STA210 SP'24 Final Project

Exploring 2023 Stop and Frisk Data in NYC

amaris

Introduction

Background:

The stop-and-frisk program in New York City, administered by the NYPD, allows officers to detain, question, and potentially search individuals suspected of carrying weapons or contraband. This initiative has sparked significant controversy due to concerns of racial profiling. In 2017, 90% of those stopped were African-American or Latino, primarily aged between 14 and 24. Despite efforts to address racial disparities, such as policy reforms, the disproportionate impact of the stop-and-frisk program persists, highlighting potential underlying factors like implicit bias.

Implicit bias, also known as implicit prejudice or implicit attitude, is a negative attitude, of which one is not consciously aware, against a specific social group. It is thought to be shaped by experience and based on learned associations between particular qualities and social categories, including race and/gender/age etc. Individuals' perceptions and behaviors can be influenced by the implicit biases they hold, even if they are unaware they hold such biases.

Dataset:

Each stop made by the NYPD requires officers to complete a detailed form, documenting various aspects of the encounter. Since 2017, these forms have been electronically recorded and stored in an NYPD database. The dataset contains information such as the stop's location, officer details, characteristics of the stopped individual (including age, race, gender, etc.), frisk/search details, and the officer's description of the individual's demeanor during the stop. Our analysis will utilize the most recently released NYPD annual report from the source: https://www.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page, containing 82 variables and 16,871 observations.

Project Motivation & Research Question:

Among the 82 variables, a variable of particular interest is "demeanor of person stopped" - where the police utilize 1 - 2 adjectives to describe stop subject "demeanor". Common adjectives include "calm", "nervous", "agitated", "aggressive", etc. These descriptions are self-generated instead of the police choosing from a pre-defined set of adjectives. We propose that these "demeanor" adjectives are indicative of the police officers' perception of the stopped subject. This project aims to investigate the relationship between physical/demographic characteristics of stopped individuals and the demeanor adjectives assigned by police officers. Specifically, we will explore:

- How do officer-assigned demeanor adjectives vary across different demographic groups (age, race, gender)?
- Are there correlations between certain physical characteristics and the types of demeanor descriptions used by officers during stops?
- Additionally, we will briefly examine whether demeanor descriptions influence subsequent police behaviors, such as frisking, searching, or requesting consent.

By analyzing these relationships, we seek to shed light on potential implicit biases affecting police interactions during stop-and-frisk encounters. Understanding these dynamics is crucial for addressing systemic biases and ensuring fair and equitable policing practices.

Variables Introduction:

Predictor Variables of Interest	
SUSPECT_REPORTED_AGE	(chr and transformed to num) the age of suspect
SUSPECT_SEX	(chr) whether the suspect is female or male
SUSPECT_RACE_DESCRIPTION	(chr) includes 7 categories: American Indian/Alaskan Native, Asian/Pacific Islander, Black, Black Hispanic, Middle Eastern/Southwest Asian, White, White Hispanic
SUSPECT_HEIGHT	(chr and transformed to num) the height of suspect by feet
SUSPECT_WEIGHT	(chr and transformed to num) the weight of suspect by pounds
SUSPECT_BODY_BUILD_TYPE	(chr) includes categories HEA(Heavy), MED(Medium), THN(Thin), U(Unknown), XXX(body type not applicable/placeholder value indicating missing data)
SUSPECT_EYE_COLOR	(chr) includes categories: BLK(Black), BLU(Blue), BRO(Brown), GRN(Green), GRY(Grey), HAZ(Hazel), MUL(Multicolored), OTH(Other), PNK(Pink)
SUSPECT_HAIR_COLOR	(chr) includes categories: BLD (Bald), BLK (Black), BLN (Blonde), BRO (Brown), GRN(Green), GRY (Gray), ORG (Orange), PLE (Purple), PNK(Pink), RED(Red), SDY(Sandy), WHI (White), XXX (Unknown/Unspecified - often used when the suspect's hair color is not recorded or unclear), ZZZ (could be an unusual or placeholder value indicating an error or missing data)
Note: *The interpretation of categorical variables is based on conventions and assumptions due to the absence of a specific codebook for the dataset. Numeric variables (age, height, weight) are obtained through suspect report, while other categorical variables may reflect subjective perceptions of police or suspect report.*	
Variables of Interest for Exploratory Analysis	
FRISKED_FLAG	(chr) indicates whether or not the suspect was frisked (N = No, Y = Yes)
SEARCH_FLAG	(chr) indicates whether or not the suspect was searched (N = No, Y = Yes)
ASK_FOR_CONSENT_FLG	(chr) indicates whether the police asked for subject consent for the frisk/search behaviors after stop (N = No, Y = Yes)

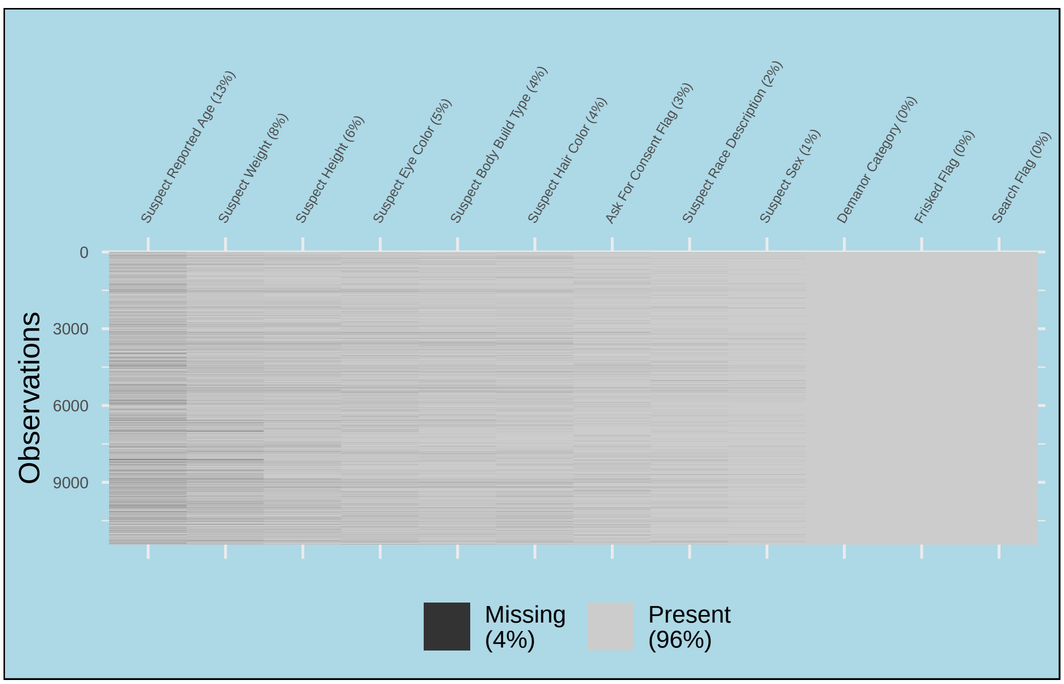
Data Cleaning & New Variable Creation

Upon reviewing the dataset, we identified a total of 1589 unique demeanor descriptions. To streamline our analysis, we focused on demeanor descriptions that appeared 10 or more times, aiming to capture meaningful trends and patterns. We then categorized these 69 demeanor descriptions into 5 broader categories based on their semantic similarities. While we recognize that the categorizations can be rather arbitrary, the groupings based on similarities in emotional or behavioral context allows for a more concise representation suitable for further analysis. Here are the definitions for the 5 categories (for specific descriptions included in each categories, see appendix):

Demeanor Categories	
Calm/Neutral	includes descriptions indicating a relaxed, cooperative, or normal state of mind
Nervous/Anxious	includes descriptions reflecting anxiety, nervousness, or apprehension
Confused/Disoriented	includes descriptions suggesting confusion, surprise, or disorientation
Indifferent	includes descriptions suggesting indifference and withdrawal
Angry/Confrontational	includes descriptions indicating anger, aggression, or hostility

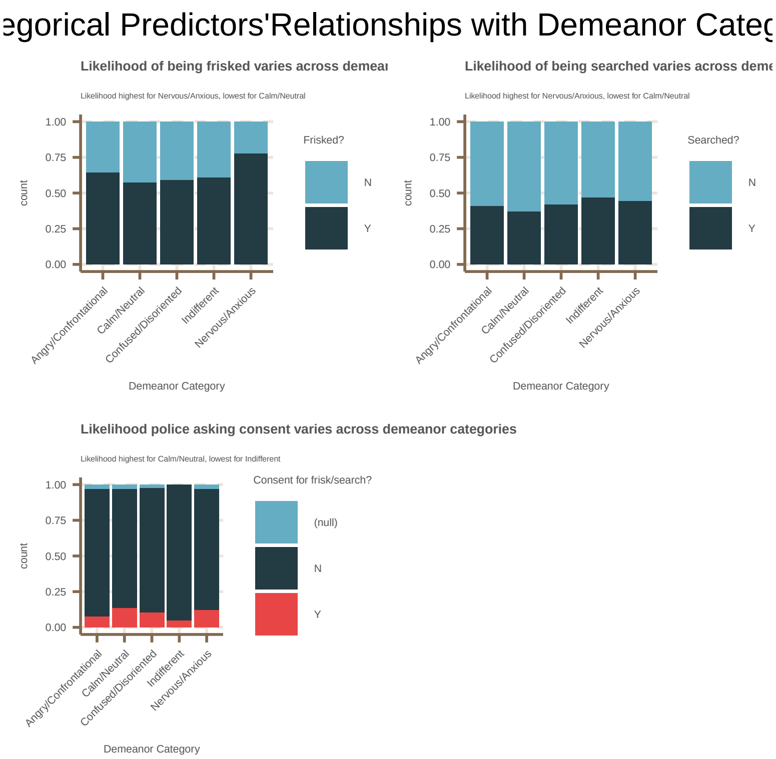
Note: *The following descriptions do not fit well into the above categories: Defensive (21), Laughing (16), Crying (14), Excited (14), Talkative (22) Given that the low relative frequencies (indicated in the brackets), we decided to remove them along with NAs (NA, N/A).*

The following graph visualizes the missingess in our newly generated data set (stop_and_frisk_cleaned) consisting of our created variable "demeanor category" and other predictors variables/variables for exploratory analysis of interest:



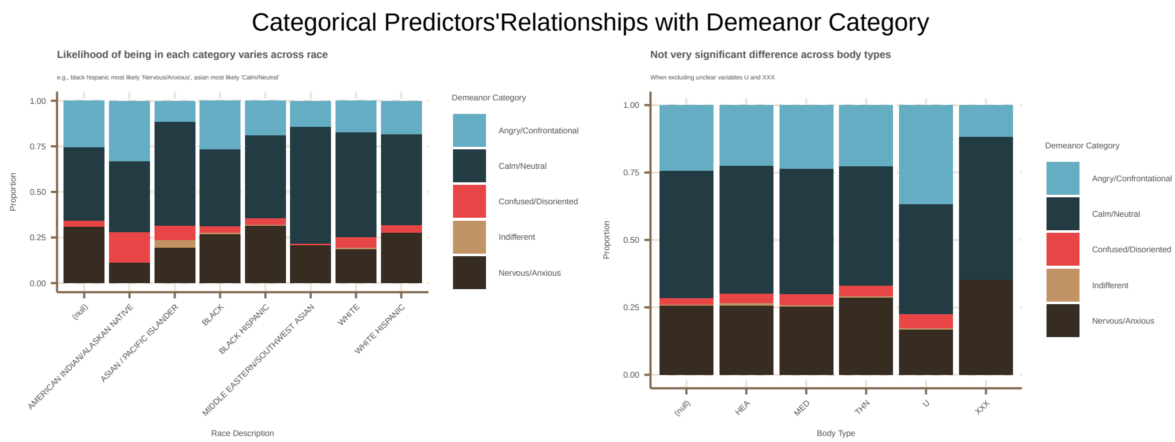
Note: Among our variables of interest, no variable contains a significant amount of missing values that would require specialized handling. Therefore, for subsequent exploratory analysis, we will use listwise deletion to remove observations with missing values (NA).

Exploratory Data Analysis

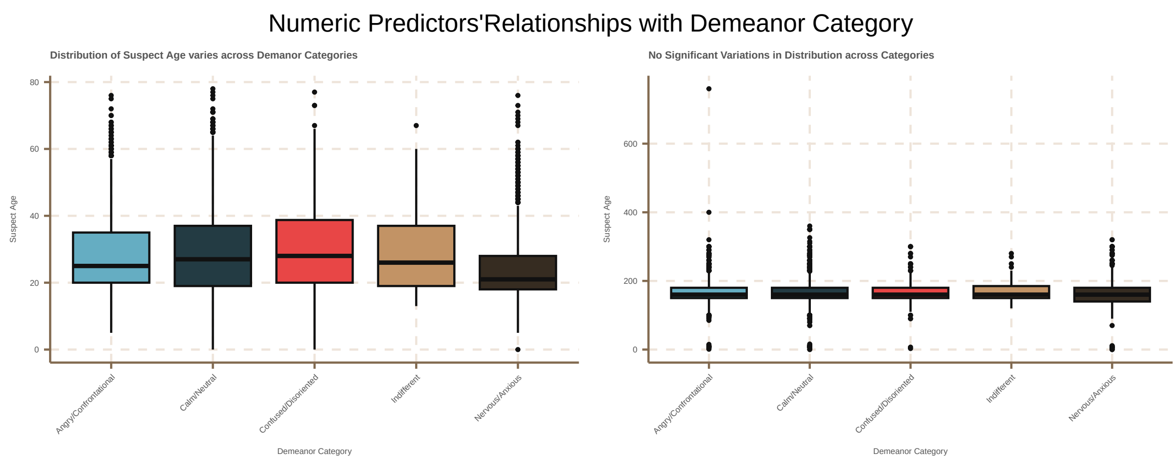


Summary: The stacked bar graphs accounts for the percentage of suspect in each demeanor category being frisked, searched, or asked for consent. The results do support our hypothesis, where membership to a particular membership demeanor category seems to affect "police behavior." For example, those in the calm/neutral group are least likely to be frisked and most likely to be asked for consent. This exploration supports the significance of our exploration.

Variable Selection



Based on the exploration above, we exclude variable "SUSPECT_BODY_BUILD_TYPE as a potential predictor variable.



Based on the exploration above, we exclude variables "SUSPECT_WEIGHT" and "SUSPECT_HEIGHT" as potential predictor variables.

Methodology:

Initial Model: mtest <- demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_DESCRIPTION + SUSPECT_EYE_COLOR + SUSPECT_HAIR_COLOR

Why Multinomial Regression Model?

The primary variable of interest, "demeanor category", consists of groups of categorical descriptors that are assigned by police officers. These descriptors are neither ordinal (they simply represent clusters of adjectives with similar characteristics) nor binary (e.g., calm vs. not calm) but rather fall into multiple distinct categories.

Multinomial regression allows us to assess how demographic/physical appearance predictors influence the likelihood of being assigned different demeanor categories compared to a reference category (set as calm/neutral). We can interpret the model coefficients to understand the direction and magnitude of these relationships.

Assessing multicolinearity & interactions:

We suspect multicolinearity between race, eye color, and hair color. Eye color and hair color also contain 9 and 14 categories respectively, largely complicating the coefficient displays of our model. To mitigate this complexity, we opted to create nested models to assess their impact on model fit and explanatory power, potentially removing unnecessary predictors.

mtest1 <- demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_DESCRIPTION

mtest2 <- demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_DESCRIPTION + SUSPECT_EYE_COLOR

mtest3 <- demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_DESCRIPTION + SUSPECT_EYE_COLOR + SUSPECT_HAIR_COLOR

mtest4 <- demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_DESCRIPTION + SUSPECT_HAIR_COLOR

Model 1 vs. Model 2 (adding suspect eye color to race):

- **P-value (Pr(Chi))**: 0.14621930 - adding `SUSPECT_EYE_COLOR` to the model (from Model 1 to Model 2) does not result in a statistically significant improvement in model fit (at the conventional significance level of 0.05).

Model 2 vs. Model 3 (adding suspect hair color to eye color & race):

- **P-value (Pr(Chi))**: 0.02794697 - adding `SUSPECT_HAIR_COLOR` to the model (from Model 2 to Model 3) results in a statistically significant improvement in model fit (at the conventional significance level of 0.05).

Model 1 vs. Model 4 (adding suspect hair color to race):

- **P-value (Pr(Chi))**: 0.03577523 - adding `SUSPECT_HAIR_COLOR` to the model results in a statistically significant improvement in model fit (at the conventional significance level of 0.05).

We decided to delete "hair color" from the predictor variables.

Similarly, we conducted nested tests to assess the interactions between remaining variables (age, sex, race, hair color). 2 interactions emerged as statistically significant:

- SUSPECT_REPORTED_AGE * SUSPECT_RACE_DESCRIPTION (p-value: 0.002785438)
- SUSPECT_SEX * SUSPECT_RACE_DESCRIPTION (p-value: 0.0153514)

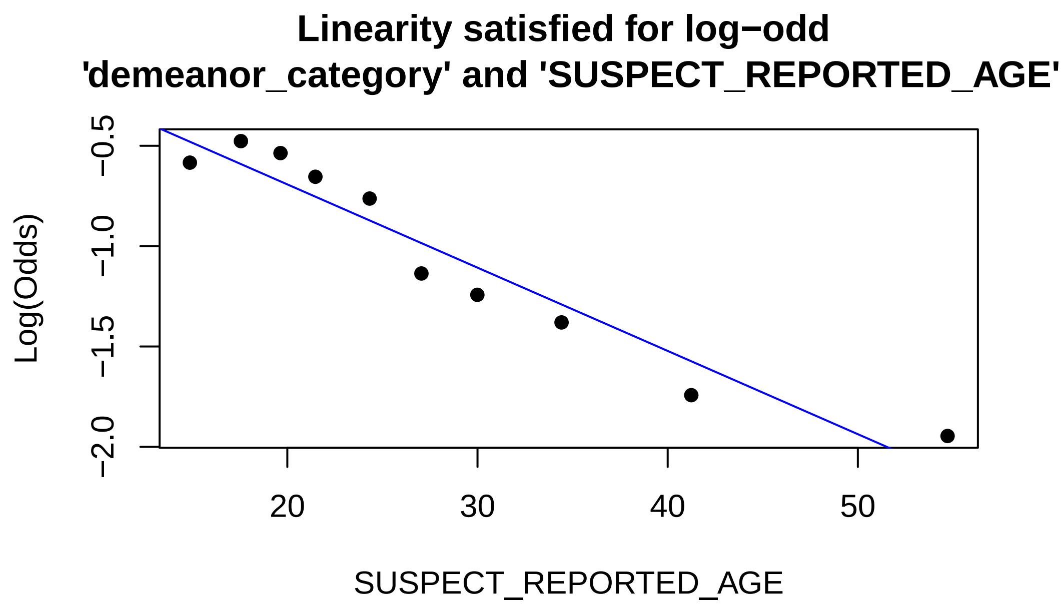
We also conducted a nested - test incorporating both interactions. When compared with original model, the p-value is 0.0003565872 - supporting the significance of adding these interaction terms.

Assumption Diagnostics

Irrelevance of Independent Alternatives Assumption?

The IIA assumption implies that the relative preference or probability of choosing one category over another is independent of the presence or characteristics of other categories in the choice set. For example the probability of police assigning an individual of given demographic/physical appearance to "Calm/Neutral" over "Nervous/Anxious" is independent from the presence/absence of the category "Indifferent". The assumption is reasonably satisfied in our model given that our categories derived from self-generated adjectives by the police officers. The original dataset was not generated in context of the 5 categories, so we could reasonably infer that the presence/absence of other categories will not affect the likelihood of police officers assigning suspects with certain characteristics to a particular category.

Linear relationship between continuous variables and the logit transformation of the outcome variable?



Result

Based on our predictor screening process, the predictors - Suspect Reported Age, Suspect Race Description, Suspect Sex Suspect Hair Color, Interaction between Age and Race, and Interaction between Sex and Race - are fitted in our final multinomial model (mfinal) to produce the following coefficients:

Demeanor Category	Nervous/Anxious		Angry/Confrontational		Confused/Disoriented		Indifferent	
Intercept	-1.358		-2.575		-4.397		-5.155	
Suspect Reported Age	0.002		0.000		0.072		-3.452	
Suspect Race Description								
Asian/Pacific Islander	2.620		-6.047		2.705		0.941	
Black	1.847		2.346	0.046	1.378		0.979	
Black/Hispanic	2.147		1.502		-5.700		-0.970	
Eastern/Southwest Asian	-6.588		-4.859		-0.116		-0.730	
White	1.584		2.389		1.106		-7.307	
White Hispanic	1.737		1.389		0.996		2.033	
Suspect Sex (Male)	0.181		1.668		0.095		1.577	
Suspect Hair Color								
Black	-0.001		0.149		0.815		-0.443	
Blonde	-0.289		-0.192		-0.344		1.443	
Brown	-0.130		0.195		0.901		-0.123	0
Grey	-0.415		0.097		0.415		-9.448	0
Orange	-0.519		0.325		1.947		-2.804	0
Purple	0.483		-6.532		-3.052		-0.577	0
Pink	-7.510		-0.298		1.899		-2.462	0
Red	-0.077		0.009		0.719		1.507	0
Sandy	-7.507		0.275		-4.916		-3.238	
White	-1.055		-0.009		1.533		-5.534	0
Unknown/Unspecified	0.353		0.522		1.002		0.232	
Error/Missing Data	0.507	0	-0.373		0.299		-5.026	0
Reported Age : Race								
Asian/Pacific Islander	-0.060		0.009		-0.066		3.549	
Black	-0.047		-0.002		-0.065		3.429	
Black Hispanic	-0.066		-0.014		-0.074		3.317	
Eastern/Southwest Asian	0.017		0.007		-0.128		-17.813	
White	-0.026		-0.013		-0.065		3.488	
White Hispanic	-0.039		-0.006		-0.056		3.416	
Sex : Race								
Asian/Pacific Islander	-0.861		4.763		-1.539		-3.999	
Black	0.066		-2.114		-0.546		-0.685	
Black Hispanic	0.380		-1.232		6.594		4.162	
Eastern/Southwest Asian	6.108		3.489		0.947		-0.724	
White	-0.708		-2.438		-0.060		5.259	
White Hispanic	-0.099		-1.504		-0.533		-2.335	

Note: The spreadsheet displays the coefficients generated by "mfinal" rounded to 3 decimal places. Coefficients highlighted in light blue has a p-value lower than the 0.05 significance threshold. P-values close to the threshold/displayed as 0.0 (likely due to scarcity of relevant observations) are indicated on the right.

Since the primary objective of our exploration is exploring existing trends rather than making predictions, we will not be assessing model predictive power through CV tests.

Key Interpretations:

As the spreadsheet suggest, our model appears to be most impactful in exploring the relationship between officer-assigned demeanor adjectives and race, particularly modulated by sex. The majority of coefficients associated with this relationship are statistically significant at the 0.05 significance threshold.

Notably, most coefficients indicating changes in log-odds of being characterized as "indifferent" relative to "calm/neutral" based on certain demographic/physical characteristics are statistically significant, with p-values reported as 0.0. However, it's important to acknowledge that we have only 64 observations in the "indifferent" category compared to over 3,000 for "nervous/anxious," which may lead to standard errors of 0 in z-score calculations to begin with.

The baseline for the "race" variable is American Indian/Alaskan Native while baseline for "sex" is female.

A few significant trends we see:

- Nervous/Anxious: Holding all other predictors constant and given that the suspect is female, the model predicts that being in all race except for Eastern/Southwest Asian correspond to a multiplicative increase in odds of being described as Nervous/Anxious rather than Calm/Neutral, compared to American Indian/Alaskan Native. Being a female Eastern/Southwest Asian (the most extreme coefficient in this category), is associated with approximately 0.002478 times the odds of being described as Nervous/Anxious rather than Calm/Neutral, compared to American Indian/Alaskan Native. When accounting for suspect sex, the overall trend remains consistent in directionality (multiplicative increase/decrease in odds). However, noticeably, being a male Eastern/Southwest Asian is associated with approximately 0.6187 times the odds of being described as Nervous/Anxious rather than Calm/Neutral, compared to American Indian/Alaskan Native - a drastic change from the previous odds ratio.

- Angry/Confrontational: Holding all other predictors constant and given that the suspect is female, the model predicts that being in all race except for Eastern/Southwest Asian or Asian/Pacific Islander correspond to a multiplicative increase in odds of being described as Angry/Confrontational rather than Calm/Neutral, compared to American Indian/Alaskan Native. When examining the interaction coefficients accounting suspect sex, previously positive race coefficients show negative interactions, while previously negative coefficients show positive interactions. This suggests that being male decreases the multiplicative difference in odds of being described as "angry/confrontational" rather than "calm/neutral" compared to American Indian/Alaskan Native across races. The overall trend remains consistent, with the most significant change observed in Eastern/Southwest Asian and Asian/Pacific Islander individuals.

- Confused/Disoriented: Overall trend is relatively similar to previous two categories. Black Hispanic is a noteworthy group. Holding all other predictors constant and given that the suspect is female, the model predicts that being in all race except for Eastern/Southwest Asian or Black Hispanic correspond to a multiplicative increase in odds of being described as Confused/Disoriented rather than Calm/Neutral, compared to American Indian/Alaskan Native.

The model predicts being a female in Black Hispanic group is associated with approximately 0.003374 times the odds of being described as Confused/Disoriented rather than Calm/Neutral, compared to American Indian/Alaskan Native. However, the direction is reversed when the suspect is male (approximately 2.4456 times the odds).

- Indifferent: As previously discussed, due to small sample size, the results in this category are relatively ambiguous. It is difficult to observe overall trends. Interestingly, female White and White Hispanic mark the two extreme ends (most positive and most negative). The difference is mitigated when accounting for sex.

These observations highlight significant trends in the relationship between officer-assigned demeanor adjectives, race, and sex, providing insights into how certain demographic/physical characteristics may influence descriptions of suspects' demeanor during police encounters.

Discussion

Our exploration underscores the significant influence of suspect demographic characteristics, particularly race and sex, on how police officers assign demeanor adjectives during stops. We observed statistically significant coefficients indicating differential odds of being described with certain demeanor traits based on race and sex. For instance, being a Black Hispanic female versus an Asian/Pacific Islander female has significant differences in the likelihood of being characterized as confused/disoriented compared with calm/neutral. Similarly, being a female Eastern/Southwest Asian versus a male of the same race showed drastic differences in the likelihood of being characterized as "nervous/anxious" or "angry/confrontational."

While concrete inferences should not be drawn, underlying stereotypes of different demographic groups could potentially contribute to these differential demeanor descriptions. For example, Eastern/Southwest Asian females are commonly portrayed in media as more levelheaded, compliant, and cooperative, whereas males are often perceived as more confrontational and aggressive. These stereotypes have the potential to subtly impact police perceptions and behaviors, even when the officer is not consciously aware at the moment.

Overall, our analysis provides insights into how implicit biases may manifest in police interactions, influencing the perceptions and subsequent actions of officers during stop-and-frisk encounters. By examining correlations between physical/demographic characteristics and demeanor descriptions, we begin to unpack the underlying factors contributing to disparities observed in policing practices.

Nevertheless, our exploration has several key limitations.

First, while the original dataset includes over 16,000 observations, the cleaned dataset we worked with consists of 10,497 observations. Some observations (approximately 4% according to our missingness visualization) were also removed when necessary in subsequent modeling due to their incomplete nature. The loss of a significant proportion of the data could undermine the representativeness of our conclusions.

"Demeanor Categories" were used to facilitate multinomial modeling and ease interpretation of the data. While we attempted to create adjective clusters based on semantic similarities, the categories remain rather arbitrary. Some adjectives could belong to multiple categories, and the categories exclude a large proportion of potentially useful data that failed to fit in. Our initial variable screening was based on simple data visualization and may have accidentally excluded significant predictor variables.

Our incomplete understanding of the dataset also posed challenges. We are uncertain of the exact data collection process—for example, whether variables such as "race description" are obtained through suspect reports or police perception. The collection method could significantly influence our interpretations regarding whether "implicit bias" is at work. Additionally, we made the assumption that demeanor adjectives to some extent are indicative of "police perception of the suspect." Lastly, while our dataset reasonably satisfied the assumption of independence (each unique police stop), we lack knowledge regarding whether observations are dependent on each other, such as being produced by the same officer.

While our exploration aimed to assess both demographic and physical characteristics of subjects, most "physical" characteristic predictors were eliminated during the process due to their complexity or lack of statistical significance. Future analysis could expand beyond sex and race to include more nuanced physical characteristics such as hair color or dressing style. The original dataset contained a variable called "other description" that captures physical appearances such as clothing color and style—this variable was excluded due to its complexity. Future exploration could also delve deeper into other police behavior-related variables (questioning, duration of stop, etc.).

Appendix

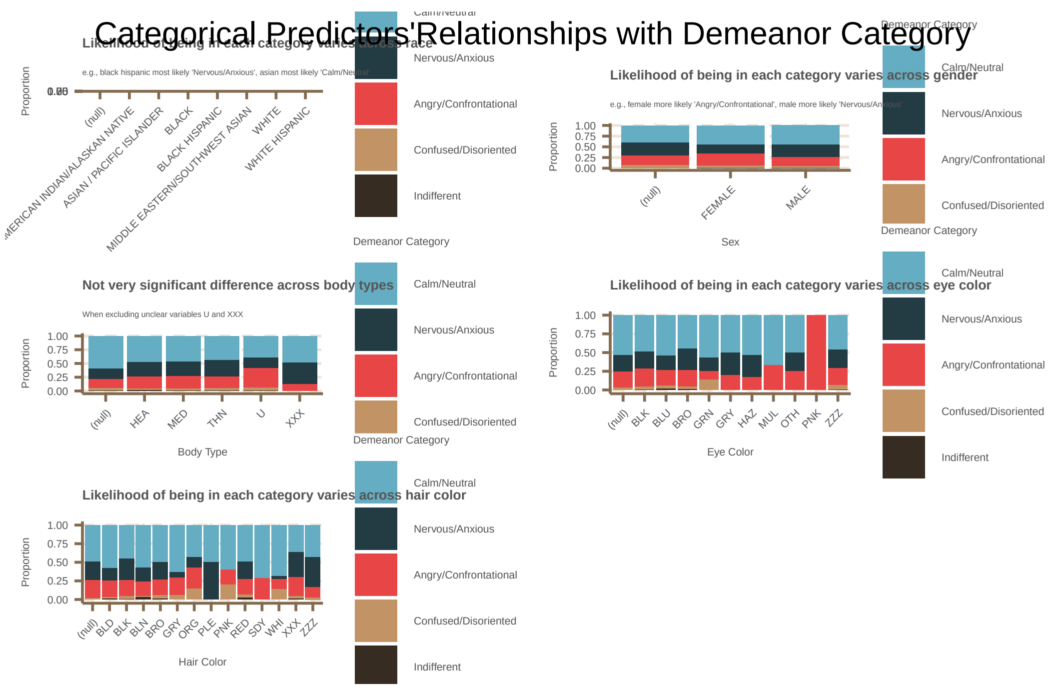
1.Descriptions allocated to each demeanor category

2.Categorical Predictors Full Exploratory Analysis

```
# Filter out rows with missing values in necessary columns
cleaned_data <- stop_and_frisk_cleaned %>%
  filter(!is.na(SUSPECT_RACE_DESCRIPTION),
         !is.na(SUSPECT_SEX),
         !is.na(SUSPECT_BODY_BUILD_TYPE),
         !is.na(SUSPECT_EYE_COLOR),
         !is.na(SUSPECT_HAIR_COLOR))

# Create individual ggplot visualizations
plot1 <- cleaned_data %>%
  ggplot(aes(x = SUSPECT_RACE_DESCRIPTION, fill = demeanor_category)) +
  geom_bar(position = "fill") +
```

```
  labs(x = "Race Description", y = "Proportion", title = "Likelihood of being in each cate
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        axis.text.y = element_text(size = 4),
        axis.title = element_text(size = 4),
        plot.title = element_text(size = 5),
        plot.subtitle = element_text(size = 3),
        legend.title = element_text(size = 4),
        legend.text = element_text(size = 4))

plot2 <- cleaned_data %>%
  ggplot(aes(x = SUSPECT_SEX, fill = demeanor_category)) +
  geom_bar(position = "fill") +
  labs(x = "Sex", y = "Proportion", title = "Likelihood of being in each category varies a
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        axis.text.y = element_text(size = 4),
        axis.title = element_text(size = 4),
        plot.title = element_text(size = 5),
        plot.subtitle = element_text(size = 3),
        legend.title = element_text(size = 4),
        legend.text = element_text(size = 4))

plot3 <- cleaned_data %>%
  ggplot(aes(x = SUSPECT_BODY_BUILD_TYPE, fill = demeanor_category)) +
  geom_bar(position = "fill") +
  labs(x = "Body Type", y = "Proportion", title = "Not very significant difference across
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        axis.text.y = element_text(size = 4),
        axis.title = element_text(size = 4),
        plot.title = element_text(size = 5),
        plot.subtitle = element_text(size = 3),
        legend.title = element_text(size = 4),
        legend.text = element_text(size = 4))

plot4 <- cleaned_data %>%
  ggplot(aes(x = SUSPECT_EYE_COLOR, fill = demeanor_category)) +
  geom_bar(position = "fill") +
  labs(x = "Eye Color", y = "Proportion", title = "Likelihood of being in each category va
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        axis.text.y = element_text(size = 4),
        axis.title = element_text(size = 4),
        plot.title = element_text(size = 5),
```

```
        legend.title = element_text(size = 4),
        legend.text = element_text(size = 4))

plot5 <- cleaned_data %>%
  ggplot(aes(x = SUSPECT_HAIR_COLOR, fill = demeanor_category)) +
  geom_bar(position = "fill") +
  labs(x = "Hair Color", y = "Proportion", title = "Likelihood of being in each category v
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 4),
        axis.text.y = element_text(size = 4),
        axis.title = element_text(size = 4),
        plot.title = element_text(size = 5),
        legend.title = element_text(size = 4),
        legend.text = element_text(size = 4))

# Arrange plots in a grid and adjust size
grid.arrange(plot1, plot2, plot3, plot4, plot5, ncol = 2, top = "Categorical Predictors'Re
```



3.Numerical Predictor Full Exploratory Analysis