Olivier Binette

olivier.binette@duke.edu / *github.com/OlivierBinette* / *olivierbinette.ca*

Summary

- 5 years of postgraduate experience in statistics research and data science with expertise in **machine learning**, **statistical evaluation**, and **entity resolution**.
- Strong applied research background with publications in top journals and several awards.
- Advanced knowledge of modern data science and software development practices.

Education

Duke University **2019 – present**
PhD Candidate, Statistical Science Department (3.9 GPA) *Durham, NC*

Université du Québec à Montréal **2014 – 2019**
BSc, Mathematics (3.97 GPA); MSc, Statistics (4.0 GPA) *Montréal, QC*

Industry Experience

Data Scientist Intern | *American Institutes for Research* **May 2022 – present**

- Led the development of statistical evaluation methodology for machine learning systems used at PatentsView.org. Resulted in an open source Python package (github.com/PatentsView/PatentsView-Evaluation) and a scientific paper:
    - **Binette, O.**, S. A. York, E. Hickerson, Y. Baek, S. Madhavan and C. Jones. (2022) Estimating the Performance of Entity Resolution Algorithms: Lessons Learned through PatentsView.org. *arXiv e-prints*. arxiv:2210.01230

Data Scientist Intern | *Intact Financial Corporation* **January 2022 – April 2022**

- Machine learning for pricing optimization within the Rating Revolution team. Developed uncertainty quantification and model-based optimization methods for improved decision-making.
- Python package development with the AI engineering team. Addressed tooling issues through improved development practices.

Open-Source Developer (part time) | *Github Sponsors* **2022 - present**

- Open-source Python and R package development on Github. Funded by G-Research and individual contributors.

Research Coordinator (part time) | *Duke Community Food Pantry* **May 2021 – May 2022**

- Developed research protocol and survey design (adopted by Duke University) to monitor food insecurity on campus.
- Record linkage, survey design, and data analysis for program assessment.

Project Assistant (part time) | *Wilson Center for Law and Justice at Duke Law* **June 2021 – August 2021**

- Wrote R bindings to C libraries for forensic fingerpring matching and data analysis.

Data Science Project Manager (part time) | *Information Initiative at Duke* **June 2021 – August 2021**

- Led a team of students building a R Shiny data visualization dashboard for the UC Davis Perinatal Origins of Disparities Center.

Data Science Project Lead (part time) | *Duke Applied Machine Learning Group* **January 2021 – May 2021**

- Led a team of students working on optical character recognition and article segmentation for historical newspaper archives.

Data Science Project Manager (part time) | *Information Initiative at Duke* **June 2020 – August 2020**

- Led a team of students analyzing time series data and geospatial data in R for the Duke River Center.

Academic Research Experience

Graduate Researcher | *Duke University, Statistical Science Department* **2019 – 2021**
Publications:

- **Binette, O.** and Steorts, R. C. (2021) On the Reliability of Multiple Systems Estimation for the Quantification of Modern Slavery. *Journal of the Royal Statistical Society Series A 185* (2), pp. 640 – 676 (American Statistical Association best paper award)
- **Binette, O.** and Steorts, R. C. (2021) Modern Bayesian Entity Resolution. *Wiley StatsRef* (to appear)
- **Binette, O.** and Steorts, R. C. (2021) (Almost) All of Entity Resolution. *Science Advances 8* (12) eabi8021

Graduate Researcher | *Université du Québec à Montréal* **2017 – 2019**
Publications:

- **Binette, O.** and Guillotte, S. (2021). Bayesian Nonparametrics for Directional Statistics. *Journal of Statistical Planning and Inference 216* pp. 118-134
- **Binette, O.** and Pati, D. and Dunson, D. B. (2020) Bayesian Closed Surface Fitting Through Tensor Products. *Journal of Machine Learning Research 21* (119) pp. 1-26
- **Binette, O.** (2019). A Note on Reverse Pinsker Inequalities. *IEEE Transactions on Information Theory 65* (7). pp.4094-4096

Leadership, Teaching, and Mentoring Experience

Instructional Teaching Assistant | *Duke University* **2020 – present**

- Teach weekly labs for *Spatio-Temporal Models*, *Bayesian and Modern Statistics*, *Entity Resolution*, and *Introduction to Data Science*.

Webmaster (part time) | *ASA Record Linkage Interest Group* **November 2021 - present**

- Website and blog development for the organization (recordlinkageig.github.io).

Mentor (part time) | *Twoples* **2020 – 2021**

- Mentored two college students in semester-long research projects.

Instructional Teaching Assistant | *Université du Québec à Montréal* **2017 – 2019**

- Teach weekly labs for *Statistics I*, *Analysis I*, *Probability II*, *Complex Analysis*, *Analysis and algebra for the actuarial sciences*, *Analysis II*.

Awards

- G-Research PhD Student Grant - Open Source Software for Big Data Integration (2000 £)
- American Statistical Association Best Paper Award (Section on Survey Methods)
- Canada Governor General's Academic Gold Medal (2020)
- Stanford University fully-funded PhD admission offer (2019)
- Faculty of Arts and Science Top Doctoral Award (University of Toronto, 2019)
- Alexander-Graham-Bell Canada Graduate Scholarship (2019; 105 000 $)
- Fonds de Recherche du Québec - Nature et Technologies Doctoral Award (2019; 84 000 $)
- Natural Sciences and Engineering Research Council of Canada Masters Award (2017; 21 000 $)
- Fonds de Recherche du Québec - Nature et Technologies Masters Award (2017; 21 000 $)

Programming and Software Development Skills

Programming: Python, R, C/C++, bash, git, Linux, Docker
R packages on Github: dgaFast, MSETools, assert, cache, fingermatchR, TessTools
Python packages on Github: PatentsView/PatentsView-Evaluation, OlivierBinette/StringCompare