Use of Force Final Project Report

Kaylhan Garcia, Adam Billen, and Phoebe Lin

2023-10-17

Use of Force Final Project Report

Introduction:

Police have existed in America in some form for nearly its entire history. Police forces themselves have their philosophical origin in Patrick Colquhoun's 1797 "Treatise on the Police of the Metropolis" (Colquhoun 1797). These ideas first took form in London in 1829, but Colquhoun himself spent his early career as a British agent for cotton manufacturers. His conception of policing grew from his time spent enforcing slave codes and working with slave patrols in Virginia. The first hiring of police through legislative means was in Boston in 1838, but before then "police" simply referred to slave patrols (Potter 2022). North Carolina's first state police force was formed in order to stamp out the publishing of "The Liberator," William Lloyd Garrison's weekly abolitionist newspaper, soon after he was almost killed in a mob attack in Boston in 1835 (Lepore 2020). The modern American police force began and grew from these dark roots of American history.

The existing literature around racialized policing has established that Black, Hispanic and American Indian/Native Americans are disproportionately likely to be killed by police in America, especially at younger ages. One study using data from between 2013 and 2018 found that Black men are ~2.5 times more and Black women ~1.4 times more likely to be killed by police in their lifetime than white men and white women, respectively (Fryer 2019). This amounts to some 96 out of 100,000 Black men and boys, between 36 and 81 American Indian/Alaskan Native men and boys, and 53 out of 100,000 Latino men and boys being killed by police over their lifetime, as compared to about 39 out of 100,000 white men and boys being killed (Fryer 2019). The risk is substantially lower for women across all racial groups. Between 2.4 and 5.4 Black women and girls, 2.4 American Indian/Alaskan Native women and girls, and 2 Latino and white women and girls out of 100,000 are expected to be killed by police throughout their lifetime (Fryer 2019). This relationship holds even when controlling for uses of police violence against criminal vs non-criminal subjects (Ross et al. 2021). This existing set of literature focuses broadly on fatal violence inflicted by American police, often using national level data collected by the Washington Post (Washington Post). This data is, however, scattered and often inconsistent. The Washington Post's database, for example, relies on local news reports (Nix 2022). Official national level data compiled by the Bureau of Justice Statistics also fails to provide a full picture. There is no federal regulation mandating the format or way in which individual police departments collect data, and so America's 18,000+ police departments each have control over how, when, and where data is collected. Matthew Hickman, a professor and former Bureau of Justicts Statistics employee, described the situation as a "huge mess" wherein police departments can each do "whatever they want" (Katie 2019).

In the past decade, however, a new political force, the Black Lives Matter movement, rose up and sparked a conservation around the appropriate use of force by the police. The public began pushing for more transparency around the use of force, and policing generalky (Schwartz, 2020). A number of local governments have, at least partially, met this demand. These local governments have mandated that their local police forces provide public data on each case of a police officer using force against a civilian. The resulting datasets provide a way to analyze use of force more broadly, rather than soley fatal violence by the police. There has not, however, been a large wave of research exploring this data. Instead, local agencies, often Police Accountability Offices, typically produce a yearly report analyzing the past year's use of force data. While these reports tend to draw lots of public attention in the news, they are are generally relatively surface level. They tend to include general information about the number of uses of force and any clear disparities in the use of force. They do not, however, include more advanced methods such as regression. To expand on this existing literature, we will emplore more advances methods such as logistic multiple regression to explore the relationship between demographic factors such as subject or officer race and the severity of a given use of force.

Hypotheses:

We hypothesize that Black/African American subjects are more likely to have serious force used against them than other racial groups, particularly white subjects. While this is our primary topic of interest, we expect that other demographic groups for both subjects and officers will have varying degrees of likelihood to have serious force used against them as well. We will also explore whether particular races of officers are more or less likely to use serious force against particular races of subjects. We hypthesize that white officers will be more likely to use serious force against black subjects than black officers.

Methods:

By including all of the demographic information provided, we will build a clearer picture of exactly what demographic groups are more likely to have severe uses of force inflicted upon them. By subsetting our data to specific officer races, we will examine which particular combinations of demographic factors result in more severe uses of force, providing a fuller picture of how officers and citizens interact. We have chosen logistic multiple regression as it is the most applicable and interpretable model for this particular data. All of the data for each city is categorical, and we have thus created dummy variables for each key cateogrical variable in each dataset. We have chosen a regular logistic regression instead of an ordinal logistic regression because each categorical independent variable is unordered. Including categorical variables such as race as an ordered set of numbers would make certain races have a larger effect on the dependent variable, hurting the accuracy and interpretability of the resulting coefficients. We have created a binary dependent variable (described) below as our output.

This study involves four unique American cities which vary widely in their locations and demographic makeup. We will study use of force data in Washington, DC (East Coast), Indianapolis (Midwest), Portland (West Coast), and Seattle (the Northwest). In order to compare these models we have created a binary Serious/non-Serious variable for each dataset. This variable is based on a pre-existing variable in the dataset for Washington, DC. This variable, "serious," is set to 1 when a use of force involves a firearm discharge, strike to the head, canine bite, neck restraint, results in a serious physical injury, or results in the risk of or loss of consciousness, death, or serious disfigurement. The data for Seattle includes a similar risk tier system from one to three, which we have simplified to a binary variable where levels two and three are serious and level one is non-serious. The data for Portland utilizes a four-tiered system, which we have simplified to make tiers two through four serious uses of force and tier four non-serious. The data for Indianapolis breaks down uses of force by individual type, which we have categorized as serious or non-serious to match the other cities. Still, these categories should not be read as precisely comparable. Systems of reporting and publishing data vary between cities, but this variable still provides a stronger point of reference between cities.

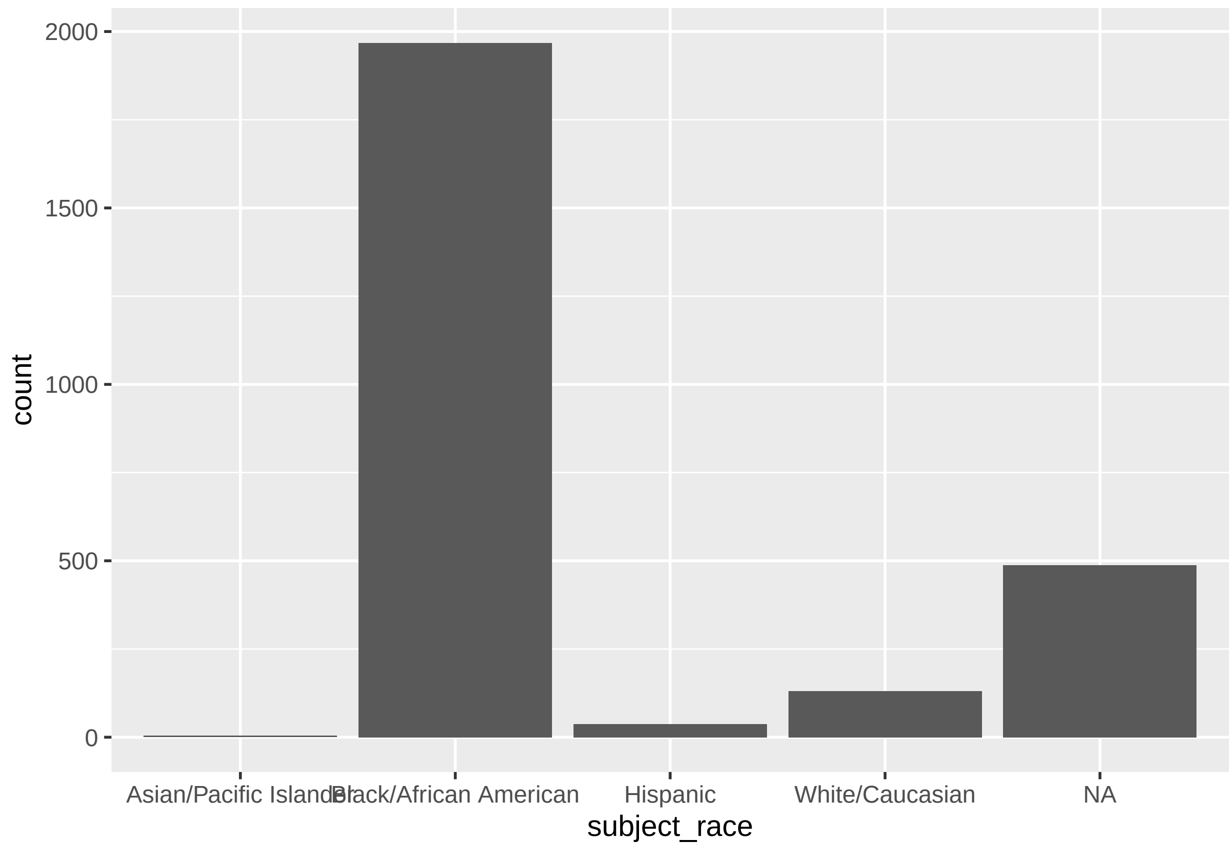
Data Exploration

Washington, DC

1. Distribution of Incident Counts by Subject Race Black/ African American have a significantly higher amount of incident counts in 2020. Conversely, Asian/Pacific Islander individuals appear to experience the least instances of use of force.

```
df <- read.csv("../data/UOF_DC_2020.csv")
mpd_dat <- df
```

```r
ggplot(data = mpd_dat, mapping = aes(x = subject_race)) +
  geom_bar()
```



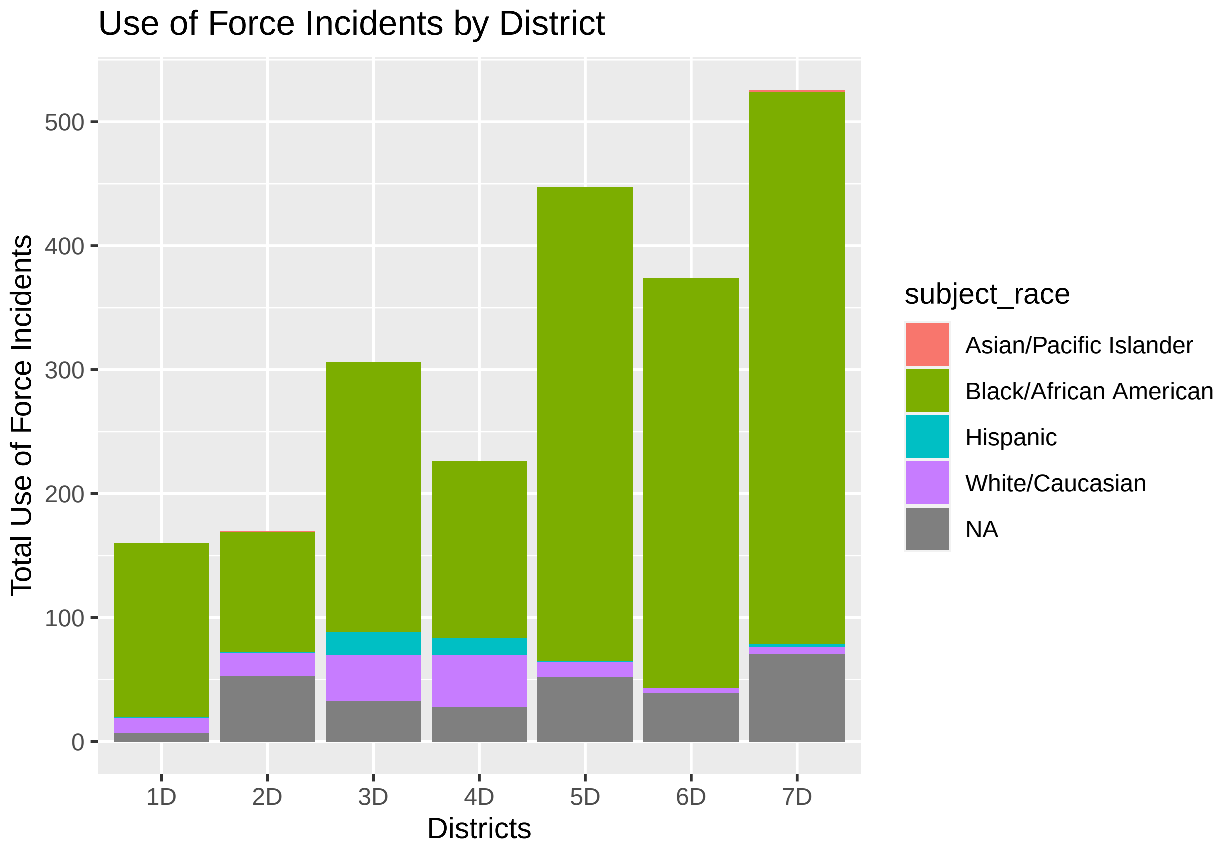
2. UOF Incident Counts by District and Subject Race Across all districts in DC, it is prevalent that Black/African Americans experience UOF at a higher rate than their counterparts.

7th District has the highest total UOF incidents out of all the DC districts.

```r
force_by_district_race <- mpd_dat %>%
group_by(OfficerAssignment, subject_race) %>% summarize(Total_Use_of_Force = n()) #total num of cases o
```

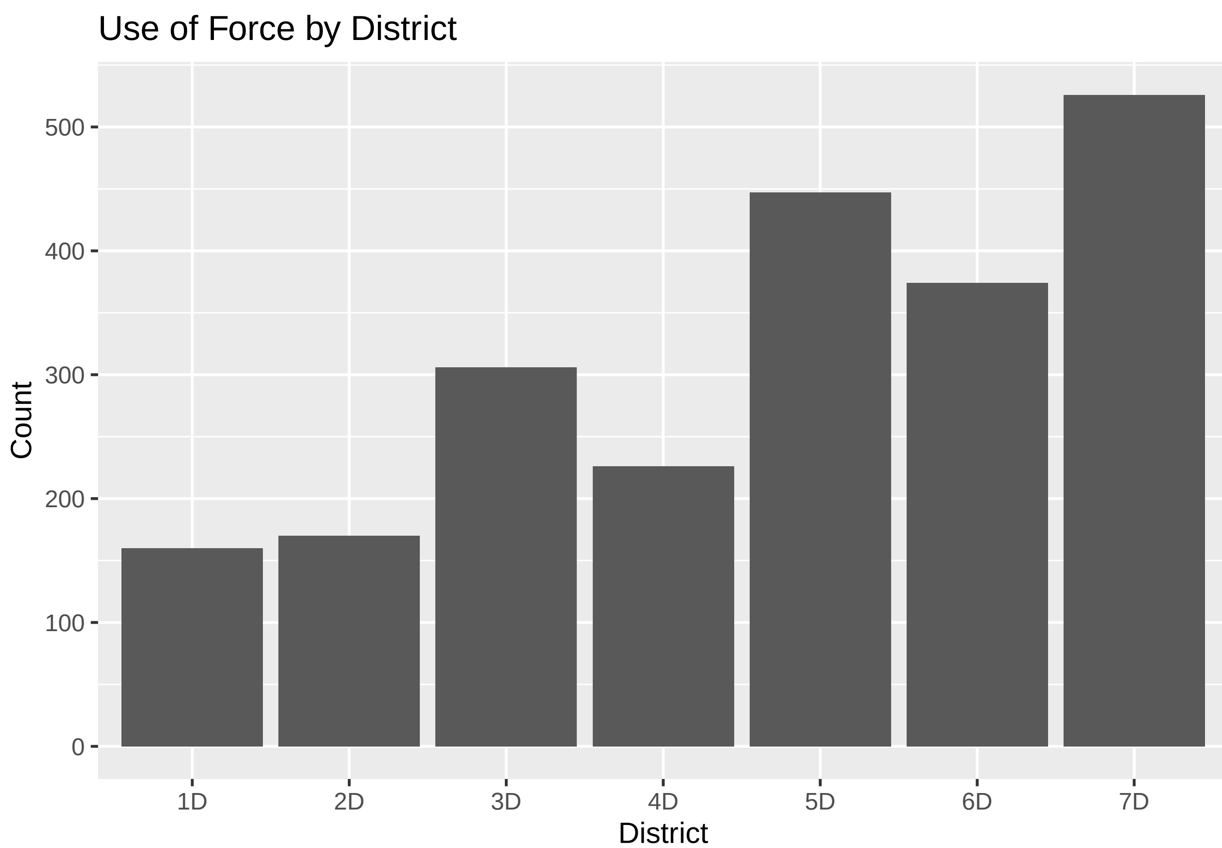
```
## `summarise()` has grouped output by 'OfficerAssignment'. You can override using
## the `.groups` argument.
```

```r
ggplot(subset(force_by_district_race, OfficerAssignment %in% c("1D", "2D", "3D", "4D", "5D", "6D", "7D"
       aes(x = OfficerAssignment, y = Total_Use_of_Force, fill = subject_race)) +
  geom_bar(stat = "identity") +
  labs(title = "Use of Force Incidents by District") +
  xlab("Districts") +
  ylab("Total Use of Force Incidents")
```



3. Use of Force by District. 1D, 2D, and 3D are shown to be the safest districts in DC, whereas 7D is deemed to be the most dangerous.

```
districts <- c("1D", "2D", "3D", "4D", "5D", "6D", "7D")
mpd_dat_filtered <- mpd_dat[mpd_dat$OfficerAssignment %in% districts, ]

uof_counts <- mpd_dat_filtered %>%
  group_by(OfficerAssignment) %>%
  summarise(Count = n())

ggplot(uof_counts, aes(x = OfficerAssignment, y = Count)) +
  geom_bar(stat = "identity") +
  labs(title = "Use of Force by District",
       x = "District",
       y = "Count")
```



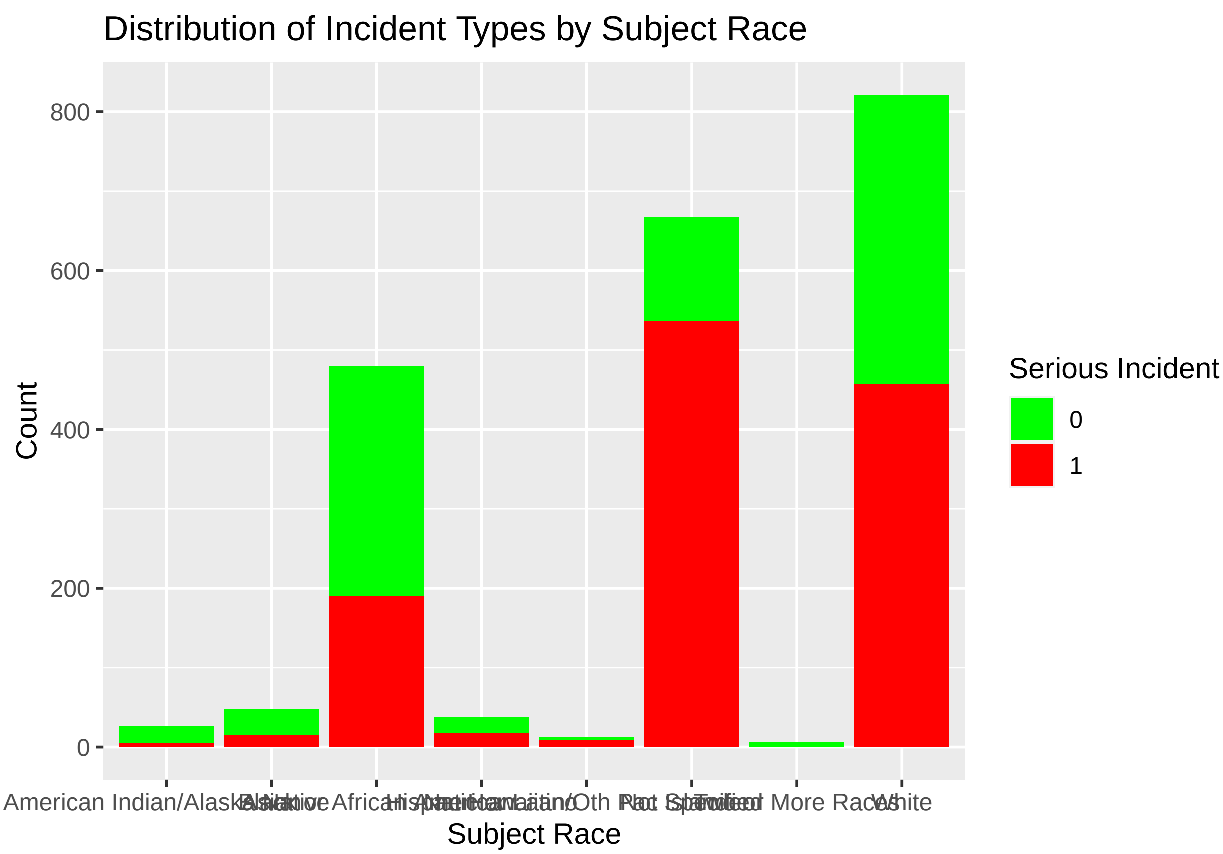
Seattle, Washington

4. Distribution of Incident Types by Subject Race There is a huge amount of 'Not Specified' values for this data visualization during the year 2020.

White people have a shared count amount of serious incidents regarding use of force my the police department.

The next following group that stands out is 'Black or African American,' with level 1 - Use of Force being greater than the other levels.

```r
ggplot(seattle_2020, aes(x = Subject_Race, fill = factor(serious))) +
  geom_bar(position = "stack") +
  labs(
    title = "Distribution of Incident Types by Subject Race",
    x = "Subject Race",
    y = "Count",
    fill = "Incident Type"
  ) +
  scale_fill_manual(values = c("0" = "green", "1" = "red"), name = "Serious Incident")
```

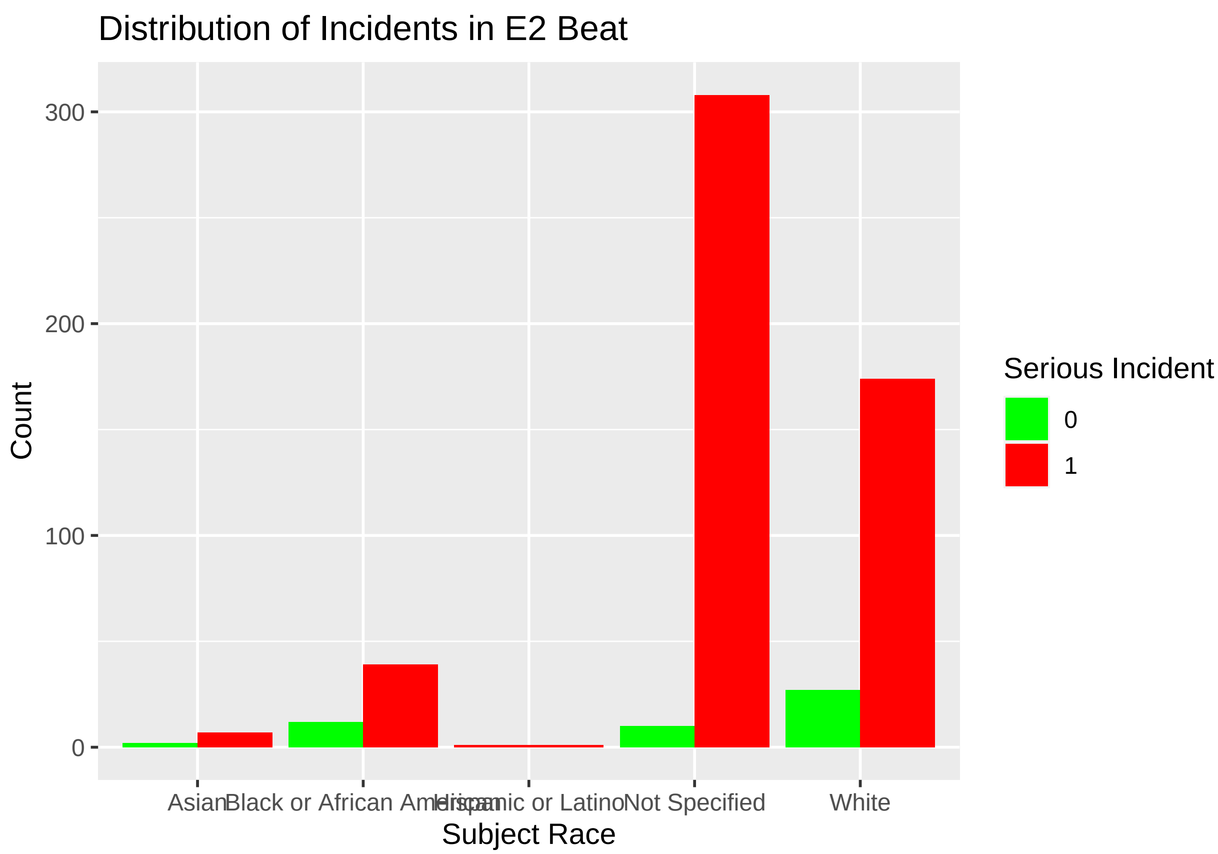


5. Comparative Analysis: Incidents in E2 by Subject Race E2 has the highest amount of incidents across all the other beats in Seattle. There are a total of 52 beats reported in the dataset.

Many of the 'Subject Race' is left unreported or 'Not Specified' in the year 2020. The next leading race that are subjected to incidents in the E2 beat are White, then it is Black or African American.

It is notable to mention that many of the serious incidents are reported as 'Not Specified' in the E2 beat.

```
e2_data <- subset(seattle_2020, Beat == "E2")

ggplot(e2_data, aes(x = Subject_Race, fill = factor(serious))) +
  geom_bar(position = "dodge", stat = "count") +
  labs(
    title = "Distribution of Incidents in E2 Beat",
    x = "Subject Race",
    y = "Count",
    fill = "Serious Incident"
  ) +
  scale_fill_manual(values = c("0" = "green", "1" = "red"), name = "Serious Incident")
```



Indianapolis, Indiana

In Indianapolis, 52.7% of the population are Whites, 28.5% are Black or African American and 3.82% are Asian. The presence of an 'Unknown' category for subject race suggests that data collection may not be complete or consistently conducted across all incidents. But overall, the "Black" category shows significantly higher number of incidents comparedito other races.

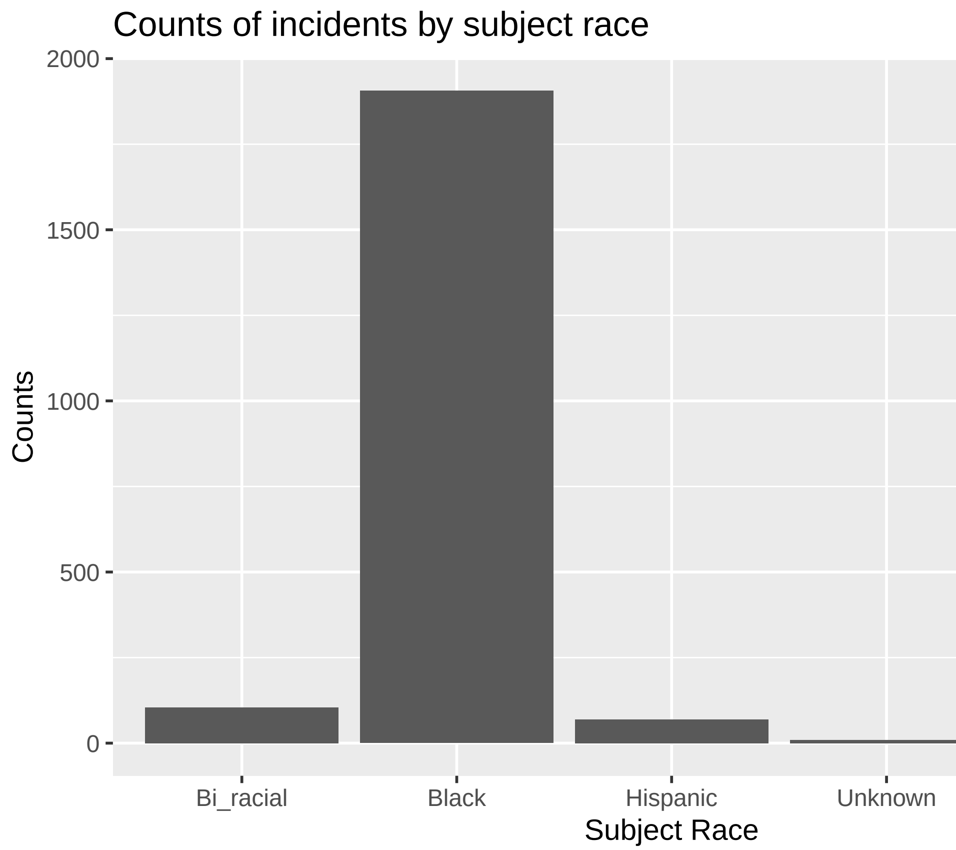
The seriousness of the cases seems to vary across races, but for those races with a significant number of incidents (Black and White categories), non-serious incidents outnumber serious incidents. This could indicate that while the use of force is more commonly reported as non-serious, serious use of force incidents still occur with notable frequency in these populations.

```
ind_dat <- read_csv("../data/ind_dat_2020_final.csv")
```

```
## Rows: 2793 Columns: 59
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (30): OCCURRED_DT, OCCURRED_TM, UDTEXT24A, UDTEXT24B, UDTEXT24C, UDTEXT...
## dbl  (27): OBJECTID, INCNUM, CITNUM, CIT_AGE, OFFNUM, OFF_AGE, OFF_YR_EMPLOY...
## lgl   (1): CIT_WEAPON_TYPE
## date  (1): datetime
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

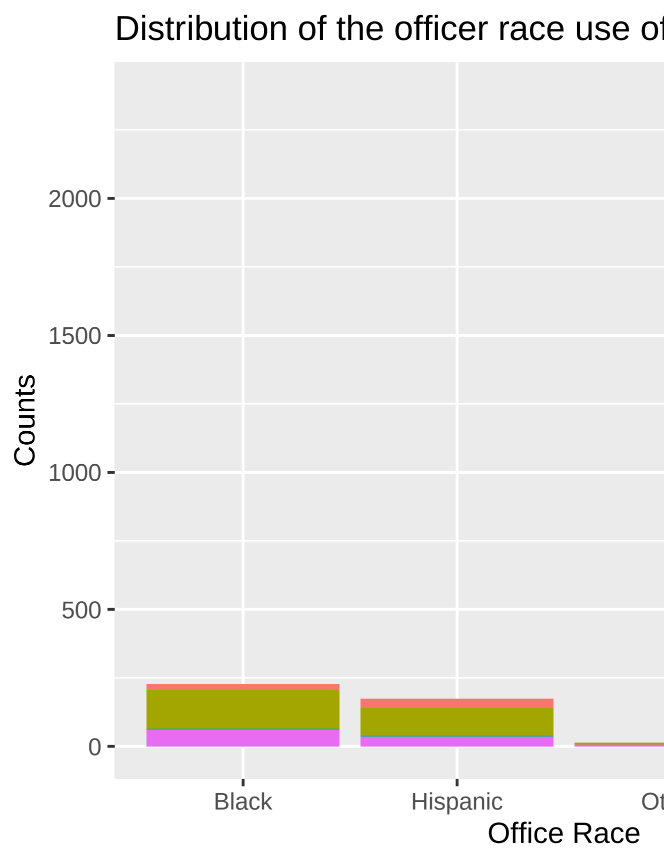
```
ind_dat$serious <- factor(ind_dat$serious)
```

```r
ggplot(data = ind_dat, mapping = aes(x = RACE)) +
  geom_bar() +
  labs(title = "Counts of incidents by subject race",
       x = "Subject Race",
       y = "Counts")
```



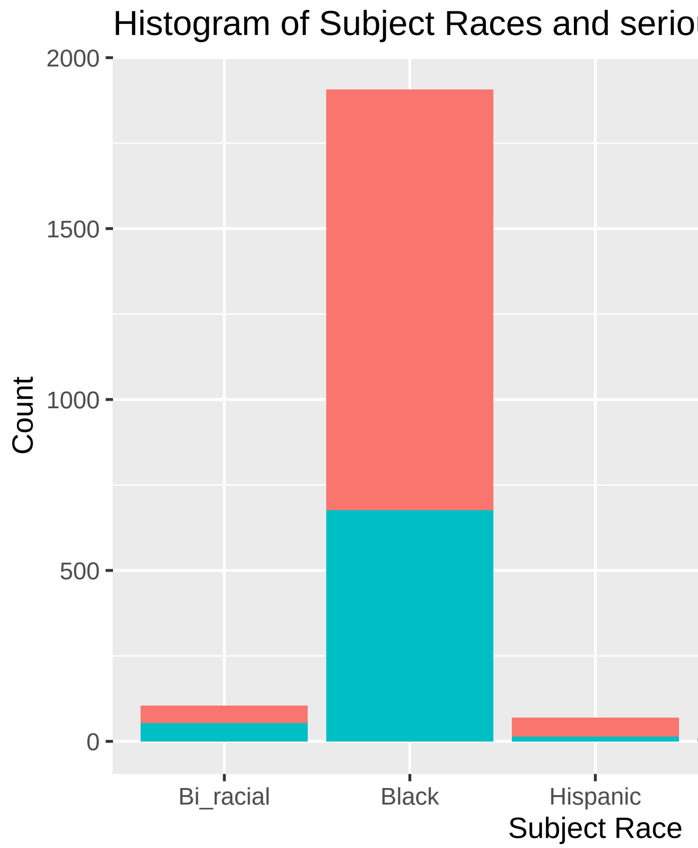
Counts of incidents by subject race
Based on the plot, the Blacks are experiencing the most UOF where over 10,000 incidents happened and followed by the White with counts less than 6,000.

```r
ggplot(data = ind_dat, mapping = aes(x = OFF_RACE, fill = RACE)) +
  geom_bar() +
  labs(title = "Distribution of the officer race use of force frequency",
       x = "Office Race",
       y = "Counts")
```



The Distribution of the officer race use of force frequency.
In this case, the White officers are using the most use of force in an incident with highest number of Black being targeted.

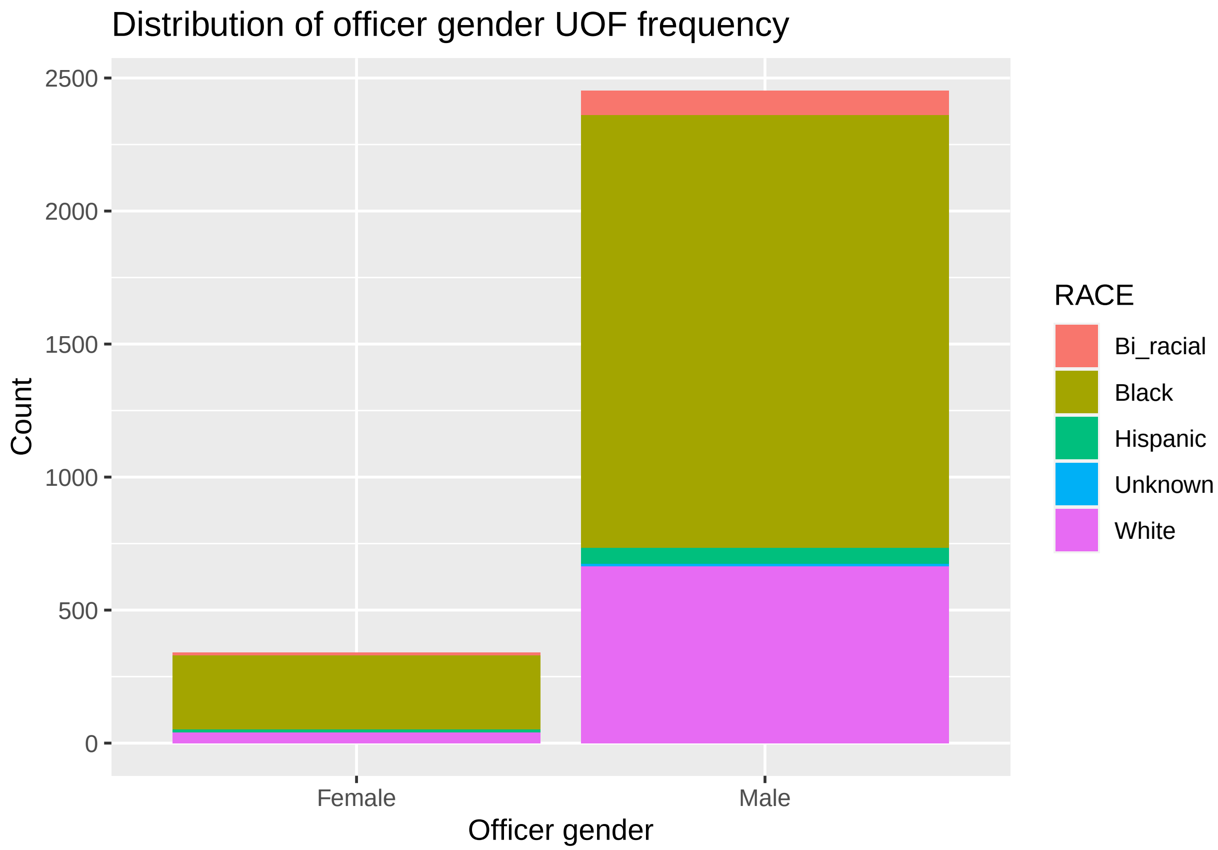
```
ggplot(data = ind_dat, mapping = aes(x = RACE, fill = serious)) +
 geom_bar() +
 labs(title = "Histogram of Subject Races and serious incidents", x = "Subject Race", y = "Count")
```



The counts of serious incidents relative to subject race.
Based on the plot, Black experiencing use of force occurred the most, over 10,000 counts. White experiencing almost less than half of the incidents of the Blacks.

```
ggplot(data = ind_dat, mapping = aes(x = SEX, fill = RACE)) +
 geom_bar() +
 labs(title = "Distribution of officer gender UOF frequency", x = "Officer gender", y = "Count")
```

The counts of officer gender use of force frequency in relation to the subject race.



This plots shows the distribution of the counts of the use of force cases among male and female officers. From the chart, it appears that there are significantly more incidents involving males officer than females officer.

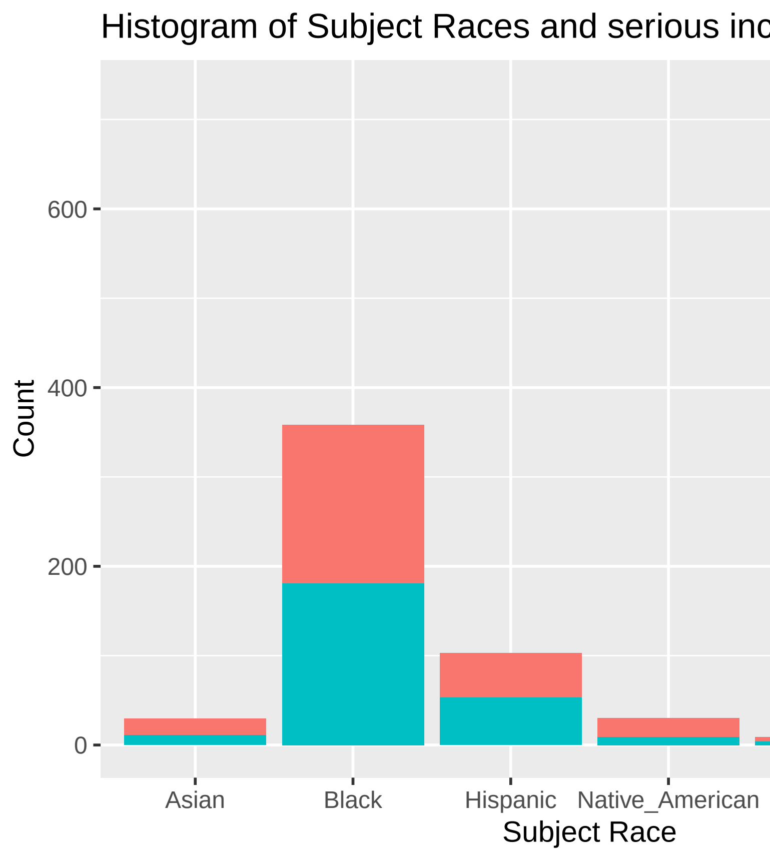
Portland, Oregon

Portland's population is made up of 635,067 of people where 73.8% are White, 5.6% are Black or African American, 8.5% are Asian. Overall, the Whites are experiencing the highest number of incidents in Portland 2020.

```
port_dat <- read_csv("../data/pdx_dat_2020_final.csv")
```

```
## Rows: 1260 Columns: 75
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (41): Blunt_Object, Category_of_Force_Event__Measured_at_Event_Level, Di...
## dbl (34): Officer, Officer_Tenure, Record_ID, Subject, Subject_Age, Subject_...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
port_dat$serious <- factor(port_dat$serious)

ggplot(data = port_dat, mapping = aes(x = Subject__Race, fill = serious)) +
 geom_bar() +
 labs(title = "Histogram of Subject Races and serious incidents", x = "Subject Race", y = "Count")
```



Histogram of Subject Races and serious incidents

Based on the plot, the race with the highest number of incidents is White, with a notable number of both serious and not serious incidents. This might due to the reason that Portland's population have over 70% Whites with only 5.6% of Blacks or African American. For all races, there are more incidents that are not serious (0) than serious (1), as indicated by the size of the colored sections.

Data driven hypotheses:

Our data exploration shows that Black/African American community members across each city are subject to a disproportionate amount of serious force, as compared to white community members. We can also see that particular wards and districts in some cities see a disproprionate amount of use of force generally. Our previous hypotheses remain relevant. We will use the methods described above to further explore these relationships. Given the disproportionate representation of Black/African Americans in cases of use of force, we expect subject race to be highly predictive in the model. We also expect some locations and districts to be highly predictive, as well as officer race.

Results ### Reading in data

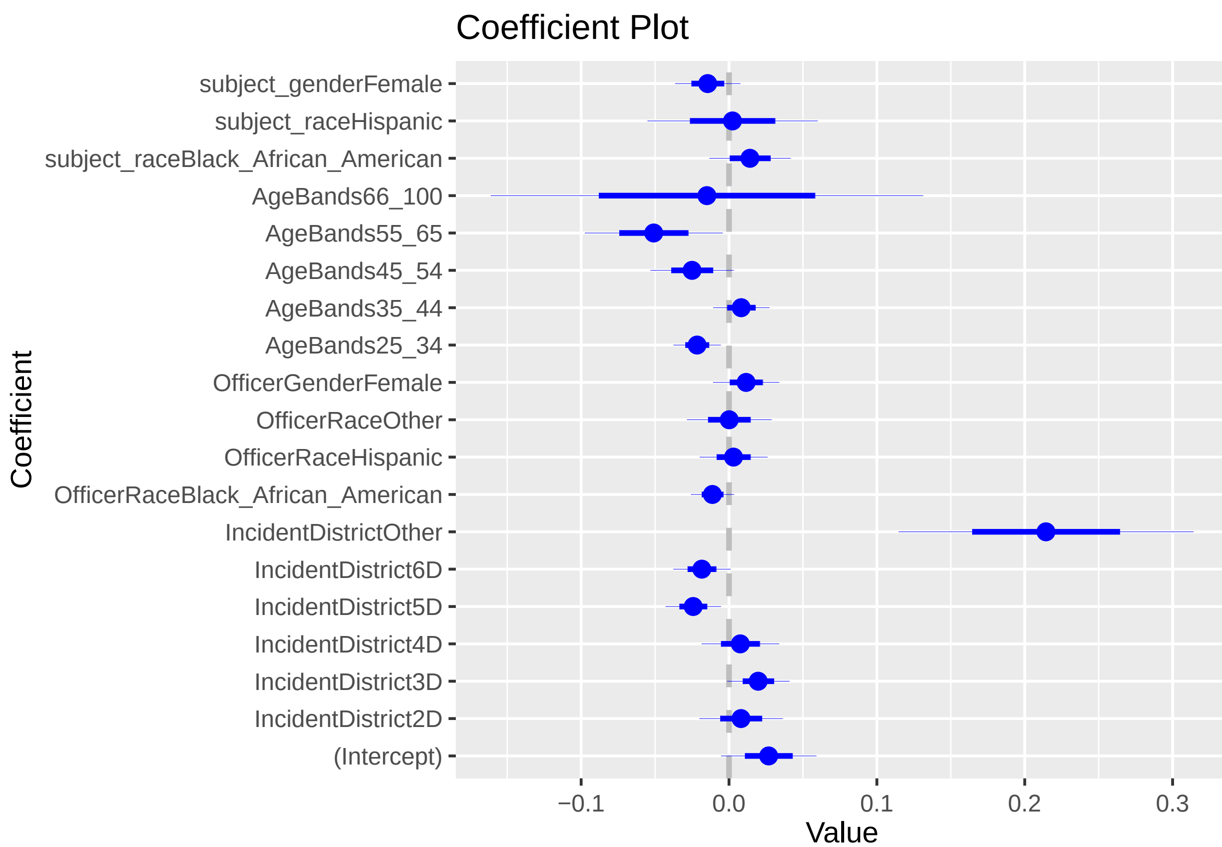
Setting up the data for DC

DC overall regression

```
model_mpd_2020 <- glm(serious ~ IncidentDistrict2D + IncidentDistrict3D + IncidentDistrict4D + Incident
               data = mpd_dat_enc_2020)

summary(model_mpd_2020)
```

```
##
## Call:
## glm(formula = serious ~ IncidentDistrict2D + IncidentDistrict3D +
##     IncidentDistrict4D + IncidentDistrict5D + IncidentDistrict6D +
##     IncidentDistrictOther + OfficerRaceBlack_African_American +
##     OfficerRaceHispanic + OfficerRaceOther + OfficerGenderFemale +
##     AgeBands25_34 + AgeBands35_44 + AgeBands45_54 + AgeBands55_65 +
##     AgeBands66_100 + subject_raceBlack_African_American + subject_raceHispanic +
##     subject_genderFemale, data = mpd_dat_enc_2020)
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.0268214  0.0160077   1.676  0.09399 .
## IncidentDistrict2D                  0.0081125  0.0140112   0.579  0.56265
## IncidentDistrict3D                  0.0197266  0.0104725   1.884  0.05976 .
## IncidentDistrict4D                  0.0076422  0.0130105   0.587  0.55701
## IncidentDistrict5D                 -0.0242273  0.0092516  -2.619  0.00889 **
## IncidentDistrict6D                 -0.0183856  0.0095872  -1.918  0.05529 .
## IncidentDistrictOther               0.2143488  0.0498187   4.303 1.77e-05 ***
## OfficerRaceBlack_African_American  -0.0112253  0.0072042  -1.558  0.11935
## OfficerRaceHispanic                 0.0030355  0.0113752   0.267  0.78961
## OfficerRaceOther                    0.0001299  0.0142572   0.009  0.99273
## OfficerGenderFemale                 0.0115627  0.0110954   1.042  0.29748
## AgeBands25_34                      -0.0215921  0.0078973  -2.734  0.00631 **
## AgeBands35_44                       0.0082553  0.0094316   0.875  0.38153
## AgeBands45_54                      -0.0250014  0.0140251  -1.783  0.07480 .
## AgeBands55_65                      -0.0509218  0.0232159  -2.193  0.02839 *
## AgeBands66_100                     -0.0149708  0.0730176  -0.205  0.83757
## subject_raceBlack_African_American  0.0141636  0.0136904   1.035  0.30100
## subject_raceHispanic                0.0023417  0.0286974   0.082  0.93497
## subject_genderFemale               -0.0144387  0.0109316  -1.321  0.18671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02095595)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 41.723  on 1991  degrees of freedom
## AIC: -2044.3
##
## Number of Fisher Scoring iterations: 2
```

```r
coefplot(model_mpd_2020)
```



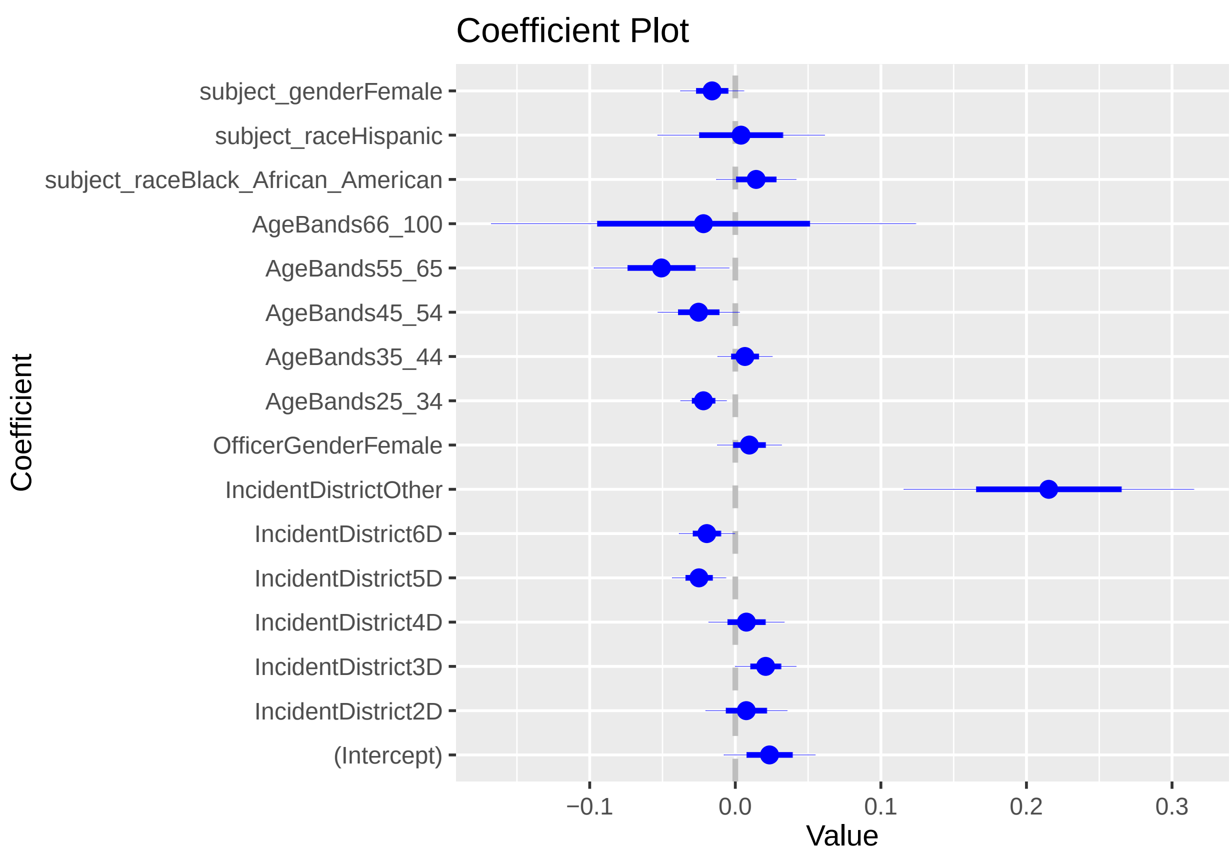
DC specific officer/subject relationships

```r
mpd_dat_blackoff <- mpd_dat_enc_2020 %>%
  filter(OfficerRace == "Black_African_American")

model_mpd_blackoff_2020 <- glm(serious ~ IncidentDistrict2D + IncidentDistrict3D + IncidentDistrict4D +
               data = mpd_dat_enc_2020)

summary(model_mpd_blackoff_2020)
```

```
##
## Call:
## glm(formula = serious ~ IncidentDistrict2D + IncidentDistrict3D +
##     IncidentDistrict4D + IncidentDistrict5D + IncidentDistrict6D +
##     IncidentDistrictOther + OfficerGenderFemale + AgeBands25_34 +
##     AgeBands35_44 + AgeBands45_54 + AgeBands55_65 + AgeBands66_100 +
##     subject_raceBlack_African_American + subject_raceHispanic +
##     subject_genderFemale, data = mpd_dat_enc_2020)
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          0.023493   0.015680   1.498  0.13421
## IncidentDistrict2D                   0.007539   0.013999   0.539  0.59026
## IncidentDistrict3D                   0.020828   0.010444   1.994  0.04626 *
```

```
## IncidentDistrict4D                     0.007617   0.012978   0.587  0.55735
## IncidentDistrict5D                    -0.024973   0.009222  -2.708  0.00683 **
## IncidentDistrict6D                    -0.019578   0.009558  -2.048  0.04066 *
## IncidentDistrictOther                  0.215305   0.049806   4.323 1.62e-05 ***
## OfficerGenderFemale                    0.009651   0.011021   0.876  0.38130
## AgeBands25_34                         -0.021905   0.007894  -2.775  0.00557 **
## AgeBands35_44                          0.006606   0.009381   0.704  0.48140
## AgeBands45_54                         -0.025221   0.014024  -1.798  0.07226 .
## AgeBands55_65                         -0.050748   0.023194  -2.188  0.02879 *
## AgeBands66_100                        -0.021889   0.072907  -0.300  0.76403
## subject_raceBlack_African_American     0.014312   0.013685   1.046  0.29577
## subject_raceHispanic                   0.003933   0.028662   0.137  0.89087
## subject_genderFemale                  -0.015959   0.010896  -1.465  0.14316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02095705)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 41.788  on 1994  degrees of freedom
## AIC: -2047.1
##
## Number of Fisher Scoring iterations: 2
```

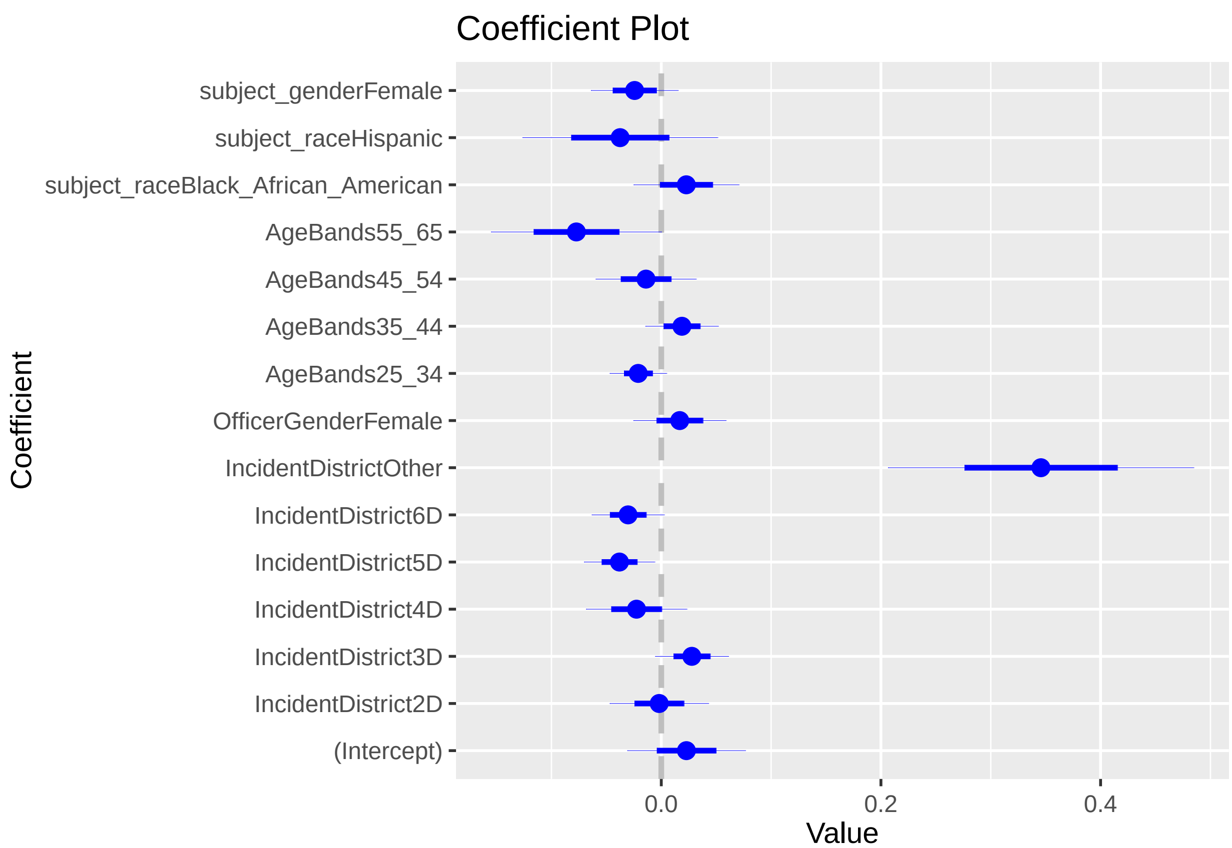
```
coefplot(model_mpd_blackoff_2020)
```



```r
mpd_dat_whiteoff <- mpd_dat_enc_2020 %>%
  filter(OfficerRace == "White_Caucasian")

model_mpd_whiteoff_2020 <- glm(serious ~ IncidentDistrict2D + IncidentDistrict3D + IncidentDistrict4D +
               data = mpd_dat_whiteoff)

summary(model_mpd_whiteoff_2020)
```

```
##
## Call:
## glm(formula = serious ~ IncidentDistrict2D + IncidentDistrict3D +
##     IncidentDistrict4D + IncidentDistrict5D + IncidentDistrict6D +
##     IncidentDistrictOther + OfficerGenderFemale + AgeBands25_34 +
##     AgeBands35_44 + AgeBands45_54 + AgeBands55_65 + AgeBands66_100 +
##     subject_raceBlack_African_American + subject_raceHispanic +
##     subject_genderFemale, data = mpd_dat_whiteoff)
##
## Coefficients: (1 not defined because of singularities)
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.022882   0.026886   0.851   0.3950
## IncidentDistrict2D                 -0.001909   0.022485  -0.085   0.9324
## IncidentDistrict3D                  0.027805   0.016692   1.666   0.0961 .
## IncidentDistrict4D                 -0.022542   0.022920  -0.984   0.3256
## IncidentDistrict5D                 -0.038091   0.016097  -2.366   0.0182 *
## IncidentDistrict6D                 -0.030231   0.016502  -1.832   0.0673 .
## IncidentDistrictOther               0.345674   0.069574   4.968 8.15e-07 ***
## OfficerGenderFemale                 0.016784   0.021076   0.796   0.4260
## AgeBands25_34                      -0.020994   0.012945  -1.622   0.1052
## AgeBands35_44                       0.018787   0.016588   1.133   0.2577
## AgeBands45_54                      -0.013932   0.022880  -0.609   0.5427
## AgeBands55_65                      -0.077323   0.038789  -1.993   0.0465 *
## AgeBands66_100                            NA         NA      NA       NA
## subject_raceBlack_African_American  0.022781   0.023960   0.951   0.3420
## subject_raceHispanic               -0.037481   0.044455  -0.843   0.3994
## subject_genderFemale               -0.024280   0.019837  -1.224   0.2213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02564137)
##
##     Null deviance: 23.341  on 873  degrees of freedom
## Residual deviance: 22.026  on 859  degrees of freedom
## AIC: -704.77
##
## Number of Fisher Scoring iterations: 2
```

```r
coefplot(model_mpd_whiteoff_2020)
```

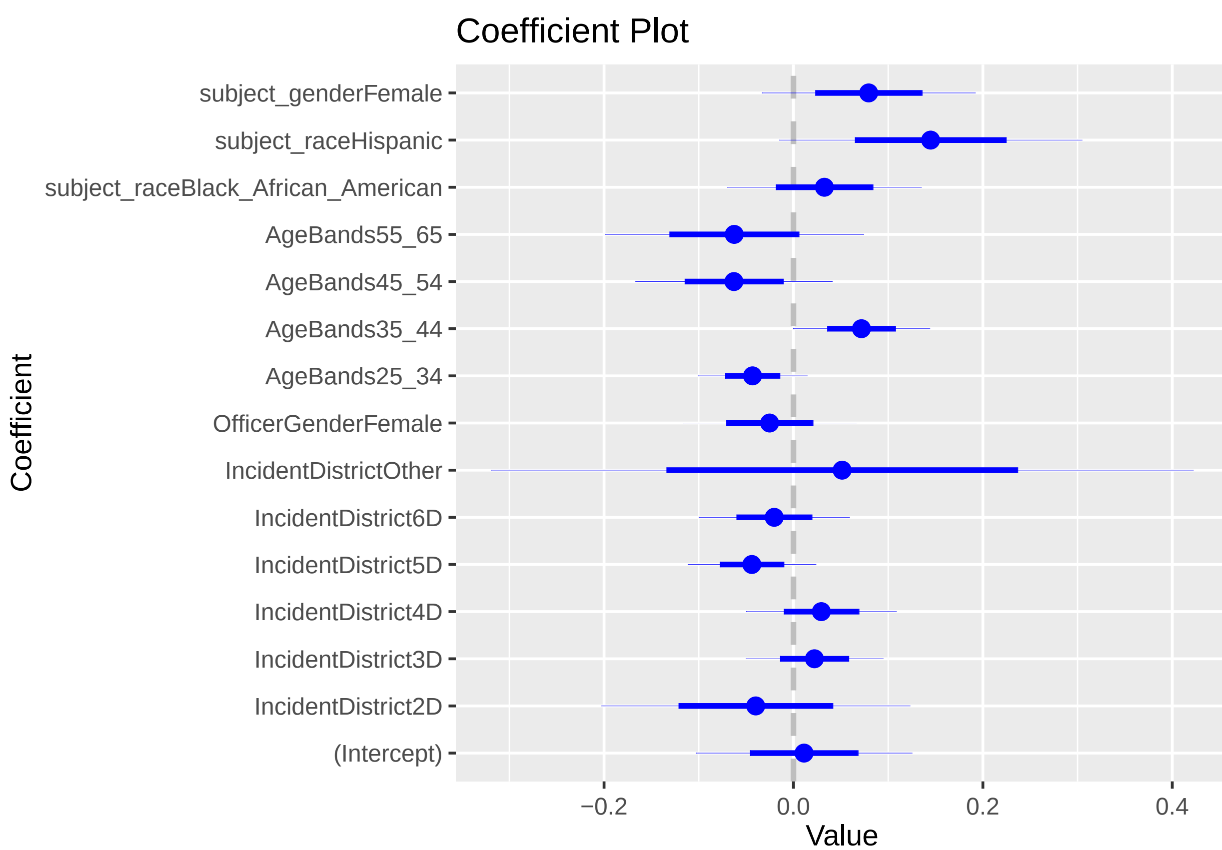



```
mpd_dat_hispanicoff <- mpd_dat_enc_2020 %>%
  filter(OfficerRace == "Hispanic")

model_dat_hispanicoff_2020 <- glm(serious ~ IncidentDistrict2D + IncidentDistrict3D + IncidentDistrict4D
              data = mpd_dat_hispanicoff)

summary(model_dat_hispanicoff_2020)
```

```
##
## Call:
## glm(formula = serious ~ IncidentDistrict2D + IncidentDistrict3D +
##     IncidentDistrict4D + IncidentDistrict5D + IncidentDistrict6D +
##     IncidentDistrictOther + OfficerGenderFemale + AgeBands25_34 +
##     AgeBands35_44 + AgeBands45_54 + AgeBands55_65 + AgeBands66_100 +
##     subject_raceBlack_African_American + subject_raceHispanic +
##     subject_genderFemale, data = mpd_dat_hispanicoff)
##
## Coefficients: (1 not defined because of singularities)
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.01117    0.05697   0.196   0.8447
## IncidentDistrict2D                 -0.03986    0.08142  -0.490   0.6250
## IncidentDistrict3D                  0.02214    0.03622   0.611   0.5417
## IncidentDistrict4D                  0.02943    0.03967   0.742   0.4590
## IncidentDistrict5D                 -0.04394    0.03378  -1.301   0.1949
## IncidentDistrict6D                 -0.02032    0.03979  -0.511   0.6101
```

```
## IncidentDistrictOther                0.05137    0.18540   0.277   0.7820
## OfficerGenderFemale                 -0.02518    0.04573  -0.551   0.5826
## AgeBands25_34                       -0.04322    0.02880  -1.500   0.1352
## AgeBands35_44                        0.07180    0.03604   1.992   0.0478 *
## AgeBands45_54                       -0.06285    0.05201  -1.208   0.2284
## AgeBands55_65                       -0.06254    0.06833  -0.915   0.3612
## AgeBands66_100                            NA         NA      NA       NA
## subject_raceBlack_African_American   0.03266    0.05123   0.637   0.5246
## subject_raceHispanic                 0.14483    0.07988   1.813   0.0714 .
## subject_genderFemale                 0.07943    0.05628   1.411   0.1598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02756711)
##
##     Null deviance: 5.8218  on 201  degrees of freedom
## Residual deviance: 5.1551  on 187  degrees of freedom
## AIC: -135.74
##
## Number of Fisher Scoring iterations: 2
```

```
coefplot(model_dat_hispanicoff_2020)
```



DC counts

```r
mpd_dat_whiteoff %>%
  filter(subject_race == "White_Caucasian") %>%
  nrow()
```

```
## [1] 55
```

```r
mpd_dat_whiteoff %>%
  filter(subject_race == "Black_African_American") %>%
  nrow()
```

```
## [1] 800
```

```r
mpd_dat_whiteoff %>%
  filter(subject_race == "Asian_Pacific_Islander") %>%
  nrow()
```

```
## [1] 2
```

```r
mpd_dat_whiteoff %>%
  filter(subject_race == "Hispanic") %>%
  nrow()
```

```
## [1] 17
```

```r
mpd_dat_blackoff %>%
  filter(subject_race == "White_Caucasian") %>%
  nrow()
```

```
## [1] 50
```

```r
mpd_dat_blackoff %>%
  filter(subject_race == "Black_African_American") %>%
  nrow()
```

```
## [1] 757
```

```r
mpd_dat_blackoff %>%
  filter(subject_race == "Asian_Pacific_Islander") %>%
  nrow()
```

```
## [1] 2
```

```r
mpd_dat_blackoff %>%
  filter(subject_race == "Hispanic") %>%
  nrow()
```

```
## [1] 7
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "White_Caucasian") %>%
  nrow()
```

```
## [1] 14
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "Black_African_American") %>%
  nrow()
```

```
## [1] 181
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "Asian_Pacific_Islander") %>%
  nrow()
```

```
## [1] 0
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "Hispanic") %>%
  nrow()
```

```
## [1] 7
```


```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "White_Caucasian") %>%
  nrow()
```

```
## [1] 14
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "Black_African_American") %>%
  nrow()
```

```
## [1] 181
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "Asian_Pacific_Islander") %>%
  nrow()
```

```
## [1] 0
```

```r
mpd_dat_hispanicoff %>%
  filter(subject_race == "Hispanic") %>%
  nrow()
```

```
## [1] 7
```

```r
matrix_data_ind <- matrix(c(
  55, 50, 14,
  800, 757, 181,
  2, 2, 0,
  17, 7, 7
), ncol = 3, byrow = TRUE)

row_names_ind <- c("White subject", "Black subject", "Hispanic subject", "Asian subject")
col_names_ind <- c("White officer", "Black officer", "Hispanic officer")

data_df_ind <- as.data.frame(matrix_data_ind)
rownames(data_df_ind) <- row_names_ind
colnames(data_df_ind) <- col_names_ind

print(data_df_ind)
```

```
##                  White officer Black officer Hispanic officer
## White subject               55            50               14
## Black subject              800           757              181
## Hispanic subject             2             2                0
## Asian subject               17             7                7
```

Seattle

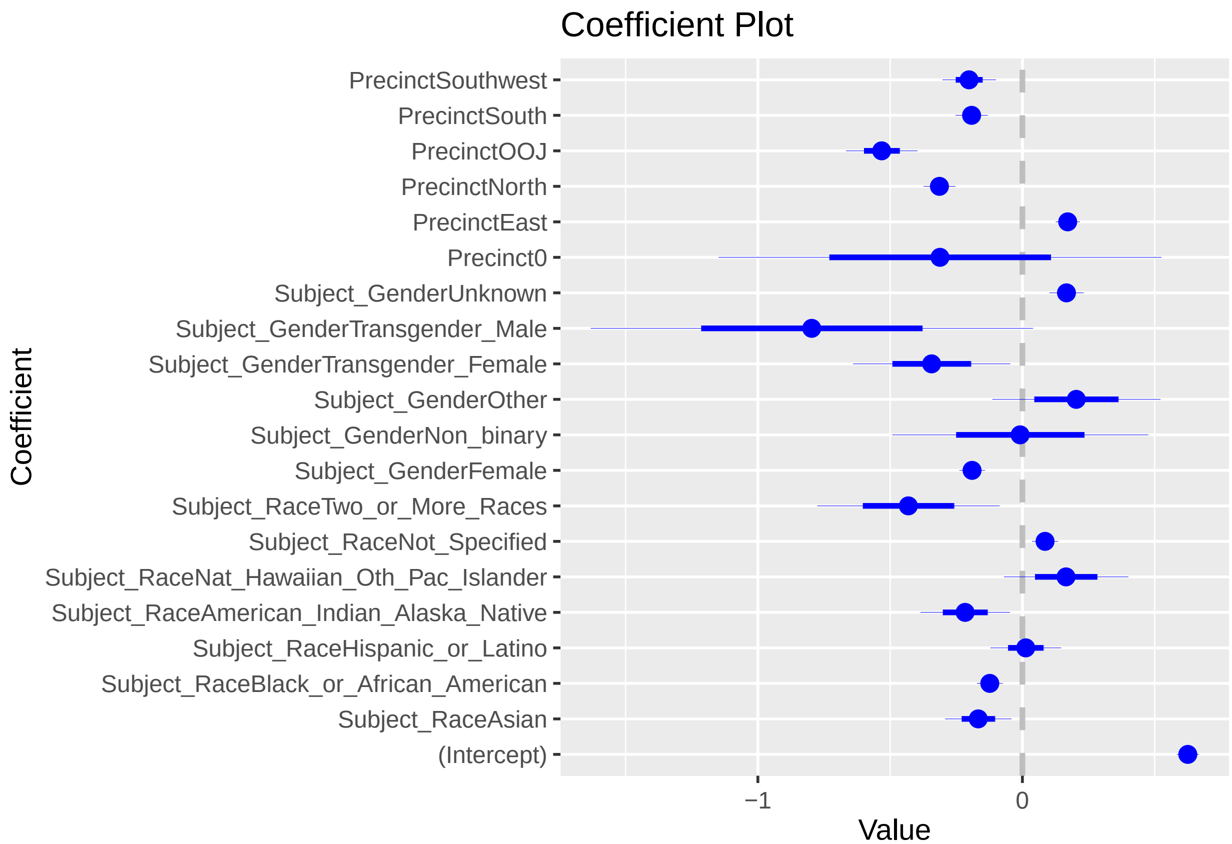
```r
sea_dat_2020 <- sea_dat_2020 %>%
  mutate(datetime = mdy_hms(Occured_date_time)) %>%
  filter(year(datetime) == 2020) %>%
  drop_na(Incident_Type, Subject_Race, Subject_Gender, Precinct, Sector, Beat) %>%
  mutate(serious = ifelse(Incident_Type == "Level 1 - Use of Force", 0, 1)) %>%
  mutate(Subject_Race = str_replace_all(Subject_Race, "/", "_")) %>%
  mutate(Subject_Race = str_replace_all(Subject_Race, " ", "_")) %>%
  mutate(Subject_Gender = str_replace_all(Subject_Gender, " ", "_")) %>%
  mutate(Subject_Gender = str_replace_all(Subject_Gender, "/", "_")) %>%
  mutate(Subject_Gender = str_replace_all(Subject_Gender, "-", "_"))

sea_dat_2020$Subject_Race <- factor(sea_dat_2020$Subject_Race)
sea_dat_2020$Subject_Race <- relevel(sea_dat_2020$Subject_Race, ref = "White")
sea_dat_2020$Subject_Gender <- factor(sea_dat_2020$Subject_Gender)
sea_dat_2020$Subject_Gender <- relevel(sea_dat_2020$Subject_Gender, ref = "Male")
sea_dat_2020$Precinct <- factor(sea_dat_2020$Precinct)
sea_dat_2020$Precinct <- relevel(sea_dat_2020$Precinct, ref = "West")

encoded_seattle_2020 <- model.matrix(~ Subject_Race + Subject_Gender + Precinct, data = sea_dat_2020)

sea_dat_2020_final <- cbind(sea_dat_2020, encoded_seattle_2020)


model_seattle_2020 <- glm(serious ~ Subject_RaceAsian + Subject_RaceBlack_or_African_American + Subject_

summary(model_seattle_2020)
```

```
##
```

```
## Call:
## glm(formula = serious ~ Subject_RaceAsian + Subject_RaceBlack_or_African_American +
##     Subject_RaceHispanic_or_Latino + Subject_RaceAmerican_Indian_Alaska_Native +
##     Subject_RaceNat_Hawaiian_Oth_Pac_Islander + Subject_RaceNot_Specified +
##     Subject_RaceTwo_or_More_Races + Subject_GenderFemale + Subject_GenderNon_binary +
##     Subject_GenderOther + Subject_GenderTransgender_Female +
##     Subject_GenderTransgender_Male + Subject_GenderUnknown +
##     Precinct0 + PrecinctEast + PrecinctNorth + PrecinctOOJ +
##     PrecinctSouth + PrecinctSouthwest, data = sea_dat_2020_final)
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                0.625230   0.020224  30.915  < 2e-16
## Subject_RaceAsian                         -0.166947   0.062229  -2.683 0.007358
## Subject_RaceBlack_or_African_American     -0.123328   0.023826  -5.176 2.48e-07
## Subject_RaceHispanic_or_Latino             0.012506   0.066177   0.189 0.850127
## Subject_RaceAmerican_Indian_Alaska_Native -0.216579   0.083953  -2.580 0.009953
## Subject_RaceNat_Hawaiian_Oth_Pac_Islander  0.164827   0.116898   1.410 0.158684
## Subject_RaceNot_Specified                  0.085625   0.024136   3.548 0.000397
## Subject_RaceTwo_or_More_Races             -0.431160   0.171830  -2.509 0.012173
## Subject_GenderFemale                      -0.190272   0.023635  -8.050 1.35e-15
## Subject_GenderNon_binary                  -0.008613   0.241461  -0.036 0.971548
## Subject_GenderOther                        0.203586   0.158326   1.286 0.198629
## Subject_GenderTransgender_Female          -0.343263   0.148017  -2.319 0.020485
## Subject_GenderTransgender_Male            -0.796388   0.417466  -1.908 0.056567
## Subject_GenderUnknown                      0.166532   0.031909   5.219 1.97e-07
## Precinct0                                 -0.311631   0.417933  -0.746 0.455962
## PrecinctEast                               0.171159   0.022341   7.661 2.77e-14
## PrecinctNorth                             -0.313694   0.029514 -10.629  < 2e-16
## PrecinctOOJ                               -0.531919   0.066854  -7.956 2.84e-15
## PrecinctSouth                             -0.192167   0.029932  -6.420 1.67e-10
## PrecinctSouthwest                         -0.201666   0.049937  -4.038 5.57e-05
##
## (Intercept)                               ***
## Subject_RaceAsian                         **
## Subject_RaceBlack_or_African_American     ***
## Subject_RaceHispanic_or_Latino
## Subject_RaceAmerican_Indian_Alaska_Native **
## Subject_RaceNat_Hawaiian_Oth_Pac_Islander
## Subject_RaceNot_Specified                 ***
## Subject_RaceTwo_or_More_Races             *
## Subject_GenderFemale                      ***
## Subject_GenderNon_binary
## Subject_GenderOther
## Subject_GenderTransgender_Female          *
## Subject_GenderTransgender_Male            .
## Subject_GenderUnknown                     ***
## Precinct0
## PrecinctEast                              ***
## PrecinctNorth                             ***
## PrecinctOOJ                               ***
## PrecinctSouth                             ***
## PrecinctSouthwest                         ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1738626)
##
##     Null deviance: 527.73  on 2160  degrees of freedom
## Residual deviance: 372.24  on 2141  degrees of freedom
## AIC: 2373.9
##
## Number of Fisher Scoring iterations: 2
```

```
coefplot(model_seattle_2020)
```




```
model_seattle_subrace <- glm(serious ~ Subject_RaceAsian + Subject_RaceBlack_or_African_American + Subje
summary(model_seattle_subrace)
```

```
##
## Call:
## glm(formula = serious ~ Subject_RaceAsian + Subject_RaceBlack_or_African_American +
##     Subject_RaceHispanic_or_Latino + Subject_RaceAmerican_Indian_Alaska_Native +
##     Subject_RaceNat_Hawaiian_Oth_Pac_Islander + Subject_RaceNot_Specified +
##     Subject_RaceTwo_or_More_Races, data = sea_dat_2020_final)
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                 0.55263    0.01606  34.415  < 2e-16
```

```
## Subject_RaceAsian                        -0.24013    0.06891  -3.485 0.000503
## Subject_RaceBlack_or_African_American    -0.17318    0.02615  -6.622 4.45e-11
## Subject_RaceHispanic_or_Latino           -0.13403    0.07260  -1.846 0.065025
## Subject_RaceAmerican_Indian_Alaska_Native -0.36032    0.09246  -3.897 0.000100
## Subject_RaceNat_Hawaiian_Oth_Pac_Islander  0.13968    0.12977   1.076 0.281899
## Subject_RaceNot_Specified                 0.24385    0.02395  10.183  < 2e-16
## Subject_RaceTwo_or_More_Races            -0.55263    0.19023  -2.905 0.003709
##
## (Intercept)                              ***
## Subject_RaceAsian                        ***
## Subject_RaceBlack_or_African_American    ***
## Subject_RaceHispanic_or_Latino           .
## Subject_RaceAmerican_Indian_Alaska_Native ***
## Subject_RaceNat_Hawaiian_Oth_Pac_Islander
## Subject_RaceNot_Specified                ***
## Subject_RaceTwo_or_More_Races            **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2155724)
##
##     Null deviance: 527.73  on 2160  degrees of freedom
## Residual deviance: 464.13  on 2153  degrees of freedom
## AIC: 2826.7
##
## Number of Fisher Scoring iterations: 2
```

```
model_seattle_subgender <- glm(serious ~ Subject_GenderFemale + Subject_GenderNon_binary + Subject_Gende
summary(model_seattle_subgender)
```

```
##
## Call:
## glm(formula = serious ~ Subject_GenderFemale + Subject_GenderNon_binary +
##     Subject_GenderOther + Subject_GenderTransgender_Female +
##     Subject_GenderTransgender_Male + Subject_GenderUnknown, data = sea_dat_2020_final)
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       0.57192    0.01217  46.988   <2e-16 ***
## Subject_GenderFemale             -0.21615    0.02590  -8.347   <2e-16 ***
## Subject_GenderNon_binary          0.09475    0.26943   0.352   0.7251
## Subject_GenderOther               0.42808    0.17662   2.424   0.0154 *
## Subject_GenderTransgender_Female -0.32192    0.16527  -1.948   0.0516 .
## Subject_GenderTransgender_Male   -0.57192    0.46634  -1.226   0.2202
## Subject_GenderUnknown             0.38175    0.03142  12.150   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2173258)
##
##     Null deviance: 527.73  on 2160  degrees of freedom
## Residual deviance: 468.12  on 2154  degrees of freedom
## AIC: 2843.2
##
```

```
## Number of Fisher Scoring iterations: 2
```

```
model_seattle_precinct <- glm(serious ~ Precinct0 + PrecinctEast + PrecinctNorth + PrecinctOOJ + Precin
summary(model_seattle_precinct)
```

```
##
## Call:
## glm(formula = serious ~ Precinct0 + PrecinctEast + PrecinctNorth +
##     PrecinctOOJ + PrecinctSouth + PrecinctSouthwest, data = sea_dat_2020_final)
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.57967    0.01636  35.422  < 2e-16 ***
## Precinct0         -0.57967    0.44185  -1.312     0.19
## PrecinctEast       0.24814    0.02294  10.819  < 2e-16 ***
## PrecinctNorth     -0.33499    0.03097 -10.816  < 2e-16 ***
## PrecinctOOJ       -0.53205    0.07007  -7.593 4.62e-14 ***
## PrecinctSouth     -0.20238    0.03134  -6.458 1.30e-10 ***
## PrecinctSouthwest -0.27967    0.05201  -5.377 8.37e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1949656)
##
##     Null deviance: 527.73  on 2160  degrees of freedom
## Residual deviance: 419.96  on 2154  degrees of freedom
## AIC: 2608.6
##
## Number of Fisher Scoring iterations: 2
```

Portland

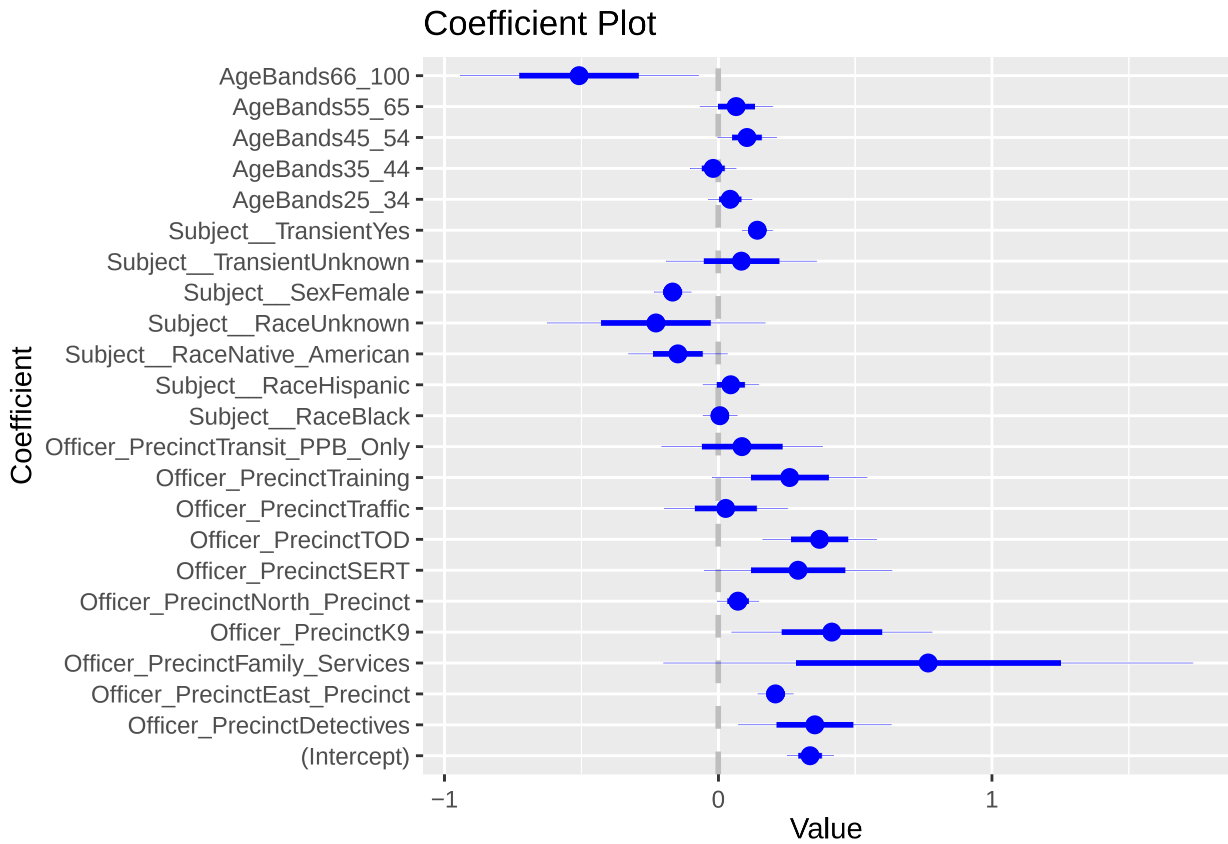
```
pdx_dat_2020 <- pdx_dat_2020 %>%
  filter(Year == 2020) %>%
  mutate_all(~str_replace_all(., "/", "_")) %>%
  mutate_all(~str_replace_all(., " ", "_")) %>%
  mutate_all(~str_replace_all(., "-", "_")) %>%
  mutate_all(~str_replace_all(., "\\(|\\)", "")) %>%
  rename_all(~str_replace_all(., " ", "_")) %>%
  rename_all(~str_replace_all(., "-", "")) %>%
  rename_all(~str_replace_all(., "/", "_")) %>%
  mutate(Subject_Age = as.numeric(Subject_Age),
    AgeBands = cut(Subject_Age, breaks = age_breaks, labels = c("18_24", "25_34", "35_44", "45_54", "55_
                   include.lowest = TRUE, right = FALSE)) %>%
  drop_na(Category_of_Force_Event__Measured_at_Event_Level, Officer_Precinct, Subject__Race, Subject__Se
  mutate(serious = ifelse(Category_of_Force_Event__Measured_at_Event_Level == "IV", 0, 1))

pdx_dat_2020$Officer_Precinct <- factor(pdx_dat_2020$Officer_Precinct)
pdx_dat_2020$Officer_Precinct <- relevel(pdx_dat_2020$Officer_Precinct, ref = "Central_Precinct")
pdx_dat_2020$Subject__Race <- factor(pdx_dat_2020$Subject__Race)
pdx_dat_2020$Subject__Race <- relevel(pdx_dat_2020$Subject__Race, ref = "White")
pdx_dat_2020$Subject__Sex <- factor(pdx_dat_2020$Subject__Sex)
pdx_dat_2020$Subject__Sex <- relevel(pdx_dat_2020$Subject__Sex, ref = "Male")
```

```r
pdx_dat_2020$AgeBands <- factor(pdx_dat_2020$AgeBands)
pdx_dat_2020$AgeBands <- relevel(pdx_dat_2020$AgeBands, ref = "18_24")

encoded_pdx_2020 <- model.matrix(~ Officer_Precinct + Subject__Race + Subject__Sex + Subject__Transient

pdx_dat_2020_final <- cbind(pdx_dat_2020, encoded_pdx_2020)

model_portland_2020 <- glm(serious ~ Officer_PrecinctDetectives + Officer_PrecinctEast_Precinct + Offic

summary(model_portland_2020)
```

```
##
## Call:
## glm(formula = serious ~ Officer_PrecinctDetectives + Officer_PrecinctEast_Precinct +
##     Officer_PrecinctFamily_Services + Officer_PrecinctK9 + Officer_PrecinctNorth_Precinct +
##     Officer_PrecinctSERT + Officer_PrecinctTOD + Officer_PrecinctTraffic +
##     Officer_PrecinctTraining + Officer_PrecinctTransit_PPB_Only +
##     Subject__RaceBlack + Subject__RaceHispanic + Subject__RaceNative_American +
##     Subject__RaceUnknown + Subject__SexFemale + Subject__TransientUnknown +
##     Subject__TransientYes + AgeBands25_34 + AgeBands35_44 + AgeBands45_54 +
##     AgeBands55_65 + AgeBands66_100, data = pdx_dat_2020_final)
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       0.335347   0.042389   7.911 5.62e-15 ***
## Officer_PrecinctDetectives        0.352590   0.139463   2.528 0.011588 *
## Officer_PrecinctEast_Precinct     0.208699   0.032361   6.449 1.61e-10 ***
## Officer_PrecinctFamily_Services   0.766666   0.483352   1.586 0.112962
## Officer_PrecinctK9                0.414573   0.182990   2.266 0.023651 *
## Officer_PrecinctNorth_Precinct    0.071772   0.038102   1.884 0.059843 .
## Officer_PrecinctSERT              0.291294   0.171438   1.699 0.089548 .
## Officer_PrecinctTOD               0.369611   0.103819   3.560 0.000385 ***
## Officer_PrecinctTraffic           0.026949   0.113118   0.238 0.811738
## Officer_PrecinctTraining          0.260655   0.141169   1.846 0.065072 .
## Officer_PrecinctTransit_PPB_Only  0.086352   0.146828   0.588 0.556562
## Subject__RaceBlack                0.005783   0.031373   0.184 0.853786
## Subject__RaceHispanic             0.045301   0.051158   0.886 0.376049
## Subject__RaceNative_American     -0.147907   0.090108  -1.641 0.100962
## Subject__RaceUnknown             -0.227911   0.199360  -1.143 0.253172
## Subject__SexFemale               -0.167067   0.033736  -4.952 8.36e-07 ***
## Subject__TransientUnknown         0.084212   0.137525   0.612 0.540428
## Subject__TransientYes             0.142413   0.028070   5.073 4.50e-07 ***
## AgeBands25_34                     0.043266   0.039677   1.090 0.275727
## AgeBands35_44                    -0.018903   0.041728  -0.453 0.650631
## AgeBands45_54                     0.104659   0.053594   1.953 0.051066 .
## AgeBands55_65                     0.065054   0.066456   0.979 0.327826
## AgeBands66_100                   -0.508783   0.217695  -2.337 0.019591 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2290336)
##
##     Null deviance: 315.00  on 1259  degrees of freedom
```

```
## Residual deviance: 283.31  on 1237  degrees of freedom
## AIC: 1743.4
##
## Number of Fisher Scoring iterations: 2
```

```
coefplot(model_portland_2020)
```




```
model_portland_precinct <- glm(serious ~ Officer_PrecinctDetectives + Officer_PrecinctEast_Precinct + Of
summary(model_portland_precinct)
```

```
##
## Call:
## glm(formula = serious ~ Officer_PrecinctDetectives + Officer_PrecinctEast_Precinct +
##     Officer_PrecinctFamily_Services + Officer_PrecinctK9 + Officer_PrecinctNorth_Precinct +
##     Officer_PrecinctSERT + Officer_PrecinctTOD + Officer_PrecinctTraffic +
##     Officer_PrecinctTraining + Officer_PrecinctTransit_PPB_Only,
##     data = pdx_dat_2020_final)
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       0.39602    0.02311  17.135  < 2e-16 ***
## Officer_PrecinctDetectives        0.31827    0.13334   2.387 0.017140 *
## Officer_PrecinctEast_Precinct     0.19828    0.03261   6.080  1.6e-09 ***
## Officer_PrecinctFamily_Services   0.60398    0.49191   1.228 0.219738
## Officer_PrecinctK9                0.46113    0.18715   2.464 0.013876 *
```

```
## Officer_PrecinctNorth_Precinct    0.07091    0.03839   1.847 0.064958 .
## Officer_PrecinctSERT              0.22898    0.17525   1.307 0.191597
## Officer_PrecinctTOD               0.38659    0.10503   3.681 0.000242 ***
## Officer_PrecinctTraffic           0.02503    0.11507   0.218 0.827807
## Officer_PrecinctTraining          0.27065    0.14371   1.883 0.059900 .
## Officer_PrecinctTransit_PPB_Only  0.14944    0.14994   0.997 0.319141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2414369)
##
##     Null deviance: 315.00  on 1259  degrees of freedom
## Residual deviance: 301.55  on 1249  degrees of freedom
## AIC: 1798
##
## Number of Fisher Scoring iterations: 2
```

```r
model_portland_race <- glm(serious ~ Subject__RaceBlack + Subject__RaceHispanic + Subject__RaceNative_Ar
summary(model_portland_race)
```

```
##
## Call:
## glm(formula = serious ~ Subject__RaceBlack + Subject__RaceHispanic +
##     Subject__RaceNative_American + Subject__RaceUnknown, data = pdx_dat_2020_final)
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.5065789  0.0181356  27.933   <2e-16 ***
## Subject__RaceBlack           -0.0009924  0.0320488  -0.031   0.9753
## Subject__RaceHispanic         0.0079842  0.0524952   0.152   0.8791
## Subject__RaceNative_American -0.2065789  0.0930649  -2.220   0.0266 *
## Subject__RaceUnknown         -0.0621345  0.1676389  -0.371   0.7110
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2499652)
##
##     Null deviance: 315.00  on 1259  degrees of freedom
## Residual deviance: 313.71  on 1255  degrees of freedom
## AIC: 1835.8
##
## Number of Fisher Scoring iterations: 2
```

```r
model_portland_sex <- glm(serious ~ Subject__SexFemale, data = pdx_dat_2020_final)
summary(model_portland_sex)
```

```
##
## Call:
## glm(formula = serious ~ Subject__SexFemale, data = pdx_dat_2020_final)
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.54694    0.01575  34.720  < 2e-16 ***
```

```
## Subject__SexFemale -0.20408    0.03342  -6.107 1.35e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2431848)
##
##     Null deviance: 315.00  on 1259  degrees of freedom
## Residual deviance: 305.93  on 1258  degrees of freedom
## AIC: 1798.2
##
## Number of Fisher Scoring iterations: 2
```

```
model_portland_transient <- glm(serious ~ Subject__TransientUnknown + Subject__TransientYes, data = pdx
summary(model_portland_transient)
```

```
##
## Call:
## glm(formula = serious ~ Subject__TransientUnknown + Subject__TransientYes,
##     data = pdx_dat_2020_final)
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.43651    0.01977  22.074   <2e-16 ***
## Subject__TransientUnknown  0.11905    0.11865   1.003    0.316
## Subject__TransientYes      0.13049    0.02817   4.632    4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2463469)
##
##     Null deviance: 315.00  on 1259  degrees of freedom
## Residual deviance: 309.66  on 1257  degrees of freedom
## AIC: 1815.4
##
## Number of Fisher Scoring iterations: 2
```

```
model_portland_age <- glm(serious ~ AgeBands25_34 + AgeBands35_44 + AgeBands45_54 + AgeBands55_65 + Age
summary(model_portland_age)
```

```
##
## Call:
## glm(formula = serious ~ AgeBands25_34 + AgeBands35_44 + AgeBands45_54 +
##     AgeBands55_65 + AgeBands66_100, data = pdx_dat_2020_final)
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.43805    0.03312  13.228  < 2e-16 ***
## AgeBands25_34   0.09260    0.04026   2.300  0.02160 *
## AgeBands35_44   0.02937    0.04241   0.692  0.48877
## AgeBands45_54   0.15286    0.05454   2.803  0.00514 **
## AgeBands55_65   0.11124    0.06773   1.642  0.10075
## AgeBands66_100 -0.43805    0.22509  -1.946  0.05186 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.247847)
##
##     Null deviance: 315.0  on 1259  degrees of freedom
## Residual deviance: 310.8  on 1254  degrees of freedom
## AIC: 1826.1
##
## Number of Fisher Scoring iterations: 2
```

Indianapolis

```
ind_dat_2020 <- read_csv("../data/UOF_Indianapolis.csv")
```

```
## Rows: 63151 Columns: 35
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (26): OCCURRED_DT, UDTEXT24A, UDTEXT24B, UDTEXT24C, UDTEXT24D, DISPOSIT...
## dbl   (7): OBJECTID, INCNUM, CITNUM, CIT_AGE, OFFNUM, OFF_AGE, OFF_YR_EMPLOY
## lgl   (1): CIT_WEAPON_TYPE
## time  (1): OCCURRED_TM
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

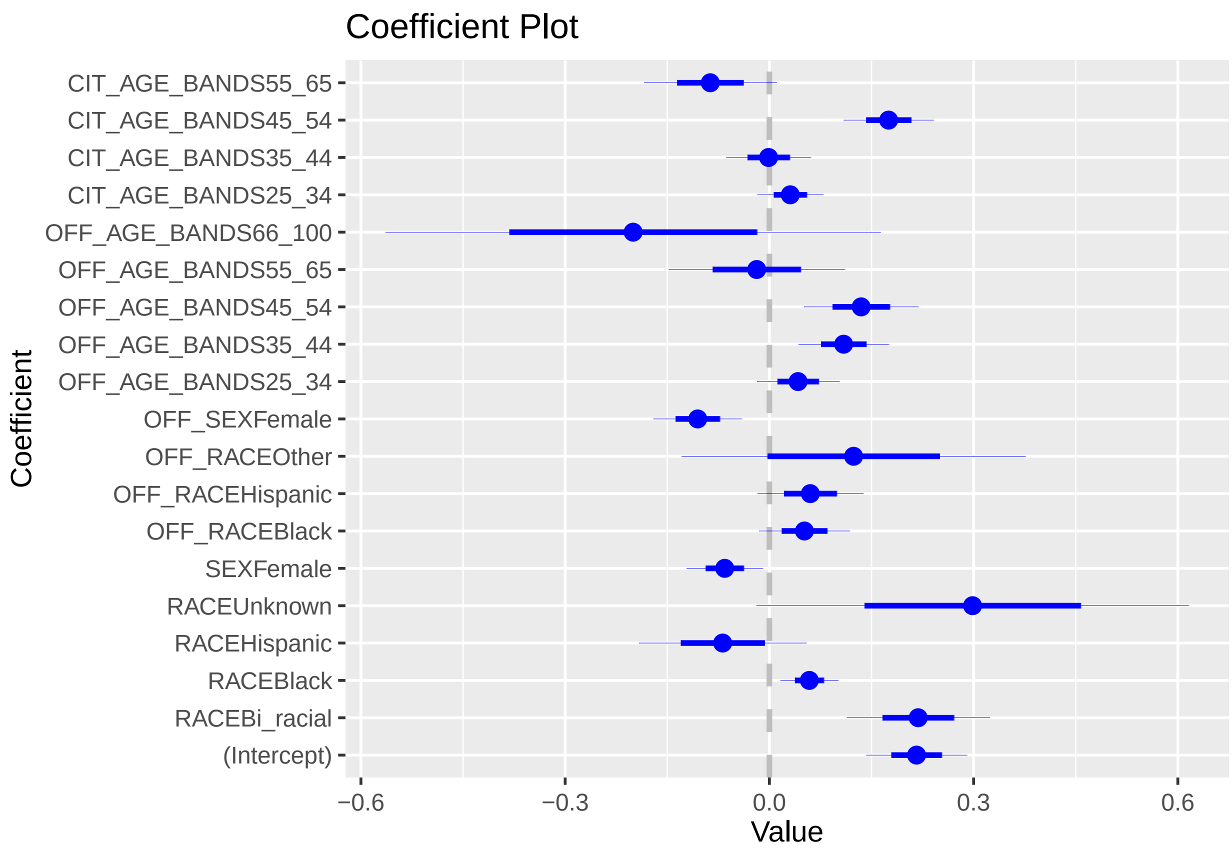
```
ind_dat_2020 <- ind_dat_2020 %>%
  mutate_all(~str_replace_all(., "/", "_")) %>%
  mutate_all(~str_replace_all(., " ", "_")) %>%
  mutate_all(~str_replace_all(., ",", "_")) %>%
  mutate_all(~str_replace_all(., "-", "_")) %>%
  mutate_all(~str_replace_all(., "\\(|\\)", "")) %>%
  mutate_all(~str_replace_all(., ":", "_")) %>%
  mutate_all(~str_replace_all(., "__", "_")) %>%
  mutate(DISTRICT = UDTEXT24B) %>%
  mutate(datetime = mdy(OCCURRED_DT)) %>%
  filter(year(datetime) == 2020) %>%
   mutate(OFF_AGE = as.numeric(OFF_AGE),
    OFF_AGE_BANDS = cut(OFF_AGE, breaks = age_breaks, labels = c("18_24", "25_34", "35_44", "45_54", "5
                   include.lowest = TRUE, right = FALSE)) %>%
  mutate(CIT_AGE = as.numeric(CIT_AGE),
    CIT_AGE_BANDS = cut(CIT_AGE, breaks = age_breaks, labels = c("18_24", "25_34", "35_44", "45_54", "5
                   include.lowest = TRUE, right = FALSE)) %>%
  drop_na(RACE, SEX, OFF_RACE, OFF_SEX, OFF_AGE, OFF_AGE_BANDS, CIT_AGE_BANDS, UOF_FORCE_TYPE)

serious_cats <- c("Less_Lethal_Pepperball", "Lethal_Handgun", "Canine_Bite", "Less_Lethal_Bean_Bag", "Pl
                  "Less_Lethal_CS_OC", "Physical_Kick", "Less_Lethal_CS_Fogger", "Physical_Leg_Sweep",
                  "Physical_Fist_Strike", "Less_Lethal_Taser", "Physical_Knee_Strike", "Physical_Elbow_S
                  "Physical_Palm_Strike", "Less_Lethal_Clearout_OC", "Lethal_Rifle", "Less_Lethal_Baton
                  "Less_Lethal_BPS_Gas", "Less_Lethal_Burning_CS", "Lethal_Shotgun", "Less_Lethal_CS_Gr

ind_dat_2020 <- ind_dat_2020 %>%
  mutate(serious = as.integer(UOF_FORCE_TYPE %in% serious_cats))
```

```r
ind_dat_2020$RACE <- factor(ind_dat_2020$RACE)
ind_dat_2020$RACE <- relevel(ind_dat_2020$RACE, ref = "White")
ind_dat_2020$SEX <- factor(ind_dat_2020$SEX)
ind_dat_2020$SEX <- relevel(ind_dat_2020$SEX, ref = "Male")
ind_dat_2020$OFF_RACE <- factor(ind_dat_2020$OFF_RACE)
ind_dat_2020$OFF_RACE <- relevel(ind_dat_2020$OFF_RACE, ref = "White")
ind_dat_2020$OFF_SEX <- factor(ind_dat_2020$OFF_SEX)
ind_dat_2020$OFF_SEX <- relevel(ind_dat_2020$OFF_SEX, ref = "Male")
ind_dat_2020$OFF_AGE_BANDS <- factor(ind_dat_2020$OFF_AGE_BANDS)
ind_dat_2020$OFF_AGE_BANDS <- relevel(ind_dat_2020$OFF_AGE_BANDS, ref = "18_24")
ind_dat_2020$CIT_AGE_BANDS <- factor(ind_dat_2020$CIT_AGE_BANDS)
ind_dat_2020$CIT_AGE_BANDS <- relevel(ind_dat_2020$CIT_AGE_BANDS, ref = "18_24")

encoded_ind_2020 <- model.matrix(~ RACE + SEX + OFF_RACE + OFF_SEX + OFF_AGE_BANDS + CIT_AGE_BANDS, data

ind_dat_2020_final <- cbind(ind_dat_2020, encoded_ind_2020)

model_indianapolis_2020 <- glm(serious ~ RACEBi_racial + RACEBlack + RACEHispanic + RACEUnknown + SEXFem

summary(model_indianapolis_2020)
```

```
##
## Call:
## glm(formula = serious ~ RACEBi_racial + RACEBlack + RACEHispanic +
##     RACEUnknown + SEXFemale + OFF_RACEBlack + OFF_RACEHispanic +
##     OFF_RACEOther + OFF_SEXFemale + OFF_AGE_BANDS25_34 + OFF_AGE_BANDS35_44 +
##     OFF_AGE_BANDS45_54 + OFF_AGE_BANDS55_65 + OFF_AGE_BANDS66_100 +
##     CIT_AGE_BANDS25_34 + CIT_AGE_BANDS35_44 + CIT_AGE_BANDS45_54 +
##     CIT_AGE_BANDS55_65, data = ind_dat_2020_final)
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.216149   0.036936   5.852 5.43e-09 ***
## RACEBi_racial        0.218580   0.052390   4.172 3.11e-05 ***
## RACEBlack            0.058655   0.021155   2.773  0.00560 **
## RACEHispanic        -0.068657   0.061390  -1.118  0.26351
## RACEUnknown          0.298466   0.158754   1.880  0.06020 .
## SEXFemale           -0.065616   0.027896  -2.352  0.01873 *
## OFF_RACEBlack        0.051423   0.033235   1.547  0.12192
## OFF_RACEHispanic     0.060095   0.038731   1.552  0.12087
## OFF_RACEOther        0.123658   0.126192   0.980  0.32721
## OFF_SEXFemale       -0.105289   0.032358  -3.254  0.00115 **
## OFF_AGE_BANDS25_34   0.042105   0.030235   1.393  0.16386
## OFF_AGE_BANDS35_44   0.109057   0.033181   3.287  0.00103 **
## OFF_AGE_BANDS45_54   0.134792   0.041981   3.211  0.00134 **
## OFF_AGE_BANDS55_65  -0.018629   0.064594  -0.288  0.77306
## OFF_AGE_BANDS66_100 -0.200181   0.181776  -1.101  0.27088
## CIT_AGE_BANDS25_34   0.030666   0.024178   1.268  0.20478
## CIT_AGE_BANDS35_44  -0.001164   0.031008  -0.038  0.97005
## CIT_AGE_BANDS45_54   0.175143   0.032939   5.317 1.14e-07 ***
## CIT_AGE_BANDS55_65  -0.086850   0.048600  -1.787  0.07404 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for gaussian family taken to be 0.2189251)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 607.30  on 2774  degrees of freedom
## AIC: 3704.5
##
## Number of Fisher Scoring iterations: 2
```

```
coefplot(model_indianapolis_2020)
```

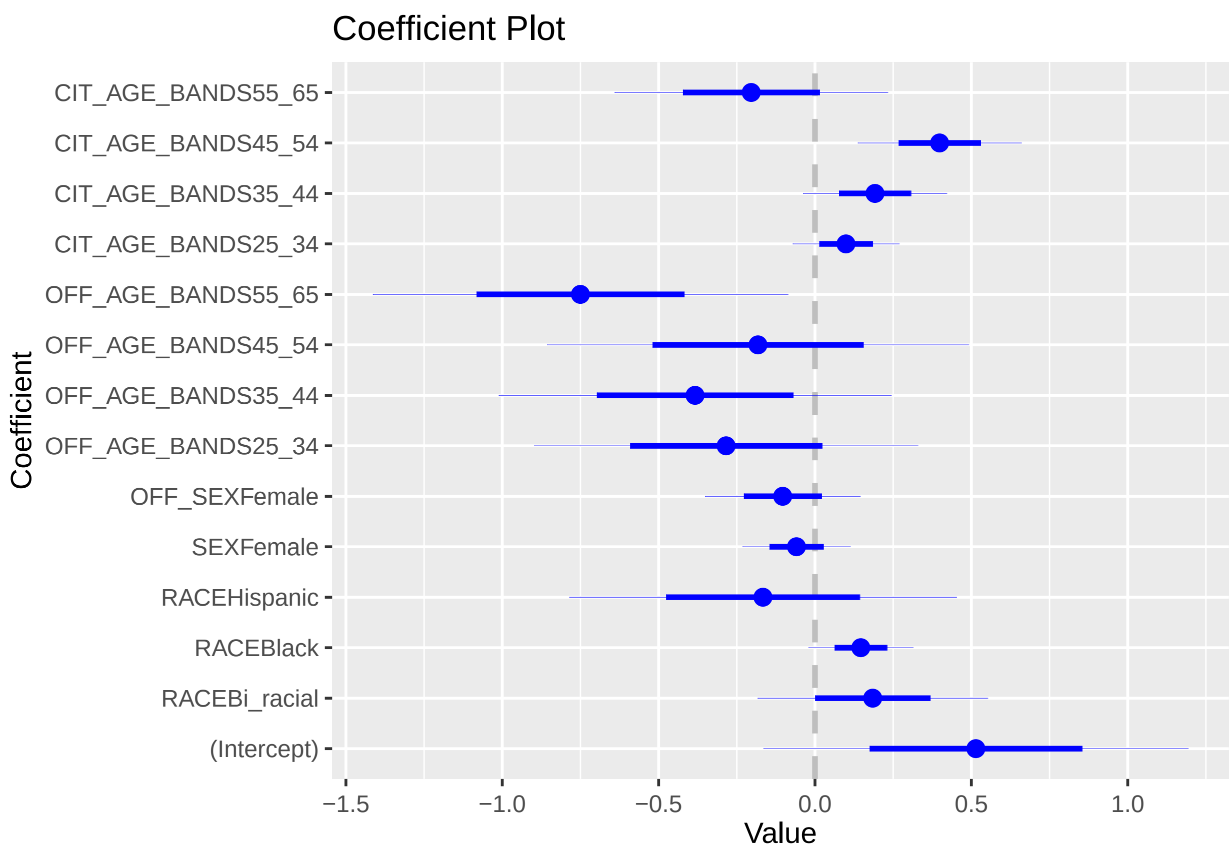



```
ind_dat_blackoff <- ind_dat_2020_final %>%
  filter(OFF_RACE == "Black")

model_ind_blackoff_2020 <- glm(serious ~ RACEBi_racial + RACEBlack + RACEHispanic + RACEUnknown + SEXFe

summary(model_ind_blackoff_2020)
```

```
##
## Call:
## glm(formula = serious ~ RACEBi_racial + RACEBlack + RACEHispanic +
##     RACEUnknown + SEXFemale + OFF_SEXFemale + OFF_AGE_BANDS25_34 +
##     OFF_AGE_BANDS35_44 + OFF_AGE_BANDS45_54 + OFF_AGE_BANDS55_65 +
##     OFF_AGE_BANDS66_100 + CIT_AGE_BANDS25_34 + CIT_AGE_BANDS35_44 +
##     CIT_AGE_BANDS45_54 + CIT_AGE_BANDS55_65, data = ind_dat_blackoff)
```

```
##
## Coefficients: (2 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.51434    0.33940   1.515   0.1311
## RACEBi_racial        0.18450    0.18383   1.004   0.3167
## RACEBlack            0.14651    0.08351   1.754   0.0808 .
## RACEHispanic        -0.16645    0.30940  -0.538   0.5912
## RACEUnknown               NA         NA      NA       NA
## SEXFemale           -0.05927    0.08610  -0.688   0.4919
## OFF_SEXFemale       -0.10331    0.12396  -0.833   0.4056
## OFF_AGE_BANDS25_34  -0.28429    0.30657  -0.927   0.3548
## OFF_AGE_BANDS35_44  -0.38362    0.31375  -1.223   0.2228
## OFF_AGE_BANDS45_54  -0.18240    0.33695  -0.541   0.5888
## OFF_AGE_BANDS55_65  -0.75013    0.33179  -2.261   0.0248 *
## OFF_AGE_BANDS66_100       NA         NA      NA       NA
## CIT_AGE_BANDS25_34   0.09898    0.08500   1.164   0.2455
## CIT_AGE_BANDS35_44   0.19177    0.11474   1.671   0.0961 .
## CIT_AGE_BANDS45_54   0.39843    0.13073   3.048   0.0026 **
## CIT_AGE_BANDS55_65  -0.20398    0.21826  -0.935   0.3511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2170585)
##
##     Null deviance: 54.520  on 226  degrees of freedom
## Residual deviance: 46.233  on 213  degrees of freedom
## AIC: 312.99
##
## Number of Fisher Scoring iterations: 2
```

```
coefplot(model_ind_blackoff_2020)
```

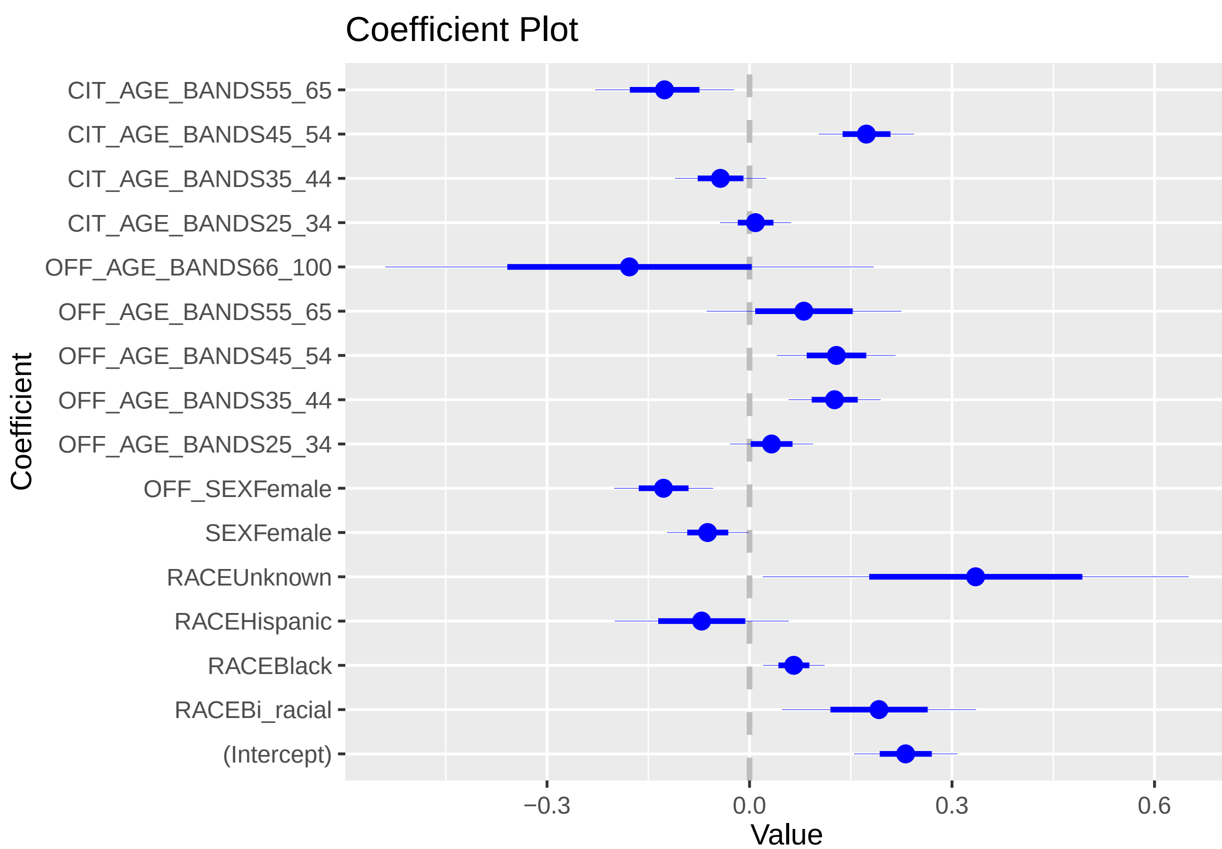



```r
ind_dat_whiteoff <- ind_dat_2020_final %>%
  filter(OFF_RACE == "White")

model_ind_whiteoff_2020 <- glm(serious ~ RACEBi_racial + RACEBlack + RACEHispanic + RACEUnknown + SEXFe

summary(model_ind_whiteoff_2020)
```

```
##
## Call:
## glm(formula = serious ~ RACEBi_racial + RACEBlack + RACEHispanic +
##     RACEUnknown + SEXFemale + OFF_SEXFemale + OFF_AGE_BANDS25_34 +
##     OFF_AGE_BANDS35_44 + OFF_AGE_BANDS45_54 + OFF_AGE_BANDS55_65 +
##     OFF_AGE_BANDS66_100 + CIT_AGE_BANDS25_34 + CIT_AGE_BANDS35_44 +
##     CIT_AGE_BANDS45_54 + CIT_AGE_BANDS55_65, data = ind_dat_whiteoff)
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.231264   0.038157   6.061 1.57e-09 ***
## RACEBi_racial        0.191726   0.071689   2.674 0.007537 **
## RACEBlack            0.065475   0.022558   2.903 0.003736 **
## RACEHispanic        -0.071005   0.064173  -1.106 0.268643
## RACEUnknown          0.334953   0.157506   2.127 0.033556 *
## SEXFemale           -0.062114   0.029927  -2.076 0.038047 *
## OFF_SEXFemale       -0.127519   0.036449  -3.499 0.000476 ***
## OFF_AGE_BANDS25_34   0.032527   0.030459   1.068 0.285683
```

```
## OFF_AGE_BANDS35_44   0.125956   0.033824   3.724 0.000201 ***
## OFF_AGE_BANDS45_54   0.128619   0.043710   2.943 0.003287 **
## OFF_AGE_BANDS55_65   0.080447   0.071921   1.119 0.263451
## OFF_AGE_BANDS66_100 -0.178009   0.180677  -0.985 0.324610
## CIT_AGE_BANDS25_34   0.008789   0.025993   0.338 0.735308
## CIT_AGE_BANDS35_44  -0.043188   0.033511  -1.289 0.197606
## CIT_AGE_BANDS45_54   0.173044   0.035131   4.926 8.99e-07 ***
## CIT_AGE_BANDS55_65  -0.125976   0.051151  -2.463 0.013855 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2143535)
##
##     Null deviance: 529.88  on 2377  degrees of freedom
## Residual deviance: 506.30  on 2362  degrees of freedom
## AIC: 3104
##
## Number of Fisher Scoring iterations: 2
```

```
coefplot(model_ind_whiteoff_2020)
```

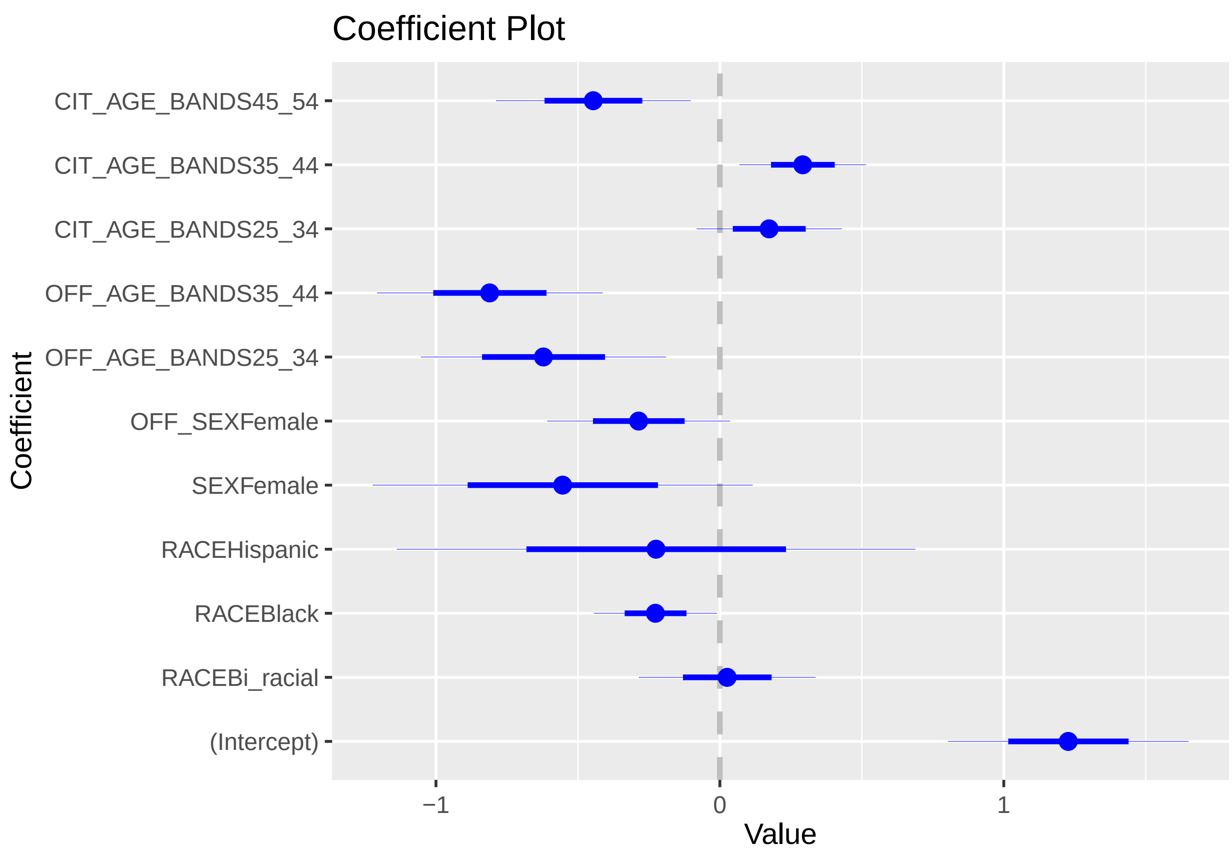



```
ind_dat_hispanicoff <- ind_dat_2020_final %>%
  filter(OFF_RACE == "Hispanic")

model_ind_hispanicoff_2020 <- glm(serious ~ RACEBi_racial + RACEBlack + RACEHispanic + RACEUnknown + SE
```

```r
summary(model_ind_hispanicoff_2020)
```

```
##
## Call:
## glm(formula = serious ~ RACEBi_racial + RACEBlack + RACEHispanic +
##     RACEUnknown + SEXFemale + OFF_SEXFemale + OFF_AGE_BANDS25_34 +
##     OFF_AGE_BANDS35_44 + OFF_AGE_BANDS45_54 + OFF_AGE_BANDS55_65 +
##     OFF_AGE_BANDS66_100 + CIT_AGE_BANDS25_34 + CIT_AGE_BANDS35_44 +
##     CIT_AGE_BANDS45_54 + CIT_AGE_BANDS55_65, data = ind_dat_hispanicoff)
##
## Coefficients: (5 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.22708    0.21108   5.813 3.14e-08 ***
## RACEBi_racial        0.02557    0.15525   0.165  0.86940
## RACEBlack           -0.22708    0.10782  -2.106  0.03674 *
## RACEHispanic        -0.22496    0.45619  -0.493  0.62259
## RACEUnknown               NA         NA      NA       NA
## SEXFemale           -0.55396    0.33409  -1.658  0.09922 .
## OFF_SEXFemale       -0.28615    0.16062  -1.781  0.07669 .
## OFF_AGE_BANDS25_34  -0.62155    0.21561  -2.883  0.00447 **
## OFF_AGE_BANDS35_44  -0.81082    0.19838  -4.087 6.84e-05 ***
## OFF_AGE_BANDS45_54        NA         NA      NA       NA
## OFF_AGE_BANDS55_65        NA         NA      NA       NA
## OFF_AGE_BANDS66_100       NA         NA      NA       NA
## CIT_AGE_BANDS25_34   0.17338    0.12745   1.360  0.17558
## CIT_AGE_BANDS35_44   0.29187    0.11119   2.625  0.00949 **
## CIT_AGE_BANDS45_54  -0.44604    0.17116  -2.606  0.01001 *
## CIT_AGE_BANDS55_65        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1975715)
##
##     Null deviance: 42.672  on 173  degrees of freedom
## Residual deviance: 32.204  on 163  degrees of freedom
## AIC: 224.26
##
## Number of Fisher Scoring iterations: 2
```

```r
coefplot(model_ind_hispanicoff_2020)
```




```r
model_indianapolis_subrace <- glm(serious ~ RACEBi_racial + RACEBlack + RACEHispanic + RACEUnknown, data
summary(model_indianapolis_subrace)
```

```
##
## Call:
## glm(formula = serious ~ RACEBi_racial + RACEBlack + RACEHispanic +
##     RACEUnknown, data = ind_dat_2020_final)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.31206    0.01786  17.473  < 2e-16 ***
## RACEBi_racial  0.20717    0.04981   4.159 3.29e-05 ***
## RACEBlack      0.04261    0.02090   2.039   0.0416 *
## RACEHispanic  -0.10916    0.05982  -1.825   0.0681 .
## RACEUnknown    0.24350    0.15907   1.531   0.1259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2248688)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 626.93  on 2788  degrees of freedom
## AIC: 3765.4
##
## Number of Fisher Scoring iterations: 2
```

```r
model_indianapolis_subsex <- glm(serious ~ SEXFemale, data=ind_dat_2020_final)
summary(model_indianapolis_subsex)
```

```
##
## Call:
## glm(formula = serious ~ SEXFemale, data = ind_dat_2020_final)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.351958   0.009612  36.616   <2e-16 ***
## SEXFemale   -0.041107   0.027510  -1.494    0.135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2265534)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 632.31  on 2791  degrees of freedom
## AIC: 3783.2
##
## Number of Fisher Scoring iterations: 2
```

```r
model_indianapolis_offrace <- glm(serious ~  OFF_RACEBlack + OFF_RACEHispanic + OFF_RACEOther, data=ind
summary(model_indianapolis_offrace)
```

```
##
## Call:
## glm(formula = serious ~ OFF_RACEBlack + OFF_RACEHispanic + OFF_RACEOther,
##     data = ind_dat_2020_final)
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.33516    0.00975  34.374   <2e-16 ***
## OFF_RACEBlack     0.06573    0.03303   1.990   0.0467 *
## OFF_RACEHispanic  0.09588    0.03734   2.568   0.0103 *
## OFF_RACEOther     0.09342    0.12745   0.733   0.4636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2260673)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 630.50  on 2789  degrees of freedom
## AIC: 3779.2
##
## Number of Fisher Scoring iterations: 2
```

```r
model_indianapolis_offsex <- glm(serious ~ OFF_SEXFemale, data=ind_dat_2020_final)
summary(model_indianapolis_offsex)
```

```
##
```

```
## Call:
## glm(formula = serious ~ OFF_SEXFemale, data = ind_dat_2020_final)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.35470    0.00943  37.615  < 2e-16 ***
## OFF_SEXFemale -0.08670    0.03152  -2.751  0.00598 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2261216)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 631.11  on 2791  degrees of freedom
## AIC: 3777.9
##
## Number of Fisher Scoring iterations: 2
```

```
model_indianapolis_offage <- glm(serious ~ OFF_AGE_BANDS25_34 + OFF_AGE_BANDS35_44 + OFF_AGE_BANDS45_54
summary(model_indianapolis_offage)
```

```
##
## Call:
## glm(formula = serious ~ OFF_AGE_BANDS25_34 + OFF_AGE_BANDS35_44 +
##     OFF_AGE_BANDS45_54 + OFF_AGE_BANDS55_65 + OFF_AGE_BANDS66_100,
##     data = ind_dat_2020_final)
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.27333    0.02736   9.991  < 2e-16 ***
## OFF_AGE_BANDS25_34   0.05889    0.02996   1.966 0.049435 *
## OFF_AGE_BANDS35_44   0.12229    0.03281   3.728 0.000197 ***
## OFF_AGE_BANDS45_54   0.16335    0.04158   3.928 8.76e-05 ***
## OFF_AGE_BANDS55_65  -0.03453    0.06403  -0.539 0.589758
## OFF_AGE_BANDS66_100 -0.27333    0.18118  -1.509 0.131497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2245323)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 625.77  on 2787  degrees of freedom
## AIC: 3762.2
##
## Number of Fisher Scoring iterations: 2
```

```
model_indianapolis_subage <- glm(serious ~ CIT_AGE_BANDS25_34 + CIT_AGE_BANDS35_44 + CIT_AGE_BANDS45_54
summary(model_indianapolis_subage)
```

```
##
## Call:
## glm(formula = serious ~ CIT_AGE_BANDS25_34 + CIT_AGE_BANDS35_44 +
##     CIT_AGE_BANDS45_54 + CIT_AGE_BANDS55_65, data = ind_dat_2020_final)
```

```
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.318739   0.019811  16.089  < 2e-16 ***
## CIT_AGE_BANDS25_34   0.026114   0.023682   1.103    0.270
## CIT_AGE_BANDS35_44  -0.006385   0.030246  -0.211    0.833
## CIT_AGE_BANDS45_54   0.158272   0.032194   4.916 9.33e-07 ***
## CIT_AGE_BANDS55_65  -0.073125   0.048562  -1.506    0.132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.224098)
##
##     Null deviance: 632.82  on 2792  degrees of freedom
## Residual deviance: 624.79  on 2788  degrees of freedom
## AIC: 3755.8
##
## Number of Fisher Scoring iterations: 2
```

```r
ind_dat_whiteoff %>%
  filter(RACE == "White") %>%
  nrow()
```

```
## [1] 601
```

```r
ind_dat_whiteoff %>%
  filter(RACE == "Black") %>%
  nrow()
```

```
## [1] 1659
```

```r
ind_dat_whiteoff %>%
  filter(RACE == "Hispanic") %>%
  nrow()
```

```
## [1] 63
```

```r
ind_dat_blackoff %>%
  filter(RACE == "White") %>%
  nrow()
```

```
## [1] 62
```

```r
ind_dat_blackoff %>%
  filter(RACE == "Black") %>%
  nrow()
```

```
## [1] 139
```

```r
ind_dat_blackoff %>%
  filter(RACE == "Hispanic") %>%
  nrow()
```

```
## [1] 3
```

```r
ind_dat_hispanicoff %>%
  filter(RACE == "White") %>%
  nrow()
```

```
## [1] 36
```

```r
ind_dat_hispanicoff %>%
  filter(RACE == "Black") %>%
  nrow()
```

```
## [1] 100
```

```r
ind_dat_hispanicoff %>%
  filter(RACE == "Hispanic") %>%
  nrow()
```

```
## [1] 3
```

```r
matrix_data <- matrix(c(
  601, 1659, 63,
  62, 139, 3,
  36, 100, 3
), ncol = 3, byrow = TRUE)

row_names <- c("White subject", "Black subject", "Hispanic subject")
col_names <- c("White officer", "Black officer", "Hispanic officer")

data_df <- as.data.frame(matrix_data)
rownames(data_df) <- row_names
colnames(data_df) <- col_names

print(data_df)
```

```
##                  White officer Black officer Hispanic officer
## White subject              601          1659               63
## Black subject               62           139                3
## Hispanic subject            36           100                3
```

```r
file_path <- "../data/ind_dat_2020_final.csv"
write.csv(ind_dat_2020_final, file = file_path, row.names = FALSE)

file_path <- "../data/sea_dat_2020_final.csv"
write.csv(sea_dat_2020_final, file = file_path, row.names = FALSE)
```

```r
file_path <- "../data/mpd_dat_enc_2020.csv"
write.csv(mpd_dat_enc_2020, file = file_path, row.names = FALSE)

file_path <- "../data/pdx_dat_2020_final.csv"
write.csv(pdx_dat_2020_final, file = file_path, row.names = FALSE)
```

Discussion:

Our research fills a gap in the existing literature by exploring the use of force in DC specifically, rather than in the US as a whole or other metro areas. Other research has explored the implementation of stop and frisk in DC, but not the use of force broadly. The history of racialized policing in America is long, and its roots lay close to Washington, DC. MPD is uniquely positioned as a majority Black police department in a demographically diverse city. By exploring how and which of its officers use force, on whom, and where we may build an understanding of trends in American policing as a whole.

Appendix All other regression outputs, including our exploration of individual relationships, are included below. We did this in order to make the results section smaller and easier to read.

Individual variable/dependent relationships for DC.

```r
model_mpd_district <- glm(serious ~ IncidentDistrict2D + IncidentDistrict3D + IncidentDistrict4D + Incid
               data = mpd_dat_enc_2020)
summary(model_mpd_district)
```

```
##
## Call:
## glm(formula = serious ~ IncidentDistrict2D + IncidentDistrict3D +
##     IncidentDistrict4D + IncidentDistrict5D + IncidentDistrict6D +
##     IncidentDistrictOther, data = mpd_dat_enc_2020)
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.024961   0.005738   4.350 1.43e-05 ***
## IncidentDistrict2D     0.005573   0.013930   0.400   0.6891
## IncidentDistrict3D     0.020022   0.010294   1.945   0.0519 .
## IncidentDistrict4D     0.003941   0.012447   0.317   0.7516
## IncidentDistrict5D    -0.022492   0.009222  -2.439   0.0148 *
## IncidentDistrict6D    -0.016674   0.009552  -1.746   0.0810 .
## IncidentDistrictOther  0.197261   0.048766   4.045 5.43e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02110705)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 42.277  on 2003  degrees of freedom
## AIC: -2041.8
##
## Number of Fisher Scoring iterations: 2
```

```r
model_mpd_offrace <- glm(serious ~ OfficerRaceBlack_African_American + OfficerRaceHispanic + OfficerRac
                         data = mpd_dat_enc_2020)
summary(model_mpd_offrace)
```

```
##
## Call:
## glm(formula = serious ~ OfficerRaceBlack_African_American + OfficerRaceHispanic +
##     OfficerRaceOther, data = mpd_dat_enc_2020)
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.027460   0.004949   5.549 3.26e-08 ***
## OfficerRaceBlack_African_American -0.013980   0.007122  -1.963   0.0498 *
## OfficerRaceHispanic                0.002243   0.011422   0.196   0.8443
## OfficerRaceOther                  -0.002036   0.014349  -0.142   0.8872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02140488)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 42.938  on 2006  degrees of freedom
## AIC: -2016.6
##
## Number of Fisher Scoring iterations: 2
```

```r
model_mpd_offgender <- glm(serious ~ OfficerGenderFemale, data = mpd_dat_enc_2020)
summary(model_mpd_offgender)
```

```
##
## Call:
## glm(formula = serious ~ OfficerGenderFemale, data = mpd_dat_enc_2020)
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.021488   0.003436   6.253  4.9e-10 ***
## OfficerGenderFemale 0.004153   0.011032   0.376    0.707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02143116)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 43.034  on 2008  degrees of freedom
## AIC: -2016.1
##
## Number of Fisher Scoring iterations: 2
```

```r
model_mpd_age <- glm(serious ~ AgeBands25_34 + AgeBands35_44 + AgeBands45_54 + AgeBands55_65 + AgeBands6
                     data = mpd_dat_enc_2020)
summary(model_mpd_age)
```

```
##
## Call:
## glm(formula = serious ~ AgeBands25_34 + AgeBands35_44 + AgeBands45_54 +
##     AgeBands55_65 + AgeBands66_100, data = mpd_dat_enc_2020)
##
```

```
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.031858   0.006142   5.187 2.35e-07 ***
## AgeBands25_34  -0.021258   0.007926  -2.682  0.00738 **
## AgeBands35_44   0.006603   0.009431   0.700  0.48393
## AgeBands45_54  -0.024340   0.014070  -1.730  0.08380 .
## AgeBands55_65  -0.031858   0.023094  -1.380  0.16789
## AgeBands66_100 -0.031858   0.073250  -0.435  0.66366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.0213115)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 42.708  on 2004  degrees of freedom
## AIC: -2023.4
##
## Number of Fisher Scoring iterations: 2
```

```
model_mpd_subrace <- glm(serious ~ subject_raceBlack_African_American + subject_raceHispanic,
                     data = mpd_dat_enc_2020)
summary(model_mpd_subrace)
```

```
##
## Call:
## glm(formula = serious ~ subject_raceBlack_African_American +
##     subject_raceHispanic, data = mpd_dat_enc_2020)
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.0223881  0.0126497   1.770   0.0769 .
## subject_raceBlack_African_American -0.0006961  0.0131012  -0.053   0.9576
## subject_raceHispanic                0.0088619  0.0288110   0.308   0.7584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02144191)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 43.034  on 2007  degrees of freedom
## AIC: -2014.1
##
## Number of Fisher Scoring iterations: 2
```

```
model_mpd_subgender <- glm(serious ~ subject_genderFemale,
                     data = mpd_dat_enc_2020)
summary(model_mpd_subgender)
```

```
##
## Call:
## glm(formula = serious ~ subject_genderFemale, data = mpd_dat_enc_2020)
##
## Coefficients:
```

```
##                       Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            0.02377    0.00344   6.911 6.44e-12 ***
## subject_genderFemale -0.01879    0.01088  -1.728   0.0841 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02140085)
##
##     Null deviance: 43.037  on 2009  degrees of freedom
## Residual deviance: 42.973  on 2008  degrees of freedom
## AIC: -2019
##
## Number of Fisher Scoring iterations: 2
```

Citation:

Schwartz G., (2020). Mapping fatal police violence across U.S. metropolitan areas: Overall

Fryer, R. G. (2019). Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. Proceedings of the National Academy of Sciences, 116(34), 16793-16798. doi:10.1073/pnas.1821204116

rates and racial/ ethnic inequities, 2013-2017. EBSCO. https://web.p.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=6ba04cd8-6484-4883-8e3e-7e0e405082c8%40redis

Potter, Gary. "The History of Police in America and the First Force." Time Magazine, time.com, 10 June 2022, https://time.com/4779112/police-history-origins/.

Ross, Cody T., et al. "Racial Disparities in Police Use of Deadly Force Against Unarmed Individuals Persist After Appropriately Benchmarking Shooting Data on Violent Crime Rates." Social Psychological and Personality Science, vol. 12, no. 3, 2021, pp. 1948550620916071, doi:10.1177/1948550620916071.

"Police Shootings Database 2015-2023: Search by Race, Age, Department." Washington Post, https://www.washingtonpost.com/graphics/investigations/police-shootings-database/. Accessed 9 Dec. 2023.

Ross, Cody T., Bruce Winterhalder, and Richard McElreath. "Racial Disparities in Police Use of Deadly Force Against Unarmed Individuals Persist After Appropriately Benchmarking Shooting Data on Violent Crime Rates." Social Psychological and Personality Science 12, no. 3 (2021): 1948–5506. doi:10.1177/1948550620916071.

Colquhoun, Patrick. A Treatise on the Police of the Metropolis. 3rd ed., J. Mawman, 1797.

Golash-Boza, T. Et al, (2022). Broken windows and order-maintenance policing in gentrifying Washington, DC. Taylor & Francis Online. https://doi.org/10.1080/10439463.2022.2085268

Hall, Katie. "Police Use-of-Force Data 'a Huge Mess' Across the U.S." TCA Regional News; Chicago, 8 Sep. 2019.

"The Origins of Modern Day Policing." NAACP, naacp.org, https://naacp.org/find-resources/history-explained/origins-modern-day-policing. Accessed 9 Dec. 2023.

Lepore, Jill. "The Invention of the Police." The New Yorker, 20 July 2020, https://www.newyorker.com/magazine/2020/07/20/the-invention-of-the-police. Accessed 9 Dec. 2023.

Metropolitan Department. (2022). 2020 and 2021 UoF_Explanatory Notesv2.

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. Proceedings of the National Academy of Sciences, 116(34), 16793–16798. https://doi.org/10.1073/pnas.1821204116

Report reveals over 1,000 uses of forces by DC Police in 2021 | wusa9.com. (n.d.). Retrieved October 17, 2023, from https://www.wusa9.com/article/news/local/dc/annual-use-of-force-report-dc-police-released-more-than-800-complaints-filed/65-9bc43d2b-059d-43bd-a785-3bed9d6ca858

Lepore, J. (2020, July 13). The New Yorker. https://www.newyorker.com/magazine/2020/07/20/the-invention-of-the-police#:~:text=It%20is%20also%20often%20said,the%20history%20of%20the%20police

Keating, D., & Uhrmacher, K. (2020, June 4). In urban areas, police are consistently much whiter than the people they serve. The Washington Post. https://www.washingtonpost.com/nation/2020/06/04/urban-areas-police-are-consistently-much-whiter-than-people-they-serve/

U.S. Census Bureau QuickFacts: District of Columbia. Accessed October 17, 2023. https://www.census.gov/quickfacts/fact/table/DC/PST045222

Velez, M. B., Lyons, C. J., & Santoro, W. A. (2015). The political context of the percent black-neighborhood violence link: A multilevel analysis. Social Problems, 62(1), 93–119. https://doi.org/10.1093/socpro/spu005

Use of Force Final Report DATA 412/612

Lit review

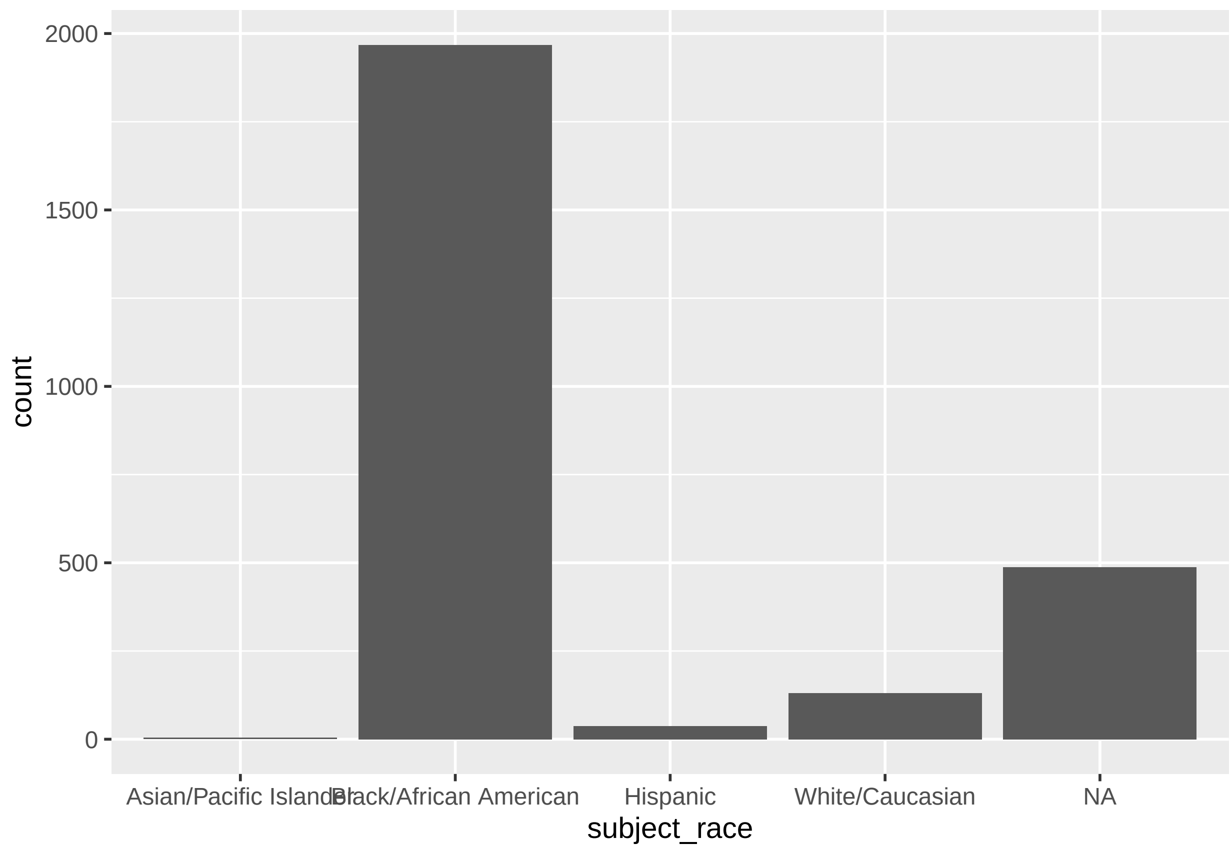
Exploratory Data Analysis

Washington, DC

1. Distribution of Incident Counts by Subject Race Black/ African American have a significantly higher amount of incident counts in 2020.

Conversely, Asian/Pacific Islander individuals appear to experience the least instances of use of force.

```
df <- read.csv("../data/UOF_DC_2020.csv")
mpd_dat <- df
ggplot(data = mpd_dat, mapping = aes(x = subject_race)) +
  geom_bar()
```



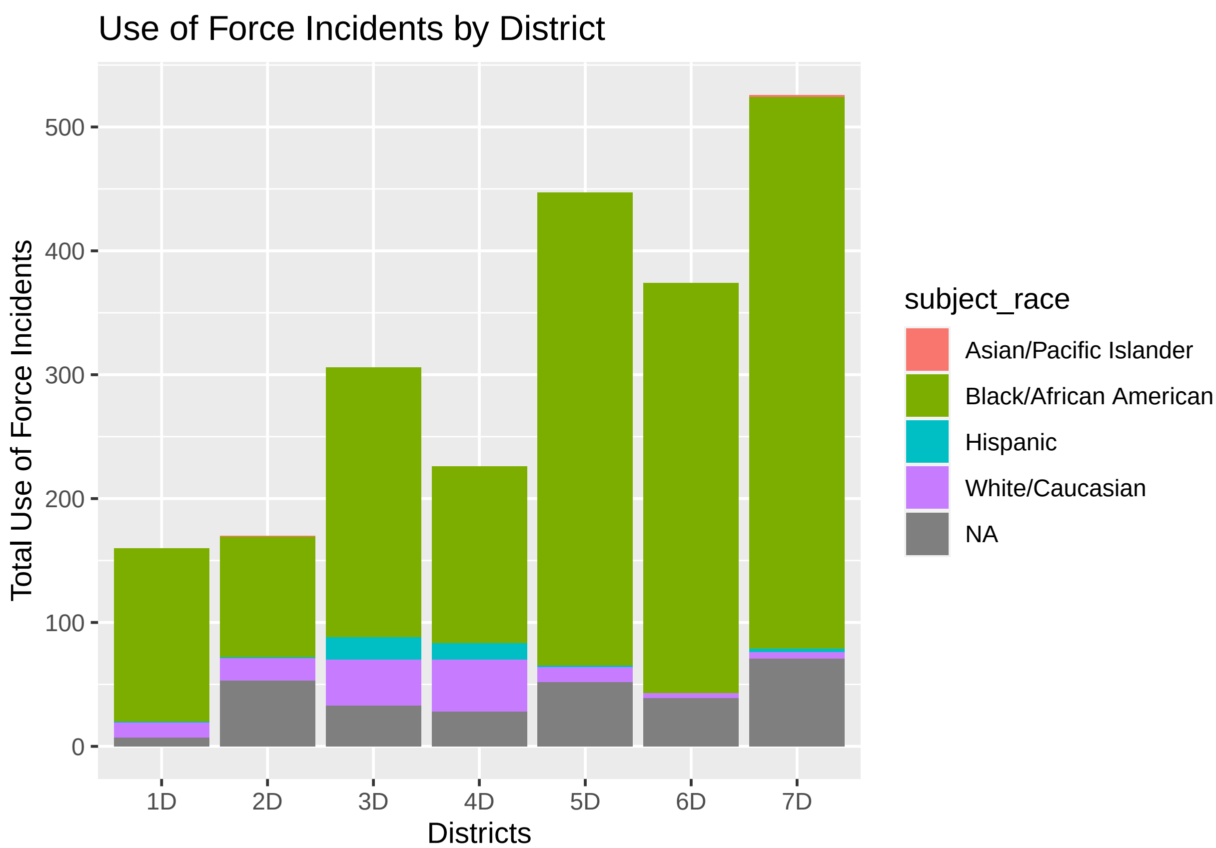
2. UOF Incident Counts by District and Subject Race Across all districts in DC, it is prevalent that Black/African Americans experience UOF at a higher rate than their counterparts.

7th District has the highest total UOF incidents out of all the DC districts.

```
force_by_district_race <- mpd_dat %>%
group_by(OfficerAssignment, subject_race) %>% summarize(Total_Use_of_Force = n()) #total num of cases o
```

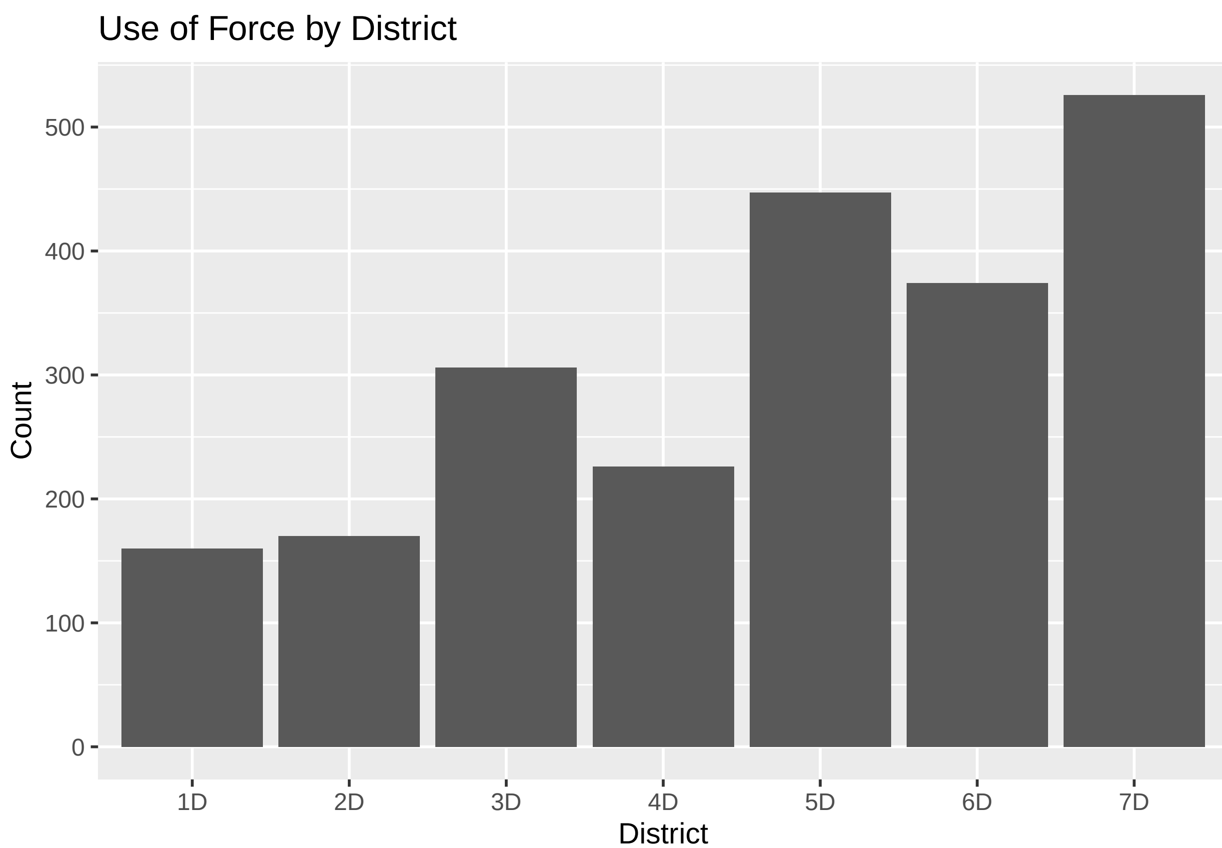
```
## `summarise()` has grouped output by 'OfficerAssignment'. You can override using
## the `.groups` argument.
```

```
ggplot(subset(force_by_district_race, OfficerAssignment %in% c("1D", "2D", "3D", "4D", "5D", "6D", "7D"
       aes(x = OfficerAssignment, y = Total_Use_of_Force, fill = subject_race)) +
  geom_bar(stat = "identity") +
  labs(title = "Use of Force Incidents by District") +
  xlab("Districts") +
  ylab("Total Use of Force Incidents")
```



3. Use of Force by District. 1D, 2D, and 3D are posed to be the safest districts in DC, whereas 7D is deemed to be the most dangerous.

```
districts <- c("1D", "2D", "3D", "4D", "5D", "6D", "7D")
mpd_dat_filtered <- mpd_dat[mpd_dat$OfficerAssignment %in% districts, ]

uof_counts <- mpd_dat_filtered %>%
  group_by(OfficerAssignment) %>%
  summarise(Count = n())

ggplot(uof_counts, aes(x = OfficerAssignment, y = Count)) +
  geom_bar(stat = "identity") +
  labs(title = "Use of Force by District",
       x = "District",
       y = "Count")
```



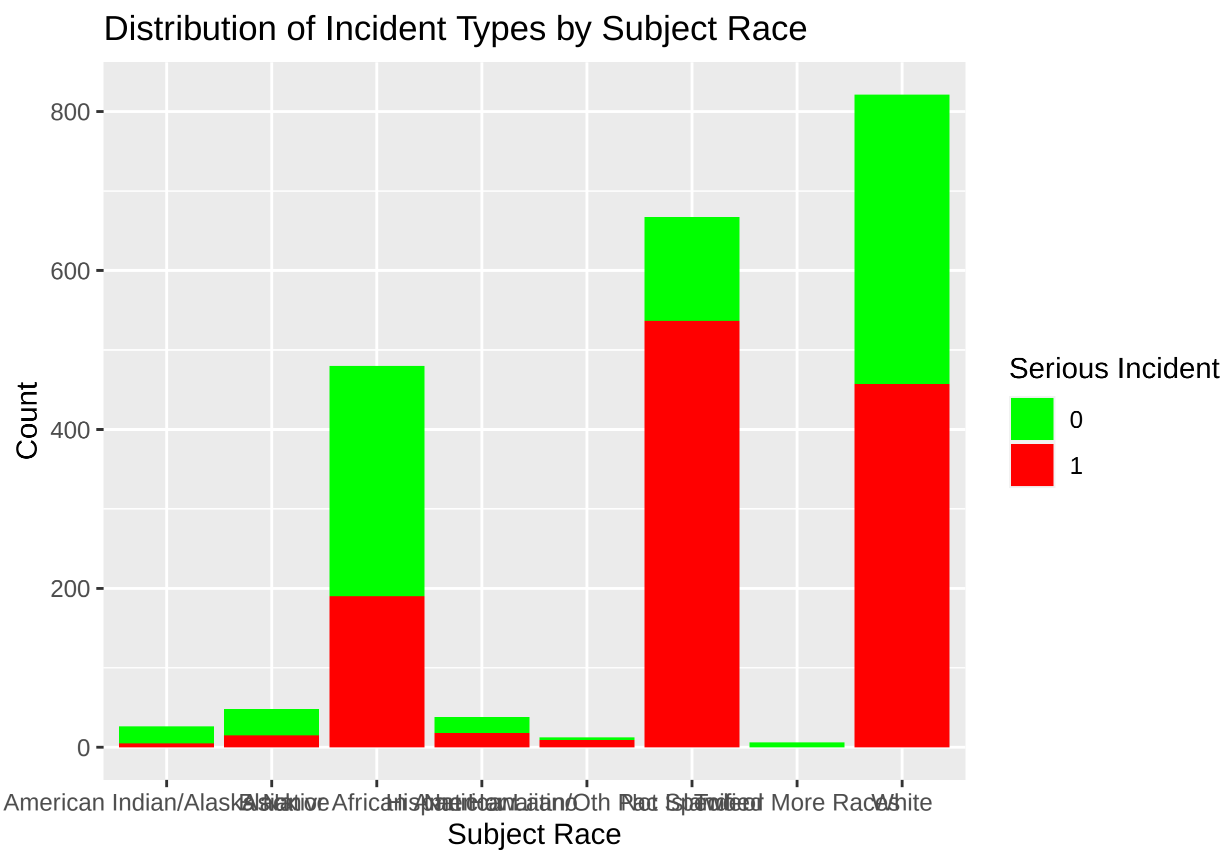
Seattle, Washington

4. Distribution of Incident Types by Subject Race There is a huge amount of 'Not Specified' values for this data visualization during the year 2020.

White people have a shared count amount of serious incidents regarding use of force my the police department.

The next following group that stands out is 'Black or African American,' with level 1 - Use of Force being greater than the other levels.

```
ggplot(seattle_2020, aes(x = Subject_Race, fill = factor(serious))) +
  geom_bar(position = "stack") +
  labs(
    title = "Distribution of Incident Types by Subject Race",
    x = "Subject Race",
    y = "Count",
    fill = "Incident Type"
  ) +
  scale_fill_manual(values = c("0" = "green", "1" = "red"), name = "Serious Incident")
```

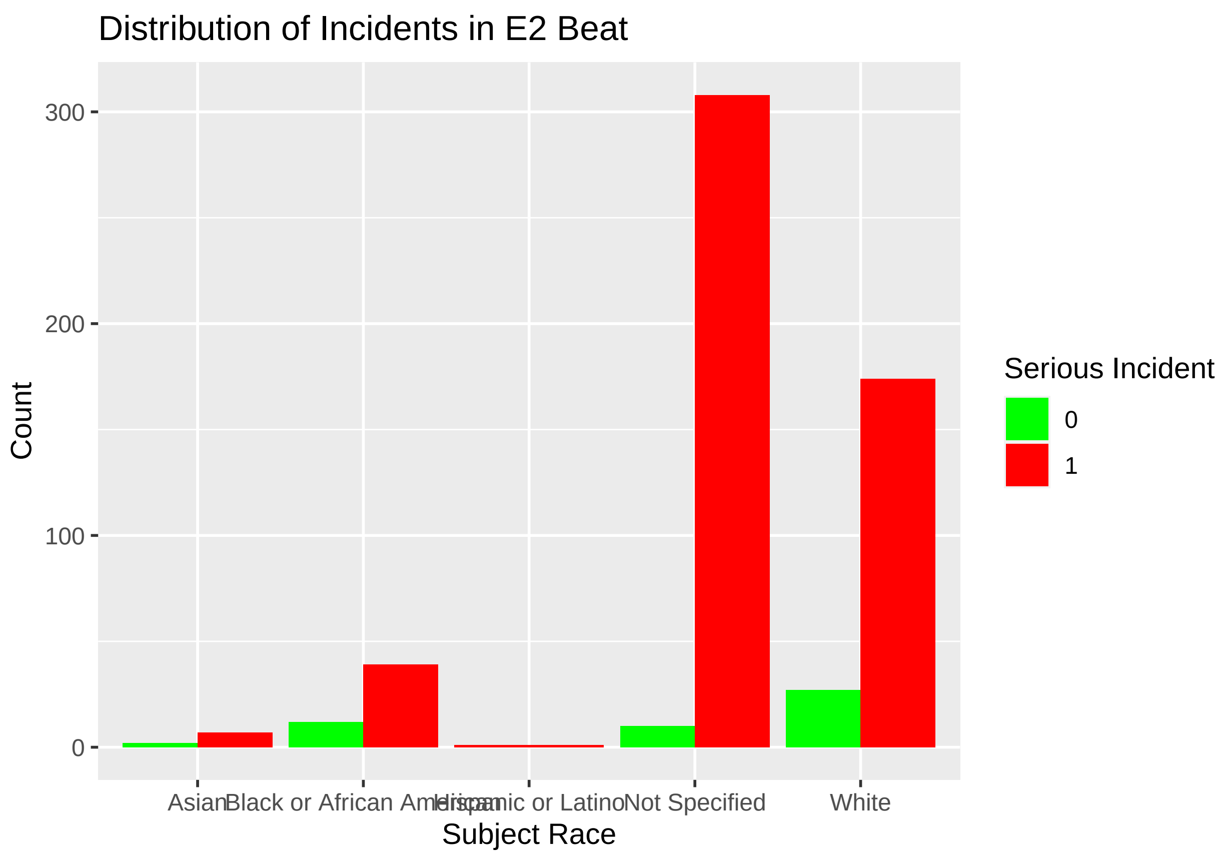


5. Comparative Analysis: Incidents in E2 by Subject Race E2 has the highest amount of incidents across all the other beats in Seattle. There are a total of 52 beats reported in the dataset.

Many of the 'Subject Race' is left unreported or 'Not Specified' in the year 2020. The next leading race that are subjected to incidents in the E2 beat are White, then it is Black or African American.

It is notable to mention that many of the serious incidents are reported as 'Not Specified' in the E2 beat.

```
e2_data <- subset(seattle_2020, Beat == "E2")

ggplot(e2_data, aes(x = Subject_Race, fill = factor(serious))) +
  geom_bar(position = "dodge", stat = "count") +
  labs(
    title = "Distribution of Incidents in E2 Beat",
    x = "Subject Race",
    y = "Count",
    fill = "Serious Incident"
  ) +
  scale_fill_manual(values = c("0" = "green", "1" = "red"), name = "Serious Incident")
```



Indianapolis, Indiana

Portland, Oregon