

Ronald J. Yurko Jr.

Curriculum Vitae

5000 Forbes Ave

Pittsburgh, PA 15213

✉ ryurko@stat.cmu.edu

🌐 <https://www.stat.cmu.edu/~ryurko/>

🐙 Github

Positions

Department of Statistics & Data Science, Carnegie Mellon University

Fall 2022 – **Assistant Teaching Professor.**

Summer 2022 **Special Faculty.**

Education

2017–2022 **PhD in Statistics**, *Carnegie Mellon University*.

Thesis supervised by Kathryn Roeder and Max G'Sell, titled "Selective inference approaches for augmenting genetic association studies with multi-omics metadata"

2017–2018 **MS in Statistics**, *Carnegie Mellon University*.

2012–2015 **BS in Statistics**, *Carnegie Mellon University*, University Honors.

Publications

Journal Articles

- 2021 **Ronald Yurko**, Kathryn Roeder, Bernie Devlin, and Max G'Sell. An approach to gene-based testing accounting for dependence of tests among nearby genes. *Briefings in Bioinformatics*, volume 22, 08 2021.
- 2021 **Ronald Yurko**, Kathryn Roeder, Bernie Devlin, and Max G'Sell. H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. *Annals of Human Genetics*, volume 85, pages 97–100. Wiley Online Library, 2021.
- 2021 Riccardo Fogliato, Natalia L Oliveira, and **Ronald Yurko**. TRAP: a predictive framework for the Assessment of Performance in Trail Running. *Journal of Quantitative Analysis in Sports*, volume 17, pages 129–143. De Gruyter, 2021.
- 2020 **Ronald Yurko**, Francesca Matano, Lee F Richardson, Nicholas Granered, Taylor Pospisil, Konstantinos Pelechrinis, and Samuel L Ventura. Going deep: models for continuous-time within-play valuation of game outcomes in american football with tracking data. *Journal of Quantitative Analysis in Sports*, volume 16, pages 163–182. De Gruyter, 2020.
- 2020 **Ronald Yurko**, Max G'Sell, Kathryn Roeder, and Bernie Devlin. A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk. *Proceedings of the National Academy of Sciences*, volume 117, pages 15028–15035. National Academy of Sciences, 2020.
- 2020 Sarah Mallepalle, **Ronald Yurko**, Konstantinos Pelechrinis, and Samuel L Ventura. Extracting NFL tracking data from images to evaluate quarterbacks and pass defenses. *Journal of Quantitative Analysis in Sports*, volume 16, pages 95–120. De Gruyter, 2020.
- 2020 Rishav Dutta, **Ronald Yurko**, and Samuel L Ventura. Unsupervised methods for identifying pass coverage among defensive backs with nfl player tracking data. *Journal of Quantitative Analysis in Sports*, volume 16, pages 143–161. De Gruyter, 2020.
- 2019 **Ronald Yurko**, Samuel Ventura, and Maksim Horowitz. nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, volume 15, pages 163–183. De Gruyter, 2019.

Invited Commentaries and Popular Press

- 2023 Ronald Yurko and Rebecca Nugent. Discussion on “Flexible marked spatio-temporal point processes with applications to event sequences from association football” by Narayanan et al. *Journal of the Royal Statistical Society, Series C*. Oxford University Press, 2023.
- 2019 Konstantinos Pelechrinis, **Ronald Yurko**, and Sam Ventura. Reducing Concussions in the NFL: A Data-Driven Approach. *CHANCE*, volume 32, pages 46–56. Taylor & Francis, 2019.

Miscellaneous Articles

- 2021 **Evaluating defender ability to limit YAC**, *Joint work with Kostas Pelechrinis*, NFL Big Data Bowl 2021 (Honorable Mention).
Kaggle: <https://www.kaggle.com/ryurko21/evaluating-defender-ability-to-limit-yac>
- 2019 **Detecting data analysis patterns in text and graphs to characterize student workflows**, *Advanced Data Analysis report*, Advised by Rebecca Nugent.

Teaching

Courses Taught at Carnegie Mellon (*ordered by course level*)

- Fall 2022 **46–926 Machine Learning I (*mini*)**.
Core course on statistical machine learning for the Master’s in Computational Finance program, including regression, classification, and tree-based methods.
- Fall 2022 **36–613 Data Visualization (*mini*)**.
Master’s level course on creating and understanding data visualizations and interactive tools from a statistical perspective; created course curriculum and materials.
- Spring 2023 **36-493 Sports Analytics Capstone**.
Undergraduate capstone course using data provided by CMU athletics for sports analytics research projects.
- Spring 2022, **36-315 Statistical Graphics and Visualization**.
Spring 2023 Undergraduate course on creating and understanding data visualizations from a statistical perspective.
- 2020–2022 **Summer Undergraduate Research Experience in Statistics, Lead Instructor and Director**.
8-10 week program with 12-16 students selected each year nationally with an emphasis on diversity. Experience includes client-facing capstone with real-world problems and datasets. Created course curriculum, materials, and advised student projects: <https://www.stat.cmu.edu/cmsac/sure/2022/materials/>
- 2015 **Introduction to Sabermetrics and Exploring Baseball Data with R**.
Instructor and created course materials in undergraduate student-taught course program.

Workshops

- 2022 **Big Data Bowl Workshop Sponsored by SumerSports**.
Created workshop materials and instructor with live coding demonstrations of accessing and analyzing player-tracking data: <https://www.stat.cmu.edu/cmsac/conference/2022/workshop/slides.html>
- 2018–2019 **Carnegie Mellon Football Analytics Workshop**.
Created workshop materials and instructor with live coding demonstrations of accessing and analyzing NFL play-by-play with introduction to ELO ratings: <https://www.stat.cmu.edu/cmsac/football/>
- Summer 2019 **Wharton Moneyball Academy and Training Camp**.
Assisted in development of course materials and course instructor for week-long introductory statistics course for high-school students.
- 2018–2019 **Carnegie Mellon Football Analytics Workshop**.
Created workshop materials, organized Q&A session with Pittsburgh Pirates, and instructor with live coding demonstrations of accessing and analyzing baseball data: <https://ryurko.github.io/Carnegie-Mellon-Baseball-Analytics-Workshop/>

Executive Education

- 2022 – **Data Science Executive Education programs**, *In progress for multiple global/national finance institutions and pharmaceutical companies (company names not allowed to be public yet)*.
Instructor role providing feedback in custom ISLE analytics platform.

2022 – **Data Science for Business Leaders**, *Optum/United Health Group*.

Instructor for data visualization content.

Courses Served as Teaching Assistant at Carnegie Mellon (*ordered by course level*)

Spring 2021 **46-927 Statistical Machine Learning II**, (*mini*).

Core course on statistical machine learning for the Master's in Computational Finance program, including clustering, classification, dimension reduction, and deep learning.

Summer 2018 **36-315 Statistical Graphics and Visualization**.

Undergraduate course on creating and understanding data visualizations from a statistical perspective.

Spring 2018 **36-462 Data Mining**.

Undergraduate course on statistical learning including clustering, classification, dimension reduction, and tree-based models.

Fall 2017 **36-350 Statistical Computing**.

Undergraduate course introducing programming for statistical analysis in R.

Fall 2013, **36-201 Statistical Reasoning and Practice**.

Fall 2014, Undergraduate introductory statistics for humanities and social sciences majors.

Spring 2014

Courses Served as Grader at Carnegie Mellon

Fall 2015 **36-225 Introduction to Probability Theory**.

Undergraduate course introducing mathematical probability theory for statistics, math, and other majors.

Spring 2015 **36-226 Introduction to Statistical Inference**.

Undergraduate course introducing mathematical statistics for statistics, math, and other majors.

Advising / Supervising

Independent Study Advisor

Fall 2022 **Jacob Muskovitz**, *Modeling Defensive Outcomes from Linear Distances Between Players*.

Advised Heinz College master's student on project working exploring and modeling spatio-temporal data provided by a professional soccer team.

Undergraduate Research Supervision

Fall 2022 **Undergraduate Research Course**, *Faculty Advisor*.

Advised two teams of undergraduate students working with spatio-temporal data from: (1) Big Data Derby and (2) Big Data Bowl.

Fall 2021 **Data Science Initiative**, *PhD Project Fellow*.

Advised two teams of undergraduate students working with United States Olympic & Paralympic Committee.

Fall 2021 **Quantitative Social Science Scholars (QSSS) Program Senior Thesis**, *Advisor*.

Advisor of undergraduate senior thesis on measuring latent attributes for NBA players.

Summer 2019 **Summer Undergraduate Research Experience in Statistics**, *Teaching Assistant*.

Advised multiple undergraduate student projects and created datasets for program labs.

Presentations

Invited Talks

Nov. 2021 **Center of Modeling, Simulation and Interactions (MSI) Seminar**, *Selective inference approaches for augmenting genetic association studies with multi-omics metadata*, Université Côte d'Azur, Virtual.

July 2020 **International Seminar on Selective Inference**, *Adaptive approaches for augmenting genetic association studies with multi-omics covariates*, Virtual, presented by Kathryn Roeder.

Oct. 2020 **UConn Sports Analytics Symposium (Keynote Speaker)**, *Going Deep: Models for Continuous-Time Within-Play Valuation of Game Outcomes in American Football with Tracking Data*, Virtual.

- Aug. 2020 **Joint Statistical Meetings**, *Going Deep: Models for Continuous-Time Within-Play Valuation of Game Outcomes in American Football with Tracking Data*, Virtual, Presented by Lee Richardson.
- Sept. 2019 **New England Symposium on Statistics in Sports (Featured Talk)**, *Going Deep: Models for Continuous-Time Within-Play Valuation of Game Outcomes in American Football with Tracking Data*, Harvard University.
- May 2019 **United States Conference on Teaching Statistics**, *Many Students, One Dataset: Using ISLE to Teach Reproducibility and the Impact of Data Analysis Decisions on Conclusions*, Pennsylvania State University, joint presentation with Rebecca Nugent, Philipp Burckhardt, Frank Kovacs.
- May 2019 **United States Conference on Teaching Statistics**, *Many Students, One Dataset: Using ISLE to Teach Reproducibility and the Impact of Data Analysis Decisions on Conclusions*, Pennsylvania State University, joint presentation with Rebecca Nugent, Philipp Burckhardt, Frank Kovacs.
- Sept. 2018 **Pittsburgh useR Group**, *Exploring NFL data with nflscrapR*, Pittsburgh, PA.
- Aug. 2018 **RIT Sports Analytics Conference**, *nflWAR: a reproducible method for offensive player evaluation in football*, Rochester Institute of Technology.
- Oct. 2017 **Carnegie Mellon Sports Analytics Conference**, *nflWAR: a reproducible method for offensive player evaluation in football*, Carnegie Mellon University.
- Oct. 2017 **Computational Sports Informatics Colloquium**, *nflWAR: a reproducible method for offensive player evaluation in football*, University of Pittsburgh.
- Sept. 2017 **New England Symposium on Statistics in Sports**, *nflWAR: a reproducible method for offensive player evaluation in football*, Harvard University.

Contributed Talks

- Aug. 2020 **Joint Statistical Meetings**, *A selective inference approach for FDR control using multi-omics covariates yields insights into disease risk*, Virtual.
- Aug. 2018 **Cascadia Symposium on Statistics in Sports**, *Multilevel models to measure player, team, and stadium effects on NFL injury risk*, Vancouver, Canada, presented by Zachary Binney.
- June 2018 **Classification Society Annual Meeting**, *A case study in reproducibility: detecting data analysis patterns in text and graphs to characterize student workflows*, Stony Brook University.
- May. 2018 **Symposium on Data Science & Statistics**, *Variable selection for consistent clustering*, Reston, VA.
- July 2017 **Great Lakes Analytics in Sports Conference**, *NFL player evaluation using expected points added with nflscrapR*, University of Wisconsin–Stevens Point.
- June 2017 **Classification Society Annual Meeting**, *Variable selection for consistent clustering*, University of California, Santa Cruz - Silicon Valley Campus.
- Apr. 2017 **UP-STAT (Second Place, Best Young Researchers' Award in Category C: Application)**, *nflscrapR: an R package for easy access to NFL data and a new model for expected points and win probability*, Canisius College.

Invited Guest Lectures

- July 2022 **PT MBA Access Weekend**, *Sports Analytics in the Post-Moneyball Era: How technology and machine learning are changing the way professional sports teams evaluate players*, Carnegie Mellon University, Tepper School of Business.
Part-time MBA Program.
- May 2022 **In-Person MSBA Immersion Campus Experience Weekend**, *Sports Analytics in the Post-Moneyball Era: How technology and machine learning are changing the way professional sports teams evaluate players*, Carnegie Mellon University, Tepper School of Business.
Online Master of Science in Business Analytics (MSBA) Program.

- Apr. 2022 **Human Genetics 2080 Statistical Genetics**, *An approach to gene-based testing accounting for dependence of tests among nearby genes*, University of Pittsburgh, Graduate School of Public Health, invited by Professor Daniel Weeks.
Graduate course in the principles and practice of statistical genetics.
- Mar. 2022 **STATS 100: The Mathematics of Sports**, *An introduction to NFL analytics research*, Stanford University, invited by Instructor Xavier Gonzalez.
Undergraduate course on statistics, probability, and mathematics in sports.
- Oct. 2021 **Foundation Seminar: Sports, Statistics and Society**, *An introduction to NFL analytics research*, Bucknell University, invited by Professor Abby Flynt.
First-year undergraduate course on statistics in sports.
- Oct. 2021 **STAT401 Sports Analytics**, *Going Deep: Models for Continuous-Time Within-Play Valuation of Game Outcomes in American Football with Tracking Data*, The Wharton School, University of Pennsylvania, invited by Professor Abraham Wyner.
Advanced undergraduate course introducing students to the growing field of sports analytics.
- Sept. 2020 **Foundation Seminar: Sports, Statistics and Society**, *An introduction to NFL analytics research*, Bucknell University, invited by Professor Abby Flynt.
First-year undergraduate course on statistics in sports.
- Mar. 2020 **Human Genetics 2080 Statistical Genetics**, *Application of post-selection inference to multi-omics data yields insights into the etiologies of human diseases*, University of Pittsburgh, Graduate School of Public Health, invited by Professor Daniel Weeks.
Graduate course in the principles and practice of statistical genetics.
- Feb. 2020 **BUSMG 7334 - Sports Analytics**, *Going Deep: Models for Continuous-Time Within-Play Valuation of Game Outcomes in American Football with Tracking Data*, Ohio State University, invited by Professor John Draper.
Course introducing students to the current state of sports analytics.
- Feb. 2018 **INFSCI 1091: Special Topics - Moneyball 2.0: Winning in Sports with Data**, *nflWAR: a reproducible method for offensive player evaluation in football*, University of Pittsburgh, invited by Professor Kostas Pelechrinis.
Special topics course introducing students to data collection, analysis, and visualization in sports.

Conference Poster Presentations

- Oct. 2019 **Annual Meeting of the American Society of Human Genetics**, *Application of post-selection inference to multi-omics data yields insights into the etiologies of human diseases*, Houston, TX.
- Nov. 2019 **Carnegie Mellon Sports Analytics Conference (Best Poster Prize)**, *TRAP: a predictive framework for the assessment of performance in trail running*, Harvard University, presented by Riccardo Fogliato.
- Sept. 2019 **New England Symposium on Statistics in Sports (Best Student Poster Prize)**, *TRAP: a predictive framework for the assessment of performance in trail running*, Harvard University, presented by Natalia L. Oliveira.
- Apr. 2018 **Pittsburgh ASA Chapter Spring Banquet**, *Variable selection for consistent clustering*, Pittsburgh, PA.
- Nov 2015 **Dietrich Undergraduate Colloquium**, *Improving predictions of ensemble methods using distributions of estimated probabilities*, Carnegie Mellon University.
- May 2015 **Meeting of the Minds (First Place, Statistics Poster Competition)**, *Classifying Kepler objects of interest*, Carnegie Mellon University.

Electronic Poster Presentations

- Oct. 2020 **Annual Meeting of the American Society of Human Genetics**, *Augmenting gene-level tests based on two-sided summary statistics with multiomics covariates*.

- June 2020 **Symposium on Data Science & Statistics**, *A selective inference approach for FDR control using multi-omics covariates yields insights into disease risk.*
- May 2018 **Electronic Conference On Teaching Statistics**, *Identifying misconceptions of introductory data science using a thinkaloud protocol*, Joint work with S. Hyun, P. Burckhardt, P. Elliott, C. Evans, K. Lin, A. Luby, C. P. Makris, J. Orellana, A. Reinhart, J. Wieczorek, G. Weinberg, R. Nugent.
- May 2018 **Electronic Conference On Teaching Statistics**, *Using text analysis to characterize student learning in an introductory statistics & data science course.*

Funding

- 2022 – 2026 **NSF: Improving Undergraduate STEM Education**, *SCORE with Data: Building a sustainable national network for developing and disseminating Sports Content for Outreach, Research, and Education in data science*, Senior Personnel.
\$1,100,000; four years

Software

- 2020 **snpcombineR: R package to combine SNP-level test statistics at various region levels.**
GitHub: <https://github.com/ryurko/snpcombineR>
- 2019 **adaptMT: Modifications including wrapper functions for XGBoost implementation with EM algorithm cross-validation tuning.**
GitHub: <https://github.com/ryurko/adaptMT>
- 2018 **fcscrapR: R package to scrape soccer commentary and statistics from ESPN.**
GitHub: <https://github.com/ryurko/fcscrapR>
- 2018 **nflWAR: An R package to compute WAR for offensive players using nflscrapR.**
GitHub: <https://github.com/ryurko/nflWAR>
- 2017 **nflscrapR: Compiling the NFL Play-by-Play API for easy use in R.**
GitHub: <https://github.com/maksimhorowitz/nflscrapR>

Student Research Experience

- 2018 – 2022 **Graduate Research Assistant**, *Applications of selective inference in statistical genetics*, Advised by Kathryn Roeder, Max G'Sell, Bernie Devlin.
Research funding provided by Simons Foundation Grant SFARI SF575097
- Fall 2015 **Undergraduate Research Assistant**, *PREDs: Prediction with Ensembles using Distribution Summaries*, Advised by Sam Ventura and Rebecca Nugent.
- Spring 2015 **Undergraduate Research Course**, *Classifying Kepler Objects of Interest*, Advised by Peter Freeman and Rebecca Nugent.
- Fall 2014 **Independent Research**, *The Science of Fooling Batters: An Objective Analysis of Pitch Sequencing*, Advised by Andrew C. Thomas.

Industry Experience

- 2021 – 2022 **Part-time Data Scientist**, *Football Strategy*, Zelus Analytics, Remote.
- 2016 – 2017 **Quantitative Analytics Associate**, *Analytics & Portfolio management*, PNC Financial Services, Pittsburgh, PA.
- 2015 **Risk Management Intern**, *Analytics & Portfolio management*, PNC Financial Services, Pittsburgh, PA.
- 2014 **Data and Analytics Intern**, *Baseball Operations*, Pittsburgh Pirates, Pittsburgh, PA.
- 2013 **Equity Research Intern**, *Equity Analysis*, Schenley Park Capital Management, Pittsburgh, PA.

Fellowships & Awards

- 2022 **Student of the Year**, *ASA Pittsburgh Chapter*.
- 2018 – 2019 **Carnegie Mellon Presidential Fellowship**.
- 2015 **Andrew Carnegie Society Scholar**.
- 2015 **Phi Kappa Phi Honor Society**.

Service

Organization

- 2017 – **Co-Organizer**, *Carnegie Mellon Sports Analytics Conference*.
200+ attendees from academia, industry, and professional sports. Responsibilities include maintaining and assessing current research in field, website/event/speaker/press management, marketing materials, budgeting, and bringing in sponsors.
- 2018 – **Co-Organizer**, *CMSAC Reproducible Research Competition*.
Inclusive conference competition to promote reproducible research with separate tracks for students and software contributions. Responsibilities include creating competition format, promoting, and organizing evaluation of submissions with review feedback.
- 2018 – 2019 **Organizer**, *Carnegie Mellon Football Analytics Workshop*.
80+ attendees from academia, industry, and professional sports. Responsibilities included creating workshop material, website/event/press management, coordinating Q&A session with NFL Director of Data and Analytics.
- 2018 **Organizer**, *Carnegie Mellon Baseball Analytics Workshop*.
50+ attendees from academia, industry, and professional sports. Responsibilities included creating workshop material, website/event/press management, and coordinating Q&A session with professional baseball team.

Department Service

- 2022 – **Director**, *Summer Undergraduate Research Experience in Statistics*.
Director of 8 week summer program with 12-16 students selected each year nationally with an emphasis on diversity. Experience includes managing admissions, organizing speakers, advising client-facing capstone projects with real-world problems and datasets.
- 2022 – **Committee Member**, *MSP Program*.
Discuss and implement changes to program curriculum.
- 2022 – **Committee Member**, *MSP Admissions*.
- 2017 – **Organizer**, *Statistics in Sports Reading Group*.
- 2020 – 2021 **Zoom Moderator**, *Statistics & Data Science Research Showcase*.
- 2019 – 2022 **Organizer**, *Statistical Genetics & Genomics Student Reading Group*.
- 2019 **Judge**, *Meeting of the Minds*.
- 2018 – 2019 **Mentor**, *Women in Statistics Matched Pairs Mentorship Program*.
- 2018 – 2019 **Judge**, *Statistical Graphics Poster Presentations*.
- 2017 – 2022 **Cohort Representative**, *Student Advisory Committee*.

University Service

- 2022 – **Faculty Advisor**, *Carnegie Mellon Sports Analytics Club*.
- 2017 – 2022 **Graduate Student Advisor**, *Carnegie Mellon Sports Analytics Club*.
- 2013 – 2016 **Co-Founder, Vice President, Editor, Writer**, *Carnegie Mellon Sports Analytics Club*.

Professional Service

- 2022 – **Associate Editor**, *Journal of Quantitative Analysis in Sports*.
- 2022 **Judge**, *SMT Data Challenge*.
- 2021 **Mentor**, *NFL Big Data Bowl Mentorship Program*.
Program designed with goal of increasing diversity within sports analytics by advising students from underrepresented groups in sports analytics / STEM fields.

2021 **Judge**, *NFL Big Data Bowl*.

2020 – **Co-Host**, *Open Source Sports*.

Podcast created to serve as public reading group discussing the latest research in sports analytics.

Peer Reviewer Big Data, Journal of Quantitative Analysis in Sports, GENETICS, PLOS Computational Biology, Journal of Sports Analytics, Journal of Business Analytics, Communications in Statistics, AStA Advances in Statistical Analysis

Media Recognition

Interviews Wharton Moneyball, Unexpected Points, Pittsburgh Post Gazette, theScore

Citations FiveThirtyEight, The Athletic, Wall Street Journal

Professional Societies

American Society of Human Genetics

American Statistical Association

Classification Society

Computer Skills

Expert R

Proficient SAS, SQL

Intermediate C++, Julia, Python

Beginner HTML, Clojure, Java

Activities

Team Sports

2017 – 2022 **Graduate Student Assembly Summer Sports**, *Softball captain*.

2013 – 2022 **Carnegie Mellon Intramural Sports**, *Flag-football captain*.

2018 co-rec champions

2013 **Carnegie Mellon University Club Baseball Team**.

Volunteering

2019 **Campaign Against Cancer**.

Charity Runs

2017 – **Pirates Home Run 5K 10K**.

Present <https://www.mlb.com/pirates/community/race>

2017 – **The Great Race**.

Present <https://www.rungreatrace.com/>

2016 – **Pittsburgh Penguins 6.6K Run & Family Walk**.

Present <https://mariolemieux.org/our-events/pittsburgh-penguins-66k/>