

# Covid-19 Impact on NYPD Shootings

05-28-2022

```
In [16]: # Download required packages
library(readr)
library(tidyverse)
library(ggplot2)
library(dplyr)
library(lubridate)
```

## Import Data

In this analysis, 2 datasets describing NYPD Shootings will be imported. The first dataset consists of shootings that occurred in New York between 2006 and 2020. The second dataset consists of more recent shootings that occurred between 2020 and 2021. The last dataset describes the number of Covid-19 cases that occurred in New York between 2020 and 2022. This dataset and the 2020-2021 NYPD Shooting datasets will be compared.

```
In [3]: # Import NYPD Historic Shooting data and NYPD 2021 Shooting Data
NYPD.shoot.hist = read.csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD")
NYPD.shoot.2021 = read.csv("https://data.cityofnewyork.us/api/views/suc2-vwe8/rows.csv?accessType=DOWNLOAD")
head(NYPD.shoot.hist)
head(NYPD.shoot.2021)

# Import New York Covid-19 Data
NY.covid19 = read.csv("https://health.data.ny.gov/api/views/xdss-u53e/rows.csv?accessType=DOWNLOAD&bom=true&format=true")
head(NY.covid19)
```

A data.frame: 6 × 19

	INCIDENT_KEY	OCCUR_DATE	OCCUR_TIME	BORO	PRECINCT	JURISDICTION_CODE	LOCATION_DESC	STATISTICAL_MURDER_FLAG	PERP_AGE_GROUP	PERP_SEX	PERP_RACE	VIC_AGE_GROUP	VIC_SEX	VIC_R
	<int>	<chr>	<chr>	<chr>	<int>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	24050482	08/27/2006	05:35:00	BRONX	52	0		true				25-44	F	BLU
2	77673979	03/11/2011	12:03:00	QUEENS	106	0		false				65+	M	WH
3	203350417	10/06/2019	01:09:00	BROOKLYN	77	0		false				18-24	F	BLU
4	80584527	09/04/2011	03:35:00	BRONX	40	0		false				<18	M	BLACK
5	90843766	05/27/2013	21:16:00	QUEENS	100	0		false				18-24	M	BLU
6	92393427	09/01/2013	04:17:00	BROOKLYN	67	0		false				<18	M	BLACK

A data.frame: 6 × 19

	INCIDENT_KEY	OCCUR_DATE	OCCUR_TIME	BORO	PRECINCT	JURISDICTION_CODE	LOCATION_DESC	STATISTICAL_MURDER_FLAG	PERP_AGE_GROUP	PERP_SEX	PERP_RACE	VIC_AGE_GROUP	VIC_SEX	VIC_R
	<int>	<chr>	<chr>	<chr>	<int>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	229643180	06/16/2021	21:34:00	BROOKLYN	73	0		false				25-44	M	BLU
2	233147632	09/03/2021	16:28:00	BRONX	43	0		false				18-24	F	BLU
3	231637053	07/31/2021	22:36:00	QUEENS	115	0		false				18-24	M	WH
4	238041594	12/17/2021	12:00:00	BRONX	50	0		false				25-44	M	BLU
5	228798560	05/27/2021	22:50:00	QUEENS	103	0		false	18-24	M	BLACK	25-44	M	BLU
6	226542151	04/05/2021	23:15:00	BROOKLYN	73	2	MULTI DWELL - PUBLIC HOUS	true	45-64	M	BLACK	<18	F	BLACK

A data.frame: 6 × 8

	Test.Date	County	New.Positives	Cumulative.Number.of.Positives	Total.Number.of.Tests.Performed	Cumulative.Number.of.Tests.Performed	Test...Positive	Geography
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	05/26/2022	Albany	114	65,615	1,017	1,297,384	10.86%	COUNTY
2	05/26/2022	Allegany	15	9,471	165	235,357	16.00%	COUNTY
3	05/26/2022	Bronx	540	411,532	11,992	7,879,283	3.81%	COUNTY
4	05/26/2022	Broome	52	50,227	737	1,093,341	7.23%	COUNTY
5	05/26/2022	Capital Region	409	237,848	4,999	4,835,163	8.33%	REGION
6	05/26/2022	Cattaraugus	23	16,522	211	286,108	15.08%	COUNTY

## Cleaning and Tidying Data

```
In [4]: # select appropriate test variables
NYPD.shoot.hist = NYPD.shoot.hist %>% select(c('OCCUR_DATE','OCCUR_TIME','BORO','PRECINCT','VIC_AGE_GROUP','VIC_SEX','VIC_RACE'))
NYPD.shoot.2021 = NYPD.shoot.2021 %>% select(c('OCCUR_DATE','OCCUR_TIME','BORO','PRECINCT','VIC_AGE_GROUP','VIC_SEX','VIC_RACE'))
```

```
In [5]: # rename variables
var.names = c('Date','Time','City','Precinct','Victim_age','Victim_sex','Victim_race')
names(NYPD.shoot.hist) = var.names ; names(NYPD.shoot.2021) = var.names

# reformat "time of crime" column (standard time)
NYPD.shoot.hist$Time = format(strptime(NYPD.shoot.hist$Time, format='%H:%M:%S'), '%r')
#NYPD.shoot.2021$Time = strptime(NYPD.shoot.2021$Time, format='%H:%M:%S')

# merge the two datasets
NYPD.shoot = rbind(NYPD.shoot.hist,NYPD.shoot.2021)

# reformat the dates
NYPD.shoot$date = mdy(NYPD.shoot$date)

# Filter out data before year 2020
NYPD.shoot = NYPD.shoot %>% filter(Date >= as.Date('2020-01-01'))

# Order data by date
NYPD.shoot = NYPD.shoot[order(as.Date(NYPD.shoot$date, format="%m/%d/%Y")),,]
head(NYPD.shoot)
```

A data.frame: 6 × 7

	Date	Time	City	Precinct	Victim_age	Victim_sex	Victim_race
	<date>	<chr>	<chr>	<int>	<chr>	<chr>	<chr>
950	2020-01-01	12:00:00 AM	BROOKLYN	77	18-24	M	BLACK
1266	2020-01-01	12:00:00 AM	MANHATTAN	24	25-44	M	BLACK
1888	2020-01-01	12:00:00 AM	BRONX	43	25-44	M	BLACK
1935	2020-01-01	12:00:00 AM	BRONX	43	25-44	M	BLACK
21	2020-01-02	12:00:00 AM	MANHATTAN	25	25-44	M	BLACK
203	2020-01-02	12:00:00 AM	BROOKLYN	67	25-44	M	BLACK

```
In [6]: # Clean New York Covid-19 Dataset

# Filter out counties
NY.covid19 = NY.covid19 %>% filter(County=="STATEWIDE") %>% select(c('Test.Date','New.Positives'))

# Filter out dates in 2022
NY.covid19$Test.Date = mdy(NY.covid19$Test.Date)
NY.covid19 = NY.covid19 %>% filter(Test.Date <as.Date('2022-01-01'))
names(NY.covid19) = c('Date','New.Positives')
head(NY.covid19)
```

A data.frame: 6 × 2

	Date	New_Positives
	<date>	<chr>
1	2021-12-31	85,476
2	2021-12-30	76,555
3	2021-12-29	74,207
4	2021-12-28	67,090
5	2021-12-27	40,780
6	2021-12-26	26,737

## Transform Data

```
In [7]: # number of shootings per each day
NYPD.shoot.day = NYPD.shoot %>% count(Date)
names(NYPD.shoot.day) = c('Date','Number_Shootings')
head(NYPD.shoot.day)
```

A data.frame: 6 × 2

	Date	Number_Shootings
	<date>	<int>
1	2020-01-01	4
2	2020-01-02	8
3	2020-01-03	1
4	2020-01-04	3
5	2020-01-05	3
6	2020-01-06	1

```
In [8]: # merge Covid-19 data and Shooting Data in NY
NY.cases = left_join(NYPD.shoot.day,NY.covid19)
NY.cases$New_Positives[is.na(NY.cases$New_Positives)] = 0

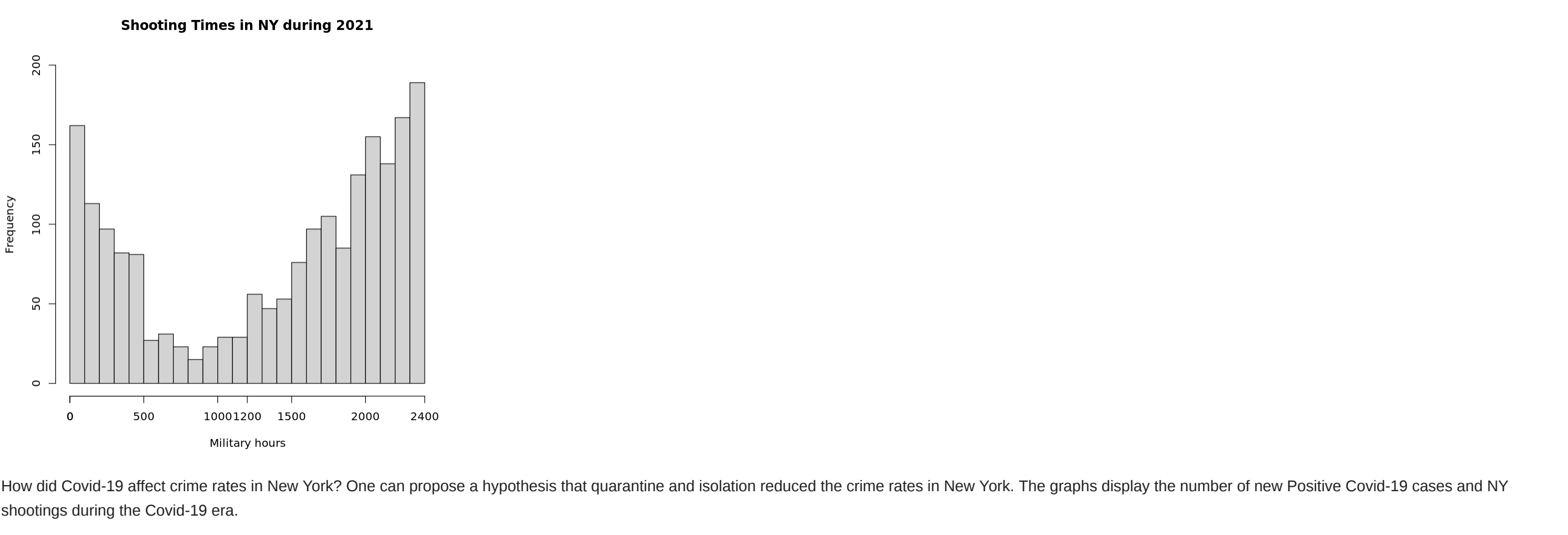
Joining, by = "Date"
```

## Visualization

Lets visualize the number of shootings that occur during different hours of the day. \ During what time of the day does crime occur the most?

```
In [9]: # Distribution of Shooting times
# convert to military hours
mil.time = gsub(':', '', NYPD.shoot.2021$Time) %>% substr(1,4)
NYPD.shoot.milhour = as.numeric(gsub(':', '', mil.time))

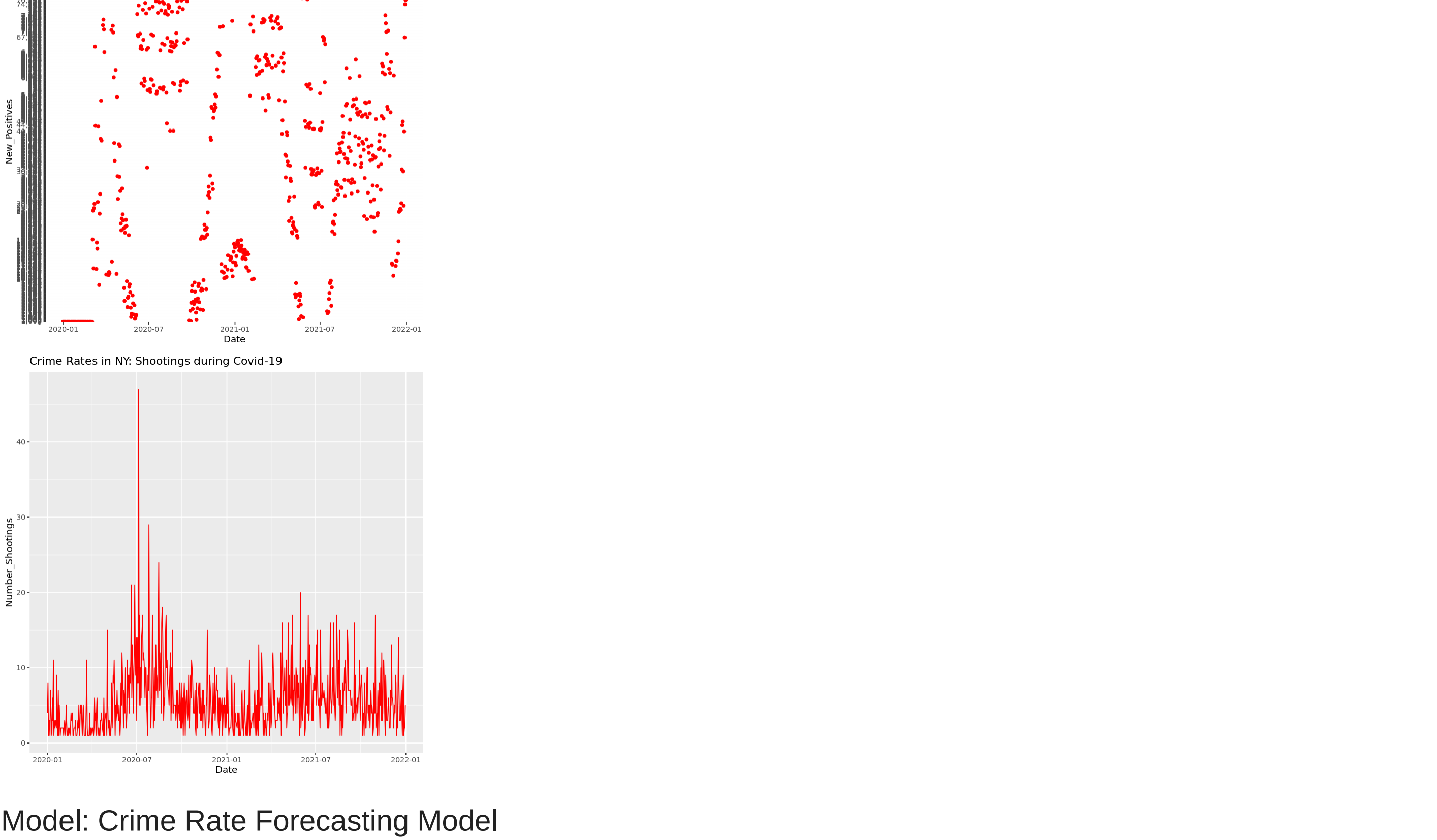
# The NYPD can instate more police officers and units during the time frame when the most crime incidents occur.
hist(NYPD.shoot.milhour,breaks = 24,main = "Shooting Times in NY during 2021", xlim = c(0,2400),ylim=c(0,200), xlab = 'Military hours',)
axis(1, at = seq(0, 2400, 1200))
```



How did Covid-19 affect crime rates in New York? One can propose a hypothesis that quarantine and isolation reduced the crime rates in New York. The graphs display the number of new Positive Covid-19 cases and NY shootings during the Covid-19 era.

```
In [10]: # Visualize the crime rate in NY during the Covid-19 pandemic
ggplot(data=NY.cases,aes(x=Date)) +
  geom_point( aes(y=New_Positives), color='red') +
  ggtitle("Covid-19 Cases in NY")

ggplot(data=NY.cases,aes(x=Date)) +
  geom_line( aes(y=Number_Shootings), color='red') +
  ggtitle("Crime Rates in NY: Shootings during Covid-19")
```



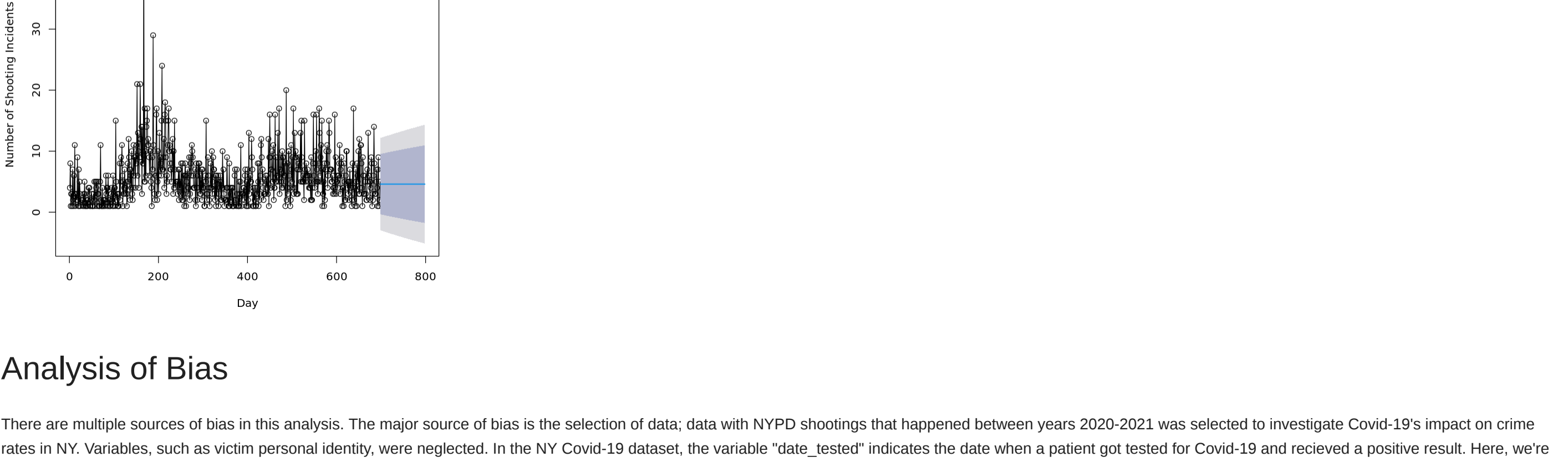
## Model: Crime Rate Forecasting Model

The number of shootings in the future can be predicted using time-series forecasting models. A SARIMA model can be used to predict the shooting frequency over a 100 day time span after 2021.

```
In [17]: # Forecasting
install.packages("forecast")
library(forecast)

Installing package into ‘/srv/rlibs’
(as ‘lib’ is unspecified)
```

```
In [15]: model = auto.arima(NYPD.shoot.day$Number_Shootings,D=1)
forecast = forecast(model,h=100)
plot(forecast,xlab='Day',ylab='Number of Shooting Incidents',main="Shooting Forecasts in New York")
```



## Analysis of Bias

There are multiple sources of bias in this analysis. The major source of bias is the selection of data; data with NYPD shootings that happened between years 2020-2021 was selected to investigate Covid-19's impact on crime rates in NY. Variables, such as victim personal identity, were neglected. In the NY Covid-19 dataset, the variable "date\_tested" indicates the date when a patient got tested for Covid-19 and recieved a positive result. Here, we're assuming that the patient first recieved Covid-19 on their test date, but the patient's true date of being infected with Covid-19 is likely 1-3 days before their test date.

## Conclusions

It appears that the most shootings in New York occur between 11 p.m. and midnight. According to the graphs, New York's crime rates started rising during May 2020. During the summer months, Covid-19 actually increased the crime rates and number of shootings. Hence, most shootings occurred during the summer of 2020.