Ronald J. Yurko, Jr.

Curriculum Vitae

Education

2017-Present **PhD in Statistics & Data Science**, *Carnegie Mellon University*.
- Dissertation: Selective inference approaches for augmenting genetic association studies with multi-omics metadata; Advisors: Kathryn Roeder and Max G'Sell

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

2012-2015 **BS in Statistics**, *Carnegie Mellon University*.
- University Honors, GPA: 3.97/4.00

Publications

2021 **An approach to gene-based testing accounting for dependence of tests among nearby genes**, *R. Yurko, K. Roeder, B. Devlin, M. G'Sell*, Briefings in Bioinformatics.
- Oxford: https://doi.org/10.1093/bib/bbab329

2020 **Going Deep: models for continuous-time within-play valuation of game outcomes in american football with tracking data**, *R. Yurko, F. Matano, L. Richardson, N. Granered, T. Pospisil, K. Pelechrinis, S. Ventura*, Journal of Quantitative Analysis in Sports.
- De Gruyter: https://doi.org/10.1515/jqas-2019-0056

Extracting NFL tracking data from images to evaluate quarterbacks and pass defenses, *S. Mallepalle, R. Yurko, K. Pelechrinis, S. Ventura*, Journal of Quantitative Analysis in Sports.
- De Gruyter: https://doi.org/10.1515/jqas-2019-0052

Unsupervised methods for identifying pass coverage among defensive backs with NFL player tracking data, *R. Dutta, R. Yurko, S. Ventura*, Journal of Quantitative Analysis in Sports.
- De Gruyter: https://doi.org/10.1515/jqas-2020-0017

A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk, *R. Yurko, M. G'Sell, K. Roeder, B. Devlin*, Proceedings of the National Academy of Sciences.
- PNAS: https://doi.org/10.1073/pnas.1918862117

H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives, *R. Yurko, K. Roeder, B. Devlin, M. G'Sell*, Annals of Human Genetics.
- Wiley: https://doi.org/10.1111/ahg.12412

TRAP: A predictive framework for the assessment of performance in trail running, *R. Fogliato, N. Oliveira, R. Yurko*, Journal of Quantitative Analysis in Sports.
- De Gruyter: https://doi.org/10.1515/jqas-2020-0013

2019 **nflWAR: a reproducible method for offensive player evaluation in football**, *R. Yurko, S. Ventura, M. Horowitz*, Journal of Quantitative Analysis in Sports.
- De Gruyter: https://doi.org/10.1515/jqas-2018-0010

Reducing concussions in the NFL: a data-driven approach, *K. Pelechrinis, R. Yurko, S. Ventura*, CHANCE.
- Taylor & Francis: https://doi.org/10.1080/09332480.2019.1695442

Presentations

Invited Talks

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

Going Deep: models for continuous-time within-play valuation of game outcomes in american football with tracking data, *Keynote Speaker*, UConn Sports Analytics Symposium.

Going Deep: models for continuous-time within-play valuation of game outcomes in american football with tracking data, *Presented by Lee Richardson*, Joint Statistical Meetings.

2019 **Going Deep: models for continuous-time within-play valuation of game outcomes in american football with tracking data**, New England Symposium on Statistics in Sports.

Many Students, One Dataset: Using ISLE to Teach Reproducibility and the Impact of Data Analysis Decisions on Conclusions, *Joint work with R. Nugent, P. Burckhardt, F. Kovacs*, USCOTS.

2018 **nflWAR: a reproducible method for offensive player evaluation in football**, RIT Sports Analytics Conference.

Exploring NFL data with nflscrapR, Pittsburgh useR Group.

2017 **nflWAR: a reproducible method for offensive player evaluation in football**, New England Symposium on Statistics in Sports.

nflWAR: a reproducible method for offensive player evaluation in football, Computational Sports Informatics Colloquium.

nflWAR: a reproducible method for offensive player evaluation in football, Carnegie Mellon Sports Analytics Conference.

Contributed Talks

2020 **A selective inference approach for FDR control using multi-omics covariates yields insights into disease risk**, Joint Statistical Meetings.

2018 **Variable selection for consistent clustering**, Symposium on Data Science & Statistics.

A case study in reproducibility: detecting data analysis patterns in text and graphs to characterize student workflows, Classification Society Annual Meeting.

Multilevel models to measure player, team, and stadium effects on NFL injury risk, *Joint work with Zachary Binney*, Cascadia Symposium on Statistics in Sports.

2017 **nflscrapR: an R package for easy access to NFL data and a new model for expected points and win probability**, UP-STAT.
- Second Place, Best Young Researchers' Award in Category C: Application

NFL player evaluation using expected points added with nflscrapR, Great Lakes Analytics in Sports Conference.

Variable selection for consistent clustering, Classification Society Annual Meeting.

Conference Poster Presentations

2019 **Application of post-selection inference to multi-omics data yields insights into the etiologies of human diseases**, Annual Meeting of the American Society of Human Genetics.

TRAP: a predictive framework for the assessment of performance in trail running, *Presented by Natalia L. Oliveira*, New England Symposium on Statistics in Sports.
- Best Student Poster Prize

TRAP: a predictive framework for the assessment of performance in trail running, *Presented by R. Fogliato*, Carnegie Mellon Sports Analytics Conference.
- Best Poster Award

2018 **Variable selection for consistent clustering**, Pittsburgh ASA Chapter Spring Banquet.

2015 **Classifying Kepler objects of interest**, *Joint work with Eric Alpert*, Meeting of the Minds.

- First Place, Statistics Poster Competition

Improving predictions of ensemble methods using distributions of estimated probabilities, Dietrich Undergraduate Colloquium.

Electronic Poster Presentations

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

A selective inference approach for FDR control using multi-omics covariates yields insights into disease risk, Symposium on Data Science & Statistics.

2018 **Identifying misconceptions of introductory data science using a think-aloud 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*, eCOTS.

Using text analysis to characterize student learning in an introductory statistics & data science course, eCOTS.

Miscellaneous Articles

2021 **Evaluating defender ability to limit YAC**, *R. Yurko and K. 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**, *R. Yurko*, Advanced Data Analysis report.

- Advisor: Rebecca Nugent

Software

R Packages

2017 **nflscrapR: Compiling the NFL Play-by-Play API for easy use in R**, *M. Horowitz, R. Yurko, S. Ventura.*

- GitHub: https://github.com/maksimhorowitz/nflscrapR

2018 **nflWAR: An R package to compute WAR for offensive players using nflscrapR**, *R. Yurko.*

- GitHub: https://github.com/ryurko/nflWAR

fcscrapR: R package to scrape soccer commentary and statistics from ESPN, *R. Yurko*.
- GitHub: https://github.com/ryurko/fcscrapR

2019 **adaptMT: Modifications including wrapper functions for XGBoost implementation with EM algorithm cross-validation tuning**, *R. Yurko*.
- GitHub: https://github.com/ryurko/adaptMT

2020 **snpcombineR: R package to combine SNP-level test statistics at various region levels**, *R. Yurko*.
- GitHub: https://github.com/ryurko/snpcombineR

Teaching Experience

Courses Taught at Carnegie Mellon

2020 - 2021 **Summer Undergraduate Research Experience in Statistics**, *Summer: 2020, 2021*.
- Assistant director and lead instructor of program, created course curriculum / materials, and advised student projects (http://www.stat.cmu.edu/cmsac/)

2015 **Introduction to Sabermetrics and Exploring Baseball Data with R**, *Fall: 2015, Spring: 2015*.
- Created course materials in student-taught course program

Courses Served as Teaching Assistant at Carnegie Mellon

2021 **Machine Learning II**, *Spring: 2021*.

2019 **Summer Undergraduate Research Experience in Statistics**, *Summer: 2019*.
- Created datasets and program materials, advised student projects

2018 **Statistical Graphics and Visualization**, *Summer: 2018*.

2018 **Data Mining**, *Spring: 2018*.

2017 **Statistical Computing**, *Fall: 2017*.

2013 - 2014 **Statistical Reasoning and Practice**, *Fall: 2013, 2014, Spring: 2014*.

Courses Served as Grader at Carnegie Mellon

2015 **Introduction to Probability Theory**, *Fall: 2015*.

2015 **Introduction to Statistical Inference**, *Spring: 2015*.

Workshops Taught

2018 - 2019 **Carnegie Mellon Football Analytics Workshop**.
- Created workshop materials and instructor of live coding demo

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 **Carnegie Mellon Baseball Analytics Workshop**.
- Created workshop materials and co-instructor of live coding demo

Research Experience

Research Assistant at Carnegie Mellon

2018-Present **Research Assistant**, *Advised by: K. Roeder, M. G'Sell, B. Devlin*, Applications of selective inference in statistical genetics.

2015 **Undergraduate Research Assistant**, *Advised by: S. Ventura, R. Nugent*, PREDS: Prediction with Ensembles using Distribution Summaries.

Other Research Experience at Carnegie Mellon

2015 **Undergraduate Research Course**, *Advised by: R. Nugent, P. Freeman*, Classifying Kepler Objects of Interest.

2014 **Independent Research**, *Advised by: Andrew C. Thomas*, The Science of Fooling Batters: An Objective Analysis of Pitch Sequencing.

Industry Experience

Part-time

2021-Present **Data Scientist**, Football Strategy, Zelus Analytics, Remote.

Full-time

2016-2017 **Quantitative Analytics Associate**, Analytics & Portfolio Management, PNC Financial Services, Pittsburgh, PA.

Internships

2015 **Risk Management Intern**, Analytics & Portfolio Management, PNC Financial Services, Pittsburgh, PA.

2014 **Data and Analytics Intern**, Baseball Operations, Pittsburgh Pirates, Pittsburgh, PA.

2013 **Intern**, Equity Analysis, Schenley Park Capital Management, Pittsburgh, PA.

Fellowships and Funding

2018-2019 **Carnegie Mellon Presidential Fellowship**.

Honors and Awards

2021 **Honorable Mention**, NFL Big Data Bowl 2021.

2019 **Best Student Poster Prize**, New England Symposium on Statistics in Sports.

2019 **Best Poster Award**, Carnegie Mellon Sports Analytics Conference.

2017 **Second Place, Best Young Researchers' Award in Category C: Application**, UP-STAT.

2015 **Andrew Carnegie Society Scholar**, Carnegie Mellon University.

2015 **Phi Kappa Phi Honor Society**.

2015 **First Place, Statistics Poster Competition**, Meeting of the Minds.

2014 **Honors courses**, *Mathematical Statistics Honors, Undergraduate Research Course*, Department of Statistics & Data Science, Carnegie Mellon University.

Professional Service

Organization

2018-Present **Co-Organizer**, CMSAC Reproducible Research Competition.
- Conference competition to promote reproducible research. Responsibilities included creating competition format, promoting, and organizing evaluation of submissions with reviewer feedback

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

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

2017-Present **Co-Organizer**, Carnegie Mellon Sports Analytics Conference.
- 200+ attendees from academia, industry, and professional sports. Responsibilities included maintaining and assessing current research in field, website/event/speaker/press management, marketing materials, budgeting

Journal 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.

Department Service

2019 **Judge**, Meeting of the Minds.

2019-Present **Organizer**, StatGen Reading Group.

2018-2019 **Mentor**, Women in Statistics Matched Pairs Mentorship Program.

2018-Present **Judge**, Statistical Graphics Poster Presentations.

2017-Present **Cohort representative**, Student Advisory Committee.

2017-Present **Organizer**, Statistics in Sports Reading and Research Group.

University Service

2017-Present **Graduate Student Advisor**, Carnegie Mellon Sports Analytics Club.

2013-2016 **Co-Founder, Vice President, Editor, and Writer**, Carnegie Mellon Sports Analytics Club.

Workshop Participation

2019 **Introduction to Bayesian Inference with Stan**, University of Pittsburgh, Pittsburgh PA.

Professional Societies

American Society of Human Genetics.

American Statistical Association.

Classification Society.

Society of American Baseball Research.

Computing Skills

Expert **R**.

Proficient **SAS**.

Intermediate **C++, Julia, Python, SQL**.

Beginner **HTML, Clojure, Java**.

Activities

Team Sports

2017-Present **Graduate Student Assembly Summer Sports**.
- Co-captain: softball

2013-Present **Carnegie Mellon Intramural Sports**.
- Captain: flag-football (2018 co-rec champions)

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

Volunteering

2019 **Campaign Against Cancer**.

Charity Runs

2017-Present **Pirates Home Run 5K 10K**.
- https://www.mlb.com/pirates/community/race

2017-Present **The Great Race**.
- http://www.rungreatrace.com/

2016-Present **Pittsburgh Penguins 6.6K Run & Family Walk**.
- https://www.mariolemieux.org/our-events/6-6k-run-and-family-walk/