Sam **Gass**

Data Scientist

1434 Carleton Street, Berkeley, CA

617-750-1121 | sdg2145@columbia.edu | sdgass13.github.io | sdgass13 | sam-gass-3488aa94

Education

Columbia University	*New York, New York*
M.A. in Quantitative Methods in the Social Sciences	*August 2017 - December 2018*

- GPA of 4.07. Recipient of QMSS Departmental scholarship

Syracuse University	*Syracuse, New York*
B.A. in Political Science and Economics	*August 2009 - May 2013*

- Graduated Magna Cum Laude with Renee Crown University Honors

Experience

Analyst Institute	*Washington, D.C.*
Data Visualization and Web-App Engineer (Contractor)	*December 2018 - June 2019*

- Built full-scale, interactive **Shiny** web app and dashboard for Analyst Institute. Application includes complex back-end which runs statistical analysis based on user input, data filtering, a connection to a remote **SQL DB**, and interactions with **Django** web-app. Front-end incorporates interconnected click-throughs and data visualizations built with **JS** giving users the ability to explore meta-analysis results, results of individual experiments, and connect to AI experiment repository.
- Ran accuracy testing on remote **AWS SQL DB**, wrote **SQL** loop to check against correct data and replace if missing or inconsistent.

Analyst, 2018 Campaign Cycle	*May 2018 - November 2018*

- Served in the primary technical role on a team that planned, supervised, and analyzed the results of randomized field experiments.
- Preparation Phase: Wrote **SQL** and **R** code pipelines to ensure proper data digestion, cleaning, and storage. Randomized universe using stratified or block randomization, ensuring balance across treatment groups.
- Analysis Phase: Conducted causal inference statistical analysis of the outcome data using **R** – identified correct models to use (**logistic regression, 2SLS, IV's**), disaggregated effects by subgroup, validated results and models to ensure soundness of analysis. Interpreted the results and presented recommendations based on experimental outcome data. Created visualizations for final presentation in **R** using **Ggplot**, drafted results memo and, along with research manager, presented results to client.
- Maintained and updated AI codebase, which is a series of **R** packages that organization uses for analysis. Built unit tests for packages to ensure functionality.

NBC-Universal	*New York, New York*
Data Scientist (Contractor)	*Jan. 2018 - May 2018*

- Used **Data Bricks** and **pySpark** environment to compile and merge multi-gigabyte datasets in order to calculate accuracy of linear forecasts. Created **AWS** hosted pipeline and automated the process for future analyses.
- Saved team hours of manual input by writing algorithm in **Python** that recreated lower level spreadsheets that were used in other departments, but unavailable to ours, from higher level aggregations we received. Automated code for future use.
- Independently researched and proposed using new deep learning techniques to improve accuracy of forecasts. Algorithms included **CNN's, RNN's, and LSTM's**. Tested methods on sample data and presented to senior staff.

The Public Good	*New York, New York*
Data Consultant	*May 2017 - Jan 2018*

- Built hosted dashboard for client using **R-Shiny**. Wrote backend that scraped, filtered, and cleaned data from multiple API's. Saved company hours of independent research.

The Institute for Social and Economic Research and Policy at Columbia University	*New York, New York*
Research Assistant	*January 2017 - September 2017*

- Worked with Professor Costas Panagopoulos on data aggregation, analysis, and writing for an upcoming book on challenger quality in Congressional races in the United States.
- Created lesson modules for web scraping, API use, and data manipulation in both **R and Python** for QMSS program director; code and notes will be used in an intro programming class for QMSS students.

Relevant Skills

Programming

Languages	**R, Python, SQL, Stata, HTML, CSS, JavaScript, unix**
Coding	**pySpark, Data Bricks, Algorithm's, API's, dplyr, pandas, GGplot, AWS, Django**

Data Skills

Modeling	**Linear Regression, GLMs, Time-Series, Panel Modeling, Multi-Level Modeling, 2SLS, IVs**
Data Viz	**Dashboards, Tableau, Shiny, d3, GGplot, Matplotlib, Plotly**
ML	**Neural Networks, PyTorch, Random Forests, SVMs, Classification, Reinforcement Learning**