# A Look at a New York Shooting Database

## The following is Herbs RMD for the Shooting assignment - Data Science

## New York Shootings were downloaded from the following:

Downloaded the link below given in course material on 8 Nov 2023 for NYDP Shooting Incident Data (Historic): https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD A glimpse of the loaded data follows:

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"

gunshots <- read_csv(url_in)
```

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## 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.
```

```
summary(gunshots)
```

```
##   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
```

```
gunshots
```

```
## # A tibble: 27,312 x 21
##    INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     LOC_OF_OCCUR_DESC PRECINCT
##           <dbl> <chr>      <time>     <chr>    <chr>                <dbl>
##  1    228798151 05/27/2021 21:30      QUEENS   <NA>                   105
##  2    137471050 06/27/2014 17:40      BRONX    <NA>                    40
##  3    147998800 11/21/2015 03:56      QUEENS   <NA>                   108
##  4    146837977 10/09/2015 18:30      BRONX    <NA>                    44
##  5     58921844 02/19/2009 22:58      BRONX    <NA>                    47
##  6    219559682 10/21/2020 21:36      BROOKLYN <NA>                    81
##  7     85295722 06/17/2012 22:47      QUEENS   <NA>                   114
##  8     71662474 03/08/2010 19:41      BROOKLYN <NA>                    81
##  9     83002139 02/05/2012 05:45      QUEENS   <NA>                   105
## 10     86437261 08/26/2012 01:10      QUEENS   <NA>                   101
## # i 27,302 more rows
## # i 15 more variables: JURISDICTION_CODE <dbl>, LOC_CLASSFCTN_DESC <chr>,
## #   LOCATION_DESC <chr>, STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>,
## #   PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>,
## #   VIC_RACE <chr>, X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>,
## #   Longitude <dbl>, Lon_Lat <chr>
```

```
write_csv(gunshots, "gunshots.csv")
```

### Peek at whats in the database using class and glimpse.

```
class(gunshots)
```

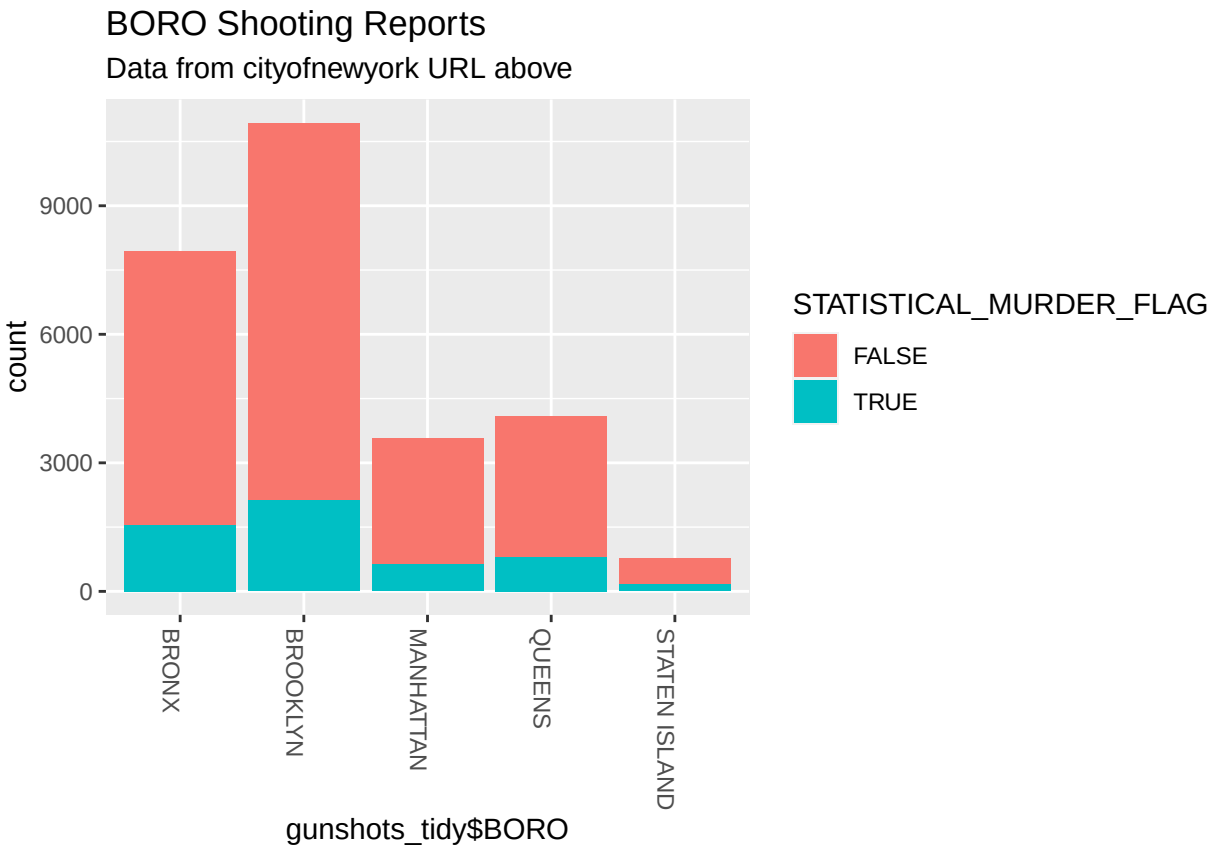
```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

```
gunshots_tidy <- gunshots %>%
  rename(
   date = "OCCUR_DATE",
   time = "OCCUR_TIME"
    )
```

##I only checked for missing in the MURDER FLAG col since that is the one col I am interested in. The other NA sun as location can be dealt with if needed by removal or placement in another table for anaylsis with out that location. Most the graphsd allow for missing data if that is what I was graphing.

### First peek at the data with bar charts...

```
gunshots_tidy |>
  ggplot(aes(x = gunshots_tidy$BORO, fill = STATISTICAL_MURDER_FLAG))+
  geom_bar()+
  theme(axis.text.x=element_text(angle = -90, hjust = 0)) +
  labs(
    title = "BORO Shooting Reports",
    subtitle="Data from cityofnewyork URL above"
  )
```




**So the above immediatly shows that there is a difference in areas reporting shootings and deaths–so I wanted to look further.**

```
freq = table(gunshots_tidy$LOCATION_DESC)
feq_df = as.data.frame(freq)
print(freq)
```

```
##
##                   (null)                      ATM                     BANK
##                      977                        1                        3
##           BAR/NIGHT CLUB       BEAUTY/NAIL SALON             CANDY STORE
##                      628                      112                        7
##              CHAIN STORE              CHECK CASH       CLOTHING BOUTIQUE
##                        5                        1                       14
##          COMMERCIAL BLDG              DEPT STORE          DOCTOR/DENTIST
##                      292                        9                        1
##              DRUG STORE      DRY CLEANER/LAUNDRY      FACTORY/WAREHOUSE
##                       14                       31                        8
##              FAST FOOD              GAS STATION          GROCERY/BODEGA
##                      104                       71                      694
##      GYM/FITNESS FACILITY                 HOSPITAL             HOTEL/MOTEL
##                        3                       65                       35
##            JEWELRY STORE              LIQUOR STORE             LOAN COMPANY
##                       12                       41                        1
```

```
##   MULTI DWELL - APT BUILD MULTI DWELL - PUBLIC HOUS                    NONE
##                      2835                      4832                     175
##          PHOTO/COPY STORE                 PVT HOUSE          RESTAURANT/DINER
##                         1                       951                     204
##                    SCHOOL                SHOE STORE            SMALL MERCHANT
##                         1                        10                      37
## SOCIAL CLUB/POLICY LOCATI          STORAGE FACILITY        STORE UNCLASSIFIED
##                        72                         1                      36
##               SUPERMARKET           TELECOMM. STORE             VARIETY STORE
##                        21                        11                      11
##               VIDEO STORE
##                         8
```

```
#see if locations have info and what it reveals

local = gunshots_tidy %>% dplyr::count(gunshots_tidy$LOCATION_DESC, gunshots_tidy$LOC_OF_OCCUR_DESC)
local = as.data.frame(local)
colnames(local) = c("location","desc","count")

local_sort = local[order(local$count,local$location, decreasing = T),]
local_sort
```

```
##                     location    desc count
## 77                      <NA>    <NA> 14977
## 50 MULTI DWELL - PUBLIC HOUS    <NA>  4559
## 47   MULTI DWELL - APT BUILD    <NA>  2664
## 2                     (null) OUTSIDE   945
## 55                 PVT HOUSE    <NA>   893
## 33            GROCERY/BODEGA    <NA>   622
## 7             BAR/NIGHT CLUB    <NA>   588
## 18           COMMERCIAL BLDG    <NA>   265
## 49 MULTI DWELL - PUBLIC HOUS OUTSIDE   226
## 58          RESTAURANT/DINER    <NA>   194
## 51                      NONE    <NA>   175
## 10         BEAUTY/NAIL SALON    <NA>   105
## 28                 FAST FOOD    <NA>    99
## 46   MULTI DWELL - APT BUILD OUTSIDE    90
## 45   MULTI DWELL - APT BUILD  INSIDE    81
## 66 SOCIAL CLUB/POLICY LOCATI    <NA>    66
## 30               GAS STATION    <NA>    61
## 32            GROCERY/BODEGA OUTSIDE    51
## 48 MULTI DWELL - PUBLIC HOUS  INSIDE    47
## 37                  HOSPITAL    <NA>    47
## 54                 PVT HOUSE OUTSIDE    38
## 43              LIQUOR STORE    <NA>    36
## 69        STORE UNCLASSIFIED    <NA>    35
## 6             BAR/NIGHT CLUB OUTSIDE    33
## 40               HOTEL/MOTEL    <NA>    32
## 1                     (null)  INSIDE    32
## 23       DRY CLEANER/LAUNDRY    <NA>    31
## 64            SMALL MERCHANT    <NA>    25
## 17           COMMERCIAL BLDG OUTSIDE    24
## 31            GROCERY/BODEGA  INSIDE    21
## 53                 PVT HOUSE  INSIDE    20
```

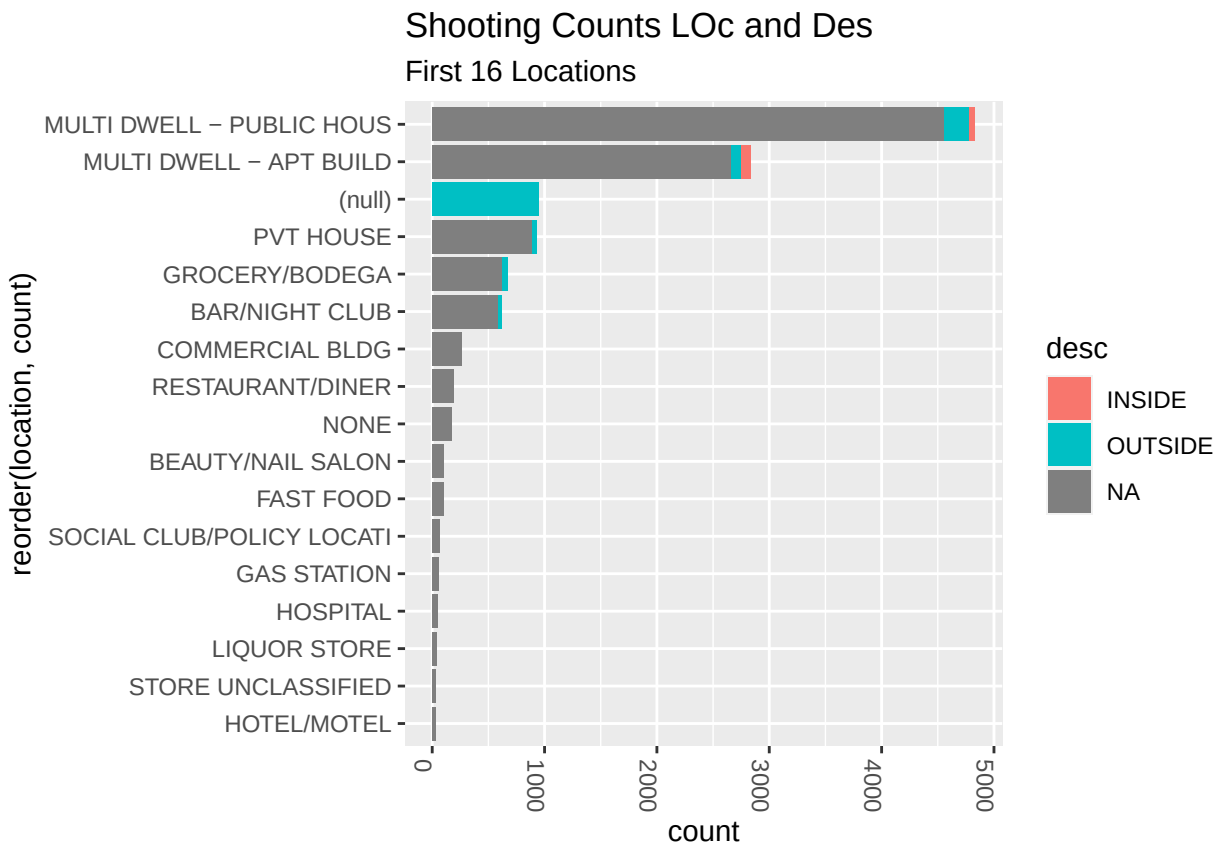
```
## 71              SUPERMARKET    <NA>    19
## 15         CLOTHING BOUTIQUE    <NA>    14
## 41             JEWELRY STORE    <NA>    12
## 74             VARIETY STORE    <NA>    11
## 36                 HOSPITAL OUTSIDE    11
## 22               DRUG STORE    <NA>    11
## 29              GAS STATION OUTSIDE    10
## 61               SHOE STORE    <NA>     9
## 19               DEPT STORE    <NA>     9
## 62            SMALL MERCHANT  INSIDE     8
## 57          RESTAURANT/DINER OUTSIDE     7
## 35                 HOSPITAL  INSIDE     7
## 5            BAR/NIGHT CLUB  INSIDE     7
## 75              VIDEO STORE  INSIDE     6
## 72           TELECOMM. STORE OUTSIDE     6
## 65 SOCIAL CLUB/POLICY LOCATI OUTSIDE     6
## 25         FACTORY/WAREHOUSE    <NA>     6
## 12              CANDY STORE    <NA>     6
## 73           TELECOMM. STORE    <NA>     5
## 42              LIQUOR STORE OUTSIDE     5
## 13              CHAIN STORE    <NA>     5
## 63            SMALL MERCHANT OUTSIDE     4
## 27                 FAST FOOD OUTSIDE     4
## 8          BEAUTY/NAIL SALON  INSIDE     4
## 56          RESTAURANT/DINER  INSIDE     3
## 34      GYM/FITNESS FACILITY    <NA>     3
## 21                 DRUG STORE OUTSIDE     3
## 16           COMMERCIAL BLDG  INSIDE     3
## 9          BEAUTY/NAIL SALON OUTSIDE     3
## 4                       BANK    <NA>     3
## 76              VIDEO STORE    <NA>     2
## 70              SUPERMARKET OUTSIDE     2
## 38              HOTEL/MOTEL  INSIDE     2
## 24         FACTORY/WAREHOUSE OUTSIDE     2
## 68        STORE UNCLASSIFIED OUTSIDE     1
## 67          STORAGE FACILITY    <NA>     1
## 60               SHOE STORE OUTSIDE     1
## 59                   SCHOOL    <NA>     1
## 52          PHOTO/COPY STORE    <NA>     1
## 44              LOAN COMPANY    <NA>     1
## 39              HOTEL/MOTEL OUTSIDE     1
## 26                FAST FOOD  INSIDE     1
## 20            DOCTOR/DENTIST    <NA>     1
## 14                CHECK CASH    <NA>     1
## 11              CANDY STORE OUTSIDE     1
## 3                       ATM    <NA>     1
```

```
#size the data down to graph main contributors without distractors...

halflocal = local_sort[2:25,]
```

```
p = ggplot(halflocal, aes(x = reorder(location,count), y = count, fill = desc)) +
  geom_bar(stat="identity") +
  theme(axis.text.x=element_text(angle = -90, hjust = 0)) +
```

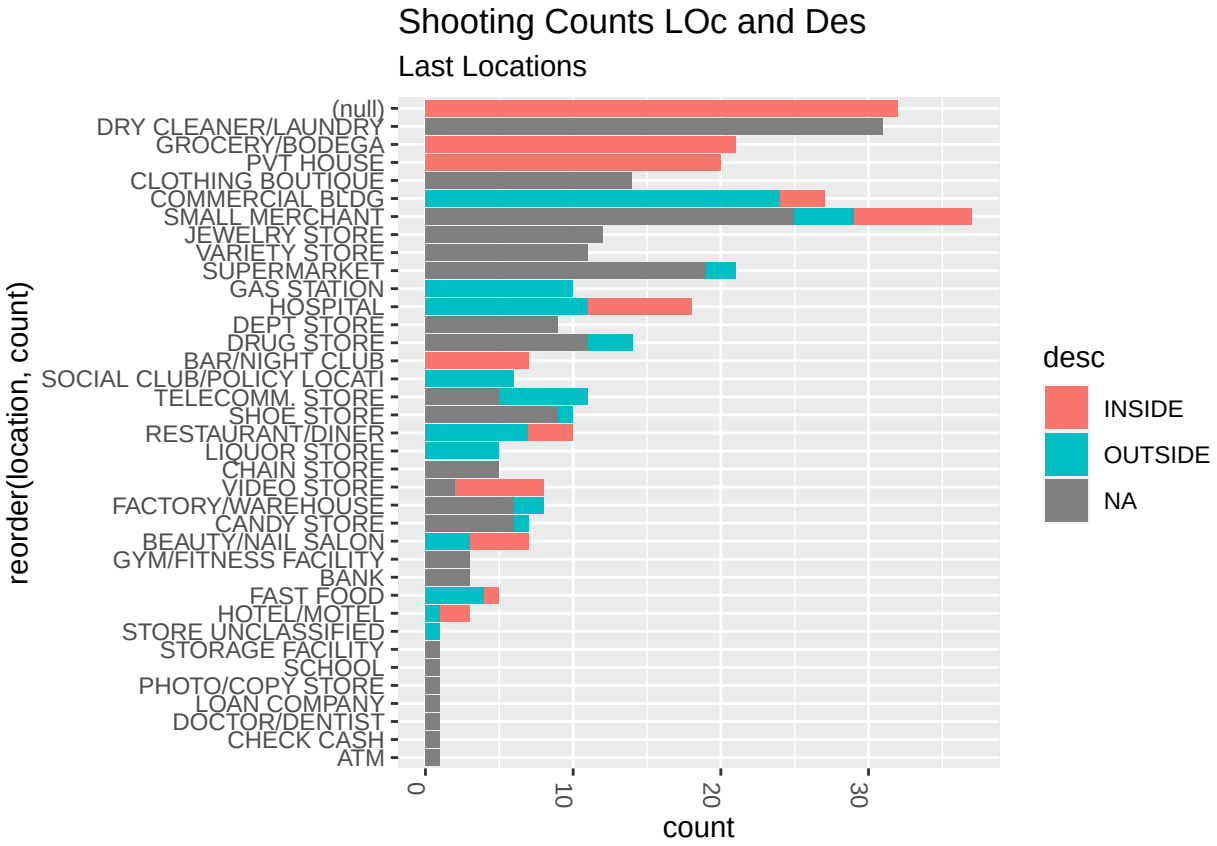
```
  labs(
    title = "Shooting Counts LOc and Des ",
    subtitle="First 16 Locations"
   )
p + coord_flip()
```




```
#plot remainder just to see whats there

otherlocal = local_sort[26:77,]

pp = ggplot(otherlocal, aes(x = reorder(location,count), y = count, fill = desc)) +
  geom_bar(stat="identity") +
  theme(axis.text.x=element_text(angle = -90, hjust = 0)) +
  labs(
    title = "Shooting Counts LOc and Des ",
    subtitle="Last Locations"
   )
pp + coord_flip()
```




**Unfortunately there was a lot of missing data - but I was able to gleam a little info that most the shootings over the years were happening within or near public and multidwelling housing. This leaps out at me like it could be a financial, population density and police services that could be affecting this...it would take a bit more reseaerch and statistics to figure that out. But I do want to pursue who is shooting who.**

**To do this I had to eliminate some confounding errors that through off the charts and data.**

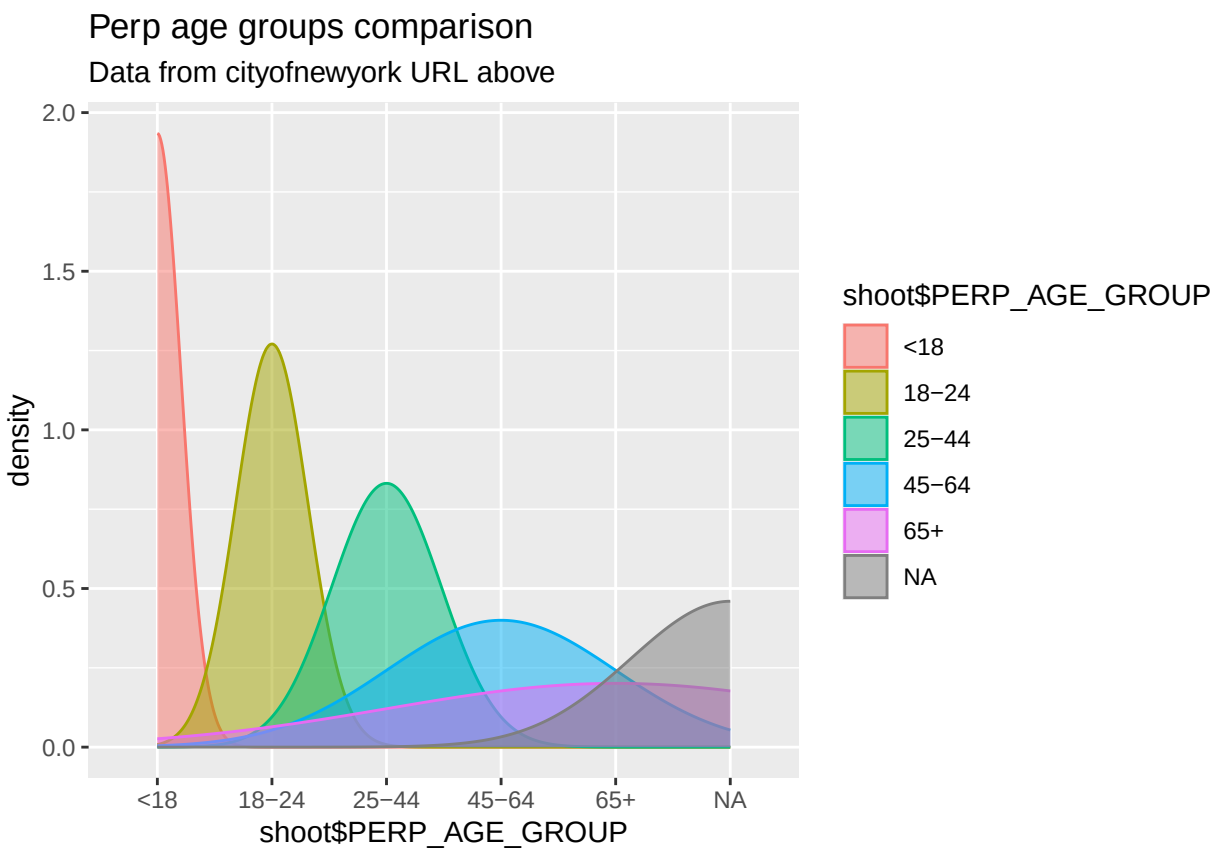
```
df1 = gunshots_tidy

shoot = df1[!df1$PERP_AGE_GROUP =="(null)",]
shoot = shoot[!shoot$PERP_AGE_GROUP =="1020",]
shoot = shoot[!shoot$PERP_AGE_GROUP =="940",]
shoot = shoot[!shoot$PERP_AGE_GROUP =="UNKNOWN",]
shoot = shoot[!shoot$PERP_AGE_GROUP == "NA",]
shoot = shoot[!shoot$PERP_AGE_GROUP == "224",]
shoot = shoot[!shoot$VIC_AGE_GROUP =="1022",]
shoot = shoot[!shoot$VIC_AGE_GROUP =="UNKNOWN",]
```

```
ggplot(shoot, aes(x = shoot$PERP_AGE_GROUP, color = shoot$PERP_AGE_GROUP, fill = shoot$PERP_AGE_GROUP )
  geom_density(alpha=.5)+
```

```
    labs(
      title = "Perp age groups comparison",
    subtitle="Data from cityofnewyork URL above"
  )
```

```
## Warning: Use of ‘shoot$PERP_AGE_GROUP‘ is discouraged.
## i Use ‘PERP_AGE_GROUP‘ instead.
## Use of ‘shoot$PERP_AGE_GROUP‘ is discouraged.
## i Use ‘PERP_AGE_GROUP‘ instead.
## Use of ‘shoot$PERP_AGE_GROUP‘ is discouraged.
## i Use ‘PERP_AGE_GROUP‘ instead.
```

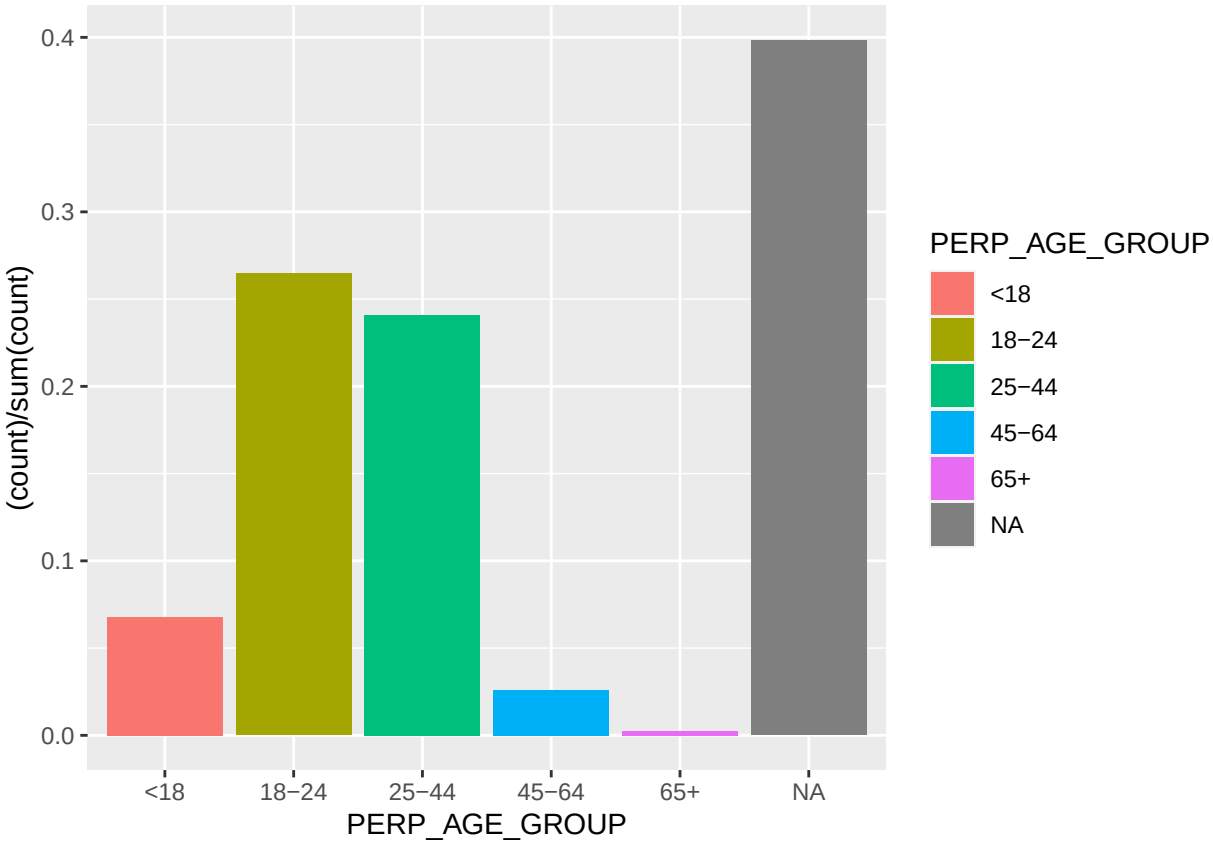



**So the above plot did not work out too well it seemed to show that the younger individuals were shooting more but – something is wrong with the calculation in density for under 18 so I decided to return to BAR CHART.**

```
ggplot(shoot, aes(x = PERP_AGE_GROUP)) +
  geom_bar(aes(y = (..count..)/sum(..count..), fill =PERP_AGE_GROUP)
)
```

```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(count)‘ instead.
## This warning is displayed once every 8 hours.
```

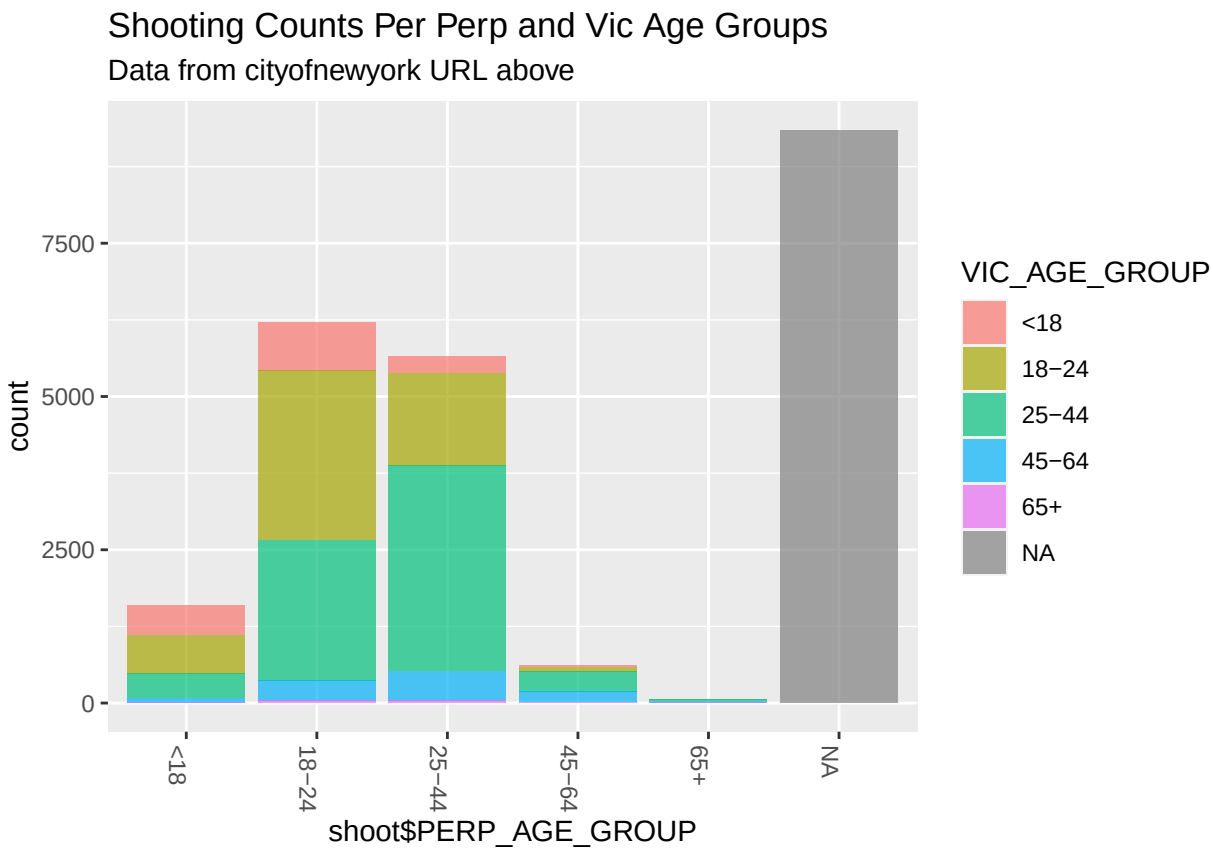
```
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```




**This one definetly shows age does appear to have different results. But, I needed more info so I added victims to the perp count and as shown below, sadly, youths were shooting youths more than older adults were shooting anyone. Again really sad info if you think about it.**

```
ggplot(shoot,aes(x = shoot$PERP_AGE_GROUP, fill = VIC_AGE_GROUP))+
  geom_bar(alpha = .7)+
  theme(axis.text.x=element_text(angle = -90, hjust = 0)) +
  labs(
    title = "Shooting Counts Per Perp and Vic Age Groups",
    subtitle="Data from cityofnewyork URL above"
  )
```

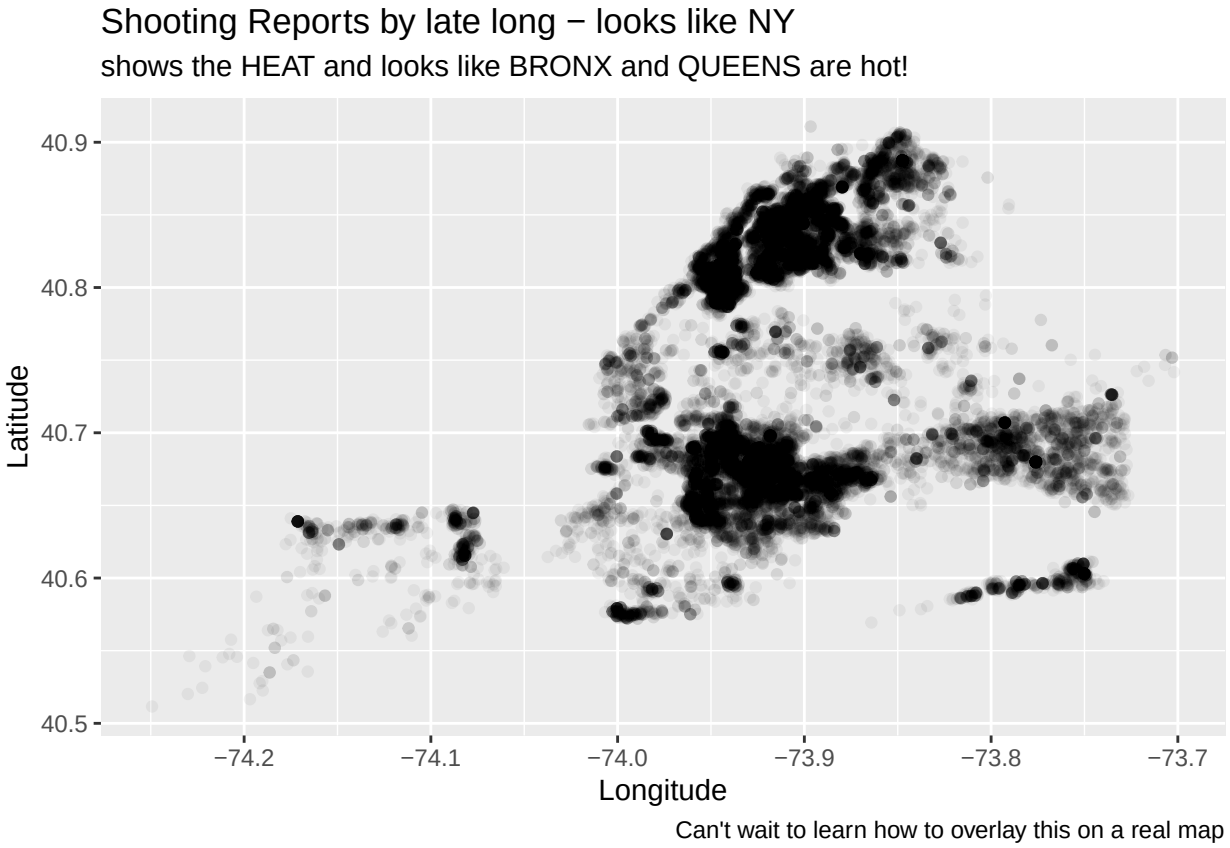
```
## Warning: Use of ‘shoot$PERP_AGE_GROUP‘ is discouraged.
## i Use ‘PERP_AGE_GROUP‘ instead.
```




**And last but not least I wanted to see if we could use the lat long data. It was easier than expected as I got a really quick "heat" like mapping. The shape follows the state of New York as you can see below.**

```
ggplot() +
  geom_point(data = gunshots_tidy, aes(x = Longitude, y = Latitude), alpha = .05)+
  labs(
  title = "Shooting Reports by late long - looks like NY",
    subtitle="shows the HEAT and looks like BRONX and QUEENS are hot!",
   caption = "Can't wait to learn how to overlay this on a real map"
  )
```

```
## Warning: Removed 10 rows containing missing values (`geom_point()`).
```




**This made me want to see more. So I visited google maps, got signed up to use their maps. . . downloaded NewYork to this markdown then plotted the above on the map – and in my amazement it worked out great. Hey becarefull when in the Bronx and Queens because the shooting activity is high there.**

```
api_key = c(1)
api_key
```

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

```
library(dotenv)
setwd(dir = "/Users/herbertschreiber/Documents/GitHub/DataSci5301-share")
list.files()
```

```
##  [1] "cred.env"                    "gunshots.csv"
##  [3] "README.md"                   "sharedSHOOTING_files"
##  [5] "sharedSHOOTING.html"         "sharedSHOOTING.nb.html"
##  [7] "sharedSHOOTING.pdf"          "sharedSHOOTING.Rmd"
##  [9] "sharedSHOOTINGnoAPI.nb.html" "sharedSHOOTINGnoAPI.Rmd"
```

```
load_dot_env("cred.env")
register_google(key = Sys.getenv("GOOGLE_MAPS_API"))
```

```
api_secret = "A" register_google(key = api_secret)
```

#The code above is my attempt to put my API google maps key into an environment and recall - I do not know if this environment travels with my repository so the image may not come out. IF YOU DO NOT GET A MAP – PLEASE – OPEN THE PNG FILE this is what the code chuck would produce if you had an API key. Sorry about any confusion.

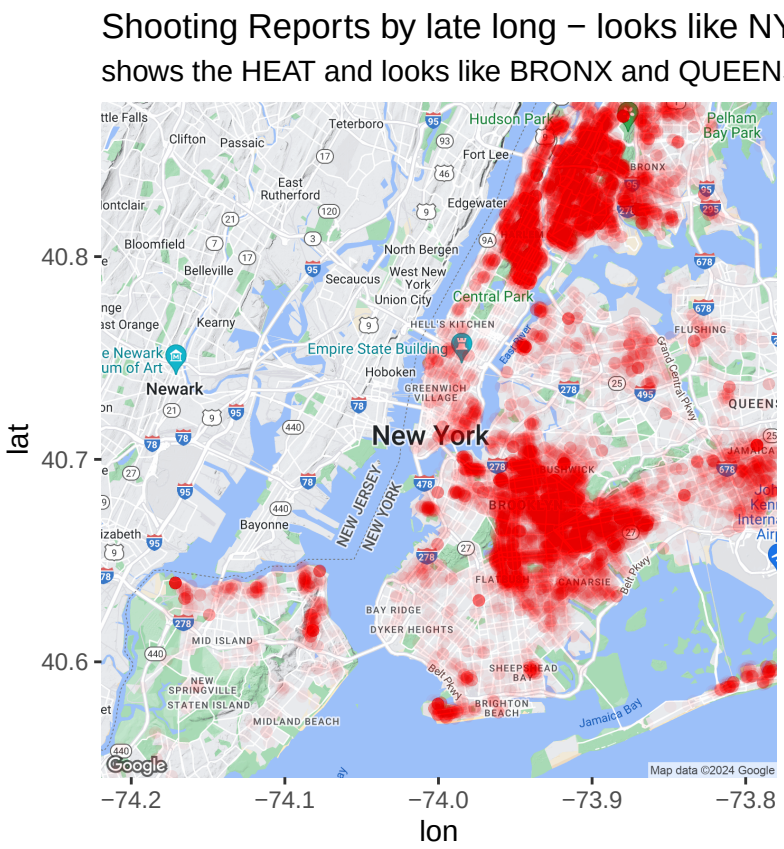
```
library(ggmap)

nyc_map <- get_map(location = c(lon = -74.00, lat = 40.71), maptype = "terrain", zoom = 11)
```

```
## i <https://maps.googleapis.com/maps/api/staticmap?center=40.71,-74&zoom=11&size=640x640&scale=2&mapty
```

```
ggmap(nyc_map) +
  geom_point(data = gunshots_tidy, aes(x = Longitude, y = Latitude), color = "red", alpha = .04)+
  labs(
  title = "Shooting Reports by late long - looks like NY",
  subtitle="shows the HEAT and looks like BRONX and QUEENS are hot!",
  )
```

```
## Warning: Removed 2512 rows containing missing values (`geom_point()`).
```



**So in conclusion, as I mentioned above this is sad data when you think about the lives and what is happening. Aside from pointing out that there appears be be a big difference in shooter age and in location there is so much that I haven't looked at.**

**The data is out their to further visualize what is happening in these locations–is it jobs, weather, density, drugs, school, the list goes on.... Another interesting idea is to pull the data time frame out and see how that changes things as the data covers before and after COVID lockdowns.**

**This has been an eye opening assignment for me as I am fairly new to R and Markdown...but now I want to learn more. Thanks. Herb.“‘{r}**