

# Daniel Klein 微担意

web: mesokurtosis.com • github: othercriteria • email: daniel.l.klein@proton.me • 952.454.4197

## Experience

### Oracle Corporation (Advertising and Customer Experience, Moat)

2019 – present

*Principal Data Scientist*

- Develop and maintain streaming pipelines for Invalid Traffic detection, consuming ~250,000 events / sec.
- Migrate high spend systems cross-cloud and cross-architecture, validating correctness and robustness.
- Ensure use of appropriate data- and statistics-driven methods within-team, cross-team, and externally.

### DMG Blockchain Solutions (Blockseer)

2018 – 2019

*Senior Software Engineer*

- Led rewrite of Bitcoin address clustering (EMR, Scala, GraphX); 1.2 TB blockchain processed in ~6 hours.
- Validated novel Bitcoin transaction classification method using custom DeepWalk implementation.

### RealScout

2016 – 2017

*Data Scientist*

- Productionized insights from agent-agent modeling into user-facing Recommended Listing feature.
- Automated schema mapping for ETL of external listing databases (MLS) using gradient boosting match model.
- Built out Bayesian models for sales price and days-on-market prediction with near-Zillow/Redfin accuracy.

### Radius Intelligence, RealScout, Neighborly (rotations)

2016

*Data Science apprentice (Catenus, 8VC)*

### TransForm Pharmaceuticals, Inc.

2006 – 2008

*Assistant Scientist, Scientific Computation*

- Delivered tooling to support computational chemistry research and data integration (HPLC, IR, UV-Vis) for lab automation platforms.

### Williams College, Dept. of Biology

2005

*Research Assistant*

- Planned and implemented experimental design and data analysis for field research project.

### University of Minnesota, Dept. of Ecology and Evolutionary Biology

2003, 2004

*Research Intern*

- Developed and analyzed numerical results from novel model for ecological community assembly.

## Education

### Brown University, Ph.D. Applied Mathematics

2008 – 2016 (did not complete)

- Advised by Matt Harrison; research on statistical inference in settings of extreme data sparsity or imbalance.
- TA for Intro Stats, Math Stats I/II, Recent Applications in Probability and Statistics.

### Williams College, B.A. Mathematics and Biology

2006

- Departmental honors in Biology, with thesis “Understanding aggregation in the membracid *Publilia concava*: using models to disentangle processes”.

## Relevant coursework

- Foundational CS: Data Structures, Algorithm Design/Analysis, PL, PL Theory.
- Bayesian Stats, Biostats, graduate Math Stats I/II, graduate Probability/Stochastics.
- Seminars in Graphical Models, Bayesian Nonparametrics, Network Models.
- Computational Biology: dynamic programming, approximation algorithms, HMMs, etc.

## Value over replacement player / comparative advantage

Relative to an engineer with the typical software engineer and data scientist competencies sketched out above:

- I’m constructively skeptical of data, design guarantees, and monitoring; these add value iff they track reality.
- I’m unsentimental about using dumb methods over theoretically optimal ones or those I’ve built myself.
- I look to the non-technical contexts of product, growth, and markets to frame technical choices. I read The Diff.