Stop Don’t Hit

Professor Morales
Topics in Applied Data Science for Social Sciences
Columbia University
Team 3: Alexandra Plassaras, Brandon Wolff, Zachary Heinemann, and Stephanie Langeland
April 26, 2017

Executive Summary

Who: Use of Police Force in Stop & Frisk Data

What: GIS Exploratory Analysis, Logit Model, Time Series Analysis, & Multi-Level Spatial Analysis

Where: Bronx, New York

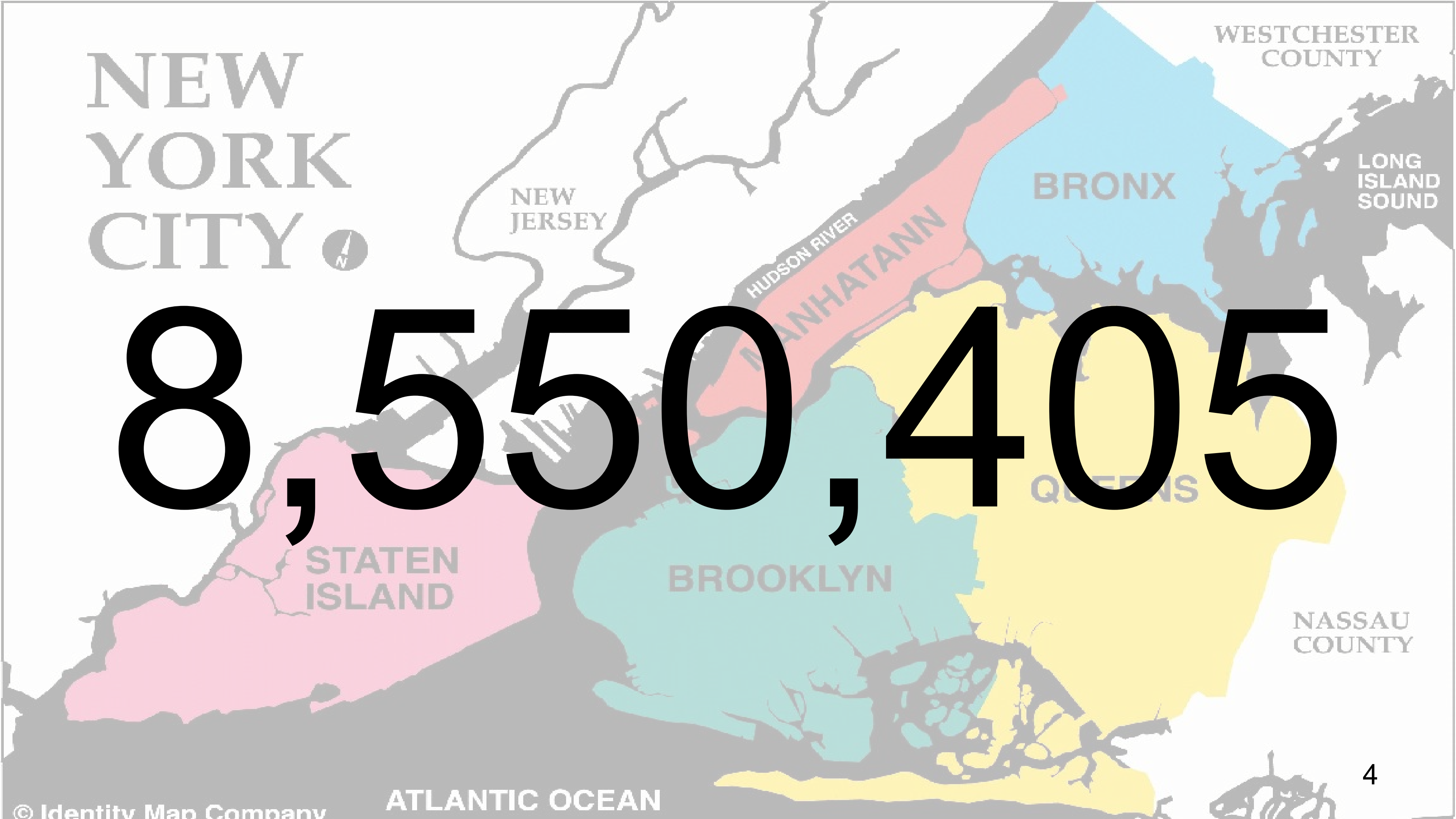
When: 2015

Why: To understand the impact of human bias in police officers when making the decision the stop, question, and frisk someone based on the demography of the vicinity.

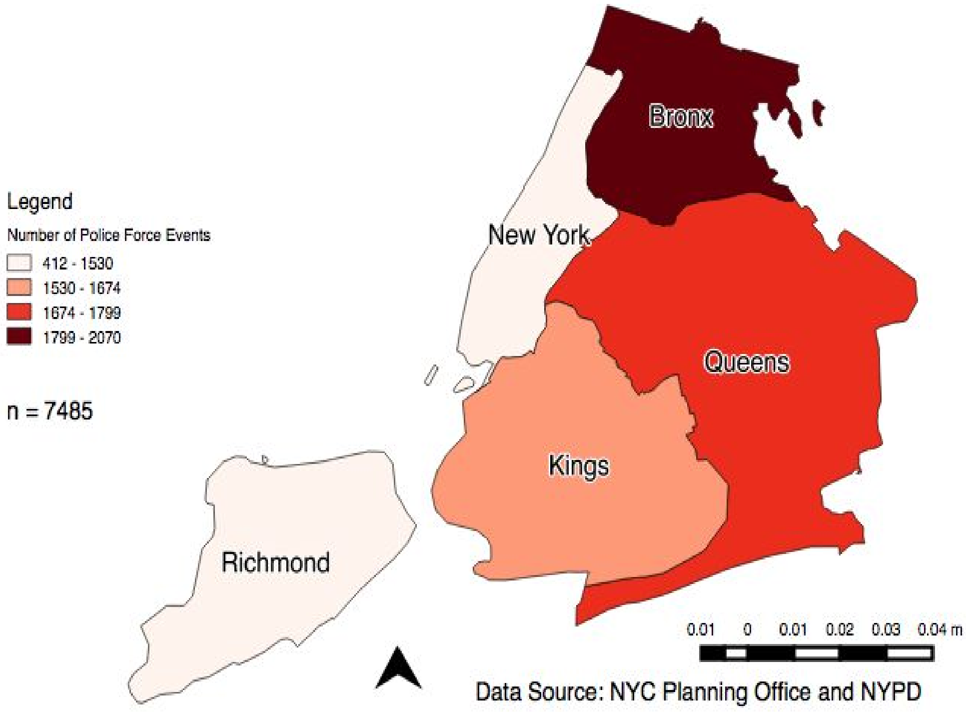
Outline

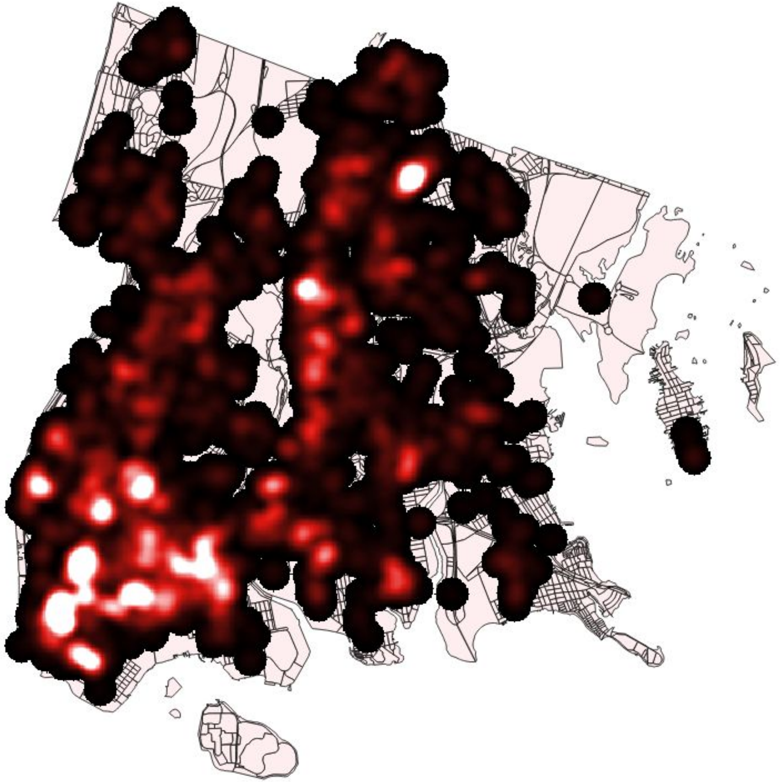
- GIS Exploratory Data Analysis
- Time Series Analysis
- Multi-Level Spatial Analysis
- Conclusion
- Takeaways
- Appendices

8,550,405


NEW YORK CITY
N
NEW JERSEY
HUDSON RIVER
MANHATANN
BRONX
WESTCHESTER COUNTY
LONG ISLAND SOUND
STATEN ISLAND
BROOKLYN
NASSAU COUNTY
ATLANTIC OCEAN
© Identity Map Company

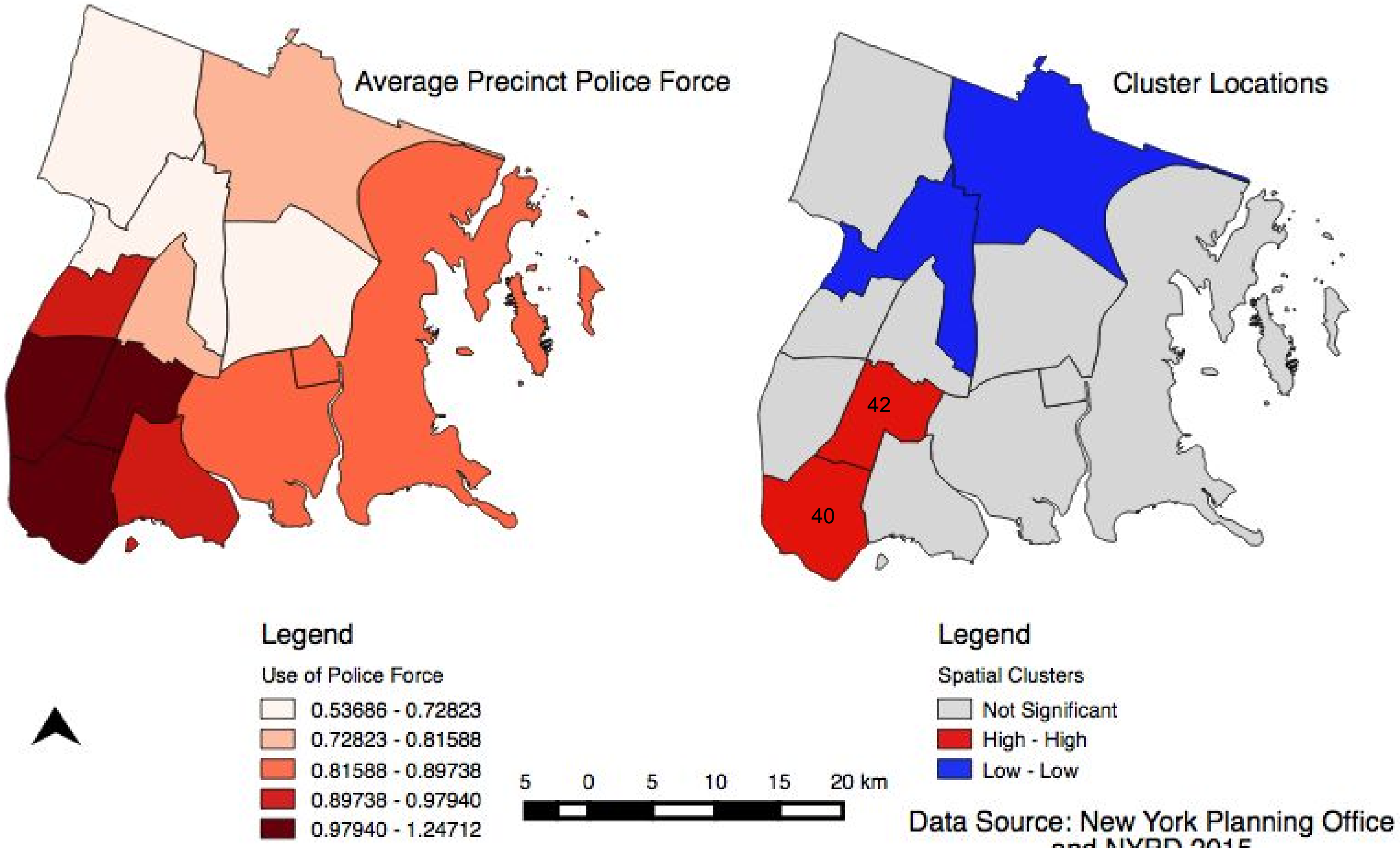
The South Bronx has the Highest Levels of Police Force





Number of unique locations: 1,199
Number of physical force events: 2,004

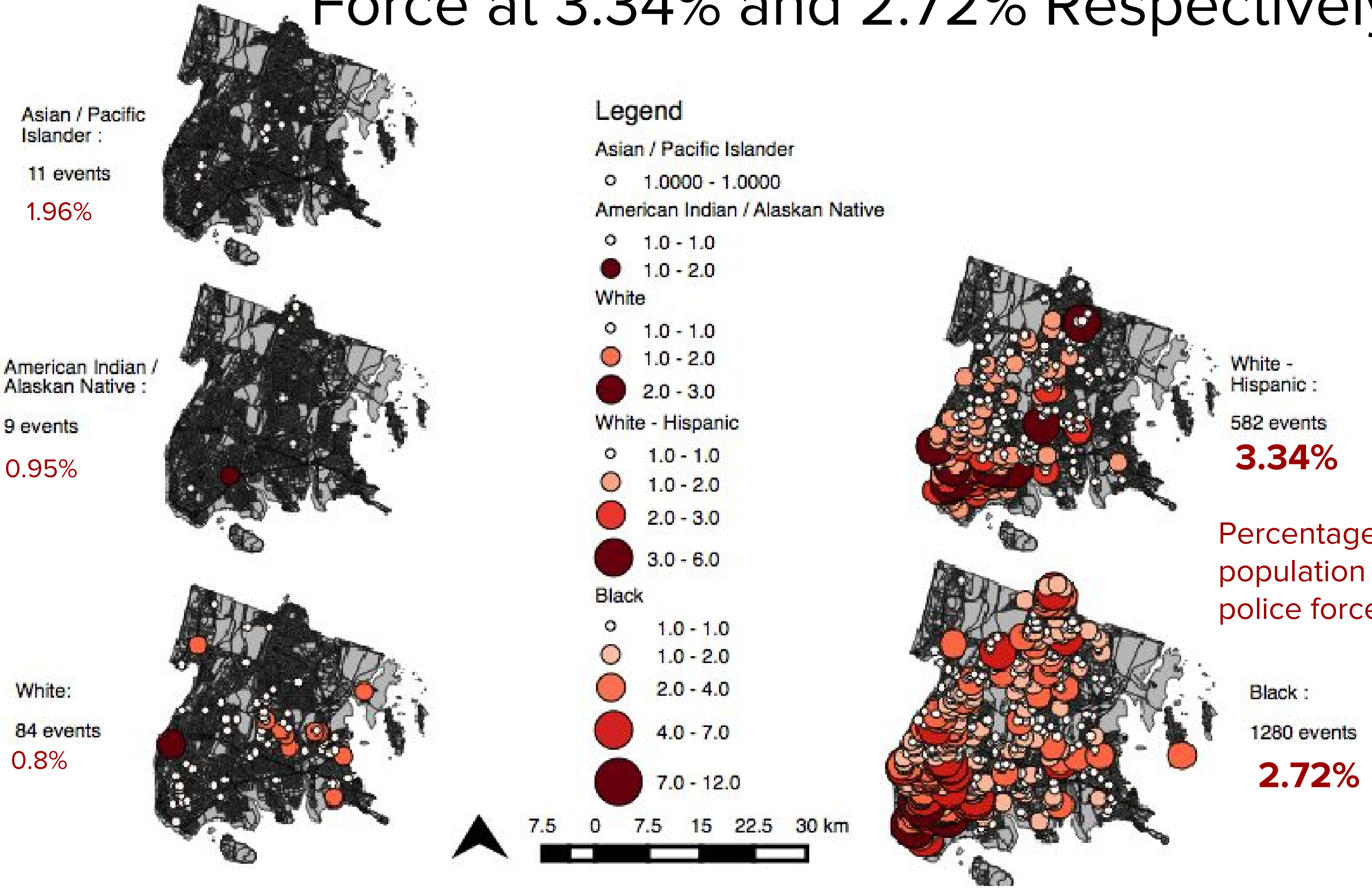
Statistically Significant Clusters of Force in Precincts 40 & 42



Moran's I = 0.46

Data Source: New York Planning Office and NYPD 2015

White Hispanics and Blacks had the Highest Instances of Police Force at 3.34% and 2.72% Respectively



Asian / Pacific Islander: 1.96%

American Indian / Alaskan Native: 0.95%

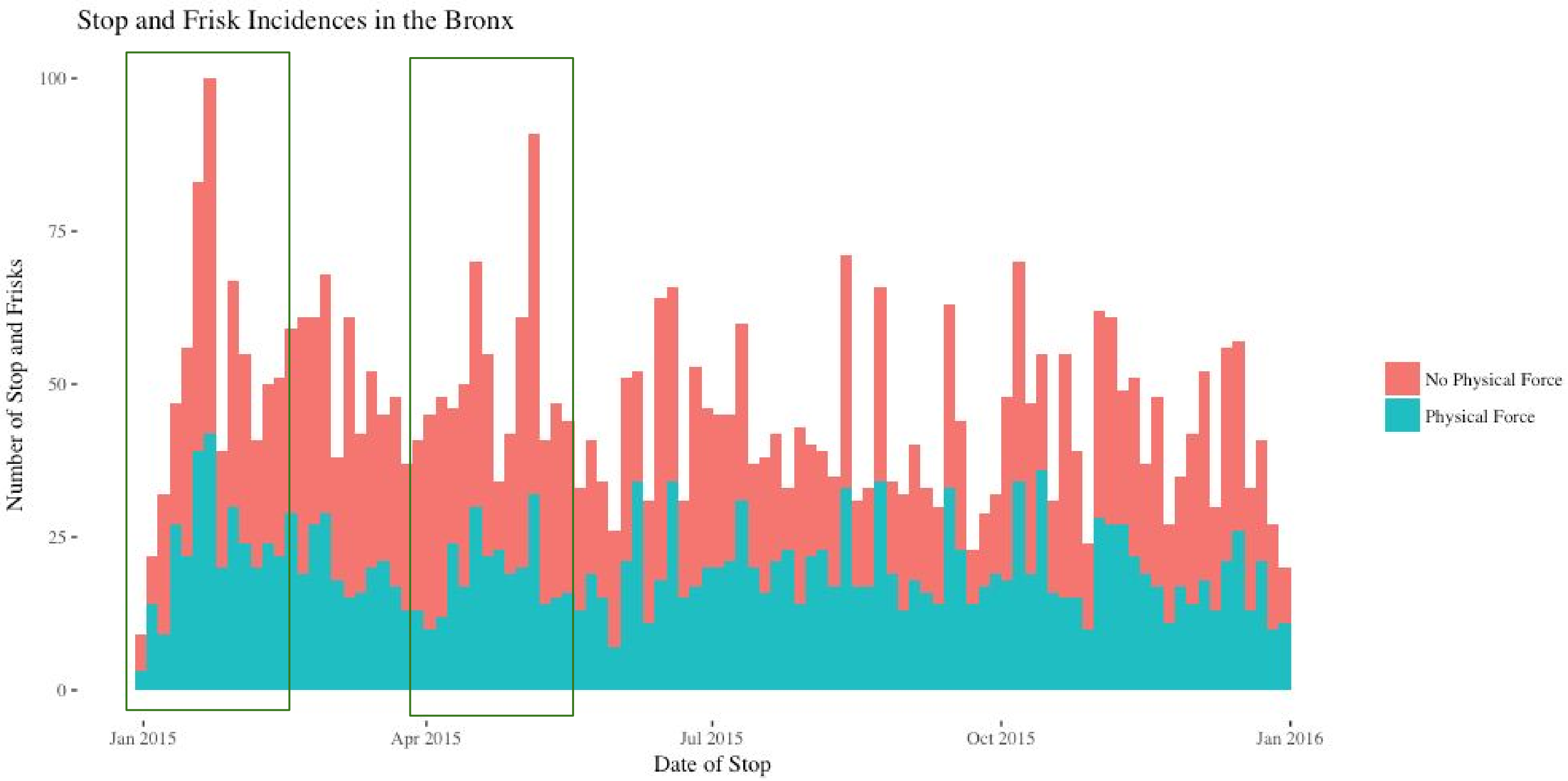
White: 0.8%

White - Hispanic: **3.34%**

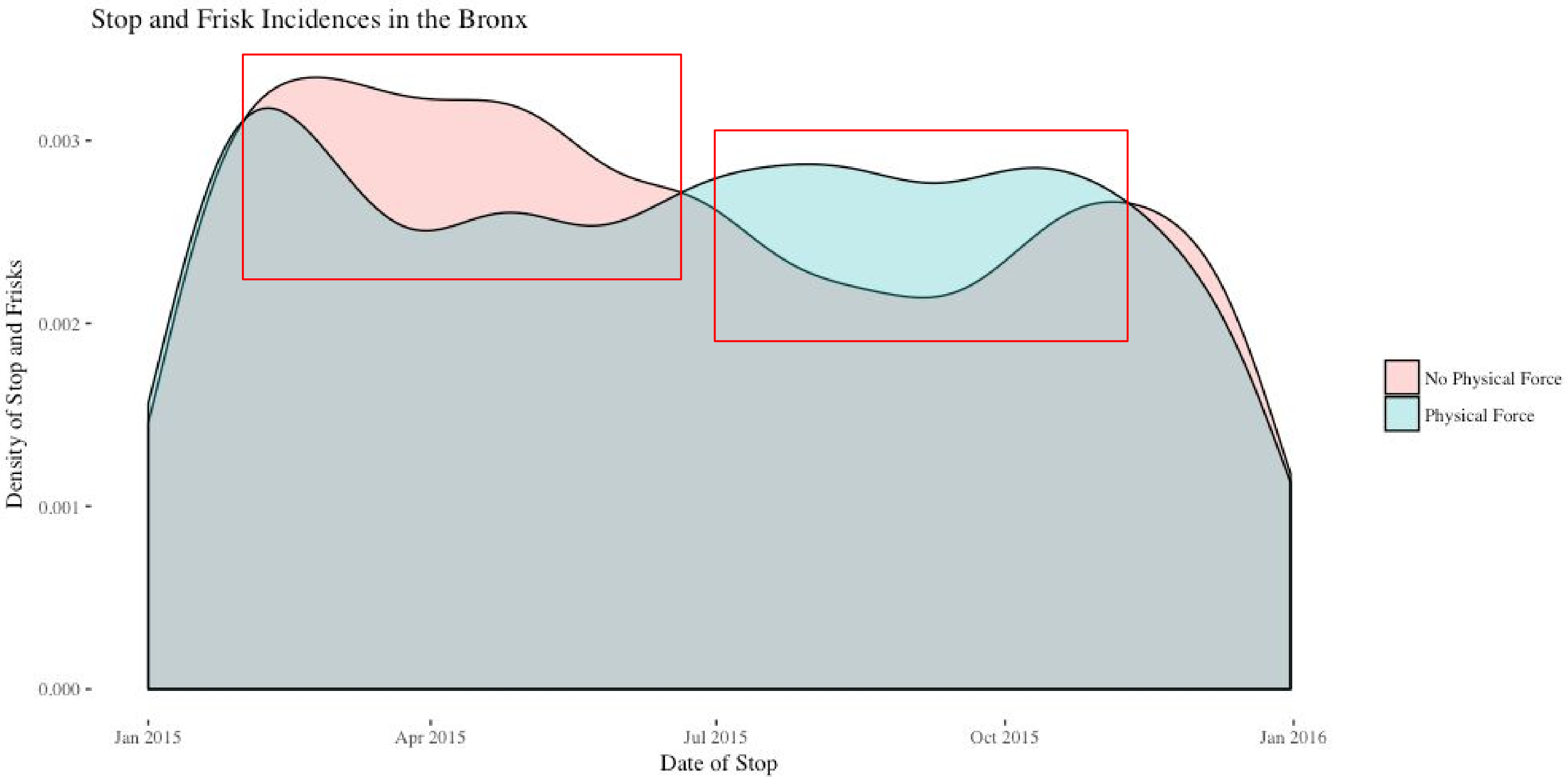
Black: **2.72%**

Percentage of each race of the Bronx population being stopped and frisked with police force

Hot Spots in Crime Appear After NYE and Summer



Force is More Prevalent in Summer, Less in Spring



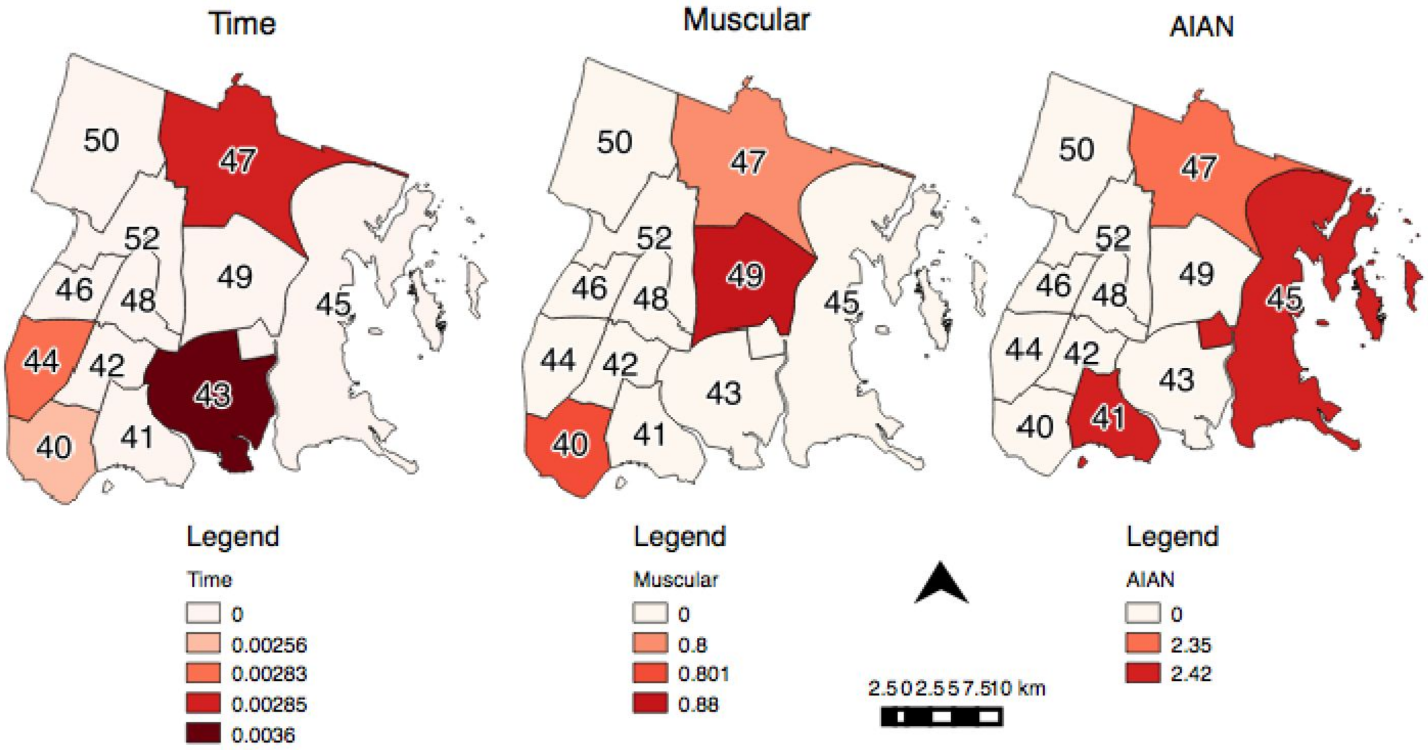
ARIMA Finds No Major Time Trends

- When accounting for time (two incidences before), the results are the same as the logistic regression!
- Force is **1.42 times more likely** for American Indian/Alaskan Natives
- Force is **1.06 more likely** in high incidence neighborhoods
- Force is **1.06 more likely** at abnormal times
- Force is **1.19 more likely** during arrests, frisks, and when the build is muscular, ***even when the suspect is not armed or does not use violence themselves***
- Interaction terms find that all of these effects are consistent over time!

Bayesian MLM Can Discern Geo-Trends

- Previous graphical analysis demonstrated potential differences in trends throughout different regions of the Bronx
- No evidence for a biasing effect ***overall,*** but what about stratified by the different precincts? Do any effects that we've observed change within ***particular precincts?***
- Began working with a Bayesian Multilevel Model nesting the previous model within precinct

MLM Finds Precinct Differences in Time, Build, and Race



Conclusion

- **Goal:** The output of this data analysis will conclude with NYPD policy recommendations to reduce human bias and increase subjectivity when making the decision to stop, question, and frisk.
- **Insight 1:** There is not significant evidence for bias on the basis of race and gender used by police in the Bronx.
- **Insight 2:** There are still spatial hotspots where more force is occurring and the data do not necessarily indicate that these hotspots are where more crime is occurring.
- **Insight 3:** Force is predominantly used on people with heavier builds, regardless of if the suspect used force or was armed.

Takeways

- **Takeaway 1:** Educate police officers on use of force when necessary (allocate resources according to clusters).
- **Takeaway 2:** Overall, the effects we observe in the Bronx in the whole are consistent throughout the Bronx, indicating most training initiatives may be streamlined.
- **Takeaway 3:** More incidences happen in the summer, it may be helpful to educate officers to be sensitive to these changes in time with end of spring non-violent stop-and-frisk training.

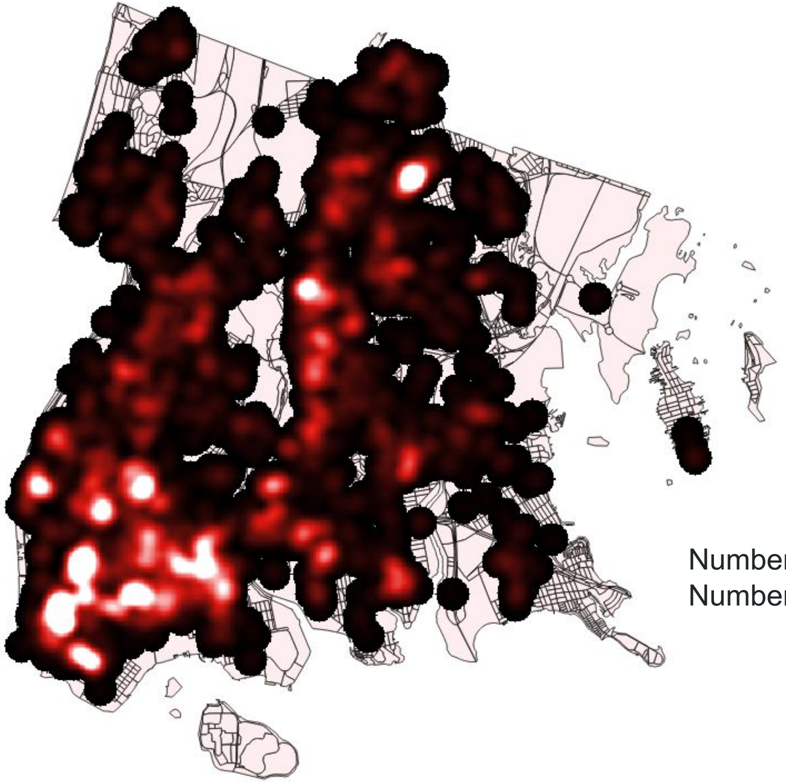
Appendices

Model Finds Results Relatively Consistent Across Bronx

- Coefficients are significant (consistently positive or negative in Bayesian term) in these **Precincts**:
- Gender: Negatively predicts force [**41**]
- American Indian/Alaska Native: None
- Black: Negatively predicts force [**40, 45, 49**]
- AAPI: None
- Hispanic: Negatively predicts force [**41, 42, 44, 45, 46**]
- Precincts with High Intercepts [**44, 45, 46**]

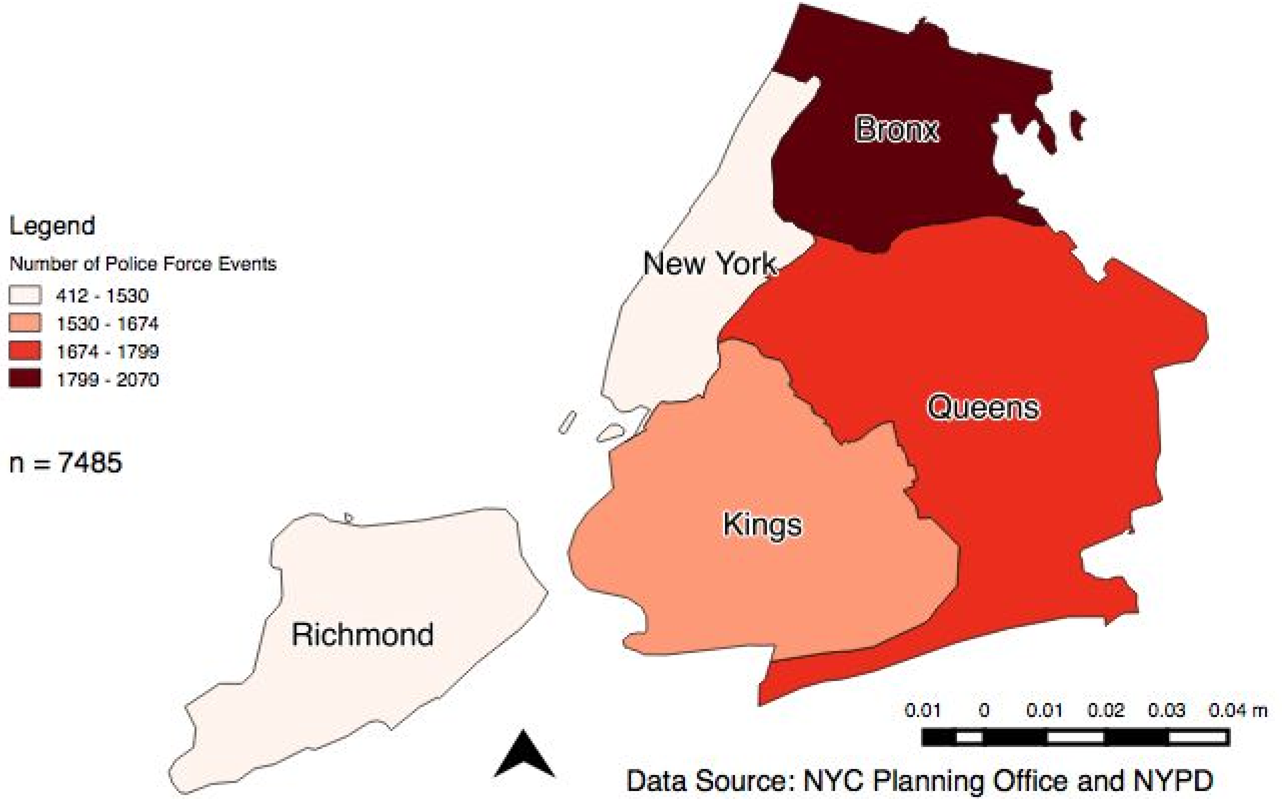
Note: This model does not control for time or build.

Lower Bronx has Higher Incidents of Police use of Physical Force



Number of unique locations: 1,199
Number of physical force events: 2,004

Bronx has the Highest Levels of Police Force Out of All Boroughs



Stop & Frisk Events by Race:

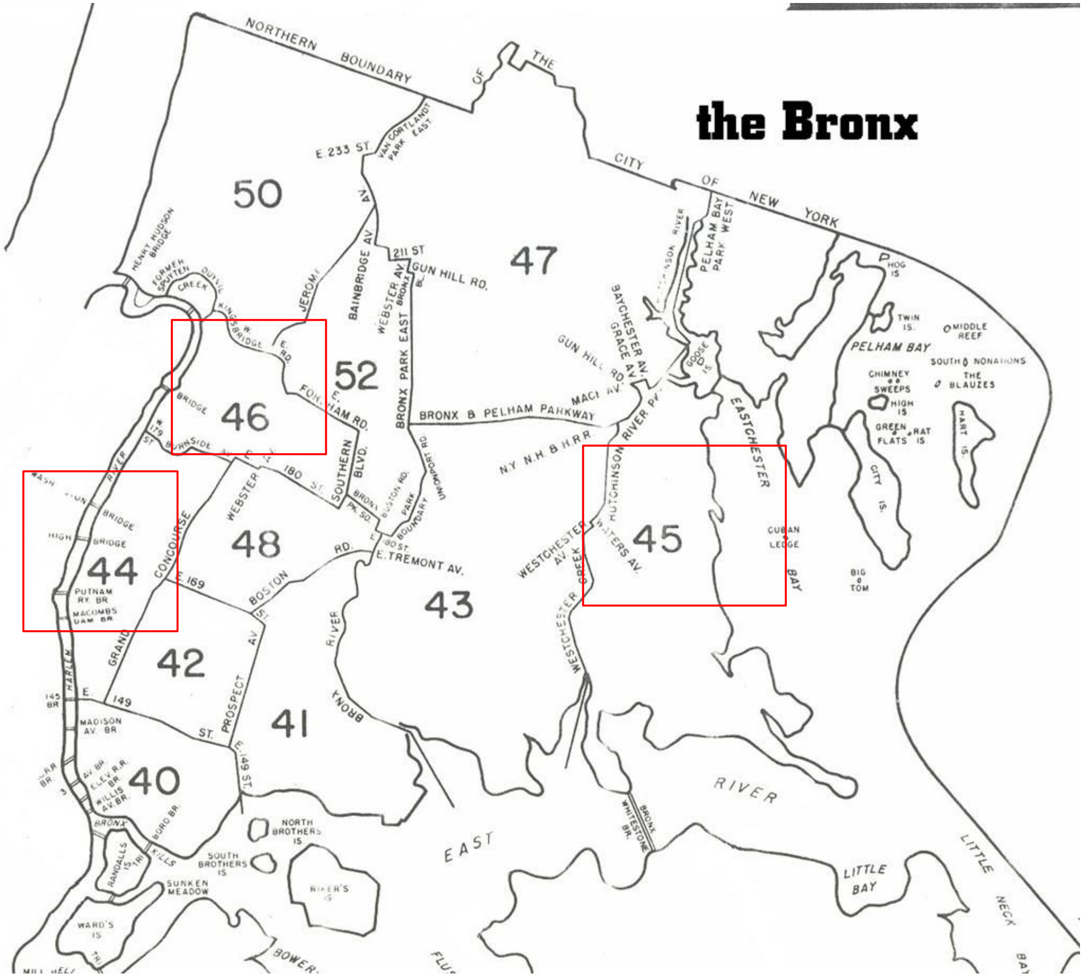
Black: 12,919

White/Hispanic: 4,907

White: 2,389

Asian/Pacific Islander: 1,043

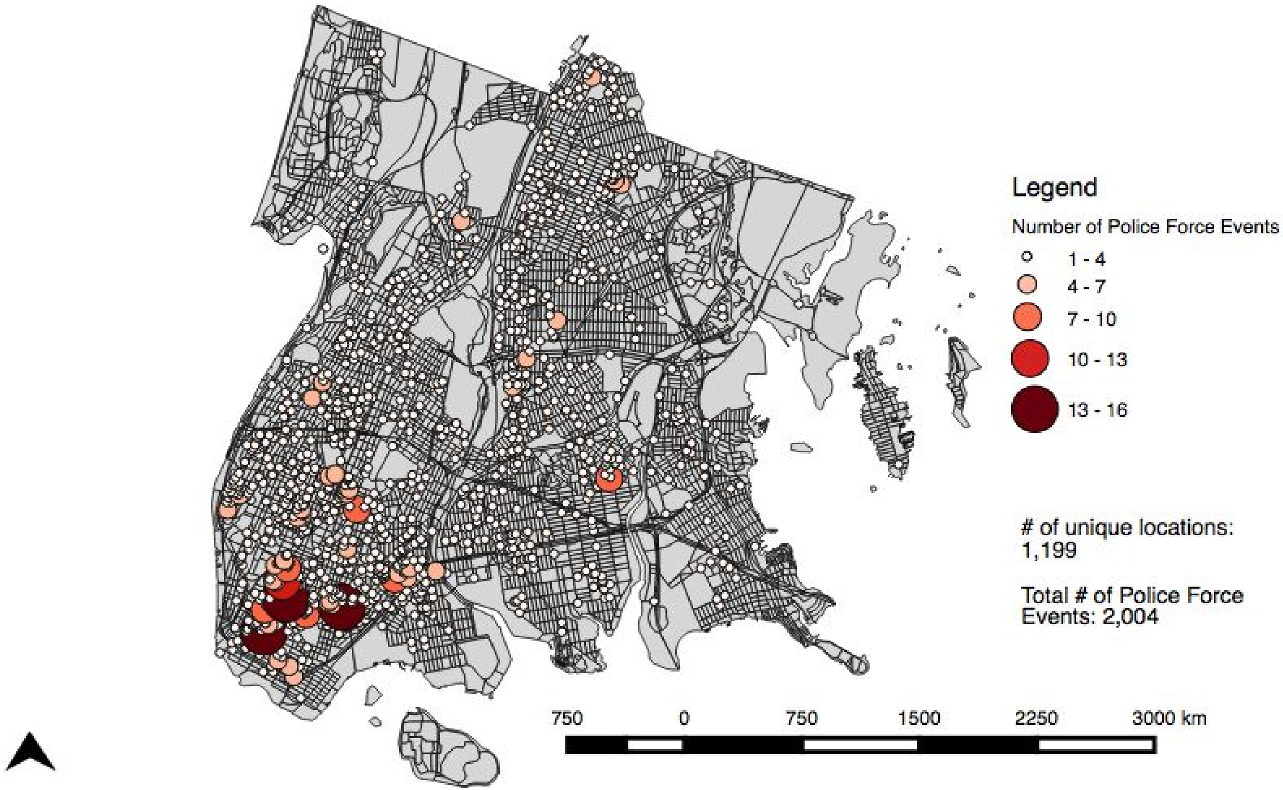
American Indian/Alaskan Native: 76


the Bronx
NORTHERN BOUNDARY OF THE CITY OF NEW YORK
50
47
52
46
48
44
45
43
42
41
40
E. 233 ST.
VAN CORTLANDT PARK EAST
211 ST
GUN HILL RD.
BAINBRIDGE AV.
WEBSTER AV.
BRONX PARK EAST
JEROME AV.
BRONX & PELHAM PARKWAY
BAYCHESTER AV.
GRACE AV.
PELHAM BAY PARK WEST
PELHAM BAY
HOG IS.
TWIN IS.
MIDDLE REEF
SOUTH NONATIONS
THE BLAUZES
CHIMNEY SWEEPS
HIGH IS
GREEN FLATS
RAT IS.
HART IS.
CITY IS.
CUBAN LEDGE
BIG TOM
EASTCHESTER BAY
GOOSE IS.
HUTCHINSON RIVER PY.
WATERS AV.
WESTCHESTER AV.
WESTCHESTER CREEK
N.Y. N.H. & H.R.R.
E. TREMONT AV.
180 ST.
BURNSIDE AV.
SOUTHERN BLVD.
BOSTON RD.
CONCOURSE
GRAND
E. 169
E. 149
PROSPECT AV.
E. 149 ST.
BRONX RIVER
HARLEM RIVER
HIGH BRIDGE
PUTNAM RY. BR.
MACOMBS DAM BR.
MADISON AV. BR.
WILLIS AV. BR.
BRONX KILLS
RANDALLS IS.
WARD'S IS.
SUNKEN MEADOW
NORTH BROTHERS IS.
SOUTH BROTHERS IS.
RIKER'S IS.
EAST RIVER
BRONX WHITESTONE BR.
LITTLE BAY
LITTLE NECK BAY
HENRY HUDSON BRIDGE
FORMER SPUYTEN DUYVIL CREEK
W. KINGSBRIDGE RD.
E. FORDHAM RD.
UNIONPORT RD.

How to Account for Time in The Data

- We ran a model with time -- nothing changed and time was not significant!
- Further analysis (Durbin Watson Test) finds that lags are necessary.
- Auto ARIMA model finds that optimal ARIMA is (2, 0, 0); two lags, no differences, no moving averages.

Prevelance of Police Force at Stop & Frisk Locations in the Bronx



Data Source: NYC Planning Office & NYPD 2015

Bayesian MLM

- Started with just race and gender terms to troubleshoot the model
- Working on running a more complicated model (this took 45 minutes to run!)

```
sf_bronx <- sf_bronx[is.na(sf_bronx$race) == FALSE & is.na(sf_bronx$sex) == FALSE, ]
mdata <- list(Force = sf_bronx$pforce, AIAN = (sf_bronx$race == "American Indian / Alaskan Native"),
AAPI = (sf_bronx$race == "Asian / Pacific Islander"), Black = (sf_bronx$race == "Black"), Hispanic =
(sf_bronx$race == "White - Hispanic"),Female = (sf_bronx$sex == "Female"), Precinct =
as.factor(sf_bronx$pct))
model.geo <- alist(
  Force ~ dbinom(1, p),
  logit(p) <- a_sub[Precinct] + b_aian[Precinct]* AIAN + b_aapi[Precinct] * AAPI + b_black[Precinct] *
Black + b_hisp[Precinct] * Hispanic + b_gender[Precinct] * Female,
  ## we would need a b_sub term for every predictor

  c(a_sub,b_aian,b_aapi,b_black,b_hisp, b_gender)[Precinct] ~
dmvnorm2(c(a,baian,baapi,bblack,bhispanic,bgender),sigma_sub,Rho),
  ## this needs to include a term for each of the model predictors

  a ~ dnorm(0,2),
  baian ~ dnorm(0,3),
  baapi ~ dnorm(0,3),
  bblack ~ dnorm(0,3),
  bhispanic ~ dnorm(0,3),
  bgender ~ dnorm(0,3),

  ##needs to be one of these specifications for each of the terms; we can play with priors

  sigma_sub ~ dcauchy(0,2),
  Rho ~ dlkjcorr(4)
)

run.model <- map2stan(model.geo,
                      data=mdata, iter=7000, warmup=2000, chains=4,
                      cores=4)
```

ARIMA Model

```
> pvalues[pvalues < .05]
                                                   ar1
                                          0.000000e+00
                                                   ar2
                                          1.881980e-02
as.factor(bronx_full$race)American Indian / Alaskan Native
                                          1.293722e-02
                                              ac_incid
                                          2.434725e-04
                                               ac_time
                                          8.795484e-03
                                              arstmade
                                          1.461054e-13
                                               frisked
                                          0.000000e+00
                      as.factor(bronx_full$build)Muscular
                                          9.973307e-03
```

Initial Logistic Model Odds-ratios

```
                (Intercept)                             raceWhite raceAmerican Indian / Alaskan Native
                  0.7150492                             1.0080599                            5.6460052
raceAsian / Pacific Islander                 raceWhite - Hispanic                              sexMale
                  1.3646671                             0.9057247                            1.1272753
```

- According to the results being American Indian/Alaskan Native increases the odds relative to being white by 5.65 net of all other factors.
- This is significant at the 0.05 level.

Logistic Model with Control Variables

```
Call:
glm(formula = pforce ~ race + sex + build + ac_incid + ac_time +
    armed + offunif + inout + arstmade + explnstp + frisked,
    family = binomial(link = "logit"), data = sf_bronx)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7594  -1.1097  -0.5523   1.1425   2.1323

Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                        -1.48237    1.44505  -1.026  0.30497
raceAmerican Indian / Alaskan Native 1.89876    0.85257   2.227  0.02594 *
raceAsian / Pacific Islander        0.09870    0.49079   0.201  0.84062
raceBlack                          -0.11834    0.16410  -0.721  0.47082
raceWhite - Hispanic               -0.17329    0.16951  -1.022  0.30665
sexMale                            -0.22598    0.13623  -1.659  0.09716 .
buildMedium                        -0.12376    0.11234  -1.102  0.27060
buildMuscular                       0.85411    0.32745   2.608  0.00910 **
buildThin                           0.10556    0.11178   0.944  0.34499
ac_incid                           -0.23459    0.07444  -3.151  0.00163 **
ac_time                             0.14961    0.07981   1.875  0.06086 .
armed                              -0.10330    0.13790  -0.749  0.45379
offunif                            -0.13206    0.07428  -1.778  0.07542 .
inoutOutside                       -0.03059    0.08451  -0.362  0.71736
arstmade                            0.56219    0.08040   6.992 2.71e-12 ***
explnstp                            0.23796    1.42243   0.167  0.86714
frisked                             1.69796    0.09041  18.780  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5958.9  on 4347  degrees of freedom
Residual deviance: 5422.0  on 4331  degrees of freedom
  (234 observations deleted due to missingness)
AIC: 5456

Number of Fisher Scoring iterations: 4
```

body build = (heavy as reference group)

"ac_incid" = whether the event took place in a high crime or not.

"ac_time" = if event took place at abnormal time.

"armed" = person was armed or not.

"offunif" = if office was in uniform or not.

"inout" = whether event took place in or outside.

"arstmade" = if arrest was made or not.

"explnstp" = if the person was given reason for the stop or not.

"frisked" = if the person was frisked or not.

Final Logistic Model Odds-ratios

```
(Intercept) raceAmerican Indian / Alaskan Native        raceAsian / Pacific Islander
  0.2270982                            6.6776329                           1.1037370
  raceBlack                  raceWhite - Hispanic                            sexMale
  0.8883948                            0.8408964                           0.7977318
buildMedium                        buildMuscular                           buildThin
  0.8835890                            2.3492942                           1.1113378
   ac_incid                              ac_time                               armed
  0.7908981                            1.1613836                           0.9018538
    offunif                          inoutOutside                           arstmade
  0.8762849                            0.9698725                           1.7545164
   explnstp                              frisked
  1.2686619                            5.4627790
```

- According to the results being American Indian/Alaskan Native increases the odds relative to being white by 6.68 net of all other factors. This is significant at the 0.05 level.

- It is also good to note that according to the results having a muscular build increases the odds relative to being heavy by 2.35 net of all other factors. This is significant at the 0.01 level.

References

- Team 3 GitHub Link with project details: https://github.com/amp5/QMSS_G5069_Applied_D_S

- https://www1.nyc.gov/site/planning/data-maps/nyc-population/population-facts.page

- http://www.nyc.gov/html/nypd/html/faq/faq_police.shtml#1