

NYPD Shooting Incidents RMD

2022-11-08

```
##NYPD Shooting Incidents Data
```

```
###Libraries I being by importing the tidyverse & lubridate libraries.
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr  0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
## Loading required package: timechange
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(ggrepel)
```

```
###Load the Data After importing my libraries, I load the data from a csv located at the following url:
https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD
```

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

```
NYPDShooting <- read.csv(url)
```

```
summary(NYPDShooting)
```

```
##   INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min.   : 9953245 Length:25596 Length:25596 Length:25596
## 1st Qu.: 61593633 Class :character Class :character Class :character
## Median : 86437258 Mode  :character Mode  :character Mode  :character
## Mean   :112382648
## 3rd Qu.:166660833
```

```
## Max. :238490103
##
## PRECINCT JURISDICTION_CODE LOCATION_DESC STATISTICAL_MURDER_FLAG
## Min. : 1.00 Min. :0.0000 Length:25596 Length:25596
## 1st Qu.: 44.00 1st Qu.:0.0000 Class :character Class :character
## Median : 69.00 Median :0.0000 Mode :character Mode :character
## Mean : 65.87 Mean :0.3316
## 3rd Qu.: 81.00 3rd Qu.:0.0000
## Max. :123.00 Max. :2.0000
## NA's :2
## PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP
## Length:25596 Length:25596 Length:25596 Length:25596
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## VIC_SEX VIC_RACE X_COORD_CD Y_COORD_CD
## Length:25596 Length:25596 Min. : 914928 Min. :125757
## Class :character Class :character 1st Qu.:1000011 1st Qu.:182782
## Mode :character Mode :character Median :1007715 Median :194038
## Mean :1009455 Mean :207894
## 3rd Qu.:1016838 3rd Qu.:239429
## Max. :1066815 Max. :271128
##
## Latitude Longitude Lon_Lat
## Min. :40.51 Min. : -74.25 Length:25596
## 1st Qu.:40.67 1st Qu.: -73.94 Class :character
## Median :40.70 Median : -73.92 Mode :character
## Mean :40.74 Mean : -73.91
## 3rd Qu.:40.82 3rd Qu.: -73.88
## Max. :40.91 Max. : -73.70
##
```

###Cleaning the Data

Next, we want to see a cleaned up version of the data by removing some columns that I won't be using, formatting my date column, and adding a few more date options.

```
tidyNYPDShooting <- NYPDShooting %>%
  mutate(date = mdy(OCCUR_DATE), time = hms(OCCUR_TIME), borough = BORO, wasMurder = STATISTICAL_MURDER,
  select(date, time, borough, wasMurder, perpAgeGroup, perpSex, perpRace, vicAgeGroup, vicSex, vicRace)

summary(tidyNYPDShooting)
```

```
## date time borough
## Min. :2006-01-01 Min. :0S Length:25596
## 1st Qu.:2009-05-10 1st Qu.:3H 23M 0S Class :character
## Median :2012-08-26 Median :15H 10M 0S Mode :character
## Mean :2013-06-13 Mean :12H 39M 17.9910923581774S
## 3rd Qu.:2017-07-01 3rd Qu.:20H 45M 0S
## Max. :2021-12-31 Max. :23H 59M 0S
## wasMurder perpAgeGroup perpSex perpRace
```

```
## Length:25596      Length:25596      Length:25596      Length:25596
## Class :character  Class :character  Class :character  Class :character
## Mode :character   Mode :character   Mode :character   Mode :character
##
##
##
## vicAgeGroup        vicSex          vicRace
## Length:25596      Length:25596      Length:25596
## Class :character  Class :character  Class :character
## Mode :character   Mode :character   Mode :character
##
##
##
```

###Group by Year & Borough

Here I am grouping the incidents by the borough where the incident occurred over each year.

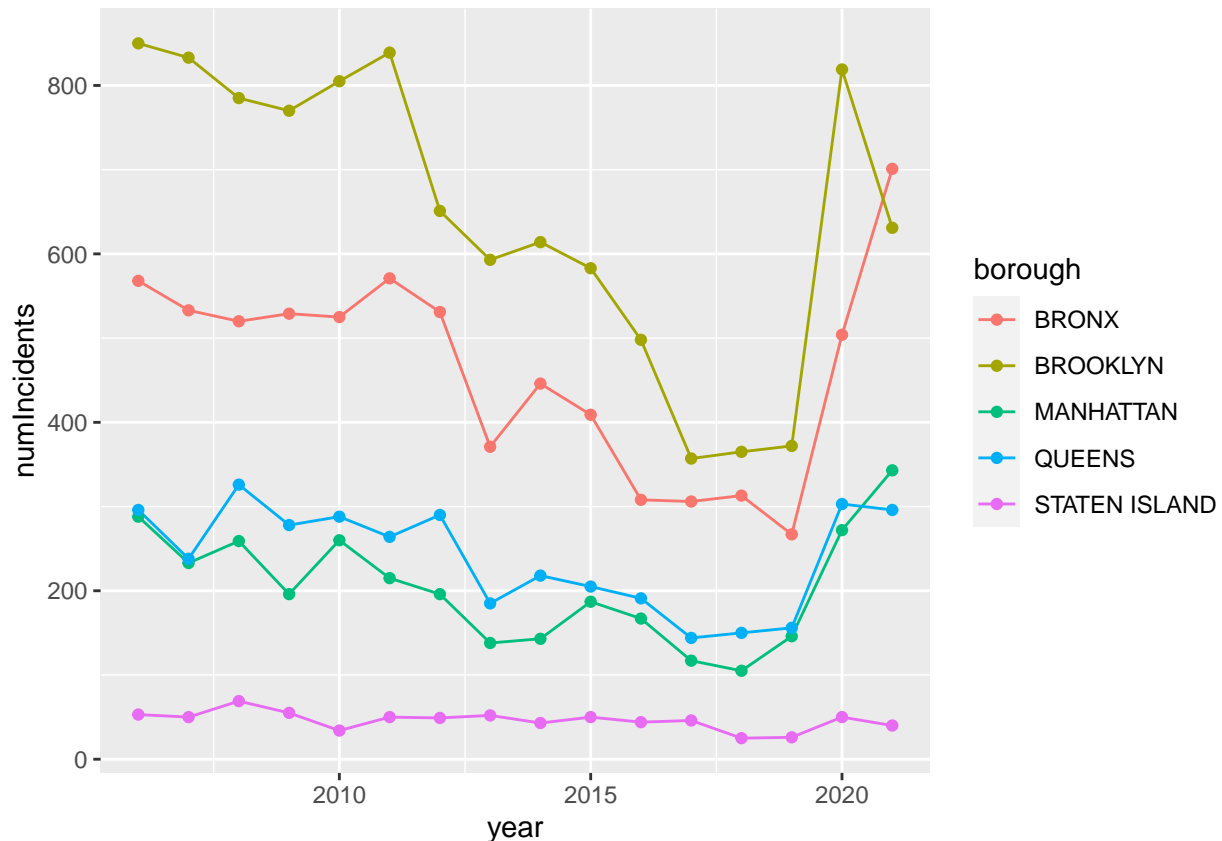
```
shootingsByYearBorough <- tidyNYPDSHooting %>%
  group_by(borough, year = year(date)) %>%
  tally() %>%
  mutate(numIncidents = n) %>%
  select(borough, year, numIncidents)

summary(shootingsByYearBorough)
```

```
##   borough      year      numIncidents
## Length:80      Min.    :2006      Min.    : 25.0
## Class :character 1st Qu.:2010      1st Qu.:143.8
## Mode :character  Median :2014      Median :275.0
##                  Mean   :2014      Mean   :319.9
##                  3rd Qu.:2017      3rd Qu.:508.0
##                  Max.    :2021      Max.    :850.0
```

###Visualize the Number of Incidents by Year & Borough Next, I want to visualize the shootings that occurred in each borough over each year. This shows a gradual decline from 2006 to 2019 and a significant increase after 2019. You can also see that Bronx & Brooklyn boroughs have more incidents each year than the other boroughs.

```
ggplot(data = shootingsByYearBorough, aes(x = year, y = numIncidents, group = borough)) +
  geom_line(aes(color = borough)) +
  geom_point(aes(color = borough))
```



Group by Month and Borough

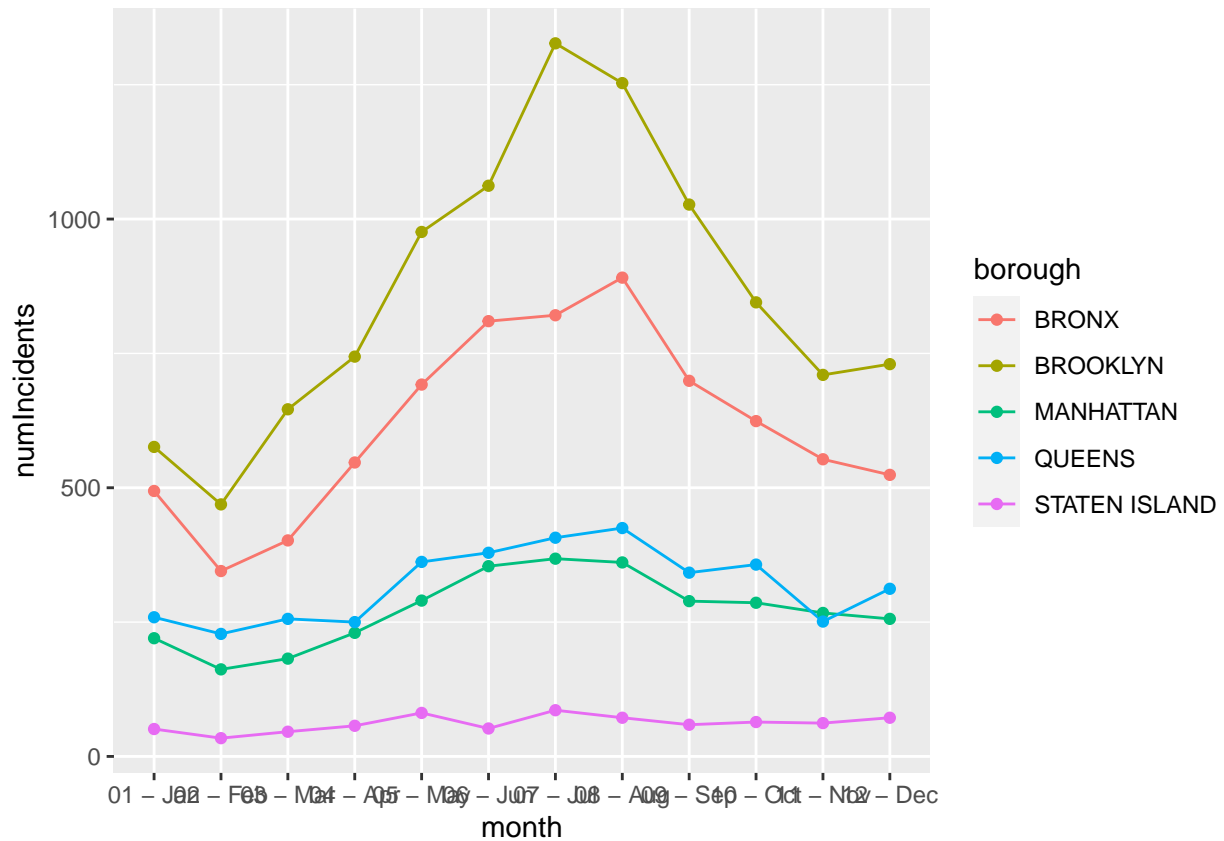
I wanted to see if there is a trend in the number of incidents based on seasonality by grouping the data by month and borough.

```
shootingsByMonthBorough <- tidyNYPDShooting %>%
  group_by(borough, month = paste(format(date, '%m'), month.abb[month(date)], sep = " - ")) %>%
  tally() %>%
  mutate(numIncidents = n) %>%
  select(borough, month, numIncidents)
```

Visualize the Number of Incidents by Month and Borough

Again, I visualize the data in a line graph. You can begin to see a trend in the seasonality of shootings now. There seem to be more shooting incidents in the warmer months, and this tends to go down in the colder months.

```
ggplot(data = shootingsByMonthBorough, aes(x = month, y = numIncidents, group = borough)) +
  geom_line(aes(color = borough)) +
  geom_point(aes(color = borough))
```



###Group by Hour & Borough

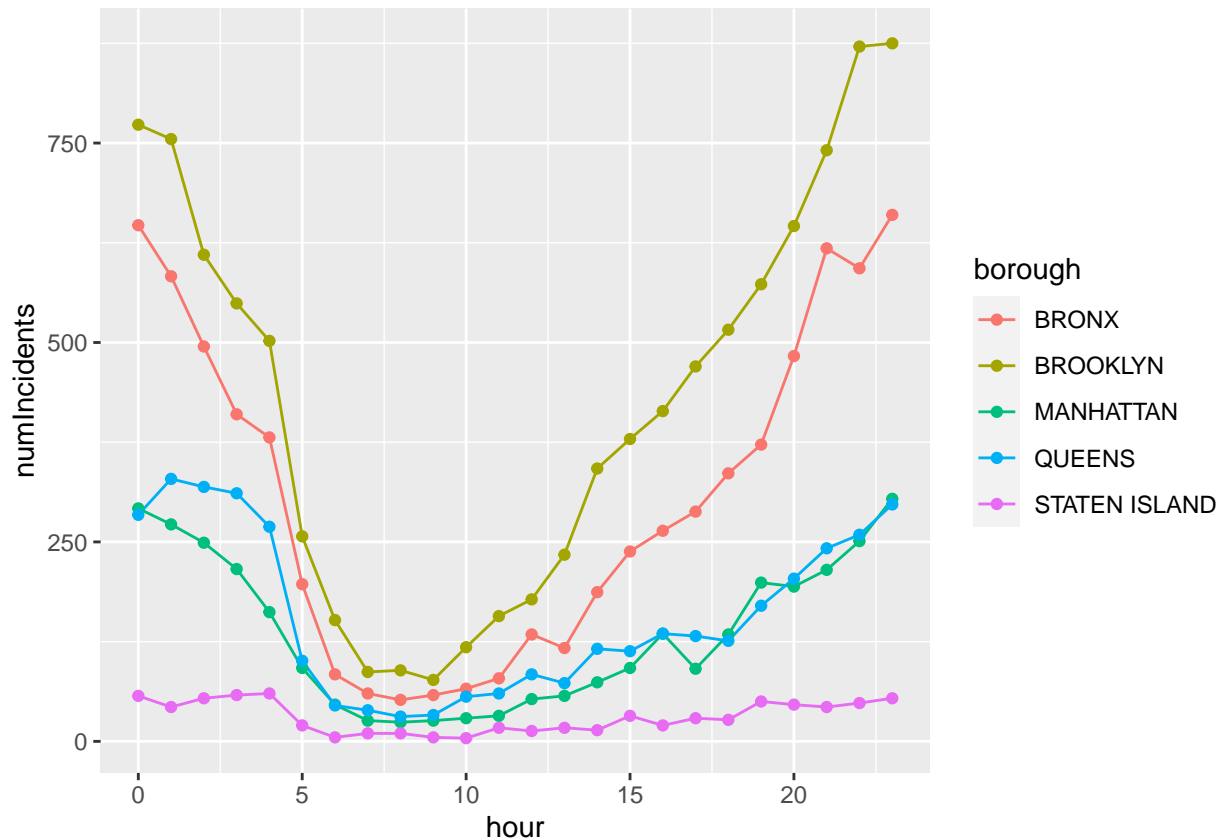
Here, I am further grouping the incidents by hour of day & borough in order to see any trends in incidents for different periods in the day.

```
shootingsByHourBorough <- tidyNYPDShooting %>%
  group_by(borough, hour = hour(time)) %>%
  tally() %>%
  mutate(numIncidents = n) %>%
  select(borough, hour, numIncidents)
```

###Visualize the Number of Incidents by Hour and Borough

Again, I am visualizing the hour data using a line graph. As you can see, there seem to be more shooting incidents in the evening, regardless of the borough.

```
ggplot(shootingsByHourBorough, aes(x = hour, y = numIncidents, group = borough)) +
  geom_line(aes(color = borough)) +
  geom_point(aes(color = borough))
```



Analyze the incidents by year, month, & hour for outliers.

I create a new dataset by grouping on year, month, & hour. Then, I look at the summary of incidents.

I can see that the minimum number of incidents is 1 and that is only 1 off from the 1st Quartile range. This tells me there are no significant outliers at the lower bound.

However, the max of 42 shooting incidents is 36 incidents more than the 3rd quartile at 8 incidents. This tells me I have some potential outliers to look at in the upper bound.

After calculating the $Q3 + 1.5IQR$ ($3 + 1.5 \cdot (3-1)$), I get an upper fence of 6 incidents. I use this upper fence as my filter on number of incidents to get my outliers.

I get a count of year month hour occurrences for each number of incidents between 6 and the max 26. I can see there is a slow decline in the occurrence of incidents greater than 15, so I further refine the outlier filter to greater than 15 incidents in for year, month, & hour.

Plotting this with ggrepel allows me to see the year month and hour that have the most occurrences. In May of 2021, there were 42 total shooting incidents at 11 pm for the entire month.

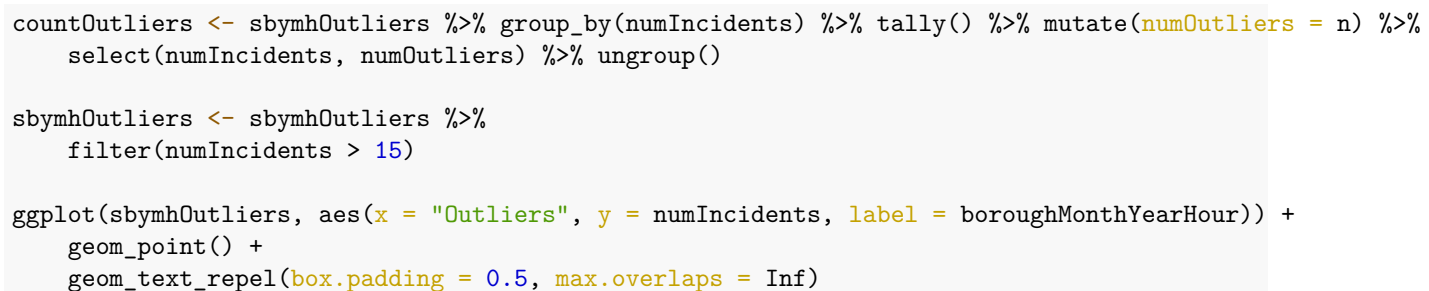
```
shootingsByBYMH <- tidyNYPDShooting %>%
  group_by(boroughMonthYearHour = paste(borough, ":", paste(format(date, '%m'), month.abb[month(date)]))
  tally() %>%
  mutate(numIncidents = n) %>%
  select(boroughMonthYearHour, numIncidents)

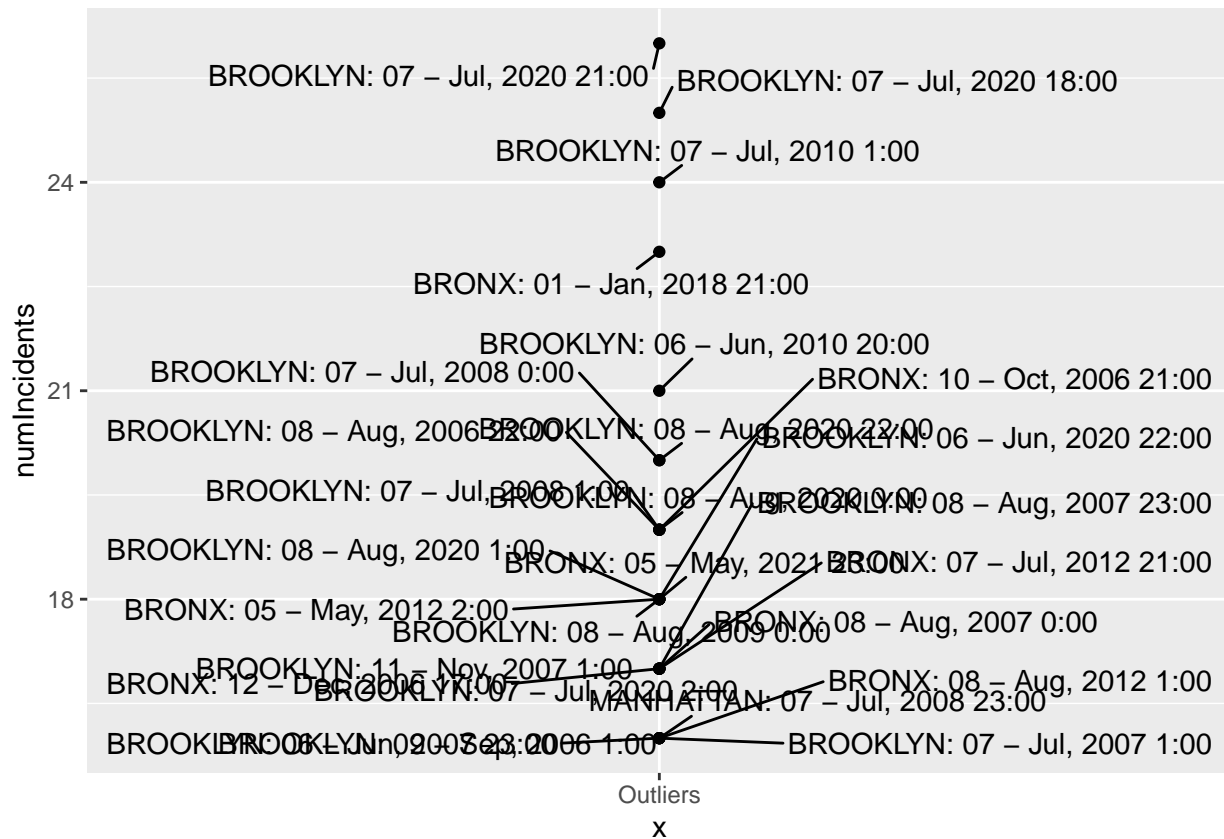
summary(shootingsByBYMH)
```

```
## boroughMonthYearHour numIncidents
```

```
sbymhOutliers <- shootingsByBYMH %>%
  filter(numIncidents > 6)

ggplot(sbymhOutliers, aes(x = "Outliers", y = numIncidents, label = boroughMonthYearHour)) +
  geom_point() +
  geom_text_repel(box.padding = 0.5, max.overlaps = Inf)
```





```
bymh26 <- NYPDShooting %>%
  mutate(date = mdy(OCCUR_DATE), time = hms(OCCUR_TIME)) %>%
  filter(BORO == "BROOKLYN" & month(date) == 7 & year(date) == 2020 & hour(time) == 21) %>%
  arrange(date,time)
```

bymh26

##	INCIDENT_KEY	OCCUR_DATE	OCCUR_TIME	BORO	PRECINCT	JURISDICTION_CODE
## 1	214983378	07/06/2020	21:45:00	BROOKLYN	77	0
## 2	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 3	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 4	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 5	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 6	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 7	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 8	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 9	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 10	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 11	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 12	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 13	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 14	215133068	07/09/2020	21:35:00	BROOKLYN	94	0
## 15	215133065	07/09/2020	21:43:00	BROOKLYN	75	2
## 16	215225152	07/10/2020	21:54:00	BROOKLYN	69	0
## 17	215223448	07/11/2020	21:21:00	BROOKLYN	75	0

## 18	215226535	07/12/2020	21:11:00	BROOKLYN	77	0
## 19	215225155	07/12/2020	21:50:00	BROOKLYN	73	0
## 20	215254848	07/13/2020	21:20:00	BROOKLYN	75	0
## 21	215255273	07/13/2020	21:45:00	BROOKLYN	69	2
## 22	215350311	07/15/2020	21:46:00	BROOKLYN	75	0
## 23	215536325	07/20/2020	21:30:00	BROOKLYN	81	0
## 24	215769254	07/25/2020	21:10:00	BROOKLYN	79	0
## 25	215770231	07/25/2020	21:30:00	BROOKLYN	81	0
## 26	215772436	07/26/2020	21:00:00	BROOKLYN	83	0
##	LOCATION_DESC	STATISTICAL_MURDER_FLAG	PERP_AGE_GROUP	PERP_SEX		
## 1		false				
## 2	MULTI DWELL - PUBLIC HOUS	true	18-24	M		
## 3	MULTI DWELL - PUBLIC HOUS	false	18-24	M		
## 4	MULTI DWELL - PUBLIC HOUS	true	25-44	M		
## 5	MULTI DWELL - PUBLIC HOUS	false	18-24	M		
## 6	MULTI DWELL - PUBLIC HOUS	true	18-24	M		
## 7	MULTI DWELL - PUBLIC HOUS	false	25-44	M		
## 8	MULTI DWELL - PUBLIC HOUS	true	18-24	M		
## 9	MULTI DWELL - PUBLIC HOUS	false	18-24	M		
## 10	MULTI DWELL - PUBLIC HOUS	true	25-44	M		
## 11	MULTI DWELL - PUBLIC HOUS	true	18-24	M		
## 12	MULTI DWELL - PUBLIC HOUS	false	25-44	M		
## 13	MULTI DWELL - PUBLIC HOUS	false	18-24	M		
## 14		false	45-64	M		
## 15	MULTI DWELL - PUBLIC HOUS	false				
## 16		false				
## 17		false				
## 18		false				
## 19		false				
## 20		false				
## 21	MULTI DWELL - PUBLIC HOUS	false				
## 22	PVT HOUSE	false				
## 23		false				
## 24		false				
## 25		false				
## 26		false	25-44	M		
##	PERP_RACE	VIC_AGE_GROUP	VIC_SEX	VIC_RACE	X_COORD_CD	Y_COORD_CD
## 1		<18	M	BLACK	1001252.8	184402.9
## 2	WHITE HISPANIC	25-44	M	WHITE HISPANIC	1021909.7	182502.6
## 3	BLACK	25-44	M	WHITE HISPANIC	1021909.7	182502.6
## 4	BLACK	25-44	M	WHITE HISPANIC	1021909.7	182502.6
## 5	BLACK	25-44	F	BLACK	1021909.7	182502.6
## 6	BLACK	25-44	M	WHITE HISPANIC	1021909.7	182502.6
## 7	BLACK	25-44	M	WHITE HISPANIC	1021909.7	182502.6
## 8	BLACK	25-44	F	BLACK	1021909.7	182502.6
## 9	WHITE HISPANIC	25-44	F	BLACK	1021909.7	182502.6
## 10	BLACK	25-44	F	BLACK	1021909.7	182502.6
## 11	WHITE HISPANIC	25-44	F	BLACK	1021909.7	182502.6
## 12	BLACK	25-44	F	BLACK	1021909.7	182502.6
## 13	WHITE HISPANIC	25-44	M	WHITE HISPANIC	1021909.7	182502.6
## 14	BLACK	25-44	M	WHITE HISPANIC	996191.2	207579.9
## 15		18-24	M	BLACK HISPANIC	1016132.9	178623.0
## 16		18-24	M	BLACK	1007165.8	171344.2
## 17		25-44	M	BLACK	1016505.6	176977.8

## 18		<18	M	BLACK	1005418.4	184431.3
## 19		18-24	M	WHITE HISPANIC	1006871.8	183675.8
## 20		18-24	M	BLACK	1014993.6	181458.4
## 21		45-64	M	BLACK	1013006.3	177560.0
## 22		18-24	M	WHITE HISPANIC	1016302.2	186693.0
## 23		<18	M	BLACK	1004121.9	186269.1
## 24		18-24	M	BLACK	1000285.8	186744.7
## 25		18-24	M	BLACK	1004121.9	186269.1
## 26	WHITE HISPANIC	25-44	M	WHITE HISPANIC	1008340.0	189454.0
##	Latitude Longitude				Lon_Lat	date
## 1	40.67281 -73.93871	POINT (-73.93870659699996	40.672812056000055)			2020-07-06
## 2	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 3	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 4	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 5	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 6	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 7	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 8	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 9	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 10	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 11	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 12	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 13	40.66753 -73.86425	POINT (-73.86424961899996	40.66753239900004)			2020-07-07
## 14	40.73644 -73.95691	POINT (-73.95691256899994	40.73643570000007)			2020-07-09
## 15	40.65691 -73.88509	POINT (-73.88509137999995	40.656906337000066)			2020-07-09
## 16	40.63696 -73.91743	POINT (-73.91743493999996	40.63695527900006)			2020-07-10
## 17	40.65239 -73.88376	POINT (-73.88375626399994	40.652389311000036)			2020-07-11
## 18	40.67288 -73.92369	POINT (-73.92368950699995	40.672880873000054)			2020-07-12
## 19	40.67080 -73.91845	POINT (-73.91845264899996	40.670803729000056)			2020-07-12
## 20	40.66469 -73.88918	POINT (-73.88918467299999	40.66469284000004)			2020-07-13
## 21	40.65400 -73.89636	POINT (-73.89636464599994	40.65399947100008)			2020-07-13
## 22	40.67906 -73.88444	POINT (-73.88444278999998	40.679056149000075)			2020-07-15
## 23	40.67793 -73.92836	POINT (-73.92835795199996	40.677928430000065)			2020-07-20
## 24	40.67924 -73.94219	POINT (-73.94218704299993	40.67924139200005)			2020-07-25
## 25	40.67793 -73.92836	POINT (-73.92835795199996	40.677928430000065)			2020-07-25
## 26	40.68666 -73.91314	POINT (-73.91313936599994	40.68665966800006)			2020-07-26
##	time					
## 1	21H 45M OS					
## 2	21H 28M OS					
## 3	21H 28M OS					
## 4	21H 28M OS					
## 5	21H 28M OS					
## 6	21H 28M OS					
## 7	21H 28M OS					
## 8	21H 28M OS					
## 9	21H 28M OS					
## 10	21H 28M OS					
## 11	21H 28M OS					
## 12	21H 28M OS					
## 13	21H 28M OS					
## 14	21H 35M OS					
## 15	21H 43M OS					
## 16	21H 54M OS					
## 17	21H 21M OS					

```
## 18 21H 11M OS
## 19 21H 50M OS
## 20 21H 20M OS
## 21 21H 45M OS
## 22 21H 46M OS
## 23 21H 30M OS
## 24 21H 10M OS
## 25 21H 30M OS
## 26 21H 0M OS
```

```
July7Shooting <- NYPDShooting %>%
  mutate(date = mdy(OCCUR_DATE), time = hms(OCCUR_TIME)) %>%
  filter(BORO == "BROOKLYN" & date == "2020-07-07" & hour(time) == 21) %>%
  arrange(desc(date), desc(time))
```

```
July7Shooting
```

##	INCIDENT_KEY	OCCUR_DATE	OCCUR_TIME	BORO	PRECINCT	JURISDICTION_CODE
## 1	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 2	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 3	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 4	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 5	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 6	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 7	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 8	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 9	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 10	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 11	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
## 12	215034244	07/07/2020	21:28:00	BROOKLYN	75	2
##	LOCATION_DESC		STATISTICAL_MURDER_FLAG	PERP_AGE_GROUP		PERP_SEX
## 1	MULTI DWELL - PUBLIC HOUS		true	18-24		M
## 2	MULTI DWELL - PUBLIC HOUS		false	18-24		M
## 3	MULTI DWELL - PUBLIC HOUS		true	25-44		M
## 4	MULTI DWELL - PUBLIC HOUS		false	18-24		M
## 5	MULTI DWELL - PUBLIC HOUS		true	18-24		M
## 6	MULTI DWELL - PUBLIC HOUS		false	25-44		M
## 7	MULTI DWELL - PUBLIC HOUS		true	18-24		M
## 8	MULTI DWELL - PUBLIC HOUS		false	18-24		M
## 9	MULTI DWELL - PUBLIC HOUS		true	25-44		M
## 10	MULTI DWELL - PUBLIC HOUS		true	18-24		M
## 11	MULTI DWELL - PUBLIC HOUS		false	25-44		M
## 12	MULTI DWELL - PUBLIC HOUS		false	18-24		M
##	PERP_RACE	VIC_AGE_GROUP	VIC_SEX	VIC_RACE	X_COORD_CD	Y_COORD_CD
## 1	WHITE HISPANIC	25-44	M	WHITE HISPANIC	1021910	182502.6
## 2	BLACK	25-44	M	WHITE HISPANIC	1021910	182502.6
## 3	BLACK	25-44	M	WHITE HISPANIC	1021910	182502.6
## 4	BLACK	25-44	F	BLACK	1021910	182502.6
## 5	BLACK	25-44	M	WHITE HISPANIC	1021910	182502.6
## 6	BLACK	25-44	M	WHITE HISPANIC	1021910	182502.6
## 7	BLACK	25-44	F	BLACK	1021910	182502.6
## 8	WHITE HISPANIC	25-44	F	BLACK	1021910	182502.6
## 9	BLACK	25-44	F	BLACK	1021910	182502.6
## 10	WHITE HISPANIC	25-44	F	BLACK	1021910	182502.6

## 11	BLACK	25-44	F	BLACK	1021910	182502.6
## 12	WHITE HISPANIC	25-44	M	WHITE HISPANIC	1021910	182502.6
##	Latitude	Longitude			Lon_Lat	date
## 1	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 2	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 3	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 4	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 5	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 6	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 7	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 8	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 9	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 10	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 11	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
## 12	40.66753	-73.86425	POINT (-73.86424961899996	40.66753239900004)		2020-07-07
##	time					
## 1	21H 28M 0S					
## 2	21H 28M 0S					
## 3	21H 28M 0S					
## 4	21H 28M 0S					
## 5	21H 28M 0S					
## 6	21H 28M 0S					
## 7	21H 28M 0S					
## 8	21H 28M 0S					
## 9	21H 28M 0S					
## 10	21H 28M 0S					
## 11	21H 28M 0S					
## 12	21H 28M 0S					