

CRIME AND POLICE SHOOTING ANALYTICS

Justin Ellis
Einstein Postdoctoral Fellow
(aspiring Data Scientist)
JPL/Caltech

SOME BACKGROUND ON ME

- Postdoctoral fellow at JPL/Caltech
- Research focus on Gravitational Wave detection. Mostly work on Bayesian time series analysis multidimensional sampling methods
- Aspiring data scientist who is very interested in machine learning, analytics, visualization etc.
- Hope to learn more through these meet ups!

SOME CAVEATS

SOME CAVEATS

- What this presentation is not:
 - A full detailed analysis of police shooting data that can give probability of being shot given x (x is some feature)
 - an indictment on any group of people based on race, socio-economic status, or ethnicity

SOME CAVEATS

- What this presentation is not:
 - A full detailed analysis of police shooting data that can give probability of being shot given x (x is some feature)
 - an indictment on any group of people based on race, socio-economic status, or ethnicity
- What this talk is:
 - An exploration of open data sets that seeks to look for correlations among police shootings and crime, populations, and poverty
 - An invitation for more ideas and interest and a call for better open data

THE DATA AND TOOLS

- Police Shooting Data:
 - Washington Post police shooting database (2015, 2016)
 - Mapping Police Violence Database (2013, 2014)
- Crime Data:
 - FBI Uniform Crime Reporting Statistics
- Population and Income data
 - US Census County Characteristics Dataset
 - US Census Income Dataset
- Tools: Python Language, Pandas for data frames, GeoPandas, shapely, and pysal for geographic data, Seaborn and Matplotlib for visualizations, Scikit-Learn for ML, nltk for text analysis

BACKGROUND

- Obviously, police brutality and police killings are a very hot topic recently, would like to try to gain some insights from data (preferably open data)

NBER WORKING PAPER SERIES

AN EMPIRICAL ANALYSIS OF RACIAL DIFFERENCES IN POLICE USE OF FORCE

Roland G. Fryer, Jr

Working Paper 22399

<http://www.nber.org/papers/w22399>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

ABSTRACT

This paper explores racial differences in police use of force. On non-lethal uses of force, blacks and Hispanics are more than fifty percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force – officer-involved shootings – we find no racial differences in either the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of which have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

Uses individualized data (mostly from Houston) and logistic regression methods.

RESEARCH ARTICLE

A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014

Cody T. Ross*

Department of Anthropology, University of California, Davis, Davis, California, United States of America

Abstract

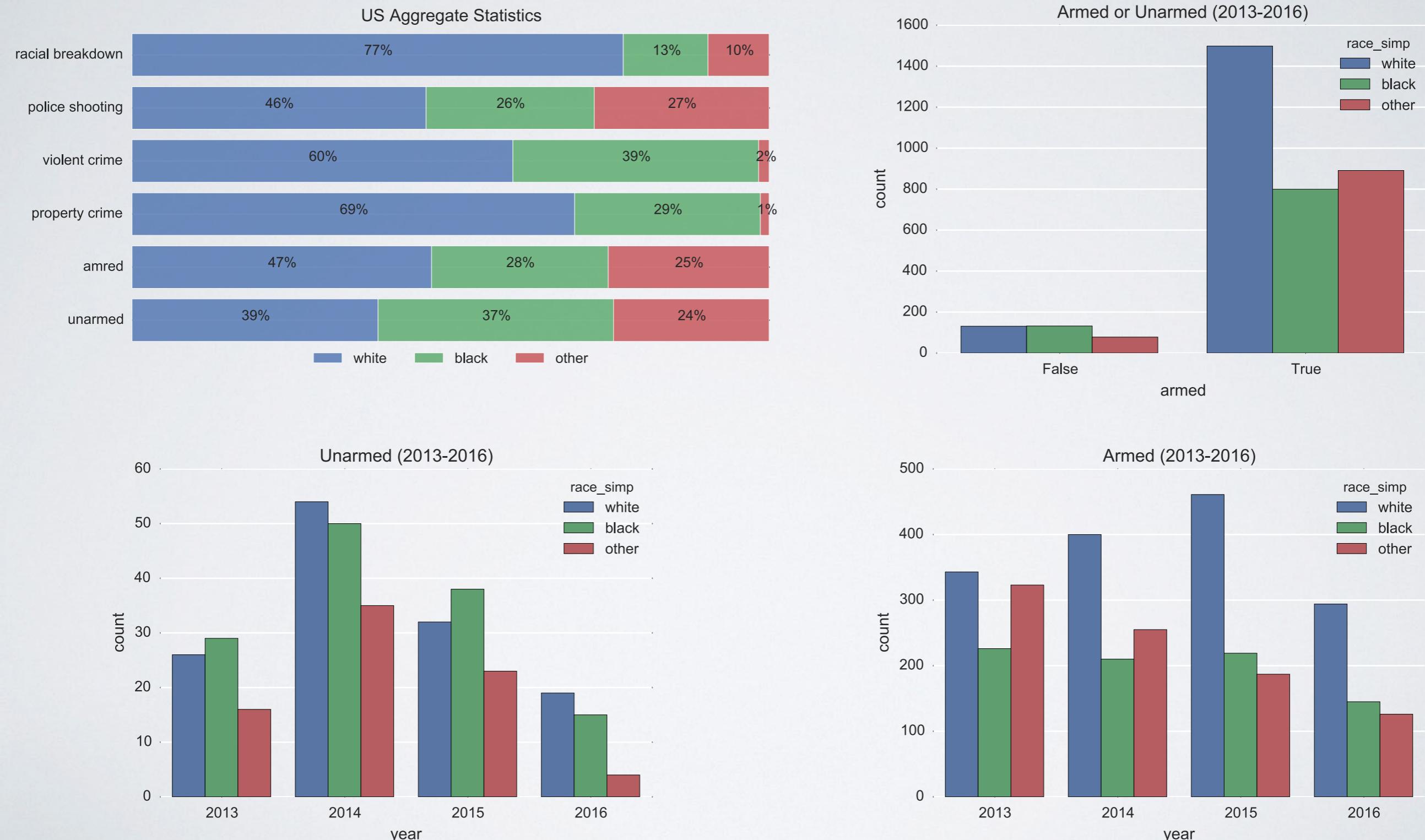
A geographically-resolved, multi-level Bayesian model is used to analyze the data presented in the U.S. Police-Shooting Database (USPSD) in order to investigate the extent of racial bias in the shooting of American civilians by police officers in recent years. In contrast to previous work that relied on the FBI's Supplemental Homicide Reports that were constructed from self-reported cases of police-involved homicide, this data set is less likely to be biased by police reporting practices. County-specific relative risk outcomes of being shot by police are estimated as a function of the interaction of: 1) whether suspects/civilians were armed or unarmed, and 2) the race/ethnicity of the suspects/civilians. The results provide evidence of a significant bias in the killing of unarmed black Americans relative to unarmed white Americans, in that the probability of being {black, unarmed, and shot by police} is about 3.49 times the probability of being {white, unarmed, and shot by police} on average. Furthermore, the results of multi-level modeling show that there exists significant heterogeneity across counties in the extent of racial bias in police shootings, with some counties showing relative risk ratios of 20 to 1 or more. Finally, analysis of police shooting data as a function of county-level predictors suggests that racial bias in police shootings is most likely to emerge in police departments in larger metropolitan counties with low median incomes and a sizable portion of black residents, especially when there is high financial inequality in that county. There is no relationship between county-level racial bias in police shootings and crime rates (even race-specific crime rates), meaning that the racial bias observed in police shootings in this data set is not explainable as a response to local-level crime rates.

Uses county level data and Bayesian statistical modeling

A FEW TERMS

- Violent crime includes Murder, Rape, Robbery, and Aggravated assault
- Property crime includes Burglary, larceny-theft, motor-vehicle theft, and arson
- I mostly categorize race in to White, Black and Other because the UCR statistics do not do a very good job tracking those of hispanic origin and other groups are largely negligible in the shooting and crime categories

NATIONAL OVERVIEW



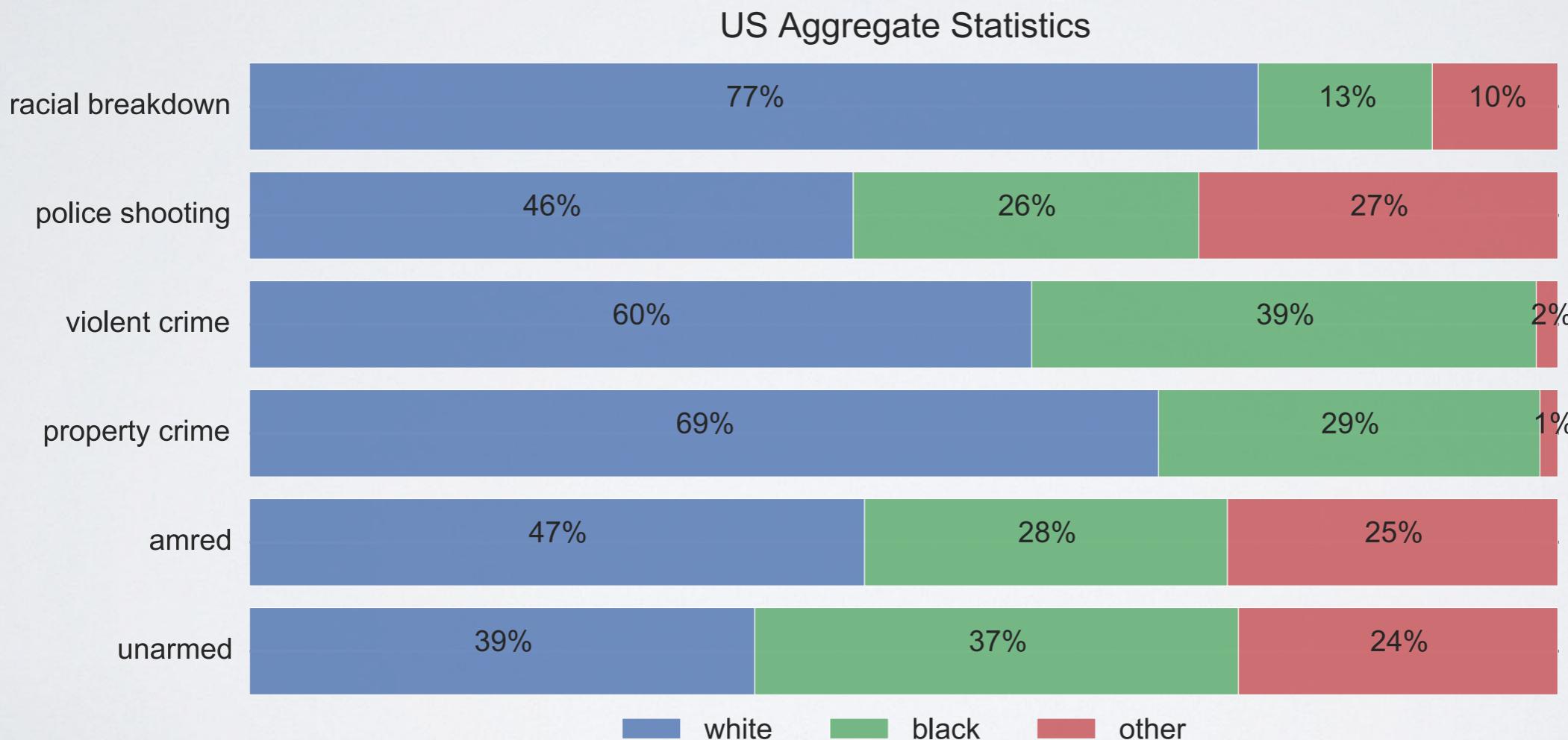
FRAMING THE QUESTIONS

FRAMING THE QUESTIONS

- What population should we compare police shootings to?

FRAMING THE QUESTIONS

- What population should we compare police shootings to?



FRAMING THE QUESTIONS

- What population should we compare police shootings to?

FRAMING THE QUESTIONS

- What population should we compare police shootings to?
- What is the best resolution level, national, state, county, intra-county?

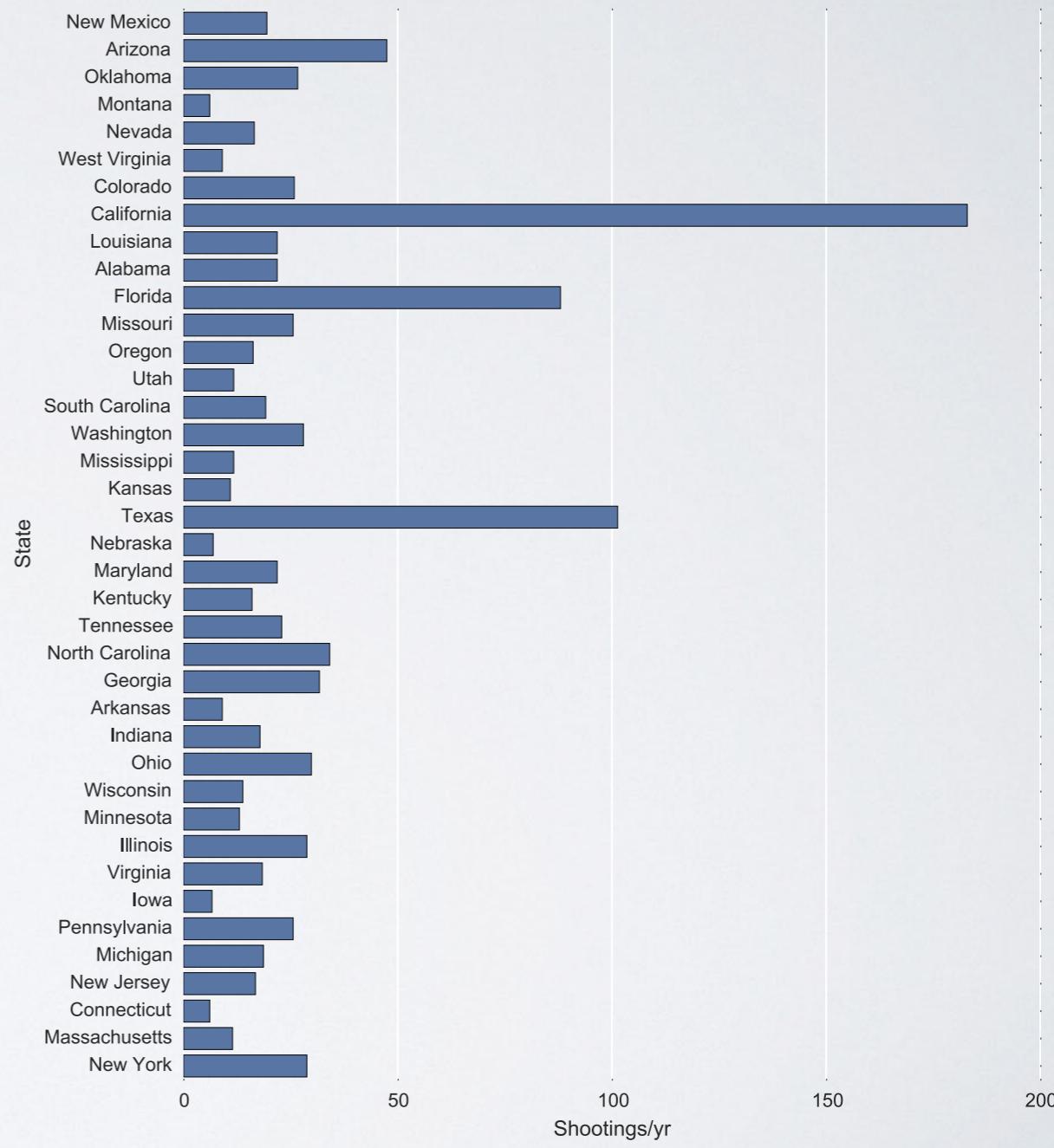
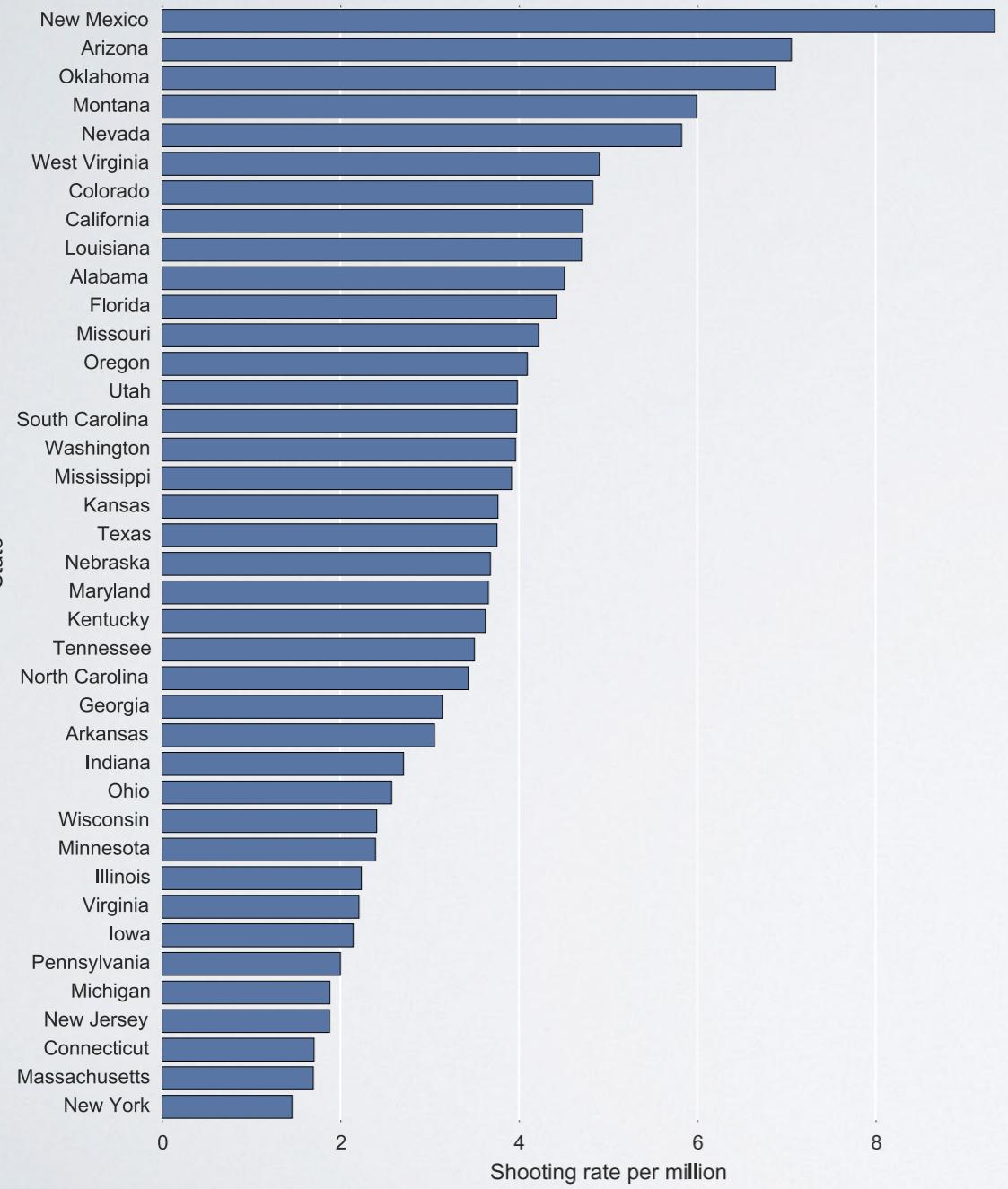
FRAMING THE QUESTIONS

- What population should we compare police shootings to?
- What is the best resolution level, national, state, county, intra-county?
- How do we identify outlier police departments or counties?

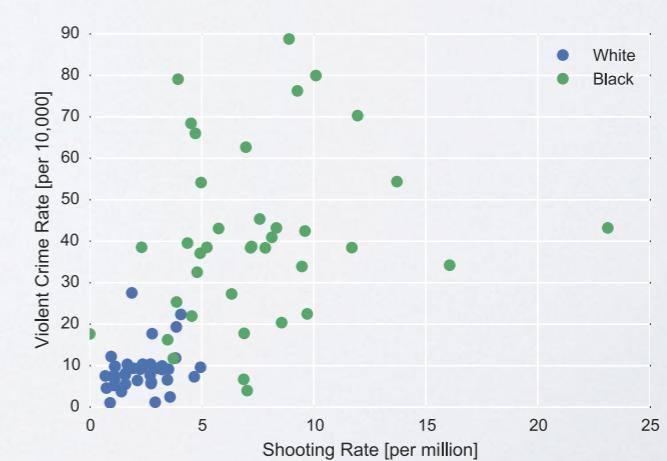
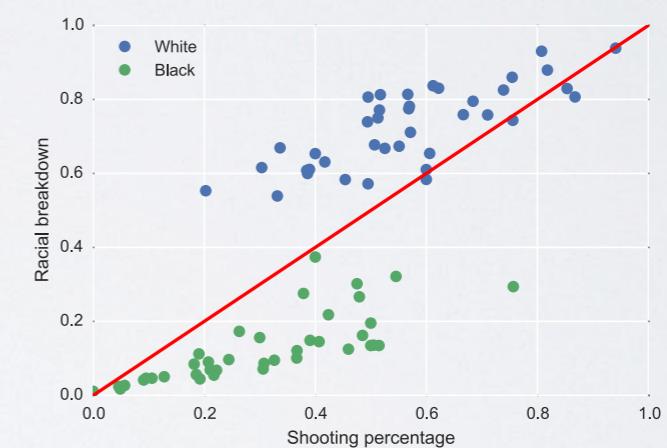
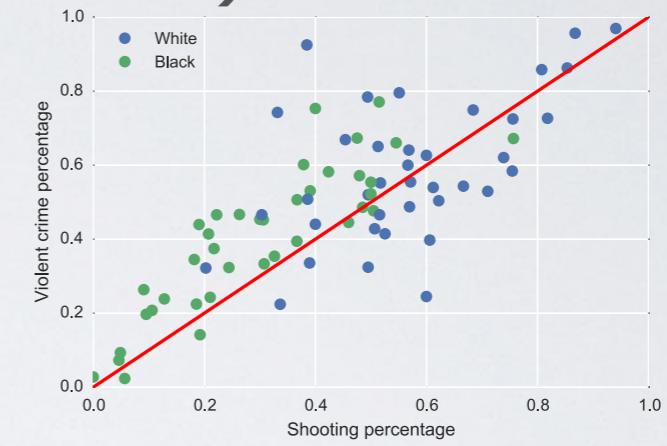
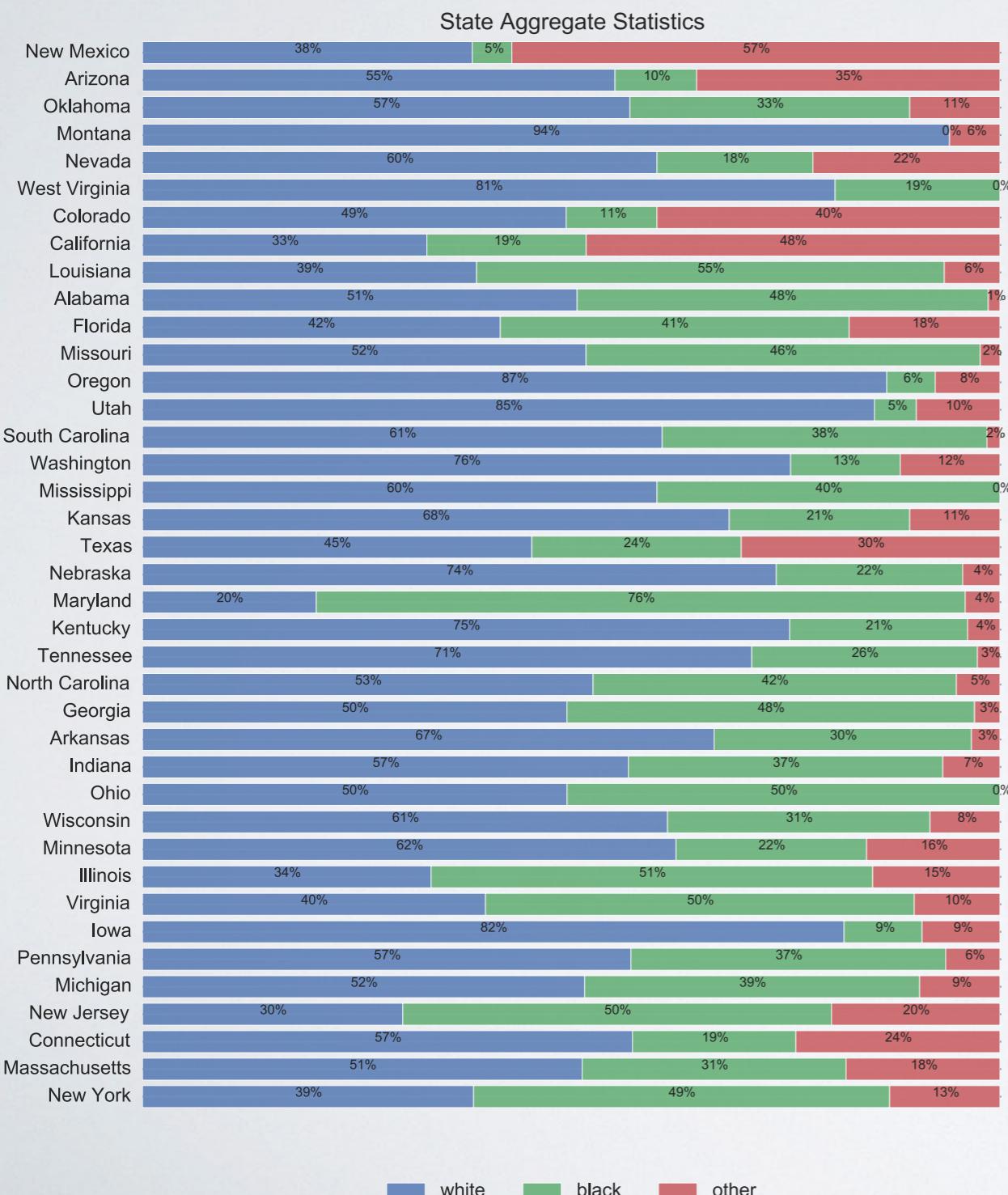
FRAMING THE QUESTIONS

- What population should we compare police shootings to?
- What is the best resolution level, national, state, county, intra-county?
- How do we identify outlier police departments or counties?
- What data do we really need to answer these questions?

STATE LEVEL ANALYSIS

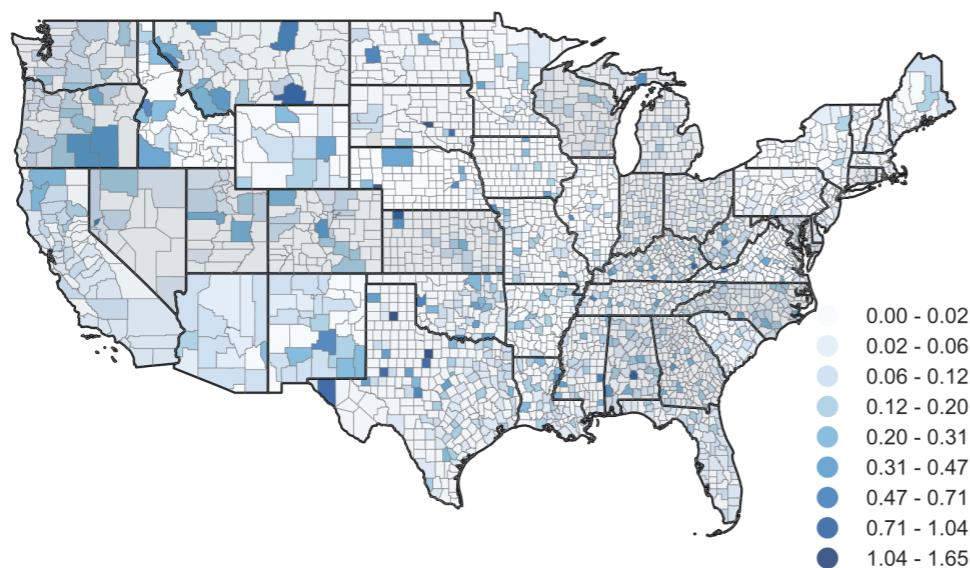


STATE LEVEL ANALYSIS (FRACTIONS)

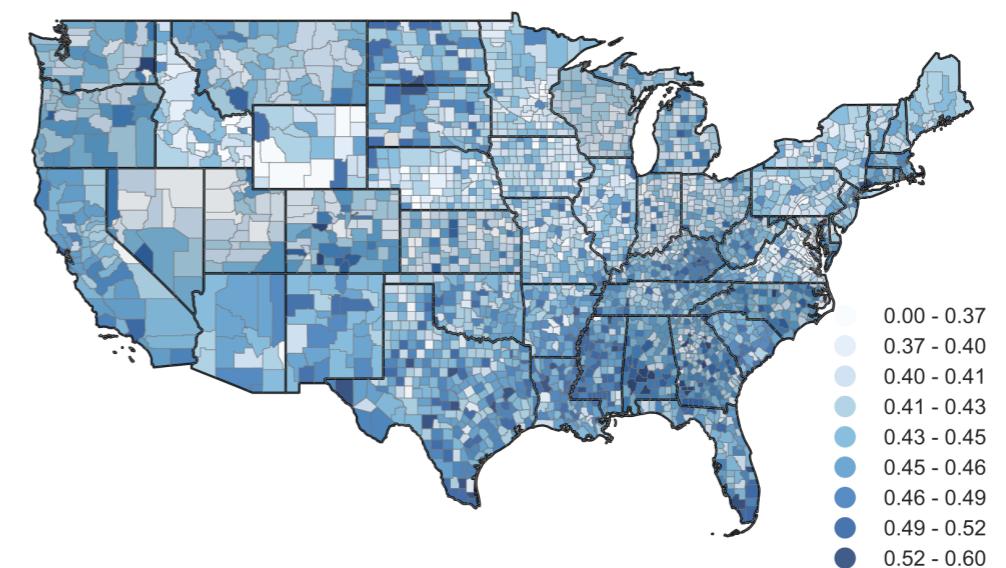


COUNTY LEVEL ANALYSIS

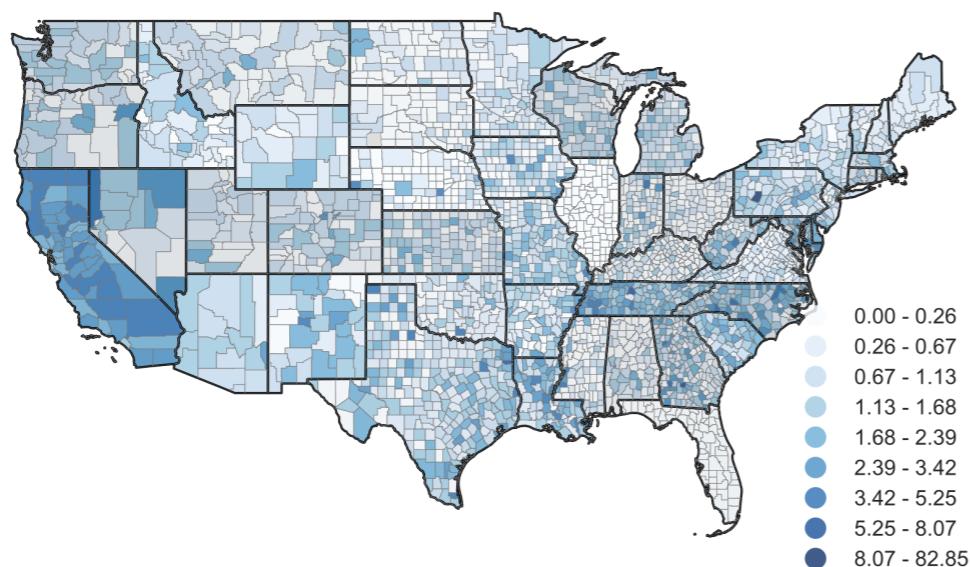
Shooting Rate



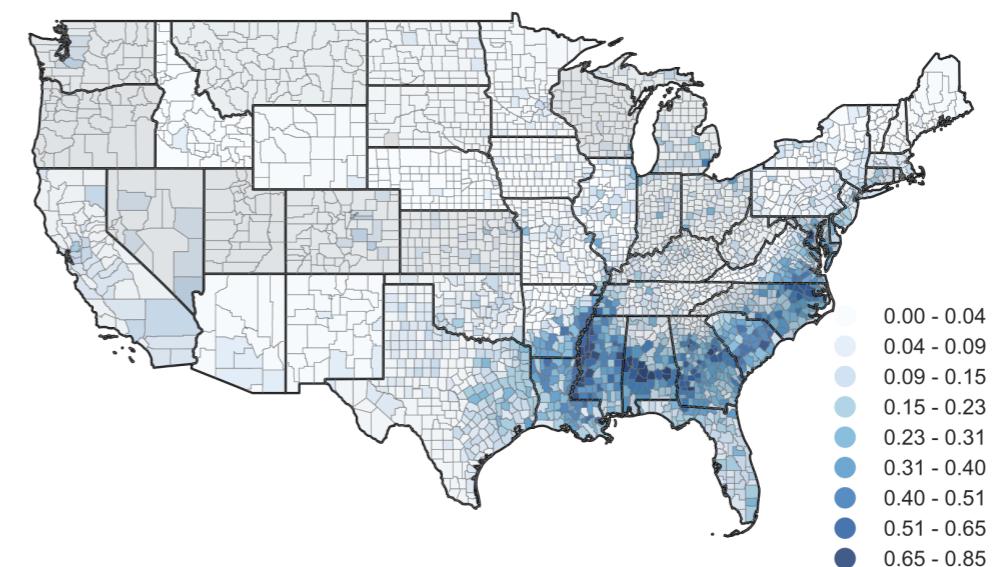
GINI Index



Violent Crime Rate

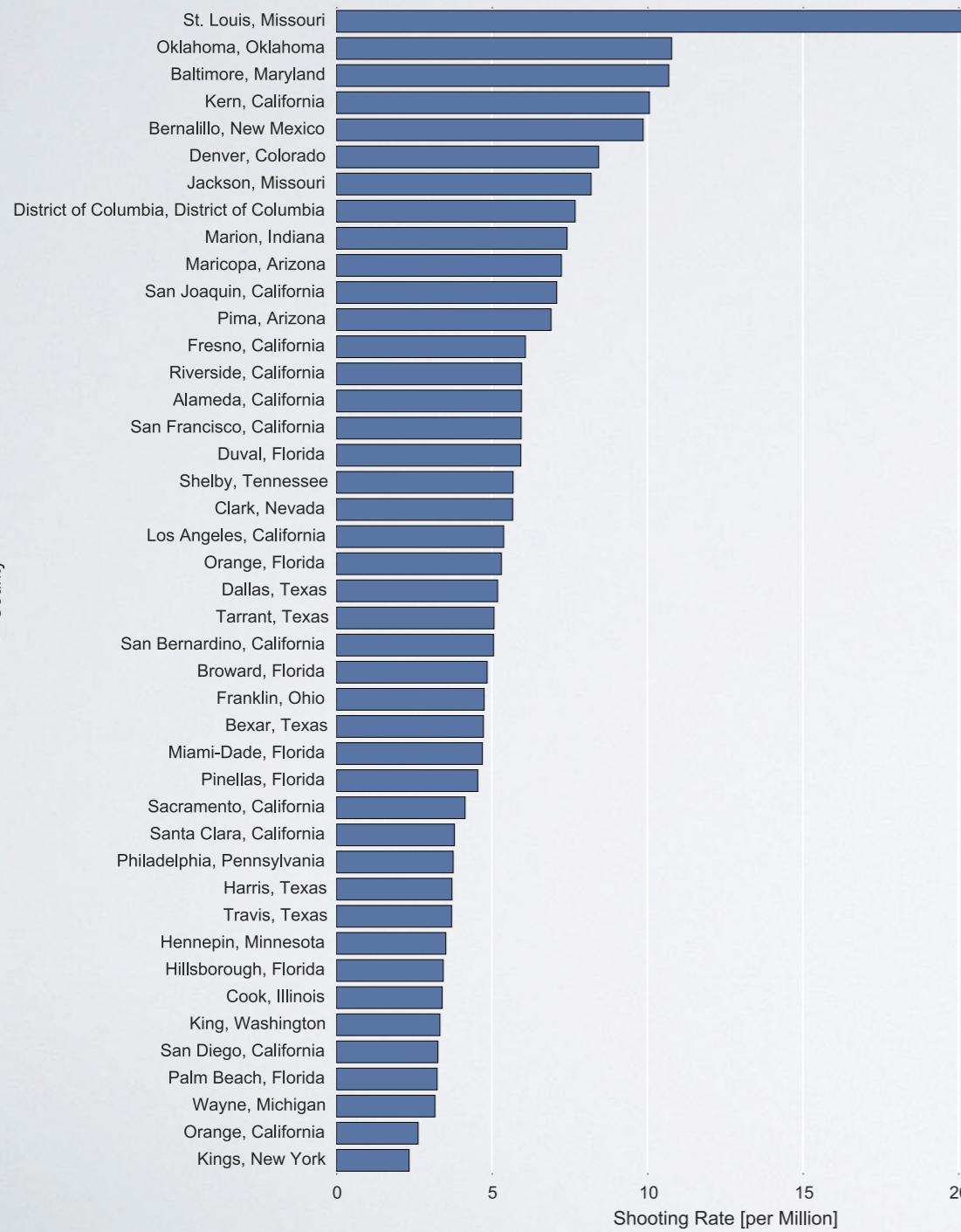


Percent Black

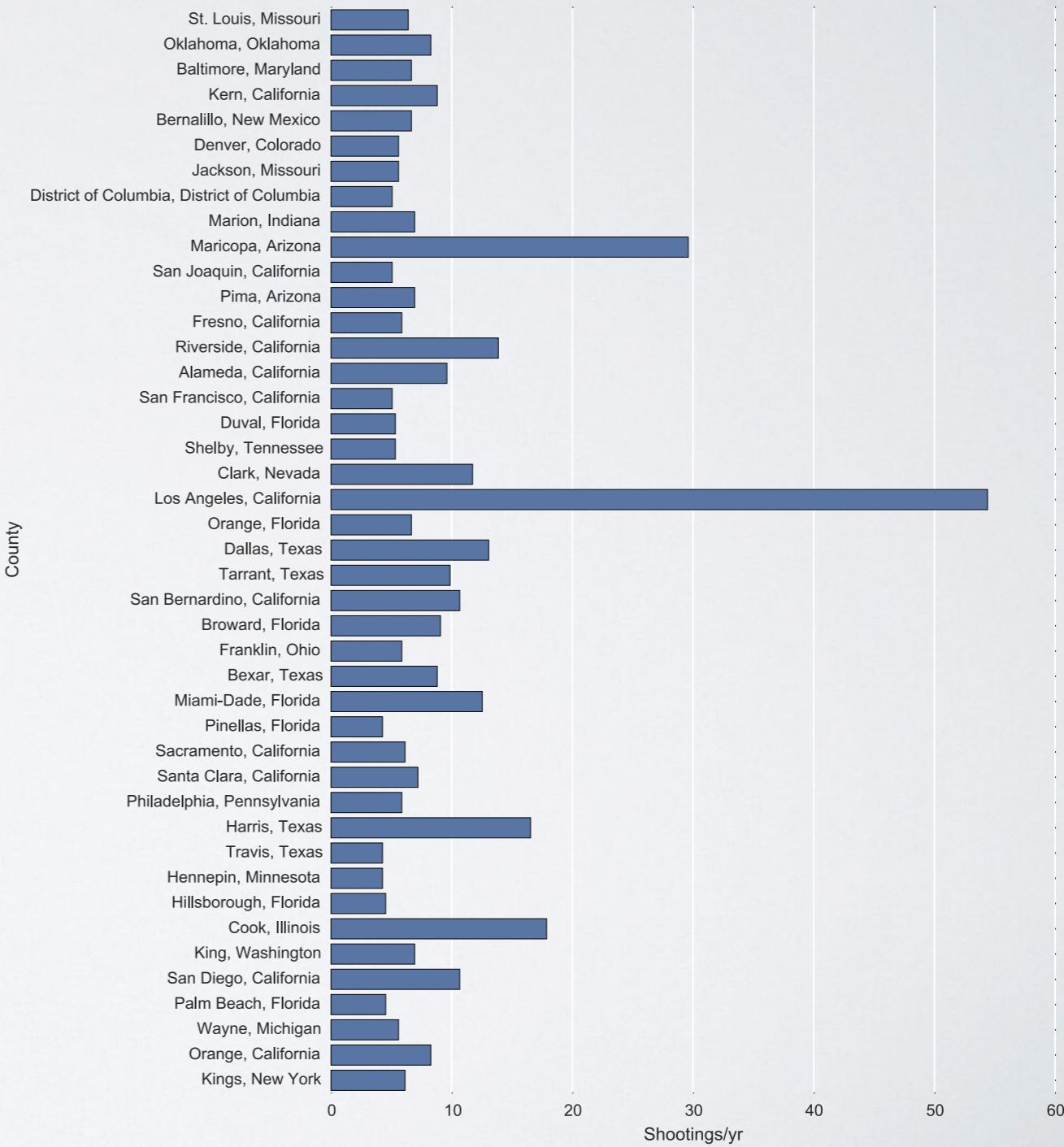


COUNTY LEVEL ANALYSIS

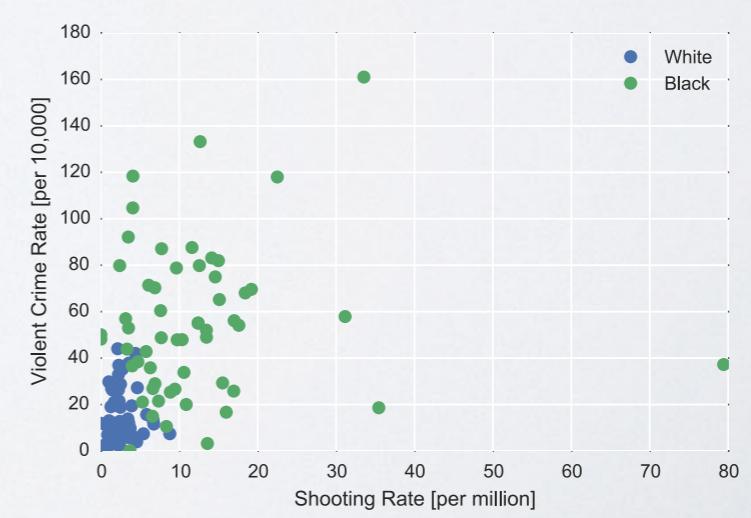
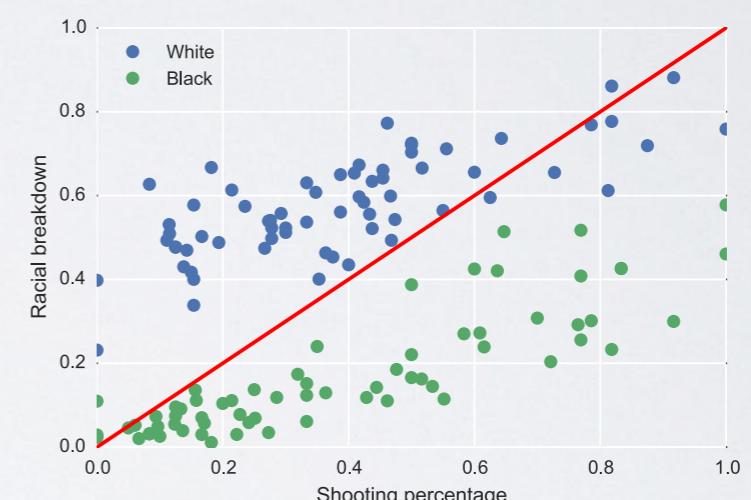
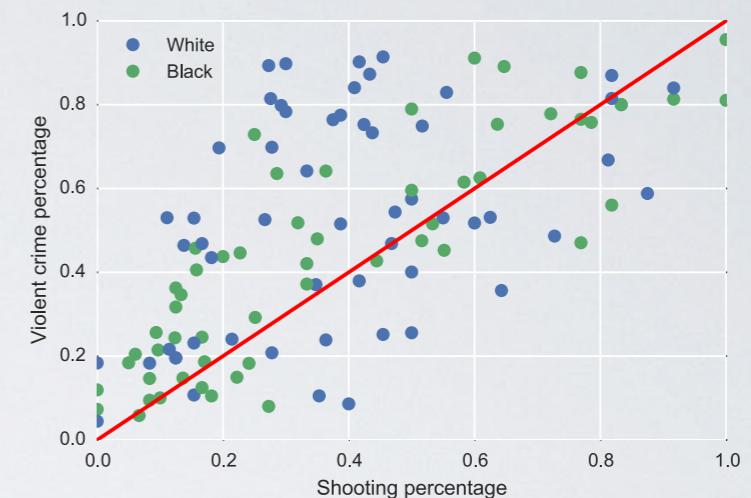
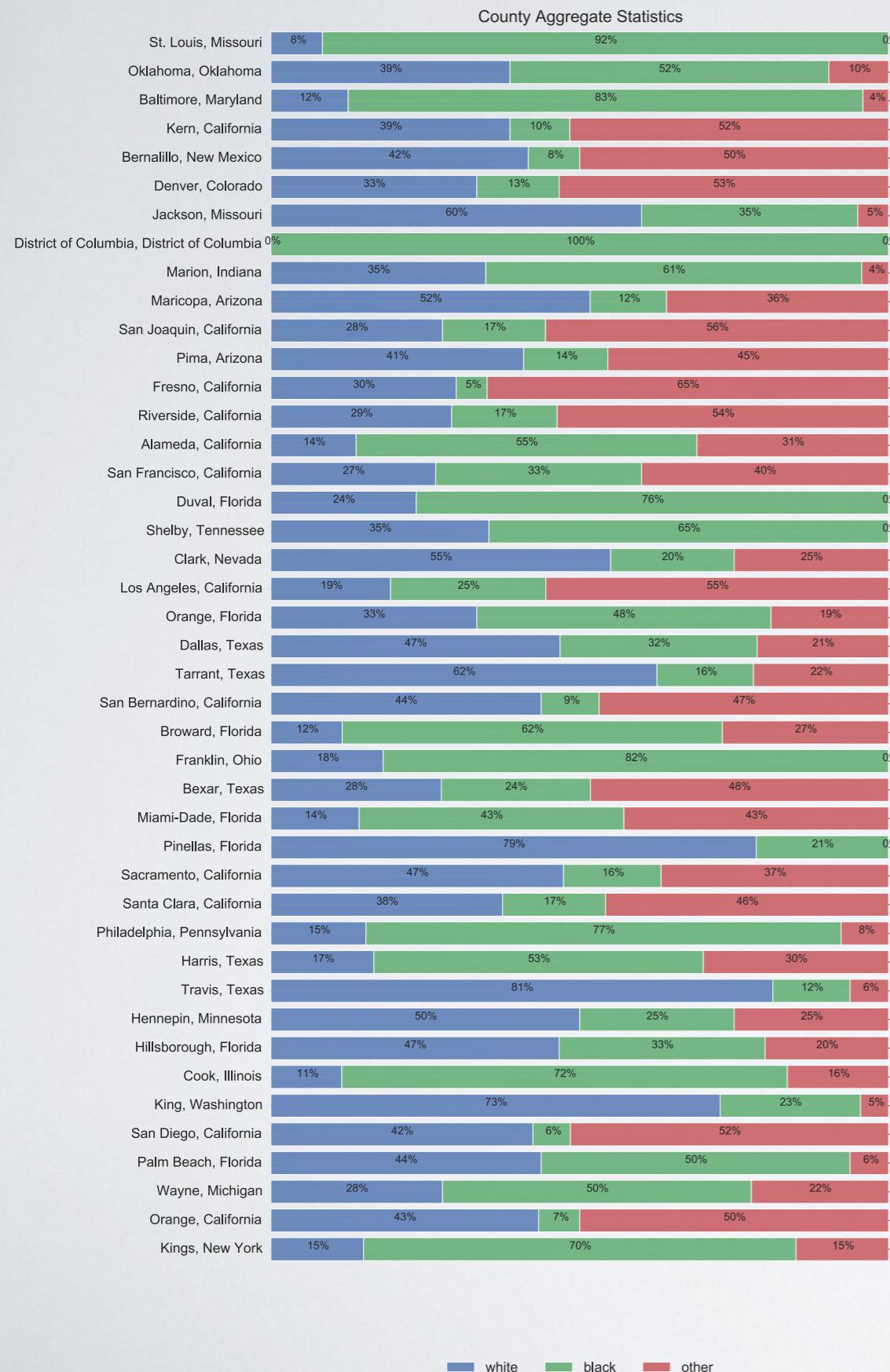
* Counties with more than 15 shootings



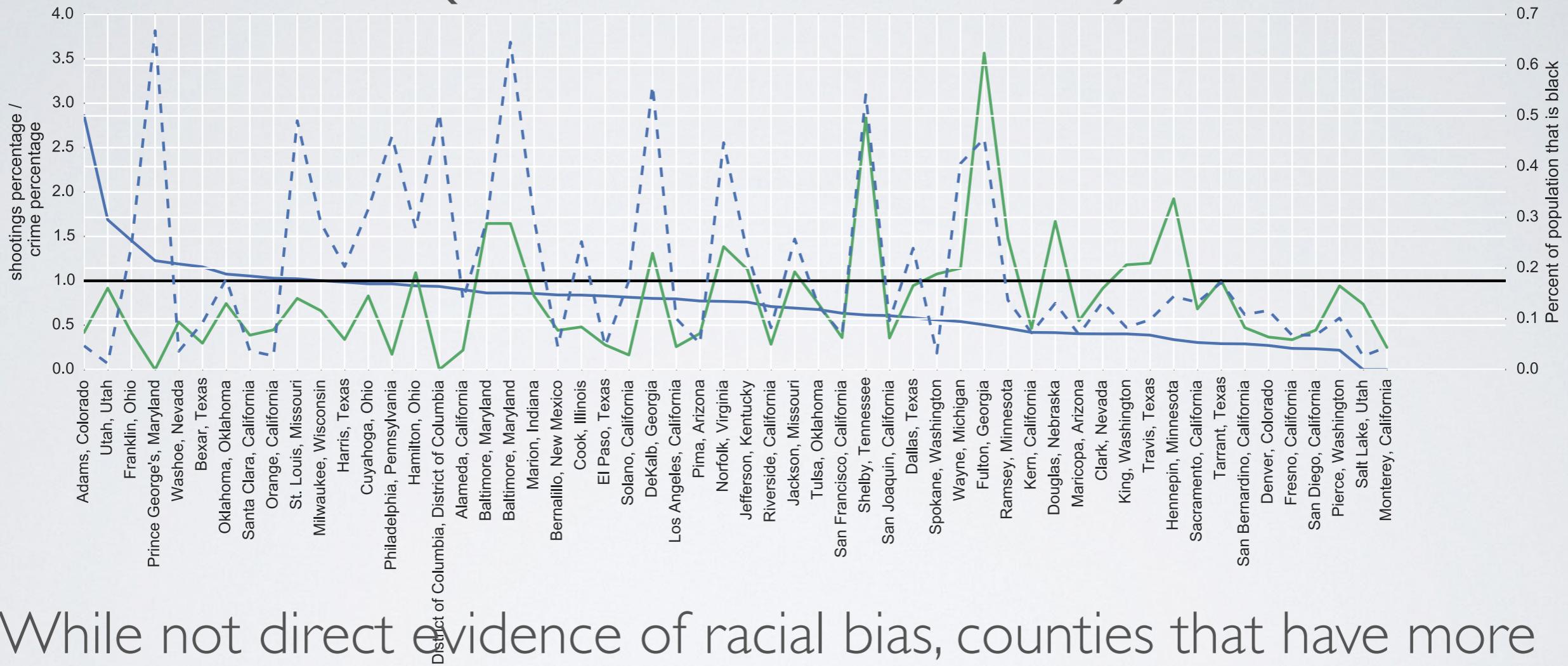
* Counties with more than 15 shootings



COUNTY LEVEL ANALYSIS



COUNTY LEVEL ANALYSIS (RACIAL BIAS)



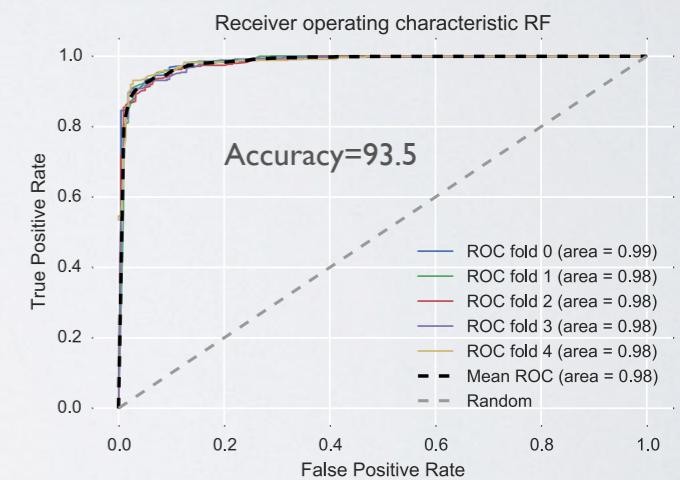
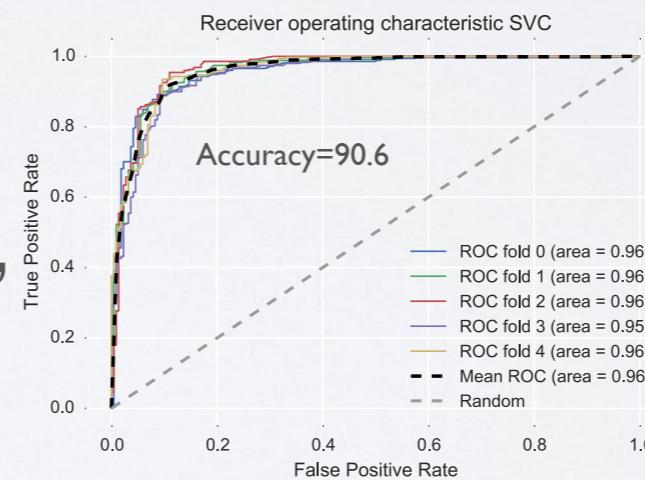
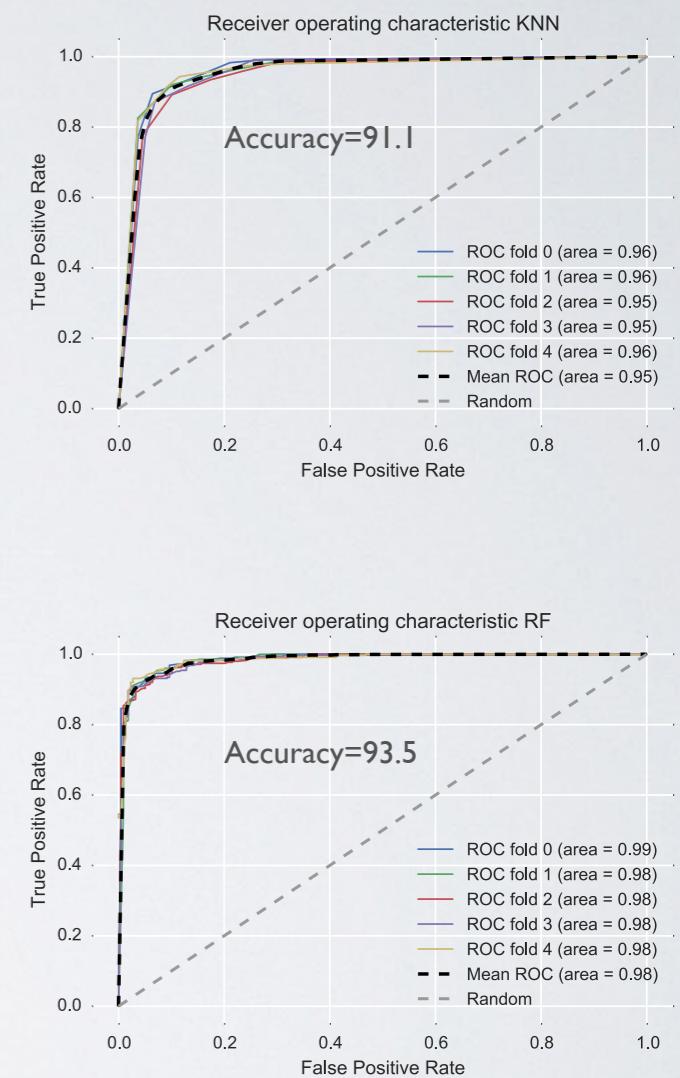
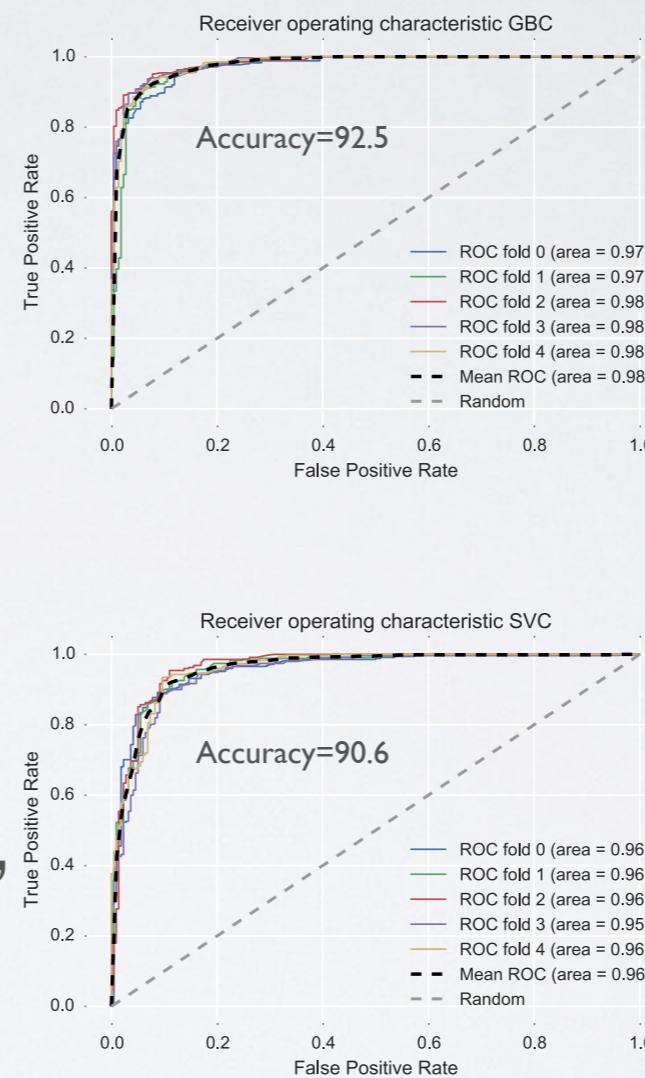
- While not direct evidence of racial bias, counties that have more shootings than one would expect based on crime rates are prime suspects.
- It also appears that these seem to happen when the racial makeup is highly skewed in one direction or the other

A BIT OF MACHINE LEARNING

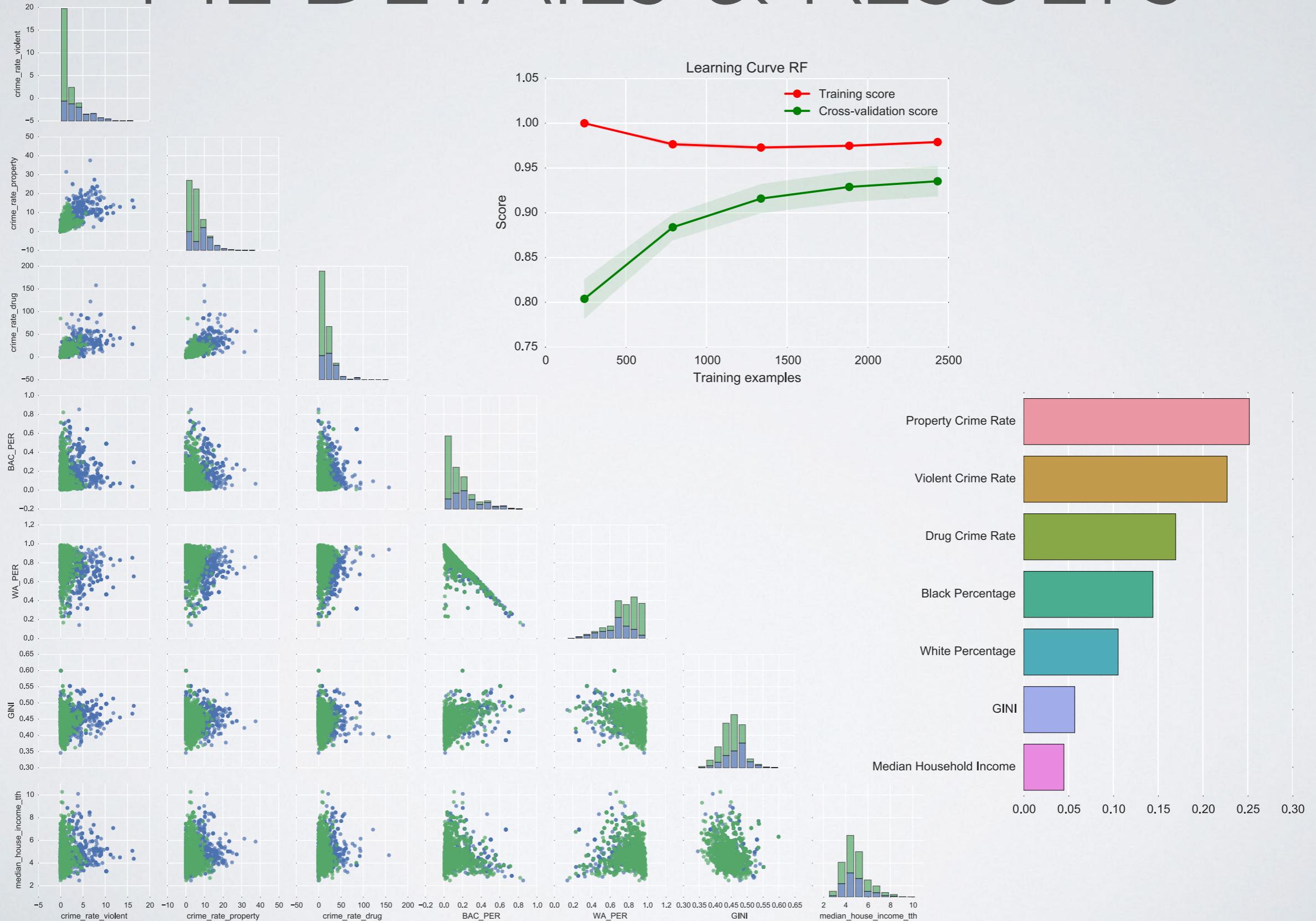
- In a perfect world we would have data on all police encounters so we would be able to do a classification problem with boolean labels shot vs not shot
- We only have fairly reliable county data on shootings but we have seen that there are various indicators that go along with police shootings
- To test this we use various features and try to predict the race of the victim.

ML DETAILS & RESULTS

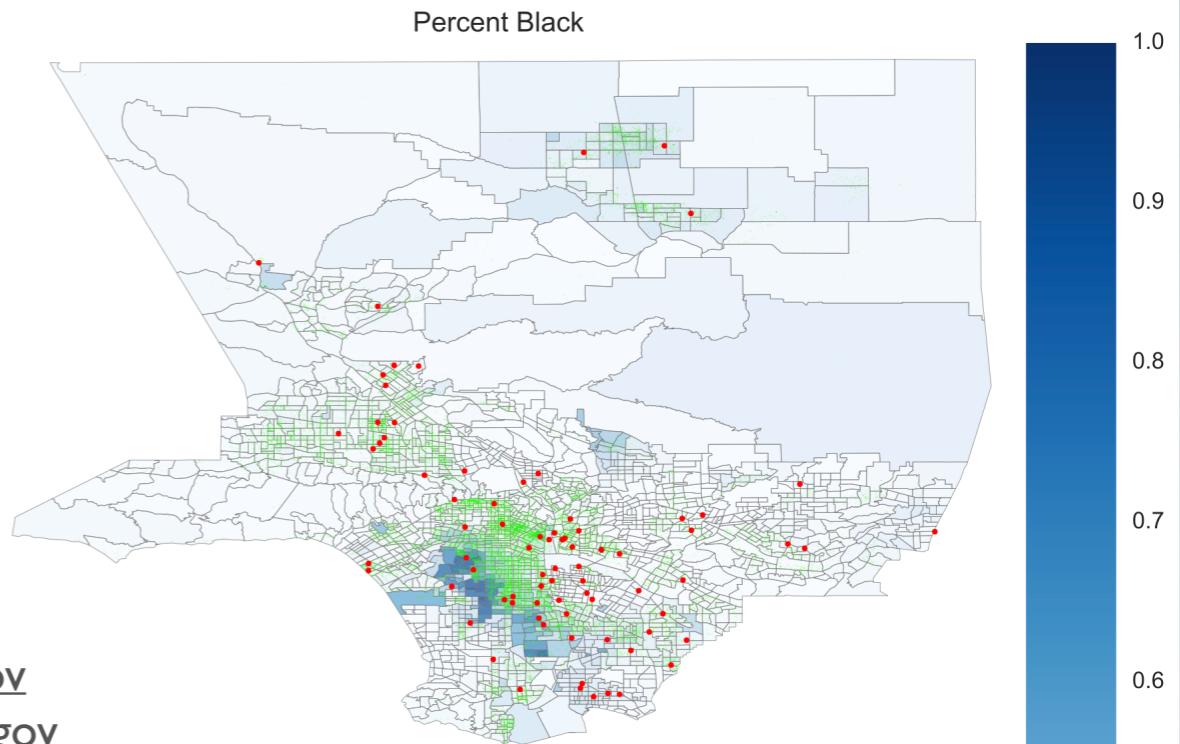
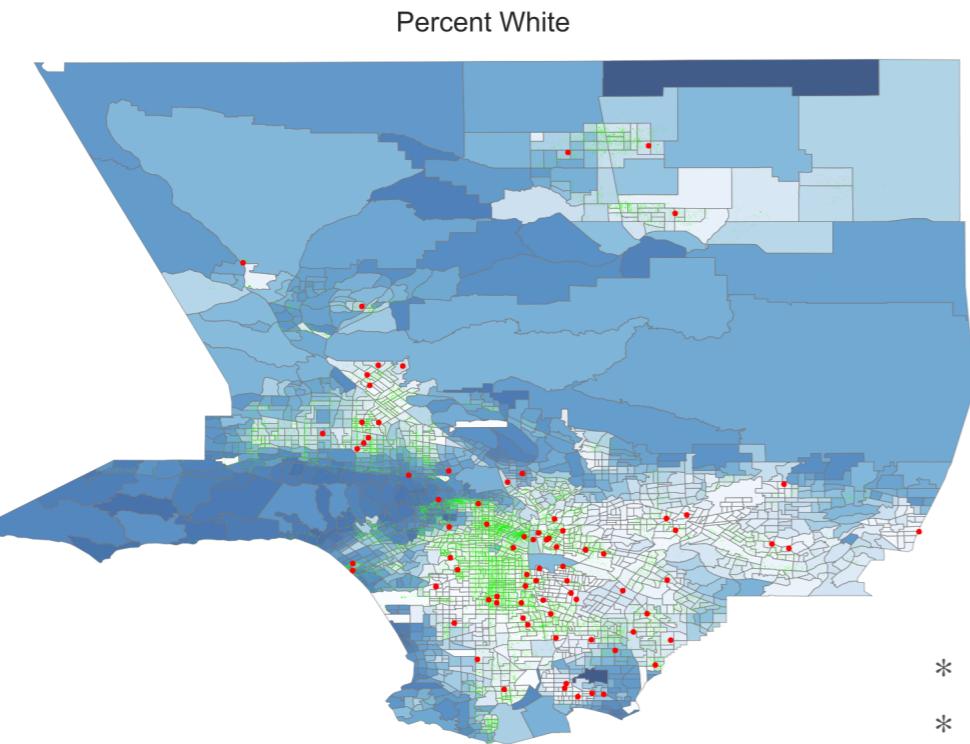
- Features: Violent crime rate, property crime rate, drug crime rate, black percentage, white percentage, GINI, median household income
- ML classifiers: Logistic Regression, Gradient Boosting, Random Forest, SVC, KNN
- Cross validation with Stratified k-folds (5 folds)



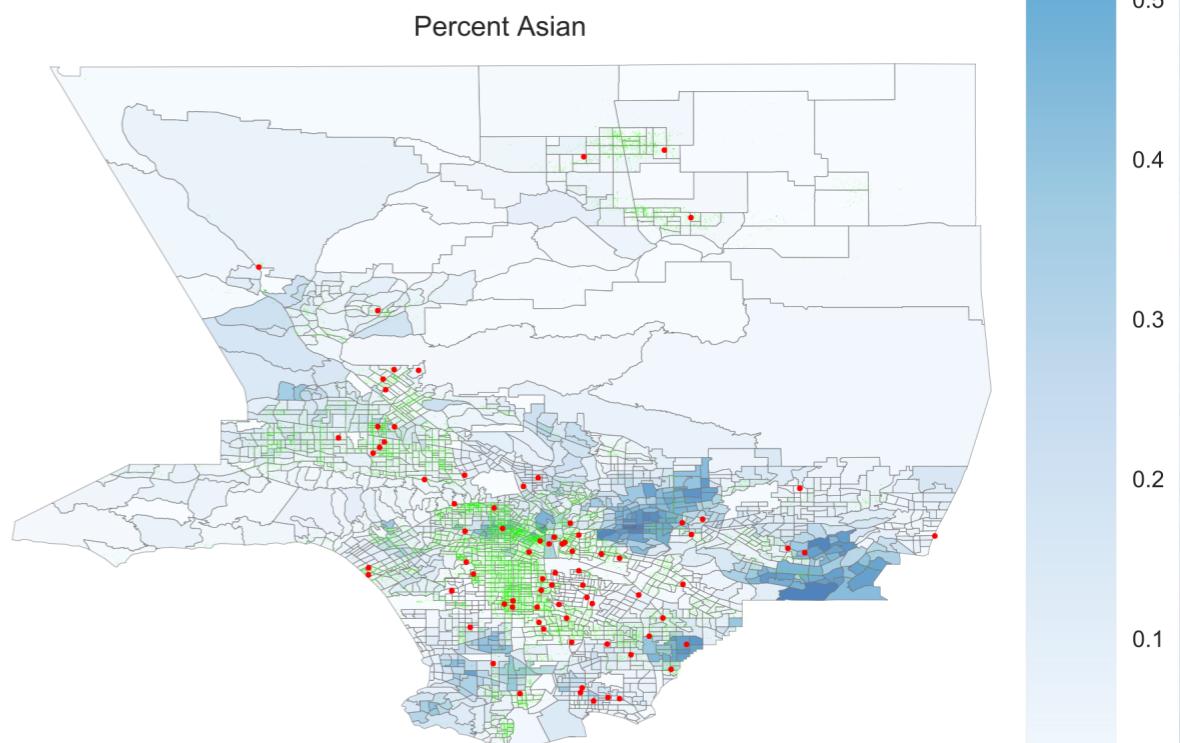
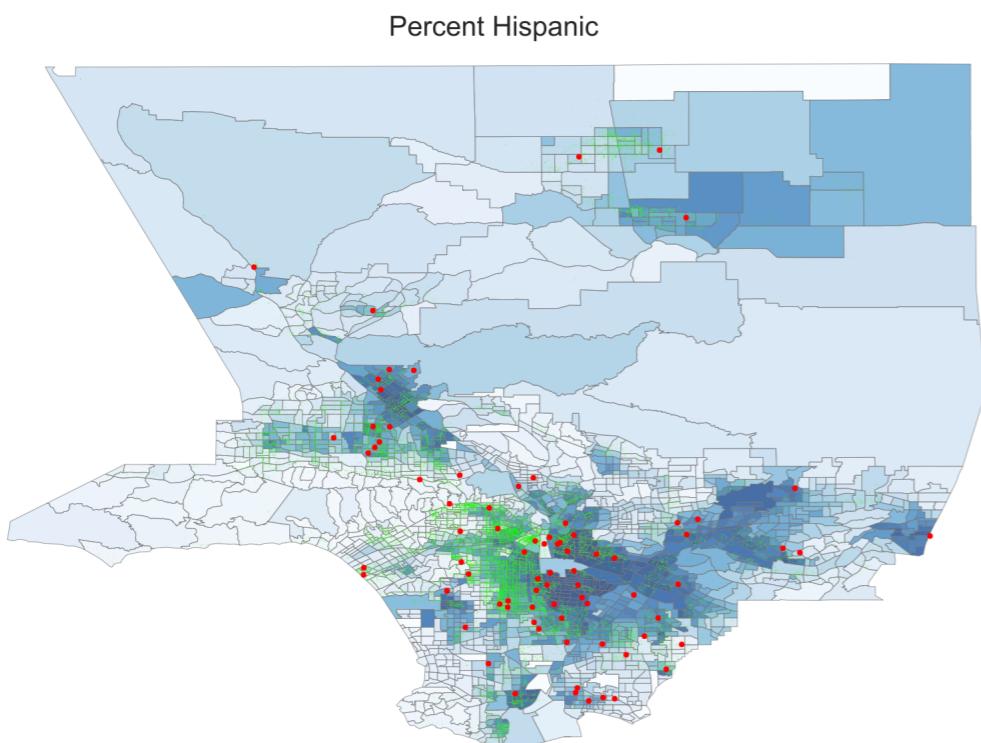
ML DETAILS & RESULTS



LOCAL LA DATA



* data.lacity.gov
* data.lacounty.gov



WHERE TO GO FROM HERE

- Use local data (only big cities have good open data) to construct crime rates and population numbers.
- Use Watson to “read” reports for all shooting incidents to gain extra features. This could be done with more personalized data like Fryer study.
- Place more emphasis on finding outliers or bad police departments
- Other thoughts?