FinalProject part 1

Maaz

2023-02-28

The following packages were used

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.0     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.1     ✓ tibble    3.1.8
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the  ]8;;http://conflicted.r-lib.org/ conflicted package ]8;;  to force all conflicts t
o become errors
```

```
library(ggplot2)
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(dplyr)
library(lubridate)
```

Data import

```
urlfile="https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
Project<-read_csv(url(urlfile))
```

```
## Rows: 25596 Columns: 19
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## 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.
```

Data cleaning

```
NYPD<-Project %>%
  rename("Date"=OCCUR_DATE, "Time" = OCCUR_TIME) %>%
  select(-c(INCIDENT_KEY, PRECINCT, JURISDICTION_CODE,
            X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat, STATISTICAL_MURDER_FLAG, LOCAT
ION_DESC))
```

Changing the variable class

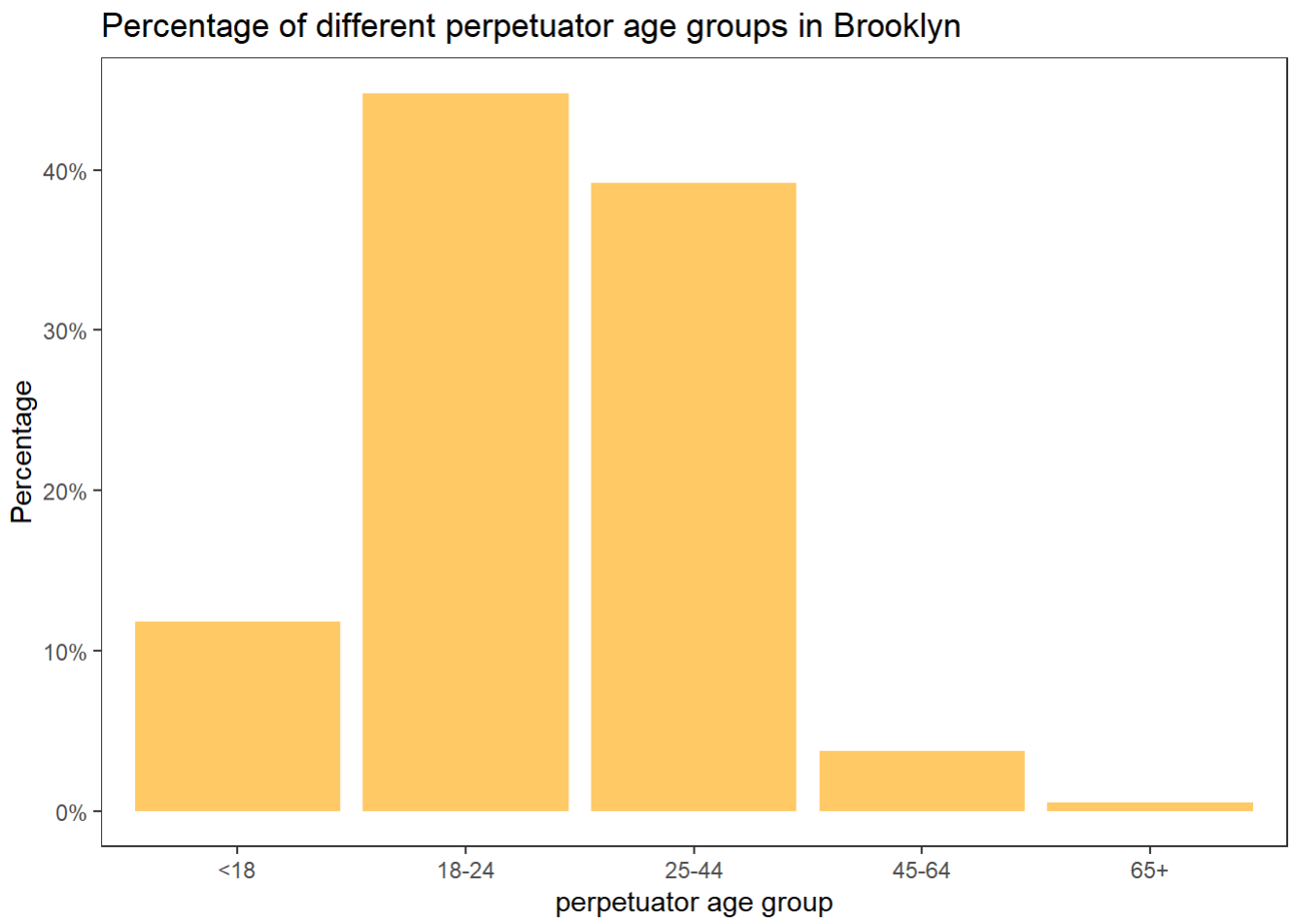
```
NYPD$PERP_AGE_GROUP<-as.factor(NYPD$PERP_AGE_GROUP)
NYPD$PERP_AGE_GROUP<-factor(NYPD$PERP_AGE_GROUP, levels = c("<18", "18-24","25-44","45-64","65
+"))
NYPD$VIC_AGE_GROUP<-as.factor(NYPD$VIC_AGE_GROUP)
NYPD$VIC_AGE_GROUP<-factor(NYPD$VIC_AGE_GROUP, levels = c("<18", "18-24","25-44","45-64","65+"))
```

Making a new dataset

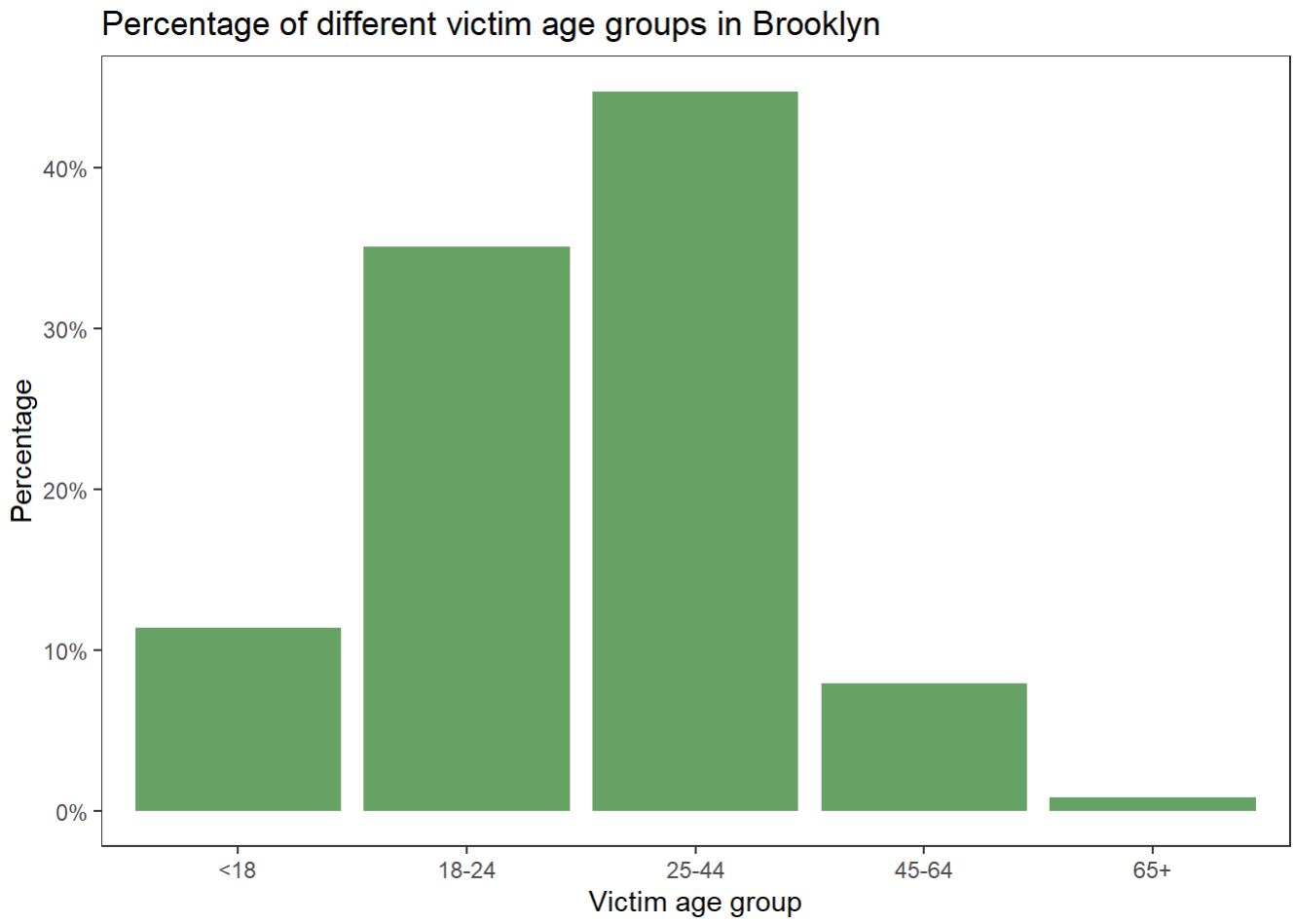
```
Brooklyn<-NYPD %>%
  filter(PERP_AGE_GROUP%in% c("<18", "18-24","25-44","45-64","65+"), BORO == "BROOKLYN", VIC_AGE
_GROUP%in% c("<18", "18-24","25-44","45-64","65+"))
```

Graphs

```
Brooklyn %>%
  ggplot(aes(PERP_AGE_GROUP, after_stat(count/sum(count))))+
  geom_bar(position="dodge", alpha=.6, fill="orange")+
  theme_bw()+
  scale_y_continuous(labels=scales::percent)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())+
  labs(title="Percentage of different perpetuator age groups in Brooklyn", x= "perpetuator age g
roup", y="Percentage")
```




```
Brooklyn %>%
  ggplot(aes(VIC_AGE_GROUP, after_stat(count/sum(count))))+
  geom_bar(position="dodge", alpha=.6, fill="darkgreen")+
  theme_bw()+
  scale_y_continuous(labels=scales::percent)+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())+
  labs(title="Percentage of different victim age groups in Brooklyn", x= "Victim age group", y
="Percentage")
```



Analysis

The graphs display the distribution of victims and perpetrators by age group in Brooklyn. As illustrated, the majority of both victims and perpetrators are in the age range of 18-44, with a greater concentration of victims between 24-44. The other age groups have a similar representation. Given that all the variables in the study are categorical, no regression model can be created.

Bias

One of the reasons why I decided to go with the age group for this study was to avoid any potential biases based on factors such as sex or race. Majority of the subjects in this study fall under my age category so I don't think there's any bias related to my findings.