Chris Salahub

chris.salahub@uwaterloo.ca | chris-salahub | Salahub | @Salahub

Summary

Technical Skills: Simulation, Data Visualization, Dimension Reduction, Regression, Prediction, Inference, Text Mining

Programming Languages: R, Python, SQL, Bash, MATLAB, C++

Domain Knowledge: Data Science, Genomics, Experimental Design, Machine Learning, Multiple Testing, Regression (GLM, GAM, Mixed Models), Process Control and Analysis, Statistical Graph Theory

Education

Ph.D. in Statistics, University of Waterloo — Waterloo, ON, Canada
Scholarship: Alexander Graham Bell Doctoral Scholarship ($105,000 over 36 months) — May 2019 – Dec. 2023
Thesis: *Explorations in Pairwise Measures of Association and Pooled Significance*

M.Sc. in Statistics, ETH Zürich (Computer Vision Specialization) — Zürich, Switzerland
Scholarship: ESOP Scholar ($48,000 for being in the top 2-3% of incoming ETH Masters Students) — Sept. 2017 – Mar. 2019
Thesis: *Seen to Be Done: A Graphical Analysis of Peremptory Challenge*

B.Math. in Statistics, Dean's Honours List, University of Waterloo — Waterloo, ON, Canada
Thesis: *About "her emails"* — Sept. 2013 – June 2017

Professional Experience

Measurement Data Scientist — Toronto, ON, Canada
Vividata — Aug. 2020 – Jan. 2022
Clustering: Pioneered a new segmentation of demographic data using multiple correspondence analysis
Data pipelines: Integrated SQL and R seamlessly to produce daily survey and click stream data reports
Survival analysis: Used Cox regression to predict panel attrition and virtually eliminate over-recruitment

Lecturer (Data Visualization) — Waterloo, ON, Canada
University of Waterloo — Sept. 2021 – Jan. 2022
Organization: Taught 80+ students and directed a team of 4+ teaching assistants
Data visualization: Presented topics including dimension reduction, spatial statistics, interactive graphics, and programming in R

R&D, Data Science Intern — Toronto, ON, Canada
Environics Analytics — May 2016 – Aug. 2016, May 2017 – Aug. 2017
Algorithm development: Reduced convergence time of a demographic microsimulation by more than 90% using a Markov Chain model
Modelling: Designed, implemented, and optimized a simulated annealing algorithm from scratch to solve a network flow problem which improved the fit of simulated data to observed data by 50% when implemented in R and MATLAB

Publications

Optimal structured matrix approximation for robustness to incomplete biosequence data. Chris Salahub and Jeffrey Uhlmann. IEEE Transactions on Computational Biology and Bioinformatics. Under review.

Impact: Reduced sensitivity of genetic correlation matrix estimates to missing data by up to 85%

Racial bias in jury selection. Chris Salahub. Significance, 20(2):16-20. April 2023.

Impact: Devised a novel plot to visualize multinomial regression and used it to detect racial patterns in jury trial data

On cycling risk and discomfort: urban safety mapping and bike route recommendations. David Castells-Graells, Christopher Salahub, and Evangelos Pournaras. Computing, 102:1259–1274. December 2019.

Impact: Combined accident, map, and insurance data in an app allowing users to plan cycling routes balancing comfort and safety

About "her emails". Christopher Salahub and R. Wayne Oldford. Significance, 15(3):34-37. June 2018.

Impact: Processed 32,795 emails using regular expressions to extract metadata and built an Rshiny web app for interactive exploration

Open Source Software Packages

PoolBal: tools to compute and control the centrality of pooled p-values in multiple testing (CRAN, Nov. 2023)

AssocBin: implements a recursive binary partitioning algorithm to measure association (CRAN, Nov. 2023)