VISUALIZATION OF HEALTHCARE QUALITY IN AMERICA

Rabih Atallah David Baugher Robert Geiger Ashli Taylor
CSE 6242, Georgia Institute of Technology

Summary

Our project offers an improvement on the current best practice in healthcare analytics for evaluating the quality of patient care. We tested whether the current standard of analyzing healthcare could be improved by including publicly available data for socioeconomic factors. The results proved that our proposed method was statistically superior. Our most significant findings are summarized in a user friendly and interactive visualization to the benefit of anyone who wants a comprehensive review of the current status of healthcare in America.

Motivation

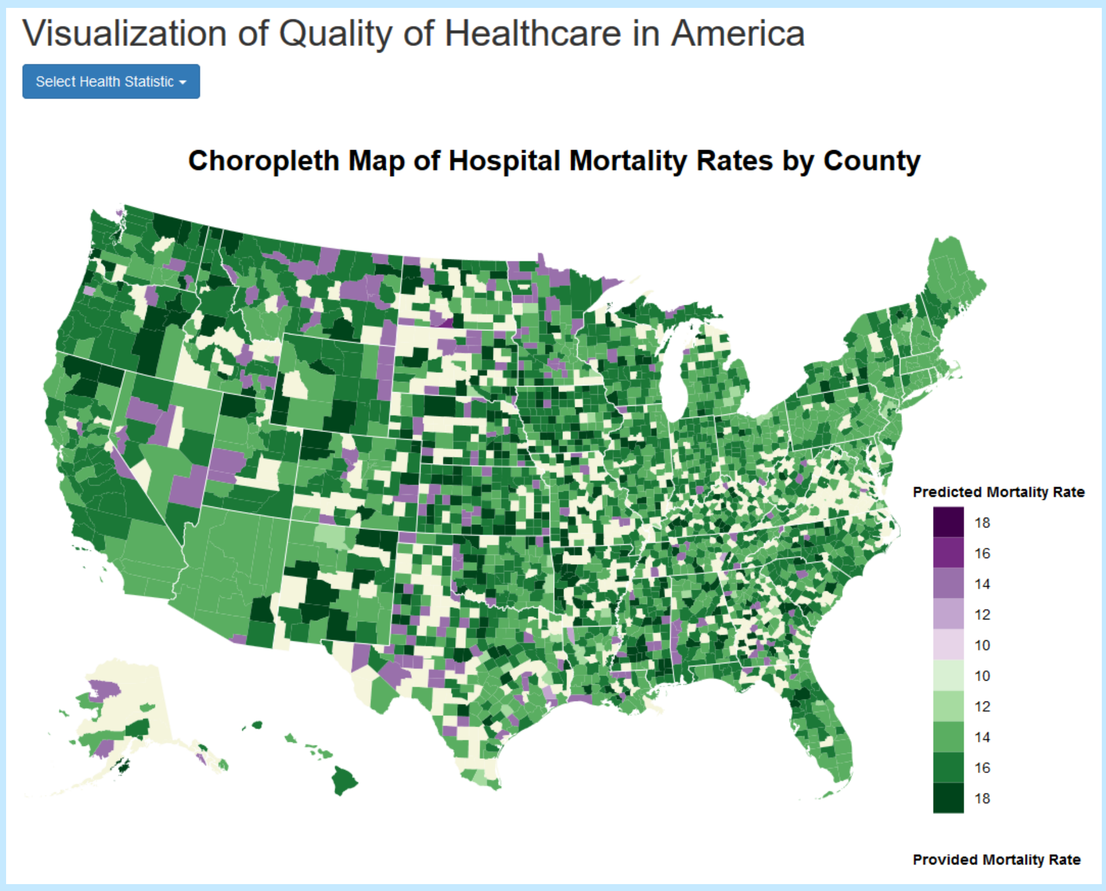
America far outpaces the developed world in healthcare expenditure, however the quality of our healthcare does not match this ranking. Research suggests that socioeconomic factors should be included when examining ways to improve healthcare quality. Healthcare analytics conventionally use a framework developed in 1966 to evaluate the quality of a medical facility. We believe that analysis methods should evolve at the same pace of industry advancements.

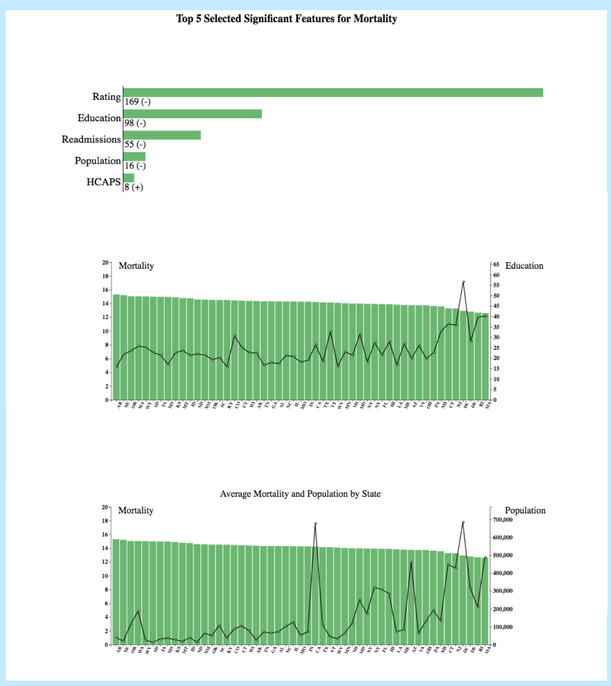
Quantifying Quality

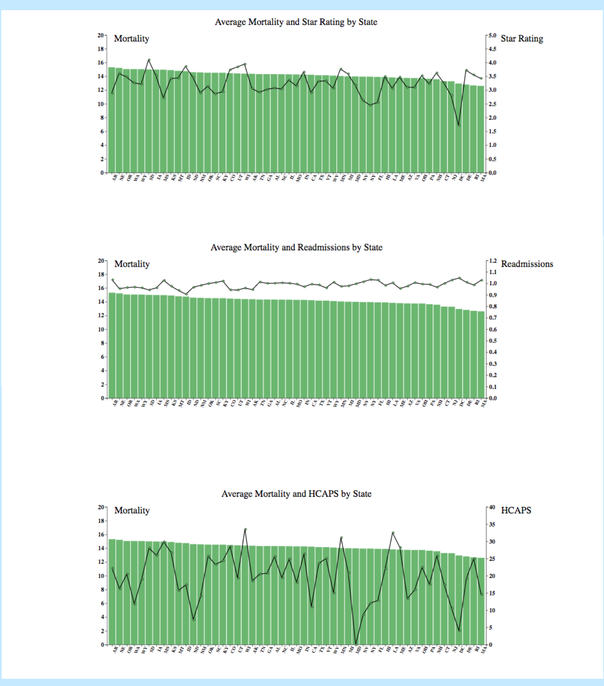
Healthcare analytics typically assess quality using different types of outcome and process measures. We focused on three such measures for the scope of this project:

- Mortality Rates – Assessed for three of the most serious pulmonary and cardiovascular events: heart attack, heart failure, and pneumonia.
- Readmission Rate – The percentage of patients who are readmitted within 30 days of being discharged.
- Facility Rating – Hospitals rated on a 1-5 scale with 5 indicating the highest level of performance.

Visualization of Quality Across America







In this example, the user selected mortality rates as the feature of interest. A choropleth map of the United States is generated that shows the values of the observed and predicted mortality rates of each county. Counties highlighted in white do not have hospitals to report on. Also shown on the visualization pane is a horizontal bar chart that orders of the top five features that are of statistical significance to mortality. They are quantified by how much they reduce the Akaike Information Criterion and show whether they positively or negatively influence mortality. Here, an increase in education was found to be consistent with a reduction in mortality rates. The next five plots show how the selected features compare against mortality across all 50 states.

Approach

1. Expand upon the standard of measures currently used to evaluate quality of healthcare by incorporating socioeconomic measures posited by our conducted literature survey.
2. Broaden the scope of analysis by aggregating and modeling patient and hospital facility data by county.
3. Replicate the current standard of modeling and test the null hypothesis that the two groups of models are statistically equivalent.
4. Visualize the significant findings via an engaging interface that presents the most significant findings.

Data

We obtained publicly available data from the Department of Health and Human Services for 4,800 hospitals and parsed them for the 3,142 counties in America.

Measure	Description	Characteristics
Process	Specific steps that are undertaken during patient care that produce positive or negative patient outcomes.	Missing data for counties without hospitals and scant reporting 60% Completeness
Outcome	Reflect the impact of the healthcare service or intervention on the health status of patients	Missing data for counties without hospitals and low number of patients to record outcomes 75% Completeness
Socioeconomic	College Education, Household Income, Poverty Rate, Race, Age, Population, Gender	Autocorrelation in Age, Race, and Gender 100% Completeness

Models

We first used a random forest model with hyperparameter tuning on each process and outcome feature to impute missing values. We then analyzed the data using the following models:

- Random Forest with recursive feature selection using the Scikit–learn library in Python to address correlated features.
- Multivariate Regression with mixed stepwise selection in R to identify the combinations of features that best captured statistical variance.

Experiments

Our experimental analysis of our null hypothesis was robust. We created 16 models to analyze 8 pairs of complementary subsets of features.

Dependent	Base Predictors	Test Predictors
Mortality Rate	Process	Process + Socioeconomic
	Process + Readmission Rate	Process + Socioeconomic + Readmission Rate
Readmissions	Process	Process + Socioeconomic
Facility Rating	Process	Process + Socioeconomic

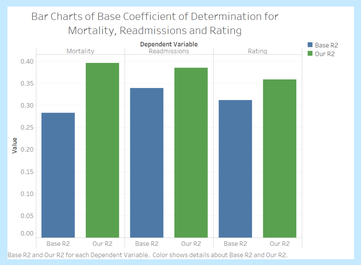
We tested our null hypothesis for the two sets of models using three statistical measures:

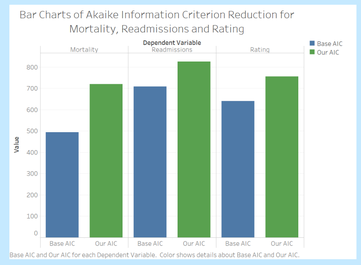
- Analysis of Variance (ANOVA)
- Akaike Information Criterion (AIC)
- Adjusted Coefficient of determination (Adjusted R^2)

Results

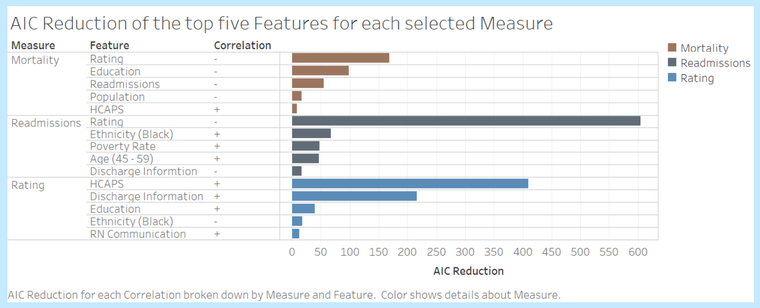
We proved that the statistical power of healthcare analytics can be improved using socioeconomic features by showing that our model

- ✓ Reduces AIC
- ✓ Increases adjusted R^2
- ✓ Rejects the ANOVA null hypothesis





We identified the most important features for understanding mortality rates, readmission rates, and facility ratings across America using the recursive feature selection tool in the Scikit–learn library. They are ranked in order of how much they reduce the AIC values of their respective models.



For each measure, we identified at least two variables of statistical significance that would not have otherwise been noted using the conventional analytics model. This kind of information has enhanced policy implications as more emphasis is placed on overhauling our healthcare system.