BRSM - Initial Project Proposal

TEAM NAME: The Police did not investigate this murder

`Brahad Kokad - 2021112004`

`Radhikesh Agrawal - 2021113013`

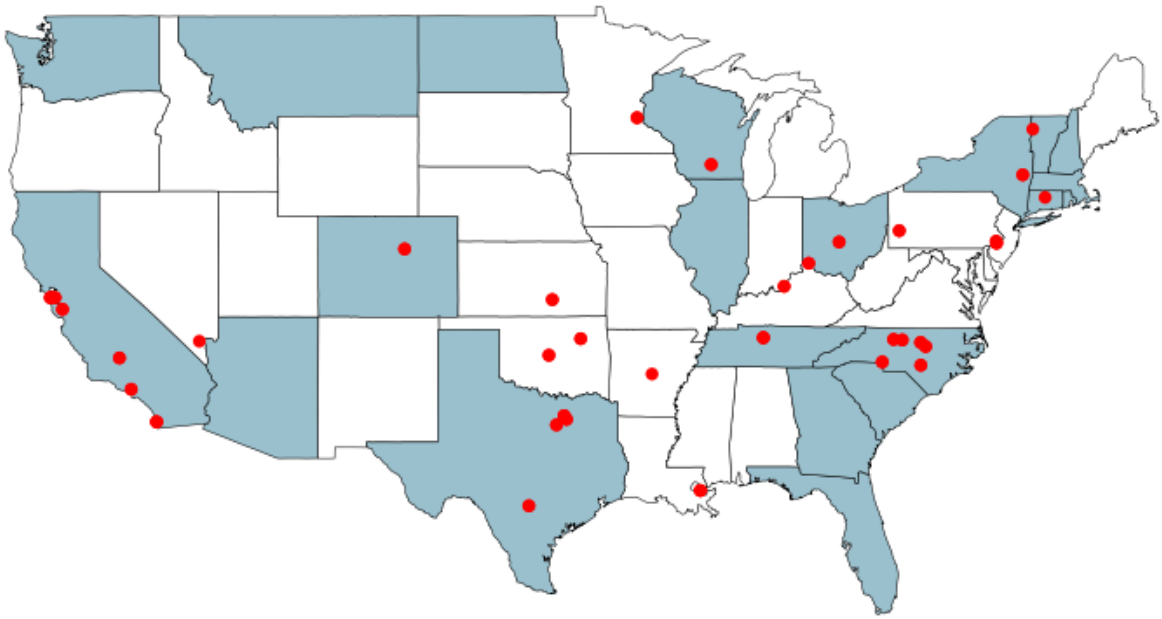
`Aarnav Nagariya - 2021112017`

Topic Selection

We intend to study racial disparities in police stops in the United States, for which we intend to use The Stanford Open Policing Project data.

Many of us might have faced some sort of discrimination in our lives, discrimination based on gender, religion, race and creed not only is a hindrance in development of as system but is also a serious threat to mankind in general. We are focusing on racial discrimination for this study. The USA seems to be a good place to base this study on because of a staggering contrast between races, and it's history of racial discrimination. Apart from getting relevant data, an inherent problem in doing a study on the India is that the Indian population by large is very diverse and the notion of contrasting races is diminished (though there is indeed discrimination in other forms, which are however, not in the scope of our study). Taking inspiration from this paper, we see that systematic racial discrimination in public services and bodies is a huge problem to justice and is a social evil at large.

About the data



Geographic coverage of compiled traffic stop data.

This dataset, encompassing nearly 100 million traffic stops across 21 US states, presents a valuable resource for understanding law enforcement interactions with drivers. Each data point

captures various factors like time, date, driver demographics (age, race, sex), citation information, vehicle type, "raw defendant race" etc.

Types of Analysis and Possible Findings

- Analyzing trends in citations issued based on driver demographics like race and sex could reveal potential biases in stop-and-search practices. The presence of "raw defendant race" raises questions about data collection protocols and potential biases embedded in them.
- Examining correlations between vehicle type and citations might show if specific vehicle types are disproportionately targeted.
- Comparing stop locations with accident data could inform targeted safety interventions.
- Correlating stop locations with socioeconomic data could explore potential connections between enforcement and certain population groups or areas.
- Examining citation types issued across different demographics could indicate socioeconomic disparities in traffic enforcement.
- The presence of "raw defendant race" suggests potential inconsistencies in data collection methods across different jurisdictions. Standardizing data collection practices would enhance the dataset's value and reduce biases.
- Ensuring data accuracy in aspects like time, date, and location is crucial for reliable analysis and drawing meaningful conclusions.

Hypotheses

Analyzing a dataset on police stops in the US for racial disparities is an important and complex task. Here are some hypotheses we could consider testing:

1. **Racial Disparity in Stop Rates (H1)**: Investigate whether certain racial or ethnic groups are stopped by police at higher rates than others, controlling for factors like location, time of day, and reason for the stop.
2. **Racial Disparity in Search Rates (H2)**: Investigate whether certain racial or ethnic groups are more likely to be subjected to searches during a police stop compared to others (even after controlling for factors like suspicion of criminal activity?).
3. **Racial Disparity in Outcome of Stops (H3)**: Examine whether the outcomes of police stops (e.g., issuance of citations, arrests) differ significantly based on the race or ethnicity of the individual stopped. Outcome includes the final decision made (and can also include police and rider's reactions?, but how? what metric?)
4. **Geographical Variation in Racial Disparities (H4)**: Investigate whether racial disparities in police stops vary significantly across different regions or jurisdictions within the US.
5. **Temporal Trends in Racial Disparities (H5)**: Analyze whether there have been changes in racial disparities in police stops over time, potentially indicating shifts in policing practices or policies.
6. **Intersectional Analysis (H6)**: Consider how the interaction between race and other factors (such as gender or socioeconomic status) may influence disparities in police stops and outcomes. (Correlation analysis?)
7. **Gender disparity in stops (H7)**: Examine Whether the reason for the police stop (e.g., traffic violation, suspicion of a more serious crime) is impacted by gender. (different levels of suspicion on women for simple crimes (signal breaking) vs serious crimes (marijuana)?)
8. **Community Demographics and Policing Practices (H8)**: Investigate whether the demographic composition of communities (e.g., racial composition, income levels) correlates with the

prevalence and nature of racial disparities in police stops. By community we mean the geosocial locality and population. Kinda related to #5.

These hypotheses can serve as points for our analysis, but it's crucial to approach the data with sensitivity, rigor, and an awareness of the broader social context surrounding policing and racial disparities in the US.

Analysis

To test our above hypothesis we plan to perform the following tests::

1. Descriptive Analysis

- **Use:** Establish baseline patterns in stop data across various groups and variables.
- **Application to Hypotheses:** Provides initial insights into disparities, trends, and distributions for all hypotheses.
- **Visualizations:** Bar charts, pie charts, and histograms for visualizing stop distributions by race, location, time, and reason.

2. Chi-Square Test of Independence

- **Use:** Determine if there's a significant association between categorical variables.
- **Application to Hypotheses:** Tests racial disparities in stop rates (H1) and search rates (H2).
- **Visualizations:** Heatmaps or stacked bar charts to visually represent the association between race and stop/search outcomes.

3. McNemar Test

- **Use:** Assess changes in paired categorical data.
- **Application to Hypotheses:** Limited direct application unless examining longitudinal changes in individuals.
- **Visualizations:** Line charts or paired bar charts to show changes in categorical outcomes over time.

4. Goodman and Kruskal's Lambda

- **Use:** Measure association strength between categorical variables.
- **Application to Hypotheses:** Useful for understanding the relationship between race and stop outcomes (H3).
- **Visualizations:** Mosaic plots or association plots to visualize the strength and direction of the association.

5. Linear Regression

- **Use:** Model relationships between variables, controlling for confounding factors.
- **Application to Hypotheses:**
  - H3: Quantify the impact of race on stop outcomes.
  - H4: Explore regional differences in racial disparities.
  - H5: Examine temporal trends in racial disparities.
  - H6: Analyze intersectional effects of race and other factors.
  - H7 & H8: Assess gender and community demographics' influence on stops and outcomes.

- **Visualizations:** Scatter plots, regression lines, or residual plots to visually represent relationships between variables

6. Partial and Semi-Partial Correlation

- **Use:** Evaluate relationships while controlling for confounding variables.
- **Application to Hypotheses:**
  - Before regression analysis (especially for H3, H4, H5, H6, H7, H8), use partial correlation to assess the relationship between race and the outcome, controlling for potential confounders.
  - Semi-partial correlation can be applied similarly but focuses on specific variables.
- **Visualizations:** Scatterplots or correlation matrices with controlled variables to understand the impact of confounders on the relationship between race and other variables.

7. Tests used by Authors

- **Veil of darkness test:** Potential bias in stop decisions is assessed using the 'veil of darkness' test developed by Grogger and Ridgeway21. This test is based on a simple premise: given the variation in sunset times throughout the year, the racial composition of stopped drivers can be examined in relation to sunlight while controlling for the time of day.
- **Threshold Test:** Potential bias in the post-stop decision to search drivers for contraband is investigated using the threshold test developed by Simoiu et al. and refined by Pierson et al. This test considers the rate and success rate of searches to infer the standard of evidence applied when determining whom to search.
- **Difference-in-differences strategy:** Analyze the impact of drug policy on racial disparities in traffic stop outcomes, focusing on a comparison between Colorado and Washington—two states that legalized recreational marijuana in 2012—and 12 states where recreational marijuana remained illegal.