

# WEEK 3 NYPD Shooting Incident Data Report

HDaher

2024-04-06

---

## Description

- Data set of every Shooting Incident in NYC since 2006 to end of previous calendar year.
- Each record includes information about the event, the location and time of occurrence.
- information related to suspect and victim demographics is also included.
- This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website.

```
url_NYPD <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
NYPD <- read.csv(url_NYPD)
```

## Libraries

```
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(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(ggplot2)
library(broom)
```

## Data Exploration

```
dim(NYPD)
```

```
## [1] 28562    21
```

```
tail(NYPD)
```

```
##      INCIDENT_KEY OCCUR_DATE OCCUR_TIME      BORO LOC_OF_OCCUR_DESC PRECINCT
## 28557      270719378 07/02/2023   21:40:00    BRONX          OUTSIDE      46
## 28558      265354835 03/19/2023   23:48:00    BRONX           INSIDE      47
## 28559      272968931 08/16/2023   02:46:00    BRONX          OUTSIDE      41
## 28560      270489846 06/27/2023   12:27:00    BRONX           INSIDE      41
## 28561      271021661 07/08/2023   11:27:00   QUEENS          OUTSIDE     102
## 28562      271818283 07/24/2023   23:38:00  MANHATTAN          OUTSIDE      28
##      JURISDICTION_CODE LOC_CLASSFCTN_DESC      LOCATION_DESC
## 28557                0          STREET              (null)
## 28558                0        COMMERCIAL      GROCERY/BODEGA
## 28559                0          STREET              (null)
## 28560                0        DWELLING  MULTI DWELL - APT BUILD
## 28561                0          STREET      BEAUTY/NAIL SALON
## 28562                2        HOUSING  MULTI DWELL - PUBLIC HOUS
##      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX      PERP_RACE
## 28557                false      (null)  (null)      (null)
## 28558                true       18-24      M          BLACK
## 28559                false      25-44      F          BLACK
## 28560                true       25-44      M          BLACK
## 28561                false      25-44      M  WHITE  HISPANIC
## 28562                false      (null)  (null)      (null)
##      VIC_AGE_GROUP VIC_SEX      VIC_RACE X_COORD_CD Y_COORD_CD
## 28557       18-24      M          BLACK  HISPANIC   1009601   247515
## 28558       18-24      M          BLACK          1025687   268586
## 28559       45-64      M          BLACK          1014639   240066
## 28560       25-44      M          BLACK          1012221   238552
## 28561        65+      M  ASIAN / PACIFIC ISLANDER  1028856   192785
## 28562       25-44      M          BLACK          997853   230889
##      Latitude Longitude      Lon_Lat
## 28557 40.84601 -73.90837 POINT (-73.908369 40.846012)
## 28558 40.90378 -73.85010 POINT (-73.850098 40.903785)
## 28559 40.82555 -73.89020 POINT (-73.890195 40.825549)
## 28560 40.82140 -73.89894 POINT (-73.898938 40.821404)
## 28561 40.69572 -73.83914 POINT (-73.839138 40.695717)
## 28562 40.80040 -73.95086 POINT (-73.950864 40.800405)
```

```
#numerical statistics summary
summary(NYPD[, sapply(NYPD, is.numeric)])
```

```
## INCIDENT_KEY      PRECINCT      JURISDICTION_CODE      X_COORD_CD
## Min. : 9953245    Min. : 1.0    Min. :0.0000    Min. : 914928
## 1st Qu.: 65439914 1st Qu.: 44.0    1st Qu.:0.0000    1st Qu.:1000068
## Median : 92711254 Median : 67.0    Median :0.0000    Median :1007772
## Mean : 127405824 Mean : 65.5    Mean :0.3219    Mean :1009424
## 3rd Qu.:203131993 3rd Qu.: 81.0    3rd Qu.:0.0000    3rd Qu.:1016807
## Max. :279758069 Max. :123.0    Max. :2.0000    Max. :1066815
##
## NA's :2
## Y_COORD_CD      Latitude      Longitude
## Min. :125757    Min. :40.51    Min. : -74.25
## 1st Qu.:182912    1st Qu.:40.67    1st Qu.: -73.94
## Median :194901    Median :40.70    Median : -73.92
## Mean :208380    Mean :40.74    Mean : -73.91
## 3rd Qu.:239814    3rd Qu.:40.82    3rd Qu.: -73.88
## Max. :271128    Max. :40.91    Max. : -73.70
##
## NA's :59    NA's :59
```

```
#categorical statistics summary
summary(NYPD[, sapply(NYPD, is.factor)])
```

```
## < table of extent 0 x 0 >
```

```
# Looking at column values
glimpse(NYPD)
```

```
## Rows: 28,562
## Columns: 21
## $ INCIDENT_KEY      <int> 244608249, 247542571, 84967535, 202853370, 270~
## $ OCCUR_DATE        <chr> "05/05/2022", "07/04/2022", "05/27/2012", "09/~
## $ OCCUR_TIME        <chr> "00:10:00", "22:20:00", "19:35:00", "21:00:00"~
## $ BORO              <chr> "MANHATTAN", "BRONX", "QUEENS", "BRONX", "BROO~
## $ LOC_OF_OCCUR_DESC  <chr> "INSIDE", "OUTSIDE", "", "", "", "", "", "", "~
## $ PRECINCT          <int> 14, 48, 103, 42, 83, 23, 113, 77, 48, 49, 73, ~
## $ JURISDICTION_CODE  <int> 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ LOC_CLASSFCTN_DESC <chr> "COMMERCIAL", "STREET", "", "", "", "", "", ""~
## $ LOCATION_DESC      <chr> "VIDEO STORE", "(null)", "", "", "", "MULTI DW~
## $ STATISTICAL_MURDER_FLAG <chr> "true", "true", "false", "false", "false", "fa~
## $ PERP_AGE_GROUP     <chr> "25-44", "(null)", "", "25-44", "25-44", "", "~
## $ PERP_SEX           <chr> "M", "(null)", "", "M", "M", "", "", "", "", "~
## $ PERP_RACE           <chr> "BLACK", "(null)", "", "UNKNOWN", "BLACK", "", ~
## $ VIC_AGE_GROUP      <chr> "25-44", "18-24", "18-24", "25-44", "25-44", "~
## $ VIC_SEX            <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "~
## $ VIC_RACE            <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ X_COORD_CD         <dbl> 986050, 1016802, 1048632, 1014493, 1009149, 99~
## $ Y_COORD_CD         <dbl> 214231.0, 250581.0, 198262.0, 242565.0, 190104~
## $ Latitude           <dbl> 40.75469, 40.85440, 40.71063, 40.83242, 40.688~
## $ Longitude          <dbl> -73.99350, -73.88233, -73.76777, -73.89071, -7~
## $ Lon_Lat            <chr> "POINT (-73.9935 40.754692)", "POINT (-73.8823~
```

```
names(NYPD)
```

```
## [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"
```

## Data Preparation

```
# Convert categorical variables to factors
NYPD <- NYPD %>%
  mutate_at(vars(BORO, PERP_SEX, PERP_RACE, VIC_AGE_GROUP, VIC_SEX, VIC_RACE), factor)
# Remove unnecessary columns
NYPD <- NYPD %>% select(-Lon_Lat)

summary(NYPD)
```

```
## INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME
## Min. : 9953245     Length:28562     Length:28562
## 1st Qu.: 65439914   Class :character   Class :character
## Median : 92711254   Mode  :character   Mode  :character
## Mean : 127405824
## 3rd Qu.:203131993
## Max. : 279758069
##
##          BORO          LOC_OF_OCCUR_DESC    PRECINCT    JURISDICTION_CODE
## BRONX      : 8376     Length:28562         Min. : 1.0    Min. :0.0000
## BROOKLYN   :11346     Class :character   1st Qu.: 44.0  1st Qu.:0.0000
## MANHATTAN  : 3762     Mode  :character   Median : 67.0  Median :0.0000
## QUEENS     : 4271
## STATEN ISLAND: 807    3rd Qu.: 81.0  3rd Qu.:0.0000
##                                     Max. :123.0  Max. :2.0000
##                                     NA's :2
## LOC_CLASSFCTN_DESC LOCATION_DESC    STATISTICAL_MURDER_FLAG
## Length:28562      Length:28562      Length:28562
## Class :character   Class :character   Class :character
## Mode :character    Mode :character    Mode :character
##
##
##
## PERP_AGE_GROUP      PERP_SEX          PERP_RACE      VIC_AGE_GROUP
## Length:28562        : 9310    BLACK          :11903    <18 : 2954
## Class :character    (null): 1141          : 9310    1022 : 1
## Mode :character     F : 444    WHITE HISPANIC: 2510  18-24 :10384
##                     M :16168  UNKNOWN         : 1837    25-44 :12973
##                     U : 1499  BLACK HISPANIC: 1392  45-64 : 1981
```

```
##                               (null)      : 1141  65+      : 205
##                               (Other)     :   469  UNKNOWN:   64
## VIC_SEX                       VIC_RACE     X_COORD_CD
## F: 2760  AMERICAN INDIAN/ALASKAN NATIVE:   11  Min.    : 914928
## M:25790  ASIAN / PACIFIC ISLANDER      :  440  1st Qu.:1000068
## U:   12  BLACK                          :20235  Median :1007772
##          BLACK HISPANIC                 : 2795  Mean   :1009424
##          UNKNOWN                        :   70  3rd Qu.:1016807
##          WHITE                          :   728  Max.   :1066815
##          WHITE HISPANIC                 : 4283
## Y_COORD_CD      Latitude      Longitude
## Min.    :125757  Min.    :40.51  Min.    :-74.25
## 1st Qu.:182912  1st Qu.:40.67  1st Qu.: -73.94
## Median :194901  Median :40.70  Median : -73.92
## Mean   :208380  Mean   :40.74  Mean   : -73.91
## 3rd Qu.:239814  3rd Qu.:40.82  3rd Qu.: -73.88
## Max.   :271128  Max.   :40.91  Max.   : -73.70
##                               NA's      :59      NA's      :59
```

```
sapply(NYPD, function(x) sum(is.na(x)))
```

```
##          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
##              59              59
```

```
names(NYPD)
```

```
## [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"
```

# Exploring Incident Time Series Data

## Total Yearly Incidents

```
# Change OCCUR_DATE to Date format
NYPD$OCCUR_DATE <- as.Date(NYPD$OCCUR_DATE, format = "%m/%d/%Y")

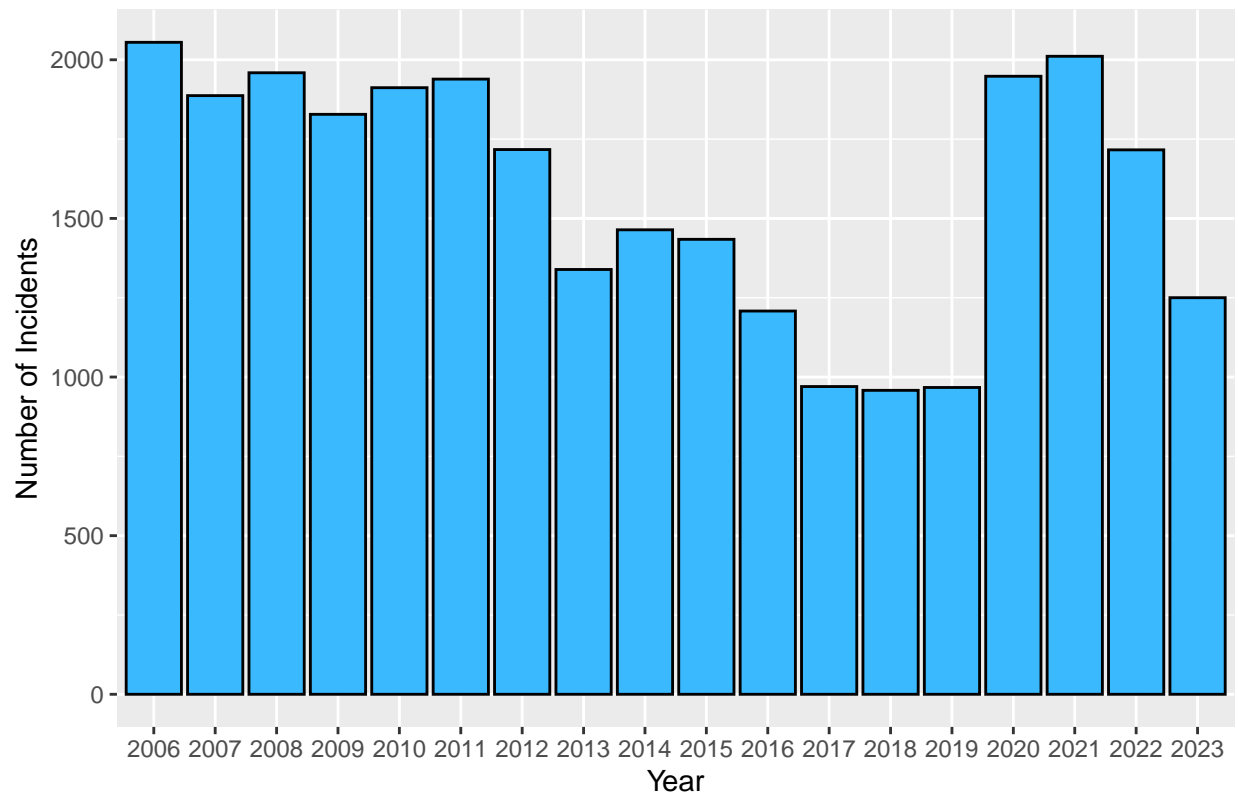
# Get year
NYPD$YEAR <- format(NYPD$OCCUR_DATE, "%Y")

# Group by year
yearly_incidents <- NYPD %>%
  group_by(YEAR) %>%
  summarise(Incidents = n())

# Plot total yearly incidents
g <- ggplot(yearly_incidents, aes(x = YEAR, y = Incidents)) +
  geom_col(fill = "#3BB9FF", colour = "black") +
  labs(title = "Total Yearly Incidents in NYC",
       x = "Year",
       y = "Number of Incidents") +
  theme(legend.position = "none")

# Print the plot
print(g)
```

## Total Yearly Incidents in NYC



## Total Daily Incidents

```
# Get day
NYPD$DAY_OF_WEEK <- weekdays(NYPD$OCCUR_DATE)

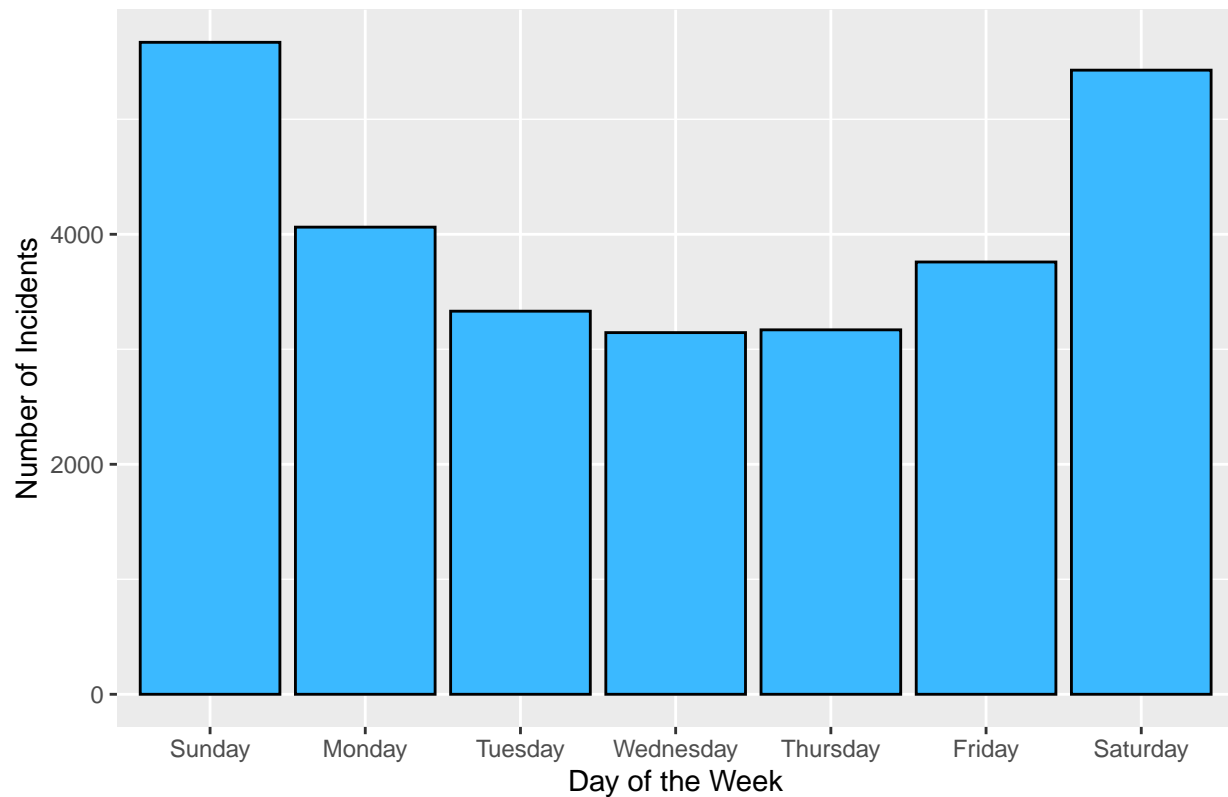
# Group totals by day of week
daily_incidents <- NYPD %>%
  group_by(DAY_OF_WEEK) %>%
  summarise(Incidents = n())

order_days <- c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")

# Plot total daily incidents
g <- ggplot(daily_incidents, aes(x = factor(DAY_OF_WEEK, levels = order_days), y = Incidents)) +
  geom_col(fill = "#3BB9FF", color = "black") +
  labs(title = "Total Incidents by Day of the Week in NYC",
       x = "Day of the Week",
       y = "Number of Incidents") +
  theme(legend.position = "none")

# Print the plot
print(g)
```

Total Incidents by Day of the Week in NYC



```
# Extract hour of the day from OCCUR_TIME
NYPD$HOUR_OF_DAY <- as.numeric(format(strptime(NYPD$OCCUR_TIME, format = "%H:%M:%S"), "%H"))

# Group data by hour of the day and count the number of incidents
hourly_incidents <- NYPD %>%
  group_by(HOUR_OF_DAY) %>%
  summarise(Incidents = n())

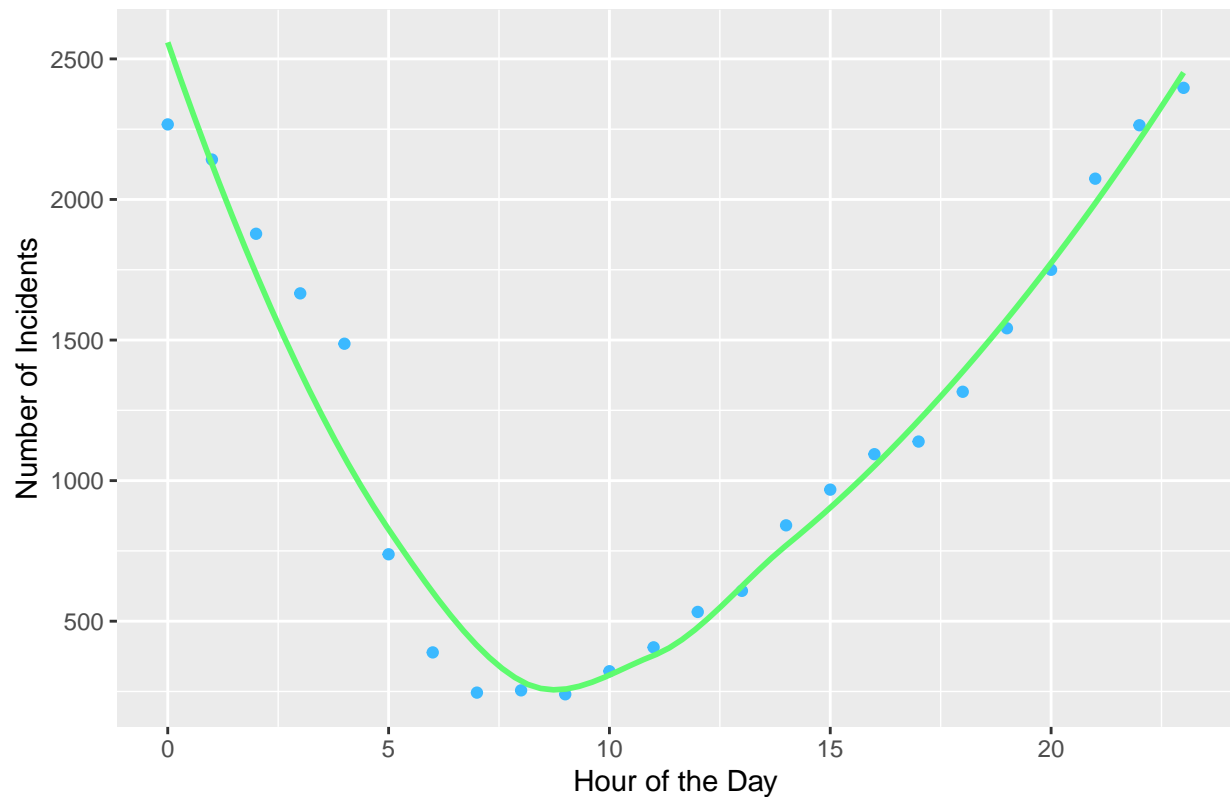
# Plot total incidents vs hour of day
g <- ggplot(hourly_incidents, aes(x = HOUR_OF_DAY, y = Incidents)) +
  geom_point(color = "#3BB9FF") +
  geom_smooth(method = "loess", se = FALSE, color = "#5EFB6E") +
  labs(title = "Total Incidents by Hour of the Day in NYC",
       x = "Hour of the Day",
       y = "Number of Incidents")

print(g)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



Total Incidents by Hour of the Day in NYC



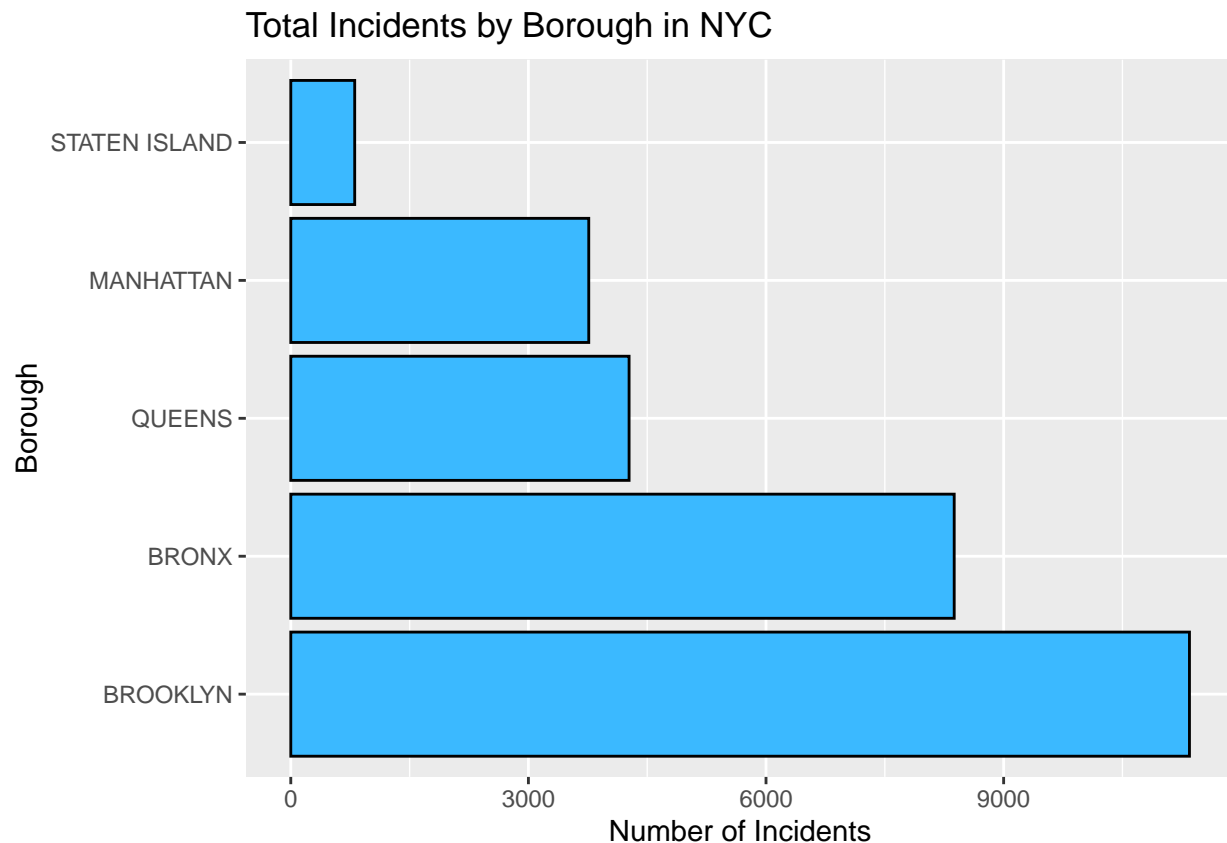
## Exploring Location Data

### Total Incidents by Borough

```
# Group data by BORO
boro_incidents <- NYPD %>%
  group_by(BORO) %>%
  summarise(Incidents = n())

# Plot total incidents by BORO
g <- ggplot(boro_incidents, aes(x = reorder(BORO, -Incidents), y = Incidents)) +
  geom_bar(stat = "identity", fill = "#3BB9FF", color = "black") +
  labs(title = "Total Incidents by Borough in NYC",
       x = "Borough",
       y = "Number of Incidents") +
  coord_flip()

print(g)
```



### Total Incidents by Location Classification

```
# Create a copy of the NYPD dataset
NYPD_copy <- NYPD

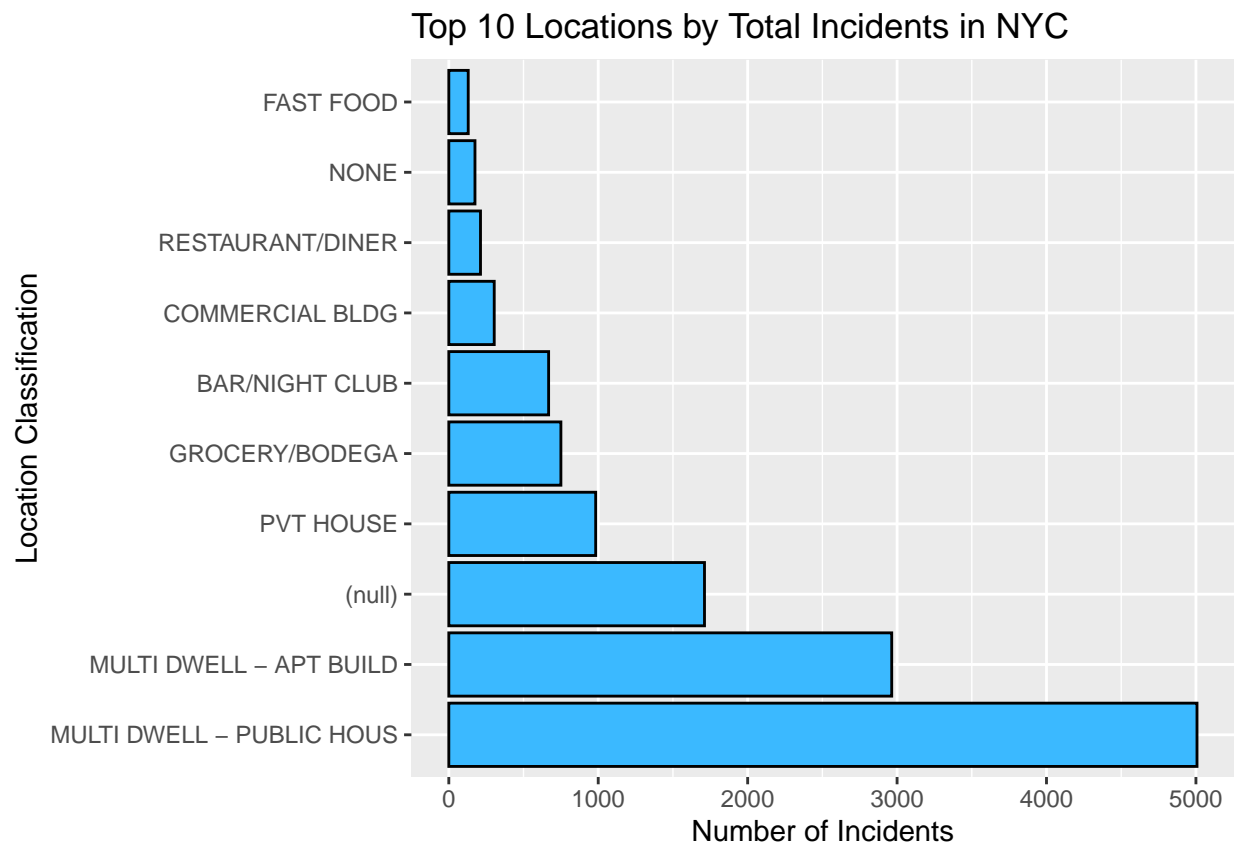
# Filter out blanks and NAs, then summarize incidents by location classification
classfctn_incidents <- NYPD_copy %>%
  filter(LOCATION_DESC != "", !is.na(LOCATION_DESC)) %>%
  group_by(LOCATION_DESC) %>%
  summarise(Incidents = n()) %>%
  arrange(desc(Incidents)) %>%
  top_n(10) # Select top 10
```

## Selecting by Incidents

```
# Plot total incidents for the top 10 location classifications
g <- ggplot(classfctn_incidents, aes(x = reorder(LOCATION_DESC, -Incidents), y = Incidents)) +
  geom_bar(stat = "identity", fill = "#3BB9FF", color = "black") +
  labs(title = "Top 10 Locations by Total Incidents in NYC ",
       x = "Location Classification",
       y = "Number of Incidents") +

  coord_flip()
```

```
# plot
print(g)
```



## Modeling Factors that can Contribute to the Fatality of a Shooting

```
NYPD$STATISTICAL_MURDER_FLAG <- as.integer(NYPD$STATISTICAL_MURDER_FLAG == "true")

# Fit the model with all specified variables
glm.fit <- glm(STATISTICAL_MURDER_FLAG ~ PERP_RACE + PERP_SEX + PERP_AGE_GROUP +
  VIC_AGE_GROUP + VIC_SEX + VIC_RACE + BORO + HOUR_OF_DAY +
  DAY_OF_WEEK +
  LOC_CLASSFCTN_DESC + LOCATION_DESC + LOC_OF_OCCUR_DESC +
  PRECINCT, data = NYPD, family = binomial)

# Summary of the model
summary(glm.fit)
```

```
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ PERP_RACE + PERP_SEX +
```

```

## PERP_AGE_GROUP + VIC_AGE_GROUP + VIC_SEX + VIC_RACE + BORO +
## HOUR_OF_DAY + DAY_OF_WEEK + LOC_CLASSFCTN_DESC + LOCATION_DESC +
## LOC_OF_OCCUR_DESC + PRECINCT, family = binomial, data = NYPD)
##
## Coefficients: (4 not defined because of singularities)
##
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.382e+01 1.475e+02 -0.094 0.925312
## PERP_RACE(null) -2.053e-01 1.109e-01 -1.852 0.064088
## PERP_RACEAMERICAN INDIAN/ALASKAN NATIVE -1.076e+01 3.753e+02 -0.029 0.977120
## PERP_RACEASIAN / PACIFIC ISLANDER 2.163e+00 4.849e-01 4.462 8.13e-06
## PERP_RACEBLACK 1.981e+00 4.506e-01 4.396 1.10e-05
## PERP_RACEBLACK HISPANIC 1.939e+00 4.548e-01 4.263 2.01e-05
## PERP_RACEUNKNOWN 1.342e+00 3.905e-01 3.437 0.000587
## PERP_RACEWHITE 2.388e+00 4.717e-01 5.063 4.14e-07
## PERP_RACEWHITE HISPANIC 2.139e+00 4.531e-01 4.721 2.34e-06
## PERP_SEX(null) NA NA NA NA
## PERP_SEXF -1.709e+00 2.876e-01 -5.941 2.83e-09
## PERP_SEXM -1.787e+00 2.646e-01 -6.752 1.46e-11
## PERP_SEXU NA NA NA NA
## PERP_AGE_GROUP(null) NA NA NA NA
## PERP_AGE_GROUP<18 -1.263e-01 4.394e-01 -0.288 0.773721
## PERP_AGE_GROUP1020 -1.247e+01 5.354e+02 -0.023 0.981414
## PERP_AGE_GROUP1028 -1.193e+01 5.354e+02 -0.022 0.982222
## PERP_AGE_GROUP18-24 -2.990e-02 4.355e-01 -0.069 0.945275
## PERP_AGE_GROUP224 -1.243e+01 5.354e+02 -0.023 0.981473
## PERP_AGE_GROUP25-44 1.849e-01 4.357e-01 0.424 0.671328
## PERP_AGE_GROUP45-64 4.787e-01 4.433e-01 1.080 0.280246
## PERP_AGE_GROUP65+ 2.470e-01 5.124e-01 0.482 0.629762
## PERP_AGE_GROUP940 -1.290e+01 5.354e+02 -0.024 0.980782
## PERP_AGE_GROUPUNKNOWN -2.686e+00 4.054e-01 -6.626 3.46e-11
## VIC_AGE_GROUP1022 -1.136e+01 5.354e+02 -0.021 0.983067
## VIC_AGE_GROUP18-24 2.556e-01 6.275e-02 4.074 4.62e-05
## VIC_AGE_GROUP25-44 5.090e-01 6.169e-02 8.251 < 2e-16
## VIC_AGE_GROUP45-64 5.509e-01 7.985e-02 6.899 5.22e-12
## VIC_AGE_GROUP65+ 9.300e-01 1.685e-01 5.518 3.44e-08
## VIC_AGE_GROUPUNKNOWN 4.567e-01 3.263e-01 1.400 0.161596
## VIC_SEXM 1.148e-01 5.284e-02 2.173 0.029743
## VIC_SEXU -2.863e-01 1.084e+00 -0.264 0.791668
## VIC_RACEASIAN / PACIFIC ISLANDER 1.190e+01 1.475e+02 0.081 0.935704
## VIC_RACEBLACK 1.174e+01 1.475e+02 0.080 0.936547
## VIC_RACEBLACK HISPANIC 1.154e+01 1.475e+02 0.078 0.937607
## VIC_RACEUNKNOWN 1.092e+01 1.475e+02 0.074 0.940955
## VIC_RACEWHITE 1.178e+01 1.475e+02 0.080 0.936317
## VIC_RACEWHITE HISPANIC 1.180e+01 1.475e+02 0.080 0.936233
## BOROBROOKLYN 2.266e-02 8.713e-02 0.260 0.794832
## BOROMANHATTAN -1.142e-01 7.384e-02 -1.546 0.122019
## BOROQUEENS -5.586e-02 1.721e-01 -0.325 0.745504
## BOROSTATEN ISLAND -1.360e-01 2.209e-01 -0.616 0.538101
## HOUR_OF_DAY -2.131e-04 1.930e-03 -0.110 0.912085
## DAY_OF_WEEKMonday 1.427e-02 5.875e-02 0.243 0.808051
## DAY_OF_WEEKSaturday -8.237e-02 5.608e-02 -1.469 0.141842
## DAY_OF_WEEKSunday 2.060e-02 5.551e-02 0.371 0.710579
## DAY_OF_WEEKThursday -2.864e-02 6.260e-02 -0.458 0.647260
## DAY_OF_WEEKTuesday -8.384e-02 6.261e-02 -1.339 0.180542

```

## DAY_OF_WEEK	Wednesday	5.175e-02	6.185e-02	0.837	0.402791
## LOC_CLASSFCN_DESC	(null)	-1.260e+01	3.398e+02	-0.037	0.970432
## LOC_CLASSFCN_DESC	COMMERCIAL	-2.238e-01	2.101e-01	-1.065	0.286868
## LOC_CLASSFCN_DESC	DWELLING	-3.290e-01	2.043e-01	-1.610	0.107295
## LOC_CLASSFCN_DESC	HOUSING	-4.979e-01	1.508e-01	-3.301	0.000965
## LOC_CLASSFCN_DESC	OTHER	-4.200e-01	4.419e-01	-0.950	0.341927
## LOC_CLASSFCN_DESC	PARKING LOT	-6.761e-01	7.903e-01	-0.855	0.392297
## LOC_CLASSFCN_DESC	PLAYGROUND	1.231e+00	3.698e-01	3.330	0.000869
## LOC_CLASSFCN_DESC	STREET	-7.828e-02	1.254e-01	-0.624	0.532382
## LOC_CLASSFCN_DESC	TRANSIT	-4.929e-01	5.964e-01	-0.826	0.408555
## LOC_CLASSFCN_DESC	VEHICLE	1.785e+00	4.091e-01	4.362	1.29e-05
## LOCATION_DESC	(null)	-1.193e-01	1.369e-01	-0.871	0.383584
## LOCATION_DESC	ATM	-1.190e+01	5.354e+02	-0.022	0.982268
## LOCATION_DESC	BANK	-1.200e+01	3.086e+02	-0.039	0.968992
## LOCATION_DESC	BAR/NIGHT CLUB	2.549e-01	1.033e-01	2.468	0.013590
## LOCATION_DESC	BEAUTY/NAIL SALON	3.645e-01	2.250e-01	1.620	0.105156
## LOCATION_DESC	CANDY STORE	8.980e-01	8.633e-01	1.040	0.298256
## LOCATION_DESC	CHAIN STORE	9.482e-01	9.009e-01	1.053	0.292540
## LOCATION_DESC	CHECK CASH	1.494e+01	5.354e+02	0.028	0.977739
## LOCATION_DESC	CLOTHING BOUTIQUE	3.192e-01	6.282e-01	0.508	0.611433
## LOCATION_DESC	COMMERCIAL BLDG	3.699e-01	1.486e-01	2.489	0.012802
## LOCATION_DESC	DEPT STORE	8.680e-01	7.146e-01	1.215	0.224464
## LOCATION_DESC	DOCTOR/DENTIST	-1.222e+01	5.354e+02	-0.023	0.981786
## LOCATION_DESC	DRUG STORE	1.157e+00	5.578e-01	2.073	0.038127
## LOCATION_DESC	DRY CLEANER/LAUNDRY	-9.006e-01	7.422e-01	-1.213	0.224995
## LOCATION_DESC	FACTORY/WAREHOUSE	8.809e-01	7.880e-01	1.118	0.263597
## LOCATION_DESC	FAST FOOD	1.770e-02	2.395e-01	0.074	0.941065
## LOCATION_DESC	GAS STATION	4.489e-01	2.833e-01	1.585	0.112998
## LOCATION_DESC	GROCERY/BODEGA	2.286e-01	9.753e-02	2.344	0.019078
## LOCATION_DESC	GYM/FITNESS FACILITY	8.080e-01	1.225e+00	0.660	0.509553
## LOCATION_DESC	HOSPITAL	-7.934e-01	4.156e-01	-1.909	0.056233
## LOCATION_DESC	HOTEL/MOTEL	1.018e+00	3.485e-01	2.922	0.003482
## LOCATION_DESC	JEWELRY STORE	-5.312e-01	7.904e-01	-0.672	0.501510
## LOCATION_DESC	LIQUOR STORE	1.095e+00	3.261e-01	3.358	0.000786
## LOCATION_DESC	LOAN COMPANY	1.479e+01	5.354e+02	0.028	0.977962
## LOCATION_DESC	MULTI DWELL - APT BUILD	6.041e-01	4.973e-02	12.147	< 2e-16
## LOCATION_DESC	MULTI DWELL - PUBLIC HOUS	1.541e-01	4.487e-02	3.435	0.000593
## LOCATION_DESC	NONE	2.879e-01	1.994e-01	1.444	0.148814
## LOCATION_DESC	PHOTO/COPY STORE	-1.219e+01	5.354e+02	-0.023	0.981830
## LOCATION_DESC	PVT HOUSE	8.525e-01	7.577e-02	11.250	< 2e-16
## LOCATION_DESC	RESTAURANT/DINER	-1.759e-01	1.980e-01	-0.888	0.374324
## LOCATION_DESC	SCHOOL	1.393e+01	5.354e+02	0.026	0.979244
## LOCATION_DESC	SHOE STORE	9.036e-01	6.553e-01	1.379	0.167938
## LOCATION_DESC	SMALL MERCHANT	9.912e-01	3.457e-01	2.867	0.004143
## LOCATION_DESC	SOCIAL CLUB/POLICY LOCATI	4.844e-01	2.746e-01	1.764	0.077745
## LOCATION_DESC	STORAGE FACILITY	-1.230e+01	5.354e+02	-0.023	0.981671
## LOCATION_DESC	STORE UNCLASSIFIED	5.456e-01	3.891e-01	1.402	0.160808
## LOCATION_DESC	SUPERMARKET	5.059e-01	5.393e-01	0.938	0.348214
## LOCATION_DESC	TELECOMM. STORE	1.736e+00	6.224e-01	2.789	0.005287
## LOCATION_DESC	VARIETY STORE	3.984e-01	8.274e-01	0.482	0.630149
## LOCATION_DESC	VIDEO STORE	3.081e+00	1.081e+00	2.849	0.004388
## LOC_OF_OCCUR_DESC	INSIDE	5.731e-01	1.683e-01	3.405	0.000662
## LOC_OF_OCCUR_DESC	OUTSIDE	NA	NA	NA	NA
## PRECINCT		-3.341e-05	2.621e-03	-0.013	0.989828

```

##
## (Intercept)
## PERP_RACE(null) .
## PERP_RACEAMERICAN INDIAN/ALASKAN NATIVE
## PERP_RACEASIAN / PACIFIC ISLANDER ***
## PERP_RACEBLACK ***
## PERP_RACEBLACK HISPANIC ***
## PERP_RACEUNKNOWN ***
## PERP_RACEWHITE ***
## PERP_RACEWHITE HISPANIC ***
## PERP_SEX(null)
## PERP_SEXF ***
## PERP_SEXM ***
## PERP_SEXU
## PERP_AGE_GROUP(null)
## PERP_AGE_GROUP<18
## PERP_AGE_GROUP1020
## PERP_AGE_GROUP1028
## PERP_AGE_GROUP18-24
## PERP_AGE_GROUP224
## PERP_AGE_GROUP25-44
## PERP_AGE_GROUP45-64
## PERP_AGE_GROUP65+
## PERP_AGE_GROUP940
## PERP_AGE_GROUPUNKNOWN ***
## VIC_AGE_GROUP1022
## VIC_AGE_GROUP18-24 ***
## VIC_AGE_GROUP25-44 ***
## VIC_AGE_GROUP45-64 ***
## VIC_AGE_GROUP65+ ***
## VIC_AGE_GROUPUNKNOWN
## VIC_SEXM *
## VIC_SEXU
## VIC_RACEASIAN / PACIFIC ISLANDER
## VIC_RACEBLACK
## VIC_RACEBLACK HISPANIC
## VIC_RACEUNKNOWN
## VIC_RACEWHITE
## VIC_RACEWHITE HISPANIC
## BOROBROOKLYN
## BOROMANHATTAN
## BOROQUEENS
## BOROSTATEN ISLAND
## HOUR_OF_DAY
## DAY_OF_WEEKMonday
## DAY_OF_WEEKSaturday
## DAY_OF_WEEKSunday
## DAY_OF_WEEKThursday
## DAY_OF_WEEKTuesday
## DAY_OF_WEEKWednesday
## LOC_CLASSFCTN_DESC(null)
## LOC_CLASSFCTN_DESCCOMMERCIAL
## LOC_CLASSFCTN_DESCDWELLING
## LOC_CLASSFCTN_DESCHOUSING ***

```

```

## LOC_CLASSFCTN_DESCOTHER
## LOC_CLASSFCTN_DESCPARKING LOT
## LOC_CLASSFCTN_DESCPLAYGROUND      ***
## LOC_CLASSFCTN_DESCSTREET
## LOC_CLASSFCTN_DESCTRANSIT
## LOC_CLASSFCTN_DESCVEHICLE        ***
## LOCATION_DESC(null)
## LOCATION_DESCATM
## LOCATION_DESCBANK
## LOCATION_DESCBAR/NIGHT CLUB      *
## LOCATION_DESCBEAUTY/NAIL SALON
## LOCATION_DESCCANDY STORE
## LOCATION_DESCCHAIN STORE
## LOCATION_DESCCHECK CASH
## LOCATION_DESCCLOTHING BOUTIQUE
## LOCATION_DESCCOMMERCIAL BLDG      *
## LOCATION_DESCDEPT STORE
## LOCATION_DESCDOCTOR/DENTIST
## LOCATION_DESCDRUG STORE          *
## LOCATION_DESCDRY CLEANER/LAUNDRY
## LOCATION_DESCFACTORY/WAREHOUSE
## LOCATION_DESCFAST FOOD
## LOCATION_DESCGAS STATION
## LOCATION_DESCGROCERY/BODEGA      *
## LOCATION_DESCGYM/FITNESS FACILITY
## LOCATION_DESCHOSPITAL            .
## LOCATION_DESCHOTEL/MOTEL         **
## LOCATION_DESCJEWELRY STORE
## LOCATION_DESCLIQUOR STORE        ***
## LOCATION_DESCLOAN COMPANY
## LOCATION_DESCMULTI DWELL - APT BUILD ***
## LOCATION_DESCMULTI DWELL - PUBLIC HOUS ***
## LOCATION_DESCNONE
## LOCATION_DESCPHOTO/COPY STORE
## LOCATION_DESCPVT HOUSE           ***
## LOCATION_DESCRESTAURANT/DINER
## LOCATION_DESCSCHOOL
## LOCATION_DESCSHOE STORE
## LOCATION_DESCSMALL MERCHANT      **
## LOCATION_DESCSOCIAL CLUB/POLICY LOCATI .
## LOCATION_DESCSTORAGE FACILITY
## LOCATION_DESCSTORE UNCLASSIFIED
## LOCATION_DESCSUPERMARKET
## LOCATION_DESCTELECOMM. STORE     **
## LOCATION_DESCVARIETY STORE
## LOCATION_DESCVIDEO STORE         **
## LOC_OF_OCCUR_DESCINSIDE           ***
## LOC_OF_OCCUR_DESCOUTSIDE
## PRECINCT
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##

```

```
## Null deviance: 28061 on 28561 degrees of freedom
## Residual deviance: 26312 on 28464 degrees of freedom
## AIC: 26508
##
## Number of Fisher Scoring iterations: 12
```

```
# Obtain tidy version of the glm model
tidy_glm <- tidy(glm.fit)

# Remove NA coefficients and sort the coefficients in descending order
sorted_tidy_glm <- tidy_glm %>%
  filter(!is.na(estimate)) %>%
  arrange(desc(estimate))

# Print the sorted table of coefficients
print(sorted_tidy_glm)
```

```
## # A tibble: 98 x 5
##   term                                estimate std.error statistic p.value
##   <chr>                                <dbl>     <dbl>     <dbl>  <dbl>
## 1 LOCATION_DESCCHECK CASH              14.9      535.      0.0279 0.978
## 2 LOCATION_DESCLOAN COMPANY            14.8      535.      0.0276 0.978
## 3 LOCATION_DESCSCHOOL                  13.9      535.      0.0260 0.979
## 4 VIC_RACEASIAN / PACIFIC ISLANDER     11.9      147.      0.0807 0.936
## 5 VIC_RACEWHITE HISPANIC               11.8      147.      0.0800 0.936
## 6 VIC_RACEWHITE                       11.8      147.      0.0799 0.936
## 7 VIC_RACEBLACK                       11.7      147.      0.0796 0.937
## 8 VIC_RACEBLACK HISPANIC               11.5      147.      0.0783 0.938
## 9 VIC_RACEUNKNOWN                     10.9      147.      0.0741 0.941
## 10 LOCATION_DESCVIDEO STORE              3.08       1.08      2.85 0.00439
## # i 88 more rows
```

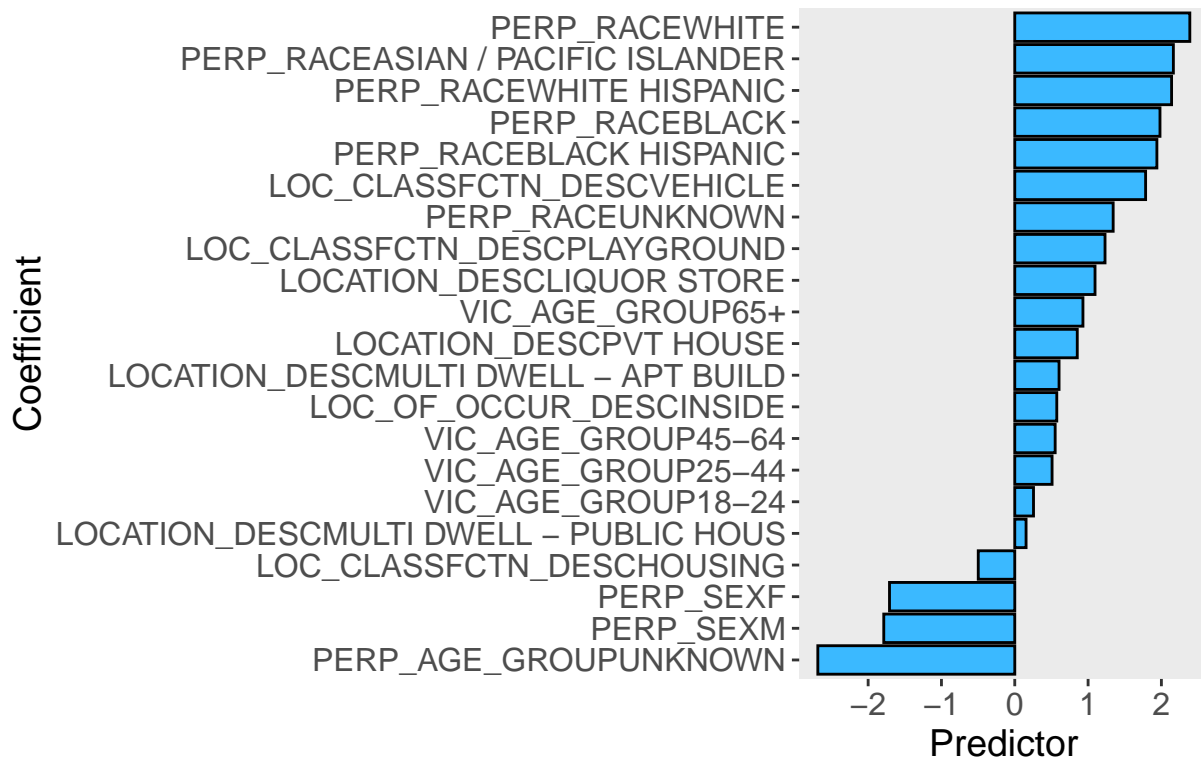
```
tidy_glm <- tidy(glm.fit)

# Remove rows with NA coefficients, non-significant p-values, and the intercept
tidy_glm <- tidy_glm %>%
  filter(!is.na(estimate) & p.value < 0.001 & term != "(Intercept)")

# Create a plot
ggplot(tidy_glm, aes(x = reorder(term, estimate), y = estimate)) +
  geom_col(fill = "#3BB9FF", color = "black", show.legend = FALSE) +
  coord_flip() +
  labs(x = "Coefficient", y = "Predictor", title = "Significant Coefficients of Predictors") +
  theme(
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12),
    axis.title = element_text(size = 14),
    plot.title = element_text(size = 16, face = "bold", hjust = 3), # Align title to the left properly
    panel.grid.major = element_blank(), # Remove major grid lines
    panel.grid.minor = element_blank(), # Remove minor grid lines
    plot.margin = margin(10, 10, 10, 10) # Adjust plot margins if needed
  )
```



## t Coefficients of Predictors



“

## Conclusion

- **NYC Incident Trends:** Declining incidents with a sharp drop present, possibly due to global factors.
- **Day of the Week:** Increased incidents on weekends
- **Hourly Distribution:** most incidents during evening hours; fewest in the early morning.
- **Borough Analysis:** Higher incident rates in Brooklyn and the Bronx; Manhattan has fewer murder-classified incidents.
- **Location:** Incidents are most common in public housing units

### Factors Which Contribute to Fatality - Logistic Regression Insights:

- **Race:** Specific racial groups such as White, Asian/Pacific Islander, Black, and Black Hispanic have higher odds of incidents being classified as murders, underscoring the influence of racial dynamics on shooting outcomes.
- **Location Type:** Video stores and drug stores, as indoor locations, show strong associations with murder classifications, likely due to the close-range nature of incidents in these settings.
- **Unknown Age Group:** The 'UNKNOWN' age category for perpetrators is significantly associated with fatal outcomes, indicating challenges or risks associated with unidentified offenders.

- **Victim Age:** Older victims, particularly those over 65, are more frequently involved in fatal shootings, pointing to their increased vulnerability.
- **Perpetrator Gender:** The presence of an “UNKNOWN” category for gender highlights potential ambiguities in the data, affecting the analysis of gender impacts on murder classifications.

Summary: The analysis reveals significant racial and locational factors in murder classifications within NYPD shooting incidents, with additional complexity introduced by unknown demographic data. These findings highlight the critical areas for law enforcement focus and further research to address and mitigate the risks associated with these key predictors.

- **Potential Biases:**

- **Data Collection Bias:** Risks due to non-representative data or reporting variances.
- **Measurement Bias:** Inconsistencies in incident classification processes.
- **Omitted Variable Bias:** Missing variables could distort the analysis.
- **Modeling Bias:** The model may not accurately capture the complexity of incident classifications.