

Sara K. Colando

scolando.github.io/
scolando@andrew.cmu.edu
Last Update: August 2025

EDUCATION

Ph.D. in Statistics, Carnegie Mellon University, Pittsburgh, PA

2024 - 2029 (*Expected*)

B.A. in Mathematics and Philosophy, Pomona College, Claremont, CA

2024

- Graduated *magna cum laude* and with departmental honors; GPA: 3.99/4.00

PUBLICATIONS AND PAPERS

- Peer-reviewed Articles

2. Colando, S., Schulz, D., Hardin, J. **Selecting ChIP-Seq Normalization Methods from the Perspective of their Technical Conditions**, *Briefings in Bioinformatics*, 2025.
<https://doi.org/10.1093/bib/bbaf431>
 1. Colando, S. & Hardin, J. **Philosophy within Data Science Ethics Courses**, *Journal of Statistics and Data Science Education*, 2024. <https://doi.org/10.1080/26939169.2024.2394542>

- In Progress

1. Colando, S., Franke E., Reinhart A. **Analyzing Statistics Students' Writing Before and After the Emergence of Large Language Models.** In progress.

RESEARCH EXPERIENCE

Carnegie Mellon University Department of Statistics and Data Science

Advanced Data Analysis Project with Dr. Nynke Niezink and Dr. Eva Jaspers

Jan 2025 - Present

- Constructing a model for the conditional probability of a student avoiding their classmate, leveraging the positive (e.g. friendship) network between classmates
 - Communicating model choices and results between statisticians and sociologists
 - Currently, working to implementing a hierarchical Bayesian Exponential Random Graph Model (ERGM) to account for between-classroom heterogeneity

Trajectory of Statistics Students' Writing since LLMs Project with Erin Franke

Jan 2025 - Present

- Conducting corpus-based analysis of student statistics reports (2021–2025) to evaluate systematic changes in writing style before and after the emergence of LLMs
 - Applying linguistic analysis techniques (lemma frequency extraction, concordance analysis, and keyness tests) to quantify changes in students' usage for ChatGPT's most frequently-used words
 - Presented findings at the 2025 United States Conference on Teaching Statistics Research Satellite Session and currently drafting a manuscript on our results

Research Assistant to Dr. Ron Yurko

May - Aug 2023

- Performed mixture-model clustering analysis to discover trends in movement patterns for injured racehorses
- Identifying horses who under-raced through residual analysis of a created expected race count model
- Engaged in 8 weeks of lectures about popular statistical analysis techniques, data engineering, and novel data science methodologies employed in sports analytics

Pomona College Department of Mathematics and Statistics

Research Assistant to Dr. Johanna Hardin

May 2022 – Aug 2025

- Analyzed ChIP-Seq normalization techniques by their underlying technical assumptions to improve the selection of differential-binding analysis methods in practice
- Created reproducible ChIP-Seq read count simulations to assess how violating the normalization methods' technical conditions impact the average false discovery rate and power of identifying genomic regions with differential DNA occupancy between experimental states

Data Science Ethics Independent Project (Supervisor: Dr. Johanna Hardin)

May – Oct 2023

- Synthesized resources and pedagogies of existing Data Science Ethics undergraduate courses to connect key ethical concepts to data science practices
- Created a website to effectively communicate data science ethics concepts and relevant background information to a broad community of decision-makers
- Aided in Pomona College's development of an ethics course for the data science minor

University of California, San Diego Halıcıoğlu Data Science Institute

Intelligence, Data, Ethics, and Society (IDEAS) Summer Institute Participant

Aug 2023

- Participated in workshops taught by experts in domains like data science, AI, philosophy, and law to improve critical and interdisciplinary thinking on data science and AI ethics topics
- Collaborated with three other students on a two-week research project, examining practical methods of implementing responsible data science and AI practices in a California wildfire resource allocation algorithm

University of Michigan School of Public Health

Big Data Summer Institute Research Assistant to Dr. Nikola Banovic

June - Aug 2022

- Collaborated with three other students to implement a partially Bayesian neural network to quantify and communicate prediction uncertainty in a supervised brain tumor segmentation model via Python's PyTorch and TensorFlow packages
- Collaborated with three other students on a two-week research project, examining practical methods of implementing responsible data science and AI practices in a California wildfire resource allocation algorithm

TEACHING EXPERIENCE

Carnegie Mellon Department of Statistics and Data Science

Graduate Teaching Assistant

Aug 2024 – Present

- 36-707: Regression Analysis (Fall 2025)
- Carnegie Mellon Sports Analytics Camp, co-taught four lectures (Summer 2025)
- 36-226: Introduction to Statistical Inference, *Head TA* (Spring 2025)
- 36-600: Overview of Statistical Learning and Models (Fall 2024)

Pomona College

Course Mentor and Grader

Jan 2021 – May 2024

- MATH158: Statistical Linear Models (Spring 2024)
- MATH058B: Introduction to Biostatistics (Spring 2023)
- MATH060: Linear Algebra (Fall 2021)
- MATH030: Calculus I, Grader Only (Fall 2022)
- ID001: First-Year Interdisciplinary Seminar: I Disagree (Fall 2023)
- BIOL040: Introductory Genetics with Lab (Fall 2022)

1-2-1 Summer Bridge Program Teaching Assistant

- Led 1.5-hour meetings, twice per week, for three small cohorts of incoming Pomona College students
- Organized group bonding by preparing icebreakers and virtual games for 40+ students in the program
- Mathematical concepts taught: logic, statistics, calculus I and II, combinatorics, and number theory

PRESENTATIONS

13. **Colando, S.** & Franke, E. "Analyzing Statistics Students' Writing Before and After the Emergence of Large Language Models" (poster), United States Conference on Teaching Statistics, July 2025
12. Franke, E. & **Colando, S.** "Teaching Data Cleaning and Wrangling with R's `data.table` Package" (poster), United States Conference on Teaching Statistics, July 2025
11. **Colando, S.** "Predicting Avoidance Ties within Social Networks (ADA Progress Update)", Networkshop Research Group, Carnegie Mellon University, March 2025
10. **Colando S.** & Franke, E. "Computational Efficiency of R's `data.table` Package", StatBytes, Carnegie Mellon University, March 2025
9. **Colando, S.** & Hardin, J. "Philosophy Within Data Science Ethics Courses", CAUSE Webinar Series, October 2024
8. **Colando, S.**, Pipping, J., Wilson K. "Clustering Race Horse Movement Profiles to Discover Trends in Injured Horses", Carnegie Mellon Sports Analytics Conference 2023, November 2023
7. **Colando, S.** "Analyzing Data Science Ethics Pedagogies" (poster), Claremont Center for Mathematical Sciences, September 2023

6. **Colando, S.** "Analyzing Data Science Ethics Pedagogies" (poster), Intensive Summer Experience Symposium, September 2023
5. **Colando, S.**, Pipping, J., Wilson K. "Clustering Race Horse Movement Profiles to Discover Trends in Injured Horses", Summer Undergraduate Research Experience 2023 Project Showcase, July 2023
4. **Colando, S.** Panelist for "Pomona Funded Summer Experiences Informational Session", Family Weekend, October 2022
3. **Colando, S.** "Analyzing ChIP-Seq Normalization Techniques through the Lens of their Biological Assumptions" (poster), Intensive Summer Experience Symposium, September 2022
2. **Colando, S.** "Analyzing ChIP-Seq Normalization Techniques through the Lens of their Biological Assumptions" (poster), Claremont Center for Mathematical Sciences, September 2022
1. Chu, C., **Colando, S.**, Nandi, D., Serrano, X. "Quantifying Uncertainty in a Tumor Segmentation Model" (poster & presentation), A Symposium on Big Data, Human Health, and Statistics. July 2022

GRANTS

Carnegie Mellon University, Institute of Complex Social Dynamics	2025-2027
Title: "Graduate Student Fellowship"	
Role: Recipient	\$6,000/year
 Carnegie Mellon University, Department of Statistics and Data Science	2025
Title: "StatBytes Travel Grant"	
Role: Co-recipient	\$500
 National Science Foundation (Award Abstract: #2303612)	2025
Title: data.table Ambassador Travel Grant	
Role: Co-recipient	\$2,700

LEADERSHIP AND SERVICE

Carnegie Mellon University Department of Statistics and Data Science

TeachStat Working Group
 Mentorship Committee
 Mentor for Undergraduate Statistics Students

Pomona College Department of Philosophy

Search and Selection Student Representative for Tenure-track Faculty Position

Pomona-Pitzer Athletics

Varsity Women's Cross Country Team Captain

HONORS AND AWARDS

Passed First-year Ph.D. Data Analysis Exam with Distinction	2025
Blair Nixon Award, Pomona College	2024
Bruce Jay Levi Senior Prize, Department of Mathematics and Statistics, Pomona College	2024
W.T. Jones Prize in Philosophy, Department of Philosophy, Pomona College	2024
Phi Beta Kappa, Southern California Chapter	2024
Sigma Xi, Claremont Colleges Chapter	2024
Pomona College Scholar, Pomona College	2020-2024
Women's Cross Country and Track and Field Varsity Letter Recipient, Pomona-Pitzer Athletics	2020-2024