MPD Use of Force Proposal

Group: Kaylhan Garcia, Adam Billen, Phoebe _____

2023-10-15

Contents

1   MPD Use of Force Final Project Proposal

2   Introduction

Police have existed in America in some form for nearly its entire history. The practice's origin has its roots in Patrick Colquhoun's 1797 "Treatise on the Police of the Metropolis," the ideas of which became reality in London in 1829. Colquhoun himself, though, spent his early career as a British agent for cotton manufacturers. His ideas of the law and policing were born of his time spent enforcing slave codes and wielding 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. 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. The modern American police force began and grew from these dark roots of American history. (I wanted some sort of introductory paragraph/early context for policing itself, but we should probably add more recent context as well, and if you guys think this is too far back that's ok too)

We are interested in studying the use of force in Washington, DC, not far from Virginia where Colquhoun spent his early career. Our data stems from

The existing literature 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 likely to be killed by police in their lifetime than white men and white women, respectively. 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 our of 100,000 white men and boys being killed. 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.

The Metrpolitan Police Department, Washington, DC's police force, is unique in a number of ways. First, 51% of its officers were Black and 34% white in 2021 compared to the 46.2% of residents who are white and

25% who are Black. This makes it one of the few police forces in the US where white officers are not over represented as compared to the community. This is commonly thought of as a positive for police forces, and studies have found that when police diversity matches a neighborhood, crime rates tend to be lower.

(could look at uses of force of white officers in majority black neighborhoods vs black officers' use of force on those areas)

The police districts listed here are roughly analogous to DC's ward system. Each district's corresponding ward is listed below. -District 1 = ward 6 -District 2 = wards 2 & 3 -District 3 = ward 1 -District 4 = ward 4 -District 5 = ward 5 -District 6 = ward 7 -District 7 = ward 8

Sources: https://www.newyorker.com/magazine/2020/07/20/the-invention-of-the-police#:~:text=It%20is%20also%20often% https://www.pnas.org/doi/10.1073/pnas.1821204116 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 2021 use of force report: https://policecomplaints.dc.gov/sites/default/files/dc/sites/ office%20of%20police%20complaints/publication/attachments/2021%20OPC%20UOF%20Report_FINAL. pdf https://www.washingtonpost.com/nation/2020/06/04/urban-areas-police-are-consistently-much- whiter-than-people-they-serve/ https://www.census.gov/quickfacts/fact/table/DC/PST045222 UMD police match diversity: https://academic.oup.com/socpro/article/62/1/93/2461537?login=true

```
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)
# https://mpdc.dc.gov/node/1635896
df <- read.csv("UoF_mpd_2021_public.csv")
#df
```

3 MPD Use of Force Data Project

```r
library(conflicted)
library(dplyr)
library(tidyverse)
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 % of white officers in use of force cases $> 34.55\%$, this would indicate over-representation.

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.
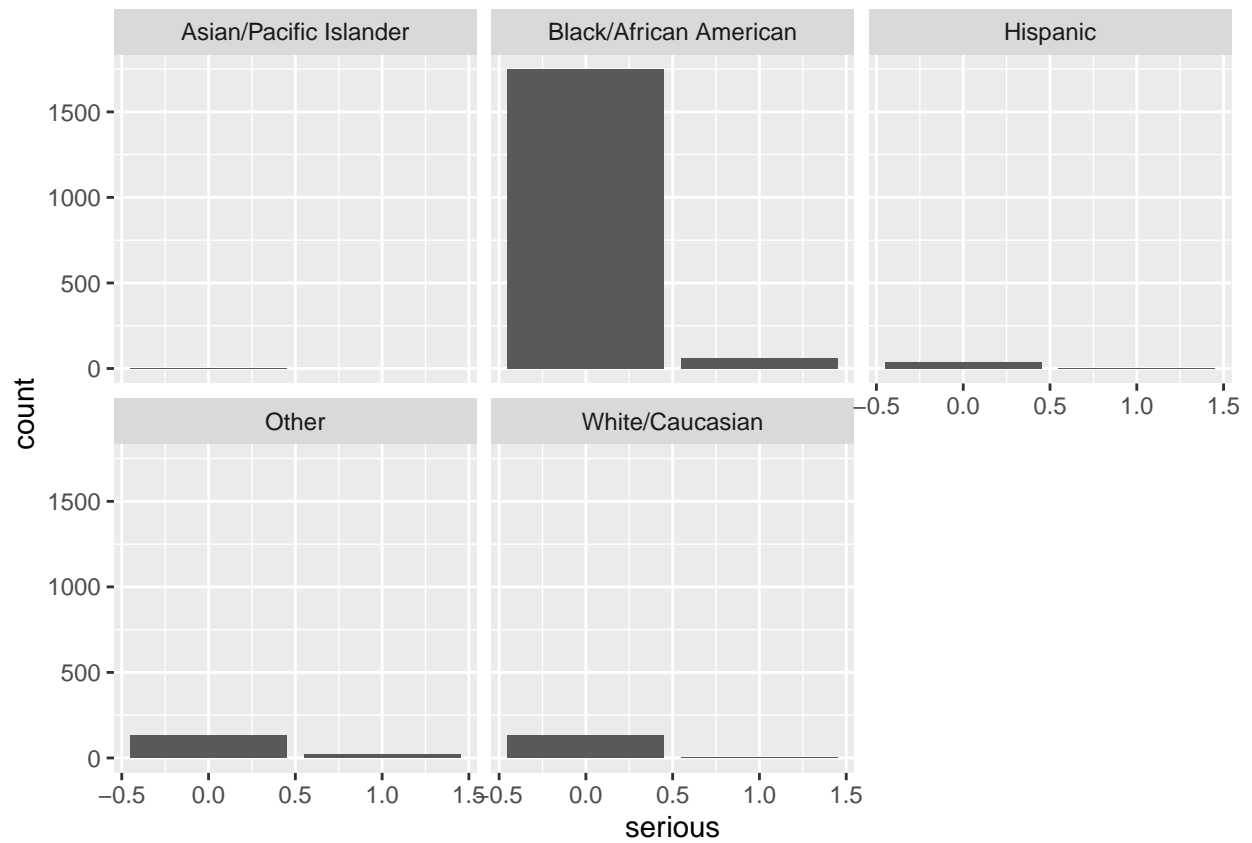
Datawise we should look at phoebe's article to find confounding/overlapping stuff

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

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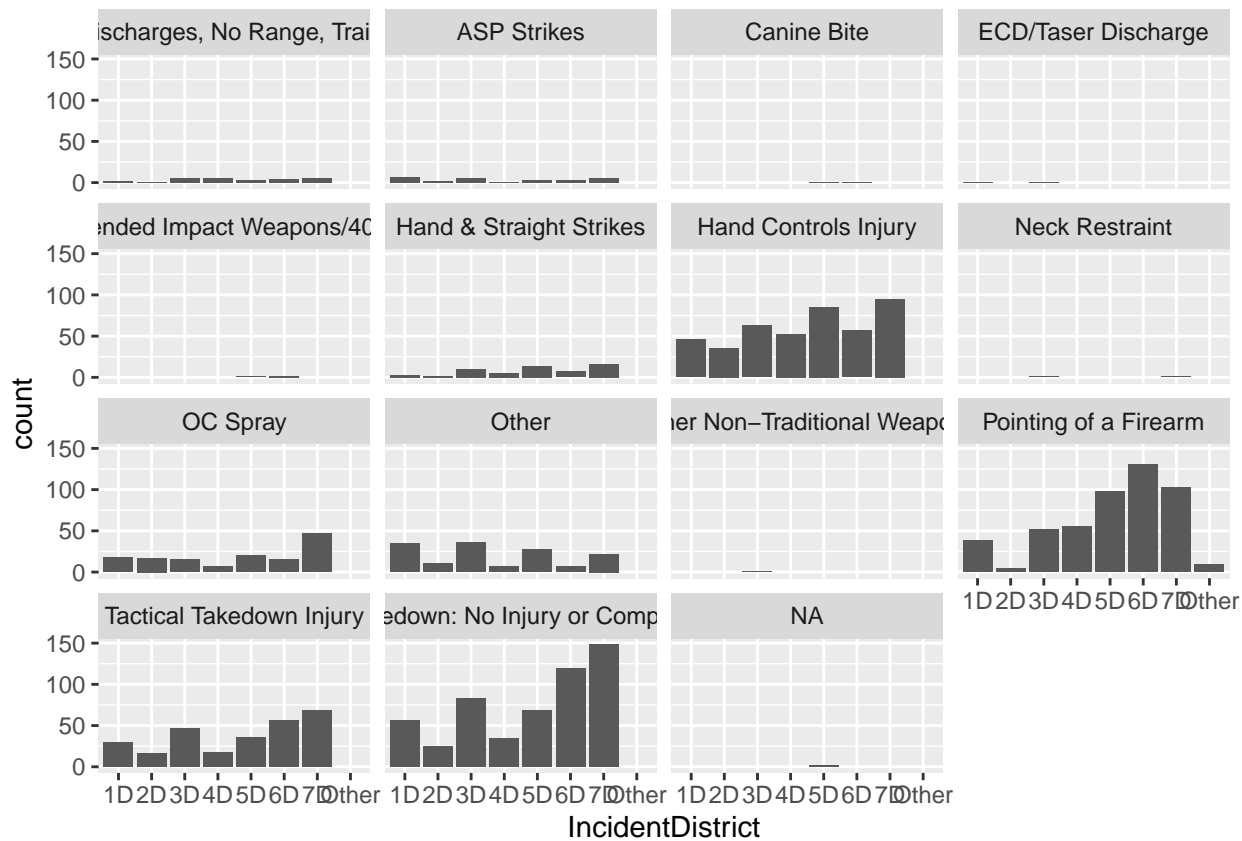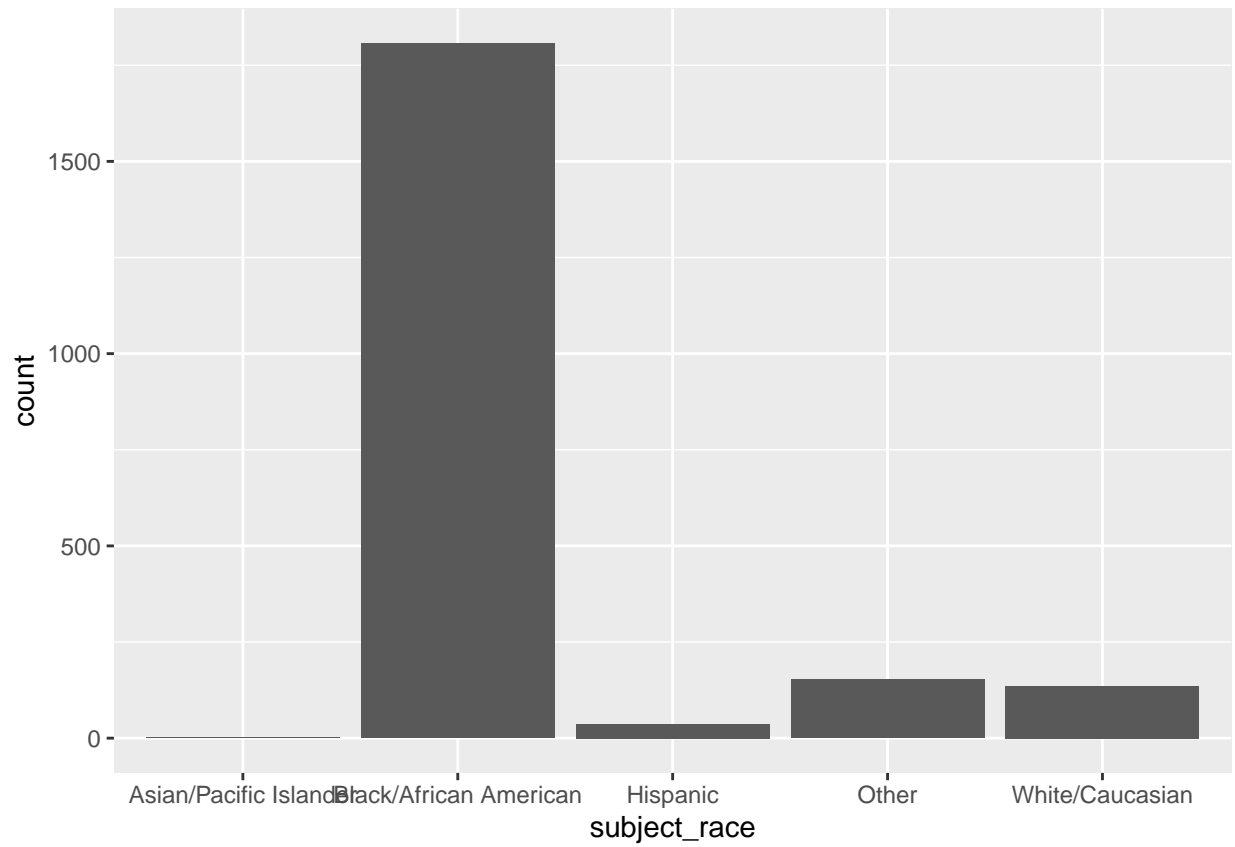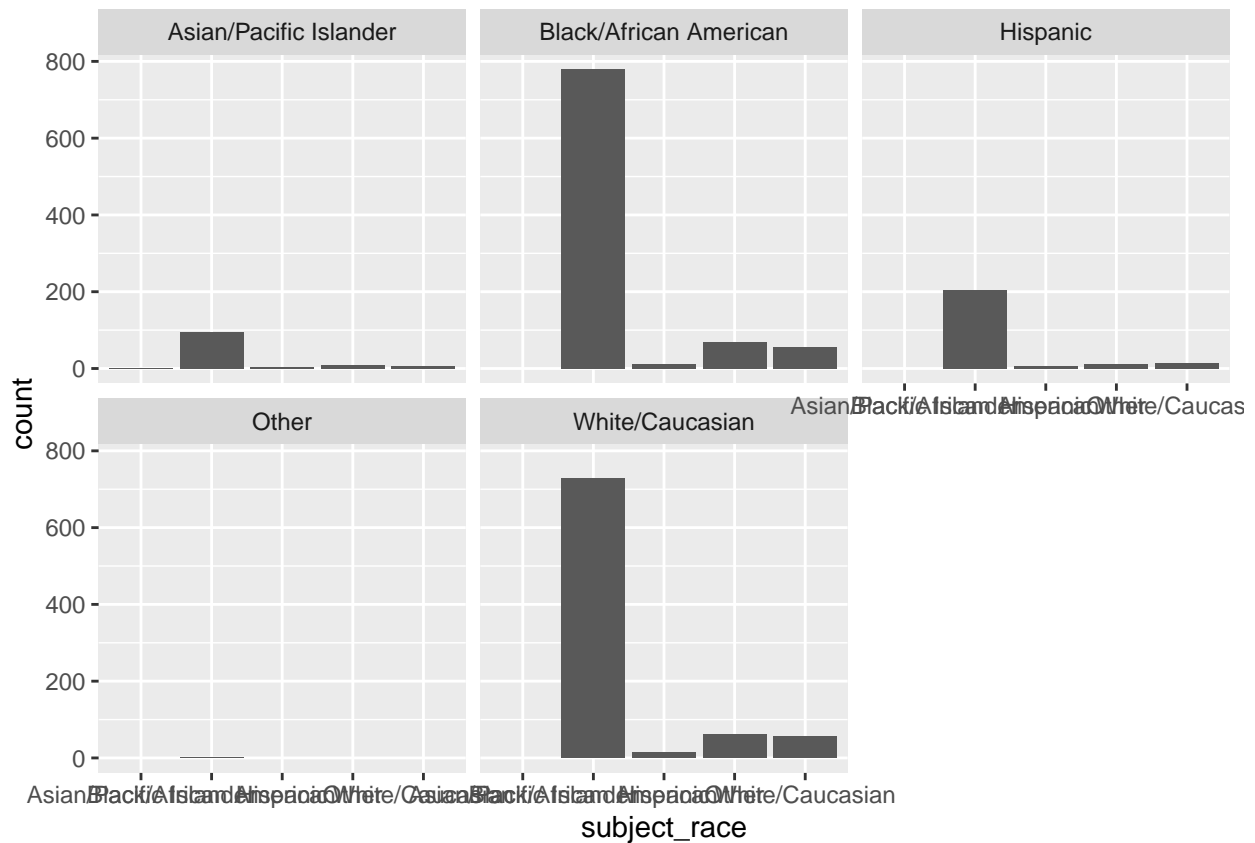
```r
ggplot(data = mpd_dat, mapping = aes(x = serious)) +
  geom_bar() +
    facet_wrap(~subject_race)
```

```
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)
```

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

```
## '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")
```

Use of Force Incidents by District
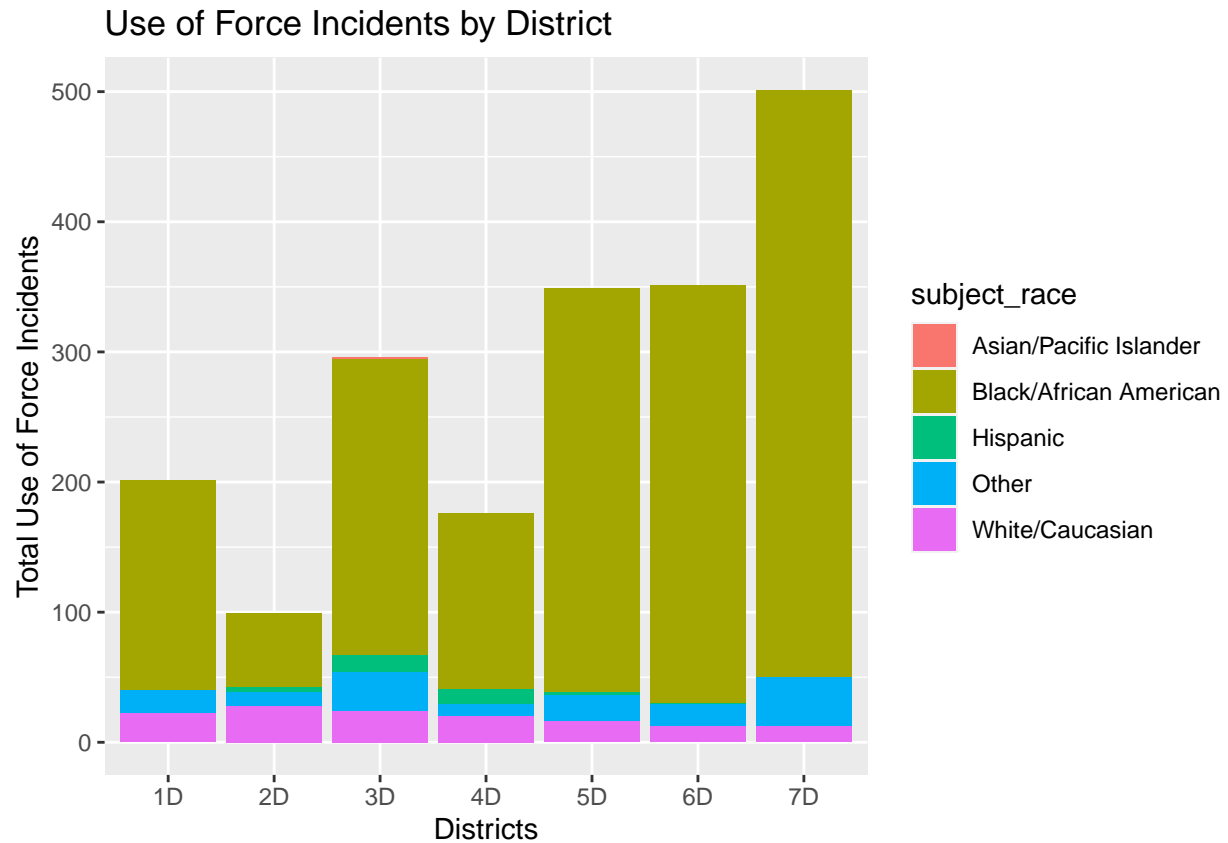

```r
#df with officer assignment, disposition, officer race
officer_by_district <- mpd_dat %>%
  group_by(OfficerAssignment, disposition, OfficerRace)


# Create a ggplot bar chart
ggplot(subset(officer_by_district, OfficerAssignment %in% c("1D", "2D", "3D", "4D", "5D", "6D", "7D")),
       aes(x = OfficerAssignment, y = OfficerRace, fill = disposition)) +
  geom_bar(stat = "identity", position = "dodge") +  # Create a bar chart
  labs(title = "Use of Force Disposition by Officers District",
       x = "Districts",
       y = "Disposition") +
  facet_wrap(~OfficerAssignment, ncol = 2)
```

Use of Force Disposition by Officers District
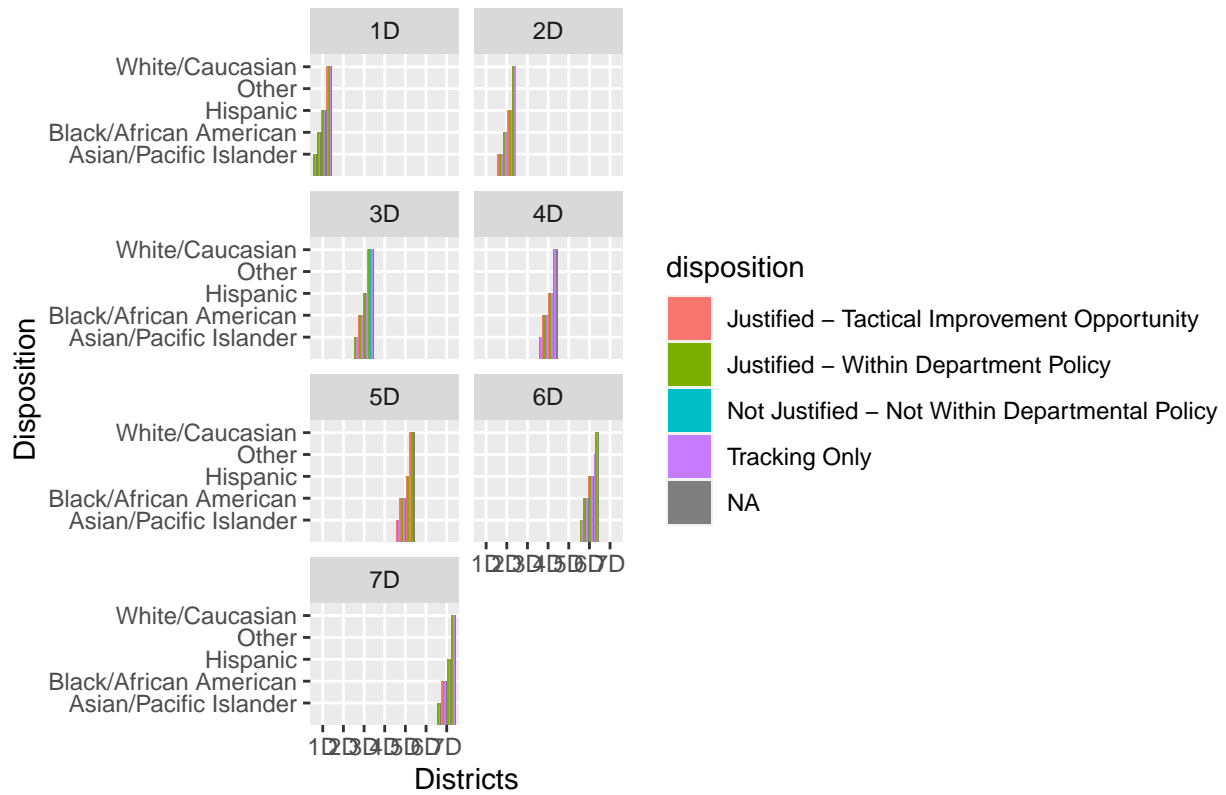

```
ggplot(mpd_dat, aes(x = disposition, fill = OfficerAssignment %in% c("1D", "2D", "3D", "4D", "5D", "6D"
  geom_bar() +
  labs(title = "Dispositions for Each District") +
  xlab("Dispositions") +
  ylab("District")
```

Dispositions for Each District