Understanding Racial Disparity In The Virginia Court System

Gaurav Anand, Shannon Paylor, Amanda West

Introduction

- Collaborated with Code for Charlottesville project with Legal Aid Justice Center (LAJC) advocating for criminal record expungement
- Focused on identifying any racial disparities in Virginia court outcomes
- Particularly interested in marijuana charges with recent legalization

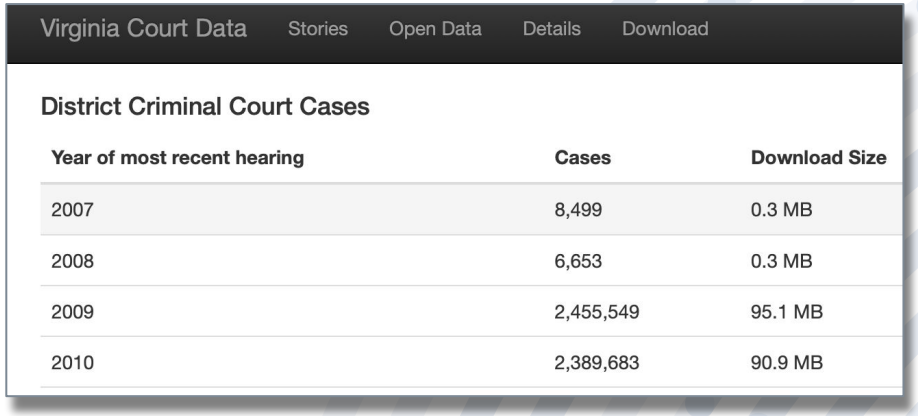
Data Source

- All Virginia court data is in the public domain
  - Ben at https://virginiacourtdata.org/ has web-scraped all cases from official sites and anonymized the data for all of 2000-2020

- Broken up by district criminal, district civil, circuit criminal, and circuit civil court cases

- Given personal address, race, gender, charge, final verdict, among other variables (names are concealed)

Data Source



Address	Gender	Race	Charge	CodeSection	CaseType
OXON HILL MD 20745	Male	Black(Non-Hispanic)	FAIL TO APPEAR,FELONY OFFENSE	19.2-128	Capias
OXON HILL MD 20745	Male	Black(Non-Hispanic)	IDENTITY FRAUD	18.2-186.3	Felony
OXON HILL MD 20745	Male	Black(Non-Hispanic)	FORGING COIN AND BANK NOTES	18.2-170	Felony
OXON HILL MD 20745	Male	Black(Non-Hispanic)	ATTEMPT - OBTAIN MONEY FALSE PRETENSES	18.2-178	Felony
TAUNTON, MA 02780	Male	White Caucasian(Non-Hispanic)	CAPIAS FAILED TO APPEAR	19.2-128	Capias
TAUNTON, MA 02780	Male	White Caucasian(Non-Hispanic)	FAIL TO STOP/ ACCIDENT - MISD	B.46.2-894	Misdemeanor

1. Exploratory Data Analysis

Exploratory Data Analysis

Demographic Information

Sex	count	Proportion
Female	21837	0.279531
Male	56283	0.720469

Gender*

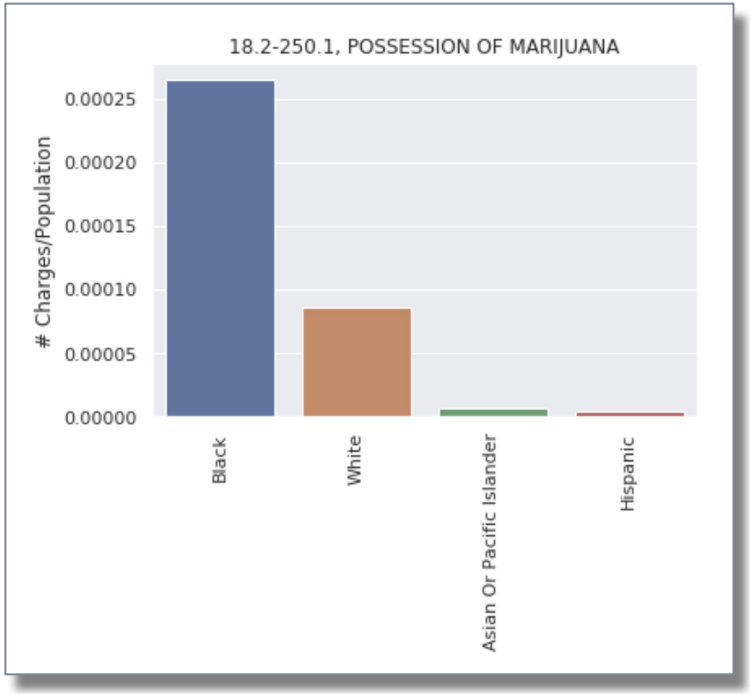
Race	count	Proportion
American Indian Or Alaskan Native	38	0.000494
White	43890	0.570318
Asian Or Pacific Islander	260	0.003379
Black	32401	0.421027
Hispanic	368	0.004782

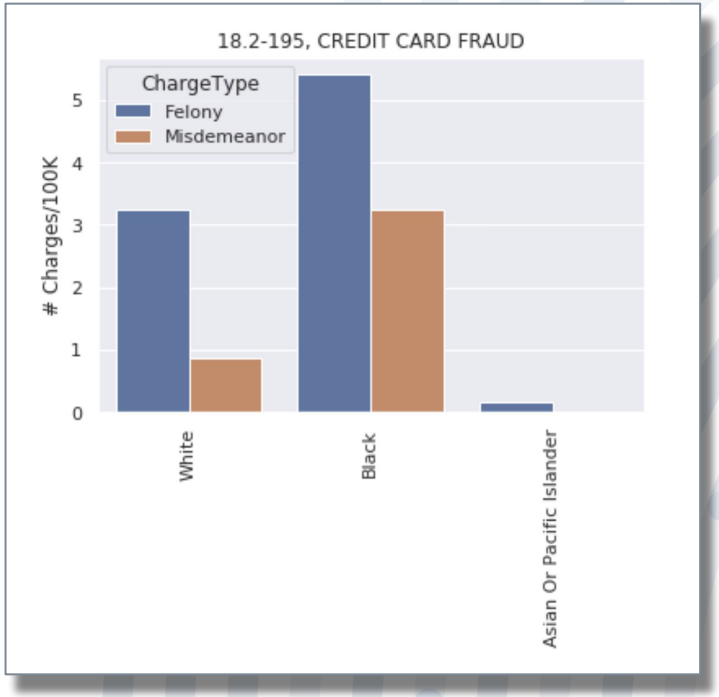
Race*

*2019 Circuit Criminal Court Data

Exploratory Data Analysis

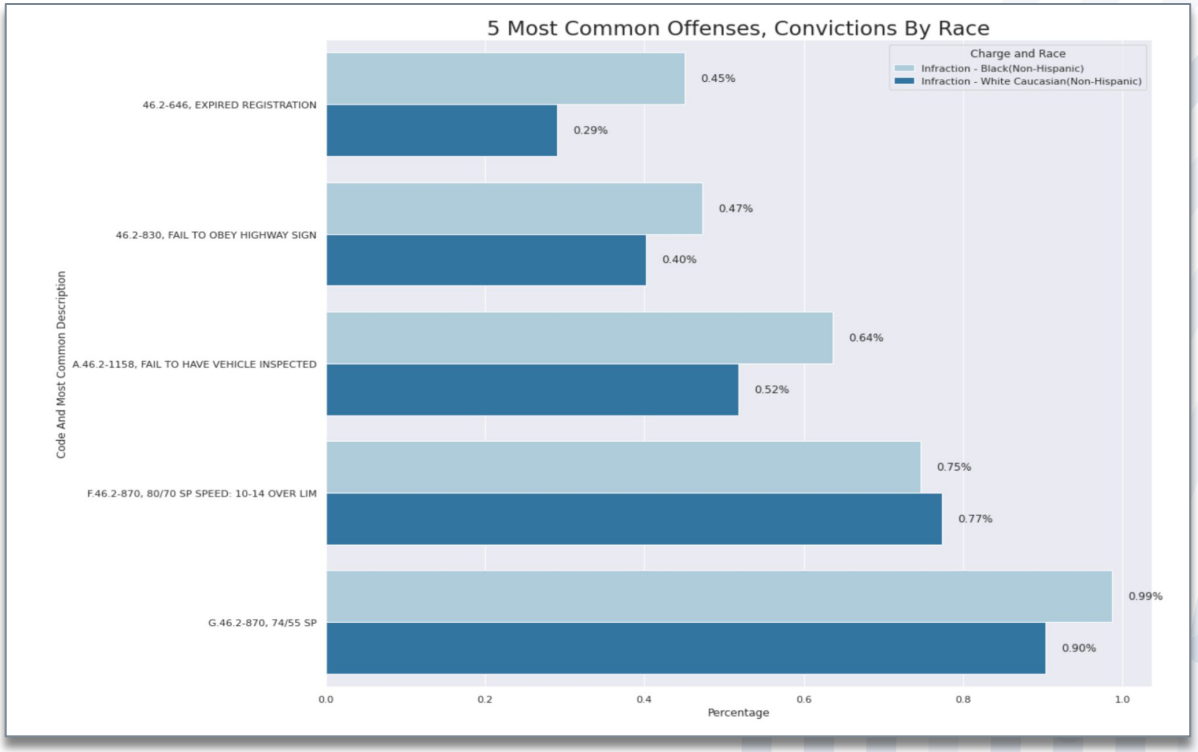
Select Charges As A Portion of Overall Population





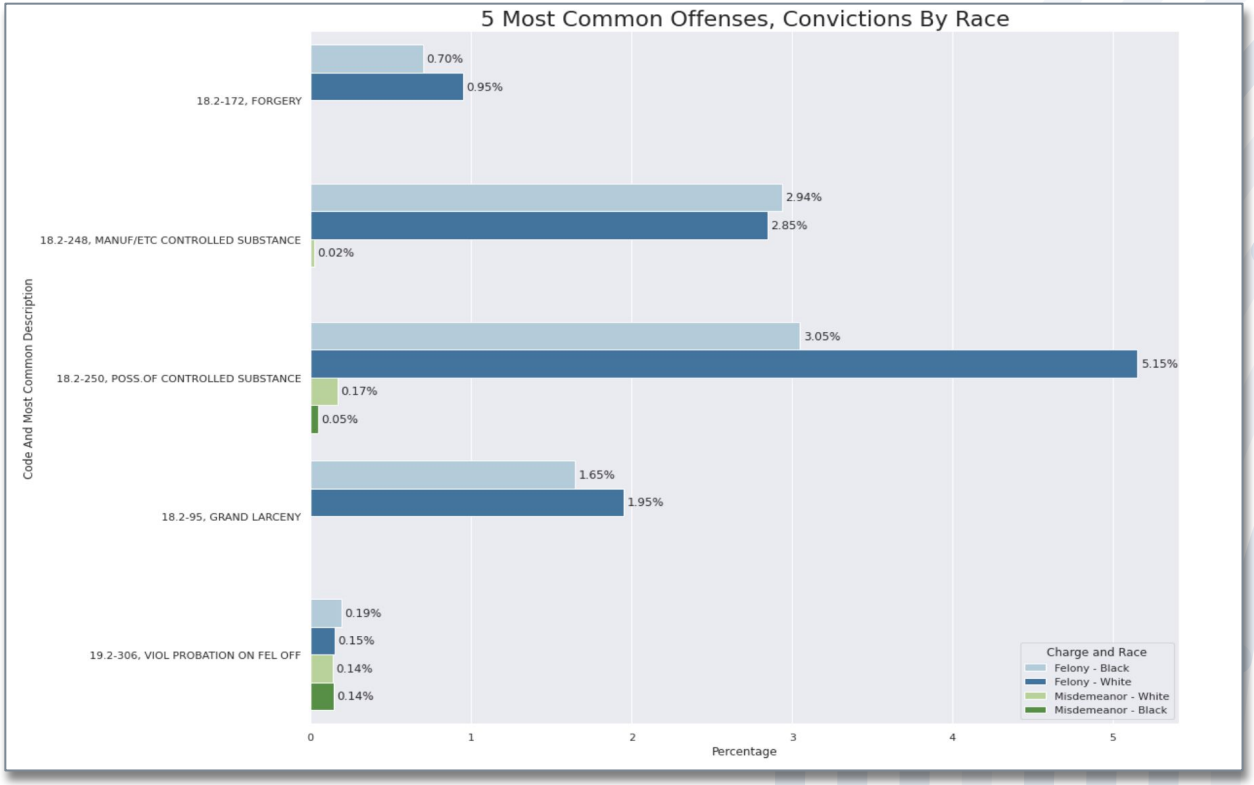
Exploratory Data Analysis

Infraction Convictions As Percentage of Population



Exploratory Data Analysis

Felony & Misdemeanor Convictions As % of Pop.



2. Random Forest for Feature Importance

Model Variables and Pre-Processing

- All Years: 2000-2020
- Response variable:
  - Misdemeanor or felony (0 or 1)
- Predictor variables:
  - Charge code, Race, Gender, County
- One-hot encoding
  - With one-hot encoding, each possible value receives an importance score
- Objective:
  - Models geared toward inference rather than prediction

Feature Importance with Random Forest

- Results:
  a. Charge code* is by far the most important feature
     - Top 35 most important features are all charge codes
  b. 5 out of top features 70 relate to counties
  c. *White* is the only race with non-zero feature importance, and is the 73rd most important
  d. Gender has a non-zero feature importance, but it is less important than other features listed

*Charge code = numeric value that accompanies a crime. For example, code 18.2-250 denotes all types of found marijuana possession

3. Logistic Regression Models

Logistic Regression Model

All Charges (Felony/Misdemeanor)

- Lasso model to encourage unnecessary predictors to be zero
- Able to successfully predict felony/misdemeanor from charge, county, race, and gender
  - AUC 96%; Precision 93%; Recall 97%
- Predictions for a few specific charges showed similar chances of felony charges for both white and Black

Predicted Likelihood of Being Charged With a Felony Marijuana Poss. w/ Intent, Albemarle County

	Black	White
Male	92.4%	92.2%
Female	91.8%	91.6%

Logistic Regression Model

Marijuana-Specific Charges (Felony/Misdemeanor)

- Used only race and gender as predictors
- Predicted likelihood of being charged with a felony was slightly higher for Black than white.

Predicted Likelihood of Being Charged With a Felony

	Black	White
Male	89.8%	88.3%
Female	89.1%	87.4%

4.
Conclusions

Conclusions

- Different races are charged with crimes at different rates relative to their population sizes
- No conclusive evidence that race plays a strong role in whether someone is charged with a felony or misdemeanor for the same crime

Next Steps

- Filter to only include first-time offenders, frequent repeat offenders may be adding unforeseen weights and incur harsher penalties than if it had been their first offense.

- Develop an interactive dashboard that allows non-technical users at the LAJC to sort and filter by crime, race, sex, county, etc. to better understand what areas are particularly susceptible to having racial disparities.

- Continue tuning machine learning model to understand the impact of race when used as a feature in crime analysis.

Thanks!

Any questions?

- Github: https://github.com/amawest/criminal-expungement
- Court Data: https://virginiacourtdata.org/
- Census Data: https://www.census.gov/quickfacts/fact/map/VA/RHI225219