AdamDataExp

Adam Billen

10/15/2023

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.4.3     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggthemes)
library(ggplot2)
library(tibble)
```

In my mind, the most interesting variables are likely: -IncidentDistrict -uof_type -IncidentDate -Disposition -department_action -OfficerGender -OfficerRace -subject_age -subject_race -subject_gender -serious -CD (civil disturbance)

In my mind, we should remove/ignore: -IncidentTime -CaseStatus (I don't think whether the case is close should be an interesting predictor of anything) -DateClosed (Again, how long ago the case closed seemingly doesn't tell us much) -Officer_id (unless we want to look at individual officers we can ignore this) -OfficerAssignment (I could be wrong on this one) -animal (we should just remove uses of force against animals)

Hypotheses: We hypothesize that Black people/African Americans are over represented in use of force cases in Washington, DC. They represented 45% of the DC population in 2021, so if $> 45\%$ of the subjects of uses of force in our data are Black/African American, they would be over represented. We also hypothesize that white police officers are over represented in use of force cases in Washington, DC. As of 2021, 34.55% of officers were white. If the race of officers in use of force cases $> 34.55\%$, this would indicate overrepresentation.

We are particularly interested in predicting the "disposition," or how the department classified a use of force, based on the variables available to us. We hypothesize that uses of force labeled unjustified may be more common in wards 7 and 8 that are more diverse and poorer. The location variable which is available, "IncidentDistrict," is police district rather than ward, but we can use the rough lines of the police district to approximate these areas. We could do the same type of prediction but for a binary variable like "serious," which represents the severity of a given use of force.

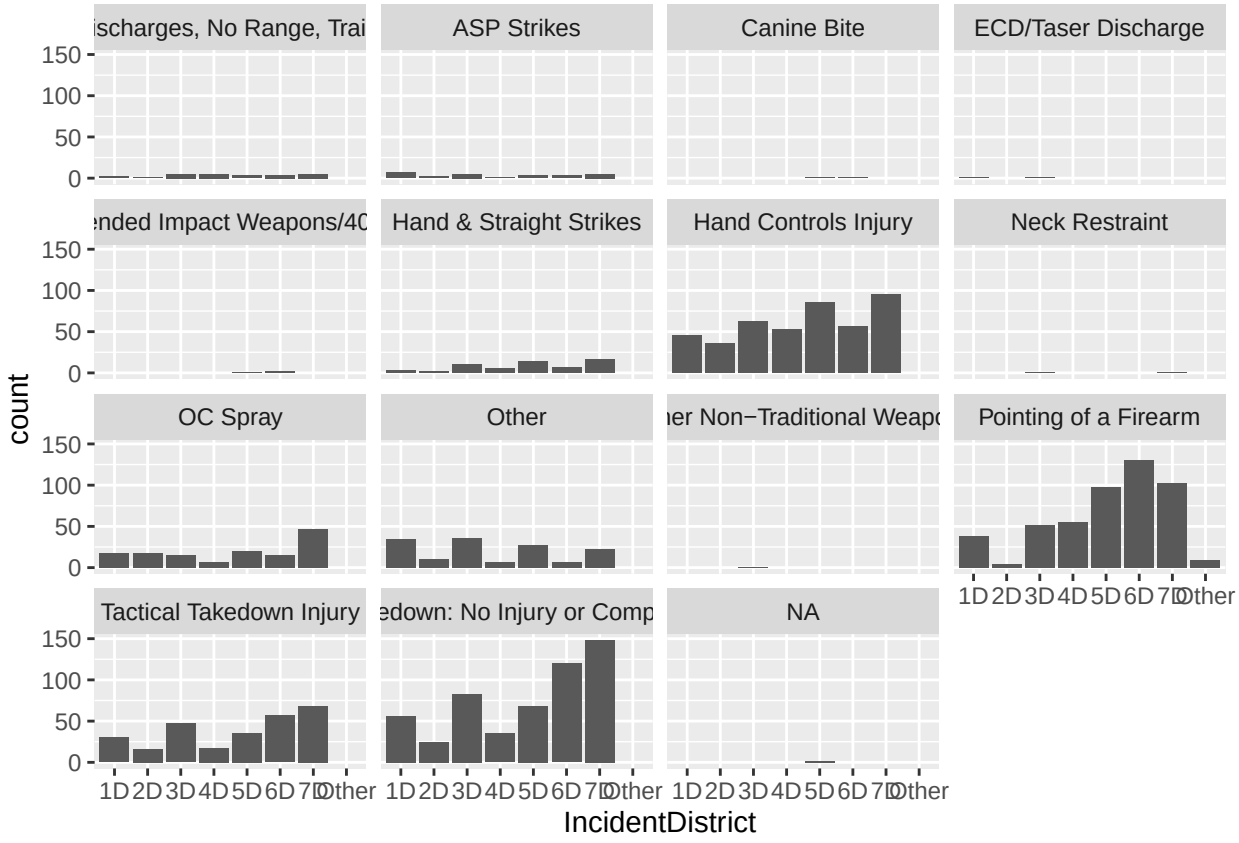
Methods: Our predictors for these models will be the police district in which the use of force took place, the date of the incident, the gender and race of the officer and the age, race, and gender of the subject. The target wil be the type of use of force, the disposition, the department action, or the severity of the force. We won't include the other potential targets as predictors for one another, because severity, for example, will likely be a very high predictor of the type of use of force, muddling our models. We will not include variables such as case status, when a case closed, the anonymized officer IDs, or officer assignments as predictors. We will remove all uses of force against animals, as we are only interested in cases involving humans.

```
mpd_dat <- read.csv('UoF_mpd_2021_public.csv')
head(mpd_dat)
```

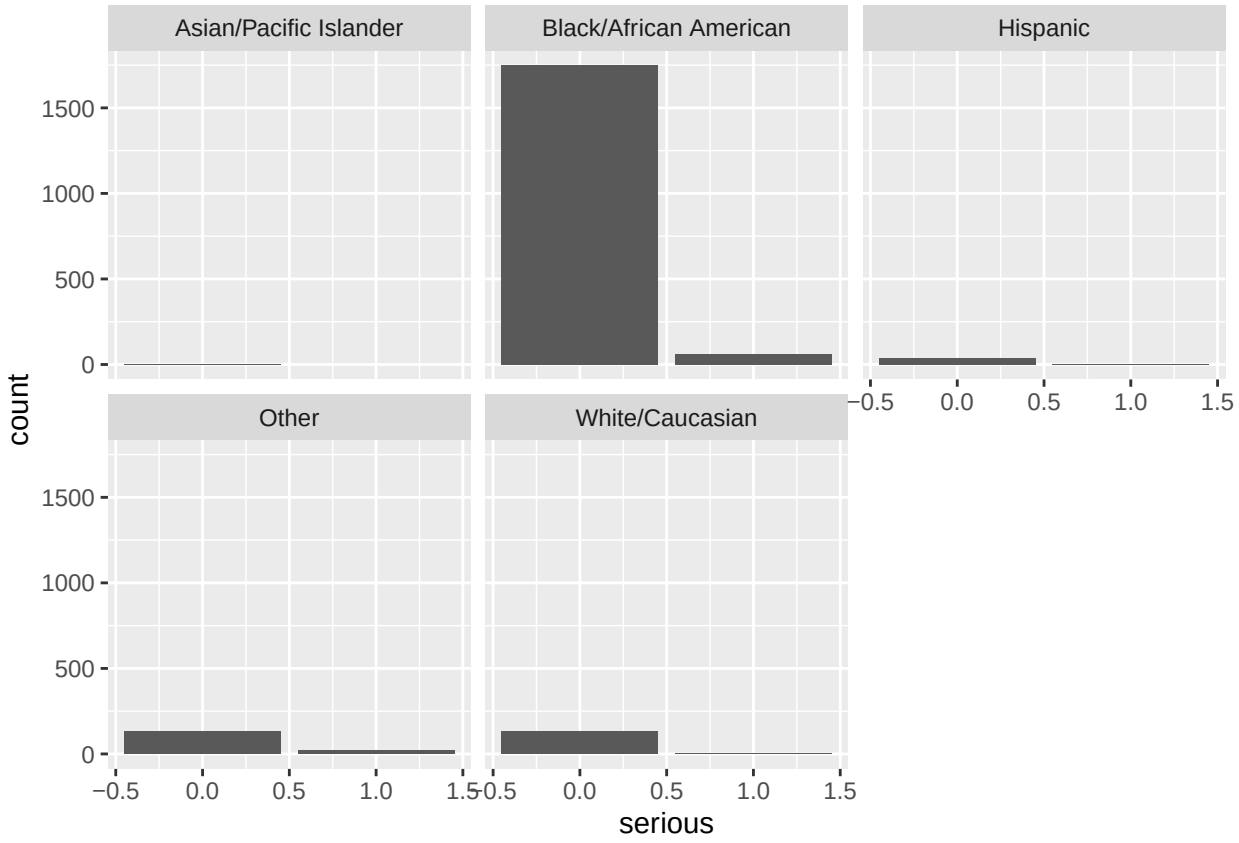
```
##                              IS_ID IncidentDate IncidentTime IncidentDistrict
## 1 ddc91c2470607ae729a0598904d4d9f3   2021-01-01     02:45:00               5D
## 2 ddc91c2470607ae729a0598904d4d9f3   2021-01-01     02:45:00               5D
## 3 ddc91c2470607ae729a0598904d4d9f3   2021-01-01     02:45:00               5D
## 4 ddc91c2470607ae729a0598904d4d9f3   2021-01-01     02:45:00               5D
## 5 ddc91c2470607ae729a0598904d4d9f3   2021-01-01     02:45:00               5D
## 6 ddc91c2470607ae729a0598904d4d9f3   2021-01-01     02:45:00               5D
##   CaseStatus DateClosed               uof_type
## 1     Closed 01/01/2022    Hand Controls Injury
## 2     Closed 01/01/2022    Hand Controls Injury
## 3     Closed 01/01/2022    Hand Controls Injury
## 4     Closed 01/01/2022    Hand Controls Injury
## 5     Closed 01/01/2022 Hand & Straight Strikes
## 6     Closed 01/01/2022    Hand Controls Injury
##                                  disposition         department_action
## 1 Justified - Tactical Improvement Opportunity         training referral
## 2 Justified - Tactical Improvement Opportunity performance documentation
## 3 Justified - Tactical Improvement Opportunity         training referral
## 4 Justified - Tactical Improvement Opportunity         training referral
## 5 Justified - Tactical Improvement Opportunity         training referral
## 6 Justified - Tactical Improvement Opportunity         training referral
##                          Officer_id OfficerAssignment OfficerGender
## 1 469fba90a485636efdee71f6d6841300                 5D          Male
## 2 68ad5fe43145b20dc953947775bdc233                 5D          Male
## 3 78b4b7169a68658581b7fd2880d8f466                 5D        Female
## 4 32ea1a3efcf01a49c0a70b793832c17b                 5D          Male
## 5 f4ef4d8e38bee0eac21eb26c99f979f9                 5D        Female
## 6 66911ba905047137d75401e8c7fb6032                 5D          Male
##              OfficerRace year_force subject_age           subject_race
## 1 Black/African American         15          31 Black/African American
## 2 Black/African American          5          31 Black/African American
```

```
## 3 Black/African American          4          31 Black/African American
## 4               Hispanic          1          31 Black/African American
## 5 Black/African American         14          31 Black/African American
## 6 Asian/Pacific Islander          2          31 Black/African American
##   subject_gender serious CD animal
## 1           Male       1  0      0
## 2           Male       1  0      0
## 3           Male       1  0      0
## 4           Male       1  0      0
## 5           Male       1  0      0
## 6           Male       1  0      0
```

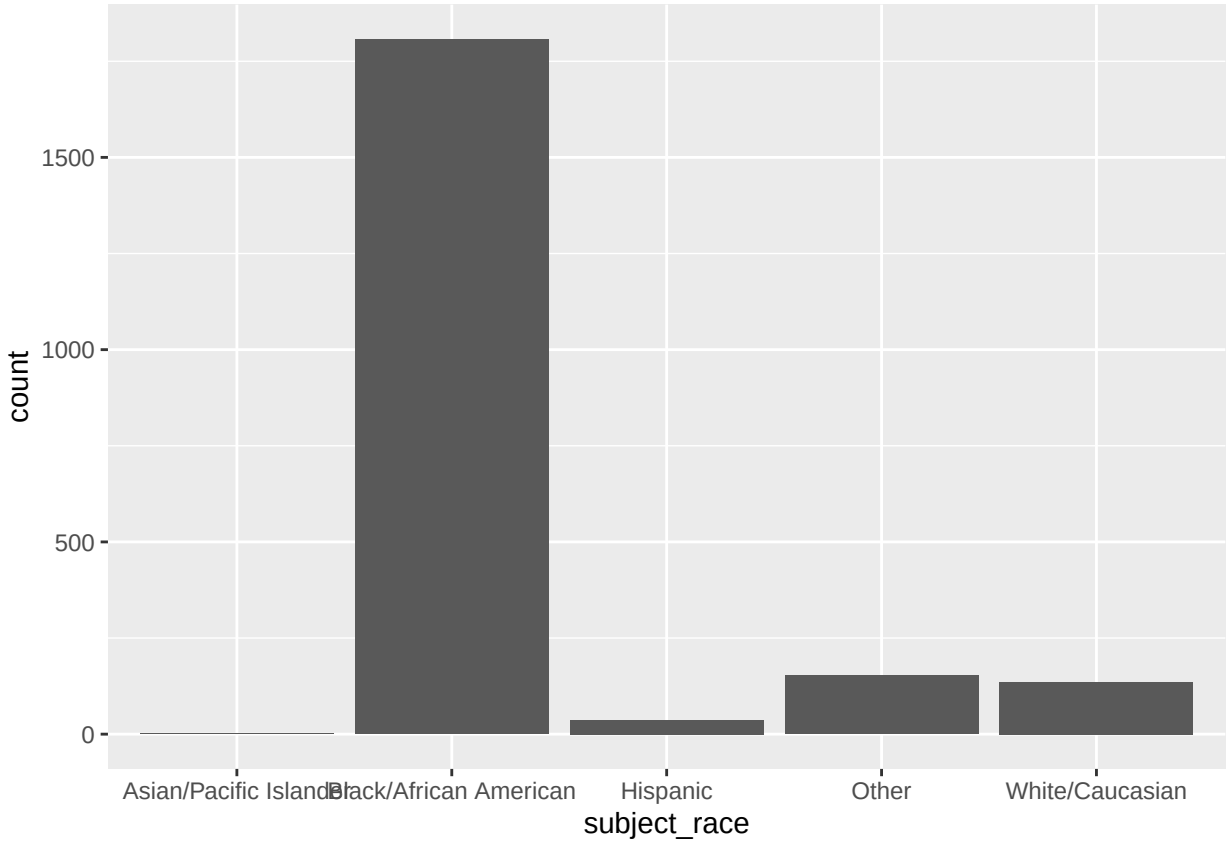
```
View(mpd_dat)

ggplot(data = mpd_dat, mapping = aes(x = IncidentDistrict)) +
  geom_bar() +
    facet_wrap(~uof_type)
```



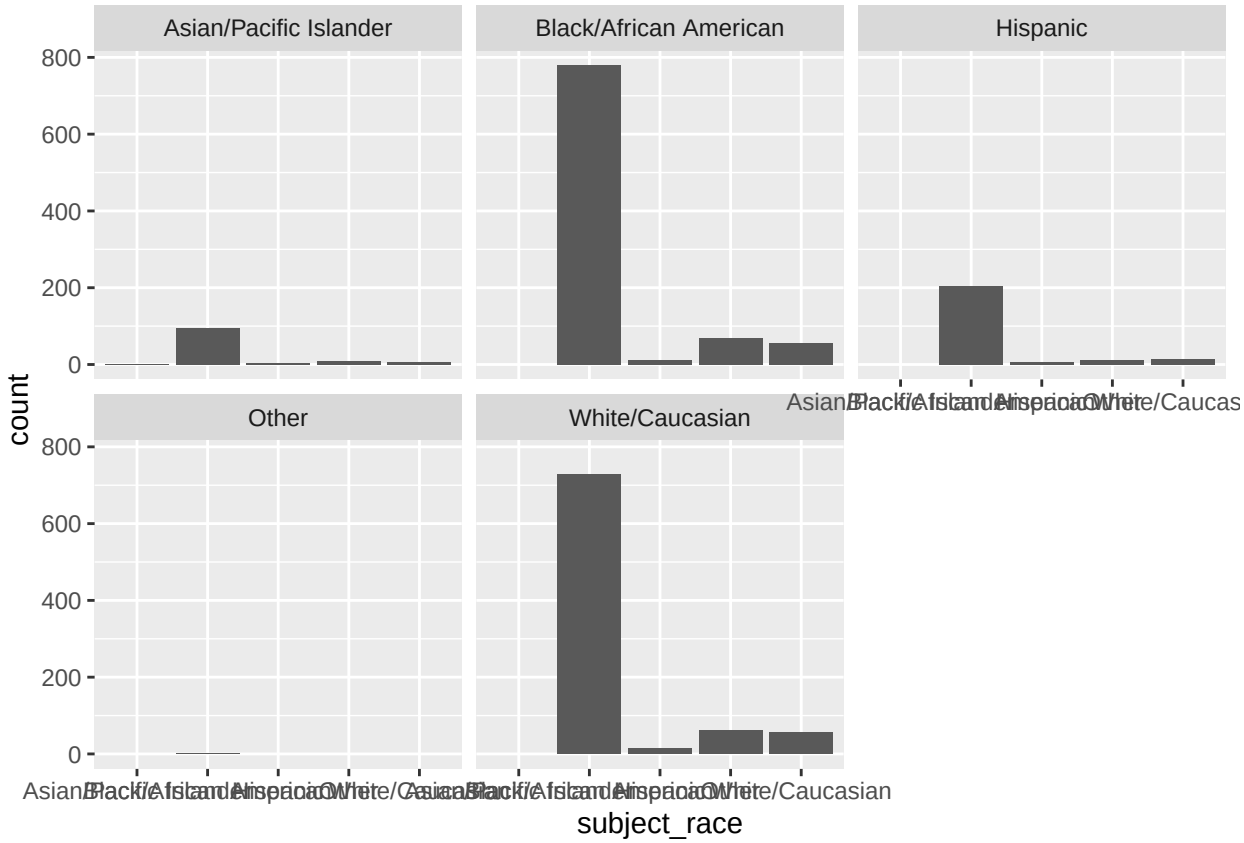

```
ggplot(data = mpd_dat, mapping = aes(x = serious)) +
  geom_bar() +
    facet_wrap(~subject_race)
```




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




```
ggplot(data = mpd_dat, mapping = aes(x = subject_race)) +
  geom_bar() +
    facet_wrap(~OfficerRace)
```


Asian/Pacific Islander
Black/African American
Hispanic
Other
White/Caucasian
count
800
600
400
200
0
subject_race