NYPD Shooting Assignment

Cody S

Import, tidy and analyze the NYPD Shooting Incident dataset obtained. Be sure your project is reproducible and contains some visualization and analysis. You may use the data to do any analysis that is of interest to you. You should include at least two visualizations and one model. Be sure to identify any bias possible in the data and in your analysis.

```
library(tidyverse)
library(lubridate)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
source_url <- 'https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAI

incident_df <- read.csv(source_url)
summary(incident_df)
```

```
  INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME            BORO
 Min.   :  9953245   Length:28562       Length:28562       Length:28562
 1st Qu.: 65439914   Class :character   Class :character   Class :character
 Median : 92711254   Mode  :character   Mode  :character   Mode  :character
 Mean   :127405824
 3rd Qu.:203131993
```

```
Max.   :279758069

LOC_OF_OCCUR_DESC     PRECINCT     JURISDICTION_CODE LOC_CLASSFCTN_DESC
Length:28562       Min.   :  1.0   Min.   :0.0000    Length:28562
Class :character   1st Qu.: 44.0   1st Qu.:0.0000    Class :character
Mode  :character   Median : 67.0   Median :0.0000    Mode  :character
                   Mean   : 65.5   Mean   :0.3219
                   3rd Qu.: 81.0   3rd Qu.:0.0000
                   Max.   :123.0   Max.   :2.0000
                                   NA's   :2
LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
Length:28562       Length:28562            Length:28562
Class :character   Class :character        Class :character
Mode  :character   Mode  :character        Mode  :character



  PERP_SEX           PERP_RACE          VIC_AGE_GROUP        VIC_SEX
Length:28562       Length:28562       Length:28562       Length:28562
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:28562       Min.   : 914928   Min.   :125757   Min.   :40.51
Class :character   1st Qu.:1000068   1st Qu.:182912   1st Qu.:40.67
Mode  :character   Median :1007772   Median :194901   Median :40.70
                   Mean   :1009424   Mean   :208380   Mean   :40.74
                   3rd Qu.:1016807   3rd Qu.:239814   3rd Qu.:40.82
                   Max.   :1066815   Max.   :271128   Max.   :40.91
                                                      NA's   :59
  Longitude        Lon_Lat
Min.   :-74.25   Length:28562
1st Qu.:-73.94   Class :character
Median :-73.92   Mode  :character
Mean   :-73.91
3rd Qu.:-73.88
Max.   :-73.70
NA's   :59
```

```r
glimpse(incident_df)
```

```
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, 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~
```

```r
desc_counts <- lapply(incident_df[, c("LOC_CLASSFCTN_DESC", "LOCATION_DESC", "PERP_RACE", "VI

print(desc_counts)
```

```
$LOC_CLASSFCTN_DESC

                (null)  COMMERCIAL    DWELLING     HOUSING       OTHER
      25596          2         208         243         460          59
PARKING LOT  PLAYGROUND      STREET     TRANSIT     VEHICLE
         15          41        1886          23          29

$LOCATION_DESC
```

	(null)	ATM
14977	1711	1
BANK	BAR/NIGHT CLUB	BEAUTY/NAIL SALON
3	668	119
CANDY STORE	CHAIN STORE	CHECK CASH
7	7	1
CLOTHING BOUTIQUE	COMMERCIAL BLDG	DEPT STORE
14	304	9
DOCTOR/DENTIST	DRUG STORE	DRY CLEANER/LAUNDRY
1	14	32
FACTORY/WAREHOUSE	FAST FOOD	GAS STATION
8	130	74
GROCERY/BODEGA	GYM/FITNESS FACILITY	HOSPITAL
750	4	77
HOTEL/MOTEL	JEWELRY STORE	LIQUOR STORE
35	14	42
LOAN COMPANY	MULTI DWELL - APT BUILD	MULTI DWELL - PUBLIC HOUS
1	2964	5007
NONE	PHOTO/COPY STORE	PVT HOUSE
175	1	983
RESTAURANT/DINER	SCHOOL	SHOE STORE
212	1	10
SMALL MERCHANT	SOCIAL CLUB/POLICY LOCATI	STORAGE FACILITY
44	73	1
STORE UNCLASSIFIED	SUPERMARKET	TELECOMM. STORE
37	21	11
VARIETY STORE	VIDEO STORE	
11	8	

$PERP_RACE

	(null)
9310	1141
AMERICAN INDIAN/ALASKAN NATIVE	ASIAN / PACIFIC ISLANDER
2	169
BLACK	BLACK HISPANIC
11903	1392
UNKNOWN	WHITE
1837	298
WHITE HISPANIC	
2510	

$VIC_RACE

```
AMERICAN INDIAN/ALASKAN NATIVE       ASIAN / PACIFIC ISLANDER
                            11                            440
                         BLACK                  BLACK HISPANIC
                         20235                           2795
                       UNKNOWN                          WHITE
                            70                            728
                 WHITE HISPANIC
                          4283

$LOC_OF_OCCUR_DESC

         INSIDE OUTSIDE
  25596     460    2506
```

```r
# Modify, reorder, and select columns in a pipeline
cleaned_df <- df %>%
  # Rename 'category' to 'type' and 'value' to 'score'
  rename(type = category, score = value) %>%

  # Reorder columns: put 'type' first, followed by 'id', and 'date' and 'score'
  select(type, id, date, score) %>%

  # Remove rows where 'score' is less than 15
  selec(score >= 15)

  # remove completely
  select(-bad_column)
```

- ☐ Keep cleaning, renaming, removing
- ☐ Aggregate by day for over-time viz
- ☐ Aggregate by month, boro
- ☐ Figure out a model?

```r
# make a nicer datetime column
clean_incident_df <- incident_df %>%
  mutate(datetime = as.POSIXct(paste(OCCUR_DATE, OCCUR_TIME), format="%m/%d/%Y %H:%M:%S")
  ) %>%


glimpse(clean_incident_df)
```

```
Rows: 28,562
Columns: 22
$ 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, 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~
$ datetime                <dttm> 2022-05-05 00:10:00, 2022-07-04 22:20:00, 201~
```

```
[1] 10
```

```
name <- 'cody'

paste('The name is',name)
```

'The name is cody'