

# Katherine Wasmer

📍 Kansas City Metropolitan Area (willing to relocate)

✉ wasmerkate@gmail.com

🌐 [linkedin.com/in/katherine-wasmer](https://www.linkedin.com/in/katherine-wasmer)

🔄 [github.com/KatherineWasmer](https://github.com/KatherineWasmer)

🔗 [katherinewasmer.github.io](https://katherinewasmer.github.io)

## SUMMARY

---

Data scientist with 6+ years of experience in mathematical and scientific computing. Implemented a Bayesian regression method to predict cross-population polygenic risk scores for autism, obtaining a mean AUROC 0.541, aligning closely with previous research.

## EDUCATION

---

**University of Michigan** Ann Arbor, MI

Aug 2025

M.S., Data Science

*Relevant Coursework:* Statistical Methods for Human Genetics, Computing with Big Data, Data Manipulation & Analysis, Data Structures & Algorithms

*Advisor:* Alexander Tsodikov, Ph.D.

**University of Iowa** Iowa City, IA

May 2021

B.S., Mathematics

*Relevant Coursework:* Introduction to Mathematical Biology, Research Data Management, Optimization Techniques, Introduction to Biostatistics, Fundamentals of Public Health

*Advisor:* Oguz Durumeric, Ph.D.

## SKILLS

---

- **Bioinformatics:** bcftools, Bioconductor, Galaxy server, PLINK
- **Machine Learning:** PyTorch, Scikit-Learn, TensorFlow, VIPRS (variational inference)
- **Tools:** Git, Google Colab, Huggingface, Linux/UNIX, SLURM, Unslloth
- **Programming:** Bash (shell scripting), C++, Java, Python, SQL

## RESEARCH EXPERIENCE

---

**Polygenic Risk Scoring of Autism Across Populations**

Ann Arbor, MI

Capstone Project, Data Science MS

May — Aug 2025

*Supervising Professor:* Jonathan Terhorst, Ph.D.

- Calculated polygenic risk scores for 320 individuals of Japanese, Saudi, or admixed American ancestry with known ASD diagnosis status (119 cases, 201 controls); implemented a variational inference framework with VIPRS.
- Obtained an AUROC of 0.541 across all populations after 5-fold cross validation and hyperparameter tuning; scores below 0.5 in the stratified groups revealed limitations of PRS for predicting autism in cross-population cohorts.
- Developed a strong understanding of human research ethics and protocols through completing HIPAA modules and submitting an IRB application for genomic data.

- Integrated Great Lakes HPC, Galaxy Bioinformatics, and parallel computing for optimizing large-scale genomic data analyses and transformations.

**Estimating the Causal Effect of Low Birth Weight on Infant Mortality**

Ann Arbor, MI

Causality and Machine Learning (Final Project)

Oct — Dec 2024

*Professor:* Maggie Makar, Ph.D.

- Built a causal inference pipeline to estimate the causal average treatment effect (CATE) of low birth weight (LBW) among same sex twins, applying an R-Learner and Double Machine Learning (DML) framework.
- Processed 2023 infant data from the National Center for Statistics with the pandas library in Python, obtaining a sample of  $n = 1210$  twins.
- Combined R-Learner and DML with a histogram gradient boosting regressor (scikit-learn) to estimate a CATE of 0.01, suggesting a slight causal effect of low birth weight on infant mortality and aligning with previous research on LBW.

**Dept. of Child Psychiatry (University of Kansas Medical Center)**

Kansas City, KS

Summer Research Intern

Jun — Aug 2024

*Supervisor:* Rene Jamison, Ph.D.

- Gauged the efficacy of a 12 week social program for girls with autism by regressing post-intervention quality of life (QoL) scores on pre-intervention scores; reported a strong positive correlation ( $r = 0.83-0.92$ ) in QoL improvement among an aggregate of 48 participants.
- Applied Welch's t-test on sex-stratified data from KUMC autism evaluations, revealing a statistically significant increase in anxiety ( $p = 0.01$ ) and social withdrawal ( $p = 0.001$ ) among adolescent girls diagnosed with ASD; contributed findings to an in-draft manuscript on the social and emotional health of girls with autism.
- Processed and transformed SPSS survey data with pandas, maintained an organized workflow for computation in VisualStudio Code, and visualized findings with matplotlib and seaborn for interpretable results.

**Predicting Cancer Survival Outcomes With Spatial Proteomics Data**

Ann Arbor, MI

Machine Learning for Health Sciences (Final Project)

Mar — Apr 2024

*Professor:* Xiang Zhou, Ph.D.

- Collaborated with two other graduate students to predict survival outcomes for 281 patients based on cell image data and average image intensities provided by the professor.
- Conducted an elastic net regression on the training dataset ( $n = 225$ ) using scikit-learn and implemented 10-fold cross-validation to improve the model's accuracy.
- Integrated Python and R to compute an average RMSE of approximately 58 months. This large error was likely due to the small sample size or a consequence of using newly published cell image data.

**Michaelson Lab (associated with SPARK for Autism)**

Iowa City, IA

Undergraduate Research Assistant

Oct — Dec 2020

*Supervisors:* Jacob Michaelson, Ph.D.; Taylor Thomas, Ph.D. Candidate

- Programmed a two-sample t-test in R with behavioral survey data, which yielded statistically significant evidence of sex differences among children with autism.
- Calculated the heritability of sex-based behavioral differences in individuals with autism by using data from a GWAS previously conducted by SPARK.

- Conducted an independent component analysis (ICA) on biomedical survey data that showed a correlation between GI problems, sensory issues, and atypical eating habits.

### **Teaching TDA Mapper Through Interactive Code**

Undergraduate Research Assistant

*Supervisors:* Isabel Darcy, Ph.D.

Iowa City, IA  
Sep 2019 — Aug 2020

- Developed a Swirl course that encompassed the fundamentals of topological data analysis and computational analysis with TDAmapper.
- Instructed other undergraduate research assistants on how to implement RStudio, Swirlify, and Swirl for designing their own lessons and courses.
- Created a lesson on performing a PCA with TDAmapper and how to visualize the results with accessible libraries such as ggplot2, scatterplot3d, and viridis.

## **PROFESSIONAL EXPERIENCE**

---

### **Shasta QA**

Automation Engineer

Iowa City, IA  
Oct 2021 — Oct 2022

- Developed a comprehensive test automation framework utilizing Selenium with Java and C for client projects; contributed to CI/CD code repositories with peer review processes across teams of 2-7 QA engineers, ensuring clean, testable, and readable code.
- Orchestrated an API regression testing suite on the IntelliJ platform using AWS infrastructure, reducing cloud computing costs through optimized test scheduling.
- Collaborated with cross-functional development teams on weekly specification reviews, translating requirements into automated test cases; reported 200+ bugs via JIRA and TestRail, improving the overall quality of client software.