Project Portfolio
Elliott Evans

✉ evanselliott1@gmail.com
⍥ elliottevans
in elliott-evans
Resume

Michigan HS Graduation Rate Study (UMich Fall Sem 2016)

- **Full Work:** [Here](#)
- **Goal:** Identify spatial patterns in Michigan High School graduation rates.
- **Results:** Two counties in Michigan had "significant spatial random effects", i.e. two counties in Michigan had lower graduation rates than we would expect after accounting for economic status and race.
- **Tools Used:**
  - Bayesian hierarchical models
  - Conditionally Autoregressive Models (CAR)
  - Spatially Autoregressive Models (SAR)

Figure: Counties in red have lower than expected graduation rates

Significant Spatial Random Effects

- ■ Sig. Negative
- □ Not Significant
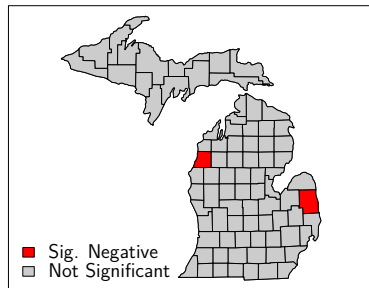
Probability of Heart Disease (UMich Winter Sem 2017)

- **Full Work:** [Here](#)
- **Goal:** Examine how different factors contribute to the probability of obtaining heart disease.
- **Results:** We found greater than nine-in-ten chances that increasing age, blood pressure, and cholesterol all increase odds of obtaining heart disease.
- **Tools Used:**
 - Bayesian logistic regression

Figure: Posterior densities for risk factors of heart disease
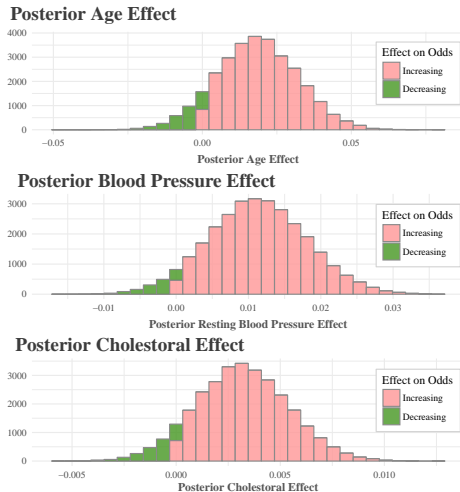
- **Full Work:** Here
- **Goal:** Identify risk factors for low rate mathematics achievement growth.
- **Results:** The biggest risk factors for mathematics achievement were race, education of the parents, and the child's participation in the government Head Start program
- **Tools Used:**
  - Hierarchical linear models

Figure: Avg. growth trajectory in Mathematics achievement over time

Average MATH Score Growth Trajectory
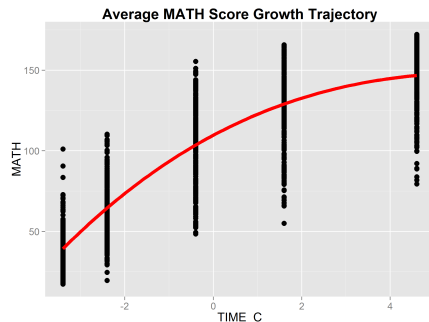
Classifying Future Income Brackets (UMich Winter Sem 2017)

- **Full Work:** Here
- **Goal:** Create a set of models that can accurately determine median future income quintiles for students in US colleges.
- **Results:** A random forest model was most successful with an out-of-bag error rate of 37%.
- **Tools Used:**
  - Support Vector Machines
  - Random forests
  - Principal Component Analysis
  - Linear/Quadratic Discriminant Analysis
  - Clustering

Figure: Tuning the number of variables on which to split nodes in the random forest model

Random Forest Parameter Tuning
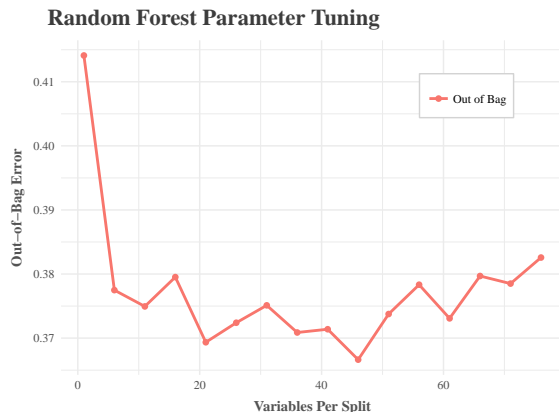
General Election 2016 Forecasts

- **Full Work:** [Here](#)
- **Goal:** Provide election analysis in a way that could be understood by everyone, in a nonpartisan way.
- **Results:** Predictions said Clinton would win popular vote *and* electoral college, when she won popular vote and *lost* electoral college.
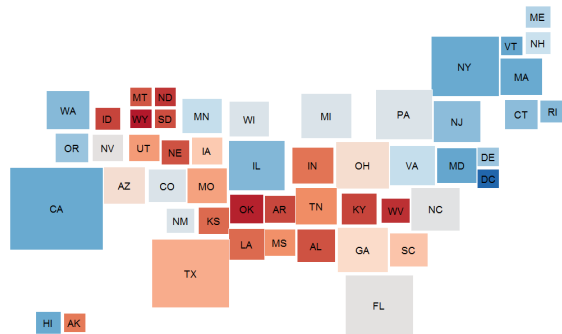- **Tools Used:**
 - Nearest-Neighbor
  - Simulation

Figure: Predicted election cartogram with state sizes reflecting electoral votes

Hall of Fame Prediction (Northwestern Fall Qtr 2013)

- **Full Work:** <u>Here</u>
- **Goal:** Create a model to accurately predict whether or not a given player will make the baseball hall of fame.
- **Results:** The model with the best possible predictive power used the statistics WAR, BsR, and wRC+.
- **Tools Used:**
  - Multi-Dimensional Gaussians

Figure: 2-Dimensional Gaussian model for Hall-of-Famers using the stats WAR and wRC+
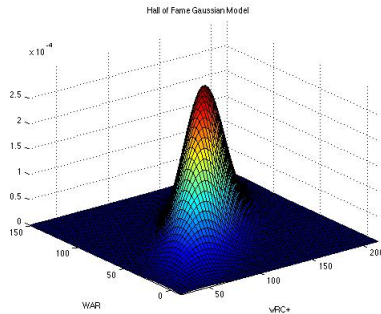
The HoF Case for Mussina & Schilling (Numberfire, 2014)

- **Full Work:** [Here](#)
- **Goal:** Determine whether pitchers Mike Mussina and Curt Schilling should be grouped together with other Hall of Fame pitchers from the 1990's.
- **Results:** Hierarchical clustering and MDS showed that Mussina and Schilling were comparable to other Hall of Famers of that time period.
- **Tools Used:**
  - Multi-dimensional Scaling (MDS)
  - Hierarchical Clustering (complete-linkage)

Name	K/9	BB/9	HR/9	RA9-WAR	CG
Curt Schilling	9.5	1.79	0.99	61.6	64
Mike Mussina	7.75	2.01	0.99	50.7	38
Pedro Martinez	10.53	2.26	0.7	73.7	41
Randy Johnson	11.95	2.61	0.83	72.7	57
Greg Maddux	6.45	1.32	0.67	68.6	45
Tom Glavine	5.44	3.1	0.75	53.5	23
Roger Clemens	9.07	3.29	0.75	56	23

Pitch Framing Analysis (Northwestern Spring Qtr 2015)

- **Full Work:** [Here](#)
- **Goal:** Determine the value of a "pitch frame" in terms of runs and wins, then find the league's best and worst pitch–framers.
- **Results:** A successful pitch frame is worth about seven-hundredths of a run and about seven-thousandths of a win. Brian McCann was the best, worth about 7.8 wins above average in five years.
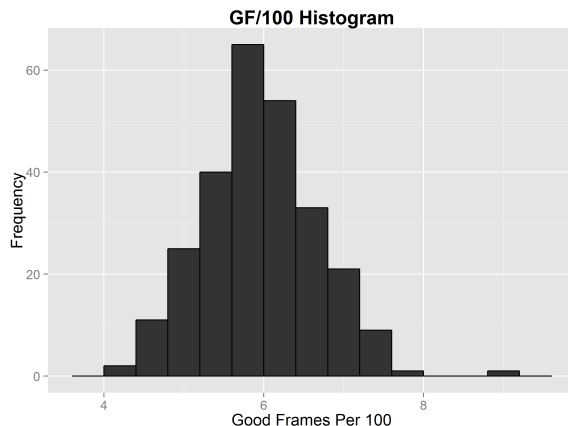- **Tools Used:**
  - Relational Database Management

Figure: Distribution of the number of "good frames" a catcher makes for every 100 pitches caught