in Mathematics, The Ohio State University, Columbus, OH — August 2013

B.S. in Mathematics, Michigan Technological University, Houghton, MI — May 2007

Selected Honors

- Runner-up, *Phil Huneke Excellence in Teaching Award*, Ohio State Mathematics. — 2010
- *Math Achievement Award* and named *Department Scholar*, Michigan Tech Math. Sciences. — 2007

Nicolas Tilmans

(240) 441-3429
www.linkedin.com/ntilmans ntilmans@gmail.com www.github.com/entee

EDUCATION

Ph.D. Biochemistry	Stanford University	2006- *2013*
B.S. Biochemistry and B.S. Computer Science	University of Maryland (UMCP)	*2001-2006*

PROFESSIONAL EXPERIENCE

Data Science Fellow *Insight Data Science* *June 2014-present*

- Built http://www.drugscope.info/ to inform drug treatment choice.
- Constructed MySQL database using OpenFDA API and PharmGKB Data.
- Developed algorithms to recognize unique drugs having different names.
- Identified feature vectors for comparing drugs and constructed distance matrices based on indication and side-effect profiles using Python.
- Identified drug interactions that might worsen side effects.
- Deployed to amazon web services.

Ph.D. Student, Postdoctoral Scholar *Harbury Lab, Stanford Biochem. Dept.* *2006-April 2014*

- Developed new technology for discovery of pharmaceutically relevant molecules.
- Developed 1,000x cheaper synthesis for cutting-edge phosphate binding molecule.
- Performed directed evolution of 1-billion member small molecule library for kinase substrates+binders, determined results by Illumina MiSeq.
- Wrote MATLAB scripts to validate proof-of-concept of technology and optimize future experimental parameters.
- Identified lead compounds by using hierarchical clustering and customized chemical similarity metrics in Python.
- Isolated and confirmed novel biologically relevant molecules.
- Published results in top journals, one paper recommended by Faculty of 1000: http://f1000.com/prime/3754956

Business Development Lead *Stanford Nitrogen Group* *2011-2012*

- Led team to create a business plan for a novel wastewater treatment technology.
- Co-led final team, presented concept at national business plan competitions, and pitched project to potential investors.
- Won "Shark Tank" Round at Rice Business Plan Competition.
- 60 second pitch featured on CNN Money: http://bit.ly/HH72Ab, http://bit.ly/18lwrwk
- Won 2012 DOE regional finals at Caltech's FLOW competition, $100,000 award.

Undergraduate Fellow *Howard Hughes Medical Institute, UMCP* *2003-2006*

- Established a classification system for low complexity protein domains
- Created rapid comparison algorithm specially adapted for low complexity domains.
- Generated characteristic fingerprints for individual domain types using algorithm.
- Built PERL and C programs to evaluate how low complexity filtering algorithms influenced real-world biologically relevant searches. Determined that algorithm choice would alter 1 in 3 in protein sequence searches.
- Presented results at June 2005 Intelligent Systems for Molecular Biology conference.

Team Leader, Engineer *CMSC435 Software Engineering, UMCP* *Fall 2005*

- Led a team of 4 computer science undergraduates to produce a PDA app for rapid database access and deployment for the Montgomery County Fire Department.
- Developed C# Windows Mobile and XP front end apps to MS SQL database.
- Delivered all products ahead of schedule despite major last minute customer requests.

SKILLS

- **Programming Languages:** Python, MySQL, prior experience: Java, C/C++, PERL, C#
- **Other:** Native speaker of French and English, and fluent in Spanish.

NICOLE H. ROMANO

(650) 862-6069 • NRomano@stanford.edu
NicoleHRomano.com • Github: NickiRom • LinkedIn: /in/nhromano

EDUCATION

Stanford University May 2014
Ph.D. Materials Science & Engineering
NSF Graduate Research Fellowship
Stanford Graduate Fellowship
US Delegate to Lindau Meeting of Nobel Laureates (2010)

Johns Hopkins University May 2008
B.S. Materials Science & Engineering, Biomaterials Track
Tau Beta Pi Engineering Honor Society
Senior Design Engineering Award
Vredenburg Fellowship (for biomedical research in Singapore)

EXPERIENCE

Insight Data Science, Fellow June 2014 – present
- Developed {SetList], a playlist curator focused on minimizing harsh song transitions and optimizing acoustic continuity. www.SetListengine.com Palo Alto, CA
- Orchestrated data from Spotify, EchoNest, and Beats Music APIs in MySQL to identify the key metrics that influence perceived continuity, including tempo, energy build, and mood profile.
- Formulated a 5-dimensional scoring algorithm in Python to rank transitions between song pairs.
- Configured optimized playlists using machine learning: depth-first search of weighted, directed graphs.
- Reported optimized playlist in a visually intuitive manner: heat maps representing the quality of song transitions in the configured playlist and the control, a randomly shuffled playlist.
- Built an AWS-deployed front end using Flask, HTML, and Javascript.

Stanford University, Materials Science & Engineering 2008 – 2014
- Engineered protein-based hydrogels to stimulate spinal cord regeneration Stanford, CA
- Established a predictive model of cranial defect regeneration following surgical intervention, using sparse data from a highly combinatorial matrix of therapeutic interventions. Used data transformation, categorization, and validation in R to identify impactful predictors in the logistic regression model.
- Developed analytic tools in MATLAB to quantify neural regeneration in terms of multi-feature co-localization, pattern regularity, n-connected components, and persistence lengths. Implemented pseudo-replication statistics to account for within-sample correlation and Kolmogorov-Smirnov tests to differentiate populations.
- Collaborated as a statistical consultant on various projects, including the dose-response relationship between drugs and cellular migration and a 3-way contingency analysis to parse the effect of micro-environmental factors on contractile behavior of intestinal tissues.
- Prepared and delivered an invited lecture on best practices in statistical analysis, including distribution analysis, predictive modelling techniques, and validation.
- Courses: Intermediate Biostatistics (SAS), Data Analysis in R, Statistical Learning (audited)

COMPUTATIONAL TOOLS

Languages R, MATLAB, Python, SQL, HTML, course-based learning: SAS, Mathematica
Packages Numpy, pandas, scipy, scikit-learn, git, MySQL, jQuery, Flask
Image analysis MatLab, ImageJ, Volocity, Photoshop; 3D, time-lapse analysis, and feature extraction

Ravi Menon

(858)-736-1769 rsmenon@icloud.com rsmenon linkedin.com/in/rsmenon
Palo Alto, CA

Experience

Fellow, Insight Data Science, Palo Alto CA *Jun. 2014 – present*

- Built searchpy.net which searches Python questions and answers on Stack Overflow and recommends relevant posts which have code similar to the user's.
- Converted and stored Stack Overflow's XML data dump in a MySQL database.
- Developed a custom vocabulary and tokenizer which improved performance over standard NLTK libraries.
- Designed an interactive front end for the web-app using Flask, Bootstrap, D3.js and the Stack Exchange API, and deployed it on Amazon EC2.

Opensource/side projects *2012 – present*

- Co-developer of MATLink (http://matlink.org), a Mathematica application for inter-operability between Mathematica and MATLAB. Designed and architected the application and authored the documentation and website (15,000+ visits in a year).
- Developed an app in Python to identify a stolen bike on Craigslist. The app continously scraped Craigslist ads and performed text analysis using scikit-learn to classify the bike by model, followed by image processing in Mathematica to narrow down results by similarity to the color of the stolen bike.
- Developed a regular-expression based Mathematica syntax highlighting plugin for for Vim (in VimL).
- Contributor to Mathematica plugin for Intellij IDEA (http://wlplugin.halirutan.de). Participated in extensive discussions with the lead developer regarding the design and development of the plugin; contributed to several bug fixes and code review discussions.

Postdoctoral Scholar, University of California San Diego, La Jolla CA *Nov. 2013 – Jun. 2014*

- Developed a framework in Mathematica to clean, unify, and analyze large volumes of seismological, oceanographic, and meteorological data from multiple datasets.
- Motivated and contributed to the development of a methodology to analyze traffic patterns in Long Beach CA, using machine learning on data from dense seismic networks (in MATLAB).
- Mentored a graduate student in formulating, analyzing and executing a research project culminating in a research publication (in production).

Graduate Researcher, University of California San Diego, La Jolla CA *Sep. 2008 – Oct. 2013*

- Led a solution-oriented research effort funded by the U.S. Navy to successfully identify and mine interesting insights (such as the presence of ships, physical properties of the environment, etc.) from background noise in the ocean. Accompanying software was developed in MATLAB.
- Developed a frequency-variant model to describe the statistics of background ocean noise and successfully validated the model using data, resulting in a 10 fold increase in the signal-to-noise ratio.
- Organized and co-chaired a technical session on "Signal processing in Underwater Acoustics" at the Acoustical Society of America.

Skills

- *Languages:* Mathematica, Python, MATLAB, SQL, Java (some experience), HTML, CSS, Javascript
- *Tools:* Flask, Bootstrap, Pandas, Scikit-Learn, Scrapy, AWS, git
- *Scientific:* Signal processing, time series analysis, Kalman and particle filters, hypothesis testing

Education

- University of California San Diego, La Jolla CA — Sep. 2008 – Oct. 2013
  Ph. D. in Electrical Engineering, M. S. in Oceanography
- Indian Institute of Technology Madras, Chennai India — Jul. 2003 – Jul. 2008
  B. Tech. & M. Tech. in Naval Architecture and Ocean Engineering

RAY MENG GAO

PALO ALTO, CA CELL: 650-285-8032
EMAIL: RAYM.GAO@GMAIL.COM PORTFOLIO: HTTP://WWW.RMGAO.COM GITHUB: HTTPS://GITHUB.COM/RAYG1234

Experience

Insight Data Science, Palo Alto, California Jun. 2014 – Present
Fellow

- Built Upfindr.com to directly find activity partners from Meetup.com without attending meetups.
- Collected half a million unique user profiles from Meetup.com in the SF Bay area using the Meetup API. Cleaned and analyzed data in Python and stored in a 2.5 GB MySQL database.
- Built unique matching engine by training a neural network (Word2Vec) on Meetup data to find relationships between people, groups, and interests.
- Created interactive frontend using Flask, Bootstrap, JQuery and D3.js deployed on AWS EC2 and RDS.

University of Toronto, Toronto Sep. 2009 – Mar. 2014
Max Planck Institute for Structure Dynamics, Hamburg
Graduate Researcher

- Led an international team of 20+ scientists on a 5-year project to create the world's first video of ultrafast organic molecular motions, resulting in a first-author Nature publication.
- Developed libraries in Python to mine TBs of noisy image data and reconstruct time-dependent 3D molecular structures. Used Cython and Numpy to dramatically speed up computationally intensive code for data processing and fitting molecular models.
- Developed novel vision algorithms with the help of Scipy.ndimg and PIL to quickly filter and identify unique patterns in large images.
- Designed a series of multi-threaded servers in C++ to enable concurrent resource sharing and remote access of scientific instruments. Currently deployed in several multi-million dollar research labs.
- Used MySQL, PHP, Javascript/D3.js, Python, and C++ to build a live data streaming application to display and monitor critical experimental parameters online. Deployed on a LAMP webserver.

University of British Columbia, Vancouver Jan. 2005 – Aug. 2005
Software Developer (Intern)

- Created embedded device on Atmel platform to allow remote interfacing with lab instruments over GPIB and Serial ports. Developed firmware in AVR C/C++. Ten of these adapters were produced and are now used throughout two separate labs at the University of British Columbia. Code and schematics available on a public SourceForge repository. (http://sourceforge.net/projects/ethernut-gpib/)

Skills

Languages Current: Python, MySQL, Javascript, HTML/CSS, Matlab **Proir:** C/C++, Java, PHP
Tools: LAMP stack, Ubuntu, Git, SVN, AWS, D3.js, Scikit-Learn, Android SDK (some experience)

Education

University of Toronto, Toronto, Canada Sep. 2008 – Mar. 2014
M.Sc. and Ph.D. Physics

University of British Columbia, Vancouver, Canada Sep. 2003 – Sep. 2008
B.Sc. Honors Computer Science, B.Sc. Honors Physics

Major Publication

Meng Gao *et al.* "Mapping molecular motions leading to charge delocalization with ultrabright electrons," **Nature** 496, pp 343-346, 2013 (Covered on front page of CBC News Science and Tech.)

Sara Salha

310 500 7384
Palo Alto, CA craigsublets.com
ssalha@gmail.com https://github.com/ssalha/

Experience

- **Insight Data Science** Palo Alto, CA
  Fellow *Jun. 2014 – present*
  - Built *craigslist++*, a web app that optimizes the search for room sublets by condensing craigslist posts into insightful graphs, and predicts the likelihood of finding a given housing option based on craigslist historical data.
  - Assembled a MySQL database to store postings from a Python web scraper and classified the information into various categories using regular expression with a naive Bayes assumption on keywords counts.
  - Deployed the web app using AWS, Flask and OpenStreetMap for front end.
- **Coherent Diffraction Imaging, University of California** Los Angeles, CA
  Graduate student researcher *2008-2014*
  - Developed a Bayesian image reconstruction algorithm and implemented novel constraints which improved on the resolution obtained in the case of sparse measurements.
  - Built an optical diffraction microscope and successfully verified a proposed imaging methodology that uses fewer measurements than conventional methods.
  - Developed codes in Matlab, C++, Python to analyze a large amount of reconstruction datasets and used Bayesian methods and machine learning techniques to extract additional information from the noisy output.
- **Lawrence Berkeley National Lab** Berkeley, CA
  Intern *Summer 2011*
  - Collaborated on the development of a phase retrieval algorithm: Saddle point optimization & Sparse Sampling.
- **Infrared Lab, University of California** Los Angeles, CA
  Graduate student researcher *2006-2007*
  - Conducted systematic analysis on the resolution limits of a new spectrometer (MOSFIRE), currently installed at the Keck observatory, using ray tracing simulations with IDL scripting.
- **Physical Optics Cooperation** Torrance, CA
  Intern *Jan - Sept 2005*
  - Analysed deep surface quality of airplane wings using interferometric probes.
- **California Institute of Technology (Caltech)** Pasadena, CA
  Intern *Summer 2003*
  - Performed simulations to study the atmospheric turbulence and improve on the adaptive optics technology using C++ .

Education

- **University of California, Los Angeles** Los Angeles,CA
  - Ph.D. in Physics 2014 – M.S. in Astronomy 2007 – B.S in Physics 2004

Skills

Languages: C/C++, Python, Matlab, IDL, shell scripting, MySQL,

Tools: Numpy, Scipy

SUMIN TANG

857-998-2775
tangsm@gmail.com
Burlingame, California
http://www.astro.caltech.edu/~stang/
http://www.linkedin.com/in/sumintang/
Github: tangsm

EXPERIENCE

FELLOW, INSIGHT DATA SCIENCE June 2014 – present

- Created Movie2Books.com, a recommendation app to suggest books you may like based on your favorite movie.
- Collected and cleaned Amazon review data from SNAP. Merged reviews with product images, genres and descriptions using Amazon, Goodreads, and Bing APIs. Stored data in a MySQL database.
- Performed natural language processing on review text using Python NLTK and Scikit-learn. Calculated Cosine and Jaccard similarities between movies and books. Performance is 20+ times better than the random baseline.
- Deployed as an interactive web app using Flask, HTML, Bootstrap, Javascript and AWS.

POSTDOCTORAL RESEARCHER, CALTECH AND UC SANTA BARBARA (joint position) 2012 – 2014

- Designed Python pipelines to identify and classify variable stars using periodic analysis and random forests. Successfully discovered the most distant R Cor Bor stars, and the most massive and fast-growing white dwarf.
- Extracted informative features and applied support vector machine to identify quasars. Successfully boosted the number of known quasars behind the Andromeda galaxy by a factor of 4, with a significantly improved precision of 90% (from 8% in traditional methods) to save telescope time (http://www.astro.caltech.edu/~stang/m31qso).
- Supervised two undergraduate students on data mining and period analysis on time series data.

GRADUATE RESEARCHER, HARVARD UNIVERSITY 2006 – 2012

- Developed automated data processing pipelines for a large set of time series data (100 TB+) using MATLAB and Python, including defect filtering for data cleaning and dynamic local-linear regression for data calibration. It provided calibrated brightness for stars, which were injected into a MySQL database.
- Performed data mining on time series data with automatic peak/trough detections, smoothing and trend analysis. Discovered new types of variable stars, and modeled their physical properties.
- Designed and implemented Monte Carlo simulations to rule out the widely accepted, yet artificial correlation between optical-to-X-ray ratio and luminosity of quasars.

EDUCATION

HARVARD UNIVERSITY, Ph.D. in Astrophysics Cambridge, MA, 2006 – 2012
TSINGHUA UNIVERSITY, M.S. in Physics with highest honors Beijing, China, 2003 – 2006
TSINGHUA UNIVERSITY, B.S. in Mathematics & Physics Beijing, China, 1999 – 2003

SKILLS

- **Languages:** Python, MySQL, MATLAB, Unix shell scripting, HTML, IDL (prior experience)
- **Tools:** Scikit-Learn, NumPy, SciPy, Matplotlib, NLTK, Flask, AWS
- Published 26 papers (12 first-authored) in top astrophysical journals with 360+ citations (www.astro.caltech.edu/~stang/pub.pdf)

HONORS AND AWARDS

- Rodger Doxsey Travel Prize (top 10% of Astronomy PhD dissertations in North America) 2012
- Chambliss Achievement Student Award, Honorable mention, American Astronomical Society 2009
- Philip Putnam Chase Memorial Fellowship, Harvard University 2006-2007

VANESSA HECKMAN

heckman@caltech.edu
(626)278-8844
San Jose, CA
www.linkedin.com/in/vanessaheckman/
Github: vanessaheckman

Experience

Insight Data Science—*Fellow* *Summer 2014*
Palo Alto, CA
- Developed Lenderstanding.com, a web app to assess default risk for P2P microloans.
- Analyzed 3.5 GB of private data from Zidisha, a microloan P2P nonprofit agency, using Python and MySQL. Scraped additional data using Beautiful Soup.
- Implemented HTTP basic authentication to restrict access to private data. *Un*: guest *Pw*: insight123
- Developed credit risk model for Zidisha to predict loan default.
- Deployed web app using AWS, Flask, Javascript, and JQuery for the front end.

California Institute of Technology—*Graduate Research Assistant* *2007 – 2013*
Pasadena, CA
- Developed two novel algorithms for detecting damage in instrumented civil structures based on the matched-filter method and anomaly detection (MATLAB).
- Conducted full-scale numerical testing running parallelized finite-element software on a computing cluster (PyLith and CUBIT).
- Conducted small-scale experimental testing (LABView), and analyzed acceleration data generated by instrumented buildings and bridges (~1 GB/day for densely-instrumented structures).
- Contributed to multi-disciplinary projects that culminated in communicating results through technical papers and presenting results at conferences and workshops.

Teaching Assistant: Linear Algebra and Applied Operator Theory, Engineering Mathematical Principles, Engineering Seismology, Experimental Methods in Earthquake Engineering

NSF JSPS EAPSI—*Fellow at Kyoto University* *Summer 2012*
Kyoto, Japan
- Applied signal processing techniques to monitor the seismic performance of an instrumented high rise, including seismic interferometry, filtering, and wavelet transforms/spectrograms (MATLAB).
- Applied system identification methods (maximum-likelihood estimation) to update model parameters using experimental data.

NASA Jet Propulsion Laboratory—*SURF Fellow* *Summer 2005*
La Cañada Flintridge, CA
- Designed and experimentally tested a piezoelectric bolt-breaking device (ANSYS, Solidworks).
- Achieved the design objective of fatigue-induced failure of a notched bolt using cyclic loading.
- Awarded a NASA Tech Brief, Bonsall Technical Writing Prize, and co-authored NASA Tech Report.

Skills

Programming languages: Python, MATLAB, MySQL, C/C++ (some experience)
Tools: scikit-learn, NumPy, Beautiful Soup, Pandas, LaTeX, SolidWorks, Adobe Illustrator
Other: Machine learning, System identification, Finite-element analysis (CUBIT, PyLith, ANSYS)
Relevant Coursework: Learning Systems (Caltech), Stochastic System Analysis and Bayesian Updating (Caltech), Machine Learning (Coursera), Data Analysis (Coursera)

Education

California Institute of Technology, Pasadena, CA
BS Engineering and Applied Science, 2006 PhD Civil Engineering, Geophysics Minor, 2014

VICKIE ZHANG

(626) 203-6218
San Francisco, CA
yzvickie@gmail.com
Github: yzvickie
https://www.linkedin.com/in/zhangvickie

Experience

Fellow, Insight Data Science **June 2014 – present**

- Created VentureNetwork, www.venturenetwork.us, a web app making personalized recommendations to help investors discover trending startups using Python, MySQL.
- Utilized AngelList API and web scraping (BeautifulSoup) to collect data on top 500 investors and over 7000 startups.
- Applied collaborative filtering techniques to recommend startups to investors using PCA. dimensionality reduction and cosine similarity algorithms in Scikit-Learn.
- Deployed as interactive web application using AWS, Flask, Javascript and JQuery for the front end.

Graduate Student Researcher, *UCSF Biomedical NMR Laboratory* **2008 – present**

- Designed and implemented multi-parametric MRI techniques to investigate post-radiotherapy response of prostate cancer.
- Analyzed metabolic biomarkers to differentiate cancer from normal tissue in patient biospies using logistic regression and k-mean cluster analysis.
- Applied nonlinear optimization algorithms in Matlab to model biological systems of preclinical animals.
- Published 8 original papers in top-rated journals such as Cancer Research and NMR in Biomedicine.
- Presented work 7 times at international scientific conferences.

Technical Lead, *ARDIS (Advanced Retinal Diagnostic Imaging Services)* **2013 – March 2014**

A venture in retinal imaging and analysis service

- Used data mining and predictive analytics with JMP (SAS) to predict disease progression and correlate therapeutic response for patients with age-related macular degeneration.
- Conducted over 50 customer interviews to identify a repeatable and scalable business model.

Application Scientist, *Gamma Medica-Ideas, Preclinical Imaging* **2007 – 2008**

- Managed clients from top global institutions, including application training and collaborative research.
- Led a team to develop, Q/A, and launch micro-volumetric gamma blood counters to market.
- Achieved completion of 6 pre-clinical imaging product sales, installations, & training – involved over $10 million in revenue.

Research Associate, *UCSF Physics Research Laboratory* **2005 – 2006**

- Designed and implemented micro-CT/SPECT acquisition techniques to characterize lung tumors and bone metastasis in transgenic mice.
- Developed an automated image-processing algorithm with Matlab that utilizes micro-CT images to quantify number, size, and shape of lung tumors in transgenic mice.

SKILLS

Languages: Python, MySQL, Matlab, javaScript
Tools: Flask, jQuery, BeautifulSoup, Numpy, Scipy, Pandas, Matplotlib, Scikit-Learn, JMP (SAS), Slicer, Matlab Imaging Processing Toolbox, Matlab Optimization Toolbox

EDUCATION

UC Berkeley – UCSF, San Francisco, CA (expected) August, 2014
Ph.D. Candidate, Joint Graduate Group in Bioengineering
NIH Training Fellowship (2009-2010), American Scandinavian Foundation Scholarship (2009)

UC Berkeley, Berkeley, CA 2006
B.S. Bioengineering

WAFA SOOFI

wafa.soofi@gatech.edu (281) 686-9156
http://www.linkedin.com/in/wsoofi South San Francisco, CA
http://github.com/wsoofi

EDUCATION

Georgia Institute of Technology and Emory University
Ph.D. in Biomedical Engineering, *May '14*
Relevant coursework: Advanced ODEs, Advanced Digital Signal Processing

Rice University
B.S. in Bioengineering, *May '08*

TECHNICAL SKILLS

Languages: MATLAB/Simulink, Python, SQL, C/C++ (some exposure), HTML, Javascript (some exposure)
Tools: Pandas, Numpy, Scipy, Scikit-Learn, MySQL, Flask, git, jQuery, SPSS
Other Skills: Data mining, Digital signal processing, Nonlinear dynamical systems theory, Numerical methods, Linear algebra, Statistics

EXPERIENCE

Fellow at Insight Data Science, Palo Alto, CA **Jun '14 – Present**

- Developed www.CrowdSkippr.com, a web application for predicting the size of crowds at national parks by analyzing photo metadata from Flickr.com.
- Utilized Flickr API and NOAA.gov to gather photo and weather data; stored NOAA data in MySQL.
- Determined significant predictors of crowd size using machine learning techniques (random forests and gradient boosting) in Python with numpy, scipy, pandas, and scikit-learn.
- Constructed an interactive front end with Flask, Bootstrap, Javascript and HTML; hosted on AWS.

Graduate Researcher, Georgia Institute of Technology/Emory University, Atlanta, GA **Aug. '08 – May '14**

- Systematically mined high-dimensional database of neuron models in MATLAB to gain insights into the effects of specific ionic currents on the output of small neuronal networks.
- Constructed real-time Linux closed-loop hybrid network of neurons in C.
- Designed algorithm to quantify effects of periodic, controlled stimulation on living neurons.
- Performed signal processing on noisy time-series data in MATLAB and statistically quantified environmental effects on neuronal output.
- Published two first-author papers in high-impact journals and one book chapter on high-dimensional model databases; presented work at several international conferences and one invited talk (Free University of Berlin).

Intern in M.D. Anderson Department of Radiation Physics, Houston, TX **Jun. – Aug. '07**

- Optimized four-dimensional dose calculations for radiation treatment plans in human cancer patients and reduced theoretical computation time by 25%. Published one first-author paper.

SELECTED FELLOWSHIPS/AWARDS

- NSF Graduate Research Fellowship (3 yrs. stipend/tuition/fees) **Awarded '08**
- NSF Integrative Graduate Education and Research Traineeship **Awarded '08**
- Georgia Tech President's Fellowship **'08 – '12**

LEADERSHIP EXPERIENCE

- Executive board member of American Society for Engineering Education (GT chapter); organized monthly meetings, symposia, and annual workshops on effective techniques in STEM education
- Served as teaching assistant for Biotransport (BMED 3300) and Biomechanics (BMED 3400)
- Interviewed, hired, and evaluated tutors for Houston Scholar Program

Woei Ling Leow

Berkeley, CA | leowl@alum.mit.edu | 617-4016634 | www.github.com/leowwl | www.linkedin.com/in/leowl

EDUCATION

***Ph.D. (Engineering Systems)*, MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT), Cambridge MA** **2007 - 2012**

● Achieved distinctions for classes on system dynamics, machine learning and probabilistic systems modeling; also attended classes on network science and engineering, inference and dynamic programming
● Awarded the Martin Family Fellowship (MIT) and the Tan Kah Kee Postgraduate Scholarship (Singapore)

***M.S. (Electrical & Computer Engineering)*, UNIVERSITY OF MASSACHUSETTS, Amherst MA** **2005 - 2007**

● Thesis research investigated performance of wireless networks and optimal traffic sensor deployment
● Accorded the Outstanding Teaching Assistant Award for excellence in instructing a course on probability and statistics

***B.Eng (Electrical)*, NATIONAL UNIVERSITY OF SINGAPORE, Republic of Singapore** **1998 - 2002**

● Graduated with a Second Class Honors (Upper Division) and placed on the Dean's List in first term of junior year

SKILLS AND TOOLS

Python, Numpy, SciPy, Pandas, Git, (My)SQL, Flask, Bootstrap, MATLAB, Virtualenv, Gunicorn
Prior Experience: C/C++, FORTRAN, Java, LonWorks, Neuron C, 80x86-based assembly language
Skills: Data analysis, (approximate) dynamic programming, machine learning, modeling, simulation, system dynamics, optimization, principal component analysis (PCA), algorithm development and experimentation, applied probability, regression

EXPERIENCE

Fellow, Insight Data Science, Palo Alto CA **June 2014 - Present**

● Developed a novel data-driven car recommender (analyticar.org) that takes into account the buyer's commute patterns and help correct misconceptions about cost of ownership of different types of cars
● Gathered different types of car data (e.g. fuel consumption, maintenance cost, etc.) from government and commercial sources, processed the raw data, and merged and stored the data in a MySQL database
● Deployed web app with Bootstrap front-end and Python-Flask-MySQL backend ran by Gunicorn on Amazon Web Services

***Energy Policy Postdoctoral Fellow*, Lawrence Berkeley National Laboratory, Berkeley CA** **2013 - Present**

● Investigated energy efficiency market trends of electrical appliances using different techniques multi-variate linear regression, PCA, hedonic regression etc. in the Python-Numpy-Pandas environment
● Developed models, based on econometrics and hedonic regression, using data obtained from the Web to characterize the relationship between price and energy efficiency in refrigerators
● Contributed to a Django-based web application to scrape, store and analyze data of electrical appliances from the Web in collaboration with co-workers using Python and Git

***Research and Teaching Assistant*, Massachusetts Institute of Technology, Cambridge MA** **2007 - 2012**

● Developed algorithms for a learning residential electricity manager to minimize heating and cooling costs using approximate dynamic programming and Monte Carlo simulation amid uncertainty arising from time-varying electricity price, weather, etc.
● Assisted in developing and instructing a new course "Modeling Risk, Dynamics, and Decisions" as a teaching assistant

***Engineering Services Intern,* EnerNOC Inc*.*, Boston MA** **Summer 2008**

● Developed inverse modeling algorithm and predictive model in MATLAB to project electricity consumption of buildings based on stochastic variables, including outdoor air temperature, building occupancy

***(Senior) Research Engineer*, Honeywell Pte. Ltd, Republic of Singapore** **2002 - 2004**

● Promoted to senior research engineer and team lead position in 2 years, tasked to lead technology development activities
● Headed team to provide in-house technical consulting and marketing services to business units in Asia-Pacific
● Certified as a Honeywell International Green Belt for Design for Six Sigma
● Led team to deliver one month ahead of schedule a LonWorks-based card reader that enabled entry into the Japanese market
● Collaborated with multi-national teams to develop advanced features for video surveillance systems implemented in C/C++

Zack Chadick

Contact zack.chadick@gmail.com http://www.linkedin.com/in/zackchadick
Redwood City, CA GitHub: zchadick
858.248.3323

Experience

Fellow *at Insight Data Science (Palo Alto, CA)* **June 2014 – Present**

Created www.StealThisHouse.info to predict home prices changes based on historical crime data.

- Created database containing previous 3 years worth of crime (crimespotter.org) and home sale data (zillow.com) and applied predictive algorithm to forecast home prices.
- Deployed tool as a web application showing historical trends and predicted future value change (AWS, Flask, Python, JavaScript, MySQL, SciKit, PyMySQL).

Postdoctoral Fellow *at University of California, San Francisco* **Jan 2011 – June 2014**

Investigated the neural circuits that underlie addiction.

- Developed software (MATLAB, SPSS, XML, Python) to integrate behavioral and physiological data into a single unified processing stream.
- Analyzed neural and behavioral data using a variety of statistical techniques (Monte-Carlo simulations, general linear models, logistic regression, principle component analysis, cluster analysis, Bayesian modeling, discriminate analysis, etc.).
- Trained and supervised laboratory assistant in experimental development/execution/analysis.
- Collaborated with lab-members on data analysis to classify neural recording based on response.
- Wrote 2 papers (one with collaborators) and presented work at 3 scientific conferences.

Ph.D. Researcher *at University of California, San Francisco* **Sept 2005 – Dec 2010**

Examined the relationship between age-related cognitive decline and decline of neural circuits.

- Deployed and debugged initial data acquisition systems for researchers including development of hardware to integrate them. Have returned to provide continuing, ongoing support.
- Wrote and standardized data processing streams that are still used by the lab (MATLAB, shell-scripting) to integrate multiple, simultaneous streams of physiological and behavioral data.
- Utilized a wide variety of statistical techniques to analyze complex neural and behavioral datasets (permutation tests, generalized linear models, coherence and power analysis, wavelet decomposition, independent component analysis).
- Collaborated closely with colleagues for design, execution, and analysis of neurophysiological experiments leading to 5 publications and presented work at 5 scientific conferences.

Research Assistant II *at Scripps Research Institute (La Jolla, CA)* **July 2003 – Aug 2005**

Developed large-scale purification protocols for multi-subunit protein complexes for structure determination using single-particle cryo-electron microscopy.

- Developed novel techniques for tagging specific protein subunits and masking data for structural analysis (SPIDER, Fortran, and clustering/classification algorithms).
- Supervised two high-school summer researchers and published 3 peer-reviewed publications.

Education

University of California, San Francisco **Dec 2010**
Ph.D. in Neuroscience

California Institute of Technology, Pasadena CA **June 2003**
B.S. in Biology and Chemistry (Graduated with honors, double major)

Skills

Languages: MATLAB, Python, SQL, SPSS, HTML, and JavaScript (familiar)
Tools: LaTeX, BeautifulSoup, PyMySQL, NumPy, SciKit (learn, stats), Flask, JQuery, AWS
Other: Resampling methods, general linear models, clustering algorithms, principle/independent component analysis, descriptive statistics, regression models

Honors & Awards

National Research Service Award (National Institute of Health) **Jan 2013 – Dec 2015**
Larry Hillblom Graduate Fellowship **Aug 2007 – Aug 2008**