Scott Lee Schwartz, Ph.D.
iam.scottschwartz@gmail.com
github/pointOfive
linkedin/scottschwartz

Customer Experience Data Scientist
Google Cloud Support
Google, New York, 10011

February 27th, 2018

To whom it may concern:

I am interested in the Customer Experience Data Scientist position at Google. I am a formally trained statistician with a strong background in data analysis pipelining; I have deep understanding and practical hands-on experience in statistical and machine learning methodology, including experience teaching regression testing, statistical modeling, machine learning, and predictive analytics, etc. to diverse audiences; I have an extensive and proven track record of success in interdisciplinary, collaborative, consultative client facing roles; and I very much enjoy and excel in written, verbal, and visual communication of complex technical concepts.

As a data scientist I am oriented towards identifying innovation opportunities and establishing best practice precedents. I work primarily in Python but I am comfortable working in other environments, such as R, C++, SQL and bash. For representative examples of my work please visit my github page which showcases my capabilities in

- data dashboarding and visualization
- the modern data science toolkit
- advanced predictive methodologies
- data infrastructures and pipelining

I'm naturally a self-starter who is motivated by taking ownership in the process of providing the highest-quality meaningful and impactful productivity that builds trust and effective relationships. I'm a happy, healthy, sincere, social, engaged and focussed team-player who brings energy, drive and determination to my work; and I am an effective listener, a thorough communicator, and I derive satisfaction from working hard in environments that value comradery, community, and mission.

Thank you for your attention and consideration – I look forward to hearing back from you.

Scott Lee Schwartz, Ph.D.

Scott Lee Schwartz, Ph.D.

iam.scottschwartz@gmail.com

github/pointOfive
linkedin/scottschwartz

Education

2010-2011	Postdoctoral	*Bioinformatics*	**Texas A&M University**
2005-2010	Ph.D., M.S.	*Statistics*	**Duke University**
2001-2005	B.S., B.A†	*Computer Science, Mathematics*	**Trinity University**

Coding

- Python
- R
- SQL
- Bash
- AWS/HPC
- C++
- git

Packages

- bokeh
- sklearn
- spark
- gensim
- statsmodels
- scipy.stats
- np/pd/plt

Machine Learning

- Ensemble Methods
- Boosting (XGBoost)
- Support Vector Machines
- Neural Networks (Keras)
- Hierarchical Clustering
- PCA/SVD/NMF
- Recommenders

Interdisciplinary Work

- Genotyping/GWAS
- Mapping Populations
- Bulk Segregation
- RAD-Seq/WGS
- RNA/TAG-Seq/DE
- Start Site Analysis
- Allele Specific RNA

Predictive Methodology

- Loss Functions/Metrics
- Regularization
- Cross Validation
- Confusion Matrices
- Cost/Benefit Matrices
- Model Interpretation

Statistical Inference

- Bayesian Analysis
- Mixture Models
- Hypothesis Testing
- Nonparametric Tests
- Regression and GLMs
- Model Diagnostics

Data Pipelining

- fastQC/Illumina
- fastX/cutadapt
- BWA/bowtie/SAM
- IGV/VCF/GFF
- SAMtools/GATK
- HTseq/TopHat

Experience

2016-2017 *Instructor, Sr. Data Scientist* **Galvanize, ATX & NYC**

Content development, lecturing, guidance, mentoring; recruiting, hiring, promotion, networking

2014-2016 *Research Associate, Integrative Biology* **University of Texas**

Bioinformatics pipeline creation and data management; mentoring and statistical analysis support

2011-2014 *Bioinformatic Analyst, Next Generation Sequencing* **Texas A&M AgriLife**

Lead bioinformatic analysis and consulting services team; managed data QC and delivery pipeline

2010-2011 *Research Associate, Postdoctoral Fellowship* **Texas A&M University**

Statistical and genomic data analysis, consulting, and project support in a basic science wet lab

2007-2010 *Instructor and Consultant, Statistical Science* **Duke University**

Experimental design and data analysis consultation; teaching and mentoring for statistics courses

2006-2007 *Research Assistant, Children's Environmental Health* **Duke University**

Collaboration with diverse interdisciplinary team in an applied and translational research setting

Scott Lee Schwartz, Ph.D.

iam.scottschwartz@gmail.com

github/pointOfive
linkedin/scottschwartz

Publications

- Wang, X, Schwartz SL, Lovell JT, Juenger T. *Photoperiod regulation of diurnal transcription in Panicum Hallii.* **In Prep**..
- Jin H, Schwartz SL, Vvedenskaya I, Malik I, Pugh BF, Nickels B, Kaplan CD. Pol II activity and promoter architecture determine transcription start site usage in Saccharomyces cerevisiae. **In Prep**..
- Lambertz IU, Luo L, Berton TR, Schwartz SL, Hursting SD, Conti CJ, Fuchs-Young R. Early Exposure to a High Fat/High Sugar Diet Increases the Mammary Stem Cell Compartment and Mammary Tumor Risk in Female Mice. **Cancer Prevention Research**, 2017.
- Lovell JT, Shakirov EV, Schwartz SL, Lowry D, Aspinwall A, Taylor S, Bonnette J, Hawkes C, Fay P, Juenger TE. Promises and challenges of eco-physiological genomics in the field: tests of drought responses in Switchgrass. **Plant Physiology**, 172(2), 2016: 734-48.
- Lovell JT, Schwartz SL, Lowry D, Shakirov E, Wang M, Johnson J, Sreedasyam A, Plott C, Jenkins J, Schmutz J, Juenger T. *Drought responsive gene expression and regulatory divergence between upland and lowland ecotypes of a perennial C4 grass.* **Genome Research**, 26(4), 2016: 510-18
- Torres M, Ghaffari N, Buiate EAS, Moore N, Schwartz SL, Johnson CD, Vaillancourt L. *A Colletotrichum graminicola mutant deficient in the establishment of biotrophy reveals early transcriptional events in the maize anthracnose disease interaction.* **BMC Genomics**, 17(202), 2016.
- Su, Z., et al. *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the SEQC Consortium.* **Nature Biotechnology**, 32, 2014: 903-14
- Clavijo A, Nikooienejad A, Shahrokh M, Metz R, Schwartz SL, Atashpaz-Gargariz E, Deliberto TJ, Lutman MW, Pedersen K, Bazan LR, Swenson SL, Koster LG, Zang M, Beckham T, Johnson C, Bonpheng M. *Identification and phylogenetic analysis of the first pandemic (H1N1) 2009 influenza virus from feral swine.* **Zoonoses Public Health**, 60(5), 2013: 327-35.
- Schwartz SL, Ivanov IV, Davidson LA, Goldsby JS, Dahl DB, Dougherty ER, Herman D, Donavan SM, and Chapkin RS. *A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response.* **Genome Biology**, 13(4), 2012.
- Shah MS, Schwartz SL, Zhao C, Davidson LA, Zhou B, Lupton JR, Ivanov I, and Chapkin RS. *Integrated microRNA and mRNA expression profiling in a rat colon carcinogenesis model: Effect of a chemoprotective diet.* **Physiological Genomics**, 43(10), 2011: 640-54.
- Schwartz SL, Li F, Reiter JP. *Sensitivity analysis for unmeasured confounding in principal stratification.* **Statistics in Medicine**, 31(10), 2012: 949-62.
- Schwartz SL, Li F, Mealli F. *Dirichlet processes for flexible modeling of continuous intermediate variables using principal stratification.* **Journal of the American Statistical Association**, 106(496), 2011: 1331-44.
- Schwartz SL, Gelfand A, Miranda ML. *Joint Bayesian analysis of birthweight and censored gestational age using finite mixture models.* **Statistics in Medicine**, 29(16), 2010: 1710-23.
- Dissertation: Bayesian Mixture Modeling Approaches for Intermediate Variables and Causal Inference. Advisors: Drs. Fan Li and Jerome P. Reiter. Duke University. 2010.

Graduated† *Summa Cum Laude*, awarded Barry M. Goldwater Scholarship and two NSF CSEMS Scholarships, selected Class of 2005 Outstanding Computer Science Student, and played collegiate soccer, winning the NCAA DIII Men's Soccer National Championship in 2003 and Academic All-American honors in 2005.