### Overview

TwoRavens is a Web-based system for statistical analysis. The goal is to allow the domain expert, in concert with our system, to complete a high quality, predictive and interpretable model without a statistical expert. To do so, the system facilitates intuitive machine learning and model interpretation, model discovery, and data exploration. As our intelligent back-end automatically seeks interesting relationships in the data and builds models to predict outcomes, researchers impart substantive knowledge about their data and own research questions to guide the automated generation of AI assistance for data analysis in an interactive paradigm we call human-guided machine learning.

TwoRavens for Event Data is an offshoot application, designed for researchers to easily structure raw event data into usable time-series formats. Researchers can browse openly available event datasets, construct queries to select types of events and sets of actors, view and download resulting time-series data, and export data to our main system for AI assisted analysis.

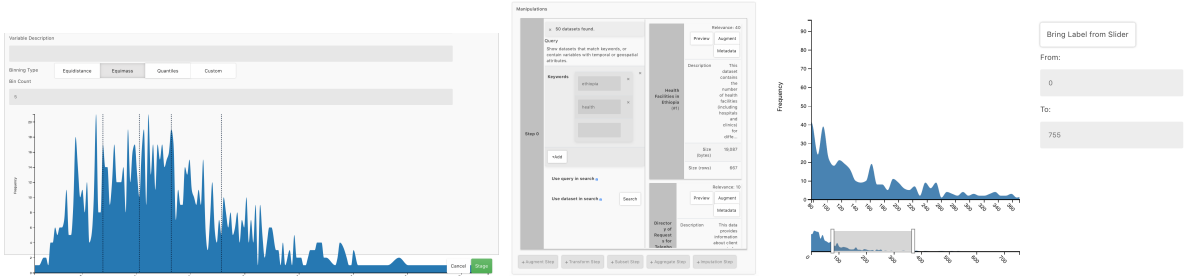
### Philosophy

In the very best research settings, where there are collaborations between domain experts and data scientists or statistical experts, exploration into the data is a joint venture where statisticians drive the computational machinery of analysis, but are directed to the interesting features by the knowledge of the domain expert. Our belief is that you can automate much of what the statistician brings, but the domain expert remains central to the task. Thus, our goal is to augment the domain expert, leading to the construction of high quality, impactful and interpretable models.

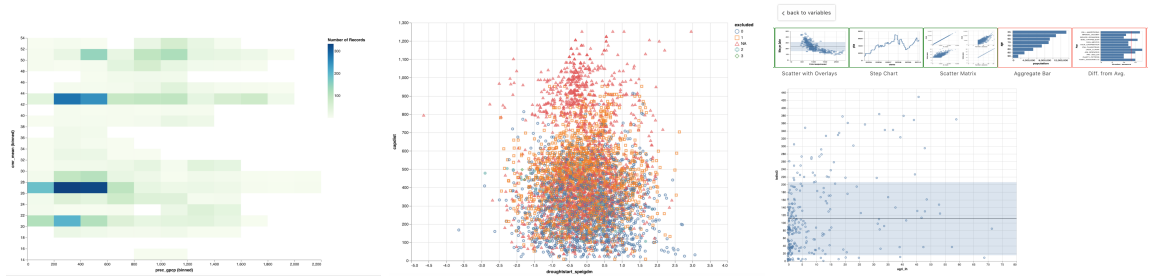
## Features

A core concept of TwoRavens is *mode*. Different modes signify different main goals a user might be attempting in the current part of their analysis. When the system is in different modes, the features available to the user are swapped or modified to better work for that intended task.

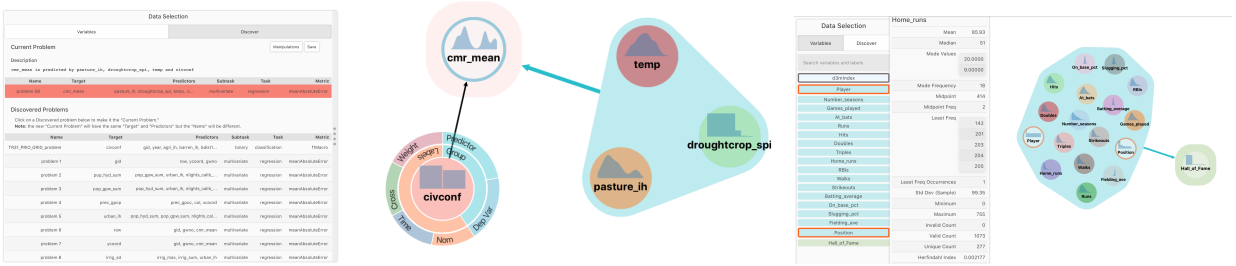
**Dataset Mode** gives different ways for a user to bring data into the system. Here a user can upload new datasets, switch between datasets already available, and access the Datamarts. Datamarts are a way to augment current datasets with additional relevant data to a problem.



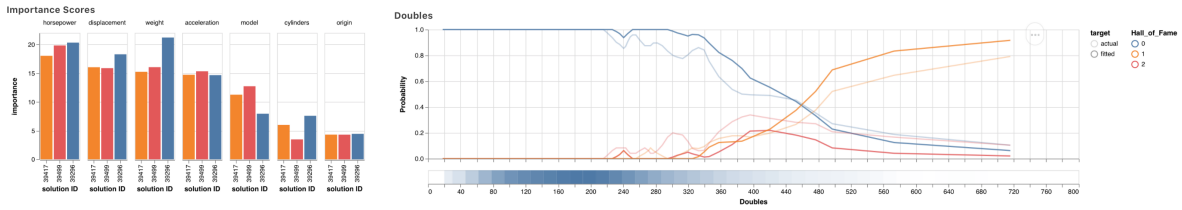
**Explore Mode** allows a user to perform exploratory data analysis by intuitively generating visualizations of relationships between variables.



**Model Mode** allows a user to construct models of interest, specify a target to predict, explore the automatically discovered problems suggested by the system, override any metadata created for features and change the problem characteristics of the model.




**Results Mode** allows a user to automatically generate machine learning solutions for a user constructed model, explore previously constructed models to gain insights into model performance, and understand the relative importance of features within and across different machine learning solutions.




## Investigators

James Honaker (`james@hona.kr`) is a Research Associate at the Center for Research on Computation and Society in the Computer Science department at the Harvard John A. Paulson School of Engineering and Applied Sciences. He studies machine learning and statistical methods for quantitative social science. He earned his PhD in Government from Harvard University in 2004.

Vito D'Orazio (`dorazio@utdallas.edu`) is an Assistant Professor of Political Science at University of Texas at Dallas. He studies conflict forecasting and computational methods. He earned his PhD in Political Science from Penn State in 2013 and was a postdoc at Harvard from 2013-2015.

For more information, see our system demos at `2ra.vn/demos/`, our documentation at `tworavens.github.io/TwoRavens/`, and feel free to email us with any questions.