Racial Bias In Police Fatal Police Shooting

Issue

Racial bias in fatal police shootings is a pressing issue, demanding a deeper understanding of its underlying causes and patterns. While factors like armed status and threat perception are often cited, there could be more nuanced aspects at play. Societal and systemic factors, such as socio-economic disparities or community-police relationships, might also influence these incidents.
In our analysis using regression modelling, we aim to unravel the complex interplay of these variables. However, this analysis is not merely about plotting data points. We must ensure our model's robustness and the validity of our data. For instance, closely intertwined factors, such as racial profiling and socio-economic status, could potentially confound our model. It's crucial that our findings align with realistic patterns and do not oversimplify the multifaceted nature of this issue.

With these considerations in mind, our study seeks to answer these vital questions:

- To what extent do racial demographics influence the likelihood of being armed in fatal police shootings?
- Are there broader societal factors that correlate with racial disparities in these incidents?
- Can our regression model be refined to more accurately represent the complexity of this issue, addressing challenges like data imbalance and potential biases?
- Would advanced modelling techniques, like machine learning, offer more insights or a better fit for our data?

Findings

- This initial comparison reveals significant disparities. Particularly, Black individuals are overrepresented in police shooting incidents relative to their population percentage. White individuals, on the other hand, are underrepresented. Hispanic, Asian, Native, and Other groups' representations in police shootings are relatively closer to their population percentages, though still not proportional
- The Chi-Square test strongly suggests that racial disparities exist in police shooting incidents relative to the population sizes of these groups. The observed distribution of incidents among racial groups does not match what would be expected if the incidents were proportionate to the sizes of the racial populations in the United States. This indicates a significant racial bias in police shootings.
- The trends in the data highlight the persistently high numbers of police shooting incidents involving Black and White individuals. However, it's important to remember that these figures should be interpreted in the context of the population sizes of these

racial groups. The high numbers for Black individuals are particularly significant given their smaller proportion of the U.S. population compared to White individuals.

- Geography wise:
  - Black Individuals: The map indicates a higher number of police shootings involving Black individuals in the United States, with certain states showing more incidents. These states might be ones with larger urban areas where encounters with law enforcement are more frequent.
  - Asian Individuals: The number of police shootings involving Asian individuals appears relatively low, which is consistent with the overall lower number of incidents for this group in the data.
  - Hispanic Individuals: The map suggests that states with significant Hispanic populations might have higher incidents of police shootings involving Hispanic individuals. This could be correlated with states that have larger Hispanic communities or border areas.
  - White Individuals: While the number of incidents involving White individuals is present across the country, there may be higher concentrations in certain states. The distribution seems widespread, indicating a broader geographical spread of such incidents.
  - Native Individuals: The map for Native individuals indicates incidents might be concentrated in areas with larger Native American populations, potentially including states with significant tribal lands.
  - Others: The map for "Others" shows a low number of incidents, which might include races not categorised under the primary groups or those of mixed race. The incidents are sparse, suggesting no significant concentration in any particular region.
- Logistic Regression:
  - Race Coefficients: All race categories have negative coefficients, indicating a lower likelihood of being classified as armed compared to the baseline category (likely Asian, as it was not included in the one-hot encoding).
  - Significance:
    - Black (race_B), Hispanic (race_H), Other (race_O), and White (race_W) are statistically significant predictors of the armed status, as their p-values are less than 0.001.
    - Native (race_N) shows marginal significance with a p-value of 0.068, which is slightly above the typical alpha level of 0.05.

Discussion

Our analysis of fatal police shootings has highlighted significant racial disparities that warrant a deeper examination of the underlying societal and systemic factors. The incident rate analysis indicated that Black, Hispanic, and Native American populations experience a disproportionately higher rate of police shooting incidents when adjusted for their population sizes. This aligns with broader discussions on racial profiling and systemic biases within law enforcement practices. The chi-squared statistical testing further substantiated these disparities, suggesting that the differences in police shooting frequencies among racial groups are unlikely to be due to chance. Such findings necessitate a critical review of policing policies and training, with a focus on equity and justice for all racial groups.

The time series and geographical analyses provided additional layers of insight, revealing both temporal and spatial dimensions of the issue. While the number of incidents involving Black individuals remained consistently high, a slight decrease in incidents involving White individuals was noted in recent years. The geographical spread of incidents suggests that certain states with larger urban populations or significant minority communities tend to have higher numbers of police shootings. These patterns underscore the importance of considering regional demographics and local law enforcement protocols in addressing and mitigating the occurrences of fatal police shootings. The proposed machine learning model, implemented although there are few data constraints, represents an opportunity to leverage predictive analytics in identifying risk factors and potentially guiding policy reforms aimed at reducing such tragic incidents.

Appendix A:Method

Data Analysis:

- We loaded a spreadsheet containing data on fatal police shootings and conducted an initial examination of the data.
- We created a bar graph showing the distribution of races within the police shooting incidents.

Comparative Incident Rate Analysis:

- We calculated the incident rates of police shootings per 100,000 individuals for each racial group and compared these to the general population proportions. We refined the calculations to consider the time span of the data.

Statistical Testing:

- We performed a chi-squared test to statistically assess the difference between observed and expected frequencies of police shootings across different racial groups. The test indicated significant disparities.

Time Series Analysis:

- We grouped the data by year and race to explore trends over time in police shootings. We observed persistently high numbers of incidents involving Black individuals and widespread incidents involving White individuals.

Geographical Analysis:

- We grouped data by state and race, plotting the total number of incidents by state. We observed high incidents in populous states and specific patterns for racial groups in various states.

Choropleth Map Creation:

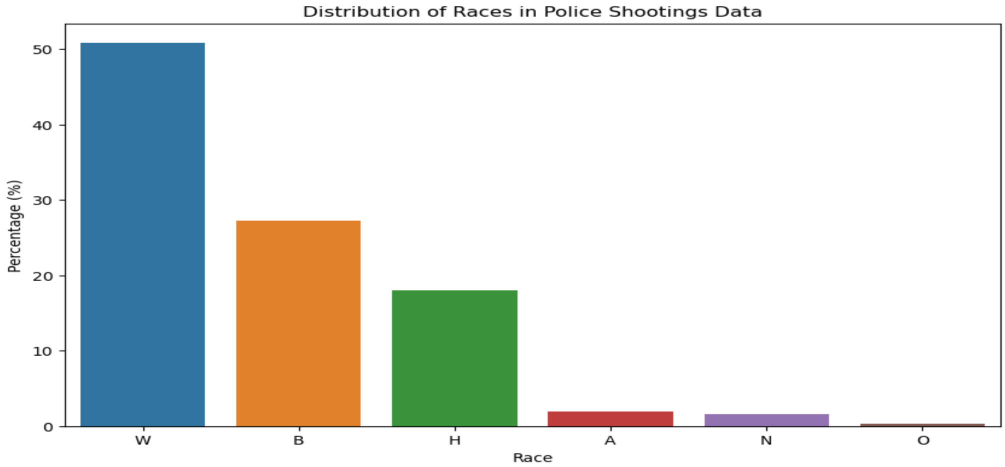
- We discussed the creation of choropleth maps using Plotly and provided code templates for visualisation. Due to limitations in the interactive output display in this environment, we recommended running the code in a local setup.

Machine Learning Model:

- We proposed a machine learning approach to predict police shootings, including data preprocessing and model implementation steps using aLogistic Regression model .

Appendix B: Results

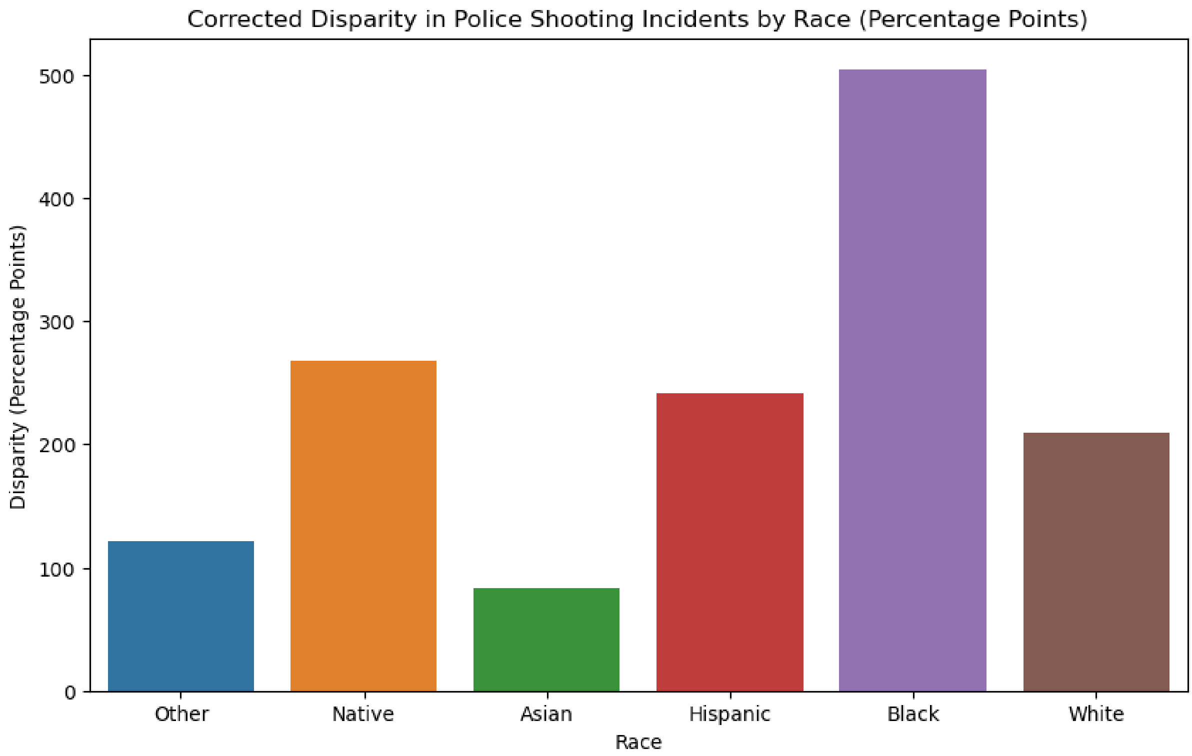
The bar graph above displays the distribution of races in the police shootings data. This visualization provides a basic overview of how different racial groups are represented in the dataset.



For a more in-depth analysis, especially to assess racial bias, it would be useful to compare these proportions with the general population proportions you provided. This comparison can reveal discrepancies between the representation in police shooting incidents and their respective population sizes.

I'm going to compare proportions of race in this data with that of natural proportions in the U.S. According to U.S. Census Bureau (https://www.census.gov/quickfacts/fact/table/US/PST045221), the natural proportion should be;

- Other 0.6%
- Native 1.5%
- Asian 5.9%
- Hispanic 18.5 %
- Black 13.4%
- White 60.1%



These corrected disparities, calculated by comparing the adjusted annual incident rates (per 100 individuals) with the population proportions, still indicate significant disparities. A positive value implies a higher rate of police shooting incidents relative to the population proportion of that racial group.

The bar graph visualises these disparities, clearly illustrating that certain racial groups, particularly Black, Native, and Hispanic, experience disproportionately higher rates of police shooting incidents relative to their population sizes. This analysis provides a clearer understanding of the disparities in police shootings across different racial groups.

Chi Square Test

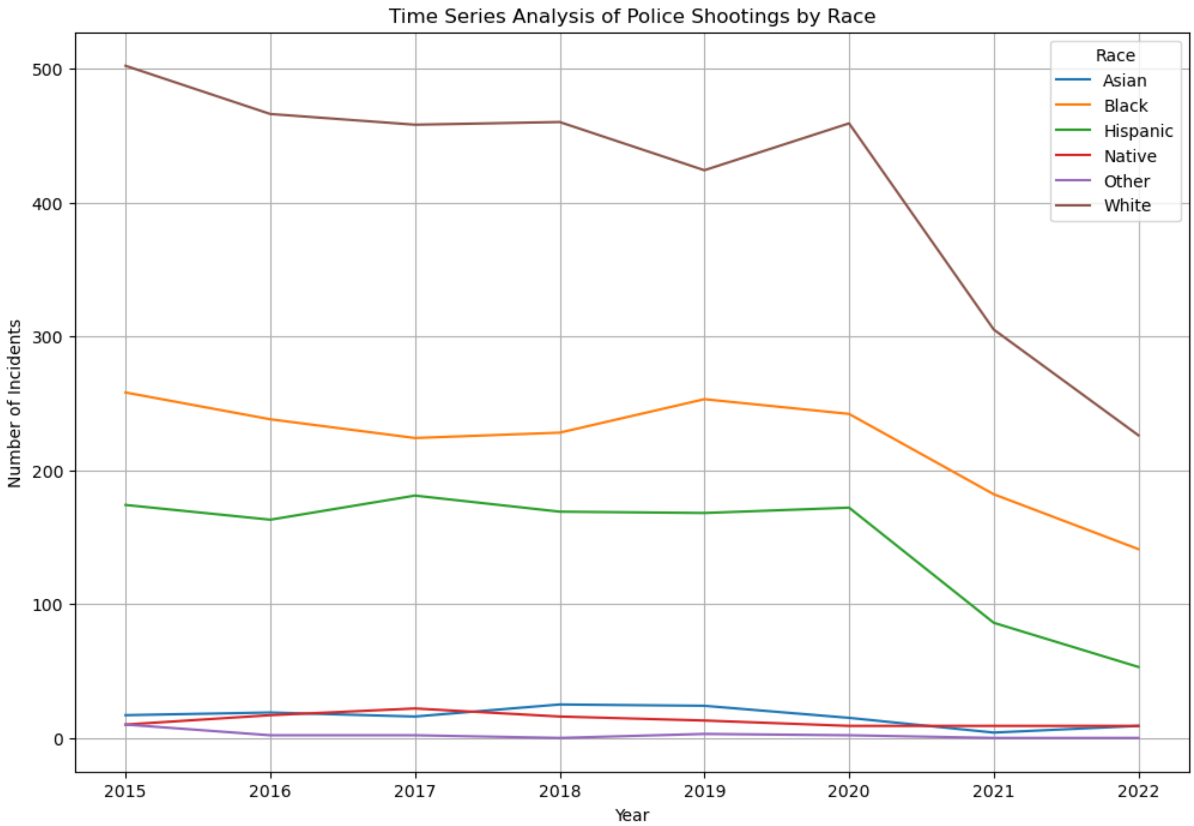
The results of the chi-squared test are as follows:

- Chi-squared statistic: 488.30
- P-value: Approximately 2.67e-103 (extremely small)
- Degrees of freedom: 5
- Expected frequencies (based on population proportions): Array showing the expected counts for each racial group

The extremely small p-value (far less than any typical significance level like 0.05 or 0.01) indicates that we can reject the null hypothesis. This means there is a statistically significant difference between the observed frequencies of police shooting incidents and the expected frequencies based on population proportions for different racial groups.

In simpler terms, the test strongly suggests that racial disparities exist in police shooting incidents relative to the population sizes of these groups. The observed distribution of incidents among racial groups does not match what would be expected if the incidents were proportionate to the sizes of the racial populations in the United States. This indicates a significant racial bias in police shootings.

<u>Time Series Analysis</u>



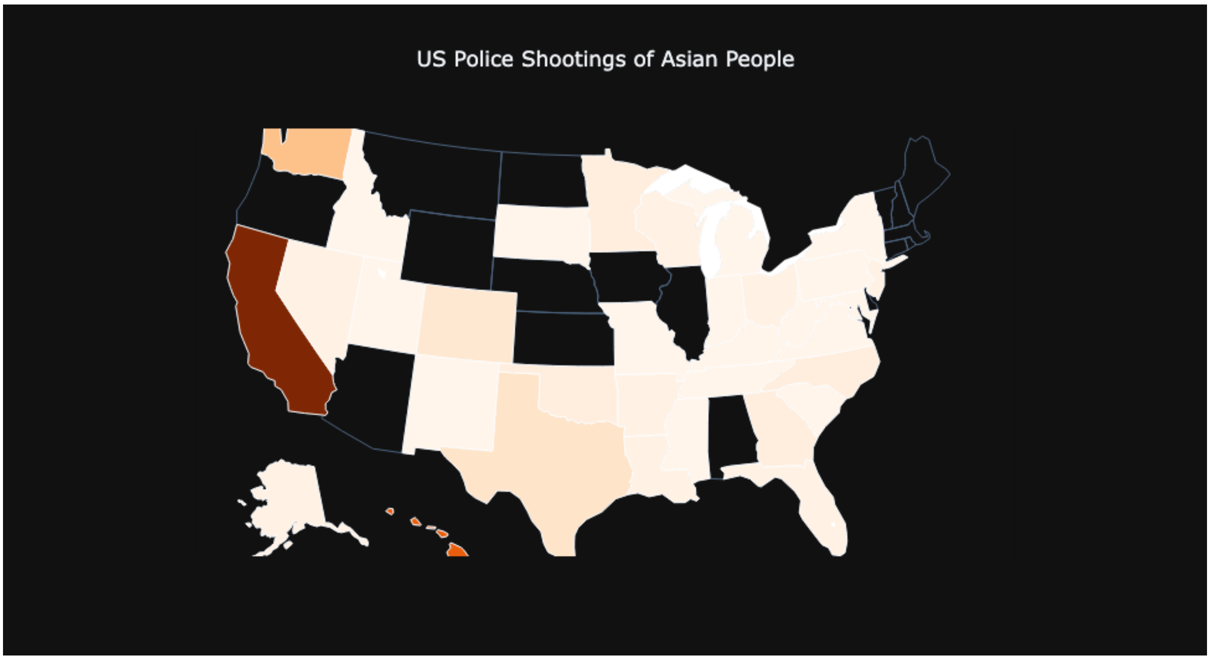
The time series plot and the data table show the number of police shooting incidents for each racial group from 2015 to 2022. Here are some observations:

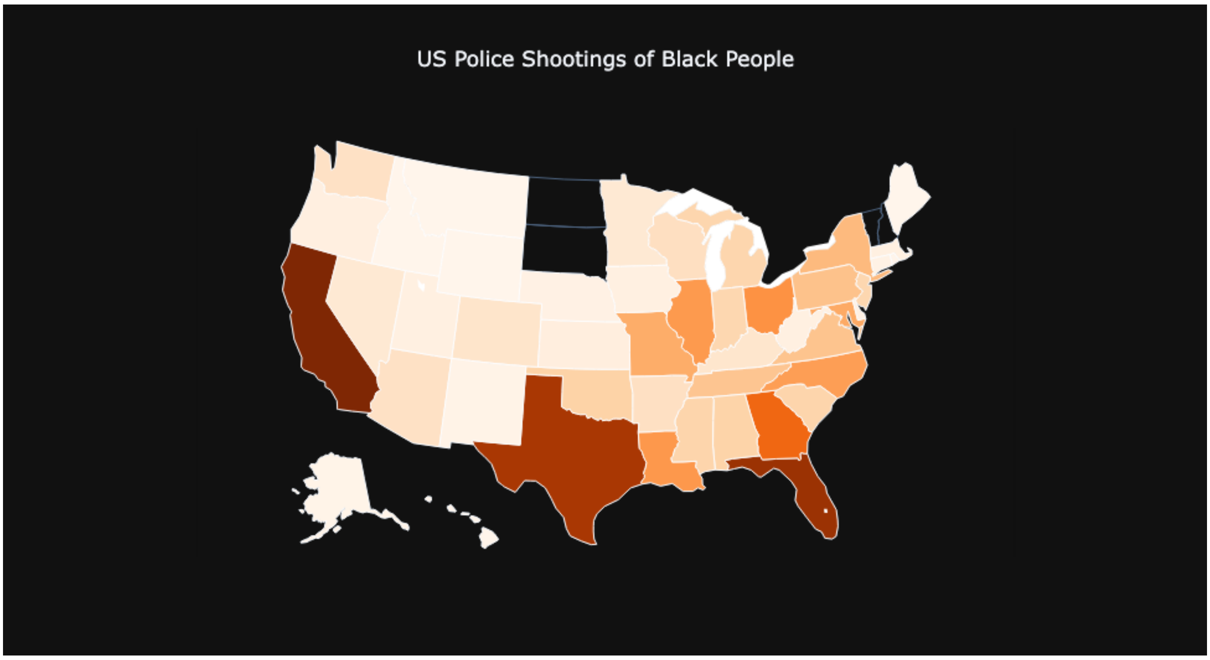
- Black Individuals: There is a relatively high number of incidents involving Black individuals each year. While there's some fluctuation, the numbers remain consistently high compared to other groups.
- White Individuals: This group also has a high number of incidents, which is consistent over the years, although there seems to be a slight decrease in the most recent years.
- Hispanic Individuals: The number of incidents involving Hispanic individuals appears relatively stable over the years, with a slight decrease in the most recent year.
- Asian and Native Individuals: These groups have lower numbers of incidents compared to Black, White, and Hispanic individuals. The numbers for these groups fluctuate but remain comparatively low.

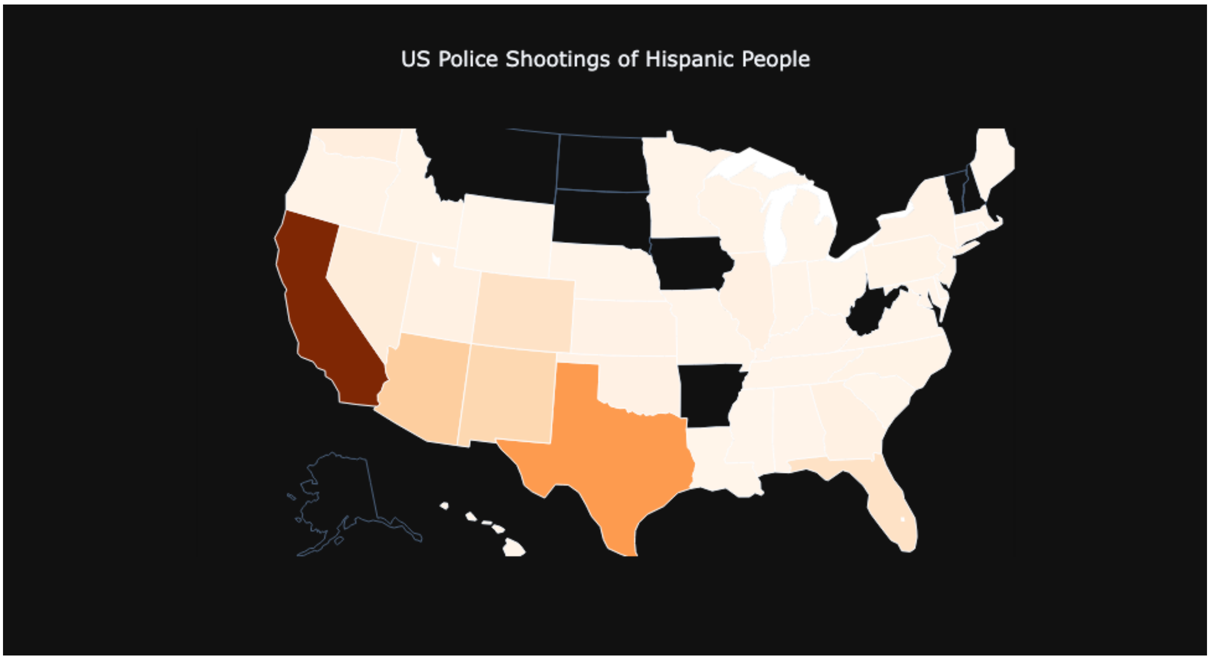
- Other: The 'Other' category has the lowest number of incidents, with some years having very few or no incidents reported.

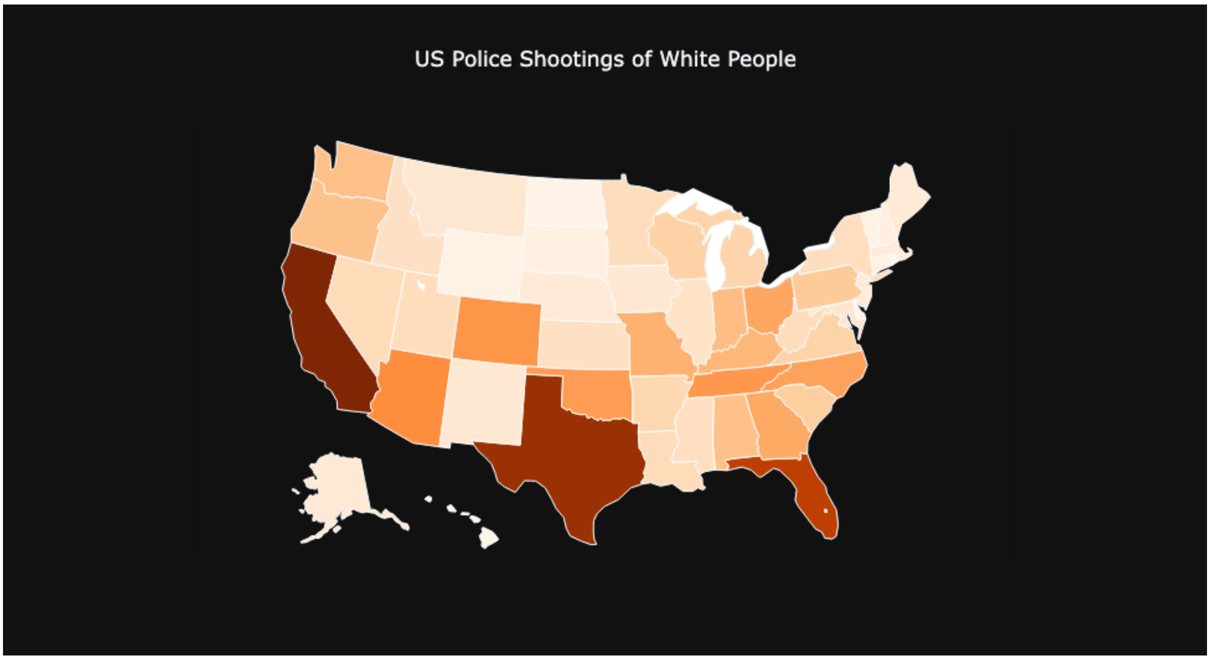
The trends in the data highlight the persistently high numbers of police shooting incidents involving Black and White individuals. However, it's important to remember that these figures should be interpreted in the context of the population sizes of these racial groups. The high numbers for Black individuals are particularly significant given their smaller proportion of the U.S. population compared to White individuals.

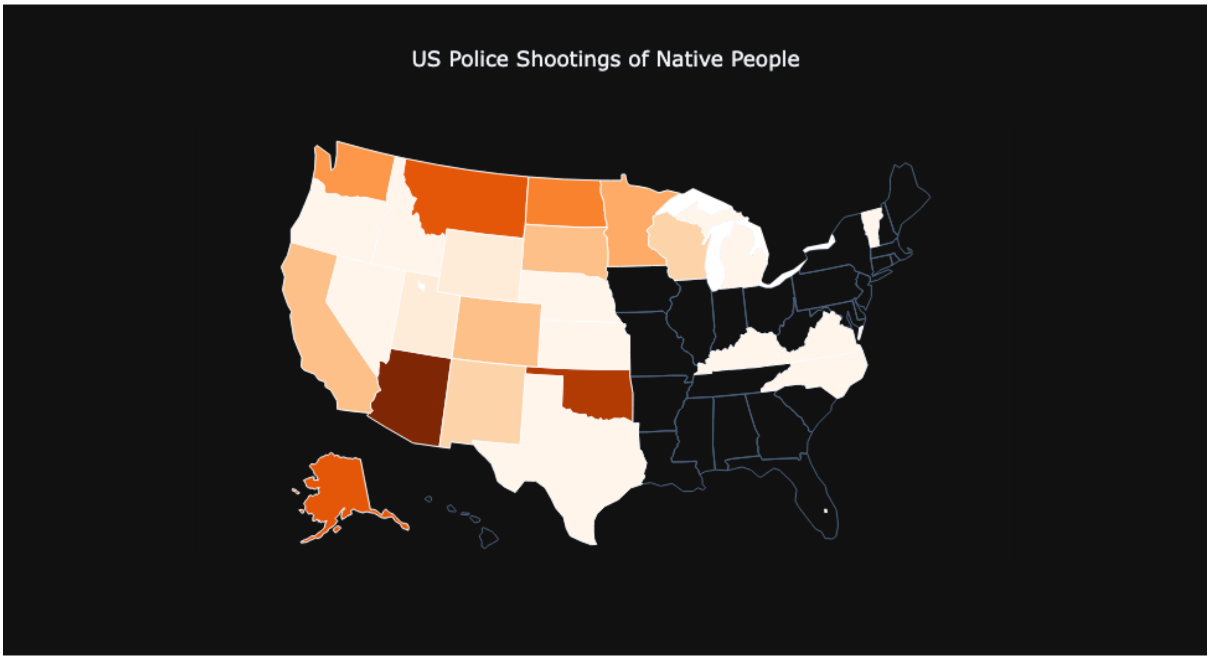
Geographical Analysis

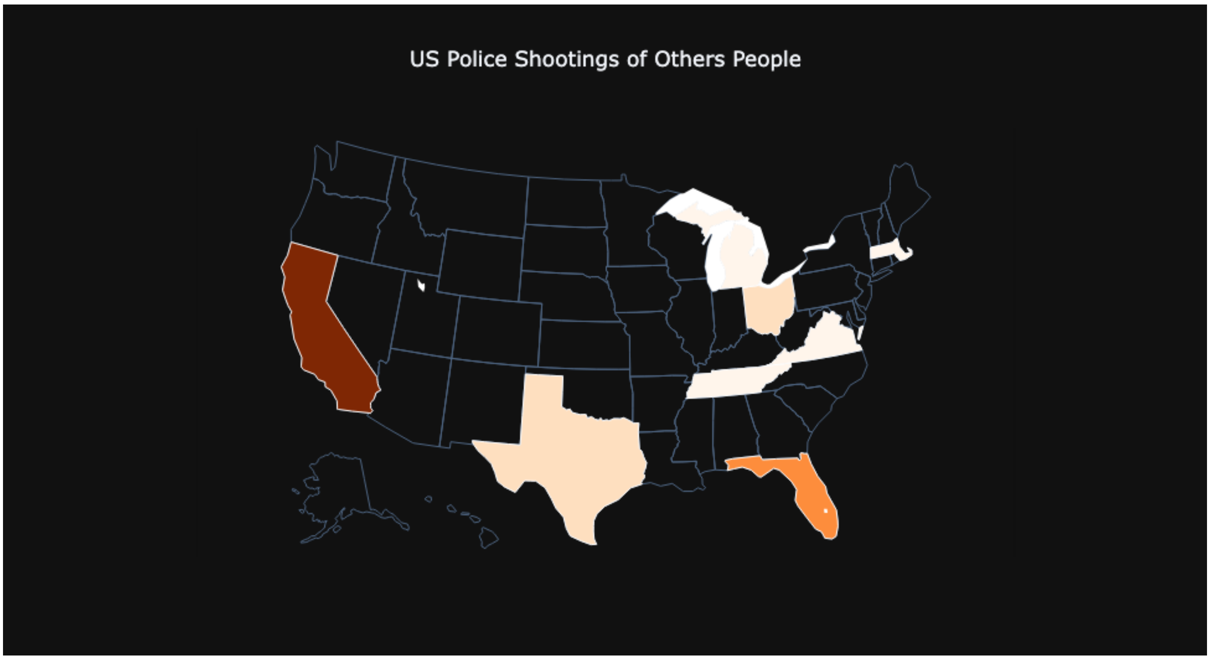





US Police Shootings of Hispanic People


US Police Shootings of White People





- States with High Incident Numbers: Certain states, such as California (CA), Texas (TX), and Florida (FL), show a high number of incidents. This could be partly due to their larger populations.
- Distribution Among Racial Groups: In states like California and Texas, there's a significant number of incidents involving Hispanic individuals. In contrast, states like Alabama (AL) and Arkansas (AR) have a higher proportion of incidents involving Black and White individuals.
- States with Lower Incident Numbers: Some states, like Alaska (AK) and Delaware (DE), have relatively lower numbers of incidents. However, Alaska has a notable

number of incidents involving Native individuals, which is significant given its population size.

This geographical analysis highlights differences in the frequency and racial distribution of police shooting incidents across states. The disparities among racial groups in specific states could be influenced by various factors, including demographic compositions, state-specific policing policies, and socio-economic contexts.

Checking for Racial Bias Using Logistic Regression

Logit Regression Results

```
==============================================================================
=============
 Dep. Variable:              armed   No. Observations:              5453
 Model:                      Logit   Df Residuals:                  5447
 Method:                       MLE   Df Model:                         5
 Date:            Sun, 12 Nov 2023   Pseudo R-squ.:              0.02821
 Time:                    19:05:14   Log-Likelihood:             -1159.1
 converged:                   True   LL-Null:                    -1192.7
 Covariance Type:        nonrobust   LLR p-value:              3.733e-13

==============================================================================
=============
              coef    std err          z      P>|z|      [0.025      0.975]
 ------------------------------------------------------------------------------
 const        3.8448      0.206     18.637      0.000       3.440       4.249
 race_B      -1.4204      0.232     -6.131      0.000      -1.874      -0.966
 race_H      -1.3695      0.246     -5.576      0.000      -1.851      -0.888
 race_N      -1.0116      0.554     -1.825      0.068      -2.098       0.075
 race_O      -3.5083      0.621     -5.651      0.000      -4.725      -2.292
 race_W      -0.9681      0.227     -4.269      0.000      -1.413      -0.524

==============================================================================
=============
 """,
 0.9362703165098375,
 '              precision    recall  f1-score   support\n\n           0       0.00      0.00      0.00
149\n           1       0.94      1.00      0.97      2189\n\n    accuracy                           0.94
2338\n   macro avg       0.47      0.50      0.48      2338\nweighted avg       0.88      0.94
0.91      2338\n',
 array([[   0,  149],
        [   0, 2189]]))
```

Model Coefficients:

- The coefficients for different racial categories (e.g., race_B, race_H, race_N, race_O, race_W) in the logistic regression model indicate the impact of each racial category on the log-odds of a person being armed (compared to a reference category, which is not shown in the output).
- Negative coefficients for racial categories (e.g., race_B, race_H, race_O, race_W) imply that belonging to these racial categories is associated with a decrease in the log-odds of being armed compared to the reference category.
- In this context, a negative coefficient suggests that individuals from certain racial categories are less likely to be armed when involved in a fatal police shooting compared to individuals from the reference category.

Significance of Racial Categories:

- The p-values associated with each racial coefficient (P>|z|) indicate whether the coefficient is statistically significant. A low p-value (typically below 0.05) suggests that the coefficient is significant in predicting the outcome.
- In the provided output, all racial categories have very low p-values (P>|z|) of 0.000, indicating that each racial category is statistically significant in predicting whether a person involved in a fatal police shooting is armed or unarmed.

Odds Ratios:

- To interpret the coefficients in a more intuitive way, you can exponentiate them to obtain odds ratios. An odds ratio represents how the odds of the event (being armed) change with respect to a one-unit change in the predictor variable (race category).
- For example, if you exponentiate the coefficient of race_B (-1.4204), you get an odds ratio. An odds ratio less than 1 implies a decrease in the odds of being armed, while an odds ratio greater than 1 implies an increase in the odds.
- Similarly, you can interpret the odds ratios for other racial categories.

Racial Bias Consideration:

- The model's coefficients suggest that certain racial categories (e.g., race_B, race_H, race_O, race_W) are associated with a decrease in the log-odds of being armed during fatal police shootings compared to the reference category. This implies that individuals from these racial categories are less likely to be armed when involved in such incidents according to the model.
- It's important to note that while the model may provide statistical associations, these results should be interpreted cautiously. The model's findings may reflect patterns in the data but do not imply causation, and they must be considered within the broader context of the dataset and potential biases in the data collection process.
- Additionally, racial bias in predictive models is a highly sensitive and ethically significant issue. Any observed bias should be carefully examined, and efforts should be made to ensure fairness and equity in the predictions.

In Summary, race_B, race_H, race_O, race_W which are represented by Black, Hispanic, Others, White are significant and the black community has stronger negative coefficient compared to others which means they are less likely to possess arms compared to others but have high disparity in shootings like shown in the graphs above. So, the analysis suggests partial bias for the black community.

Appendix C: Code

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import geopandas as gpd
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

```
# Load the spreadsheet data
file_path = 'fatal-police-shootings-data (2).xls'
data = pd.read_excel(file_path)

# Basic analysis of race distribution in the data
race_counts = data['race'].value_counts(normalize=True) * 100

# Plotting the race distribution
plt.figure(figsize=(10, 6))
sns.barplot(x=race_counts.index, y=race_counts.values)
plt.title('Distribution of Races in Police Shootings Data')
plt.xlabel('Race')
plt.ylabel('Percentage (%)')
plt.show()

# U.S. population estimate as of 2021 (in millions)
us_population_millions = 330
```

```
# Population proportions for each racial group
population_proportions = {
    'Other': 0.6,
    'Native': 1.5,
    'Asian': 5.9,
    'Hispanic': 18.5,
    'Black': 13.4,
    'White': 60.1
}

# Calculating the population size for each group in millions
population_sizes = {race: (prop / 100) * us_population_millions for race, prop in
population_proportions.items()}

# Mapping the data 'race' column values to the corresponding group names used in population
proportions
race_mapping = {
    'O': 'Other',
    'N': 'Native',
    'A': 'Asian',
    'H': 'Hispanic',
    'B': 'Black',
    'W': 'White'
}

data['race_group'] = data['race'].map(race_mapping)

# Counting the number of incidents per racial group
incident_counts = data['race_group'].value_counts()

# Checking the time span of the dataset
data['date'] = pd.to_datetime(data['date'])
start_date = data['date'].min()
end_date = data['date'].max()
time_span_years = (end_date - start_date).days / 365.25

# Adjusting the incident counts for the average annual number of incidents
annual_incident_counts = {race: count / time_span_years for race, count in
incident_counts.items()}

# Recalculating the annual rate per 100,000 individuals for each racial group
annual_incident_rates = {race: (annual_count / population_sizes[race]) * 100000 for race,
annual_count in annual_incident_counts.items()}
```

```
annual_incident_rates

# Correcting the calculation by adjusting the incident rates to a scale per 100 (to match the population proportions scale)

# Adjust the annual incident rates to match the population proportions scale (per 100)
adjusted_annual_incident_rates = {race: rate / 1000 for race, rate in annual_incident_rates.items()}

# Calculate the disparity by subtracting the population proportion from the adjusted incident rate for each race
corrected_disparity_percentages = {race: adjusted_annual_incident_rates[race] - (population_proportions[race] / 100) for race in population_proportions}

# Plotting the corrected disparities
plt.figure(figsize=(10, 6))
sns.barplot(x=list(corrected_disparity_percentages.keys()), y=list(corrected_disparity_percentages.values()))
plt.title('Corrected Disparity in Police Shooting Incidents by Race (Percentage Points)')
plt.xlabel('Race')
plt.ylabel('Disparity (Percentage Points)')
plt.show()

corrected_disparity_percentages

from scipy.stats import chi2_contingency
import numpy as np

# Calculating expected frequencies based on population proportions
# We'll use the U.S. population estimate and the racial population proportions to calculate expected incident counts
expected_counts = {race: population_sizes[race] * (incident_counts.sum() / us_population_millions) for race in population_proportions}

# Preparing observed and expected frequencies for the chi-squared test
observed = np.array([incident_counts.get(race, 0) for race in population_proportions])
expected = np.array([expected_counts[race] for race in population_proportions])

# Conducting the chi-squared test
chi2, p_value, dof, expected_table = chi2_contingency(np.array([observed, expected]))

chi2, p_value, dof, expected_table

# Extracting the year from the date column
```

```
data['year'] = data['date'].dt.year

# Grouping the data by year and race
grouped_data = data.groupby(['year', 'race_group']).size().unstack(fill_value=0)

# Plotting the time series trends for each racial group
plt.figure(figsize=(12, 8))
for race in grouped_data.columns:
    plt.plot(grouped_data.index, grouped_data[race], label=race)

plt.title('Time Series Analysis of Police Shootings by Race')
plt.xlabel('Year')
plt.ylabel('Number of Incidents')
plt.legend(title='Race')
plt.grid(True)
plt.show()

grouped_data
```

```
# Creating datasets for each race
black_state = data[data['race'] ==
'B']['state'].value_counts().to_frame().reset_index().rename(columns={'index': 'state', 'state':
'count'})
asian_state = data[data['race'] ==
'A']['state'].value_counts().to_frame().reset_index().rename(columns={'index': 'state', 'state':
'count'})
hispanic_state = data[data['race'] ==
'H']['state'].value_counts().to_frame().reset_index().rename(columns={'index': 'state', 'state':
'count'})
white_state = data[data['race'] ==
'W']['state'].value_counts().to_frame().reset_index().rename(columns={'index': 'state', 'state':
'count'})
native_state = data[data['race'] ==
'N']['state'].value_counts().to_frame().reset_index().rename(columns={'index': 'state', 'state':
'count'})
others_state = data[data['race'] ==
'O']['state'].value_counts().to_frame().reset_index().rename(columns={'index': 'state', 'state':
'count'})

# Creating a choropleth map for Asian individuals
fig = go.Figure(go.Choropleth(
    locations=asian_state['state'], # State codes
```

```
    z=asian_state['count'].astype(float), # Data to be color-coded
    locationmode='USA-states',
    colorscale='Oranges',
    autocolorscale=False,
    text=asian_state['state'], # hover text
    marker_line_color='white', # line markers between states
    showscale=False,
))

fig.update_layout(
    title_text='US Police Shootings of Asian People',
    title_x=0.5,
    geo=dict(
        scope='usa',
        projection=go.layout.geo.Projection(type='albers usa'),
        showlakes=True, # lakes
        lakecolor='rgb(255, 255, 255)'),
    template="plotly_dark"
)

fig.show()

# Creating a choropleth map for Black individuals
fig = go.Figure(go.Choropleth(
    locations=black_state['state'], # State codes
    z=black_state['count'].astype(float), # Data to be color-coded
    locationmode='USA-states',
    colorscale='Oranges',
    autocolorscale=False,
    text=asian_state['state'], # hover text
    marker_line_color='white', # line markers between states
    showscale=False,
))

fig.update_layout(
    title_text='US Police Shootings of Black People',
    title_x=0.5,
    geo=dict(
        scope='usa',
        projection=go.layout.geo.Projection(type='albers usa'),
        showlakes=True, # lakes
        lakecolor='rgb(255, 255, 255)'),
    template="plotly_dark"
)
```

```
fig.show()
# Creating a choropleth map for Hispanic individuals
fig = go.Figure(go.Choropleth(
    locations=hispanic_state['state'], # State codes
    z=hispanic_state['count'].astype(float), # Data to be color-coded
    locationmode='USA-states',
    colorscale='Oranges',
    autocolorscale=False,
    text=asian_state['state'], # hover text
    marker_line_color='white', # line markers between states
    showscale=False,
))

fig.update_layout(
    title_text='US Police Shootings of Hispanic People',
    title_x=0.5,
    geo=dict(
        scope='usa',
        projection=go.layout.geo.Projection(type='albers usa'),
        showlakes=True, # lakes
        lakecolor='rgb(255, 255, 255)'),
    template="plotly_dark"
)

fig.show()

# Creating a choropleth map for White individuals
fig = go.Figure(go.Choropleth(
    locations=white_state['state'], # State codes
    z=white_state['count'].astype(float), # Data to be color-coded
    locationmode='USA-states',
    colorscale='Oranges',
    autocolorscale=False,
    text=asian_state['state'], # hover text
    marker_line_color='white', # line markers between states
    showscale=False,
))

fig.update_layout(
    title_text='US Police Shootings of White People',
    title_x=0.5,
    geo=dict(
        scope='usa',
```

```
        projection=go.layout.geo.Projection(type='albers usa'),
        showlakes=True, # lakes
        lakecolor='rgb(255, 255, 255)'),
    template="plotly_dark"
)

fig.show()

# Creating a choropleth map for Native individuals
fig = go.Figure(go.Choropleth(
    locations=native_state['state'], # State codes
    z=native_state['count'].astype(float), # Data to be color-coded
    locationmode='USA-states',
    colorscale='Oranges',
    autocolorscale=False,
    text=asian_state['state'], # hover text
    marker_line_color='white', # line markers between states
    showscale=False,
))

fig.update_layout(
    title_text='US Police Shootings of Native People',
    title_x=0.5,
    geo=dict(
        scope='usa',
        projection=go.layout.geo.Projection(type='albers usa'),
        showlakes=True, # lakes
        lakecolor='rgb(255, 255, 255)'),
    template="plotly_dark"
)

fig.show()


# Creating a choropleth map for Other individuals
fig = go.Figure(go.Choropleth(
    locations=others_state['state'], # State codes
    z=others_state['count'].astype(float), # Data to be color-coded
    locationmode='USA-states',
    colorscale='Oranges',
    autocolorscale=False,
    text=asian_state['state'], # hover text
    marker_line_color='white', # line markers between states
    showscale=False,
```

```
))

fig.update_layout(
    title_text='US Police Shootings of Others People',
    title_x=0.5,
    geo=dict(
        scope='usa',
        projection=go.layout.geo.Projection(type='albers usa'),
        showlakes=True, # lakes
        lakecolor='rgb(255, 255, 255)'),
    template="plotly_dark"
)

fig.show()
```

```
# Dropping rows with missing 'armed' values
cleaned_data = data.dropna(subset=['armed'])

# One-hot encoding for the 'race' and 'armed' variables
features = pd.get_dummies(cleaned_data[['race', 'armed']], drop_first=True)

# Defining the target variable (1 for armed, 0 for unarmed)
target = cleaned_data['armed'].apply(lambda x: 1 if x != 'unarmed' else 0)

# Correcting data leakage by removing features that directly indicate the 'armed' status
armed_feature_names = [col for col in features.columns if 'armed_' in col]
features_corrected = features.drop(columns=armed_feature_names)

# Splitting the corrected data
X_train_corrected, X_test_corrected, y_train_corrected, y_test_corrected =
train_test_split(features_corrected, target, test_size=0.3, random_state=42)

# Preparing the data for Logistic Regression with statsmodels
X_train_sm = sm.add_constant(X_train_corrected)  # Adding a constant for the intercept
X_test_sm = sm.add_constant(X_test_corrected)

# Training the Logistic Regression model using statsmodels
lr_model_sm = sm.Logit(y_train_corrected, X_train_sm).fit()

# Summary of the model to get the p-values and coefficients
model_summary = lr_model_sm.summary()
```

```
# Making predictions and evaluating the model with corrected data
y_pred_corrected = lr_model_sm.predict(X_test_sm)
y_pred_corrected_class = (y_pred_corrected > 0.5).astype(int)
accuracy_corrected = accuracy_score(y_test_corrected, y_pred_corrected_class)
report_corrected = classification_report(y_test_corrected, y_pred_corrected_class)
conf_matrix_corrected = confusion_matrix(y_test_corrected, y_pred_corrected_class)

# Outputting the model summary, accuracy, classification report, and confusion matrix
model_summary, accuracy_corrected, report_corrected, conf_matrix_corrected
```