Created by

Asif Newaz

This notebook provides analysis on the fatal shooting in the USA by the police.

The data was curated by the Washington Post. It is available here - https://github.com/washingtonpost/data-police-shootings.

Importing Data

```
head(data)
```

ans = 8×17 table

. . .

|	id	name	date	manner_of_death	armed	age
1	3	"Tim Elliot"	2015-01-02	shot	gun	53
2	4	"Lewis Lee Lembke"	2015-01-02	shot	gun	47
3	5	"John Paul Quintero"	2015-01-03	shot and Tasered	unarmed	23
4	8	"Matthew Hoffman"	2015-01-04	shot	toy weapon	32
5	9	"Michael Rodriguez"	2015-01-04	shot	nail gun	39
6	11	"Kenneth Joe Brown"	2015-01-04	shot	gun	18
7	13	"Kenneth Arnold Buck"	2015-01-05	shot	gun	22
8	15	"Brock Nichols"	2015-01-06	shot	gun	35

The data contains a total of 8002 records and 17 attributes (columns). We will analyze the records gradually.

Primary Analysis

Locations of police shootings

The following figure displays the overall distribution across the USA from 2015-2022. Later, we will see yearly distribution and how it changed over the years.

```
geoscatter(data.latitude,data.longitude,'.')
geobasemap topographic
```
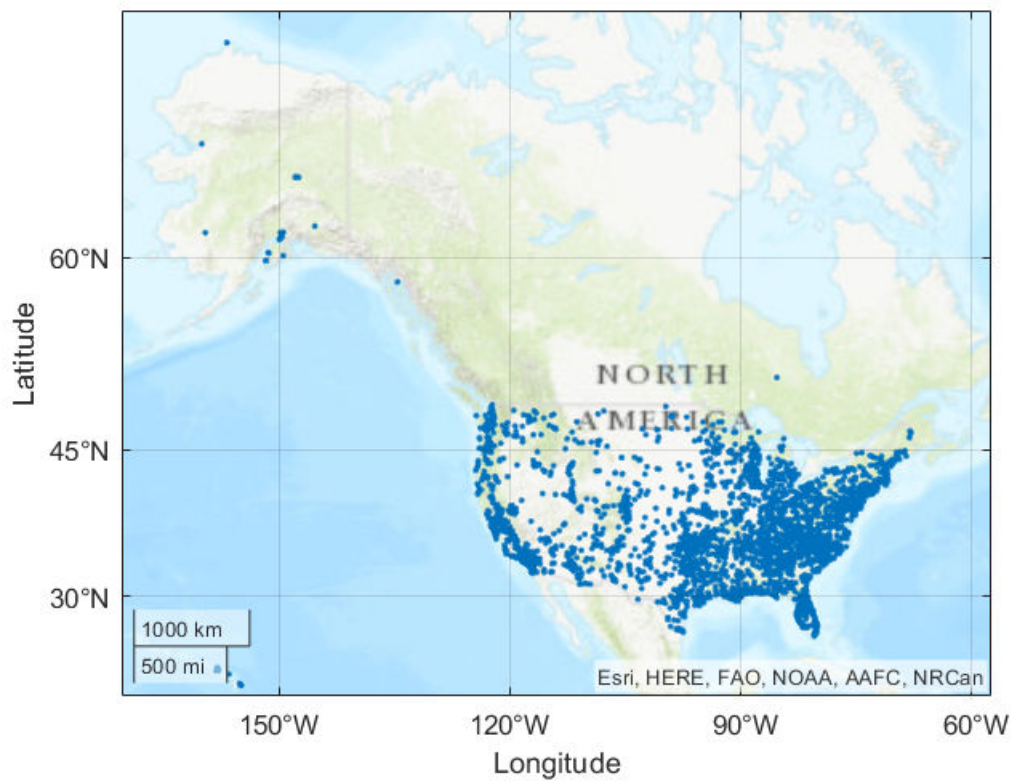
Manner of death

```
summary(data.manner_of_death)
```

```
    shot                    7664
    shot and Tasered         338
```

Circumstances of the shooting

```
summary(data.armed)
```

```
    Airsoft pistol              6
    BB gun                     19
    BB gun and vehicle          1
    Gun                         5
    Taser                      39
    air conditioner             1
    air pistol                  2
    ax                         30
    ax and machete              1
    barstool                    1
    baseball bat               20
    baseball bat and bottle     1
    baseball bat and firepl…    1
    baseball bat and knife      1
    baton                       6
    bean-bag gun                1
    beer bottle                 3
    binoculars                  1
```

blunt object	6
bottle	3
bow and arrow	2
box cutter	14
brick	2
car, knife and mace	1
carjack	1
chain	3
chain saw	2
chainsaw	1
chair	4
claimed to be armed	1
contractor's level	1
cordless drill	1
crossbow	10
crowbar	5
fireworks	1
flagpole	1
flare gun	1
flashlight	2
garden tool	4
glass shard	5
gun	4573
gun and car	15
gun and knife	27
gun and machete	3
gun and sword	1
gun and vehicle	27
guns and explosives	3
hammer	22
hammer and garden tool	1
hand torch	1
hatchet	16
hatchet and gun	2
ice pick	1
incendiary device	2
incendiary weapon	1
knife	1142
knife and vehicle	2
knife, hammer and gasol…	1
lawn mower blade	2
machete	65
machete and gun	1
machete and hammer	1
meat cleaver	6
metal hand tool	4
metal object	7
metal pipe	18
metal pole	5
metal rake	1
metal stick	3
microphone	1
motorcycle	1
nail gun	2
oar	1
pair of scissors	14
pellet gun	4
pen	1
pepper spray	2
pick-axe	4
piece of wood	9
pipe	8
pitchfork	2
pole	3

```
pole and knife                 2
railroad spikes                1
rock                           8
samurai sword                  5
screwdriver                   18
sharp object                  25
shovel                         8
spear                          3
stake                          1
stapler                        1
stone                          1
straight edge razor            5
sword                         27
tire iron                      6
toy weapon                   248
unarmed                      460
undetermined                 338
unknown weapon               121
vehicle                      284
vehicle and gun               10
vehicle and machete            1
walking stick                  1
wasp spray                     1
wrench                         1
<undefined>                  211
```

```
histogram(data.armed)
```
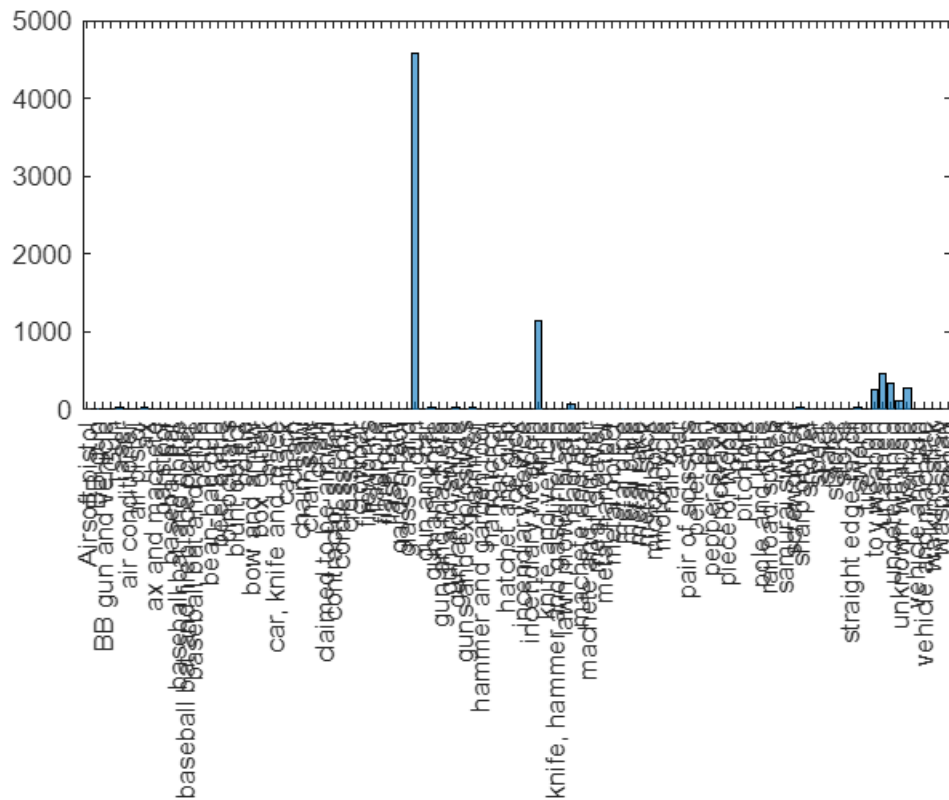
As we can see, in most cases, they were armed or had a knife. However, there were 460 cases where the person was unarmed but still shot to death. We want to investigate those specific cases further.

```
cir= groupcounts(data,"armed")
```

cir = 107×3 table

	armed	GroupCount	Percent
1	Airsoft pistol	6	0.0750
2	BB gun	19	0.2374
3	BB gun and vehicle	1	0.0125
4	Gun	5	0.0625
5	Taser	39	0.4874
6	air conditioner	1	0.0125
7	air pistol	2	0.0250
8	ax	30	0.3749
9	ax and machete	1	0.0125
10	barstool	1	0.0125
11	baseball bat	20	0.2499
12	baseball bat and ...	1	0.0125
13	baseball bat and ...	1	0.0125
14	baseball bat and ...	1	0.0125

$\vdots$

```
sortrows(cir,"GroupCount","descend")
```

ans = 107×3 table

	armed	GroupCount	Percent
1	gun	4573	57.1482
2	knife	1142	14.2714
3	unarmed	460	5.7486
4	undetermined	338	4.2239
5	vehicle	284	3.5491
6	toy weapon	248	3.0992
7	<undefined>	211	2.6368
8	unknown weapon	121	1.5121
9	machete	65	0.8123
10	Taser	39	0.4874

5

	armed	GroupCount	Percent
11	ax	30	0.3749
12	gun and knife	27	0.3374
13	gun and vehicle	27	0.3374
14	sword	27	0.3374

:
:

Age distribution

```
histogram(data.age, 90)
```
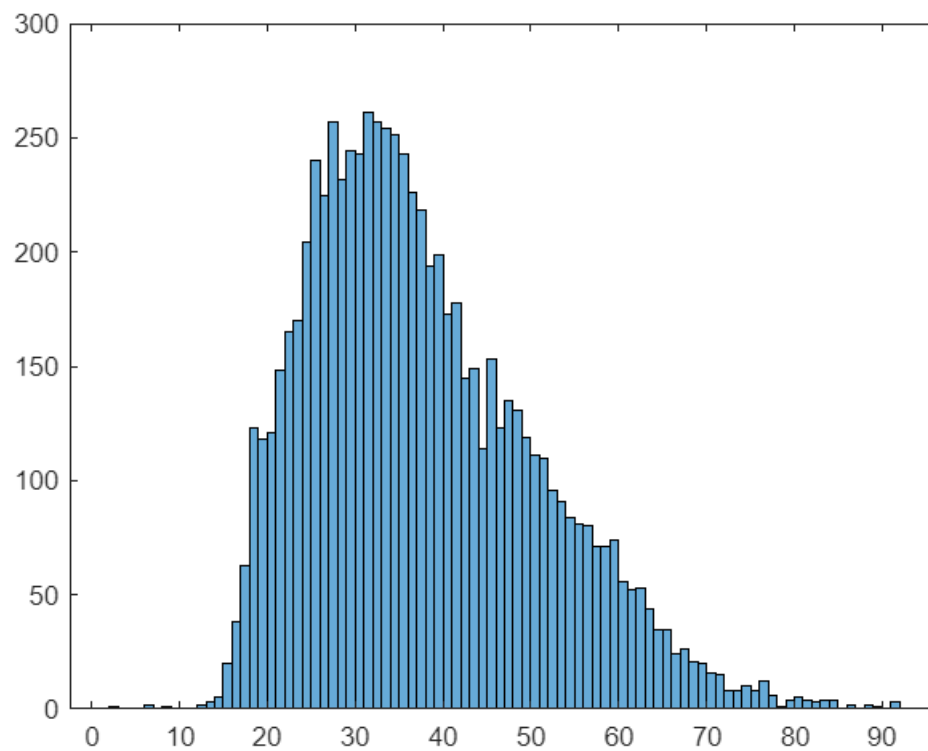

Surprisingly, there were several shootings of underage children as well as some very elderly people. Need further investigation in these particular cases.
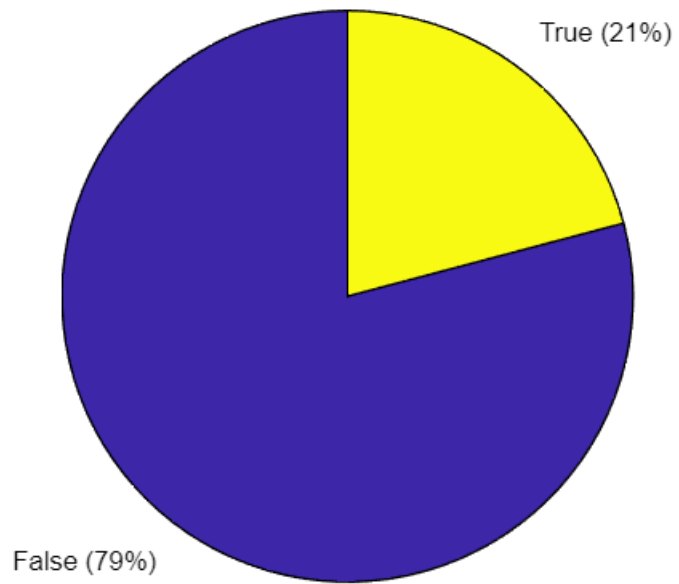
Gender Distribution

```
pie(data.gender)
```

F (4%)

M (96%)
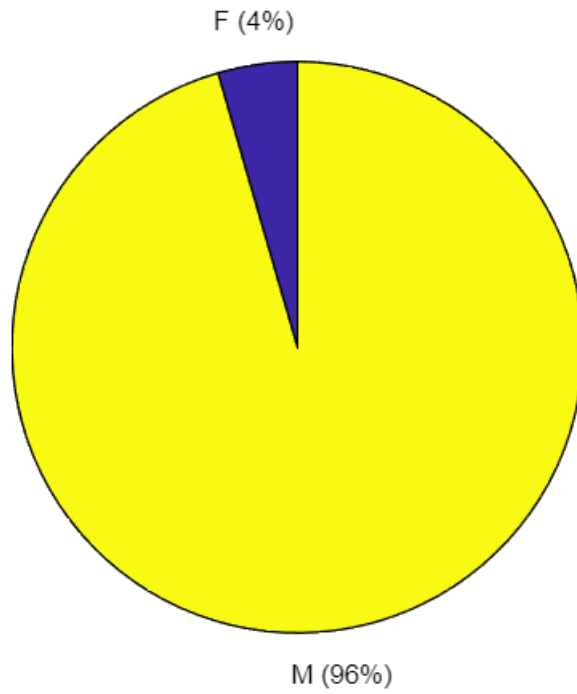
Racial Distribution

```
histogram(data.race)
xticklabels({'Asian', 'Black', 'Hispanic', 'Native American', 'Other', 'White'})
```
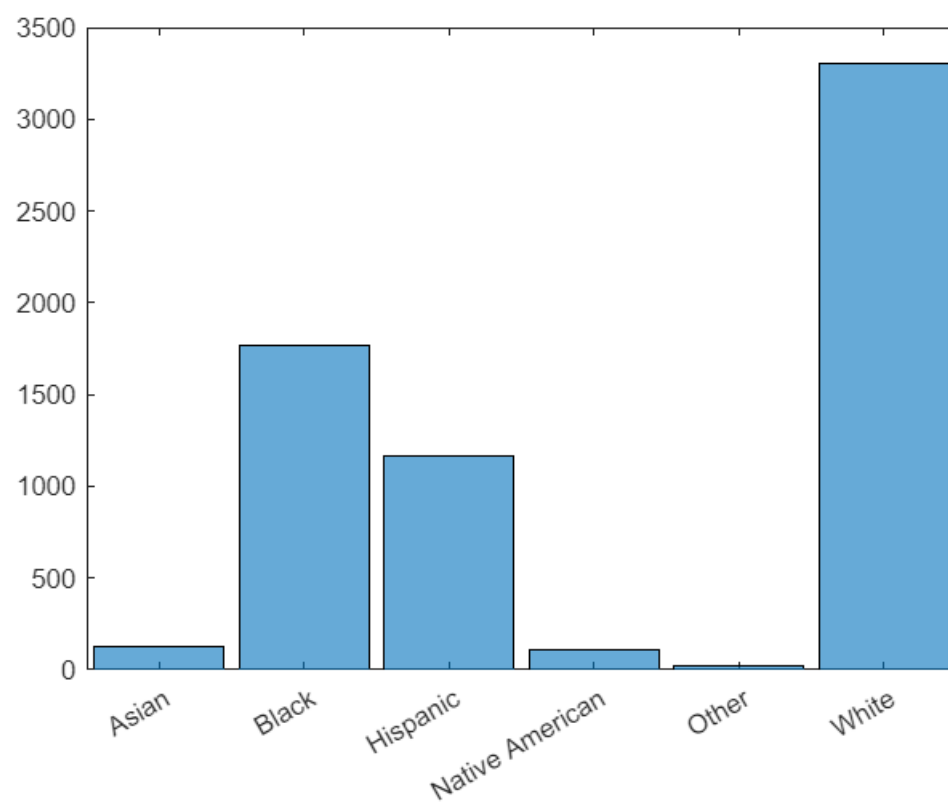
Mental illness

```
pie(data.signs_of_mental_illness)
```

True (21%)

False (79%)

Its surprising to see around 21% of the people shot dead had displayed mental illness (The attribute indicates - whether news reports have indicated the victim had a history of mental health issues, expressed suicidal intentions or was experiencing mental distress at the time of the shooting).

Threat Level

```
pie(data.threat_level)
```

undetermined (4%)

other (33%)

attack (63%)
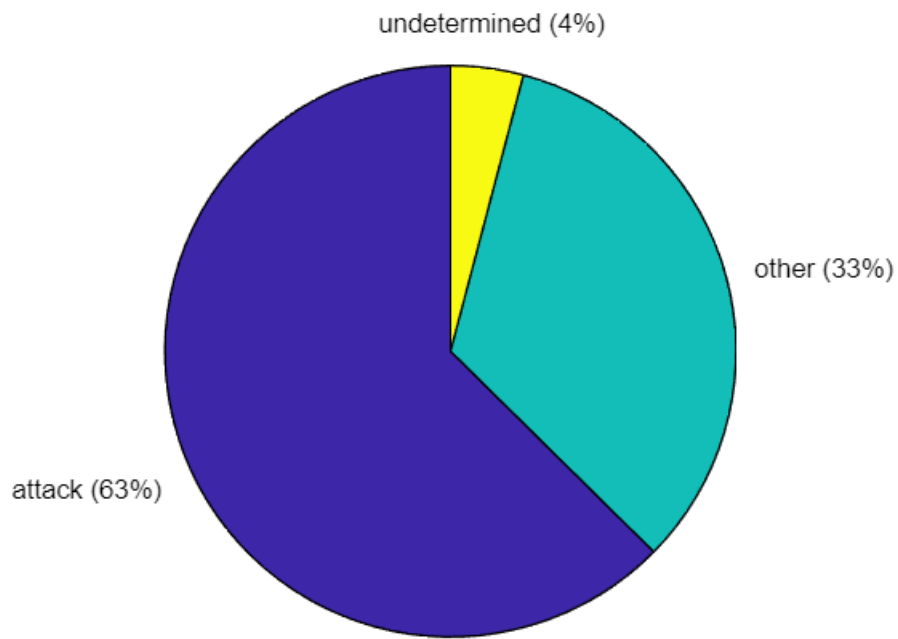
This attribute has been updated in the version-2 of the dataset with more information.

Was the person fleeing?

```
pie(data.flee)
```
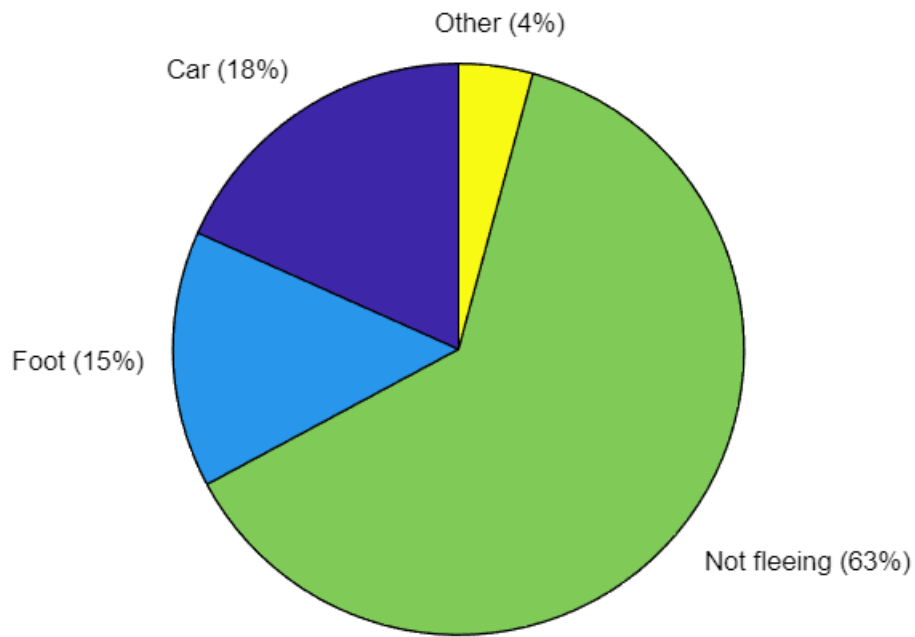
In 63% of the cases, the person was not fleeing. Need to analyze relation with threat level and circumstances of the shooting.

Body camera (was the incident recorded?)

```
pie(data.body_camera)
```

True (14%)

False (86%)
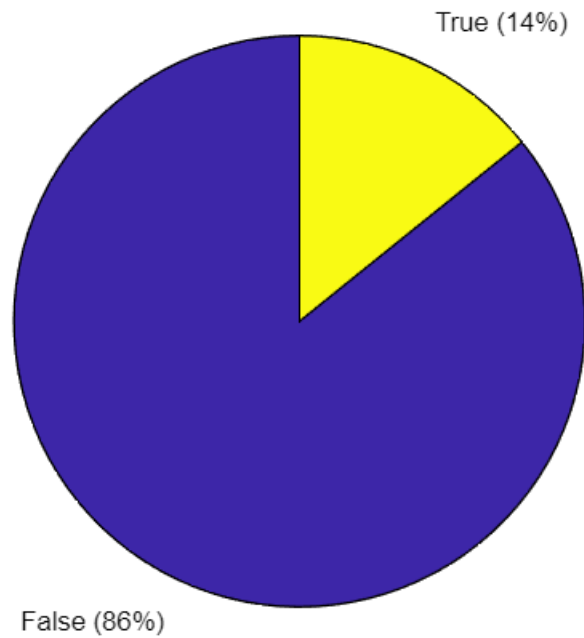
In most cases, the shootings were not recorded in the body camera of the police.

Yearly distribution

The date column is a datetime variable with the exact date. For the year information only, a new column is created.

```
data.year= year(data.date);
tail(data)
```

ans = 8×18 table

. . .

	id	name	date	manner_of_death	armed	age
1	8689	""	2022-11-29	shot	gun	NaN
2	8691	""	2022-11-29	shot	gun	NaN
3	8693	"Reymundo Ricardo Flores"	2022-11-29	shot	gun	33
4	8690	""	2022-11-30	shot	vehicle	NaN
5	8692	""	2022-11-30	shot	knife	NaN
6	8694	""	2022-12-01	shot	gun	38
7	8695	""	2022-12-01	shot	gun	NaN
8	8696	""	2022-12-01	shot	knife	NaN

```
histogram(data.year)
```
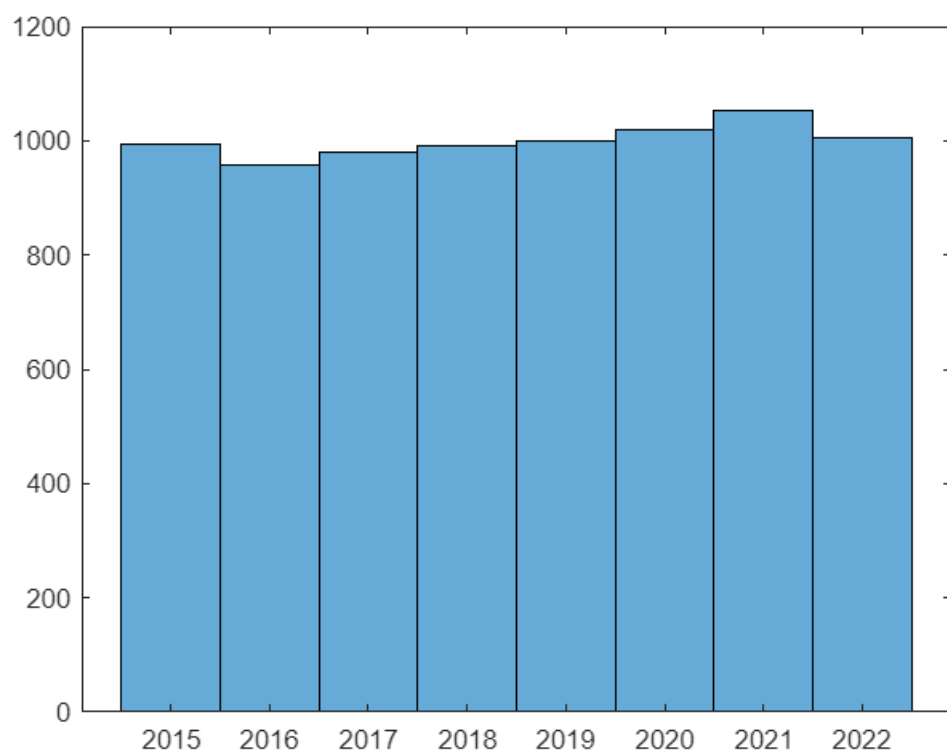

State - wise

```
histogram(data.state)
```
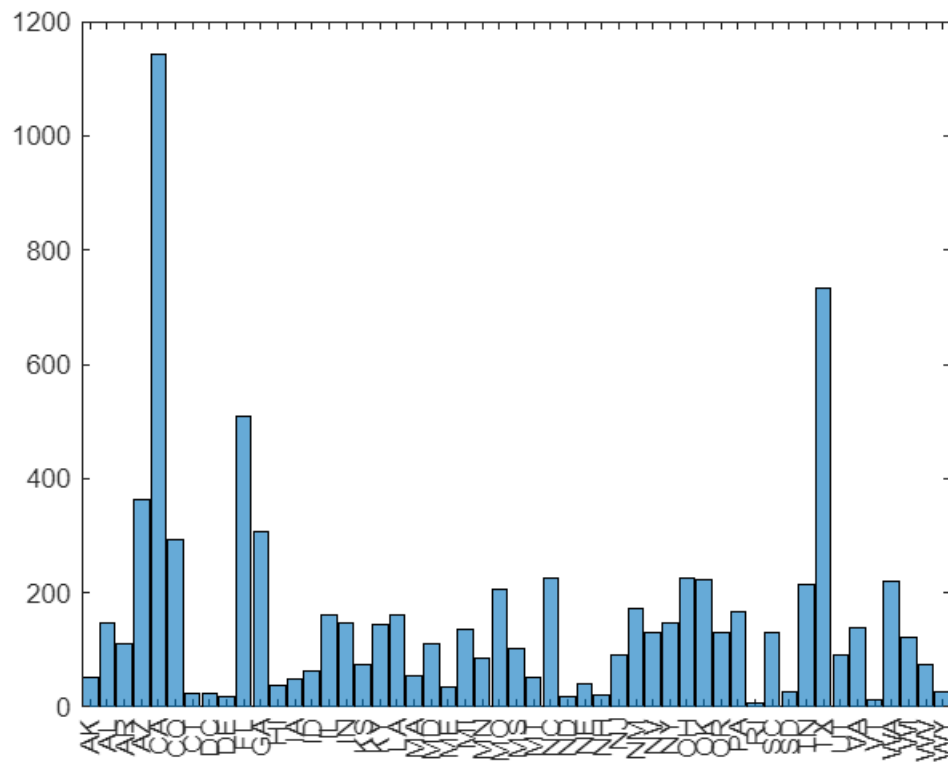
California, Texas and Florida has the highest number of shootings.

```
st=groupcounts(data,"state");
sortrows(st,'GroupCount')
```

ans = 51×3 table

	state	GroupCount	Percent
1	RI	6	0.0750
2	VT	13	0.1625
3	DE	17	0.2124
4	ND	18	0.2249
5	NH	22	0.2749
6	CT	23	0.2874
7	DC	24	0.2999
8	SD	28	0.3499
9	WY	28	0.3499
10	ME	34	0.4249
11	HI	37	0.4624
12	NE	40	0.4999

	state	GroupCount	Percent
13	IA	49	0.6123
14	MT	51	0.6373

⋮
⋮

Whereas, Vermont, Delaware, North Dakota had the least amount of such shootings.

In-depth Analysis

Shootings over the years

```matlab
% excluding alaska for now

% Define the latitude and longitude boundaries for the continental USA (excluding Alaska)
lat_min = 24.396308;    % Southernmost point (Florida)
lat_max = 49.384358;    % Northernmost point (Minnesota)
lon_min = -125.0;       % Westernmost point (California)
lon_max = -66.93457;    % Easternmost point (Maine)

% Apply the filter for continental USA, excluding Alaska
in_usa = (data.latitude >= lat_min) & (data.latitude <= lat_max) & (data.longitude >= lon_min)

% Extract the valid locations
df= data(in_usa,:);
```

```matlab
    % Initialize GIF
    filename = 'police_killings.gif';
    years= 2015:2022;

    for i = 1:length(years)
        currentYear = years(i);
        subset = df(df.year == currentYear, :);

        figure;
        geoscatter(subset.latitude, subset.longitude, '.');
        geobasemap topographic
        title(sprintf('Police Killings in %d', currentYear));

        % Capture the plot as an image
        frame = getframe(gcf);
        im = frame2im(frame);
        [imind, cm] = rgb2ind(im, 256);

        % Write to the GIF file
        if i == 1
```
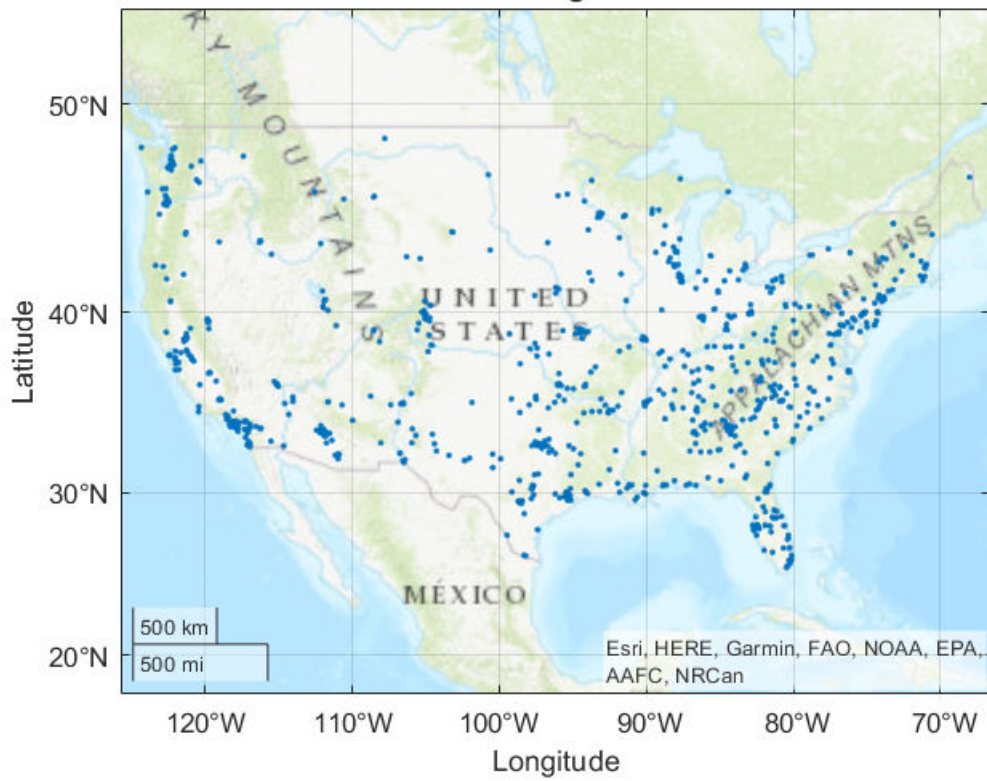
```
        imwrite(imind, cm, filename, 'gif', 'Loopcount', inf, 'DelayTime', 1);
    else
        imwrite(imind, cm, filename, 'gif', 'WriteMode', 'append', 'DelayTime', 1);
    end

    pause(1);
end
```
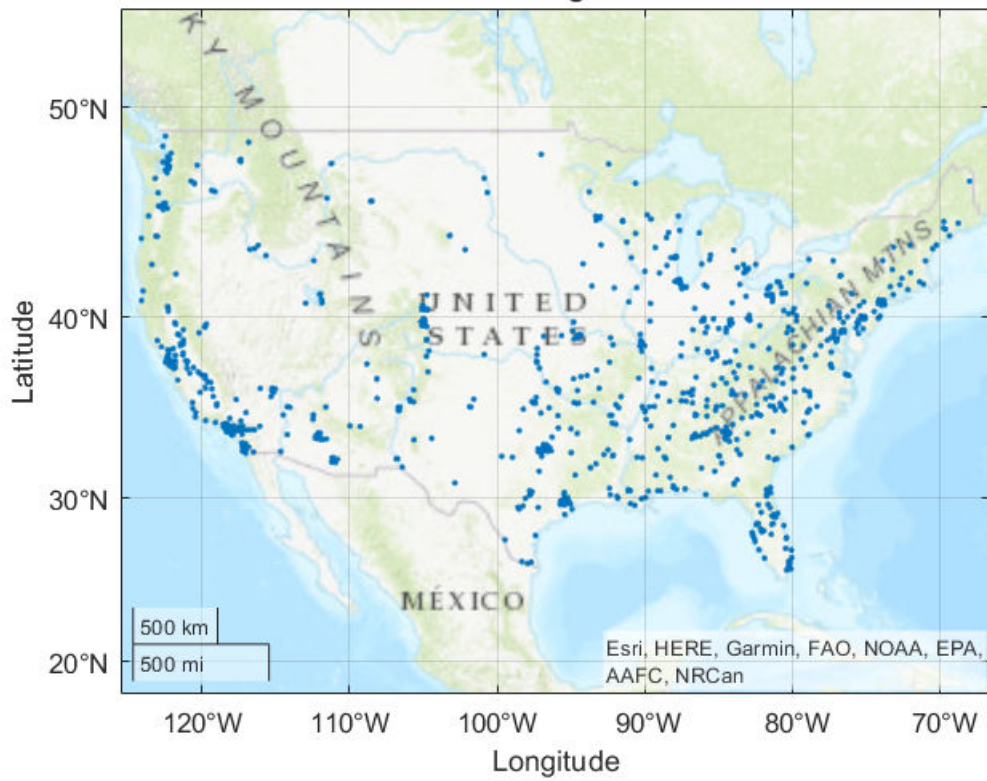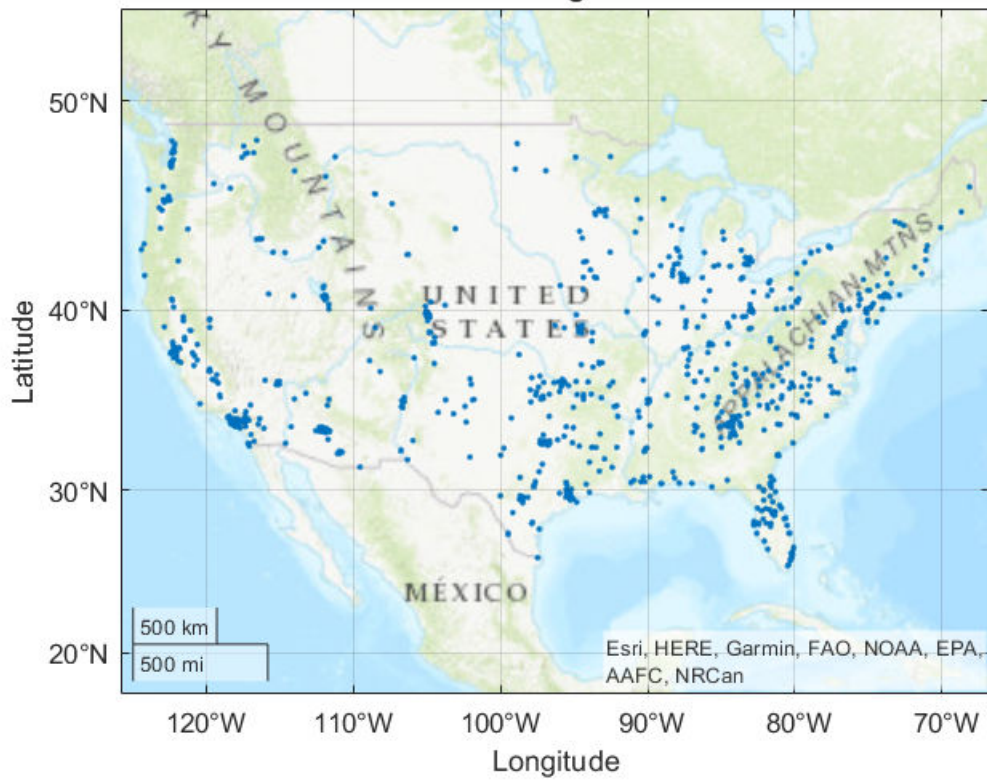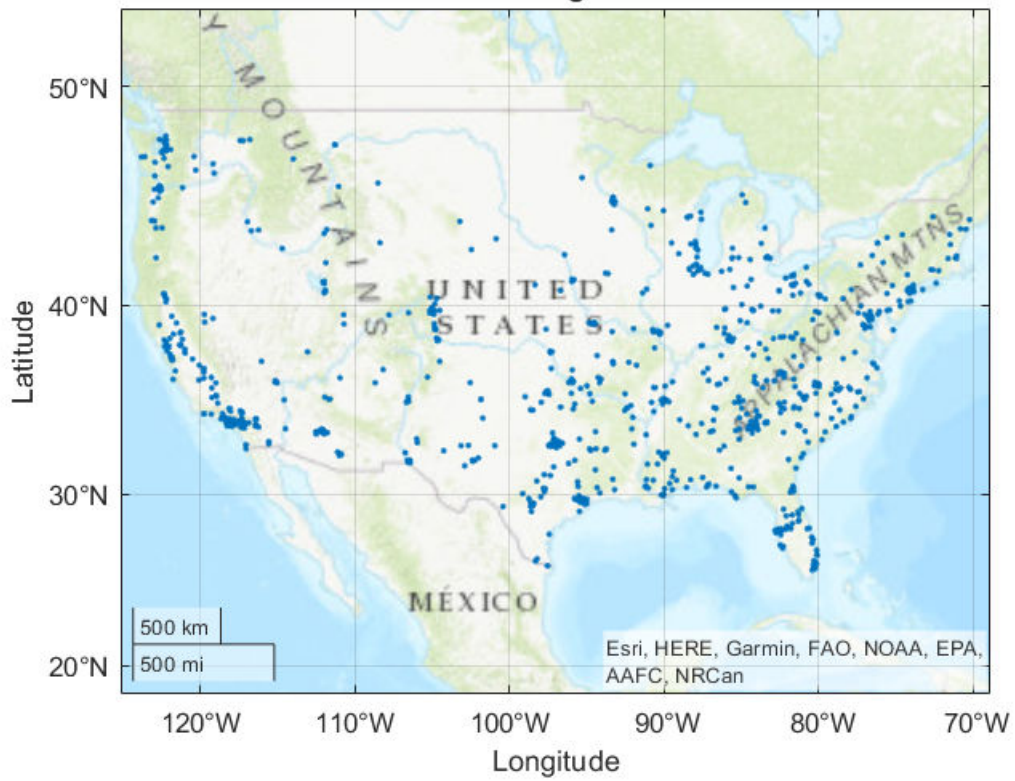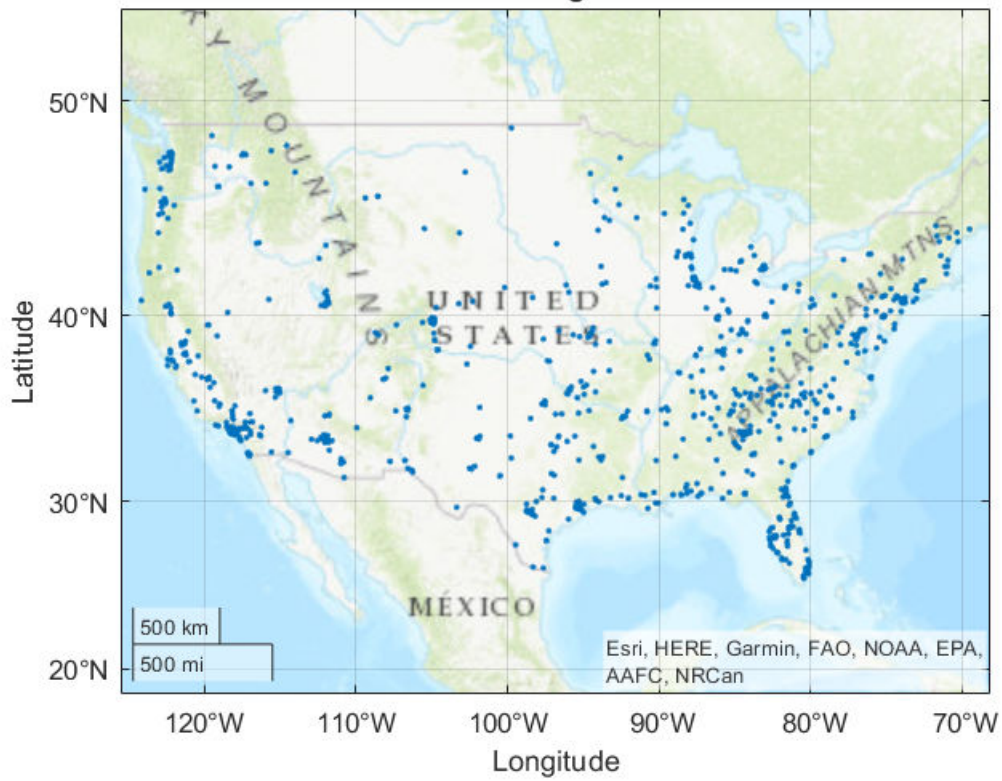
Police Killings in 2015
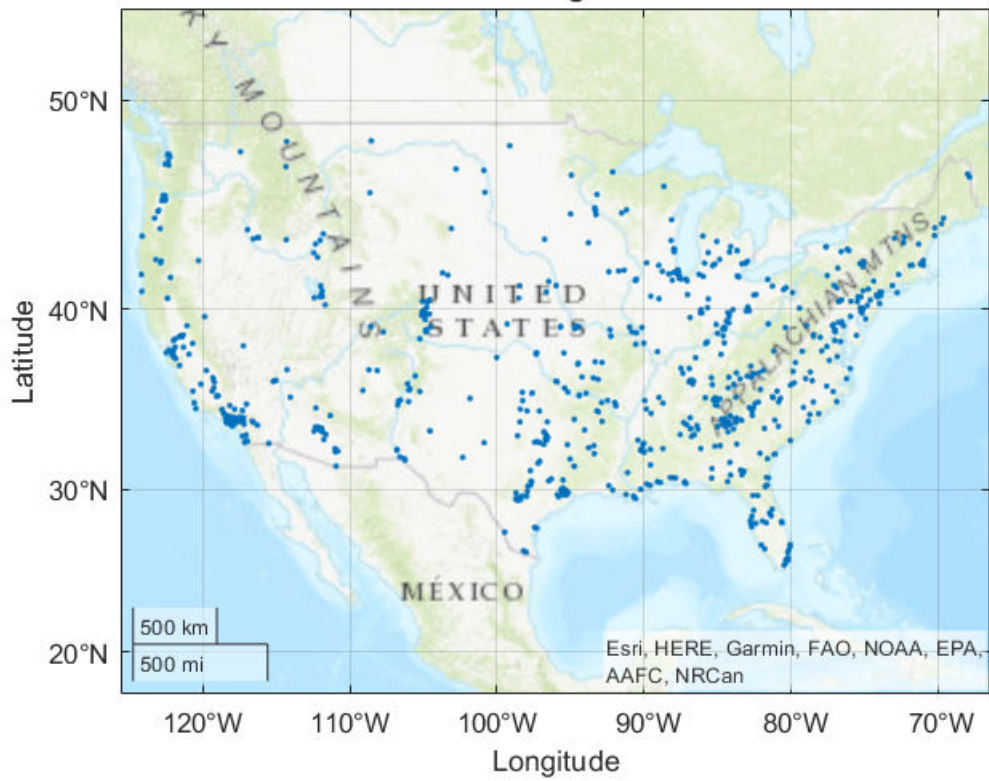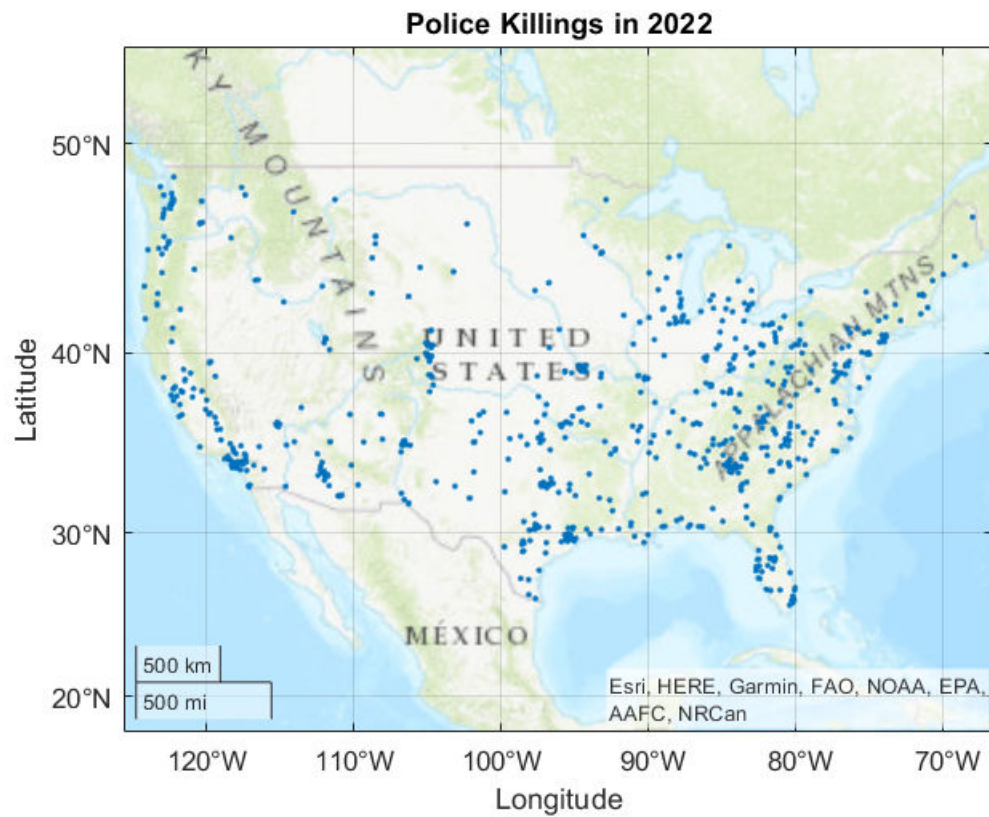
Police Killings in 2016

Police Killings in 2017

17

Police Killings in 2018

Police Killings in 2019

Police Killings in 2020

Police Killings in 2021

Police Killings in 2022
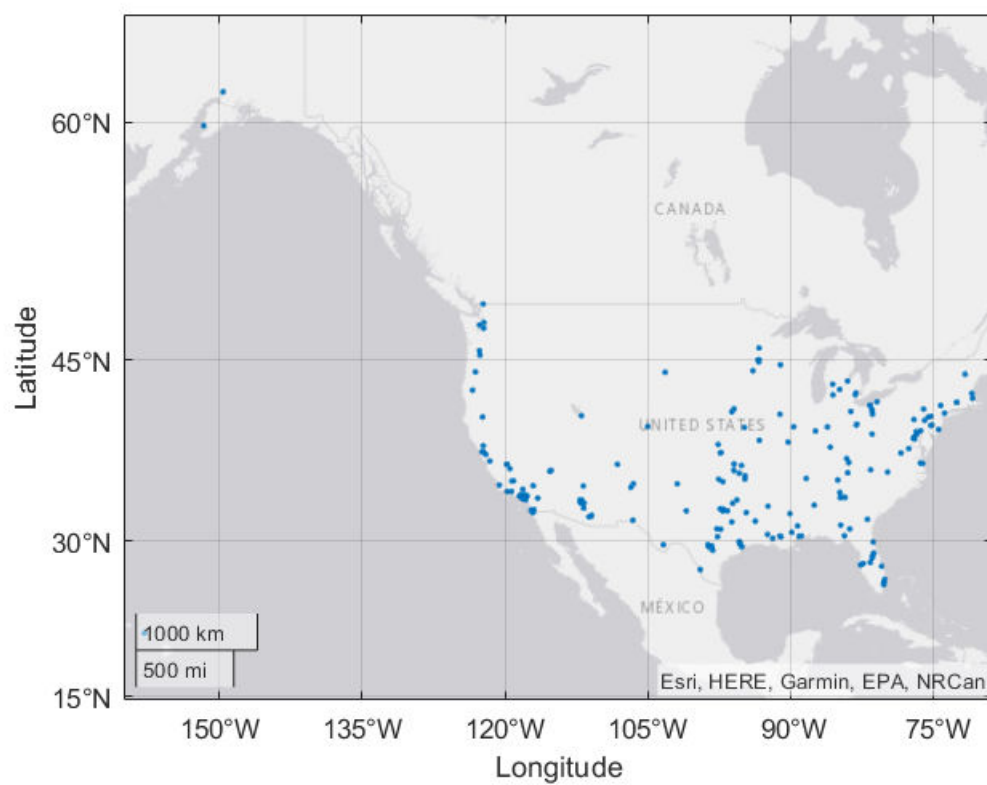
Investigation into unarmed shooting cases

```
una= data(data.armed == "unarmed",:);
```
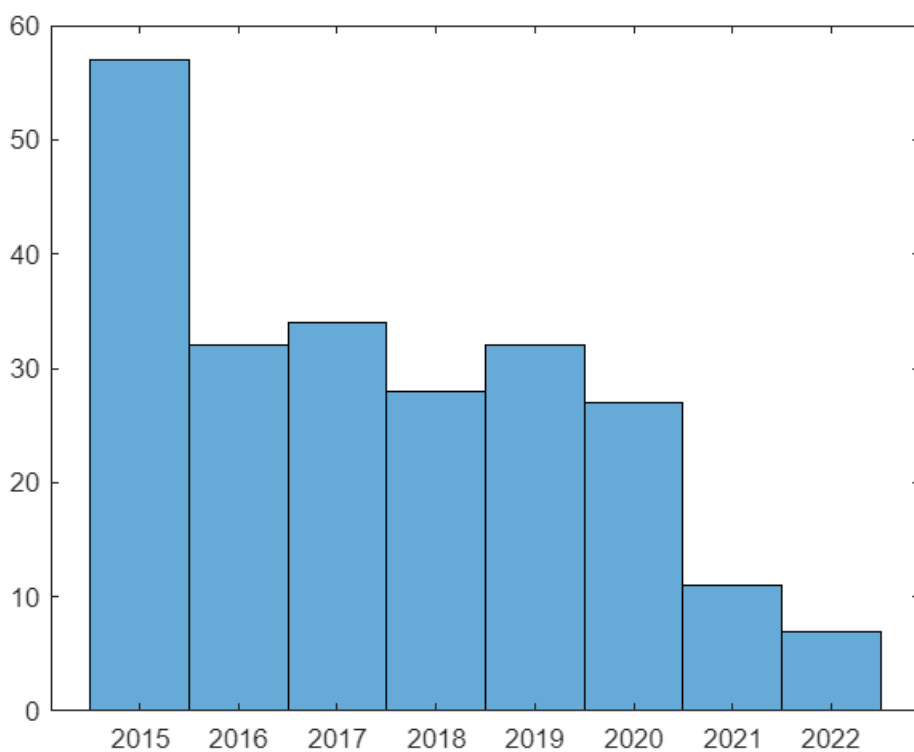
There are 460 such events.

```
% Unarmed and not fleeing
una_nf = una(una.flee == "Not fleeing",:);
```

There are 228 such cases.

```
geoscatter(una_nf.latitude,una_nf.longitude,'.')
```

```
histogram(una_nf.year)
```

This is quite a positive sign that such cases has reduced over the years. Lets see where such incidents are taking place.

```matlab
x= df(df.armed == "unarmed",:);
xx = x(x.flee == "Not fleeing",:);


% Initialize GIF
    filename = 'police_killings_unarmed_nf.gif';
    years= 2015:2022;

    for i = 1:length(years)
        currentYear = years(i);
        subset = xx(xx.year == currentYear, :);

        figure;
        geoscatter(subset.latitude, subset.longitude,  '.');
        geobasemap topographic;
        title(sprintf('Police Killings in %d', currentYear));

        % Capture the plot as an image
        frame = getframe(gcf);
        im = frame2im(frame);
        [imind, cm] = rgb2ind(im, 256);

        % Write to the GIF file
        if i == 1
            imwrite(imind, cm, filename, 'gif', 'Loopcount', inf, 'DelayTime', 1);
        else
            imwrite(imind, cm, filename, 'gif', 'WriteMode', 'append', 'DelayTime', 1);
        end

        pause(1);
    end
```
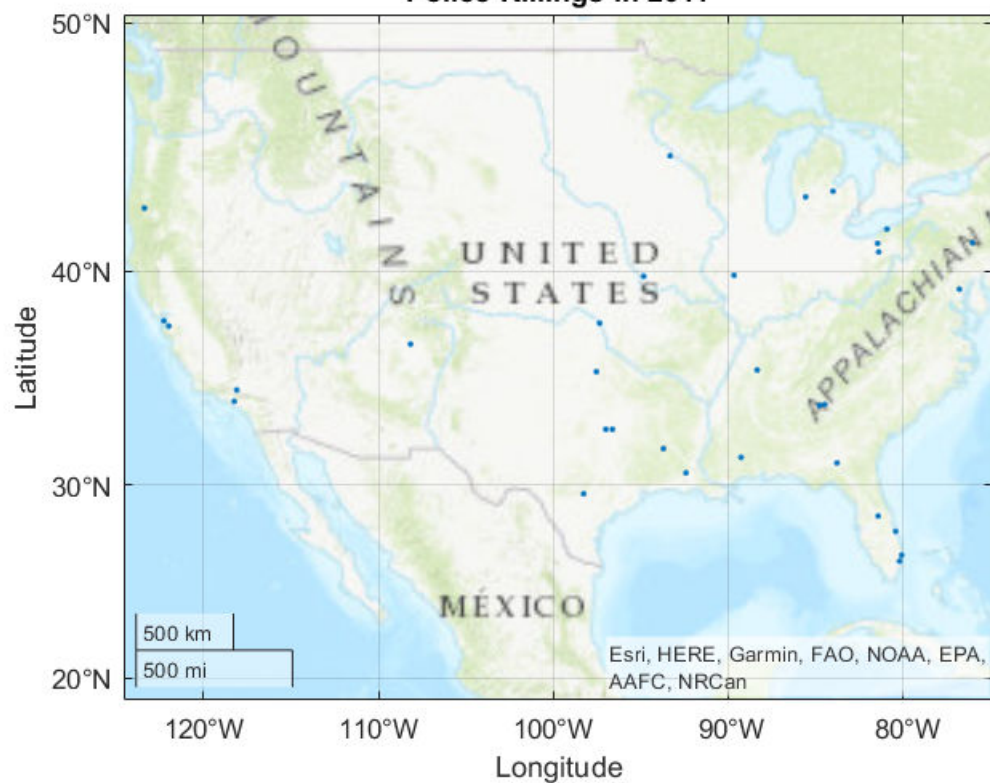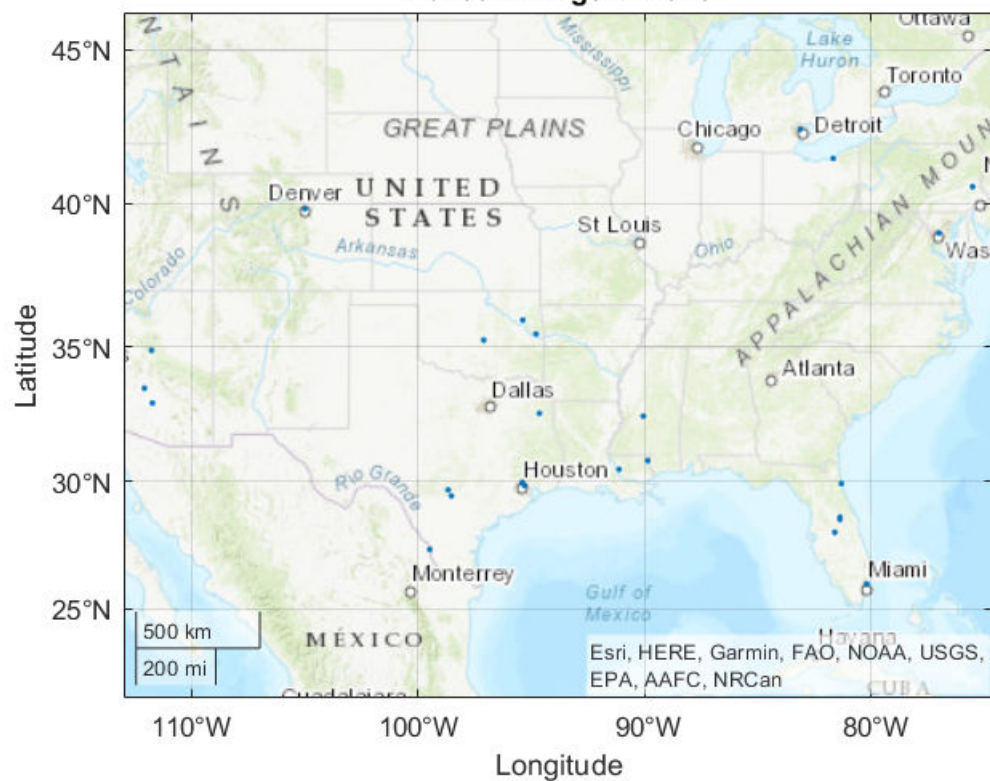
Police Killings in 2015
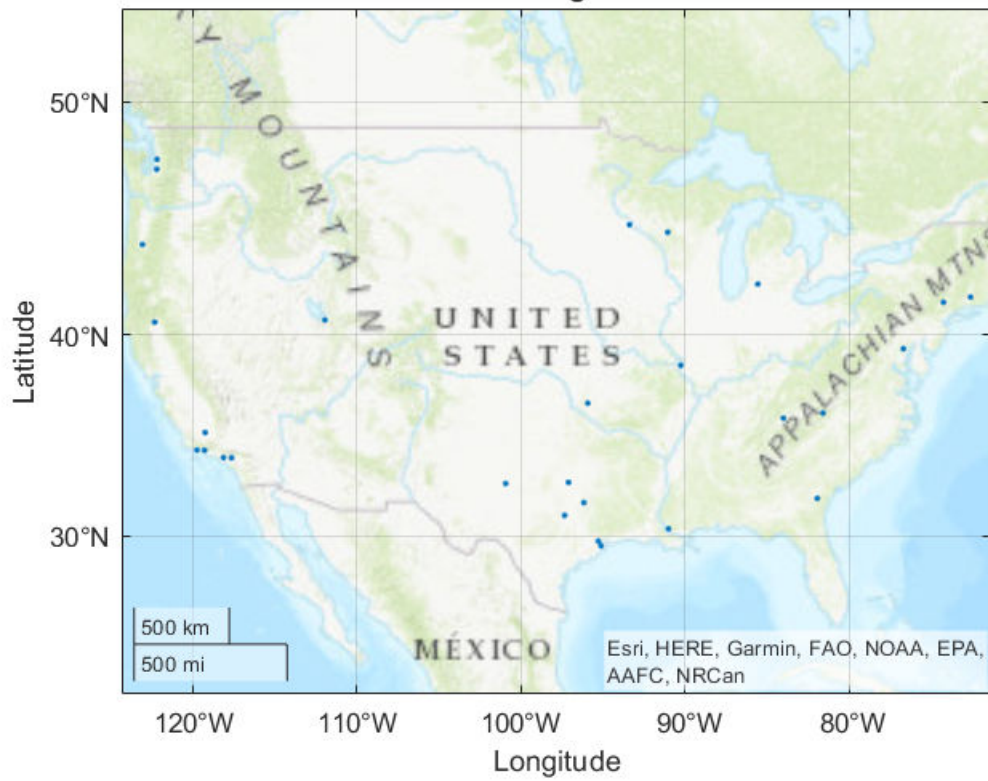

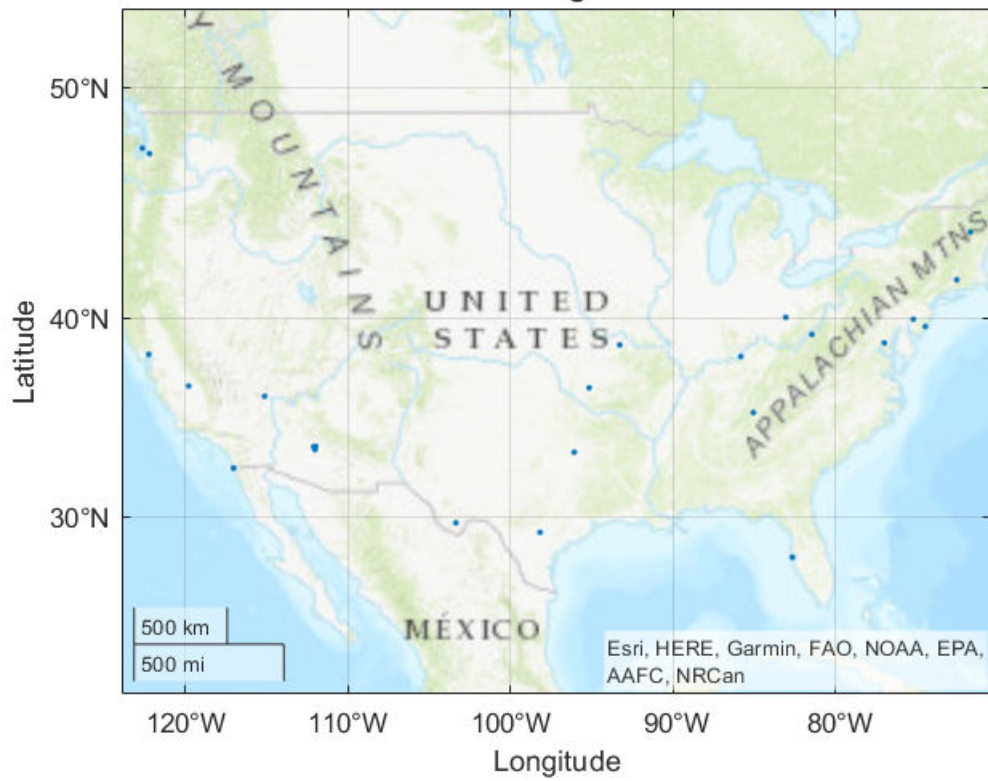## Police Killings in 2016
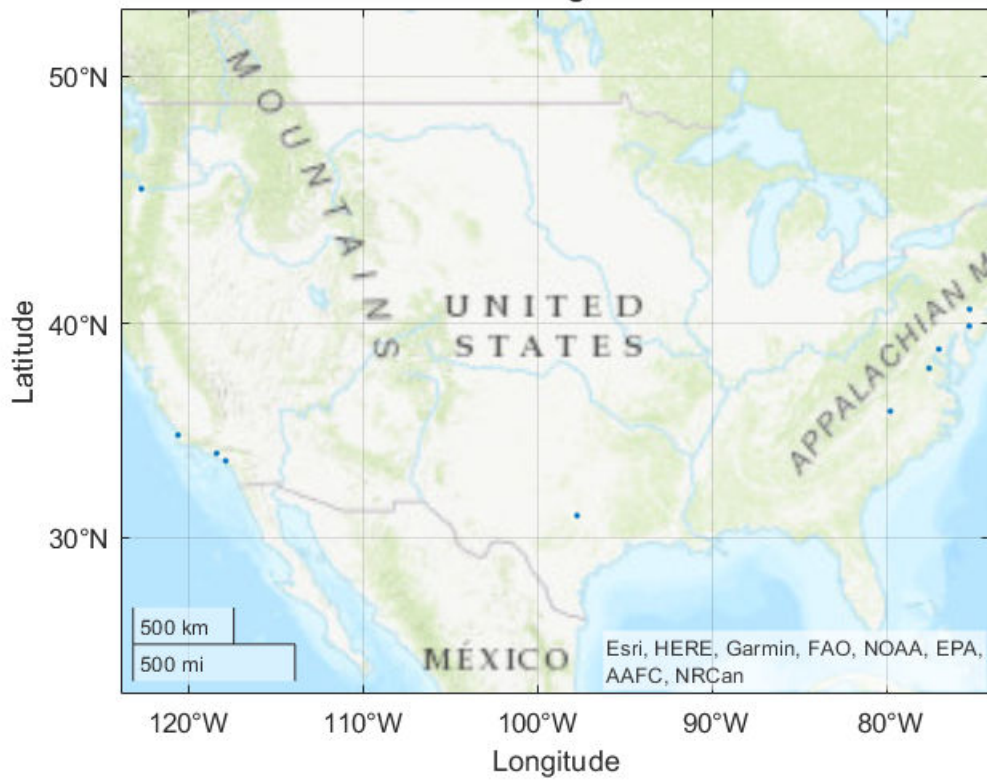
Police Killings in 2017

Police Killings in 2018

Police Killings in 2019

Police Killings in 2020

Police Killings in 2021

Police Killings in 2022

Investigation into Underage shooting

```
ug= data(data.age<18,:);
```

There are 135 such cases.

```
% underage but had a gun or weapon
ugw= groupcounts(ug, "armed");
sortrows(ugw,"GroupCount", 'descend')
```

ans = 14×3 table

	armed	GroupCount	Percent
1	gun	58	42.9630
2	unarmed	18	13.3333
3	knife	16	11.8519
4	toy weapon	13	9.6296
5	undetermined	11	8.1481
6	vehicle	7	5.1852
7	<undefined>	5	3.7037
8	BB gun	1	0.7407
9	crowbar	1	0.7407
10	gun and knife	1	0.7407
11	gun and vehi...	1	0.7407
12	machete	1	0.7407
13	pair of scis...	1	0.7407
14	unknown weapon	1	0.7407

Although, they were underage but in most cases, they were armed with gun or a knife. There are 18 unarmed cases.

```
% lowest age
head(sortrows(ug,"age"),15)
```

ans = 15×18 table

...

	id	name	date	manner_of_death	armed	age
1	7981	"Clesslynn Jane Crawford"	2022-03-26	shot	unarmed	2
2	980	"Jeremy Mardis"	2015-11-03	shot	unarmed	6
3	3229	"Kameron Prescott"	2017-12-21	shot	unarmed	6
4	7322	"Fanta Bility"	2021-08-27	shot	unarmed	8

|	id	name	date	manner_of_death	armed	age
5	1165	"Ciara Meyer"	2016-01-11	shot	unarmed	12
6	7785	"Thomas Joseph Siderio"	2022-03-01	shot	gun	12
7	1883	"Tyre King"	2016-09-14	shot	toy weapon	13
8	6729	"Adam Toledo"	2021-03-29	shot	undetermined	13
9	8130	"Andre Hernandez"	2022-06-03	shot	undetermined	13
10	1776	"Jesse James Romero"	2016-08-09	shot	gun	14
11	3117	"Jason Ike Pero"	2017-11-08	shot	knife	14
12	4386	"Antonio Arce"	2019-01-15	shot	toy weapon	14
13	7450	"Valentina Orellana-Peralta"	2021-12-23	shot	unarmed	14
14	7797	"Juan Herrera"	2022-03-03	shot	undetermined	14

⋮

There are a few infant who were killed, maybe a bystander. Need more information on individual cases.

For instance, Jeremy Mardis, a 6 year old was shot to death. The police officers were sentenced to prison. More information - Killing of Jeremy Mardis - Wikipedia.

Investigation into Racial bias

```
gscatter(df.longitude,df.latitude,df.race)
```
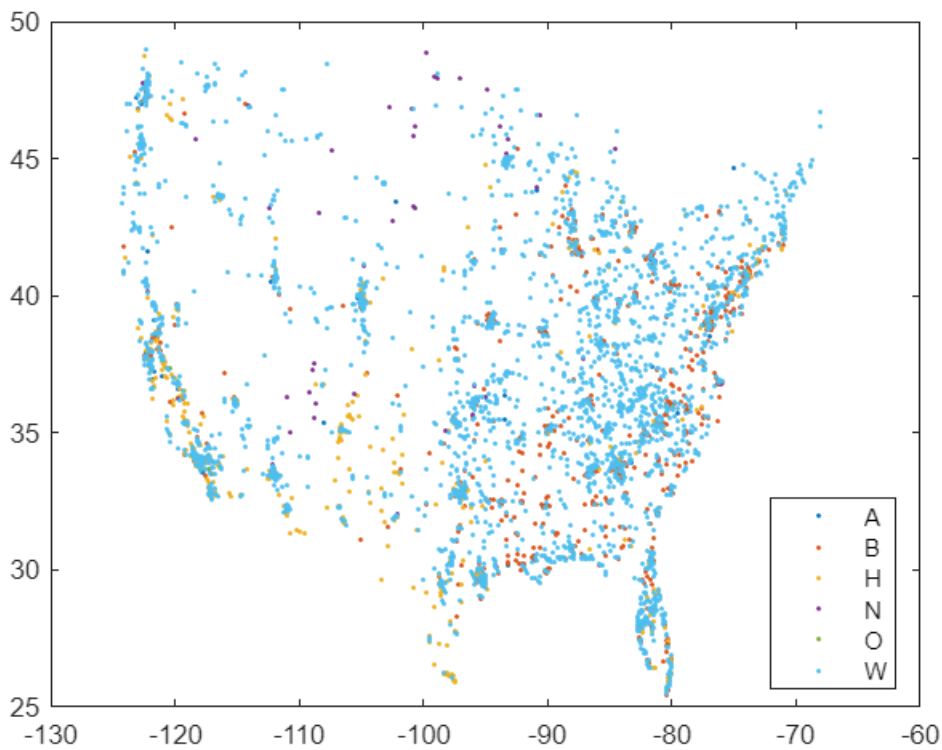
Difficult to differentiate. Let's look at yearly information.
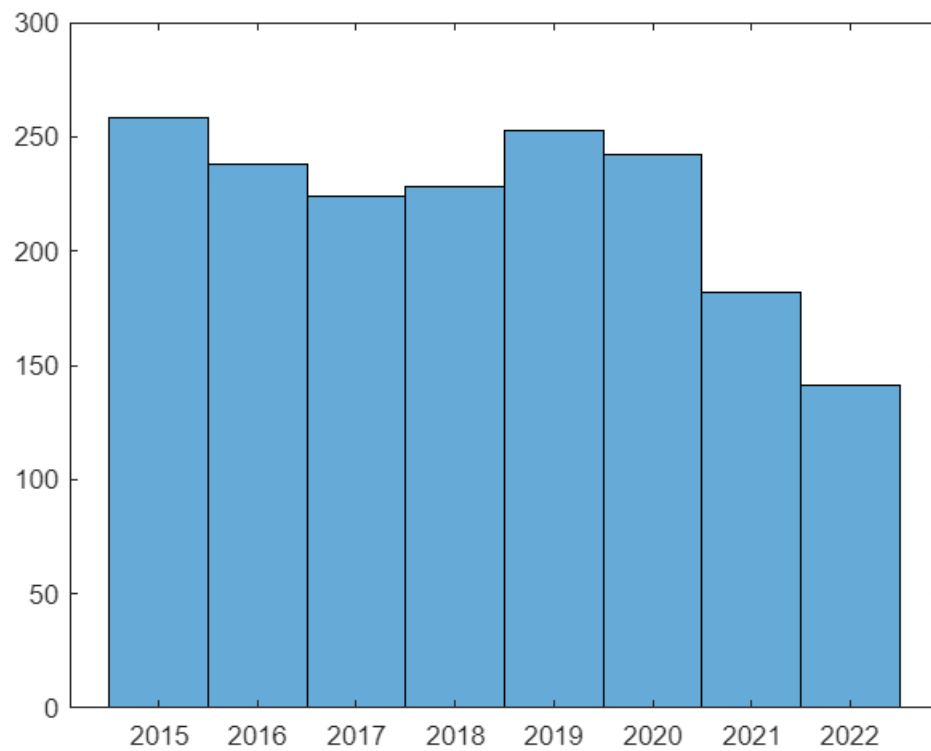
```
groupcounts(data,"race")
```

ans = 7×3 table

	race	GroupCount	Percent
1	A	129	1.6121
2	B	1766	22.0695
3	H	1166	14.5714
4	N	105	1.3122
5	O	19	0.2374
6	W	3300	41.2397
7	<undefined>	1517	18.9578

```
data_2020= df(df.year == 2020,:);
data_2021= df(df.year == 2021,:);
data_2022= df(df.year == 2022,:);
```
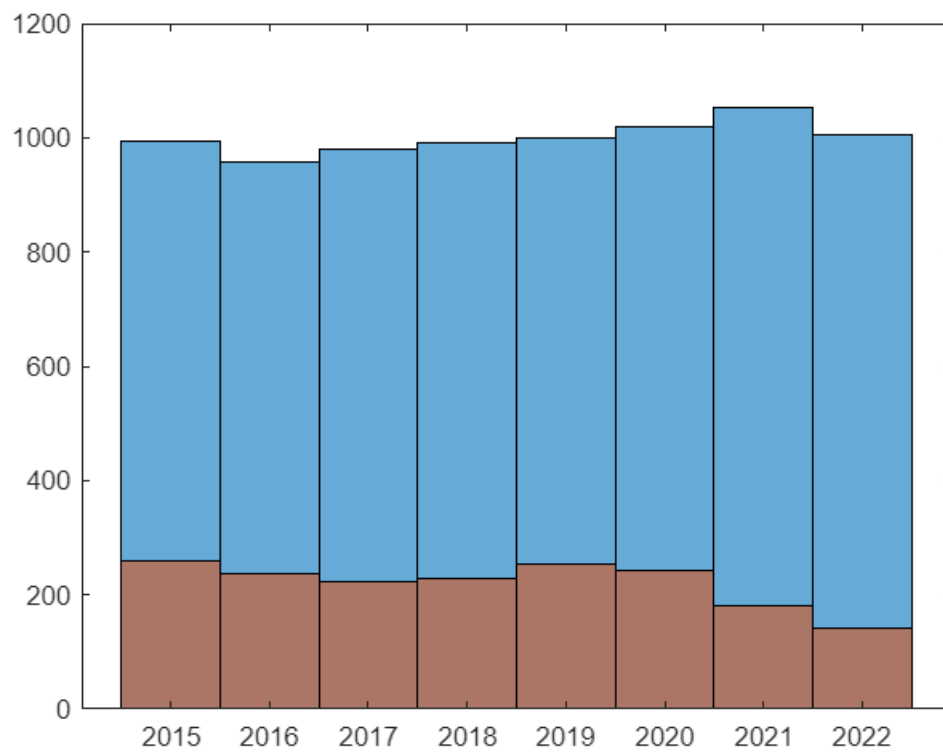
```
data_b= data(data.race == "B",:);
```

There are 1766 (22%) incidents happened to African-American people.

```
histogram(data_b.year)
```


Number of such incidents reduced in year 2021, 2022. lets see the overall percentage.

```
figure
histogram(data.year)
hold on
histogram(data_b.year)
hold off
```

In which states, there are more such shootings?

```
histogram(data_b.state)
```