

# Gregory Faletto

Email: [Gregory.Faletto@marshall.usc.edu](mailto:Gregory.Faletto@marshall.usc.edu)

Website: <https://gregoryfaletto.com/>

GitHub: <https://github.com/gregfaletto>

## Education

Ph.D. Data Sciences and Operations (Statistics Track), University of Southern California Marshall School of Business (2018 - )

Advisor: Dr. Jacob Bien.

B.A. Physics, Washington University in St. Louis (2009).

Graduated Phi Beta Kappa.

Graduated with College Honors in Arts and Sciences.

Recipient of William George Eliot Scholarship.

Recipient of National Merit Scholarship.

## Publications

Salganik, M. J. *et al.* (2020). **Measuring the predictability of life outcomes with a scientific mass collaboration.** *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403.

<https://doi.org/10.1073/PNAS.1915006117>

## Current Research

Feature Selection Research Project (2018 - )

Developing a stable feature selection procedure for predictive modeling in the setting of highly correlated covariates.

Industry Collaboration (2019 - )

Collaborating with a private firm on an applied novel research problem (see below).

## Industry Employment and Collaborations

Undisclosed Technology Firm (2019 - )

Developing a novel recommendation system for a non-standard problem.

[ZipRecruiter](#), Data Analytics Research Intern (2017 - 2018)

Developed a model that predicts whether a job seeker will apply to a given job listing to infer preferences for forms of compensation. Chose observational study design and model form. Accessed data via **SQL**. Conceived outcome metric, constructed from raw data. Worked mostly independently under light supervision from ZipRecruiter and Prof. Sanjog Misra (University of Chicago).

[Live Nation](#), Machine Learning Intern (2017)

Working remotely, independently created an algorithm in R to predict future concert set lists of bands based on past set list data. Code and description: <https://github.com/gregfaletto/setlistpredictor>.

## Independent Projects

[2020 COVID-19 Computational Challenge](#) (2020; Hosted by RMDS and the City of Los Angeles)

Awarded second place (tied) in the competition. Using daily neighborhood test results across Los Angeles, developed a GLM panel data model (Poisson autoregressive count model) to forecast new infections of COVID-19 a day ahead in each neighborhood in Los Angeles. Used infection predictions to generate daily forecasted risk scores for each neighborhood. Joint work with [Mohammad Mehrabi](#). Full solution and writeup available at <https://grmds.org/2020challenge#solutions>.

[Orange County R Users Group Hackathon 2019](#) (2019)

Participated in team that won "Best Model" at the Orange County R Users Group Hackathon 2019. Over the course of two days, developed a model associating health outcomes in California counties with levels of water pollutants. More information at [gregoryfaletto.com/2019/05/19/our-entry-in-the-ocrug-hackathon-2019/](http://gregoryfaletto.com/2019/05/19/our-entry-in-the-ocrug-hackathon-2019/).

[Fragile Families Challenge](#) (2017)

Cleaned data and developed models to predict three different outcome variables in a high-dimensional social science data set (2,121 observations, nearly 13,000 covariates) with missing data. Code and description: <https://github.com/gregfaletto/fragilefamilies>. Resulting publication listed above.

## Invited Talks

[Copenhagen Workshop On Algorithmic Fairness \(University of Copenhagen\)](#) (Nov. 13 2020)

Talk title: "Fair Decisions, Hard and Soft." Presentation on accounting for more than one measure of algorithmic fairness. Joint work with [Prof. Kenneth Silver](#) (Trinity College Dublin).

[IM Data Conference 2020](#) (Nov. 3 2020)

Talk title: "Data-Driven Forecasts of Local COVID-19 Risk in Los Angeles." Presenting our winning solution from the 2020 COVID-19 Computational Challenge hosted by RMDS and the City of Los Angeles. Joint work with [Mohammad Mehrabi](#).

[Orange County R Users Group](#) (May 2019)

Presented models that won "Best Model" at the Orange County R Users Group Hackathon 2019.

## Programming Languages/Software Packages

R (proficient), SQL (intermediate), Matlab (intermediate), Excel (intermediate), Python (some experience), Stata (some experience).

## Coursework

### *Graduate Coursework (USC)*

Fall 2020:

[Econ 615: "Applied Econometrics"](#) (Instructor: Dr. Cheng Hsiao).

[Math 525A: "Real Analysis"](#) (Instructor: Dr. Kenneth Alexander).

[DSO 699: "Statistics Methods on Causal Inference and Deep Learning"](#) (Instructor: Dr. Yingying Fan).

Spring 2020:

[Math 425: "Fundamental Concepts of Analysis"](#) (Instructor: Dr. Andrew Manion).

[DSO 677: "Dynamic Programming and Markov Decision Processes"](#) (Instructor: Dr. Paat Rusmevichientong).

Fall 2019:

[Math 547: "Mathematical Foundations of Statistical Learning Theory"](#) (Instructor: Dr. Steven Heilman).

[Math 541B: "Introduction to Mathematical Statistics"](#) (Instructor: Dr. Stanislav Minsker).

[GSBA 604: "Regression and Generalized Linear Models for Business Applications"](#) (Instructor: Dr. Gourab Mukherjee).

Spring 2019:

[Math 541A: "Introduction to Mathematical Statistics"](#) (Instructor: Dr. Steven Heilman).

[ISE 620: "Foundations of Stochastic Processes"](#) (Instructor: Dr. Sheldon Ross).

[DSO 607: "High Dimensional Statistics and Big Data Problems"](#) (Instructor: Dr. Jinchi Lv).

Fall 2018:

[Math 505A: "Applied Probability"](#) (Instructor: Dr. Sergey Lototsky).

[Econ 613: "Economic and Financial Time Series I"](#) (Instructor: Dr. Hashem Pesaran).

[EE-588: "Optimization for the Information and Data Sciences"](#) (Instructor: Dr. Mahdi Soltanolkotabi).

### *Selected Online Courses (MOOCs)*

["Databases: Introduction and Relational Databases"](#) (Widom, Jennifer), Stanford Lagunita. *Completed January 2017.*

["Intro to Python for Data Science"](#) (Schouwenaars, Filip), Datacamp. *Completed December 2016.*

["Data Visualization with ggplot2 \(Part 1\)"](#) (Scavetta, Rick), Datacamp. *Completed December 2016.*

## Service

Served as a reviewer for Sankhya: The Indian Journal of Statistics (Series A).