

Anna Neufeld

Contact Information	Public Health Sciences Division Fred Hutchinson Cancer Center 1100 Fairview Ave. N., M1-B514 Seattle, WA, 98109	aneufeld@fredhutch.org https://anna-neufeld.github.io 781-392-4894
Employment	Assistant Professor of Statistics Department of Mathematics and Statistics Williams College, Williamstown, MA	2024 - Present
	Postdoctoral researcher Biostatistics Program (Public Health Sciences Division) Fred Hutchinson Cancer Center, Seattle, WA <ul style="list-style-type: none">• Advisor: Dr. Jeffery Leek	2023-2024
Education	University of Washington , Seattle, Washington PhD in Statistics <ul style="list-style-type: none">• Advisor: Dr. Daniela Witten• Thesis: Addressing double dipping via selective inference and data thinning Williams College , Williamstown, MA BA in Mathematics (Highest Honors) and Computer Science. <i>Summa Cum Laude</i> <ul style="list-style-type: none">• Advisor: Dr. Brianna Heggeseth• Thesis: Regression trees for longitudinal data	2018-2023 2014-2018
Teaching Experience	Visiting Lecturer Williams College Department of Computer Science <ul style="list-style-type: none">• Co-instructed CS 374T, Machine Learning, with Professor Andrea Danyluk• Tutorial style course (modeled after Oxford University tutorials), which involves meeting with two students at a time. Ten total students enrolled in the course. Instructor of Record University of Washington Department of Statistics <ul style="list-style-type: none">• Stat 311: Elements of Statistical Methods• Fully responsible for course with 60 students. Head Teaching Assistant University of Washington Department of Statistics <ul style="list-style-type: none">• Stat 311: Elements of Statistical Methods• Developed lab assignments, maintained course website, helped write assignments and exams, served as liaison between professor and other TAs. There were 180 total students in the course.• Led two lab sections of 30 undergraduates each. Graduate Teaching Assistant University of Washington Department of Statistics <ul style="list-style-type: none">• Stat 570: Regression Methods for Independent Data (Autumn 2021)	Spring 2021 Summer 2019 Autumn 2019, Winter 2020 2018-2021

- Stat 527: Nonparametric Regression (Spring 2021)
- Stat 311: Elements of Statistical Methods (Autumn 2018, Spring 2020)
- Stat 423: Applied Regression and Analysis of Variance (Winter 2019)
- CSE/Stat 416: Introduction to Machine Learning (Spring 2019)
- Responsible for lab sections of around 30 undergraduates each. Also held office hours and graded assignments. Courses listed above range from undergraduate introductory courses (Stat 311 and Stat 416) to core courses for the statistics PhD program (Stat 570).

Undergraduate Teaching Assistant September 2015 - May 2018

Williams College Departments of Computer Science, Mathematics, and Statistics

- Data Structures and Advanced Programming (Fall 2015), Linear Algebra (Spring 2016, Spring 2017), Abstract Algebra (Fall 2016), Regression and Forecasting (Fall 2017, Spring 2018)
- Duties included grading homework, holding office hours, and running review sessions.

Peer Tutor and Workshop Leader 2016-2018

Williams College Office of Academic Resources

- Held one-on-one and drop-in tutoring sessions for microeconomics, macroeconomics, calculus, linear algebra, real analysis, statistics, and computer science.
- Assisted biology research students with data analyses in R.
- Developed and taught a series of workshops in R for students from a variety of departments as part of a pilot for a winter study coding workshop program.

Preprints

Anna Neufeld, Joshua Popp, Lucy L. Gao, Alexis Battle, and Daniela Witten (2023) Negative binomial count splitting for single-cell RNA sequencing data. <https://arxiv.org/pdf/2307.12985>.

Publications

Ameer Dharamshi, **Anna Neufeld**, Keshav Motwani, Lucy L. Gao, Daniela Witten, and Jacob Bien (2024) Generalized data thinning using sufficient statistics. *Journal of the American Statistical Association*. *SLDS student paper award awarded to Ameer Dharamshi at JSM, 2024*.

Anna Neufeld, Ameer Dharamshi, Lucy L. Gao, and Daniela Witten (2024) Data thinning for convolution-closed distributions. *Journal of Machine Learning Research*.

Anna Neufeld, Lucy L. Gao, Joshua Popp, Alexis Battle, and Daniela Witten (2022) Inference after latent variable estimation for single cell RNA-sequencing data. *Biostatistics*. *Winner of best student paper award at WNAR, 2022*.

Anna C. Neufeld, Lucy L. Gao, and Daniela M. Witten (2022) Tree-Values: selective inference for regression trees. *Journal of Machine Learning Research*.

Anna Neufeld and Daniela Witten (2021). Discussion of Breimans Two Cultures: From Two Cultures to One. *Observational Studies*.

Maxian, O*, **Neufeld, A.***, Talis, E. J.*, Childs, L. M., & Blackwood, J. C. (2017). Zika virus dynamics: when does sexual transmission matter?. *Epidemics*.
(* denotes equal contribution)

Software

datathin: splitting a random variable into independent training and test components. R package available on [github](#), with tutorials available on our [website](#).

countsplit: splitting integer-valued count matrices. R package available on [CRAN](#), with tutorials available on our [website](#).

treevalues: selective inference for regression trees. R package available on [github](#), with tutorials available on our [website](#).

splinetree: longitudinal trees and forests using a spline projection method. R package available on [CRAN](#). Tutorials available on our [website](#).

Invited Presentations

- ASA Section on Statistical Genetics and Genomics monthly webinar, April 2024, virtual. *Data thinning to avoid double dipping*.
- Banff International Research Station Workshop on Statistical Aspects of Trustworthy Machine Learning. February, 2024. Banff, AB. *Data thinning to avoid double dipping*. [\[slides\]](#).
- University of Washington Department of Genome Sciences and Computational Molecular Biology Program Combi seminar, January 2024, Seattle, WA. *Avoiding double dipping in the analysis of scRNA-seq data*.
- Channing Division Methods Meeting, Brigham and Women's Hospital and Harvard Medical School, December 2023, virtual. *Avoiding double dipping in the analysis of scRNA-seq data*.
- Williams College Statistics Colloquium, November 2023, Williamstown, MA. *Data thinning to avoid double dipping*. *Data thinning to avoid double dipping*. [\[slides\]](#).
- Reed College Mathematics Colloquium, October 2023, Portland, OR. *Data thinning to avoid double dipping*.
- North American Machine Learning, Optimization, and Statistics Symposium (NAMOS), June 2023, Vancouver, BC. *Data thinning*.
- Private Brands Casual Science Seminar at Amazon, April 2023, virtual. *Data thinning*. [\[slides\]](#).
- Fred Hutchinson Cancer Center, group of Dr. Mike Wu, April 2023, virtual. *Data thinning*.
- Electronic Undergraduate Statistics Research Conference, November 2022, virtual. Panelist for graduate school information panel.
- Williams College Statistics Colloquium, October 2022, virtual. *Avoiding “double dipping” in the analysis of single cell RNA-sequencing data*.
- Joint Statistical Meetings, August 2022. Panelist for: *Leo Breiman's Two Cultures: Introspection, Debate, and Discussion 20 Years Later*.
- International Seminar on Selective Inference, June 2022, virtual. *Inference after latent variable estimation for single cell RNA-sequencing data*. [\[recording\]](#), [\[slides\]](#).
- Boston Children's Hospital Vascular Anomalies Group Meeting, June 2022, virtual. *Statistical issues when analyzing single cell RNA-sequencing data*.

Contributed Presentations

- WNAR 2022 Student Paper Competition Session, June 2022, virtual. *Inference after latent variable estimation for single cell RNA-sequencing data*.

Mentoring

NSF INSPIRE U2 REU, Spelman College

Summer 2023

(Increasing Statistical Preparation in Research Education for Underrepresented Undergraduates)

- Virtual RStudio mentor for two undergraduate researchers; Allyanna Lewis and Sasha Villefranche. Both students will be presenting at the Annual Biomedical Research Conference for Minoritized Scientists (ABRCMS).

University of Washington Statistics Directed Reading Program (DRP)

- Co-founded a program that pairs undergraduates with PhD student mentors for independent studies. Modeled after successful [Directed Reading Programs](#) (DRPs) in mathematics departments at several universities. More information can be found at [our website](#).
- Served as the graduate student coordinator from 2020-2023. Managed admissions, recruitment, and scheduling for 95 undergraduate independent studies. Several undergraduates have gone on to graduate school in statistics or related fields.
- Served as the instructor of record for Stat 499 (the associated course that undergraduate participants can register for) from 2021-2023.
- Served as a mentor for the following eight undergraduates:
 - Winter 2020, Christina Nick, Statistical Natural Language Processing
 - Spring 2020, Rachael Ren, Infectious Disease Modeling with Differential Equations. See [project writeup](#) and [presentation](#). Now a PhD student in statistics at the University of Texas.
 - Autumn 2020, Harper Zhu, Infectious Disease Modeling on a Network. See [presentation](#) and [shiny app](#). Now a MS student in data science at Cornell Tech.
 - Spring 2021, Kayla Kenyon, Infectious Disease Modeling. See [presentation](#).
 - Autumn 2021, Cathy Qi, Multiple Testing.
 - Winter 2022, Hisham Bhatti, Statistical Concepts in Clinical Trials.
 - Spring 2022, Wei Jun Tan [[presentation](#)] and Iris Zhao [[presentation](#)], Introduction to Computational Biology, co-mentored with Alan Min.

Professional Service

Manuscript reviewer: *Statistical Science*, *Journal of Computational and Graphical Statistics*, *Japanese Journal of Statistics and Data Science*, *Electronic Journal of Statistics*, *Biometrics*, *Bernoulli*, and *Genetic Epidemiology*.

University Service

University of Washington

- Lead consultant and volunteer recruitment coordinator for UW StatCom (statistics in the community non-profit consulting program), 2022-2023.
- Member of UW Statistics Undergraduate Curriculum Committee, 2021-2022.
- UW Statistics Graduate Student Representative (elected and paid position that involves attending faculty meetings, serving on committees, and planning orientation for new students). 2020-2021.
- Student interviewer for tenure track assistant professor search (2020) and teaching professor search (2021).
- First round reviewer for PhD Admissions, 2020-2023.
- UW Statistics Diversity, Inclusion, Community, and Equity (DICE) Committee, 2019-2023.

- Chair of the UW Statistics Fun Committee, 2019-2020.
- Organizer and Founder: [Statistics Education Reading Group](#), 2019-2020.

Williams College

- Member of the Committee on Undergraduate Life (committee consisted of students, faculty, and staff), 2015-2016.
- Member of the Great Ideas Committee (managed a budget of \$10,000 per year to implement improvements to undergraduate life), 2015-2017.

Outreach

- Lesson leader at [UW Math Day](#), March, 2023.
- Lesson leader at [UW GEAR UP's](#) (Gaining Early Awareness and Readiness of Undergraduate Education) after school program (May 2021) and statistics summer camp (July 2022).

Other Experience

SMALL Research Experience for Undergraduates June 2016-August 2016

Williams College

- Advisors: Dr. Julie Blackwood and Dr. Lauren Childs
- Worked with a team of undergraduates in applied mathematics to model the dynamics of Zika Virus. Work resulted in co-first-author publication.