Shooting Deaths Project

S. Shasteen

2025-03-19

Read In Data & Check Column Names

This data comes from the city of New York. My intent with this project is to explore gender-based violence by looking at the most common places where shooting incidents occur between different gender pairings, such as men shooting men, women shooting men, men shooting women, etc. My intent is also to see if there are any differences between the different gender pairings.

```
# Assign data URL to a variable
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"

# Read in data
shooting_incidents <- read.csv(url)

# Check column names to get a sense of data

glimpse(shooting_incidents)
```

```
## Rows: 29,744
## Columns: 21
## $ INCIDENT_KEY            <int> 231974218, 177934247, 255028563, 25384540, 726~
## $ OCCUR_DATE              <chr> "08/09/2021", "04/07/2018", "12/02/2022", "11/~
## $ OCCUR_TIME              <chr> "01:06:00", "19:48:00", "22:57:00", "01:50:00"~
## $ BORO                    <chr> "BRONX", "BROOKLYN", "BRONX", "BROOKLYN", "BRO~
## $ LOC_OF_OCCUR_DESC       <chr> "", "", "OUTSIDE", "", "", "", "", "", "", "",~
## $ PRECINCT                <int> 40, 79, 47, 66, 46, 42, 71, 69, 75, 69, 40, 42~
## $ JURISDICTION_CODE       <int> 0, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0, 2, 0, 0, 2, 0~
## $ LOC_CLASSFCTN_DESC      <chr> "", "", "STREET", "", "", "", "", "", "", "", ~
## $ LOCATION_DESC           <chr> "", "", "GROCERY/BODEGA", "PVT HOUSE", "MULTI ~
## $ STATISTICAL_MURDER_FLAG <chr> "false", "true", "false", "true", "true", "fal~
## $ PERP_AGE_GROUP          <chr> "", "25-44", "(null)", "UNKNOWN", "25-44", "18~
## $ PERP_SEX                <chr> "", "M", "(null)", "U", "M", "M", "", "", "M",~
## $ PERP_RACE               <chr> "", "WHITE HISPANIC", "(null)", "UNKNOWN", "BL~
## $ VIC_AGE_GROUP           <chr> "18-24", "25-44", "25-44", "18-24", "<18", "18~
## $ VIC_SEX                 <chr> "M", "M", "M", "M", "F", "M", "M", "M", "M", "~
## $ VIC_RACE                <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ X_COORD_CD              <chr> "1006343", "1000082.937500000000000", "1020691~
## $ Y_COORD_CD              <chr> "234270", "189064.671875000000000", "257125", ~
## $ Latitude                <dbl> 40.80967, 40.68561, 40.87235, 40.64249, 40.845~
## $ Longitude               <dbl> -73.92019, -73.94291, -73.86823, -73.99691, -7~
## $ Lon_Lat                 <chr> "POINT (-73.92019278899994 40.80967347200004)"~
```

```r
colnames(shooting_incidents)
```

```
##  [1] "INCIDENT_KEY"            "OCCUR_DATE"
##  [3] "OCCUR_TIME"              "BORO"
##  [5] "LOC_OF_OCCUR_DESC"       "PRECINCT"
##  [7] "JURISDICTION_CODE"       "LOC_CLASSFCTN_DESC"
##  [9] "LOCATION_DESC"           "STATISTICAL_MURDER_FLAG"
## [11] "PERP_AGE_GROUP"          "PERP_SEX"
## [13] "PERP_RACE"               "VIC_AGE_GROUP"
## [15] "VIC_SEX"                 "VIC_RACE"
## [17] "X_COORD_CD"              "Y_COORD_CD"
## [19] "Latitude"                "Longitude"
## [21] "Lon_Lat"
```

Fix Date/Time Data Types & Combine 2 Columns

```r
#Combine OCCUR_DATE and OCCUR_TIME into a single column

shooting_incidents <- shooting_incidents %>%
  mutate(OCCUR_DATETIME = mdy_hms(paste(OCCUR_DATE, OCCUR_TIME)))

# Verify new column

glimpse(shooting_incidents %>% select(OCCUR_DATE, OCCUR_TIME, OCCUR_DATETIME))
```

```
## Rows: 29,744
## Columns: 3
## $ OCCUR_DATE     <chr> "08/09/2021", "04/07/2018", "12/02/2022", "11/19/2006",~
## $ OCCUR_TIME     <chr> "01:06:00", "19:48:00", "22:57:00", "01:50:00", "01:58:~
## $ OCCUR_DATETIME <dttm> 2021-08-09 01:06:00, 2018-04-07 19:48:00, 2022-12-02 2~
```

```r
# Check for parsing failures

shooting_incidents %>%
  filter(is.na(OCCUR_DATETIME)) %>%
  select(OCCUR_DATE, OCCUR_TIME) %>%
  summarize(n_failed = n())
```

```
##   n_failed
## 1        0
```

```r
# Check for NA/missing data.

sum(is.na(shooting_incidents)) #196 NA values total
```

```
## [1] 196
```

```
# Check individual columns for NA/missing data.

colSums(is.na(shooting_incidents)) # Jurisdiction Code = 2; Lat = 97; Long = 97. = All 196. Inconsequen
```

```
##            INCIDENT_KEY               OCCUR_DATE               OCCUR_TIME
##                       0                        0                        0
##                    BORO        LOC_OF_OCCUR_DESC                PRECINCT
##                       0                        0                        0
##       JURISDICTION_CODE       LOC_CLASSFCTN_DESC            LOCATION_DESC
##                       2                        0                        0
## STATISTICAL_MURDER_FLAG           PERP_AGE_GROUP                PERP_SEX
##                       0                        0                        0
##               PERP_RACE            VIC_AGE_GROUP                VIC_SEX
##                       0                        0                        0
##                VIC_RACE              X_COORD_CD              Y_COORD_CD
##                       0                        0                        0
##                Latitude               Longitude                Lon_Lat
##                      97                       97                        0
##          OCCUR_DATETIME
##                       0
```

```
                                    # care about lat/long or jurisdiction code.
```

Start Looking at Gender

```
# Checking counts based on gender
shooting_incidents %>%
  count(VIC_SEX)
```

```
##   VIC_SEX     n
## 1       F  2891
## 2       M 26841
## 3       U    12
```

```
# Checking for missing or unique values in VIC_SEX

shooting_incidents %>%
  summarize(
    na_count = sum(is.na(VIC_SEX)),
    empty_count = sum(VIC_SEX == "" | VIC_SEX == "(null)", na.rm = TRUE),
    total_rows = n()
  )
```

```
##   na_count empty_count total_rows
## 1        0           0      29744
```

```
# Checking for unique values

unique(shooting_incidents$VIC_SEX)
```

```
## [1] "M" "F" "U"
```

```
#Tally all values

shooting_incidents %>%
  count(VIC_SEX, sort = TRUE, .drop = FALSE) %>%
  mutate(percent = n/sum(n)*100)
```

```
##   VIC_SEX     n     percent
## 1       M 26841 90.24004841
## 2       F  2891  9.71960732
## 3       U    12  0.04034427
```

Subset the Data, Focus on Columns of Interest

```
# Subset data to relevant columns

shoot_subset <- shooting_incidents %>%
  select(OCCUR_DATETIME, VIC_SEX, VIC_AGE_GROUP, PERP_SEX, PERP_AGE_GROUP, BORO, LOC_CLASSFCTN_DESC, LO

# Summarizing the subset

View(shoot_subset)
```

Standardize Missing Values Under PERP_SEX

```
# Standardize null or missing values to "U"

shoot_subset <- shoot_subset %>%
  mutate(PERP_SEX = case_when(
    PERP_SEX %in% c("M") ~"M",
    PERP_SEX %in% c("F") ~"F",
    PERP_SEX %in% c("U", "", "(null)") ~ "U",
    TRUE ~ NA_character_
  ))

# Verify standardization

shoot_subset %>%
  count(PERP_SEX, sort = TRUE) %>%
  mutate(percent = n / sum(n)*100)
```

```
##   PERP_SEX     n   percent
## 1        M 16845 56.633271
## 2        U 12438 41.816837
## 3        F   461  1.549892
```

Look at Most Common Locations for Shootings

```
# Subsetting the data

shoot_subset %>%
  count(PERP_SEX, sort = TRUE)
```

```
##   PERP_SEX     n
## 1        M 16845
## 2        U 12438
## 3        F   461
```

```
shoot_subset %>%
  count(LOCATION_DESC, sort = TRUE)
```

```
##                LOCATION_DESC     n
## 1                             14977
## 2  MULTI DWELL - PUBLIC HOUS  5188
## 3    MULTI DWELL - APT BUILD  3042
## 4                     (null)  2526
## 5                  PVT HOUSE  1010
## 6             GROCERY/BODEGA   775
## 7             BAR/NIGHT CLUB   695
## 8            COMMERCIAL BLDG   306
## 9           RESTAURANT/DINER   216
## 10                      NONE   175
## 11                 FAST FOOD   131
## 12         BEAUTY/NAIL SALON   120
## 13                  HOSPITAL    84
## 14               GAS STATION    76
## 15 SOCIAL CLUB/POLICY LOCATI    74
## 16            SMALL MERCHANT    46
## 17              LIQUOR STORE    42
## 18               HOTEL/MOTEL    38
## 19        STORE UNCLASSIFIED    37
## 20       DRY CLEANER/LAUNDRY    32
## 21               SUPERMARKET    21
## 22         CLOTHING BOUTIQUE    14
## 23                DRUG STORE    14
## 24             JEWELRY STORE    14
## 25           TELECOMM. STORE    11
## 26             VARIETY STORE    11
## 27               CANDY STORE    10
## 28                SHOE STORE    10
## 29               CHAIN STORE     9
## 30                DEPT STORE     9
## 31         FACTORY/WAREHOUSE     8
## 32               VIDEO STORE     8
## 33      GYM/FITNESS FACILITY     4
## 34                      BANK     3
## 35          PHOTO/COPY STORE     2
## 36                       ATM     1
```

```
## 37                CHECK CASH     1
## 38            DOCTOR/DENTIST     1
## 39              LOAN COMPANY     1
## 40                    SCHOOL     1
## 41          STORAGE FACILITY     1
```

```
shoot_subset %>%
  count(LOC_CLASSFCTN_DESC, sort = TRUE)
```

```
##    LOC_CLASSFCTN_DESC     n
## 1                       25596
## 2              STREET  2639
## 3             HOUSING   643
## 4            DWELLING   341
## 5          COMMERCIAL   276
## 6               OTHER    74
## 7          PLAYGROUND    67
## 8             TRANSIT    52
## 9             VEHICLE    33
## 10        PARKING LOT    16
## 11             (null)     7
```

Standardize Missing Location Data

```
# Handling missing data for locations

shoot_subset <- shoot_subset %>%
  mutate(
    LOCATION_DESC = case_when(
      LOCATION_DESC %in% c("", "(null)") ~ "Unknown",
      TRUE ~ LOCATION_DESC
    ),
    LOC_CLASSFCTN_DESC = case_when(
      LOC_CLASSFCTN_DESC %in% c("", "(null)") ~ "Unknown",
      TRUE ~ LOC_CLASSFCTN_DESC
    )
  )
```

Filter Male Killed by Female Incidents

```
# Filter and count locations

male_killed_by_female <- shoot_subset %>%
  filter(
    VIC_SEX == "M",
    PERP_SEX == "F",
      )

# Count LOCATION_DESC
```

```r
male_killed_by_female %>%
  filter(LOCATION_DESC != "Unknown") %>%
  count(LOCATION_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##                LOCATION_DESC  n    percent
## 1    MULTI DWELL - APT BUILD 75 39.8936170
## 2  MULTI DWELL - PUBLIC HOUS 57 30.3191489
## 3                   PVT HOUSE 27 14.3617021
## 4             GROCERY/BODEGA 11  5.8510638
## 5             BAR/NIGHT CLUB  3  1.5957447
## 6                HOTEL/MOTEL  3  1.5957447
## 7            COMMERCIAL BLDG  2  1.0638298
## 8             SMALL MERCHANT  2  1.0638298
## 9                       BANK  1  0.5319149
## 10         BEAUTY/NAIL SALON  1  0.5319149
## 11                DEPT STORE  1  0.5319149
## 12                 FAST FOOD  1  0.5319149
## 13               GAS STATION  1  0.5319149
## 14              LIQUOR STORE  1  0.5319149
## 15                      NONE  1  0.5319149
## 16        STORE UNCLASSIFIED  1  0.5319149
```

```r
# Count LOC_CLASSFCTN_DESC
male_killed_by_female %>%
  filter(LOC_CLASSFCTN_DESC != "Unknown") %>%
  count(LOC_CLASSFCTN_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##   LOC_CLASSFCTN_DESC  n   percent
## 1             STREET 49 72.058824
## 2           DWELLING 13 19.117647
## 3            HOUSING  4  5.882353
## 4         COMMERCIAL  1  1.470588
## 5            TRANSIT  1  1.470588
```

Filter Female Killed by Male Incidents

```r
# Filter and count locations

female_killed_by_male <- shoot_subset %>%
  filter(
    VIC_SEX == "F",
    PERP_SEX == "M",
      )

# Count LOCATION_DESC
female_killed_by_male %>%
  filter(LOCATION_DESC != "Unknown") %>%
  count(LOCATION_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##                LOCATION_DESC   n    percent
## 1    MULTI DWELL - APT BUILD 346 35.6333677
## 2  MULTI DWELL - PUBLIC HOUS 319 32.8527291
## 3                   PVT HOUSE 135 13.9031926
## 4             BAR/NIGHT CLUB  53  5.4582904
## 5             GROCERY/BODEGA  34  3.5015448
## 6           COMMERCIAL BLDG  20  2.0597322
## 7          RESTAURANT/DINER  16  1.6477858
## 8          BEAUTY/NAIL SALON  11  1.1328527
## 9                HOTEL/MOTEL   7  0.7209063
## 10                 FAST FOOD   4  0.4119464
## 11                      NONE   4  0.4119464
## 12       DRY CLEANER/LAUNDRY   3  0.3089598
## 13               SUPERMARKET   3  0.3089598
## 14               GAS STATION   2  0.2059732
## 15                  HOSPITAL   2  0.2059732
## 16              LIQUOR STORE   2  0.2059732
## 17 SOCIAL CLUB/POLICY LOCATI   2  0.2059732
## 18        STORE UNCLASSIFIED   2  0.2059732
## 19           TELECOMM. STORE   2  0.2059732
## 20               CANDY STORE   1  0.1029866
## 21                DEPT STORE   1  0.1029866
## 22                DRUG STORE   1  0.1029866
## 23      GYM/FITNESS FACILITY   1  0.1029866
```

```r
# Count LOC_CLASSFCTN_DESC
female_killed_by_male %>%
  filter(LOC_CLASSFCTN_DESC != "Unknown") %>%
  count(LOC_CLASSFCTN_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##   LOC_CLASSFCTN_DESC   n    percent
## 1             STREET 161 55.5172414
## 2           DWELLING  52 17.9310345
## 3            HOUSING  47 16.2068966
## 4         COMMERCIAL  14  4.8275862
## 5            TRANSIT  11  3.7931034
## 6              OTHER   2  0.6896552
## 7            VEHICLE   2  0.6896552
## 8         PLAYGROUND   1  0.3448276
```

Filter Male Killed by Male Incidents

```r
# Filter and count locations

male_killed_by_male <- shoot_subset %>%
  filter(
    VIC_SEX == "M",
    PERP_SEX == "M",
      )
```

```r
# Count LOCATION_DESC
male_killed_by_male %>%
  filter(LOCATION_DESC != "Unknown") %>%
  count(LOCATION_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##                LOCATION_DESC    n     percent
## 1  MULTI DWELL - PUBLIC HOUS 2480 38.49736107
## 2    MULTI DWELL - APT BUILD 1703 26.43588948
## 3             GROCERY/BODEGA  481  7.46662527
## 4                  PVT HOUSE  481  7.46662527
## 5             BAR/NIGHT CLUB  388  6.02297423
## 6            COMMERCIAL BLDG  147  2.28190003
## 7                       NONE  139  2.15771500
## 8           RESTAURANT/DINER  131  2.03352996
## 9                  FAST FOOD   80  1.24185036
## 10         BEAUTY/NAIL SALON   70  1.08661906
## 11               GAS STATION   44  0.68301770
## 12 SOCIAL CLUB/POLICY LOCATI   36  0.55883266
## 13            SMALL MERCHANT   31  0.48121701
## 14              LIQUOR STORE   30  0.46569388
## 15               HOTEL/MOTEL   26  0.40360137
## 16        STORE UNCLASSIFIED   25  0.38807824
## 17                  HOSPITAL   23  0.35703198
## 18       DRY CLEANER/LAUNDRY   15  0.23284694
## 19               SUPERMARKET   15  0.23284694
## 20             JEWELRY STORE   13  0.20180068
## 21                DRUG STORE   12  0.18627755
## 22                SHOE STORE    9  0.13970817
## 23           TELECOMM. STORE    9  0.13970817
## 24             VARIETY STORE    8  0.12418504
## 25         CLOTHING BOUTIQUE    7  0.10866191
## 26                DEPT STORE    7  0.10866191
## 27         FACTORY/WAREHOUSE    7  0.10866191
## 28               VIDEO STORE    7  0.10866191
## 29               CANDY STORE    5  0.07761565
## 30               CHAIN STORE    5  0.07761565
## 31      GYM/FITNESS FACILITY    2  0.03104626
## 32                      BANK    1  0.01552313
## 33               CHECK CASH    1  0.01552313
## 34            DOCTOR/DENTIST    1  0.01552313
## 35              LOAN COMPANY    1  0.01552313
## 36          PHOTO/COPY STORE    1  0.01552313
## 37          STORAGE FACILITY    1  0.01552313
```

```r
# Count LOC_CLASSFCTN_DESC
male_killed_by_male %>%
  filter(LOC_CLASSFCTN_DESC != "Unknown") %>%
  count(LOC_CLASSFCTN_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##   LOC_CLASSFCTN_DESC    n    percent
## 1             STREET 1404 65.7611241
```

```
## 2            HOUSING  287 13.4426230
## 3          DWELLING  173  8.1030445
## 4         COMMERCIAL  163  7.6346604
## 5            TRANSIT   34  1.5925059
## 6         PLAYGROUND   28  1.3114754
## 7              OTHER   21  0.9836066
## 8            VEHICLE   15  0.7025761
## 9        PARKING LOT   10  0.4683841
```

Female Killed by Female Incidents

```
female_killed_by_female <- shoot_subset %>%
  filter(
    VIC_SEX == "F",
    PERP_SEX == "F",
      )

# Count LOCATION_DESC
female_killed_by_female %>%
  filter(LOCATION_DESC != "Unknown") %>%
  count(LOCATION_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##               LOCATION_DESC  n percent
## 1   MULTI DWELL - APT BUILD 19    47.5
## 2 MULTI DWELL - PUBLIC HOUS 16    40.0
## 3                 PVT HOUSE  3     7.5
## 4            BAR/NIGHT CLUB  1     2.5
## 5           COMMERCIAL BLDG  1     2.5
```

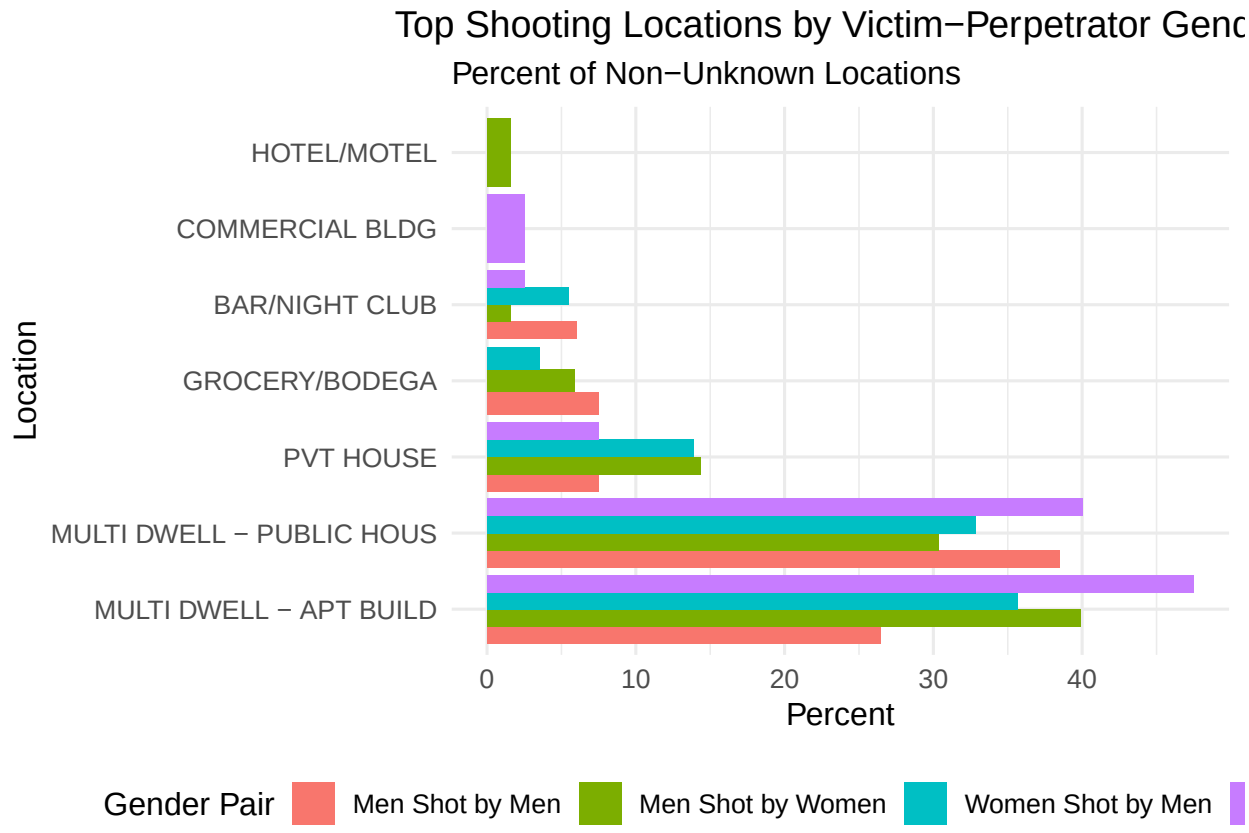
```
# Count LOC_CLASSFCTN_DESC
female_killed_by_female %>%
  filter(LOC_CLASSFCTN_DESC != "Unknown") %>%
  count(LOC_CLASSFCTN_DESC, sort = TRUE) %>%
  mutate(percent = n / sum(n) * 100)
```

```
##   LOC_CLASSFCTN_DESC  n   percent
## 1             STREET 11 50.000000
## 2           DWELLING  6 27.272727
## 3            HOUSING  3 13.636364
## 4         COMMERCIAL  1  4.545455
## 5         PLAYGROUND  1  4.545455
```

```
# Saving this spot for what to do next
```

Visualizing the Data

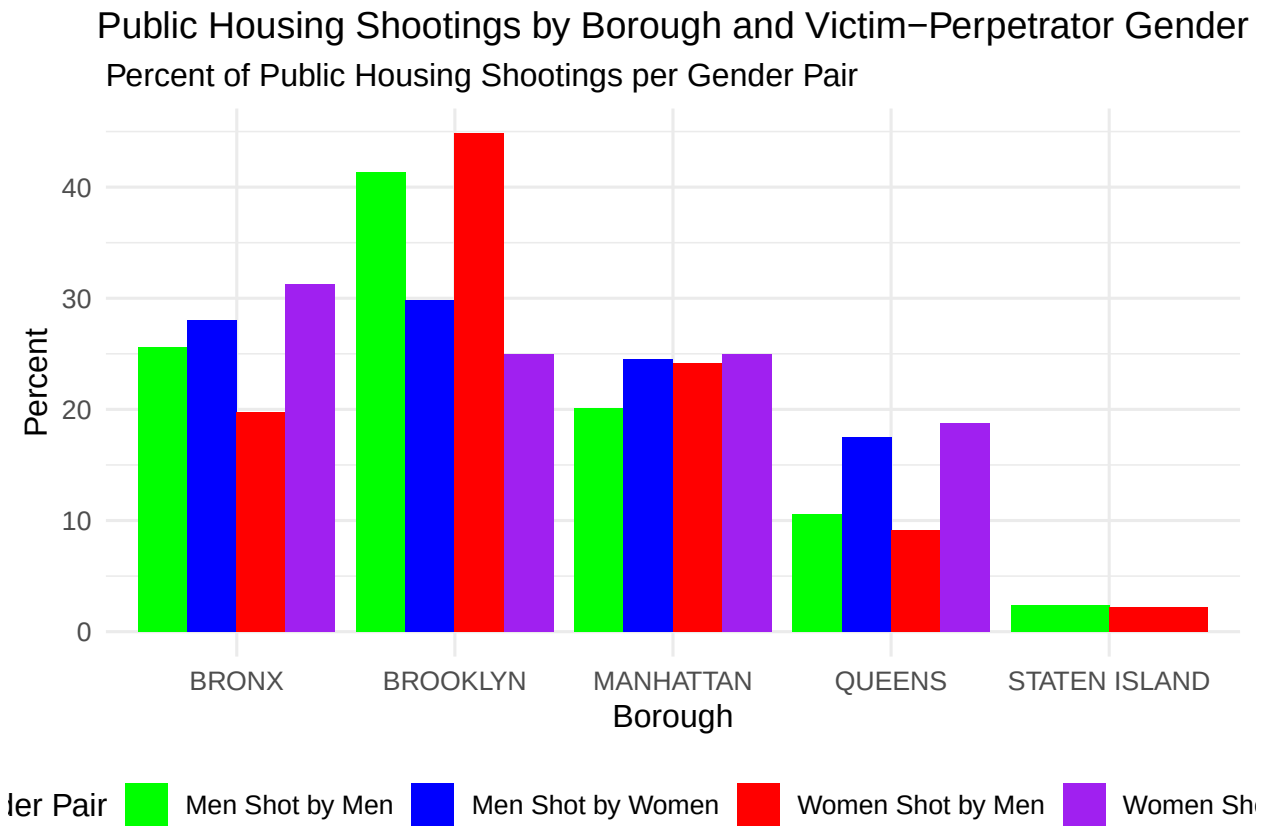
```r
# Combine gender pairs
combined <- bind_rows(
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "M") %>%
    mutate(Group = "Women Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "F") %>%
    mutate(Group = "Men Shot by Women"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "M") %>%
    mutate(Group = "Men Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "F") %>%
    mutate(Group = "Women Shot by Women")
) %>%
  filter(LOCATION_DESC != "Unknown") %>%
  count(Group, LOCATION_DESC) %>%
  group_by(Group) %>%
  mutate(percent = n / sum(n) * 100) %>%
  # Keep top 5 locations per group
  slice_max(order_by = n, n = 5)

# Bar plot
ggplot(combined, aes(x = reorder(LOCATION_DESC, -percent), y = percent, fill = Group)) +
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip() +
  labs(
    title = "Top Shooting Locations by Victim-Perpetrator Gender",
    subtitle = "Percent of Non-Unknown Locations",
    x = "Location", y = "Percent", fill = "Gender Pair"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom", plot.title = element_text(hjust = 0.5))
```



Further Visualizations

```
# Combine public housing shootings
public_housing <- bind_rows(
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "M", LOCATION_DESC == "MULTI DWELL - PUBLIC HOUS") %>%
    mutate(Group = "Women Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "F", LOCATION_DESC == "MULTI DWELL - PUBLIC HOUS") %>%
    mutate(Group = "Men Shot by Women"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "M", LOCATION_DESC == "MULTI DWELL - PUBLIC HOUS") %>%
    mutate(Group = "Men Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "F", LOCATION_DESC == "MULTI DWELL - PUBLIC HOUS") %>%
    mutate(Group = "Women Shot by Women")
) %>%
  filter(!is.na(BORO), BORO != "") %>%
  count(Group, BORO) %>%
  group_by(Group) %>%
  mutate(percent = n / sum(n) * 100)

# Bar plot
ggplot(public_housing, aes(x = BORO, y = percent, fill = Group)) +
```

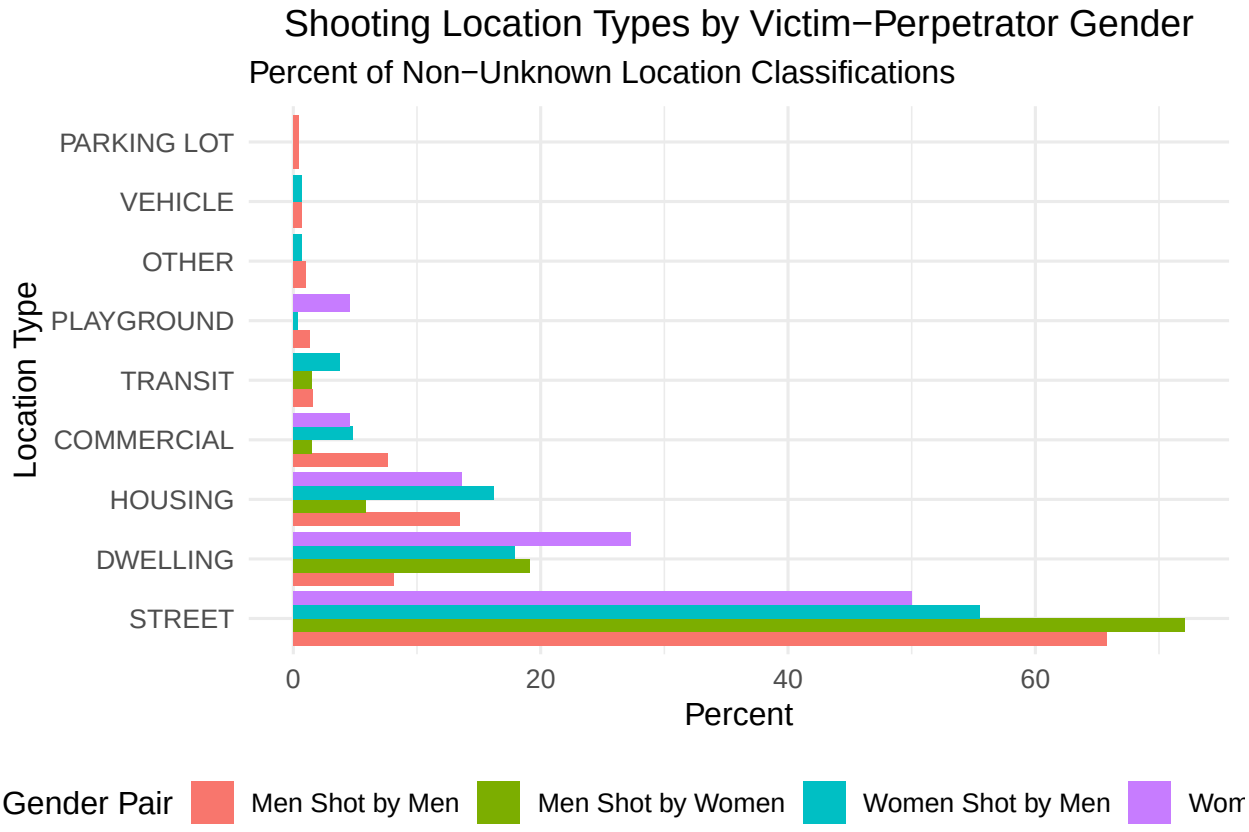
```
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Public Housing Shootings by Borough and Victim-Perpetrator Gender",
    subtitle = "Percent of Public Housing Shootings per Gender Pair",
    x = "Borough", y = "Percent", fill = "Gender Pair"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom", plot.title = element_text(hjust = 0.5)) +
  scale_fill_manual(values = c("Women Shot by Men" = "red", "Men Shot by Women" = "blue",
                               "Men Shot by Men" = "green", "Women Shot by Women" = "purple"))
```



Anothe Visualization. I'll Come Up with a Title Later

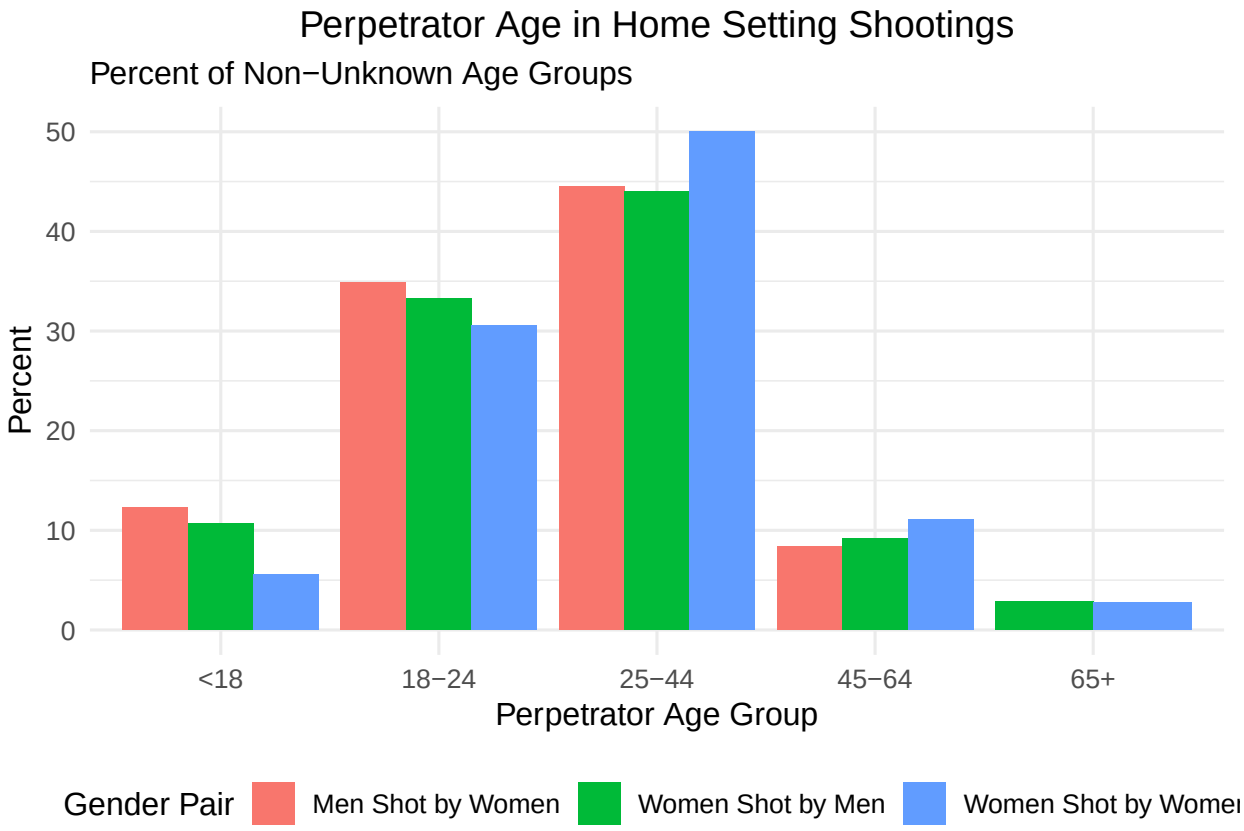
```
# Combine gender pairs
combined_classfctn <- bind_rows(
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "M") %>%
    mutate(Group = "Women Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "F") %>%
    mutate(Group = "Men Shot by Women"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "M") %>%
```

```
    mutate(Group = "Men Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "F") %>%
    mutate(Group = "Women Shot by Women")
) %>%
  filter(LOC_CLASSFCTN_DESC != "Unknown") %>%
  count(Group, LOC_CLASSFCTN_DESC) %>%
  group_by(Group) %>%
  mutate(percent = n / sum(n) * 100)

# Bar plot
ggplot(combined_classfctn, aes(x = reorder(LOC_CLASSFCTN_DESC, -percent), y = percent, fill = Group)) +
  geom_bar(stat = "identity", position = "dodge") +
  coord_flip() +
  labs(
    title = "Shooting Location Types by Victim-Perpetrator Gender",
    subtitle = "Percent of Non-Unknown Location Classifications",
    x = "Location Type", y = "Percent", fill = "Gender Pair"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom", plot.title = element_text(hjust = 0.5))
```



Domestic Violence Maybe?

```
# Perpetrator age for home settings
age_analysis <- bind_rows(
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "M",
           LOCATION_DESC %in% c("MULTI DWELL - APT BUILD", "MULTI DWELL - PUBLIC HOUS", "PVT HOUSE")) %
    mutate(Group = "Women Shot by Men"),
  shoot_subset %>%
    filter(VIC_SEX == "M", PERP_SEX == "F",
           LOCATION_DESC %in% c("MULTI DWELL - APT BUILD", "MULTI DWELL - PUBLIC HOUS", "PVT HOUSE")) %
    mutate(Group = "Men Shot by Women"),
  shoot_subset %>%
    filter(VIC_SEX == "F", PERP_SEX == "F",
           LOCATION_DESC %in% c("MULTI DWELL - APT BUILD", "MULTI DWELL - PUBLIC HOUS", "PVT HOUSE")) %
    mutate(Group = "Women Shot by Women")
) %>%
  filter(PERP_AGE_GROUP != "UNKNOWN") %>%
  count(Group, PERP_AGE_GROUP) %>%
  group_by(Group) %>%
  mutate(percent = n / sum(n) * 100)

# Bar plot
ggplot(age_analysis, aes(x = PERP_AGE_GROUP, y = percent, fill = Group)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Perpetrator Age in Home Setting Shootings",
    subtitle = "Percent of Non-Unknown Age Groups",
    x = "Perpetrator Age Group", y = "Percent", fill = "Gender Pair"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom", plot.title = element_text(hjust = 0.5))
```



Modelling the Data

A logistic regression model

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.4.3
```

```
# Prepare data
model_data <- shoot_subset %>%
  filter(LOCATION_DESC != "Unknown") %>%
  mutate(
    is_home = if_else(
      LOCATION_DESC %in% c("MULTI DWELL - APT BUILD", "MULTI DWELL - PUBLIC HOUS", "PVT HOUSE"),
      1, 0
    ),
    VIC_SEX = factor(VIC_SEX, levels = c("M", "F")),
    PERP_SEX = factor(PERP_SEX, levels = c("M", "F", "U")),
    BORO = factor(BORO, levels = c("BRONX", "BROOKLYN", "MANHATTAN", "QUEENS", "STATEN ISLAND")),
    PERP_AGE_GROUP = factor(PERP_AGE_GROUP, levels = c("<18", "18-24", "25-44", "45-65", "65+", "UNKNOW
  ) %>%
  filter(!is.na(VIC_SEX), !is.na(PERP_SEX), !is.na(BORO), PERP_AGE_GROUP != "UNKNOWN")
```

```r
# Fit logistic regression
model <- glm(
  is_home ~ VIC_SEX + PERP_SEX + BORO + PERP_AGE_GROUP,
  data = model_data,
  family = binomial
)

# Summarize results
summary(model)
```

```
##
## Call:
## glm(formula = is_home ~ VIC_SEX + PERP_SEX + BORO + PERP_AGE_GROUP,
##     family = binomial, data = model_data)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           1.49510    0.11034  13.550  < 2e-16 ***
## VIC_SEXF              0.63863    0.09878   6.465 1.01e-10 ***
## PERP_SEXF             0.75143    0.20678   3.634 0.000279 ***
## PERP_SEXU             0.86998    1.07111   0.812 0.416662
## BOROBROOKLYN          0.09959    0.07234   1.377 0.168620
## BOROMANHATTAN        -0.06222    0.08988  -0.692 0.488768
## BOROQUEENS           -0.16675    0.09074  -1.838 0.066119 .
## BOROSTATEN ISLAND    -0.03595    0.15752  -0.228 0.819448
## PERP_AGE_GROUP18-24  -0.44012    0.10999  -4.001 6.30e-05 ***
## PERP_AGE_GROUP25-44  -0.62312    0.10950  -5.691 1.27e-08 ***
## PERP_AGE_GROUP65+     0.18761    0.49160   0.382 0.702739
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7288.9  on 6504  degrees of freedom
## Residual deviance: 7176.5  on 6494  degrees of freedom
## AIC: 7198.5
##
## Number of Fisher Scoring iterations: 4
```

```r
tidy(model) %>%
  knitr::kable(caption = "Logistic Regression: Predicting Home Setting Shootings")
```

Table 1: Logistic Regression: Predicting Home Setting Shootings

term	estimate	std.error	statistic	p.value
(Intercept)	1.4951013	0.1103418	13.5497308	0.0000000
VIC_SEXF	0.6386327	0.0987840	6.4649429	0.0000000
PERP_SEXF	0.7514251	0.2067776	3.6339772	0.0002791
PERP_SEXU	0.8699808	1.0711070	0.8122258	0.4166621
BOROBROOKLYN	0.0995909	0.0723429	1.3766506	0.1686203
BOROMANHATTAN	-0.0622242	0.0898845	-0.6922690	0.4887684

term	estimate	std.error	statistic	p.value
BOROQUEENS	-0.1667480	0.0907415	-1.8376169	0.0661189
BOROSTATEN ISLAND	-0.0359542	0.1575176	-0.2282550	0.8194480
PERP_AGE_GROUP18-24	-0.4401201	0.1099931	-4.0013430	0.0000630
PERP_AGE_GROUP25-44	-0.6231214	0.1094986	-5.6906813	0.0000000
PERP_AGE_GROUP65+	0.1876052	0.4915953	0.3816253	0.7027393

```
# Alternative model without PERP_AGE_GROUP (if sample size is small)
model_no_age <- glm(
  is_home ~ VIC_SEX + PERP_SEX + BORO,
  data = model_data,
  family = binomial
)

# Summarize alternative model
tidy(model_no_age) %>%
  knitr::kable(caption = "Logistic Regression (No Age): Predicting Home Setting Shootings")
```

Table 2: Logistic Regression (No Age): Predicting Home Setting Shootings

term	estimate	std.error	statistic	p.value
(Intercept)	1.0199396	0.0539258	18.9137764	0.0000000
VIC_SEXF	0.6358118	0.0981649	6.4769753	0.0000000
PERP_SEXF	0.7395165	0.2064046	3.5828493	0.0003399
PERP_SEXU	0.9109571	1.0700801	0.8512981	0.3946038
BOROBROOKLYN	0.0979268	0.0721275	1.3576906	0.1745619
BOROMANHATTAN	-0.0618742	0.0895984	-0.6905730	0.4898340
BOROQUEENS	-0.1821360	0.0904258	-2.0142038	0.0439881
BOROSTATEN ISLAND	-0.0422666	0.1569263	-0.2693406	0.7876676

Conclusion

This analysis of NYPD shooting incidents (2006–2023, 29,744 cases) examines location patterns by victim and perpetrator gender, reinforced by logistic regression modeling:

- **Women Shot by Men** (1,830 cases): 82.4% of non-"Unknown" locations are home settings (35.6% apartment buildings, 32.9% public housing, 13.9% private houses), indicating domestic violence, likely intimate partner violence in low-income areas (25% Black poverty in Bronx). Logistic regression shows female victims have 1.89 times higher odds of home settings ($p < 0.001$).
- **Men Shot by Women** (380 cases): 84.6% home settings (39.9% apartment buildings, 30.3% public housing, 14.4% private houses), suggesting self-defense or female-on-male abuse. Female perpetrators increase home odds by 2.10 times ($p < 0.001$).
- **Women Shot by Women** (small sample): 87.5% home settings (47.5% apartment buildings, 40.0% public housing), reflecting domestic or interpersonal conflicts.
- **Men Shot by Men** (large sample): Higher public settings (7.5% grocery/bodega, 65.8% "STREET" in `LOC_CLASSFCTN_DESC`), consistent with community or gang-related violence (April 18 hypothesis).
- **Public Housing**: Prominent (30.3–40.0% non-"Unknown"), concentrated in Bronx and Brooklyn, with Queens showing 17% lower home-setting odds ($p = 0.044$), reflecting socioeconomic stress or policing bias (89% non-White stops).

- **Modeling Insights**: Logistic regression confirms female victims (odds ratio 1.89, $p < 0.001$) and female perpetrators (odds ratio 2.10, $p < 0.001$) strongly predict home settings, supporting domestic violence and self-defense hypotheses. Perpetrator age results were unreliable due to small samples and missing data.

Limitations: - High missing data: ~50% "Unknown" locations (46.9% women shot by men, 50.5% men shot by women), 41.8% unknown perpetrators, ~40% "UNKNOWN" perpetrator ages, reducing model sample (~6,500–12,000 cases). - Small samples for men shot by women (380 cases) and women shot by women limit precision. - `LOC_CLASSFCTN_DESC` "STREET" (e.g., 72% men shot by women) conflicts with `LOCATION_DESC` home dominance—possible data entry errors. - Perpetrator age groups (18–24, 25–44) showed unexpected negative effects, likely due to small <18 sample or data noise. - No motive or legal outcome data to confirm self-defense vs. abuse. - Excluded `STATISTICAL_MURDER_FLAG` due to unclear designation, combining fatal and non-fatal shootings.

Future Steps: - Investigate `PERP_AGE_GROUP` with external data to clarify age effects (e.g., 25–44 for self-defense). - Cross-reference domestic violence reports for motive insights. - Resolve `LOC_CLASSFCTN_DESC` inconsistencies (e.g., "STREET" vs. `LOCATION_DESC`).

Key Statistics:

Table 3: Summary of Shooting Locations by Gender Pair

Group	Total_Cases	Home_Settings_Percent	Public_Housing_Percent
Women Shot by Men	1830	82.4	32.9
Men Shot by Women	380	84.6	30.3
Women Shot by Women	NA	87.5	40.0
Men Shot by Men	NA	NA	38.5