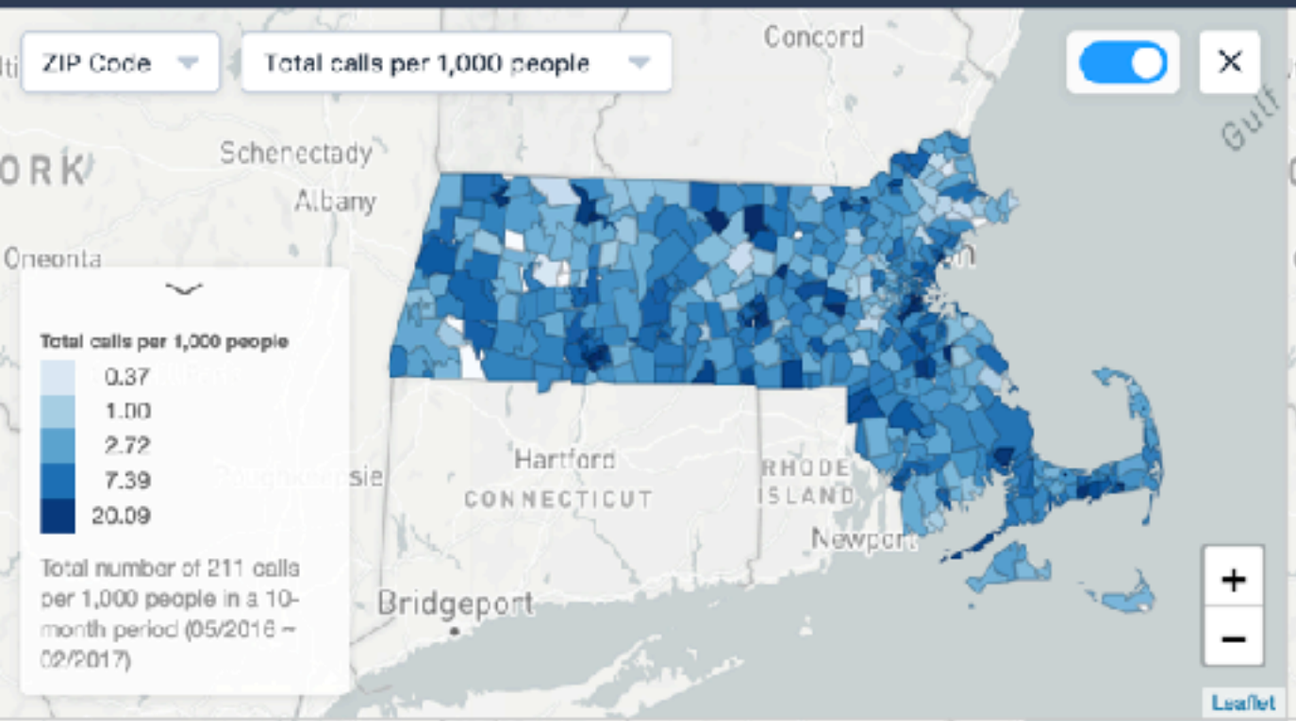
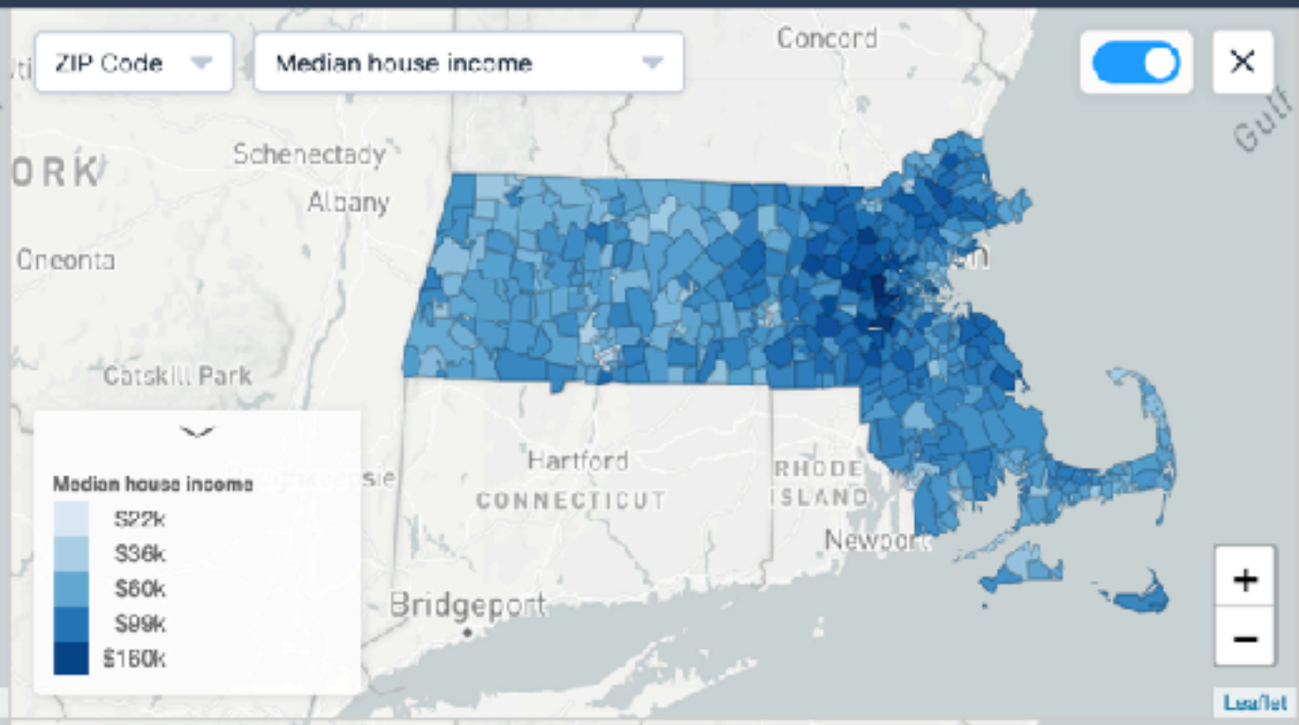
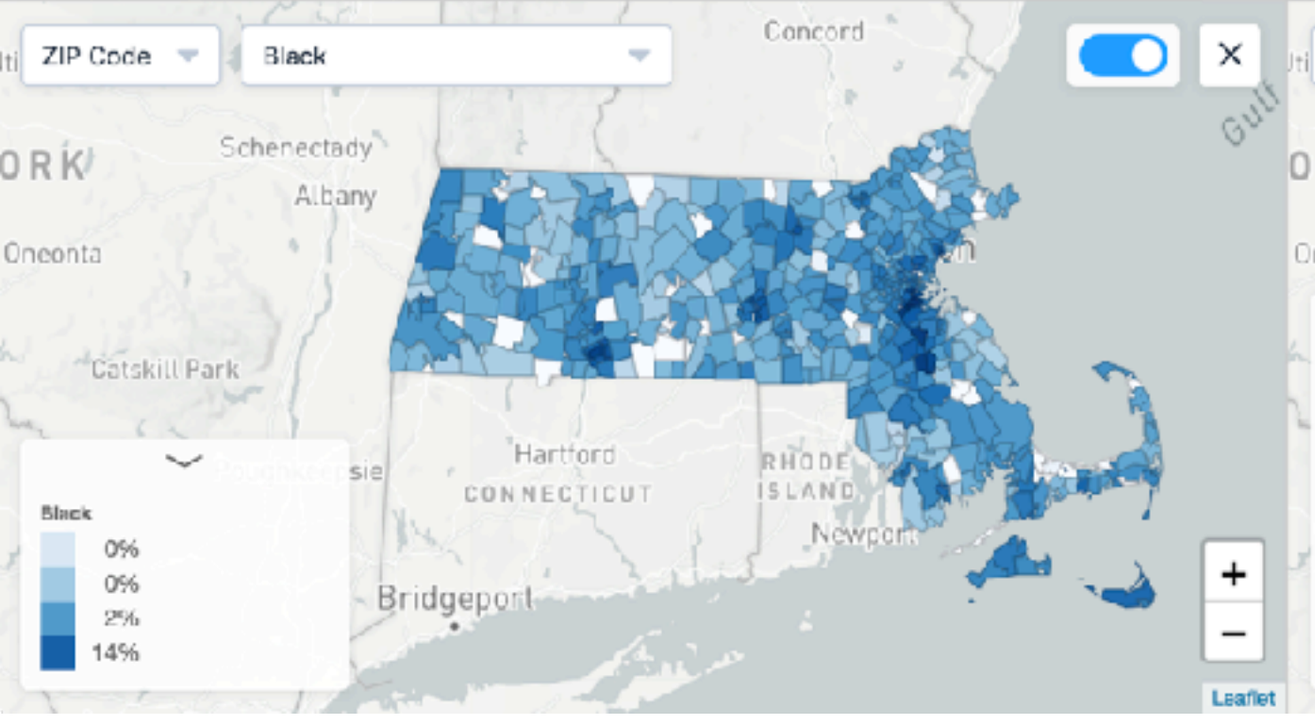
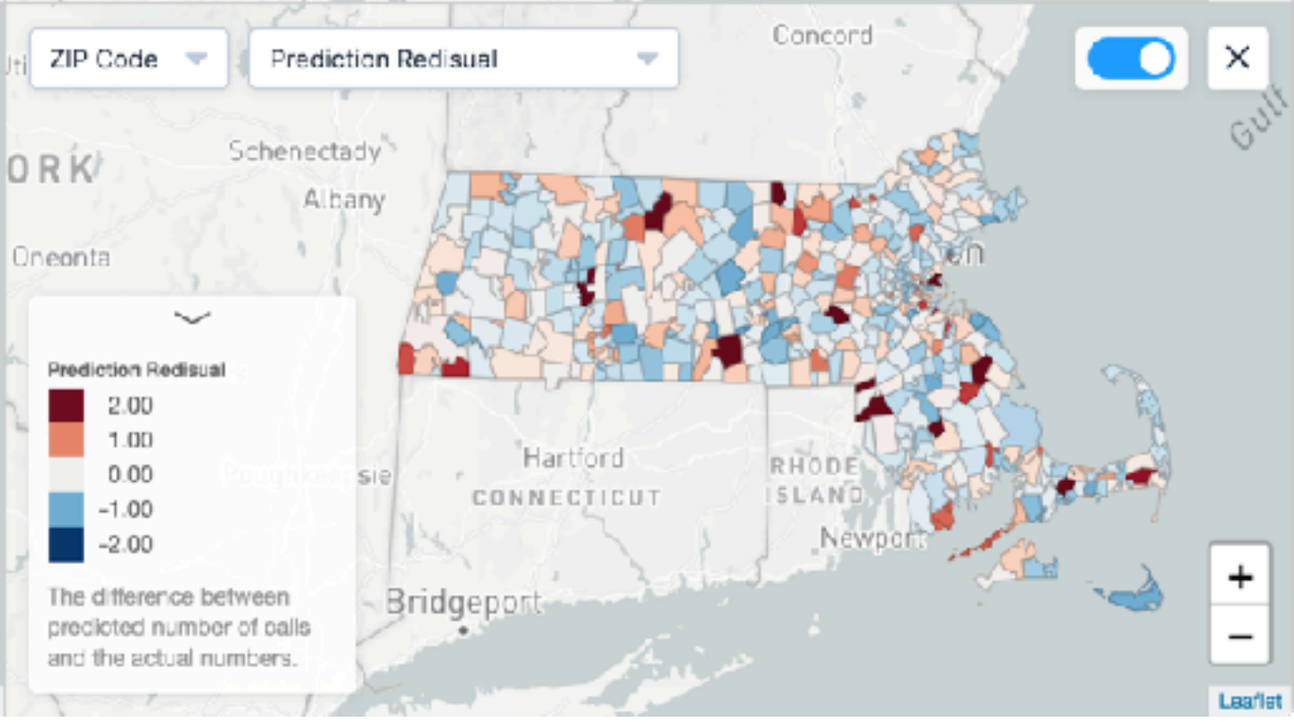
Demands Resources









Mass 211 Data Analysis

An Overview

Objectives

- Create tools and high-level models to *predict* and *evaluate* the volume of human service demands based on neighborhood characteristics.
- Investigate potential public health issues underlying 2-1-1 calls.
- Identify underserved neighborhoods where 2-1-1 services would add the most value.
- Track acute social problems where specific demands are high (e.g., housing for people with mental health issues.)

Challenges

- No way to personally identify callers
  - Only 5% of all calls have phone number recorded
- Lack of corroborative data
  - Follow ups? Multi-needs? Problem resolved? Does it persist?
  - How representative are the calls to the population?
- Relatively short period of reliable longitudinal data
  - Only about 1 year of data in new format
  - You don’t know whether it’s seasonal fluctuations or real pressure erupting
- Non-normality and loss of details in geographical aggregation
  - Highly-skewed distribution in both outcome variable and predictors
  - Concentration of population does not strictly align with population density
  - 90% of the population live in “urban” areas

Data (By)products

- Taxonomy of Human Services *by 2-1-1 LA County*
  - + Hand picked major service topics
  - an alternative: Open Referral (sounds nice at least!)
- 2-1-1 Map / Demands and Resources
- Massachusetts Health Status Indicators Dataset
- (Robustly) map ZIP code to (official) town names
- Configurable and extendable census data scraper
- Preliminary 211 data analysis results

Access to

Private Health Insurance

is a good proxy for

211 needs

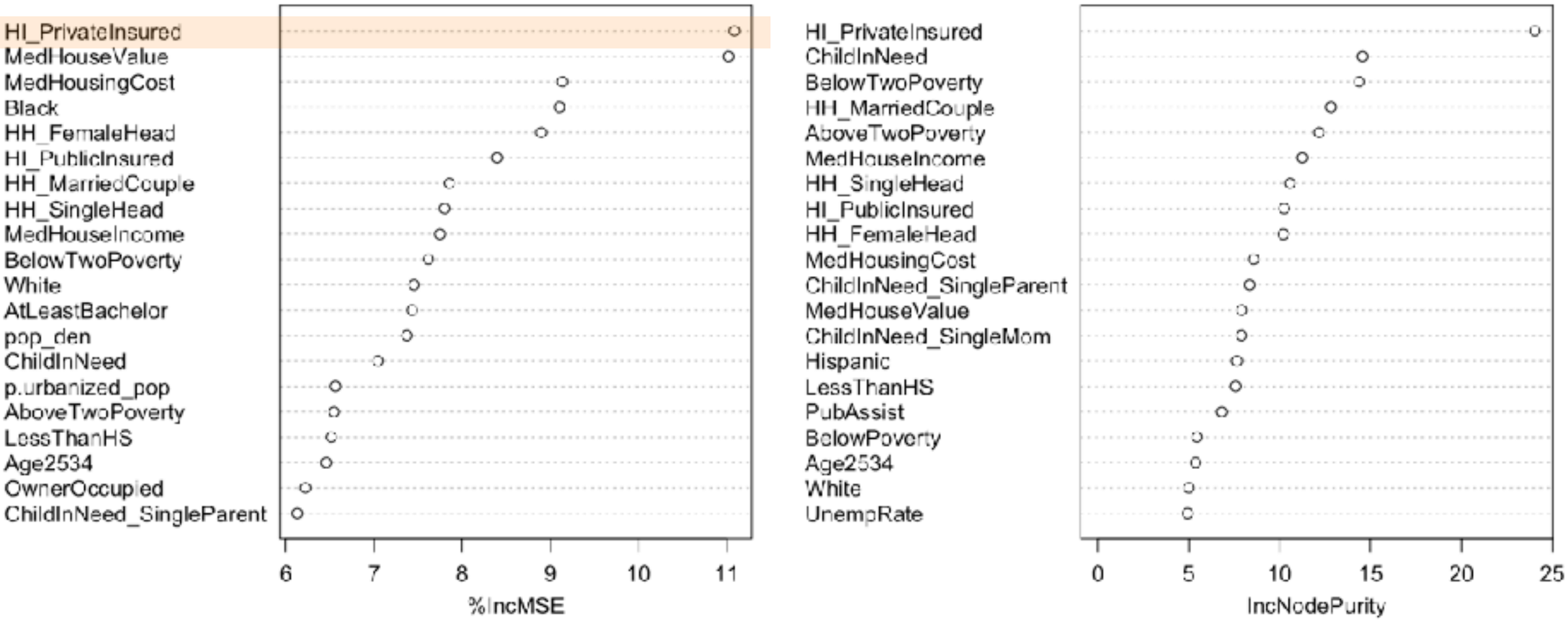
	Dependent variable:
	p_call
MedHouseValue	-0.003
	(0.003)
MedHouseIncome	-0.062***
	(0.018)
AtLeastBachelor	-8.191***
	(3.147)
Black	54.200***
	(4.554)
Hispanic	24.229***
	(3.316)
Constant	15.087***
	(1.153)
Observations	487
R^2	0.546
Adjusted R^2	0.542
Residual Std. Error	7.217 (df = 481)
F Statistic	115.881*** (df = 5; 481)
Note:	*p<0.1; **p<0.05; ***p<0.01

standard error

	Dependent variable:
	p_call
HI_PrivateInsured	-37.429***
	(5.167)
MedHouseIncome	-0.005
	(0.018)
AtLeastBachelor	-3.951
	(2.531)
Black	45.792***
	(4.527)
Hispanic	4.920
	(4.086)
Constant	38.496***
	(3.382)
Observations	490
R^2	0.577
Adjusted R^2	0.573
Residual Std. Error	7.008 (df = 484)
F Statistic	132.135*** (df = 5; 484)
Note:	*p<0.1; **p<0.05; ***p<0.01

HI_PrivateInsured even made other variables obsolete

As verified by RandomForest



No similar effect was found on **HI_PublicInsured**

Simple model works well enough

	Dependent variable:
	p_call
HI_PrivateInsured	-41.715***
	(2.960)
Black	46.130***
	(4.466)
MedHouseValue	-0.006***
	(0.002)
Constant	42.236***
	(2.178)
Observations	495
R^2	0.577
Adjusted R^2	0.575
Residual Std. Error	7.017 (df = 491)
F Statistic	223.645*** (df = 3; 491)
Note:	*p<0.1; **p<0.05; ***p<0.01

A linear model with just 3 variables can already explain 58% of the variance.

* Note: 7 observations have NA's for *MedHouseValue*, which are filled with the median of known values.

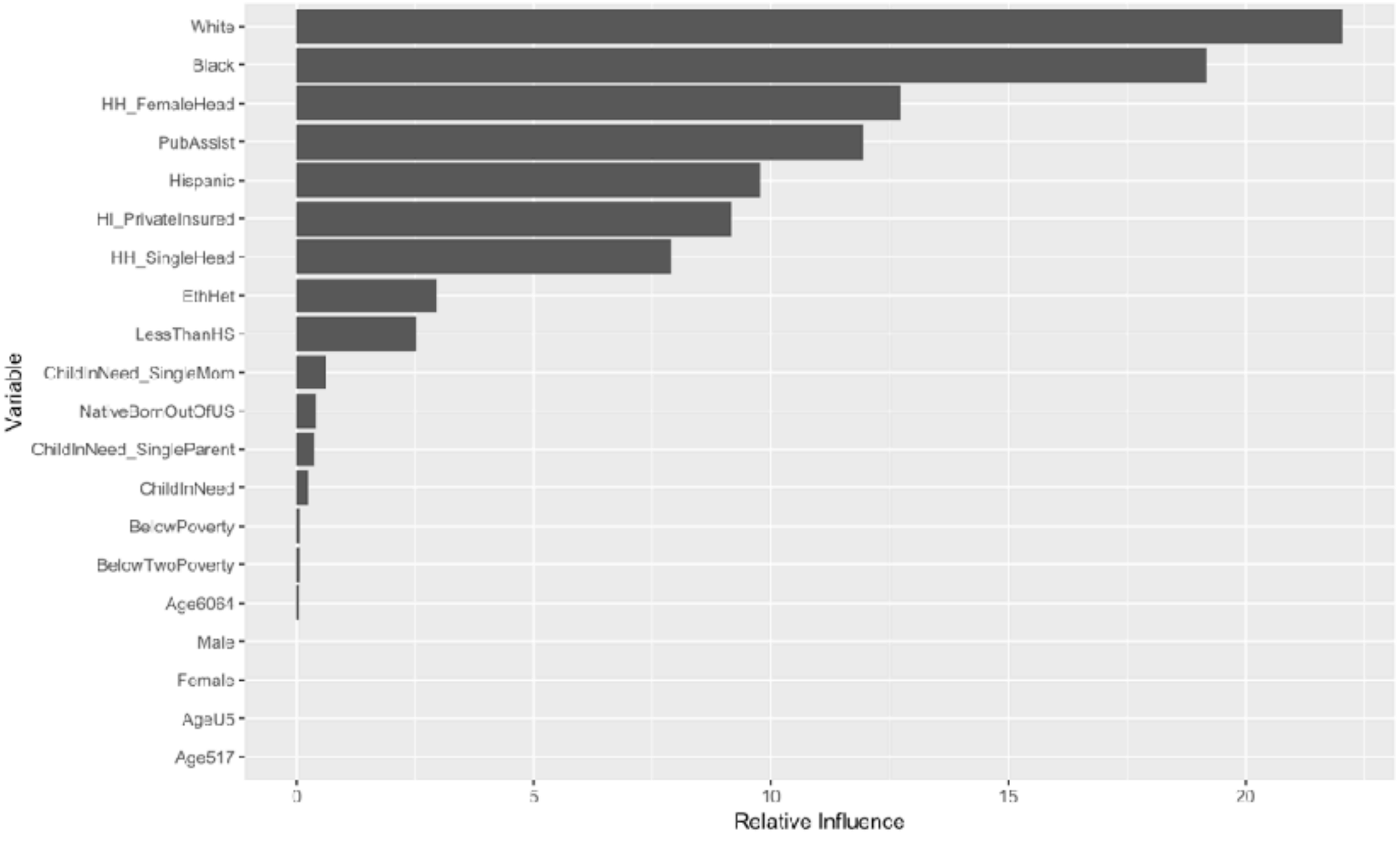
“**Urbanness**” measurements improve the fit.

“Urbanness” measurements improve the fit

	Dependent variable:
	p_call
HI_PrivateInsured	-41.545***
	(2.898)
Black	40.672***
	(4.521)
MedHouseValue	-0.008***
	(0.002)
p.urbanized_pop	0.041***
	(0.009)
Constant	40.064***
	(2.180)
Observations	495
R^2	0.596
Adjusted R^2	0.593
Residual Std. Error	6.869 (df = 490)
F Statistic	180.653*** (df = 4; 490)
Note:	*p<0.1; **p<0.05; ***p<0.01

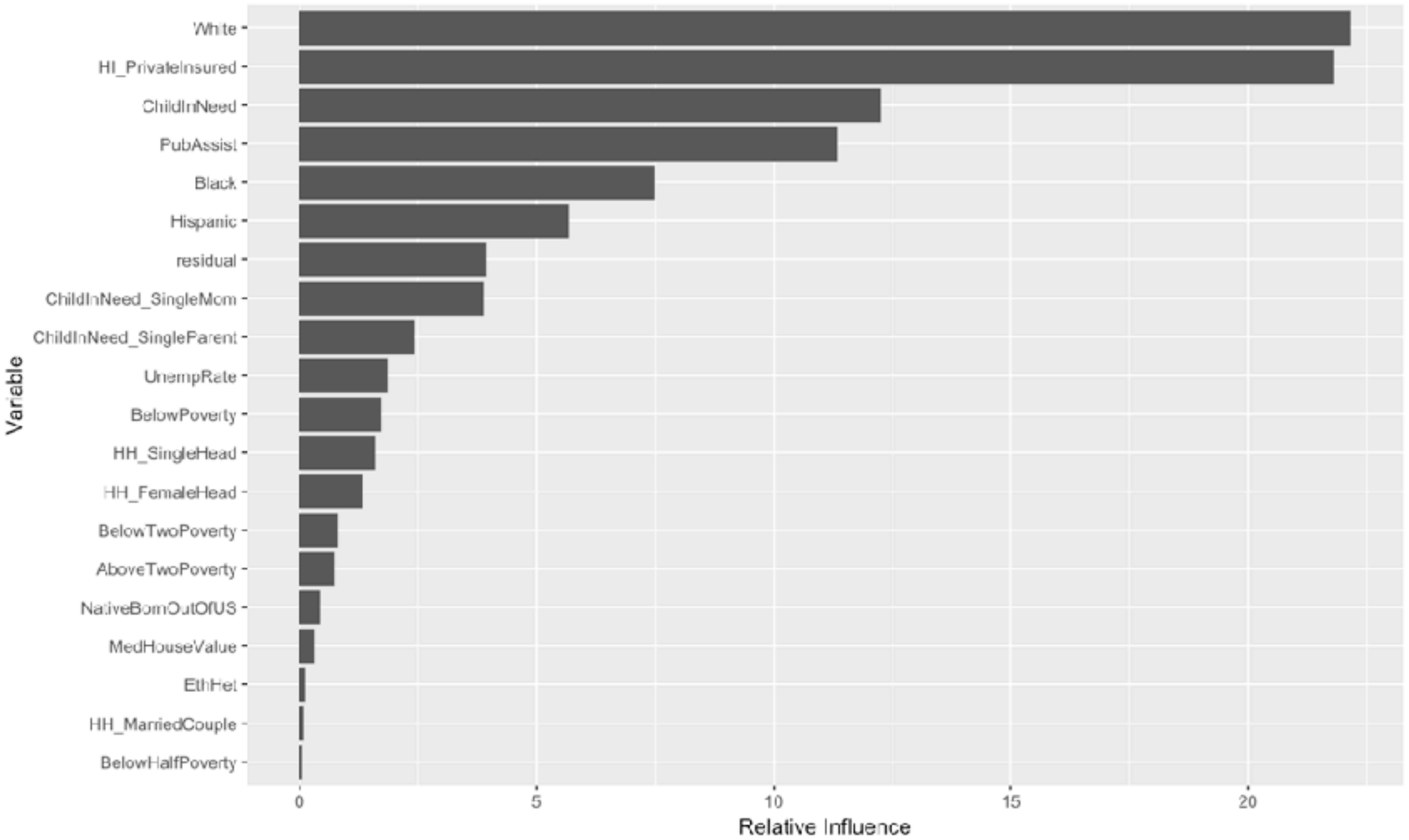
- *p.urbanied_pop* = proportion of population living in an Urban Area or Urban Cluster
- More powerful than simple population density by ZIP code land areas.
- Median housing value may also be an differentiator for rural v.s. urban areas.

Poisson regression tells a different story

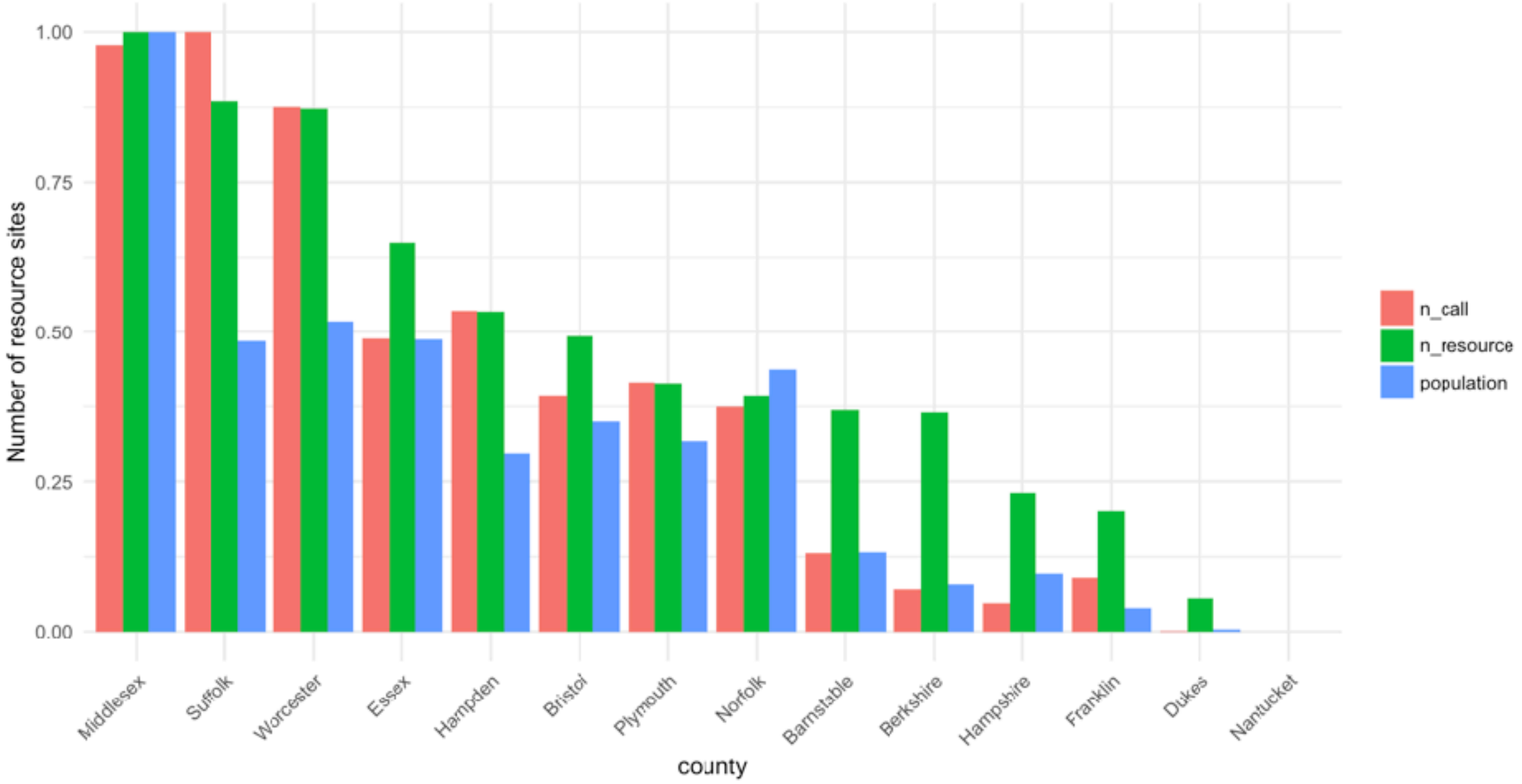


Poisson regression on number of calls with offset by total population

Gradient Boosting assuming Gaussian distribution



Resource Disparity



Resource Disparity

- **Resources have a relatively even distribution** comparing to the other two variables—critical resources must be available to all places, even for remote areas with small population.

- **Resources tend to agglomerate in big cities**, and poor people in big cities are more likely to make 2-1-1 calls–either because there are more resources readily available to them, or because they are more poor.

- Counties surrounding the Suffolk county–Middlesex, Norfolk and Essex, all have a relatively high population and number of resources, but low number of calls.

Outlier Analysis

Neighborhoods with decent amount of inhabitants, but no 211 calls

- **02575** - 2,564 people: Town of *West Tisbury* on *Martha's Vineyard*, where rich people build mansions.
- **01731** - 2,082 people: *Hanscom* Air Force Base
- **01434** - 1,658 people: *Fort Devens*, US Army Reserve Force training area
- **01262** - 1,239 people: *Town of Stockbridge*, located at westernmost of MA

Outlier Analysis

Neighborhoods with unusually high amount of calls

- **01009** - 429 people, 31 calls: Town of Bonsville, part of Palmer, near Springfield, 100% white.
  - Turns out, this is a PO Box! (Removing these ZIP code areas gives an even better model.)
- **01109** - Eastern part of Springfield. 37% hispanics, 14% Purto Ricans, 37% black, median house income 34k.
- **01605** - Worchester, neighborhoods around Green Hill Park.
- **02121** - Roxbury, Boston, predominantly black and hispanic neighborhood, median house income 26k
- **02126** - Mattapan, Boston, predominantly black neighborhood, median house income 46k

Future steps // no additional data needed

- **Temporal Visualization:**
  - Although initial attempts are not enticing…
- **Resource Database:**
  - Run reverse geocoding for more accurate markers
  - Try Ohana Web Search by *Code for America*
  - Or facilitate Mass211 to *fully join* the Open211 initiative?
- **Coverage area analysis:**
  - Use NLP techniques to extract coverage areas from agency description?
- **Caller segments:**
  - While incomplete and (possibly) unreliable, caller profile data may still be useful in some unknown ways.

Future steps // needs additional data and resources

- **Experiments!** Premeditated experiments are the only way to verify the real impact of policies.
  - E.g. Promoting Health by Addressing Basic Needs
- **Track the callers:**
  - 2-1-1 should at least record the number of the callers.
  - ? *Privacy concerns?*
- **Referral interactions:**
  - Is giving more referrals better? Should 2-1-1 maintain a list of "preferred" agencies?