

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. It should be noted that 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): 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.

1. Calm/Neutral Demeanor: This category includes descriptions indicating a relaxed, cooperative, or normal state of mind.

- CALM - NORMAL - APPARENTLY NORMAL - RELAXED - QUIET - UNDERSTANDING - CALM AND COOPERATIVE - CALM AND COMPLIANT - CALM AND UNDERSTANDING - CALM COOPERATIVE - CALMED - NEUTRAL - CALM COMPLIANT - CALM UNDERSTANDING - APP NORMAL - COMPLIANT - APPARENT NORMAL

2. Nervous/Anxious Demeanor: Descriptions reflecting anxiety, nervousness, or apprehension.

- NERVOUS - ANXIOUS - VERY NERVOUS - EXTREMELY NERVOUS - PHYSICALLY NERVOUS - NERVOUS SCARED - NERVOUS OUT OF BREATH - AGGITATED - SCARED - SUSPICIOUS - APPREHENSIVE - WORRIED - NERVOUSE

3. Angry/Confrontational Demeanor: This category comprises descriptions indicating anger, aggression, or hostility.

- UPSET - ANNOYED - ANGRY - AGITATED - AGGRESSIVE - COMBATIVE - IRATE
- IRRITATED - AGGRAVATED - HOSTILE - MAD - AGGRESSIVE/NERVOUS - UN-
COOPERATIVE - IRRATE - AGGRESIVE - ARGUMENTATIVE - DEFENSIVE - NON
COMPLIANT

4. Confused/Disoriented Demeanor: Descriptions suggesting confusion, surprise, or dis-orientation.

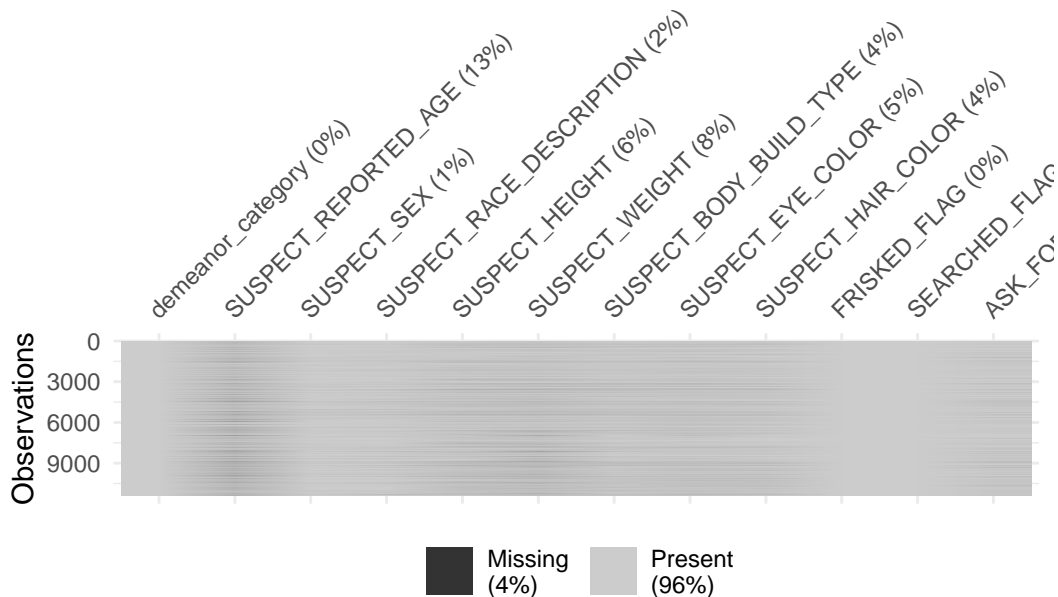
- CONFUSED - SURPRISED - SHOCKED - INTOXICATED - INTOX - ERRATIC - OUT
OF BREATH

5. Indifferent Demeanor: Descriptions suggesting withdrawal

- INDIFFERENT - EVASIVE - TIRED

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).

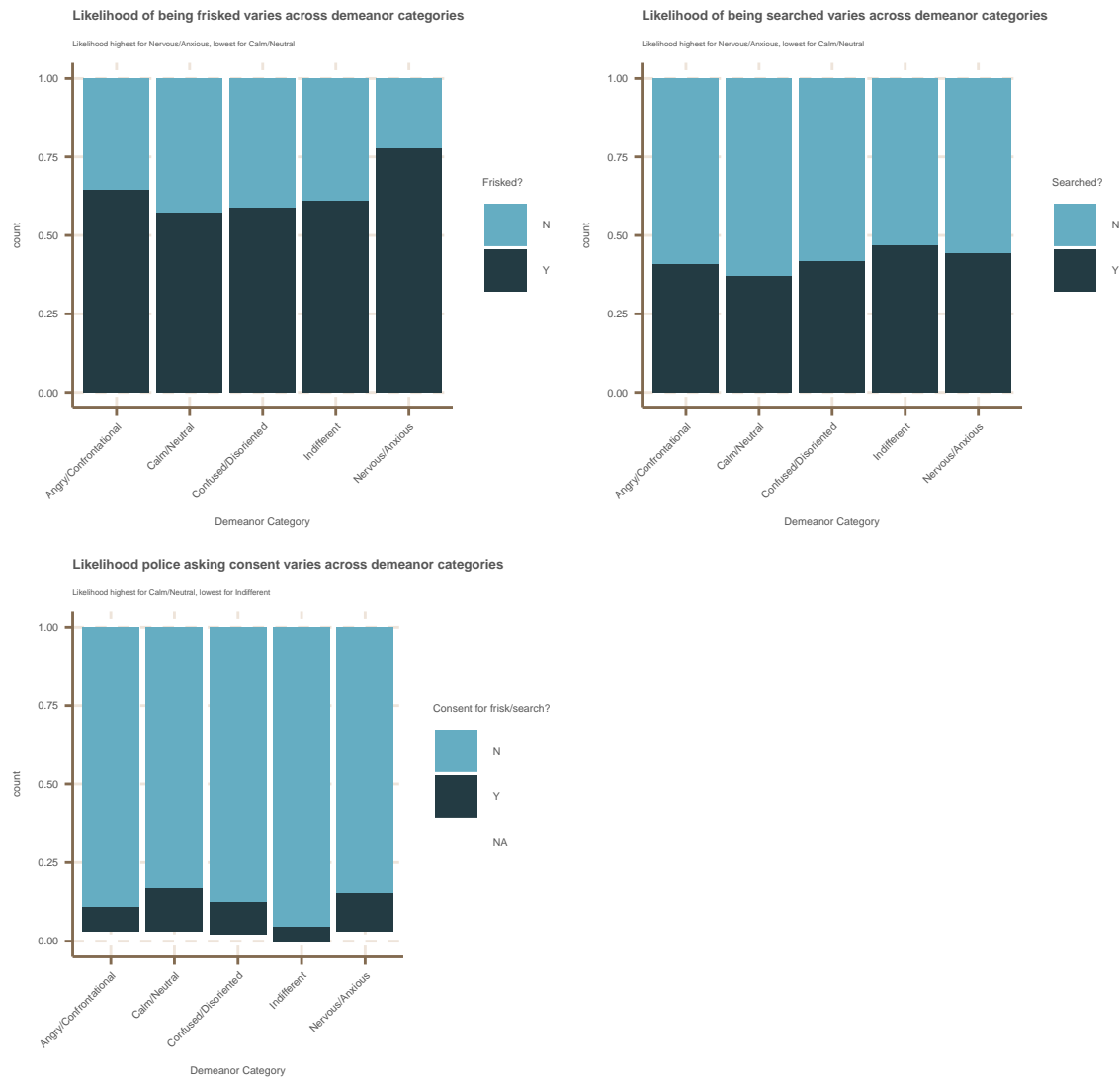
The following graph visualizes the missingness in our newly generated data set (stop_and_frisk_cleaned) consisting of our created variable “demeanor category” and other predictors variables/variables for exploratory analysis of interest:



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

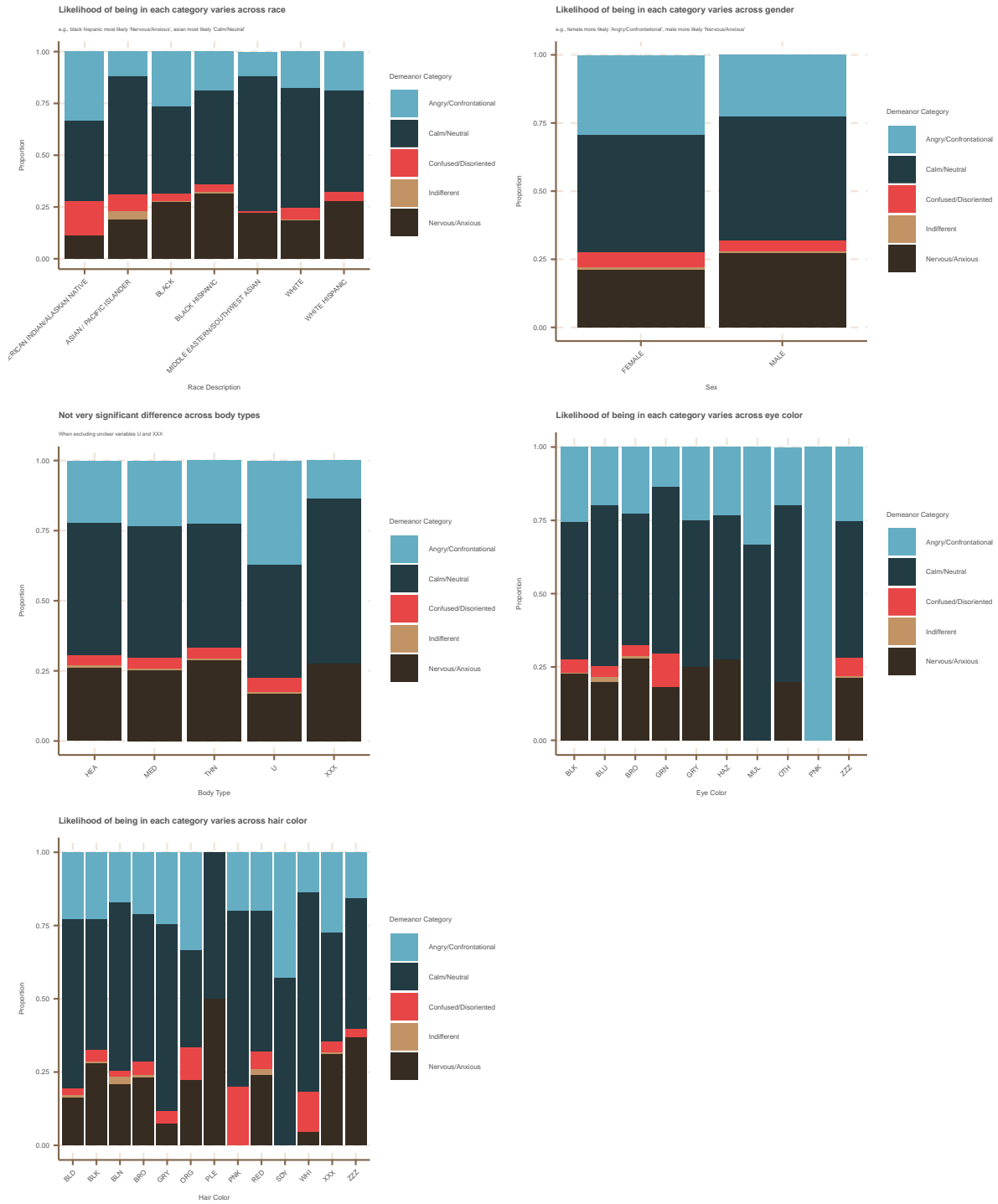
Categorical Predictors' Relationships with Demeanor Category



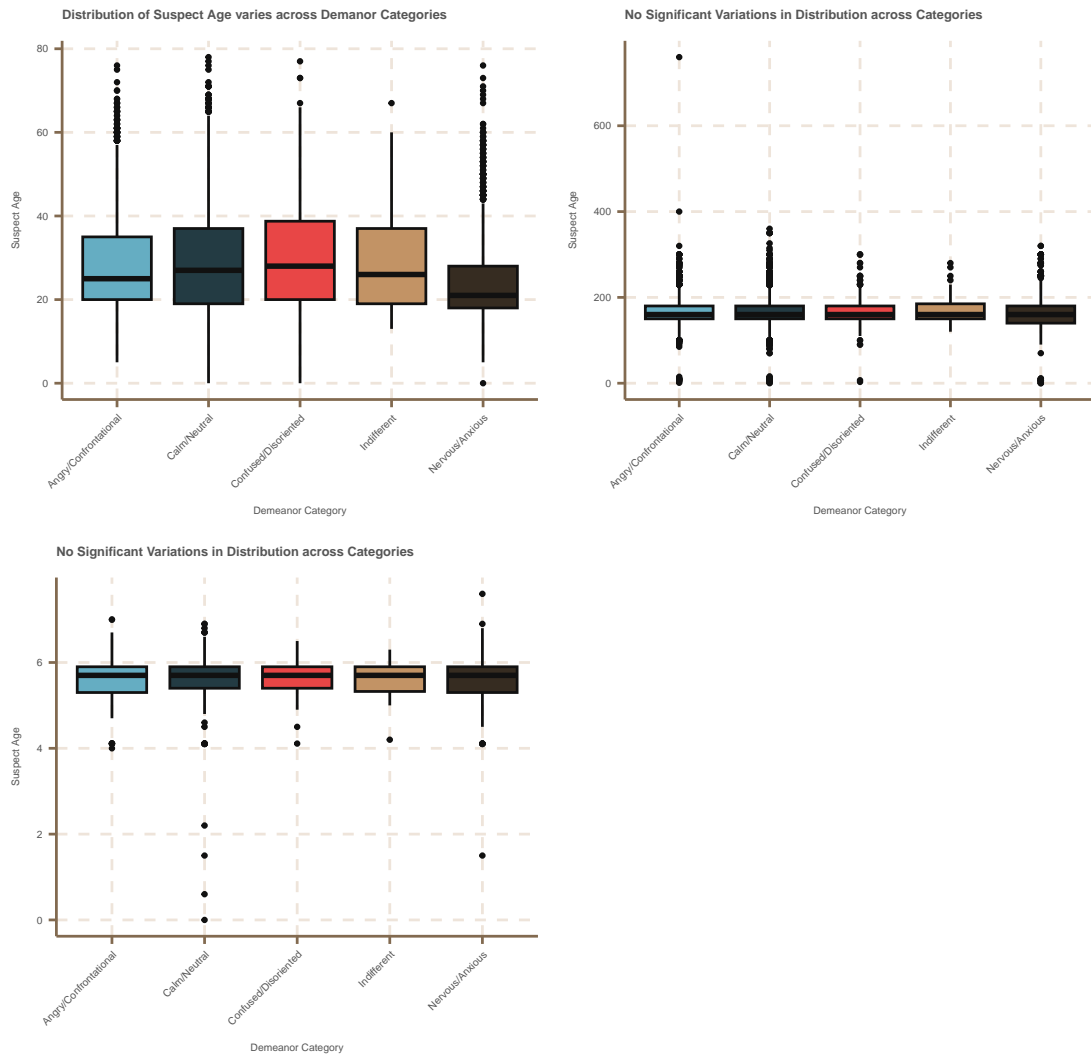
Summary:

Variable Selection

Categorical Predictors' Relationships with Demeanor Category



Numeric Predictors' Relationships with Demeanor Category



Methodology:

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.

```
library(nnet)
# Set calm/neutral as baseline
stop_and_frisk_cleaned$demeanor_category <- factor(stop_and_frisk_cleaned$demeanor_category, levels=c("calm/neutral", "agitated", "hostile"))
mtest <- multinom(demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_DE

# weights:  155 (120 variable)
initial  value 15370.132064
iter   10 value 12354.940189
iter   20 value 11564.898190
iter   30 value 11306.709240
iter   40 value 11242.182943
iter   50 value 11201.239959
iter   60 value 11190.982313
iter   70 value 11190.551769
iter   80 value 11190.320487
iter   90 value 11190.206823
final   value 11190.203908
converged
```

Assessing multicollinearity & interactions:

Excluding incidents of hair dye and contact lenses, basing off common sense, we suspect a multicollinearity between race and eye color or hair color. Eye color and hair color also contain 9 and 14 categories respectively, largely complicating the coefficient displays of our model.

```
stop_and_frisk_cleaned$demeanor_category <- factor(stop_and_frisk_cleaned$demeanor_category, levels=c("calm/neutral", "agitated", "hostile"))
stop_and_frisk_cleaned <- na.omit(stop_and_frisk_cleaned)

mtest1 <- multinom(demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_D

# weights:  50 (36 variable)
initial  value 14641.056689
iter   10 value 11441.484538
iter   20 value 10887.247851
iter   30 value 10744.997398
iter   40 value 10705.238769
iter   50 value 10705.065536
iter   60 value 10705.025915
```



```
iter 60 value 10705.025831
iter 60 value 10705.025831
final value 10705.025831
converged
```

```
mtest2 <- multinom(demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_D
```

```
# weights: 95 (72 variable)
initial value 14641.056689
iter 10 value 11399.953060
iter 20 value 10976.798446
iter 30 value 10745.458705
iter 40 value 10703.245151
iter 50 value 10685.675253
iter 60 value 10682.722193
iter 70 value 10682.584887
iter 80 value 10682.563648
final value 10682.563455
converged
```

```
mtest3 <- multinom(demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_D
```

```
# weights: 155 (120 variable)
initial value 14641.056689
iter 10 value 11440.374708
iter 20 value 10945.222495
iter 30 value 10748.546495
iter 40 value 10684.488511
iter 50 value 10657.258032
iter 60 value 10649.522304
iter 70 value 10648.733008
iter 80 value 10648.497591
iter 90 value 10648.355485
final value 10648.350799
converged
```

```
anova_result1 <- anova(mtest1, mtest2, mtest3)
```

```
mtest4 <- multinom(demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_D
```

```
# weights: 110 (84 variable)
initial value 14641.056689
iter 10 value 11412.309157
iter 20 value 10971.484114
iter 30 value 10733.951681
iter 40 value 10701.155682
iter 50 value 10672.995207
iter 60 value 10671.772974
iter 70 value 10671.523296
iter 80 value 10671.489499
final value 10671.489148
converged
```

```
anova_result2 <- anova(mtest1, mtest4)
```

Interpretation:

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

- **P-value (Pr(Chi)):** 0.14621930
 - The p-value (0.14621930) suggests that 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
 - The p-value (0.02794697) indicates that 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
 - The p-value (0.02794697) indicates that 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.

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”.

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

```
install.packages("devtools")
```

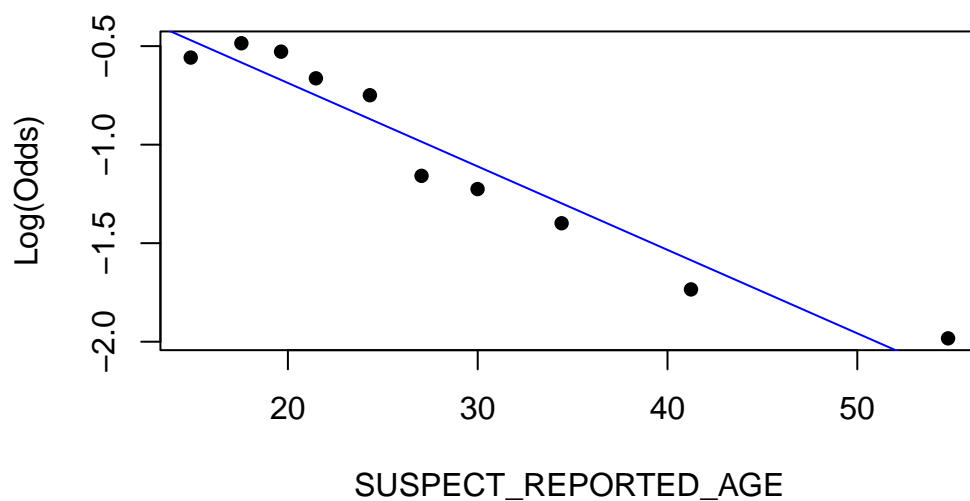
Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)

```
devtools::install_github("statmanrobin/Stat2Data")
```

Skipping install of 'Stat2Data' from a github remote, the SHA1 (3fe987c7) has not changed since last install. Use `force = TRUE` to force installation

```
library(Stat2Data)
emplogitplot1(demeanor_category ~ SUSPECT_REPORTED_AGE,
              data = stop_and_frisk_cleaned,
              ngroups = 10,
              main = "Linearity satisfied for log-odd demeanor_category and SUSPECT_REPO
              )
```

satisfied for log-odd demeanor_category and SUSPECT_RE



Result

```
mfinal <- multinom(demeanor_category ~ SUSPECT_REPORTED_AGE + SUSPECT_SEX + SUSPECT_RACE_D
```

```
# weights: 110 (84 variable)
initial value 14641.056689
iter 10 value 11412.309157
iter 20 value 10971.484114
iter 30 value 10733.951681
iter 40 value 10701.155682
iter 50 value 10672.995207
iter 60 value 10671.772974
iter 70 value 10671.523296
iter 80 value 10671.489499
final value 10671.489148
converged
```

```
summary(mfinal)
```

Call:

```
multinom(formula = demeanor_category ~ SUSPECT_REPORTED_AGE +
  SUSPECT_SEX + SUSPECT_RACE_DESCRIPTION + SUSPECT_HAIR_COLOR,
  data = stop_and_frisk_cleaned)
```

Coefficients:

	(Intercept)	SUSPECT_REPORTED_AGE	SUSPECT_SEXMALE
Nervous/Anxious	0.01695811	-0.042543582	0.1143035
Angry/Confrontational	-0.45182058	-0.002831759	-0.3300198
Confused/Disoriented	-1.83744294	0.008042736	-0.3245742
Indifferent	-9.10596437	-0.004146931	-0.1340090
	SUSPECT_RACE_DESCRIPTIONASIAN / PACIFIC ISLANDER		
Nervous/Anxious			0.1165216
Angry/Confrontational			-1.0479628
Confused/Disoriented			-0.9372544
Indifferent			7.1236466
	SUSPECT_RACE_DESCRIPTIONBLACK		
Nervous/Anxious		0.5354716	
Angry/Confrontational		0.1608729	
Confused/Disoriented		-1.3286137	
Indifferent		5.3758331	
	SUSPECT_RACE_DESCRIPTIONBLACK HISPANIC		
Nervous/Anxious		0.6391787	
Angry/Confrontational		-0.2016313	
Confused/Disoriented		-1.5938041	
Indifferent		5.6712591	
	SUSPECT_RACE_DESCRIPTIONMIDDLE EASTERN/SOUTHWEST ASIAN		
Nervous/Anxious			-0.2004389
Angry/Confrontational			-1.2276365
Confused/Disoriented			-2.8653023
Indifferent			-6.4644813
	SUSPECT_RACE_DESCRIPTIONWHITE		
Nervous/Anxious		0.1742579	
Angry/Confrontational		-0.4321404	
Confused/Disoriented		-1.1952678	
Indifferent		4.8371091	
	SUSPECT_RACE_DESCRIPTIONWHITE HISPANIC		
Nervous/Anxious		0.4660528	
Angry/Confrontational		-0.3369109	
Confused/Disoriented		-1.4243220	
Indifferent		4.5341510	
	SUSPECT_HAIR_COLORBLK SUSPECT_HAIR_COLORBLN		
Nervous/Anxious		0.003441623	-0.2356335

Angry/Confrontational	0.147871965	-0.1612911
Confused/Disoriented	0.817635265	-0.3565815
Indifferent	-0.407188405	1.1786388
	SUSPECT_HAIR_COLORBRO	SUSPECT_HAIR_COLORGRY
Nervous/Anxious	-0.11379642	-0.35523873
Angry/Confrontational	0.18945534	0.07698733
Confused/Disoriented	0.90119749	0.43950101
Indifferent	0.07125463	-7.21547027
	SUSPECT_HAIR_COLORORG	SUSPECT_HAIR_COLORPLE
Nervous/Anxious	-0.4571677	0.7987439
Angry/Confrontational	0.3392415	-17.9510093
Confused/Disoriented	1.9444618	-12.5993116
Indifferent	-11.4537715	-6.7841897
	SUSPECT_HAIR_COLORPNK	SUSPECT_HAIR_COLORRED
Nervous/Anxious	-17.5826083	-0.076175594
Angry/Confrontational	-0.3568012	-0.008934897
Confused/Disoriented	1.8965336	0.587060019
Indifferent	-12.4007672	0.890066416
	SUSPECT_HAIR_COLORSDY	SUSPECT_HAIR_COLORWHI
Nervous/Anxious	-18.4759158	-0.8560116
Angry/Confrontational	0.2491789	-0.1615604
Confused/Disoriented	-14.7638634	1.5919234
Indifferent	-12.9824825	-13.9364234
	SUSPECT_HAIR_COLORXXX	SUSPECT_HAIR_COLORZZZ
Nervous/Anxious	0.3701923	0.5167834
Angry/Confrontational	0.5243158	-0.3546765
Confused/Disoriented	1.0036870	0.3337953
Indifferent	0.3455830	-10.0921077

Std. Errors:

	(Intercept)	SUSPECT_REPORTED_AGE	SUSPECT_SEXMALE
Nervous/Anxious	0.8581425	0.002722397	0.1190838
Angry/Confrontational	0.7410952	0.002439957	0.1093863
Confused/Disoriented	0.9532281	0.004708373	0.2079957
Indifferent	0.8277054	0.011618762	0.5122228
	SUSPECT_RACE_DESCRIPTIONASIAN / PACIFIC ISLANDER		
Nervous/Anxious			0.8498220
Angry/Confrontational			0.7552202
Confused/Disoriented			0.8700772
Indifferent			0.3825180
	SUSPECT_RACE_DESCRIPTIONBLACK		
Nervous/Anxious		0.8262199	
Angry/Confrontational		0.7087544	

Confused/Disoriented	0.8215503	
Indifferent	0.2565377	
	SUSPECT_RACE_DESCRIPTIONBLACK	HISPANIC
Nervous/Anxious	0.8292082	
Angry/Confrontational	0.7138933	
Confused/Disoriented	0.8421220	
Indifferent	0.3606831	
	SUSPECT_RACE_DESCRIPTIONMIDDLE	EASTERN/SOUTHWEST ASIAN
Nervous/Anxious		8.723853e-01
Angry/Confrontational		8.028861e-01
Confused/Disoriented		1.299085e+00
Indifferent		2.044484e-06
	SUSPECT_RACE_DESCRIPTIONWHITE	
Nervous/Anxious	0.8341582	
Angry/Confrontational	0.7178333	
Confused/Disoriented	0.8415953	
Indifferent	0.5069771	
	SUSPECT_RACE_DESCRIPTIONWHITE	HISPANIC
Nervous/Anxious	0.8274433	
Angry/Confrontational	0.7108710	
Confused/Disoriented	0.8273135	
Indifferent	0.3956554	
	SUSPECT_HAIR_COLORBLK	SUSPECT_HAIR_COLORBLN
Nervous/Anxious	0.1637925	0.3146318
Angry/Confrontational	0.1495769	0.3148147
Confused/Disoriented	0.3775459	0.8197258
Indifferent	0.6473282	0.9123472
	SUSPECT_HAIR_COLORBRO	SUSPECT_HAIR_COLORGRY
Nervous/Anxious	0.1839946	0.347244895
Angry/Confrontational	0.1730203	0.245383924
Confused/Disoriented	0.4086463	0.535993104
Indifferent	0.7439750	0.001578182
	SUSPECT_HAIR_COLORORG	SUSPECT_HAIR_COLORPLE
Nervous/Anxious	1.179602e+00	1.518935e+00
Angry/Confrontational	9.352750e-01	1.336934e-08
Confused/Disoriented	1.221828e+00	3.389185e-07
Indifferent	1.101218e-06	5.099504e-05
	SUSPECT_HAIR_COLORPNK	SUSPECT_HAIR_COLORRED
Nervous/Anxious	2.255176e-08	0.4143470
Angry/Confrontational	1.172001e+00	0.4174201
Confused/Disoriented	1.227510e+00	0.8397188
Indifferent	3.816108e-07	1.2567769
	SUSPECT_HAIR_COLORSDY	SUSPECT_HAIR_COLORWHI

Nervous/Anxious	3.977977e-09	1.079802e+00
Angry/Confrontational	8.930904e-01	6.659059e-01
Confused/Disoriented	5.257464e-08	7.517723e-01
Indifferent	2.351685e-07	1.896014e-07
	SUSPECT_HAIR_COLORXXX	SUSPECT_HAIR_COLORZZZ
Nervous/Anxious	0.1832616	4.188355e-01
Angry/Confrontational	0.1727214	5.352358e-01
Confused/Disoriented	0.4186227	1.096198e+00
Indifferent	0.7313541	2.232503e-05

Residual Deviance: 21342.98
AIC: 21510.98

```
#Statistical Significance?

coefficients_final <- coef(mfinal)
standard_errors_final <- sqrt(diag(vcov(mfinal)))

z <- coefficients_final/ standard_errors_final
p <- (1 - pnorm(abs(z), 0, 1)) * 2

print(p)
```

	(Intercept)	SUSPECT_REPORTED_AGE	SUSPECT_SEXMALE
Nervous/Anxious	0.9842337	0.9589336	0.89012933
Angry/Confrontational	0.0000000	0.9972752	0.04391886
Confused/Disoriented	0.0000000	0.9926442	0.30225961
Indifferent	0.0000000	0.9960334	0.46641066
	SUSPECT_RACE_DESCRIPTIONASIAN / PACIFIC ISLANDER		
Nervous/Anxious			0.7372024
Angry/Confrontational			0.3743236
Confused/Disoriented			0.5372036
Indifferent			0.0000000
	SUSPECT_RACE_DESCRIPTIONBLACK		
Nervous/Anxious		0.1962441	
Angry/Confrontational		0.0000000	
Confused/Disoriented		0.2185385	
Indifferent		0.0000000	
	SUSPECT_RACE_DESCRIPTIONBLACK HISPANIC		
Nervous/Anxious		0.1269887	
Angry/Confrontational		0.7855666	

Confused/Disoriented	0.0000000	
Indifferent	0.0000000	
	SUSPECT_RACE_DESCRIPTION	MIDDLE EASTERN/SOUTHWEST ASIAN
Nervous/Anxious		7.906978e-01
Angry/Confrontational		8.325501e-02
Confused/Disoriented		5.979249e-05
Indifferent		8.881784e-16
	SUSPECT_RACE_DESCRIPTION	WHITE
Nervous/Anxious	8.081950e-01	
Angry/Confrontational	5.432520e-01	
Confused/Disoriented	1.332268e-15	
Indifferent	0.000000e+00	
	SUSPECT_RACE_DESCRIPTION	WHITE HISPANIC
Nervous/Anxious	0.007067836	
Angry/Confrontational	0.169753964	
Confused/Disoriented	0.127785931	
Indifferent	0.000000000	
	SUSPECT_HAIR_COLOR	BLK SUSPECT_HAIR_COLORBLN
Nervous/Anxious	0.9976570	0.1724920
Angry/Confrontational	0.7231499	0.7631508
Confused/Disoriented	0.3599227	0.7083464
Indifferent	0.5408816	0.0000000
	SUSPECT_HAIR_COLOR	BRO SUSPECT_HAIR_COLORGRY
Nervous/Anxious	0.5843036	0.7845050
Angry/Confrontational	0.8276274	0.9271129
Confused/Disoriented	0.2726644	0.5952533
Indifferent	0.9325689	0.0000000
	SUSPECT_HAIR_COLOR	ORG SUSPECT_HAIR_COLORPLE
Nervous/Anxious	0.5770437167	0
Angry/Confrontational	0.4064486629	0
Confused/Disoriented	0.0002858754	0
Indifferent	0.0000000000	0
	SUSPECT_HAIR_COLOR	PNK SUSPECT_HAIR_COLORRED
Nervous/Anxious	0.00000000	5.517498e-11
Angry/Confrontational	0.39403562	9.860829e-01
Confused/Disoriented	0.08361223	1.248514e-01
Indifferent	0.00000000	5.213601e-04
	SUSPECT_HAIR_COLOR	SDY SUSPECT_HAIR_COLORWHI
Nervous/Anxious	0	0.18604281
Angry/Confrontational	0	0.85944389
Confused/Disoriented	0	0.03237465
Indifferent	0	0.00000000
	SUSPECT_HAIR_COLOR	XXX SUSPECT_HAIR_COLORZZZ

Nervous/Anxious	0.000000	0.0000000
Angry/Confrontational	0.000000	0.0000000
Confused/Disoriented	0.000000	0.6480972
Indifferent	0.783335	0.0000000

Key Interpretations:

Overall trend in model:

Criteria - Salient Slope + Statistical Significance:

Note: Primary objective is prediction, so we will not be assessing model predictive power through CV tests.

Discussion

Overall conclusions from analysis are clearly described, and the model results are put into the larger context of the subject matter and original research question. There is thoughtful consideration of potential limitations of the data and/or analysis, and ideas for future work are clearly described.

Pattern

Evaluate Model - testing + training

limitations:

-missingness - excluding a huge proportion of data...

-definition of categories

-“reported age”, weight - rather arbitrary when considering what’s a “physical characteristic”

proportions + numbers (problem with representation)

variable selection process - based on visualizations

incomplete understanding of dataset

e.g., potential violation of independence

Ideas for future work:

more focus on behavior vs. characteristics

explore other behavior variables

implicit bias is hard to measure

less focused aspects like hair color + extending to accessory + outfits? (detailed description variable)