An Interactive Application for Identifying and Explaining Mental and Behavioral Health Provider Shortages By Population

Team 03: Mohit Aggarwal, James Boyle, Nathan Cook, Andrew Doss, Suzi Pike (suzipike@gatech.edu)
CSE6242 - Fall 2018, Georgia Institute of Technology

PROBLEM STATEMENT

Studies [2,3] have shown that shortages of mental and behavioral health providers are limiting patients' access to required treatment. In addition to lower provider supply, sociodemographics such as age, race, income and insurance, are associated with access to care [1,4]. Our web application was developed to deliver insights to healthcare stakeholders [5] striving to understand and remedy the underlying causes of provider shortages in order to increase provider access for their communities.

DATASETS

We downloaded the following data:

1. Provider locations (source: National Plan and Provider Enumeration System (NPPES)) with 5725854 records, 6.7 GB on disk.
2. Urbanicity data (source: United States Department of Agriculture Economic Research Service (ERS)) with 72153 records, 4.6 MB on disk.
3. Sociodemographic features (source: United States Census Bureau American Community Survey (ACS) 2011, 2016) with 76200 records, 747.4 MB on disk.

APPROACH

1. Geocode provider addresses to census geographies.
2. Integrate provider data with urbanicity data and sociodemographic features data in a SQLite database providing dozens of attributes across 70,000 US census tracts and geographic coordinates for 200,000 plus mental health providers.
3. Explore data through overlaid histograms, box plots, correlation matrices and PCA.
   a. Identify interesting feature combinations and structures w.r.t. provider density.
   b. Identify collinearity and remove redundant features.
4. Select features using Random Forest, Lasso and Ridge regression models to identify subsets of features with the greatest predictive power of provider density.
   a. Different approaches provide distinct rankings of feature importance.
   b. Not reliable to estimate predictive power in presence of multicollinearity.
5. Transform features using PCA to understand how the relevant features relate to each other and whether a transformed feature space may enable useful visualizations (Fig. 1-2).
   a. Confirm presence of and address issue of multicollinearity in data.
   b. Transform original features into new features to reliably assess for independent predictive power.
6. Build an interactive visualization tool using D3.js with multiple usage pathways to identify provider shortage areas and derive unique patterns and associations within provider shortage areas (Fig. 3-5, 7-8).

Fig. 1: PCA - projecting original features on first two principal components

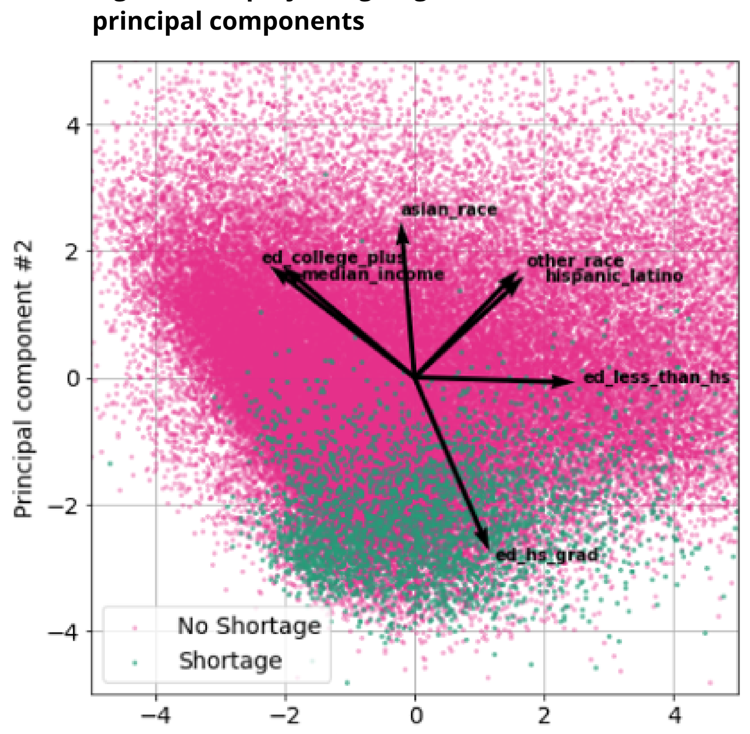


Fig. 2: Distribution of variance explained per Principal Component from actual dataset (green) vs uncorrelated data (orange)

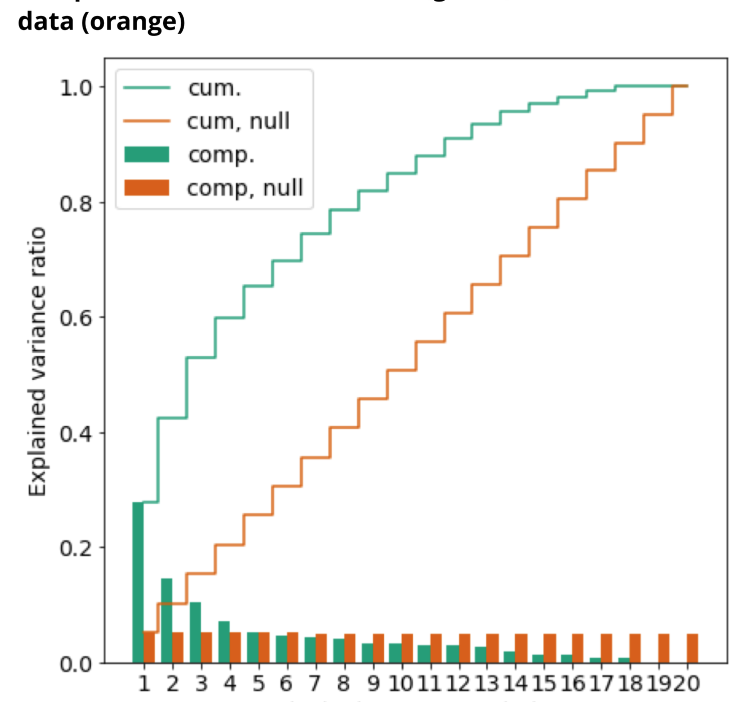
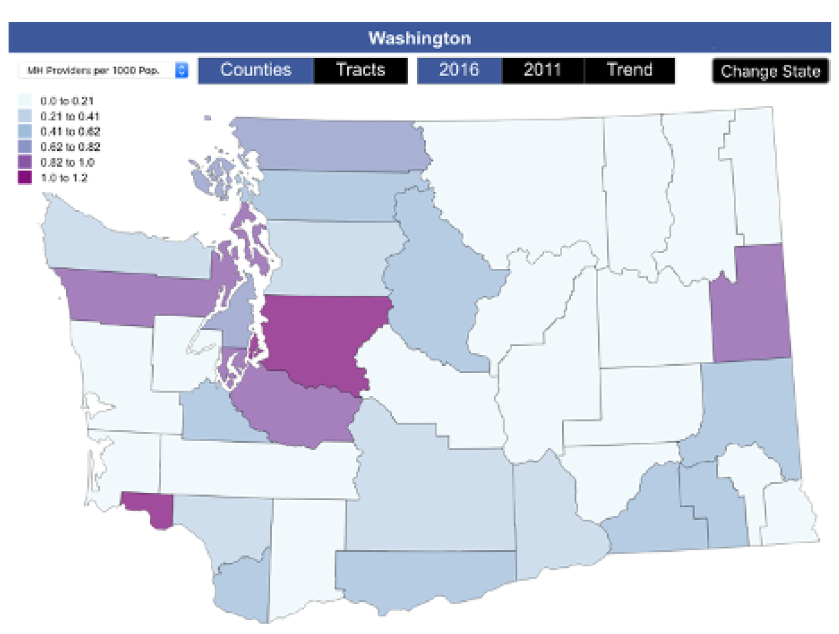


Fig. 3: Web Application showing map of United States allows user to graphically select any US state

Fig 4: Heat map of Washington showing provider density by county for 2016. User able to select from 19 sociodemographic features, at the county and tract level for 2011, 2016 and trend.



EXPERIMENTS AND RESULTS

1. Use k-means clustering to identify natural subgroups with varying provider density.
   a. Evaluate using elbow plots and gap statistics.
   b. No significant natural clusters identified at county or tract level.
2. Use PCA transformed dataset to run random forests, neural networks and regularized linear models to predict provider density.
   a. Models fit to all counties and tracts for selected states.
   b. Hyper-parameters tuned using 10 fold cross validation grid searches.
   c. Random forest and neural network are powerful but not enough data at state level to train non-linear models.
   d. Ridge regression outperforms with minimum MSE and error reduction (Fig. 6).
   e. Setup pipeline to automatically fit best ridge regression model and store predictive power rank for each principal component.
   f. Use scatter-plot to show which feature combination are more predictive of provider density (Fig. 8).

Fig. 5: Heat-map of provider density for Spokane County

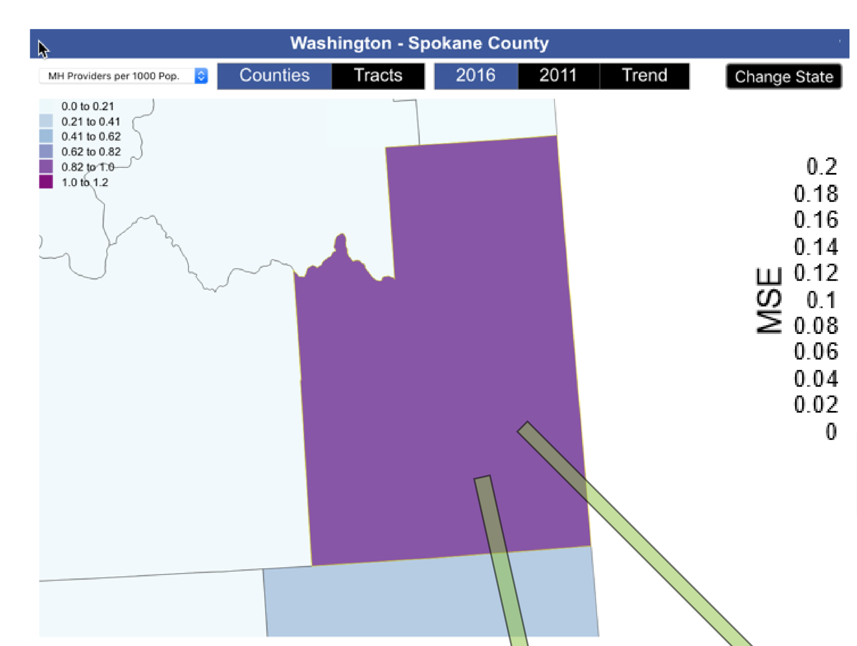


Fig. 6: Experimental Model Results

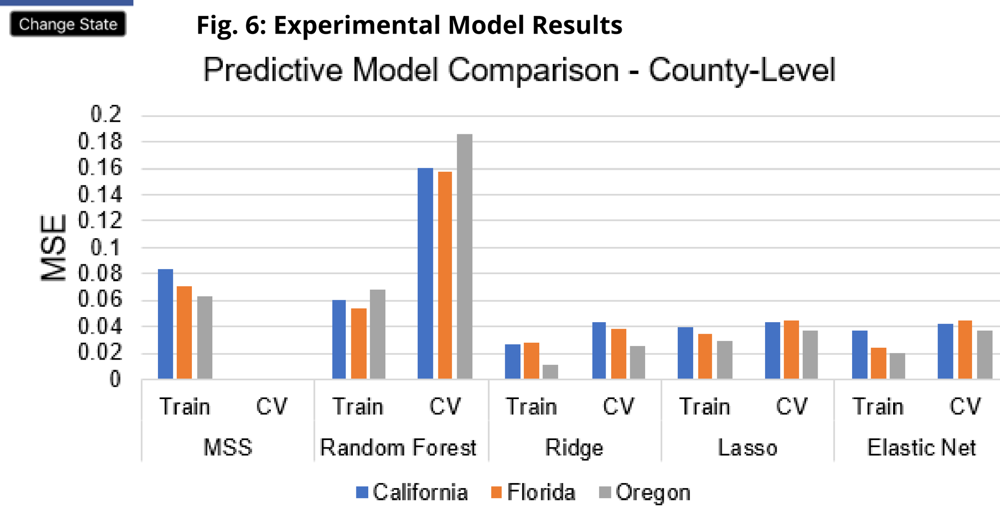
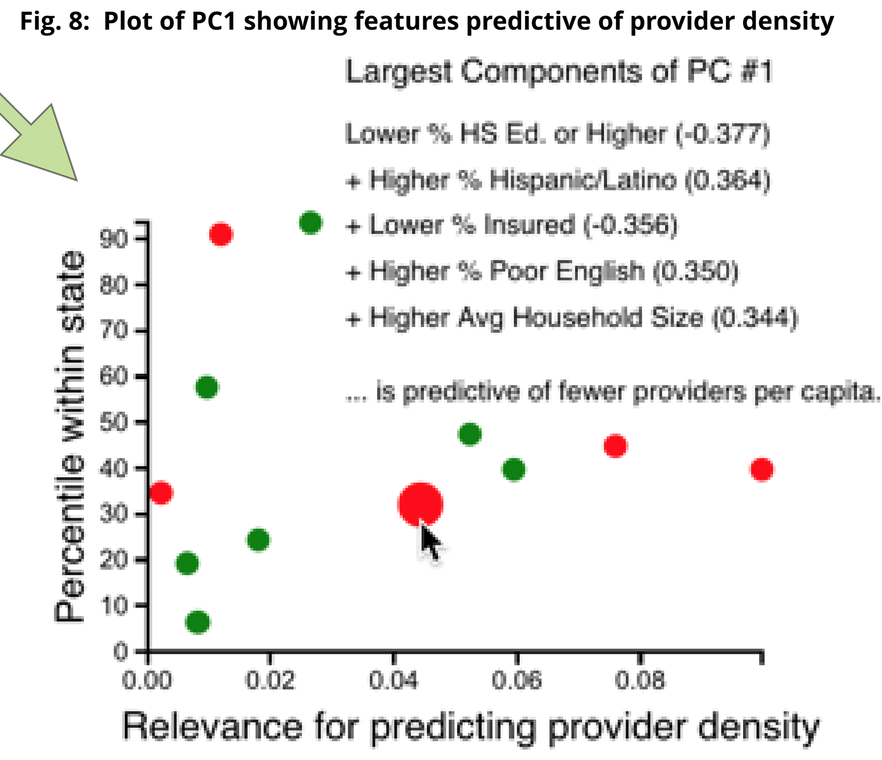


Fig. 7: Sortable table showing comparison of attributes for 2016, 2011 and Trend

Attribute	2016	2011	Trend
MH Providers per 1000 Pop.	0.90561	NaN	NaN
% Asian	2.3000	2.3000	NaN
% Bachelors	NaN	18.200	NaN
% Black	1.8000	1.7000	0.10000
% Education < HS	6.9000	7.5000	-0.60000
% Graduate Degree	NaN	10.400	NaN
% Hawaiian/PI	0.40000	0.40000	NaN

Fig. 8: Plot of PC1 showing features predictive of provider density



INNOVATIONS

1. Integrated multiple data sources to reveal relationships between a variety of socioeconomic measures and their association to provider shortages.
2. Provided feature rich interactive web application to analyze and compare community needs metrics across all US geographies.
3. Applied ML to identify complex patterns not accessible through descriptive statistics to explain relationships between sociodemographics and provider shortages.

CONCLUSION AND FUTURE WORK

1. Augments research capabilities of healthcare stakeholders by providing convenient access and summarization of multiple data sources.
2. Enables prioritizing interventions around addressing under-served areas with behavioral health provider shortages by focusing on relevant statistics.
3. Provides an extensible infrastructure capable of exploring and analyzing other aspects of healthcare supply and demand.

REFERENCES

[1] T. Arcury et. al., The Effects of Geography And Spatial Behavior On Health Care Utilization Among The Residents Of A Rural Region. *Health Services Research, 40*(1), 135–156 (2005).

[2] Cunningham et. al., Primary Care Physicians' Perspectives on Access to Mental Health Care. *Health Affairs,28*(3), 490-501(2009).

[3] S. Kataoka et. al., Unmet Need for Mental Health Care Among U.S. Children: Variation by Ethnicity and Insurance Status. American Journal of Psychiatry 2002 159:9, 1548-1555 (2002).

[4] R. Kathol et. al. , Barriers to Physical and Mental Condition Integrated Service Delivery. *Psychosomatic Medicine,72*(6), 511-518 (2010).

[5] Community Health Needs Assessment PeaceHealth (2016). Sacred Heart Medical Center. https://www.peacehealth.org/about-peacehealth/Pages/community-health-needs-assessment