Sam Schwartz

Ph.D. Candidate

CV Last Updated September 2023

430 W. 12th Ave., Apt. #1.
Eugene, OR 97401
United States

☎ +1 801 739 3520
✉ sam@cs.uoregon.edu
schwartzstuff.com

Sam in a Nutshell for Busy Search Committees

- **Education**: PhD candidate in computer science at the University of Oregon, graduating spring 2024. Dual undergraduate major in mathematics and computer science with minors in Spanish and organizational communication at Utah State University. Graduated cum laude. Also graduated with a master's degree in mathematics from Utah State University.

- **Teaching**: Loves teaching undergraduates. Student course evaluations indicate undergraduates love Sam back. Has experience teaching multiple classes in math and computer science as full instructor of record. Made it a point to be the graduate assistant for as many different math and computer science classes as possible. Ready to teach a full range of undergraduate classes in math, statistics, data science and computer science. Ready to teach master's-level classes in data science and computer science. Also taught high school English in Chile before graduate school.

- **Work in Industry**: Spent the last five summers working with a variety of scientists in multiple national laboratories. Before that, worked with dozens of start ups and small businesses in getting their websites and servers set up, and also spent a summer at Amazon.

- **Research**: Interdisciplinary track record, with an overarching theme of using machine learning (ML) and data science techniques to answer domain-specific research questions. Domains include his MS thesis in math (Can we use ML to predict the solution to this NP-hard problem?), scientific visualization – especially particle advection, student sense of belonging, ecology, supercomputing memory management, and more. His PhD dissertation is focused on software development in research institutions (like national laboratories) and doing empirical mining and analysis on their software repositories. This work allows policymakers the information and justification to dedicate money towards things like software sustainability. It also lays the groundwork for the empirical understanding of the career arcs of research software engineers, as well as curated data collection for machine learning tools in this space.

- **Service**: Enjoys committee work and volunteers for it. Strong believer in shared governance.

- **Mentorship**: Made it a point as a second-year PhD student to help mentor an undergraduate through a multi-month thesis-like experience. Regrets that time and resource constraints in the PhD program hasn't allowed for more 1-on-1 undergraduate thesis mentorship.

- **Languages and Skills**: Speaks Spanish and has familiarity with a smorgasboard of various programming languages and frameworks.

Contents

Education

Expected Graduation Spring 2024 — **Ph.D. in Computer Science**, *University of Oregon*, Eugene, Oregon.
Research focus in tools and software engineering processes for advanced scienctific computing.
Advisors: Stephen Fickas, Ph.D. and Boyana Norris, Ph.D.

Graduated Jun. 2019 — **Master of Science in Mathematics**, *Utah State University*, Logan, Utah.
Research focus in machine learning techniques for non-numerical training sets.
Advisor: David E. Brown, Ph.D.

Graduated Dec. 2016 — **Bachelor of Science in Computer Science**, *Utah State University*, Cum Laude.
Five-time "Dean's List" recipient. Departmental honors student.

Graduated Dec. 2016 — **Bachelor of Science in Mathematics**, *Utah State University*, Cum Laude.
Minors in Spanish and Organizational Communication.

Completed Aug. 2014 — **Study Abroad**, *Pontificia Universidad Católica de Valparaíso*, Valparaíso, Chile.
Study abroad semester with a focus on Latin American history and Chilean politics.
Conversationally proficient in Spanish due to the immersion experience.

Work Experience: Academic and Pedagogical

Oct. 2019 – Present — **Graduate Employee**, *University of Oregon*, Eugene, Oregon.
In this role, my responsibilities include research activites – full time since Spring Term 2022 – and the teaching of undergraduate computer science labs and classes. I have been assigned as the instructor of record or laboratory instructor for the following courses, ordered by course number:

- CIS 122 - Intro to Programming and Problem Solving* . . . . . Spring 2020
- CIS 210 - Computer Science I* . . . . . Fall 2019, Winter 2020
- CIS 211 - Computer Science II* . . . . . Winter 2021
- CIT 281 - Web Applications Development I* . . . . . Spring 2021
- CIS 510 - Game Programming† . . . . . Spring 2022
- Undergraduate Python Bootcamp . . . . . Fall 2020, Fall 2021

The primary objective of this two-week bootcamp was to ensure that transfer students who have experience with an object oriented programming language are able to use Python. Python is the lingua franca used in upper division classes at the University of Oregon. The target audience of the bootcamp was transfer students. However, any interested undergraduate was welcome to attend. The lecture materials, homework, course objectives, and overall curriculum were all created solely by me.

* Lab Instructor, † Graduate Teaching Employee

Additionally, in summer 2021 I was hired by the director of the data science program to create all projects for a forthcoming capstone data science class, to be first taught in January 2022. These projects include practical applications of linear regression, neural networks, random forests, and other key machine learning techniques.

Aug. 2017 – May 2019 **Graduate Student Instructor**, *Utah State University*, Logan, Utah.
In this role, my responsibilities include researching the mathematics of machine learning techniques and teaching undergraduate mathematics classes.
I have been assigned as the instructor of record, recitation leader, or graduate teaching assistant for the following courses, ordered by course number:

- MATH 0995 - College Mathematics Preparation . . . . . Summer 2018
- MATH 1051 - Classical Algebra for Teachers . . . . . Spring 2018
- MATH 1060 - Trigonometry . . . . . Fall 2018, Spring 2019
- MATH 1100 - Calculus Techniques* . . . . . Fall 2017
- MATH 2270 - Linear Algebra* . . . . . Spring 2019
- MATH 3310 - Discrete Mathematics* . . . . . Fall 2018
- STAT 6910 - Neural Networks† . . . . . Spring 2019

* Recitation Leader, † Graduate Teaching Assistant

Mar. 2017 – Jul. 2017 **English & Technology Instructor**, *Chilean Ministry of Education*, Chillán, Chile.
One of the highlights of my professional and personal life was the opportunity I had to teach English as a second language for an academic semester in a Chilean high school. This volunteer experience was paid for via sponsorship from the Government of Chile and the United Nations. In this capacity, I also ran an extracurricular club called the *Technical English Club* to facilitate learning English based programming languages for students whose primary language was Spanish. I also facilitated a two-week robotics summer camp through sponsorship from Microsoft, Telefónica, and the Government of Chile.

Work Experience: Software Development Roles in Industry

May. 2023 – Aug. 2023 **Software Development Summer Intern**, *Argonne National Laboratory (ANL)*, Greater Chicago Area, Illinois.
Working with a software development team focused on multi-physics spacetime code development as a case study for the creation of software engineering tools for broader, cross-cutting advanced scientific computing contexts.

Jun. 2022 – Sep. 2022 **Software Development Summer Intern**, *Lawrence Livermore National Laboratory (LLNL)*, Remote; Eugene, Oregon.
Worked with the Umpire team at LLNL to develop a prototype machine learning tool for runtime memory managment and allocation decision making in cross-platform high performance computing contexts.

Jun. 2021 – Sep. 2021 **Software Development Summer Student**, *Department of Energy Exascale Computing Project vis-à-vis the University of Oregon*, Eugene, Oregon.
This project involved integrating several complex scientific visualization libraries designed to run on supercomputers into an existing simulation software package. The underlying simulation is used to understand the behavior of electricity generating windmill farms under a variety of weather conditions.

Jun. 2020 – Sep. 2020 **Software Development Summer Intern**, *Oak Ridge National Laboratory (ORNL)*, Remote*; Eugene, Oregon.
This internship focused on developing optimization techniques which utilize machine learning to profile compute-intensive visualization algorithms on supercomputers and subsequently maximize their performance. Data was drawn from Summit, the world's fastest supercomputer at the time.
**Due to the COVID-19 pandemic, this internship was formally executed as a subcontract through the University of Oregon to allow for remote work within budget constraints.*

Jul. 2019 – Sep. 2019 **Software Development Summer Student**, *VTK-m Project vis-à-vis the University of Oregon*, Eugene, Oregon.
This project implemented several dozen mesh quality metrics into the VTK-m framework. A mesh, in this context, is the data structure which undergirds simulation data to be visualized in high performance computing settings. The project involved initial implementation in C/C++, unit testing, and integration with an established code base.

May 2016 – Aug. 2016 **Software Development Summer Intern**, *Amazon.com*, Seattle, Washington.
As a summer intern at Amazon, my responsibilities revolved around creating automated data cleansing software to integrate legacy marketing information into newer systems. This was done by creating an in-house dashboard to minimize manual data manipulation and cleaning.

Oct. 2013 – May 2016 **Software Development & IT Consultant**, *Small Business Administration - SBDC*, Logan, Utah.
The Small Business Development Center (SBDC) program of the Small Business Administration works with start-ups across the nation to cultivate ingenuity, secure bank loans, develop marketing strategies, and provide information about community resources.

My responsibilities included consulting with entrepreneurs about the latest market trends in technology, developing e-commerce websites, advising business owners on system architecture, and providing support for the variety of IT related concerns new companies have. Projects involved industries as diverse as retail to bio-fuel.

Dec. 2010 – Aug. 2013 **Internal Applications Programmer**, *Worker's Compensation Fund*, Sandy, Utah.
My primary project with WCF was developing security software which allowed the internal audit department to review building, network, claims, underwriting and subsidiary system access levels and history while simultaneously notifying system administrators of change in access requests from managers throughout the company. Other projects included developing internal account lookup and modification tools for their helpdesk.

Research Publications, Grants, and Presentations

2023 **Temporal Classification of Allocations for Reduced Memory Usage**, *Kristi Belcher, David Beckingsale, Samuel D. Schwartz, Marty McFadden*, Conference Poster, Supercomputing (In Review).
Note: This poster is a teaser for a forthcoming paper on which Sam is the first author.

2023 **How many of us are out there? An Inventory of Research and Research-Adjacent Open Source Software Development Activity**, *Samuel D. Schwartz, Stephen F. Fickas, Boyana Norris, Anshu Dubey*, Conference Presentation, Supercomputing, Workshop: RSE-HPC (In Review).

2023 **MeerCat: a Pull Request Assistant**, *Stephen Fickas, Samuel D. Schwartz*, Boyana Norris, Jason Prideaux, Pablo Flores, Anshu Dubey, Ellie Kobak, Jared O'Neal, Akash Dhruv*, Conference Presentation, Supercomputing, Workshop: RSE-HPC (In Review).

2023 **A Five Year Survey of Literature in Software Engineering and Repository Mining Research**, *Samuel D. Schwartz*, PhD Candidacy Qualifying Milestone.
Qualifier awarded "with distinction" status by department evaluators, the highest rating available. Historically awarded to the top ∼25% of PhD candidates.

2023 **A Systematic Comparison of Class Imbalance Methods and Machine Learning Algorithms on Species Distribution Model Performance**, *Donald J. Benkendorf, Samuel D. Schwartz, D. Richard Cutler, Charles P. Hawkins*, Journal Article, Ecological Modelling.

2023 **Technical Report on Grade Inflation and Geographic Distribution of Top Students at the University of Oregon**, *Samuel D. Schwartz*, Technical Report, Submitted to the University (Faculty) Senate and the Board of Trustees of the University of Oregon.

2022 **What Predicts Community in Virtual Computing Classrooms? A Machine Learning-Based Approach**, *Samuel D. Schwartz and Joanna Goode*, Conference Paper, RESPECT.

2021 **An Entropy-Based Approach for Identifying User-Preferred Camera Positions** , *Nicole Marsaglia, Yuya Kawakami*, Samuel D. Schwartz*, Stefan Fields, Hank Childs*, Conference Paper, LDAV.
** These authors contributed equally.*

2021 **Machine Learning Based Autotuning for Parallel Particle Advection**, *Samuel D. Schwartz, Hank Childs, David Pugmire*, Conference Paper, EGPGV.

2019 **Machine Learning Techniques Applied to Discrete & Combinatorial Structures**, *Samuel D. Schwartz*, MS Thesis, Utah State University.

2019 **Image Completion Techniques Applied to Matrix Factorization in the Boolean Semiring**, *Samuel D. Schwartz*, Conference Presentation, SIAM Wasatch Student Chapters Conference.

2019 **Software Development in Production Environments for Researchers**, *Samuel D. Schwartz*, Invited Talk, Data Analytics Club; Hosted for and by PhD Candidates, Utah State University.

2018 **Modified Kernel Principal Component Analysis Results on NP-Class Problems in Digraphs**, *Samuel D. Schwartz and David E. Brown*, Technical Report.

2018 **How to Type Math: [Teaching] LaTeX** , *Samuel D. Schwartz and Kaitlin S. Murphy*, Conference Presentation, Together we Teach (TweT).

2015 **Analysis of Statistical Components of Tournaments Modeling Patterns in Voting and Social Choice Theory.**, *Samuel D. Schwartz*, Grant, US$1,000, College of Science Undergraduate Minigrant Program, Utah State University.

Academic and Community Service

Summer 2023 – Present **Board Member; State of Oregon Electronic Government Portal Advisory Board (EPAB)**, Appointed by Governor Tina Kotek to the EPAB, which advises the State Chief Information Officer of the State of Oregon about the state's web services, websites and e-commerce. The board also holds binding votes on end user transaction fees on state websites, which provide a multi-million dollar revenue stream for the state to pay for the hosting and development of web services. Began board service with two state senators, two state representatives, one university president, the president of the League of Women Voters, and other senior civil servants as fellow board members.

Fall 2021 – Fall 2023 **Graduate Representative; Computer Science Graduate Education Committee**, Represented the graduate student voice in the University of Oregon's computer science department's monthly faculty committee meeting on graduate education affairs.

Fall 2022 – Spring 2023 **Guest Participant; Tuition and Fee Advisory Board (TFAB)**, Full participant in the dozen+ 8:15am meetings of the board which put together the tuition and fee proposals for all programs of the University of Oregon. Wrote a 28 page minority report submitted to the president and Board of Trustees. The report concurred with undergraduate tuition proposals and some graduate program tuition proposals; dissented with School of Law proposals. It also provided analysis of the overall buisness model of the university and provided suggestions for the tuition setting process going forward.

Spring 2022 & Spring 2021 **Visit Day Committee**, Facilitated recruitment for new graduate students to the Department of Computer and Information Science, University of Oregon.

Spring 2021 **Judge; High School and Middle School Science Fair – State Regionals**, Undergraduate Research Symposium., Central Western Oregon Science Expo (CWOSE).

Spring 2020 **Judge; Oral Defenses and Poster Presentations in STEM Fields.**, Undergraduate Research Symposium, University of Oregon.

Oct. 2019 – Mar. 2020 **Orchestrator; "Together We Teach" series**, Department of Computer and Information Science at the University of Oregon, Started a weekly meeting among graduate employees and faculty in the department to discuss ways to improve teaching in CS. Over 10 external speakers were invited to talk, and the series prompted the recreation of a for-credit graduate-level "Teaching Effectiveness" seminar the following year.

Nov. 2019 – Jun. 2020 **Member; Graduate Employee Professional Development Committee**, Worked with the Teaching Engagement Program, the Graduate School, the graduate employee labor union, human resources, and university information services to develop training, especially teaching training, for all graduate employees at the University of Oregon.

Sep. 2019 – Jun. 2021 **Steward (Department-Level Elected Graduate Employee Labor Union Representative)**, Department of Computer and Information Science., Graduate Teaching Fellows Federation at the University of Oregon.

Spring 2019 **Judge; Oral Defenses and Poster Presentations in STEM Fields**, Undergraduate Research Symposium., Utah State University.

Spring 2019 **Instructor; Pro Bono Six Day Python Bootcamp for Graduate Students.**, Dept. Mathematics and Statistics, Utah State University.

Student Mentorship

2021 **Ben Backen**, *University of Oregon Undergraduate Honors Student*, Presented at the undergraduate research symposium in 2021 on predicting stock market fluctuations via machine learning using numerical and natual language data from Reddit. Working towards an undergraduate honors thesis in the areas of cybersecurity and machine learning.

Languages

English, Native.

Spanish, Conversationally Proficient. (Becoming less fluent with each passing month.)

Skills

Languages, technologies, and frameworks with which I have worked include:

- C/C++
- C#
- Git
- HTML
- Java
- Javascript
- Julia
- Keras
- LaTeX
- LDAP
- Linux
- Maple
- MySQL
- Numpy
- Pandas
- PHP
- Perl
- Python
- Pytorch
- Qualtrics
- Quickbooks
- R
- Sage
- TensorFlow
- VTK, VTK-m
- Visit

Appendix A: Research with Abstracts

2023 Conference Poster

Temporal Classification of Allocations for Reduced Memory Usage, *Kristi Belcher, David Beckingsale, Samuel D. Schwartz, Marty McFadden*, Supercomputing (In Review).

Abstract: Umpire, a data and memory management API created at LLNL, provides memory pools which enable less expensive ways to allocate large amounts of memory in HPC environments. Memory pools commonly contain both allocations that persist for only a portion of the program (temporary) and those that persist for the entire program (permanent). However, too much of a mix of both allocation types can lead to pool fragmentation and cause the pool to perform poorly. Umpire created a tool that uses a machine learning model to perform temporal classifications and categorize allocations as either temporary or permanent. We conducted experiments using trace files from two LLNL applications to study how much memory can be saved when those allocations are separated into distinct pools. We found that our ML tool accurately classifies memory allocations and that separating these allocation types into distinct pools reduces overall memory usage significantly (up to 29.5%).

Note: This poster is a teaser for a forthcoming paper on which Sam is the first author.

2023 Conference Presentation

How many of us are out there? An Inventory of Research and Research-Adjacent Open Source Software Development Activity, *Samuel D. Schwartz, Stephen F. Fickas, Boyana Norris, Anshu Dubey*, Supercomputing, Workshop: RSE-HPC (In Review).

Abstract:

In recent years, the engineering and sustainability of research software has emerged as an important element of scientific discovery. As more software projects move to the open source model at research institutions, particularly at the US national laboratories, a census of such projects can lead to many insights about the work being done and where gaps exist.

As an exploratory exercise in discovering all open-source research software projects that various national laboratories are engaged in, we inventoried all commits to GitHub stored in Google's BigQuery GitHub database. We then filtered and analyzed this data, which includes terabytes of commit data going back to the 1990s.

We found that there are collectively 3,124 distinct lab email addresses committing code to open-source repositories. We also found lab-affiliated committers have contributed to over 15,472 distinct open-source repositories, ranging from well-known projects like Docker or Linux, to the obscure. To get a better handle on how many of these repositories were lab projects, and not just large open-source projects which happened to have a lab employee contributing, we analyzed the percentage of lab committers on each project. After some preliminary histogram plotting, we chose 5% as a threshold. Using that threshold, we found 2,540 repositories (16.4%) had lab-affiliated emails contributing more than 5% of all commits.

Looking more broadly at all repositories where government employees contribute, we discovered that lab-affiliated emails compose 42.6% of all commits from accounts with a ".gov" email address, with accounts using a nasa.gov email address as the most common non-lab government entity contributing to open source code repositories in the United States.

We believe that these preliminary findings are of interest to the community, and the results invite questions about the role repository mining can play in empirically assessing the activities of research software engineers, their induction and persistence within the community over time, and so forth.

This work is part of the University of Oregon IDEAS project, which in turn is part of the larger Exascale Computing Project's Interoperable Design of Extreme-scale Application Software (ECP IDEAS) Productivity project. The research we report on has close collaboration with Argonne National Laboratory.

2023 Conference Presentation

MeerCat: a Pull Request Assistant, *Stephen Fickas, Samuel D. Schwartz*, Boyana Norris, Jason Prideaux, Pablo Flores, Anshu Dubey, Ellie Kobak, Jared O'Neal, Akash Dhruv*, Supercomputing, Workshop: RSE-HPC (In Review).

Abbreviated Abstract:

The idea of the Pull Request Assistant (PRA), and its requirements to be a useful tool, came from working with PR reviewers on GitHub. Given their limited time, reviewers rely on PR authors to give them clean and well-thought-out code. Unfortunately, this is often not the case. Reviewers may spend their time carrying on conversations not about substantive issues but instead about low-level problems with code, documentation and even testing. In essence, to help PR reviewers, we turned our attention to PR authors. The PRA attempts to work with the PR author to both diagnose low-level problems and aid in fixing those problems before a reviewer begins a review. Specifically, the analysis tools of the PRA are as follows: When a PR is created, MeerCat is called into service through a GitHub webhook. The MeerCat PRA focuses on 4 areas: Structured documentation, Code quality, Testing impacts, Robust discussion. In general, the PRA is a post-CI tool. Once mandated tests are passed, it attempts to further aid a developer in achieving a ready-to-review state by pointing out issues and offering suggestions. It also attempts to add other voices to the discussion as appropriate, enhancing the review process. In summary, the goal is a clean, well-thought-out PR that moves its way to formal review. **Primary presenter.*

2023 PhD Candidacy Qualifying Milestone

A Five Year Survey of Literature in Software Engineering and Repository Mining Research, *Samuel D. Schwartz*.

Abstract: This literature review surveys the extant research from 2018 through October 2022 in three key conferences related to software engineering and repository mining: Mining Software Repositories (MSR), the International Conference on Software Engineering (ICSE), and the International Conference on Software Maintenance and Evolution (ICSME). The review taxonifies hundreds of research papers, and provides summary and analysis of noteworthy papers related to code writing and refactoring, code comprehension, smells and code quality, aspects of software development related to an entire software project, and human dynamics within that project. The review also explores cross cutting themes among these topics, such as bots, machine learning, and the influence of different environments – such as free open source projects vs start ups vs established industry – on the development process. Underlying this work is an eye to research applicable to the advanced scientific computing space, which, in the United States, is often performed at national laboratories. *Qualifier awarded "with distinction" status by department evaluators, the highest rating available. Historically awarded to the top ∼25% of PhD candidates.*

2023 Journal Article

A Systematic Comparison of Class Imbalance Methods and Machine Learning Algorithms on Species Distribution Model Performance, *Donald J. Benkendorf, Samuel D. Schwartz, D. Richard Cutler, Charles P. Hawkins*, Ecological Modelling.

Abstract: Numerous methods have been developed to combat the unwanted effects of imbalanced training data on the performance of machine learning based predictive models. These methods attempt to balance the trade-off between sensitivity and specificity. However, the effects of specific imbalance methods on the performance of different machine learning algorithms is not well understood for ecological data. In this study, we used four machine learning algorithms (random forest, artificial neural network, gradient boosting, support vector machine) and five imbalance methods (base algorithm, cutoff, up-sampling, down-sampling, weighting) to produce species distribution models for 15 freshwater macroinvertebrate genera that varied from 2.4 – 29.4% in prevalence. All imbalance methods substantially improved average model performance over the base machine learning algorithms, except when up-sampling was applied to random forest. In contrast to up-sampling, down-sampling applied to all of the machine learning algorithms performed consistently well across all prevalence ranges. Choice of machine learning algorithm had little effect on model performance, although it was slightly more important when modeling taxa with the most imbalanced datasets. Our results suggest that species distribution model performance can be improved by implementing imbalance methods.

2023 Technical Report

Technical Report on Grade Inflation and Geographic Distribution of Top Students at the University of Oregon, *Samuel D. Schwartz*, Submitted to the University (Faculty) Senate and the Board of Trustees of the University of Oregon.

Abbreviated Abstract: This report aims to answer the following research questions (RQs) about undergraduate students at the University of Oregon (UO):

- RQ1: Has UO seen grade inflation?
- RQ2: Are academically high performing students more likely to be men or women at UO?
- RQ3: Which majors produce students with higher GPAs?
- RQ4: When considering students from Oregon, which geographic areas of the state produce academically high performing students? Similarly, which places don't produce high performing students? How does this production intersect with the racial and gender demographics of a student's community of origin?

Here are the takeaways (TA) from our investigation:

- TA1: The percentage of full time enrolled students with term GPAs of 3.75 or above has increased from 7% of total enrollment in Spring Term 2012 to 23% in Spring Term 2022.
- TA2: Our findings suggest that women at UO are consistently more likely than men to be among the top performing students in the past decade. On average, women are overrepresented among top performing students by an additional 6%, even when controlling for their higher enrollment numbers.
- TA3: These majors each produced more than 65 Dean's List students during at least one term since Winter Term 2020 and have also more than doubled their number of students on the list since pre-COVID times. A major's percentage increase in Dean's List students compared to pre-COVID times is in parentheses: Business Administration (up by 293%), Family and Human Services (271%), Advertising (167%), Cinema Studies (152%), Public Relations (151%), Journalism (139%), Psychology (116%), Political Science (104%).
- TA4: We have an urban-rural divide and a racial gap in the places where our top performing in-state students are sourced from.
  - There is a racial gap in zip codes which produce the top performers at UO. Zip codes in the top 50% of Dean's List producers are ∼6% more white than zip codes in the bottom 50%.
  - Rural areas are particularly unlikely to produce Dean's List students. Eastern Oregon fares worse than Western Oregon.
  - Of note is the Warm Springs Indian Reservation, which is a rural community of color. There hasn't been a single student from that community on the Dean's List in the past decade.
  - Within the Willamette Valley, rural areas and smaller towns with comparatively higher proportions of kids of color are less likely to have students on the Dean's Lists. Getting more kids on the Dean's List from smaller towns and rural areas slightly north of Salem, as well as Multnomah county in areas east of I205, would likely go far in closing the geographical racial gap in the zip codes of UO's top performers.

2022 Conference Paper

What Predicts Community in Virtual Computing Classrooms? A Machine Learning-Based Approach, *Samuel D. Schwartz and Joanna Goode*, RESPECT.

Abstract: The physical separation of people in online classes can produce feelings of isolation and disconnectedness in some students. Feelings of community – that is, a sense of belonging and that people in the community matter to each other – can help ameliorate these feelings of isolation and disconnectedness, and lead to more effective learning outcomes. We investigated what behavioral and demographic factors contributed to students' perception of classroom community in virtually conducted computing courses at a large public university in the Pacific Northwest region of the United States. We deployed a survey composed of self-reported demographic and behavior questions, and items from a previously validated instrument on classroom community for online courses. The questions were coded into 53 features used to predict a student's sense of classroom community. We used three machine learning models to do this prediction: a random forest, a gradient boosting machine, and a neural network. These models were validated and then analyzed with SHAP, a relatively new model-agnostic technique which uses principles from game theory to explain how data features influence a complex model's output. The most impactful prediction features for an individual student's self-assessed sense of classroom community across all three models were whether the student was an international student (negative impact), and participation behavior, including whether a student spoke, used the chat feature, interacted in breakout rooms, or turned on a webcam (all positive impact).

2021 Conference Paper

An Entropy-Based Approach for Identifying User-Preferred Camera Positions , *Nicole Marsaglia, Yuya Kawakami*, Samuel D. Schwartz*, Stefan Fields, Hank Childs*, LDAV.

Abstract: Viewpoint Quality (VQ) metrics have the potential to predict user preferences for camera placement. With this study, we introduce new VQ metrics that incorporate entropy, and explore how they can be used in combination. Our evaluation involves three phases: (1) creating a corpus of visualization imagery from ten large, scientific data sets, (2) conducting a user study with approximately 30 large data visualization experts who provided over 1000 responses, and (3) analyzing how our entropy-based VQ metrics compared with existing VQ metrics in predicting user preference. In terms of findings, we find that our entropy-based metrics are able to predict participant preferences 68% of the time, while existing VQ metrics perform much worse (52%). This finding, while valuable on its own, also opens the door for future work on in situ camera placement. Finally, as another important contribution, this work has the most extensive evaluation to date of existing VQ metrics to predict human preference for visualizations of large, scientific data sets.
** These authors contributed equally.*

2021 Grant Proposal (Unfunded)

A Machine Learning Based Tool for the Analyisis of Tumor Identification and Growth, Samuel D. Schwartz, Eleven page grant proposal for new research software for Oregon Health and Science University's Department of Radiology and Medical Imaging.

Overarching Software Objective: The objective of the software is to identify cancerous tumors from patient scans. Furthermore, the software is to predict whether the cancer is progressing, in remission, or stable based on scans of the patient taken at different points in time. In doing so, the software will calculate metrics and statistics about the patient's condition to assist medical personnel in making treatment decisions.

2021 Conference Paper

Machine Learning Based Autotuning for Parallel Particle Advection, *Samuel D. Schwartz, Hank Childs, David Pugmire*, EGPGV.

Abstract: Data-parallel particle advection algorithms contain multiple controls that affect their execution characteristics and performance, in particular how often to communicate and how much work to perform between communications. Unfortunately, the optimal settings for these controls vary based on workload, and, further, it is not easy to devise straight-forward heuristics that automate calculation of these settings. To solve this problem, we investigate a machine learning-based autotuning approach. During a pre-processing step, we train multiple machine learning techniques using a corpus of performance data that includes results across a variety of workloads and control settings. The best performing of these techniques is then used to form an oracle, i.e., a module that can determine control settings during run-time. To evaluate this approach, we assessed the ability of seven machine learning models to capture particle advection performance behavior and then ran experiments for 108 particle advection workloads on 64 GPUs of a supercomputer. Our findings show that our machine learning-based oracle achieves good speedups, especially relative to the available gains.

2019 MS Thesis

Machine Learning Techniques Applied to Discrete & Combinatorial Structures, Samuel D. Schwartz.

Abstract: Machine learning techniques, while well established for many observation types, have only recently come onto the scene for graphs and other combinatorial objects. Further, the use and efficacy of machine learning techniques in predicting computationally difficult invariants on discrete combinatorial objects is next to unknown. This thesis outlines methodologies useful for articulating discrete structures in the paradigm of many machine learning algorithms. Moreover, we examine several NP-hard problems in different articulations. We then report on the results of various techniques and methodologies in solving certain families of these problems.

2019 Conference Presentation

Image Completion Techniques Applied to Matrix Factorization in the Boolean Semiring, *Samuel D. Schwartz*, SIAM Wasatch Student Chapters Conference.

Abstract:

Matrix factorization under the reals is a solved problem. Decompositions, such as LU and QR, are taught in foundational Linear Algebra courses. Computationally, the algorithms to find a factorization of a real-valued matrix are polynomial in their time complexity. Once we restrict ourselves to the elements and operations in the Boolean semiring, however, the computational complexity to find any decomposition of a matrix becomes NP Complete. This presentation will discuss ways in which we can articulate the Boolean matrix factorization problem as a partially complete image. We will then show how machine learning techniques, such as random forests and neural networks, can be applied to this representation. In the processes of doing so, we will motivate how we can find polynomial time decompositions in the amortized case.

2019 Invited Talk

Software Development in Production Environments for Researchers, *Samuel D. Schwartz*, Data Analytics Club; Hosted for and by PhD Candidates, Utah State University.

Abstract:

Much of what we do as advanced undergraduates and graduate students in STEM revolves around writing programs to solve acute research problems.

Once the research problem is ameliorated by our code-fu skills, however, we are still left with the non-trivial task of publishing our software for use by others. This presentation discusses generalized practices for researchers which can facilitate this process.

2018 Technical Report

Modified Kernel Principal Component Analysis Results on NP-Class Problems in Digraphs, *Samuel D. Schwartz and David E. Brown*.

Abstract:

Kernel methods, while well established for many observation types, have only recently come onto the scene for digraphs, and their use in predicting computationally difficult graph invariants is next to unknown. This paper outlines an algorithm which utilizes a modified kernel principal component analysis using a Ye-Wilson-Hancock kernel to generate contrived graph predictors. Those predictors are, by design, in a space which can then be readily analyzed and partitioned using traditional supervised learning approaches such as k-Nearest Neighbors or Support Vector Machines. Two NP-class problems are considered, with remarkably good prediction resulting.

2018 Conference Presentation

How to Type Math: [Teaching] LaTeX , *Samuel D. Schwartz and Kaitlin S. Murphy*, Together we Teach (TweT).

Abstract:

In the modern era, typewritten work is the norm in every discipline and mathematics and statistics are no exception. Nevertheless, many undergraduate institutions provide little to no training to their students on how to typeset mathematical documents. Indeed, many instructors gladly accept handwritten work even in upper-level division mathematics courses. Contrasted with many upper division humanities classes, where a handwritten essay would likely be rejected upon receipt, mathematics and statistics departments nationwide do a poor job in preparing their students for the reality that both academia and industry expect professional mathematics to be typed. Given that a non-negligible number of incoming graduate students have no experience in using the most common framework used to typeset mathematical expressions, we present a ready-made workshop framework which will ameliorate the initially steep learning curve of LaTeX.

2015 Grant, US$1,000

Analysis of Statistical Components of Tournaments Modeling Patterns in Voting and Social Choice Theory., *Samuel D. Schwartz*, College of Science Undergraduate Minigrant Program, Utah State University.