

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 statistics, machine learning, evaluation, computational social science, and record linkage/entity resolution.
- Strong research background with publications in top journals and several awards.
- Advanced knowledge of modern data science and software development practices.

Education

Duke University <i>PhD Candidate, Statistical Science Department (3.9 GPA)</i>	2019 – present Durham, NC
Université du Québec à Montréal <i>BSc, Mathematics (3.97 GPA); MSc, Statistics (4.0 GPA)</i>	2014 – 2017 Montréal, QC

Data Science and Applied Research Experience

Data Scientist Intern <i>American Institutes for Research</i>	May 2022 – present
<ul style="list-style-type: none">• Led the development of statistical evaluation methodology for the entity resolution systems used at PatentsView.org. Resulted in an open source Python package (github.com/PatentsView/PatentsView-Evaluation) and a scientific paper:<ul style="list-style-type: none">◦ Binette, O., S. A. York, E. Hickerson, Y. Baek, S. Madhavan and C. Jones. (2022) Practical Performance Evaluation of Entity Resolution Algorithms: Lessons Learned at PatentsView.org. <i>To appear</i>.	
Data Scientist Intern <i>Intact Financial Corporation</i>	January 2022 – April 2022
<ul style="list-style-type: none">• 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) <i>Github Sponsors</i>	2022 – present
<ul style="list-style-type: none">• Open-source Python and R package development on Github. Funded by G-Research and individual contributors.	
Research Coordinator (part time) <i>Duke Community Food Pantry</i>	May 2021 – present
<ul style="list-style-type: none">• Developed research protocol (adopted by Duke University) to monitor food insecurity on campus.• Record linkage, survey design, and data analysis for program assessment.	
Project Assistant (part time) <i>Wilson Center for Law and Justice at Duke Law</i>	June 2021 – August 2021
<ul style="list-style-type: none">• Wrote R bindings to C libraries for forensic fingerprint matching and data analysis.	
Data Science Project Manager (part time) <i>Information Initiative at Duke</i>	June 2021 – August 2021
<ul style="list-style-type: none">• 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 <i>Duke Applied Machine Learning Group</i>	January 2021 – May 2021
<ul style="list-style-type: none">• Led a team of students working on optical character recognition and article segmentation for historical newspaper archives.	
Data Science Project Manager (part time) <i>Information Initiative at Duke</i>	June 2020 – August 2020
<ul style="list-style-type: none">• Led a team of students analyzing time series data and geospatial data in R for the Duke River Center.	

Peer-Reviewed Academic Research Experience

Research Assistant <i>Duke University, Statistical Science Department</i>	2019 – 2021
Publications:	
<ul style="list-style-type: none">• Binette, O. and Steorts, R. C. (2021) On the Reliability of Multiple Systems Estimation for the Quantification of Modern Slavery. <i>Journal of the Royal Statistical Society Series A</i> 185 (2), pp. 640 – 676 (American Statistical Association best paper award)• Binette, O. and Steorts, R. C. (2021) Modern Bayesian Entity Resolution. <i>Wiley StatsRef</i> (to appear)	

- **Binette, O.** and Steorts, R. C. (2021) (Almost) All of Entity Resolution. *Science Advances* 8 (12) eabi8021

Research Assistant | 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

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 | Duke University

2020 – 2021

- Lead weekly labs for *Bayesian and Modern Statistics*, *Entity Resolution*, and *Introduction to Data Science*.

Instructional Teaching Assistant | Université du Québec à Montréal

2017 – 2019

- Lead 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
- 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: R, Python, C/C++, Java, bash, git, Linux, Docker, DevOps basics, Systems Engineering basics

R packages on Github: dgaFast, MSETools, assert, cache, fingermatchR, TessTools

Python packages on Github: PatentsView/PatentsView-Evaluation, StringCompare