

# Shootings in NYC

James Boyd

2023-04-17

## NYPD Shooting Incident Data (Historic)

Let's import the data.

```
library(tidyverse)
library(lubridate)
url_nypd <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
shootings <- read_csv(url_nypd)
```

Let's take a peak at the data.

```
summary(shootings)
```

```
## INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min.   : 9953245   Length:27312   Length:27312   Length:27312
## 1st Qu.: 63860880  Class :character  Class1:hms     Class :character
## Median : 90372218  Mode  :character  Class2:difftime Mode  :character
## Mean   :120860536                      Mode  :numeric
## 3rd Qu.:188810230
## Max.   :261190187
##
## LOC_OF_OCCUR_DESC  PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
## Length:27312      Min.   : 1.00   Min.   :0.0000   Length:27312
## Class :character  1st Qu.: 44.00  1st Qu.:0.0000   Class :character
## Mode  :character  Median : 68.00  Median :0.0000   Mode  :character
##                      Mean   : 65.64   Mean   :0.3269
##                      3rd Qu.: 81.00   3rd Qu.:0.0000
##                      Max.   :123.00   Max.   :2.0000
##                      NA's    :2
## LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
## Length:27312      Mode :logical      Length:27312
## Class :character  FALSE:22046         Class :character
## Mode  :character  TRUE :5266          Mode  :character
##
##
##
## PERP_SEX          PERP_RACE          VIC_AGE_GROUP      VIC_SEX
## Length:27312      Length:27312      Length:27312      Length:27312
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
```

```
##
##   VIC_RACE           X_COORD_CD       Y_COORD_CD       Latitude
## Length:27312      Min.   : 914928    Min.   :125757    Min.   :40.51
## Class :character   1st Qu.:1000028    1st Qu.:182834    1st Qu.:40.67
## Mode  :character   Median :1007731    Median :194487    Median :40.70
##                               Mean  :1009449    Mean  :208127    Mean  :40.74
##                               3rd Qu.:1016838    3rd Qu.:239518    3rd Qu.:40.82
##                               Max.   :1066815    Max.   :271128    Max.   :40.91
##                               NA's   :10
##
##   Longitude         Lon_Lat
## Min.   : -74.25      Length:27312
## 1st Qu.: -73.94      Class :character
## Median : -73.92      Mode  :character
## Mean   : -73.91
## 3rd Qu.: -73.88
## Max.   : -73.70
## NA's   :10
```

## Summary of Data

It appears that the shooting project data set is a distillation of who shot who, when, where in the city, and some other information, per shooting, such as precinct, jurisdiction code, descriptions of the area, ages of the shooters and those shot, ethnicities, and sexes. The clericality of the following analysis does no justice to the unfathomable horror behind each row of data.

According to the website from which the data is downloaded: “This is a breakdown of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year.”

Let’s clean up our dataset by:

- converting some categorical variables into factors.
- converting some characters strings, of dates, into Date objects.
- getting rid of unneeded columns.

```
shootings <- shootings %>%
  mutate(BORO = factor(BORO)) %>%
  mutate(VIC_SEX = factor(VIC_SEX)) %>%
  mutate(VIC_RACE = factor(VIC_RACE)) %>%
  mutate(VIC_AGE_GROUP = factor(VIC_AGE_GROUP)) %>%
  mutate(PERP_SEX = factor(PERP_SEX)) %>%
  mutate(PERP_RACE = factor(PERP_RACE)) %>%
  mutate(PERP_AGE_GROUP = factor(PERP_AGE_GROUP)) %>%
  mutate(PRECINCT = factor(PRECINCT)) %>%
  mutate(JURISDICTION_CODE = factor(JURISDICTION_CODE)) %>%
  mutate(days = factor(OCCUR_DATE)) %>%
  rename(date = 'OCCUR_DATE') %>%
  mutate(date = as.Date(date, format = "%m/%d/%Y")) %>%
  select(-c(INCIDENT_KEY, X_COORD_CD, Y_COORD_CD, Latitude, Longitude))

shootings <- shootings %>%
  mutate(year = format(as.Date(shootings$date, format="%m/%d/%Y"), "%Y"))
```

Let’s confirm that our transformations worked and that no data is missing.

```
summary(shootings)
```

```
##           date           OCCUR_TIME           BORO
```

```

## Min.      :2006-01-01   Length:27312   BRONX      : 7937
## 1st Qu.   :2009-07-18   Class1:hms   BROOKLYN   :10933
## Median    :2013-04-29   Class2:difftime MANHATTAN   : 3572
## Mean      :2014-01-06   Mode :numeric QUEENS      : 4094
## 3rd Qu.   :2018-10-15           STATEN ISLAND: 776
## Max.      :2022-12-31
##
## LOC_OF_OCCUR_DESC      PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
## Length:27312          75      : 1557   0      :22809      Length:27312
## Class :character      73      : 1452   1      : 74      Class :character
## Mode  :character      67      : 1216   2      : 4427   Mode  :character
##                               44      : 1020   NA's:    2
##                               79      : 1012
##                               47      : 953
##                               (Other):20102
## LOCATION_DESC          STATISTICAL_MURDER_FLAG PERP_AGE_GROUP      PERP_SEX
## Length:27312          Mode :logical      18-24 :6222   (null): 640
## Class :character      FALSE:22046          25-44 :5687   F      : 424
## Mode  :character      TRUE :5266          UNKNOWN:3148 M      :15439
##                               <18      :1591   U      : 1499
##                               (null) : 640   NA's   : 9310
##                               (Other): 680
##                               NA's     :9344
## PERP_RACE              VIC_AGE_GROUP      VIC_SEX
## BLACK                  :11432   <18      : 2839   F: 2615
## WHITE HISPANIC: 2341   1022      : 1      M:24686
## UNKNOWN                : 1836   18-24    :10086   U: 11
## BLACK HISPANIC: 1314   25-44    :12281
## (null)                 : 640   45-64    : 1863
## (Other)                : 439   65+      : 181
## NA's                   : 9310   UNKNOWN: 61
## VIC_RACE              Lon_Lat              days
## AMERICAN INDIAN/ALASKAN NATIVE: 10   Length:27312   07/05/2020: 47
## ASIAN / PACIFIC ISLANDER          : 404   Class :character 09/04/2011: 31
## BLACK                             :19439   Mode  :character 07/26/2020: 29
## BLACK HISPANIC                    : 2646           08/11/2007: 26
## UNKNOWN                           : 66           08/27/2022: 25
## WHITE                             : 698           09/04/2006: 25
## WHITE HISPANIC                    : 4049           (Other)   :27129
## year
## Length:27312
## Class :character
## Mode  :character
##
##
##
##

```

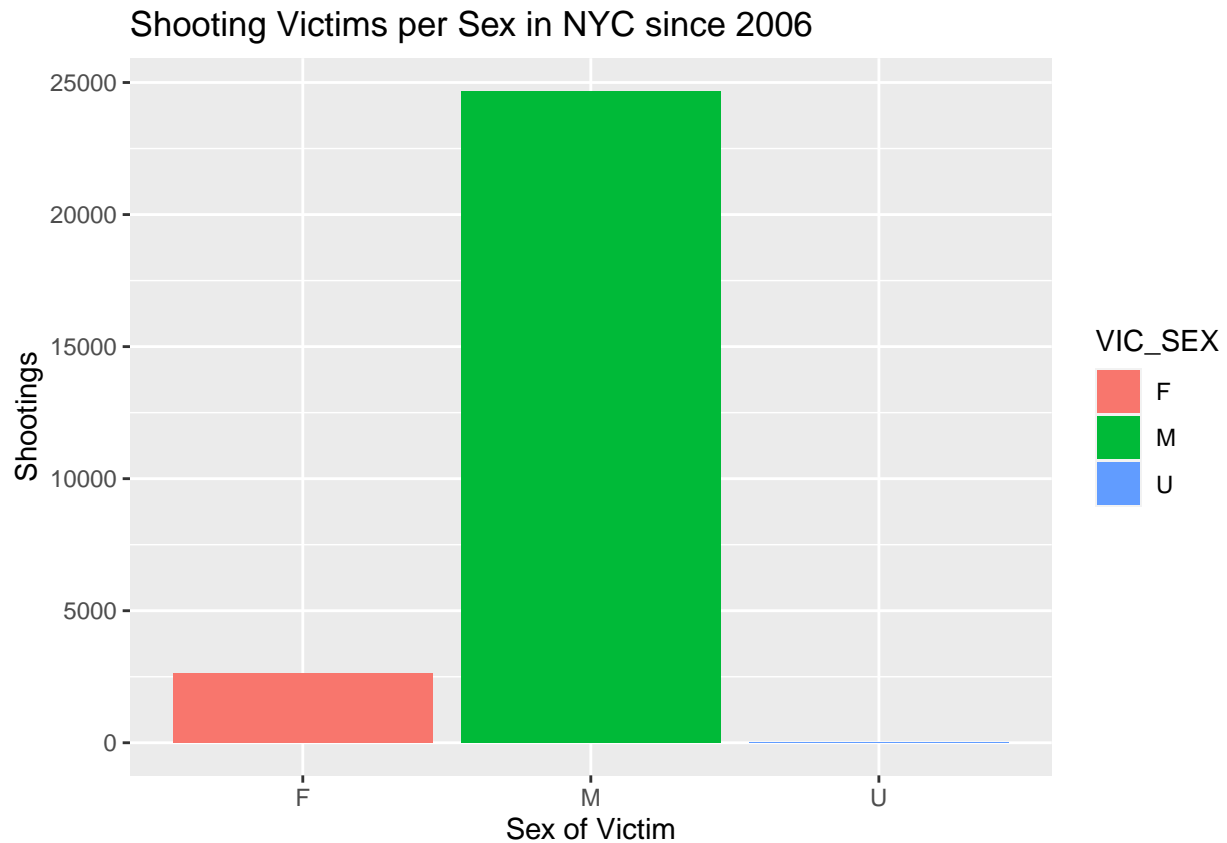
It seems that our transformations worked, and that no data is missing, except for null and NA values, which we'll keep because that information is significant, shining light on the unknowingness and chaos associated with shootings.

Since 2006, in NYC, which sex:

- committed the most shootings?

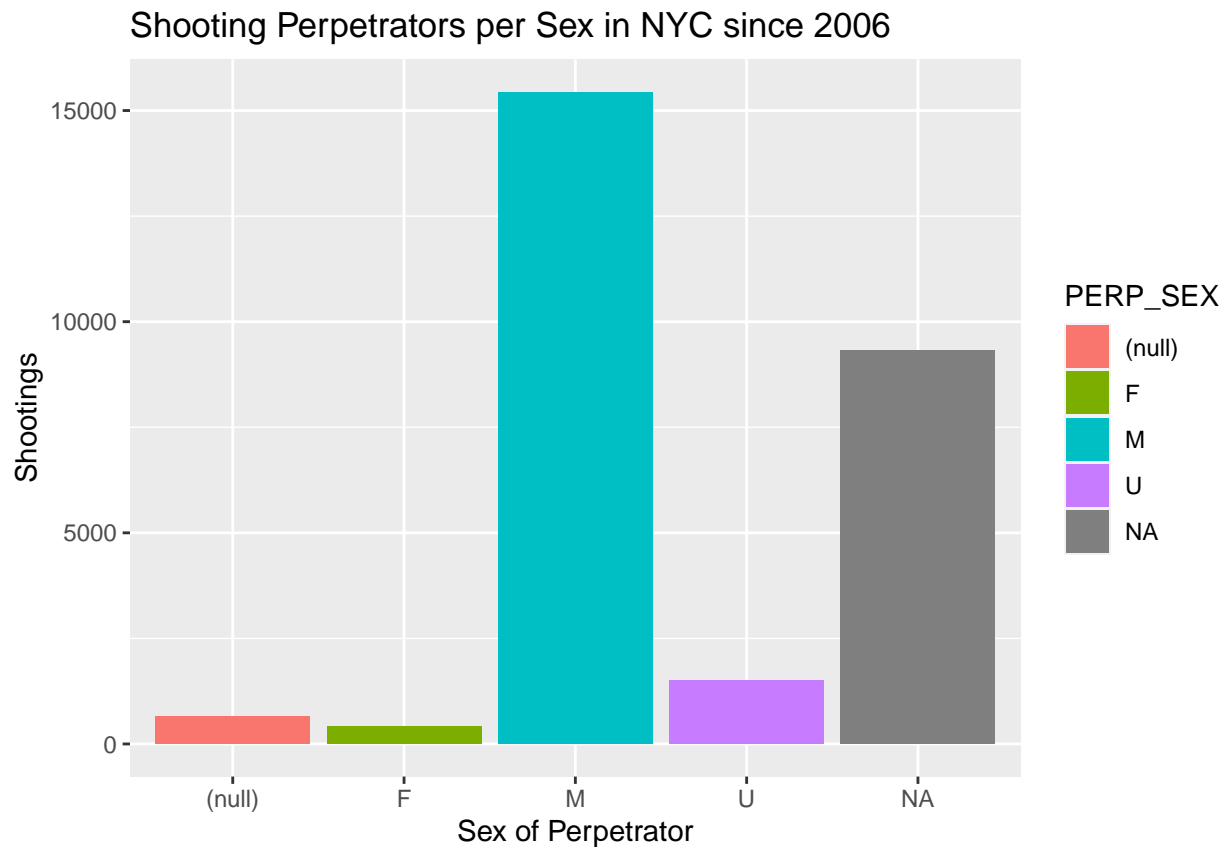
- was shot the most?

```
plot <- ggplot(shootings, aes(VIC_SEX, fill = VIC_SEX))
plot +
  geom_bar() +
  ggtitle("Shooting Victims per Sex in NYC since 2006") +
  xlab("Sex of Victim") +
  ylab("Shootings")
```



Males are by far the most shot sex in NYC, being shot almost nine times as often as females. The unknown segment is barely visible.

```
plot <- ggplot(shootings, aes(PERP_SEX, fill = PERP_SEX))
plot +
  geom_bar() +
  ggtitle("Shooting Perpetrators per Sex in NYC since 2006") +
  xlab("Sex of Perpetrator") +
  ylab("Shootings")
```

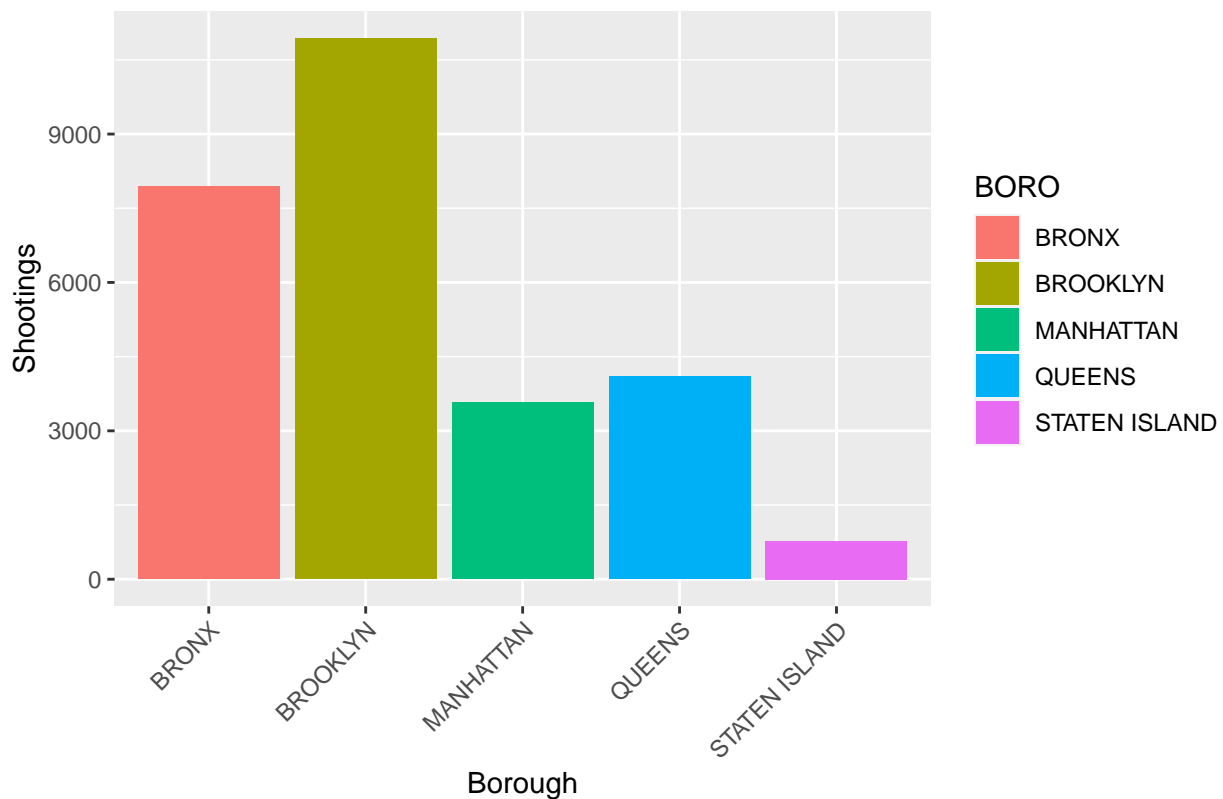


Males appear to account for the vast majority of the shooting perpetrators, as well.

Since 2006, in NYC, which borough had the most shootings?

```
plot <- ggplot(shootings, aes(BORO, fill = BORO)) +  
  geom_bar() +  
  ggtitle("Shootings per Borough since 2006") +  
  xlab("Borough") +  
  ylab("Shootings") +  
  theme(axis.text.x = element_text(angle=45, hjust=1))  
plot
```

### Shootings per Borough since 2006



```
summary(shootings['BORO'])
```

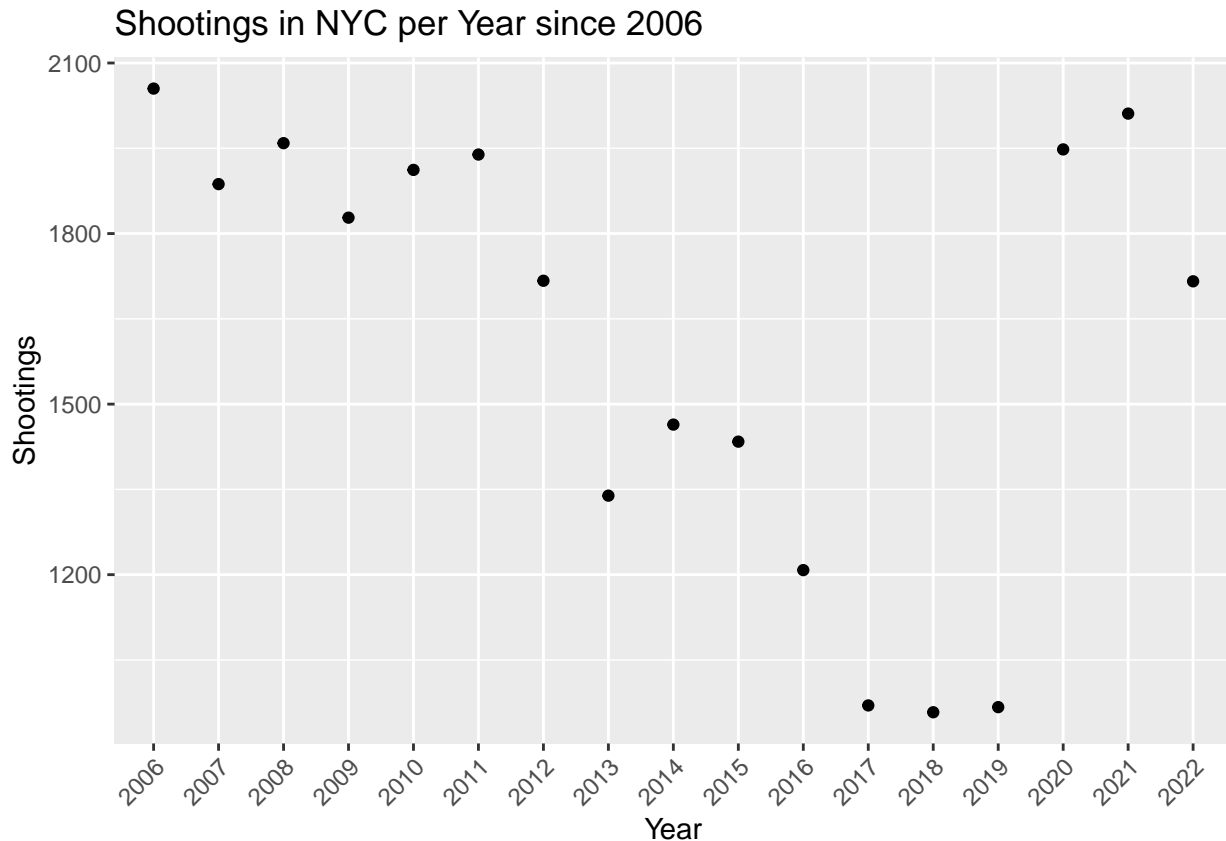
```
##          BORO
##  BRONX      : 7937
##  BROOKLYN   :10933
##  MANHATTAN  : 3572
##  QUEENS     : 4094
##  STATEN ISLAND: 776
```

It appears that Brooklyn has had the most shootings, since 2006, at 10,933 shootings. Staten Island had the least, at 776 shootings.

Have shootings in NYC gone up, or down, since 2006?

```
perYear <- shootings %>%
  group_by(year) %>%
  summarize(Tot = n())

plot <- ggplot(perYear, aes(year, Tot)) +
  geom_point() +
  ggtitle("Shootings in NYC per Year since 2006") +
  xlab("Year") +
  ylab("Shootings") +
  theme(axis.text.x = element_text(angle=45, hjust=1))
plot
```



It appears that shootings were on their way down, since 2006. Then, they spiked back up in 2020. It may have been related to COVID-19. However, 2022 was still fewer than any year between 2006 and 2011.

## Questions

Did COVID-19, or the circumstances surround it, cause a spike in shootings in NYC? What have people been doing right, specifically, to press the shooting count down? What was causing the shooting count to go down in 2006-2019? Why do men do so much shooting? Why do they get shot so much? These questions merit further research.

## Conclusion

In addition to the aforementioned questions that arise from this analysis, there are many more pieces of insight to be gained around this data. There are, in fact, professionals, working this particular subject, around the clock, in NYC. One particular question is what insight can be gained from this data that could reduce, or prevent, one or more shootings? Does lending a generous hand to one's neighbor help in trying times, such as COVID-19? Do men stand to benefit from considering taking a step back from violent situations, if given the chance, when those situations approach? It's a compelling data set.

Potential biases include those related to the unknowns in regards to the one-to-one mapping of perpetrators to victims. Are the counts actually more? Are they actually less? Is someone trying to inflate or deflate the numbers? Another potential point of bias is the racial labeling in this dataset, in a 23 and me world, in which many people, perhaps all people, are not just exclusively this ethnicity or that ethnicity. One personal bias of mine is that I'm all but in denial that this much violence is occurring in such a small land area. Part of me doesn't want to examine this subject. However, means by which I have addressed this bias are by facing these numbers as more of a scientist than a moralist, who wishes these numbers were lower or zero. This dataset will continue to be studied for decades.

```
sessionInfo()
```

```
## R version 4.2.3 (2023-03-15)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] lubridate_1.9.2 forcats_1.0.0  stringr_1.5.0  dplyr_1.1.1
## [5] purrr_1.0.1    readr_2.1.4    tidyr_1.3.0    tibble_3.2.1
## [9] ggplot2_3.4.2  tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] highr_0.10      pillar_1.9.0    compiler_4.2.3  tools_4.2.3
## [5] bit_4.0.5       digest_0.6.31   timechange_0.2.0 evaluate_0.20
## [9] lifecycle_1.0.3 gtable_0.3.3    pkgconfig_2.0.3 rlang_1.1.0
## [13] cli_3.6.1       rstudioapi_0.14 curl_5.0.0      parallel_4.2.3
## [17] yaml_2.3.7      xfun_0.38       fastmap_1.1.1   withr_2.5.0
## [21] knitr_1.42      generics_0.1.3  vctrs_0.6.1     hms_1.1.3
## [25] bit64_4.0.5     grid_4.2.3      tidyselect_1.2.0 glue_1.6.2
## [29] R6_2.5.1        fansi_1.0.4     vroom_1.6.1     rmarkdown_2.21
## [33] farver_2.1.1    tzdb_0.3.0      magrittr_2.0.3  scales_1.2.1
## [37] htmltools_0.5.5 colorspace_2.1-0 labeling_0.4.2   utf8_1.2.3
## [41] stringi_1.7.12  munsell_0.5.0   crayon_1.5.2
```