

Kyle A. Chezik

☎ (206) 832-5883 📍 Seattle, Washington ✉ kchezik@gmail.com
🐙 github.com/kchezik in linkedin.com/in/kyle-chezik 🏠 kchezik.github.io

Skills

| | |
|------------------|---|
| Languages | Python, R, SQL, Unix/Linux, Git, Regex, Stan |
| Machine Learning | Supervised, unsupervised & deep learning, logistic regression, hierarchical & generalized linear models, time series analysis & forecasting, simulation, classification, random forest, clustering, feature selection & feature engineering |
| Statistics | A/B Testing, probability, likelihood, inference |
| Tools | Jupyter, Pandas, NumPy, Scikit-learn, R-Studio, Tidyverse, Caret, LME4, RStan, <i>web-dev</i> : HTML, CSS, AWS, Flask, <i>data-vis</i> : GGPlot2, Matplotlib, Seaborn, R-Shiny |

Experience

- 2019 **Data Science Consultant**, Seattle WA, USA
- Built a stochastic consumption **Bayesian structural time series** model in **pyStan** to predict stock-out for Bottomless, a Y-Combinator backed company providing precise coffee re-supply.
 - Incorporated an **Economic Order Quantity model** to dynamically estimate re-order points that limit over-stock and stock-out risk in an uncertain delivery environment.
 - Extracted 180k records from 350⁺ users held in a **MongoDB** and developed a data cleaning algorithm that combined probabilistic and logical processes resulting in full automation.
- 2019 **Insight Data Science Fellow**, Seattle WA, USA
- Developed an **interactive recommender** using **deep learning**, **computer vision** and cosine similarity to help gardeners find native plants that meet their aesthetic tastes.
 - Scraped and decomposed **1400⁺** plant images into 512 features using the **convolutional neural network** ResNet18 in **PyTorch**, and combined meta-data from multiple databases.
- 2013-19 **Research Assistant**, Simon Fraser University, Burnaby BC, Canada
- Automated error identification in time series data using a **Bayesian Hidden Markov model** with 84% accuracy across **1 Million⁺** records. Significantly reduced human work hours, and earned the SFU KEY Big Data Graduate Scholarship.
 - Identified novel river-network properties using **linear regression**, **simulations**, **ARIMA** processes and **parametric bootstrapping**. Achieved 98% certainty.
 - Used **periodic time series** and **generalized hierarchical spatial network models** to determine stream temperature drivers and assess salmon heat risk.
 - **Feature engineering** for gridded data with **GIS** (e.g., *GDAL*, *OSGEO*) and **parallel computing** within Python (*WhiteboxGAT*) and on the command line (*GNU parallel*).
- 2016 **Data Engineer Contractor**, ESSA Technologies Ltd., Vancouver BC, Canada
- Constructed and managed a relational database of **~4 million records**, for a river network model.
 - Aggregated messy data from multiple sources. Developed and packaged R functions for **end-to-end re-productibility** and improved data acquisition efficiency.
 - Used **feature engineering** of river flow, temperature and landscape data to improve model accuracy.
-

Education

- 2019 **Ph.D.** Biological Sciences, Simon Fraser University, Burnaby BC, Canada
- 2013 **M.Sc.** Conservation Biology, University of Minnesota, St. Paul MN, USA
- 2009 **B.A.** Biology, St. Olaf College, Northfield MN, USA