An Analysis of Fatal Police Shootings in the US (2015-Present)

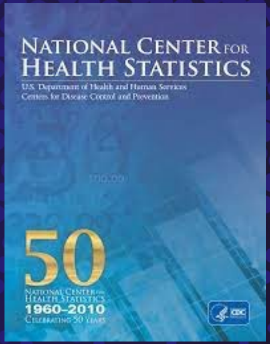
Andrea Perera-Ortega
CISC 878 Course Project

Introduction

- Use of deadly force by police is a high-profile and controversial issue in the US
- What is justifiable/unjustifiable
- Lack of data makes analysis difficult

Introduction

- 1994: U.S. Congress mandates that AG collect data on excessive force by police
  - Bureau of Justice Statistics stopped keeping count in March 2014
- National Center for Health Statistics: National Vital Statistics System (NVSS)
  - Does not document whether killing was legally justified or law enforcement involved
- FBI: Uniform Crime Reporting System (UCR)
  - 2007-2012, more than 550 homicides by country's 105 LE agencies missing from FBI records



Introduction

- If we can't rely on government to be consistent, then what can we do?

- 2014: Well publicized cases of Eric Garner, Michael Brown, and John Crawford III

- Emergence of crowd sourced projects to collect data
    - Fatal Encounters
    - U.S. Police Shootings Data at Deadspin
    - "Killed by Police" (FB page)
    - Mapping Police Violence
    - and more...

The Data

- The Washington Post
- Tracking fatal shootings in US, 2015-Present
- Hosted on Kaggle, GitHub
- 6,163 cases reported

The Data

```
id                          int64
name                       object
date                       object
manner_of_death            object
armed                      object
age                       float64
gender                     object
race                       object
city                       object
state                      object
signs_of_mental_illness      bool
threat_level               object
flee                       object
body_camera                  bool
longitude                 float64
latitude                  float64
is_geocoding_exact           bool
dtype: object
```

→ shot, shot_and_tasered

→ gun, knife, undetermined, etc.

→ White (NH), Black (NH), Hispanic, Asian, Native American, Other

→ attack, other, undetermined

→ foot, car, not_fleeing, other

The Data

- Null values per attribute
- Pretty significant # of race values missing
- Dropped everything missing except name (11 cases)

```
id                         0
name                     227
date                       0
manner_of_death            0
armed                    208
age                      275
gender                     2
race                     633
city                       0
state                      0
signs_of_mental_illness    0
threat_level               0
flee                     360
body_camera                0
longitude                301
latitude                 301
is_geocoding_exact         0
dtype: int64
```

The Data

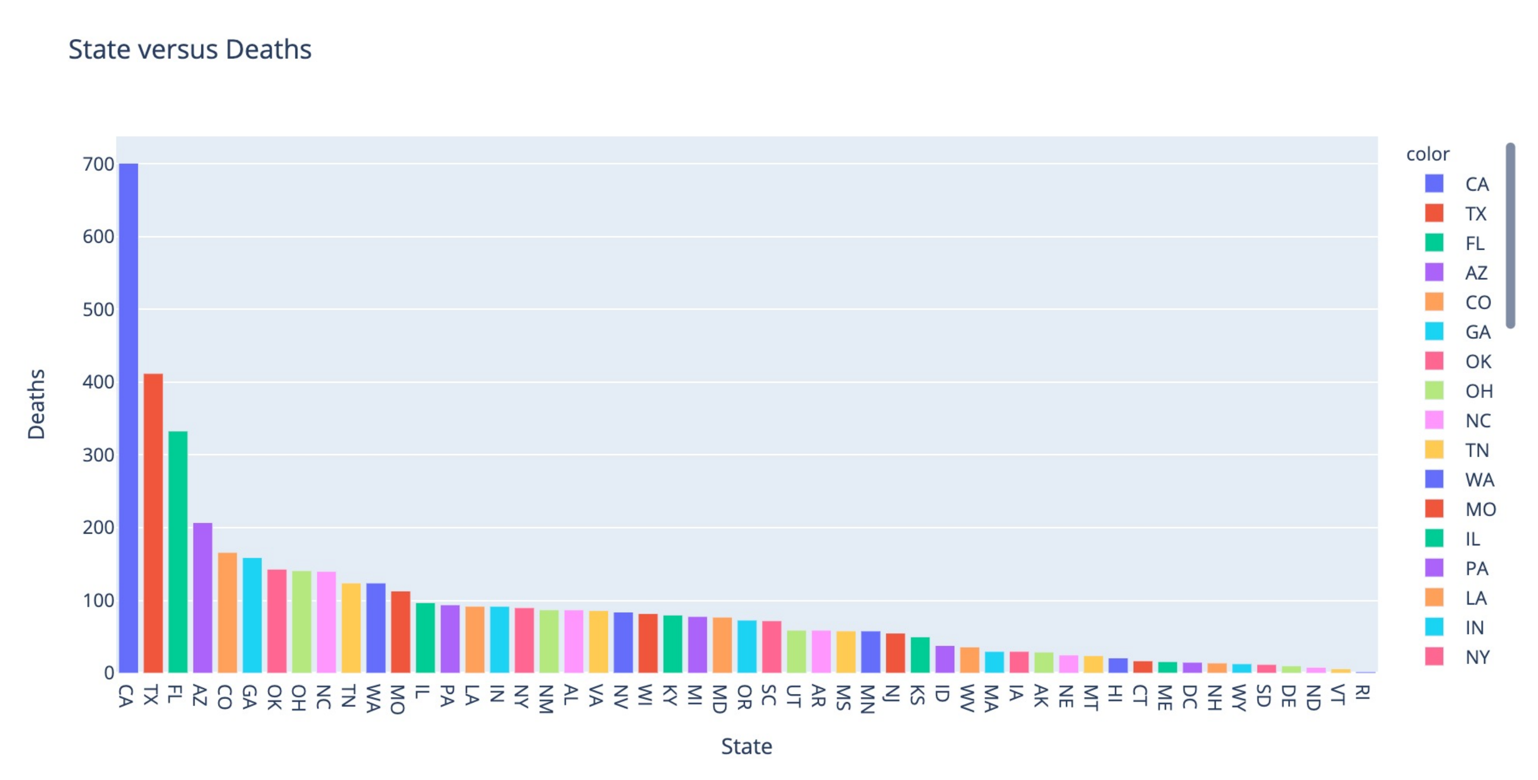
- 4719 x 17

	id	name	date	manner_of_death	armed	age	gender	race	city	state	signs_of_mental_illness	threat_level	flee	body_camera	longitud
0	3	Tim Elliot	2015-01-02	shot	gun	53.0	M	A	Shelton	WA	True	attack	Not fleeing	False	-123.12
1	4	Lewis Lee Lembke	2015-01-02	shot	gun	47.0	M	W	Aloha	OR	False	attack	Not fleeing	False	-122.89
2	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23.0	M	H	Wichita	KS	False	other	Not fleeing	False	-97.28
3	8	Matthew Hoffman	2015-01-04	shot	toy weapon	32.0	M	W	San Francisco	CA	True	attack	Not fleeing	False	-122.42
4	9	Michael Rodriguez	2015-01-04	shot	nail gun	39.0	M	H	Evans	CO	False	attack	Not fleeing	False	-104.69
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
4	6699	Christopher Ruffin	2021-03-14	shot	gun	28.0	M	B	Palm Bay	FL	False	attack	Foot	False	-80.61
5	6712	John Doe	2021-03-16	shot	gun	37.0	M	H	Los Angeles	CA	True	attack	Not fleeing	False	-118.25
6	6715	Leonard John Popa	2021-03-18	shot	gun	79.0	M	W	Pasadena	MD	True	attack	Not fleeing	False	-76.44
7	6719	Juan Jimenez-Salas	2021-03-18	shot	gun	46.0	M	H	Arlington	TX	False	attack	Not fleeing	False	-97.08
8	6709	Eduardo Parra	2021-03-21	shot	unarmed	24.0	M	H	Sylvania Township	OH	False	other	Foot	False	-83.66

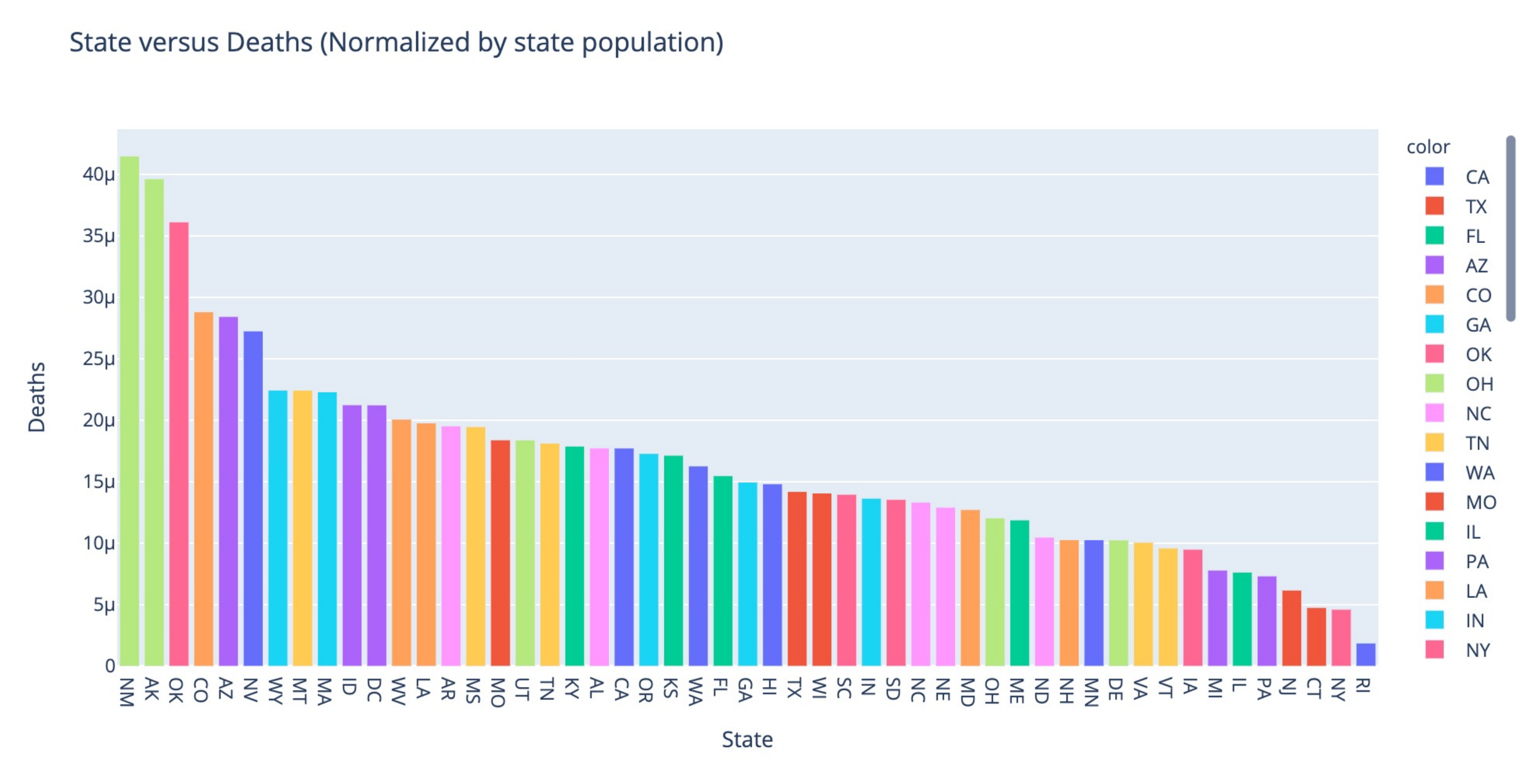
Exploratory Data Analysis (EDA)

- What we'll look at:
  - Distribution of location, time, age, gender, mental illness, race, etc.
  - Relationship and proportions between attributes

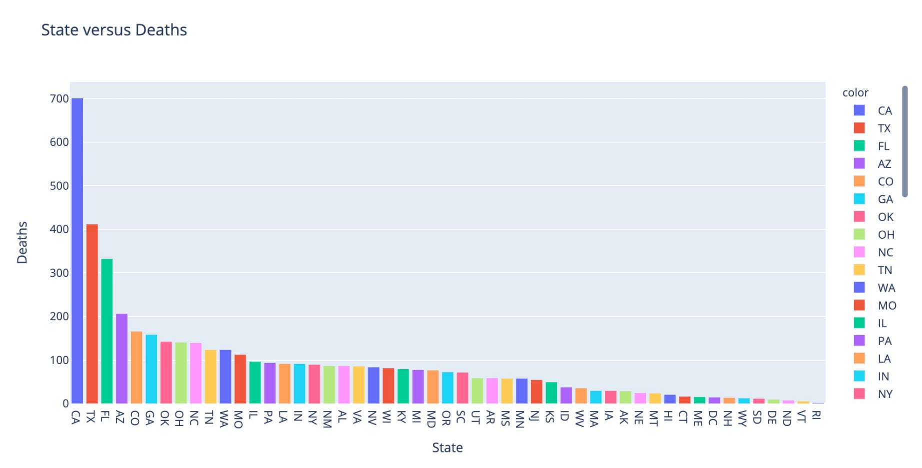
EDA – Location (State)



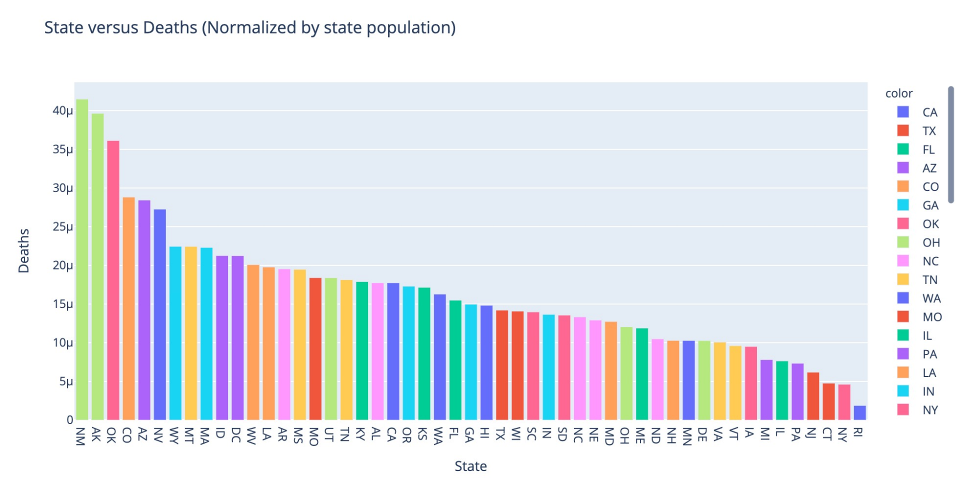
EDA – State: Normalized



EDA – State: Normalized (+ Gun ownership)



- California (42), Texas (18), Florida (24), Arizona (25), Colorado (21), Georgia (28), Oklahoma (29), North Carolina (33), Tennessee (15), Washington (34)

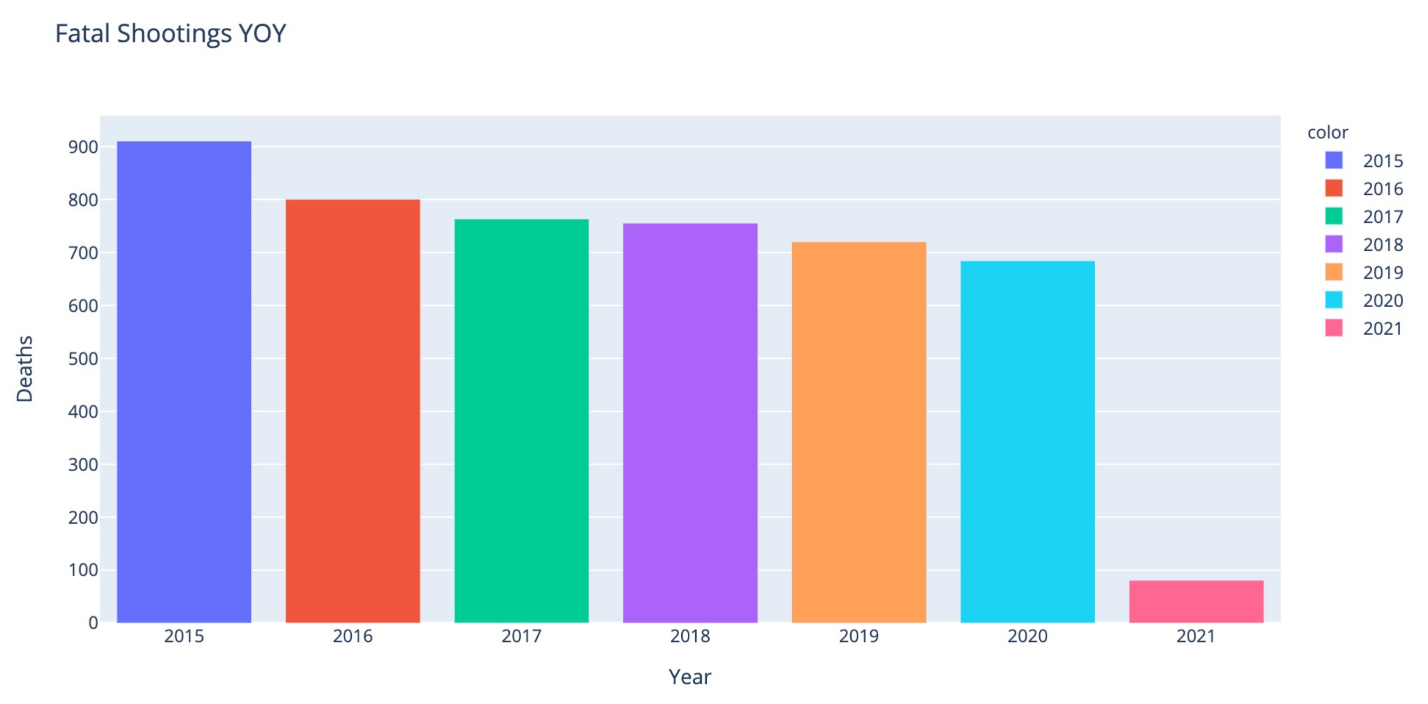


- *New Mexico(7), Alaska (1),* Oklahoma (29), Colorado (21), Arizona (25), Nevada (16), *Wyoming (5), Montana (6),* Maine (40), *Idaho (3)*

https://worldpopulationreview.com/state-rankings/gun-ownership-by-state

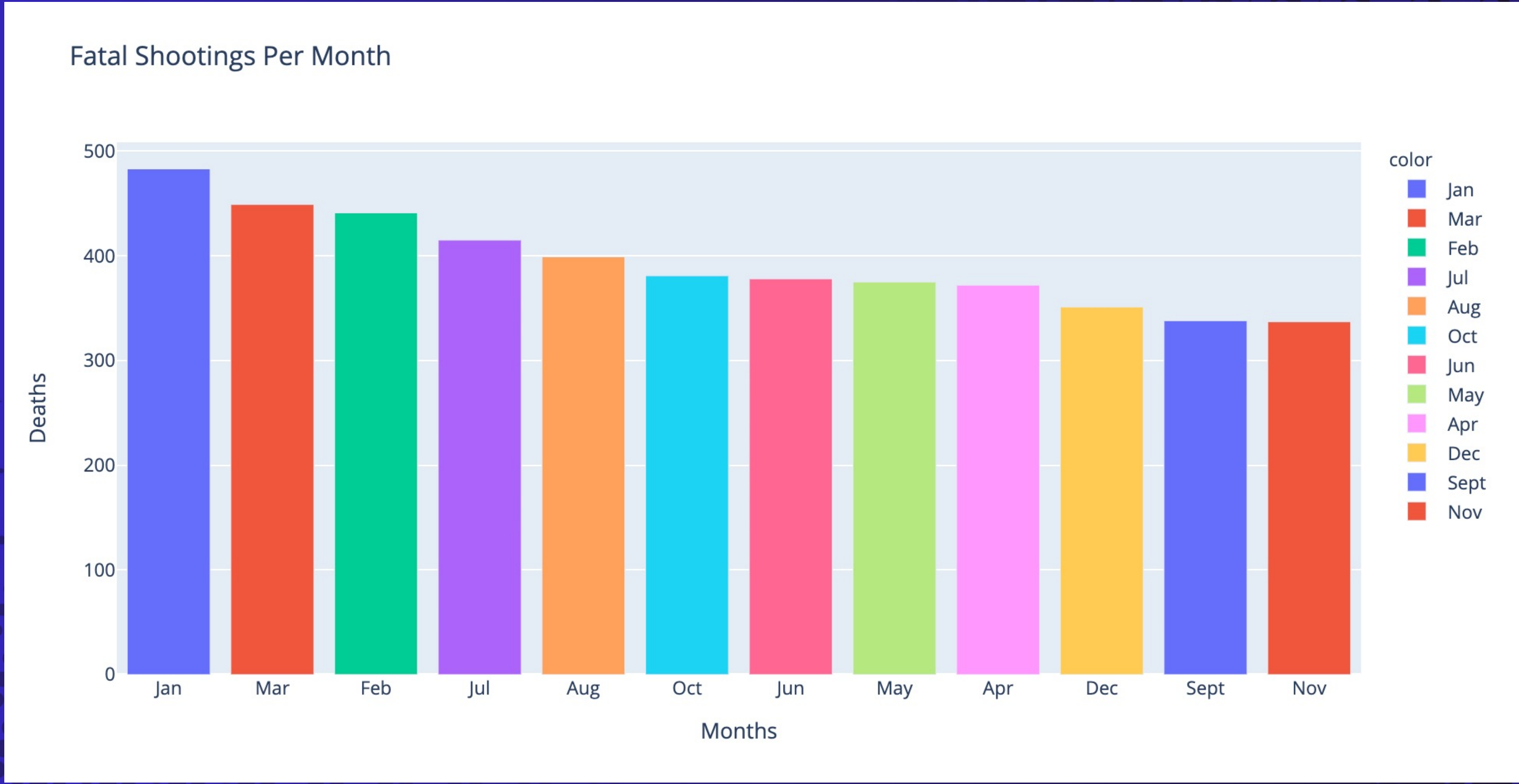
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6391295/

EDA – Time



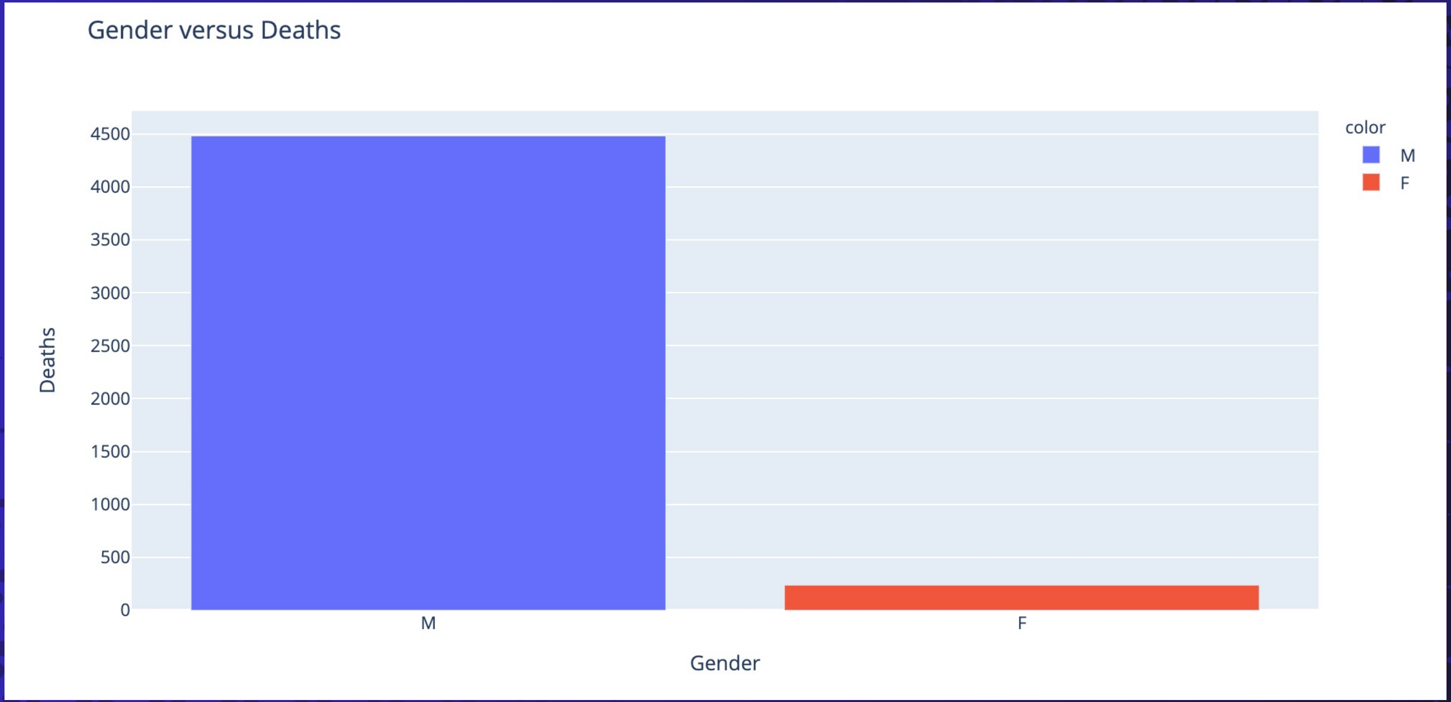
Fatal Shootings YOY
900
800
700
600
500
400
300
200
100
0
Deaths
2015
2016
2017
2018
2019
2020
2021
Year
color
2015
2016
2017
2018
2019
2020
2021

EDA – Time



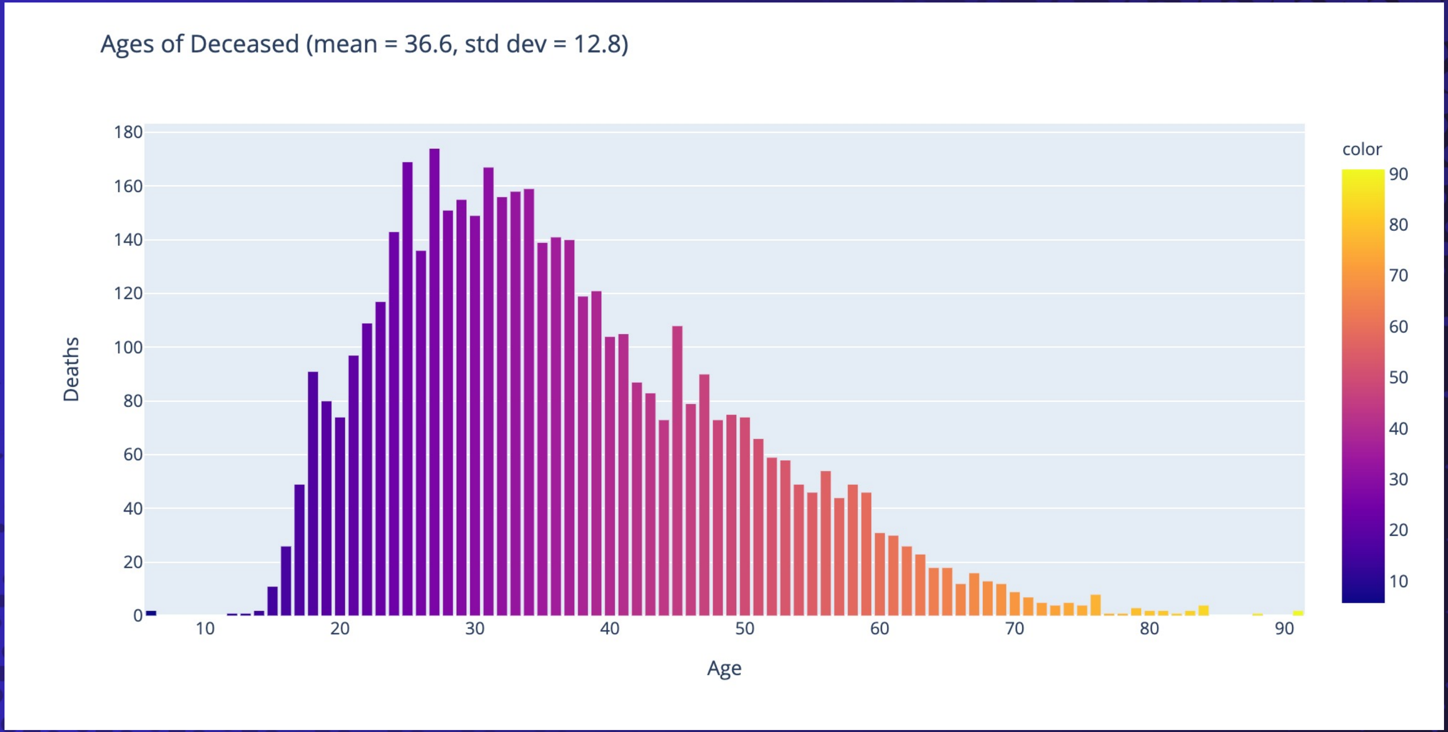
Fatal Shootings Per Month
500
400
300
200
100
0
Deaths
Jan
Mar
Feb
Jul
Aug
Oct
Jun
May
Apr
Dec
Sept
Nov
Months
color
Jan
Mar
Feb
Jul
Aug
Oct
Jun
May
Apr
Dec
Sept
Nov

EDA – Gender



Gender versus Deaths
color
M
F
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Deaths
M
F
Gender

EDA – Age

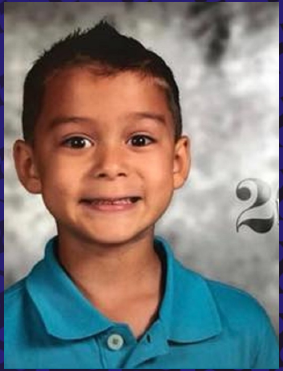


Ages of Deceased (mean = 36.6, std dev = 12.8)
Deaths
Age
color
0
20
40
60
80
100
120
140
160
180
10
20
30
40
50
60
70
80
90

EDA – Age

```
data_nonull['age'].describe()
```

```
count    4708.000000
mean       36.654630
std        12.782008
min         6.000000
25%        27.000000
50%        34.000000
75%        45.000000
max        91.000000
Name: age, dtype: float64
```



```
# initially I thought that the min value being 6 was an error, but it turns out that
# it is not an error.
print(data_nonull[data_nonull['age'] == 6])
```

```
        id              name        date manner_of_death    armed  age gender  \
776    980     Jeremy Mardis  2015-11-03            shot  unarmed  6.0      M
2462  3229  Kameron Prescott  2017-12-21            shot  unarmed  6.0      M

     race        city state  signs_of_mental_illness threat_level  \
776     W  Marksville    LA                    False        other
2462    W    Schertz    TX                    False        other

             flee  body_camera  longitude  latitude  is_geocoding_exact
776           Car         True    -92.050    31.125                True
2462  Not fleeing        False    -98.257    29.552                True
```

EDA – Minors in the Data

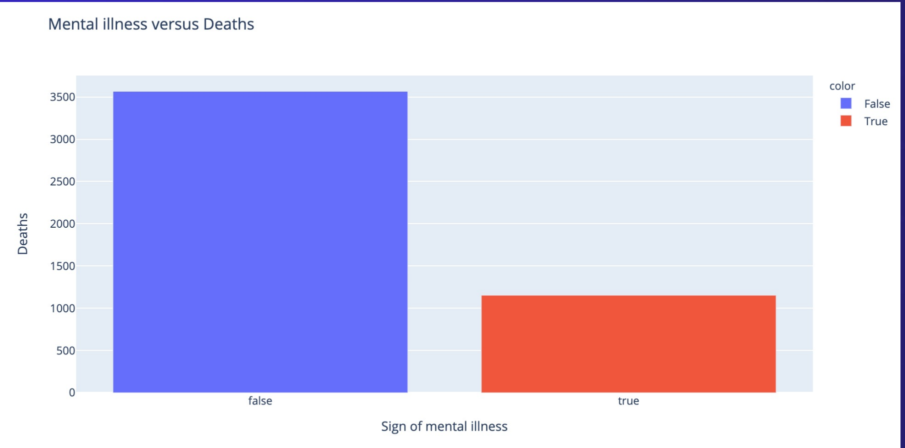
```
data_minor['age'].value_counts()
```

```
17.0    49
16.0    25
15.0    11
14.0     2
6.0      2
13.0     1
12.0     1
Name: age, dtype: int64
```

- 92 minors total

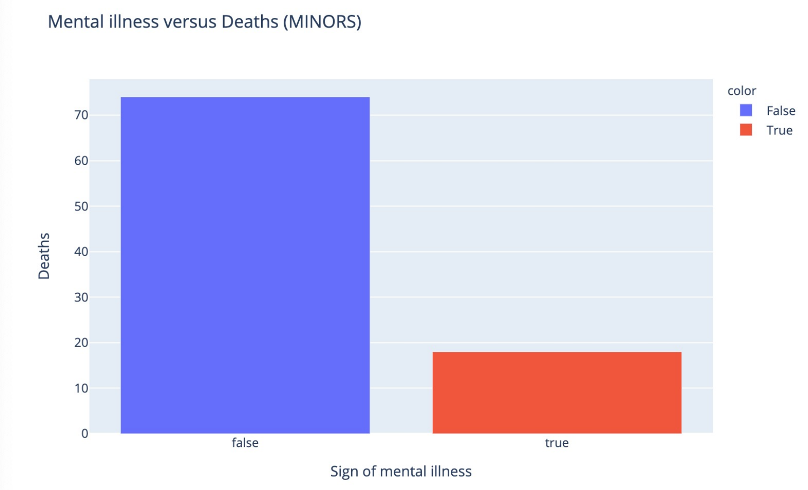
EDA – Mental Illness (Total vs. Minors)

TOTAL



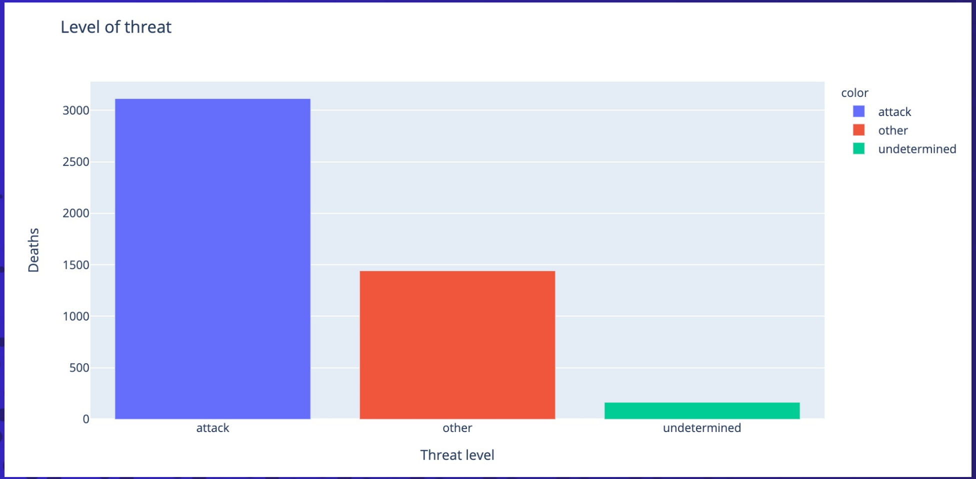

```
False percentage =  0.755668573850392
True percentage =  0.24433142614960796
```

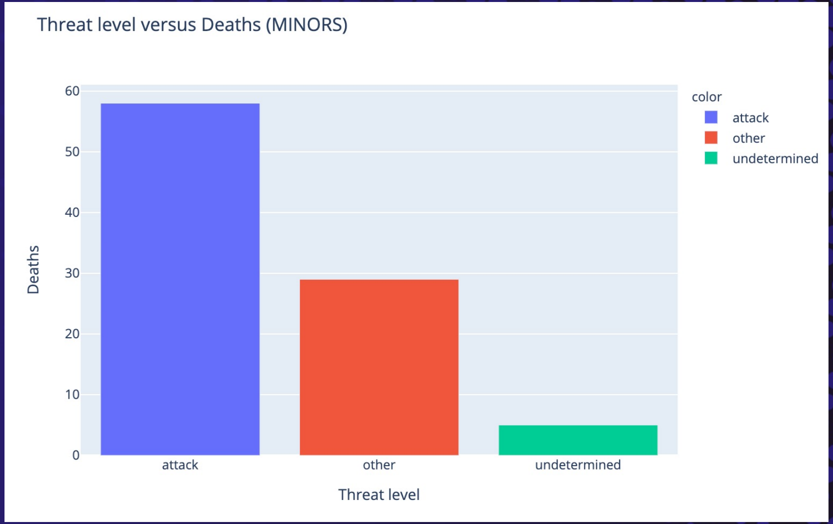
MINORS



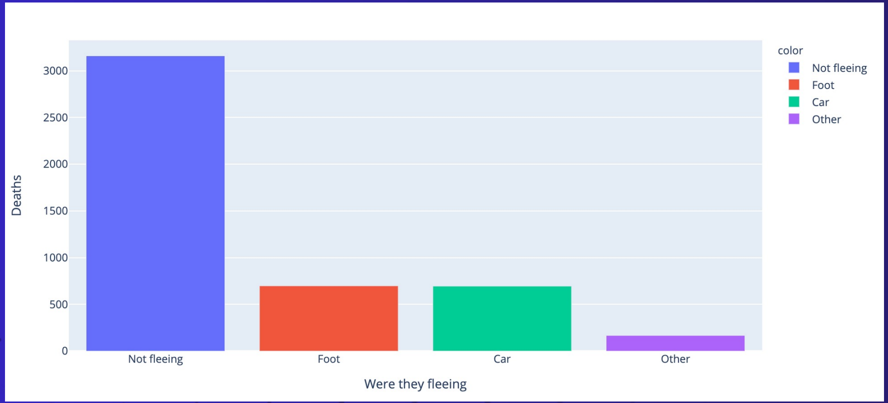

```
False percentage =  0.8043478260869565
True percentage =  0.1956521739130435
```

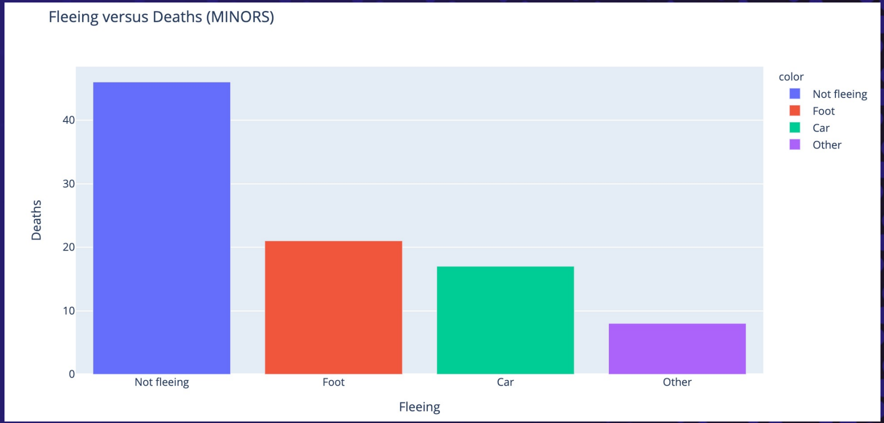
EDA – Threat Level (Total vs. Minors)



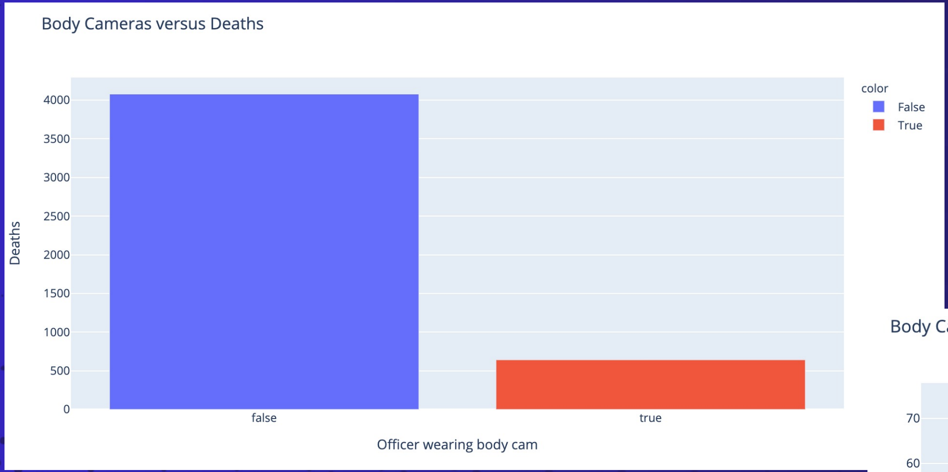


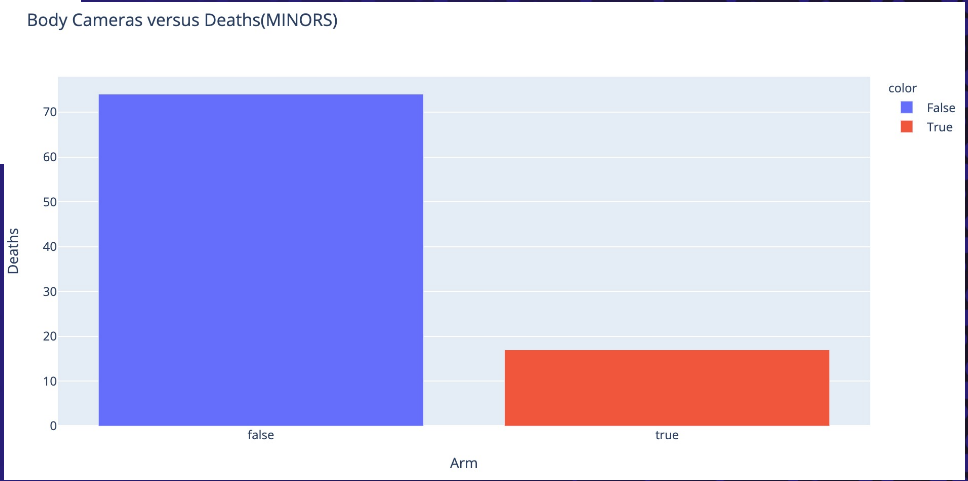
EDA – Fleeing (Total vs. Minors)



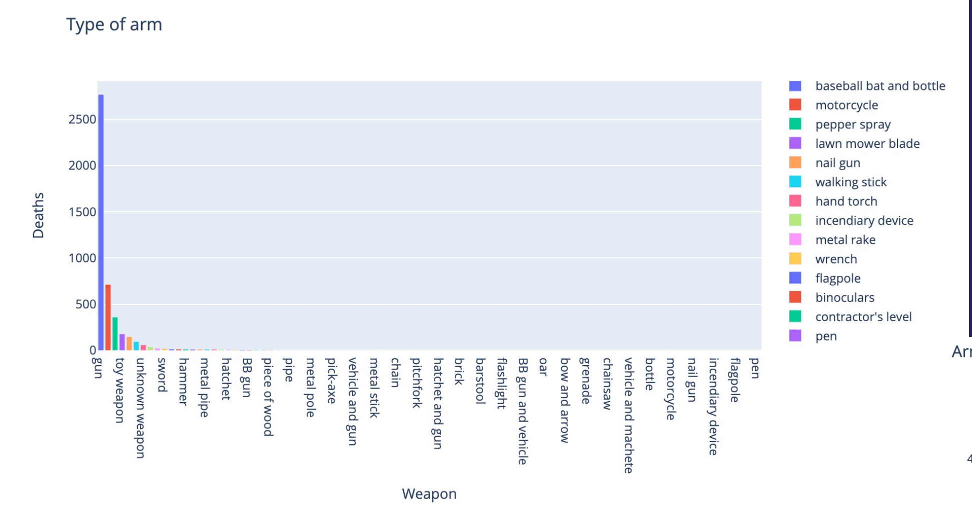


EDA – Body Cameras (Total vs. Minors)

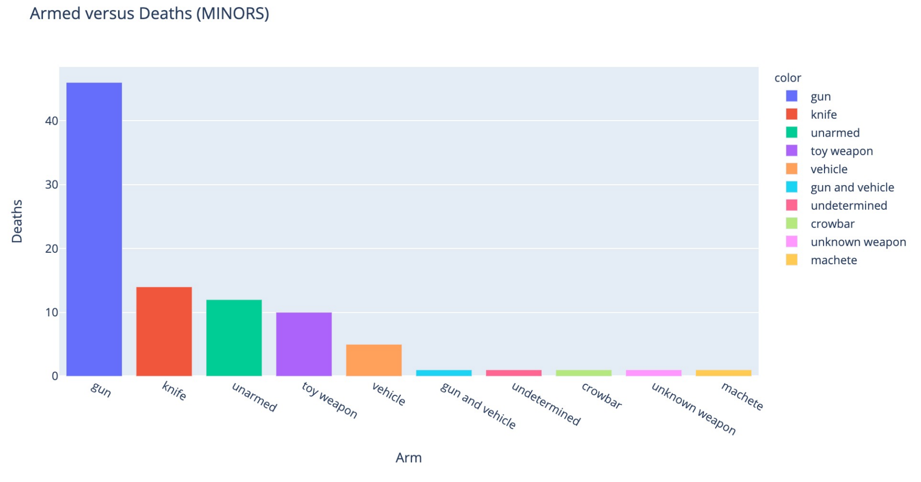




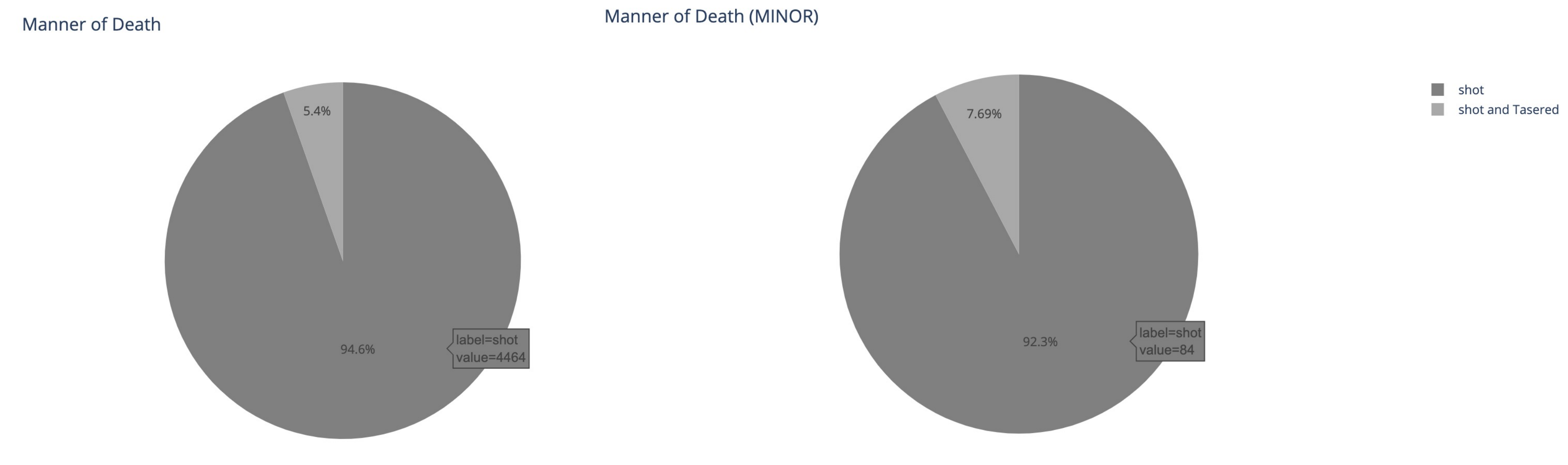
EDA – Armed? (Total vs. Minors)



Note that 362 individuals were unarmed. 97 were undetermined.

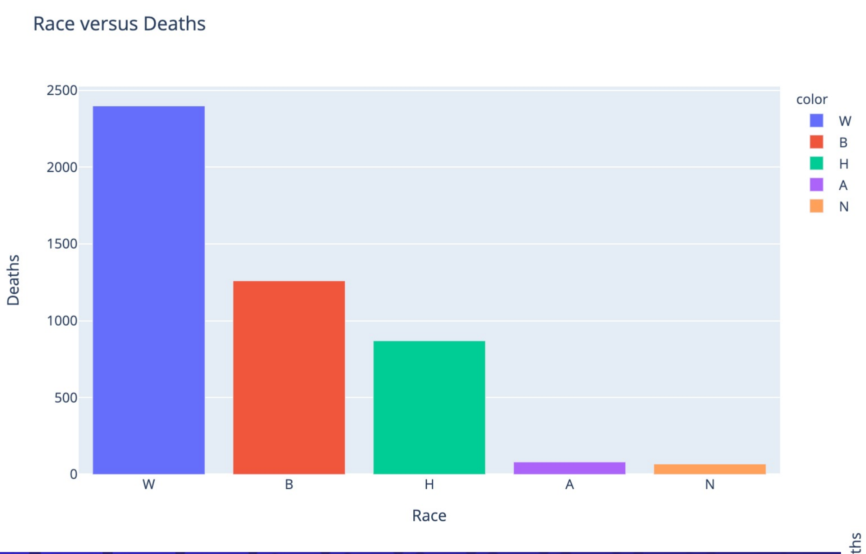


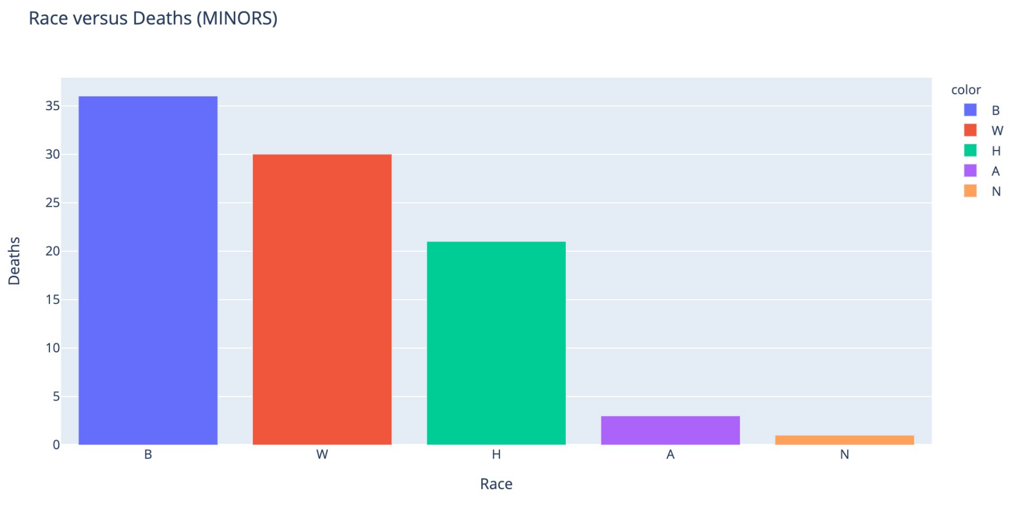
EDA – Manner of Death (Total vs. Minors)



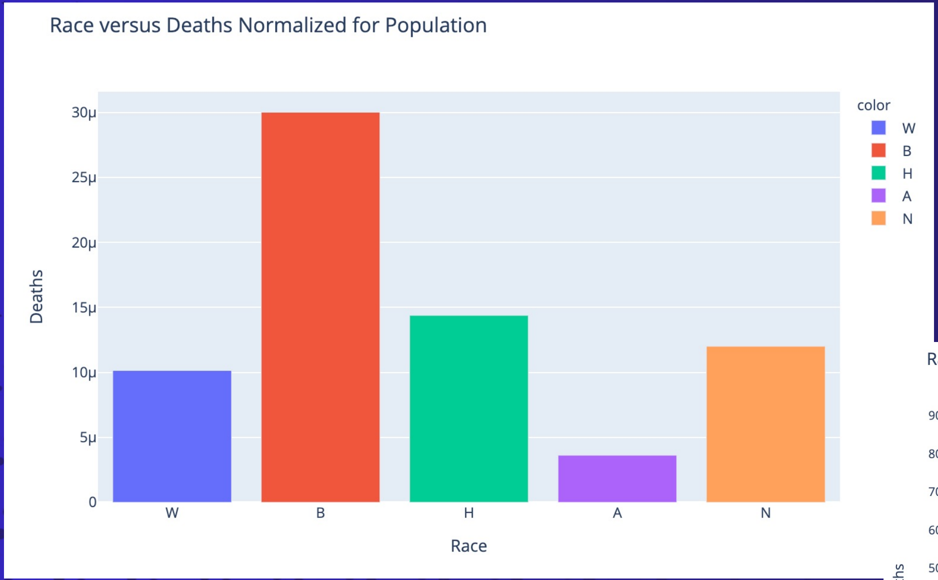
EDA – Race (Total vs. Minors)

- W: White (Non-Hispanic)
- B: Black (NH)
- H: Hispanic
- A: Asian
- N: Native American
- O: Other (omitted here)

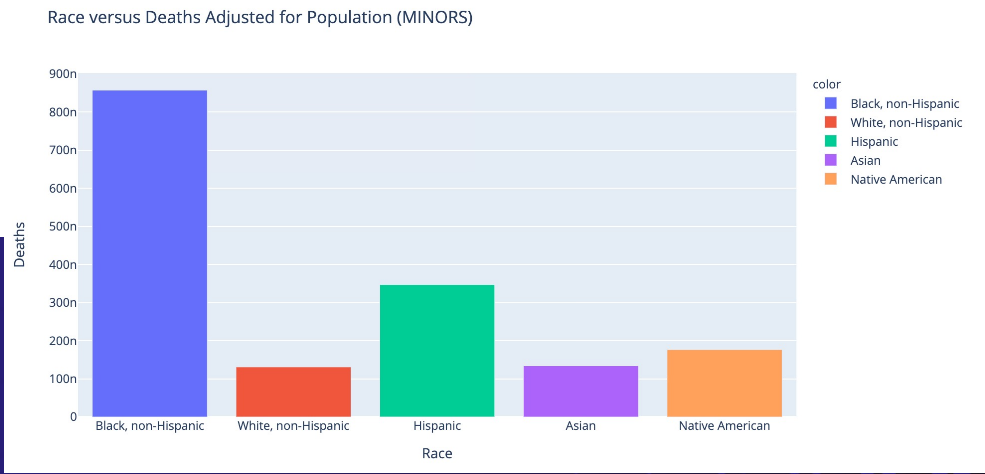




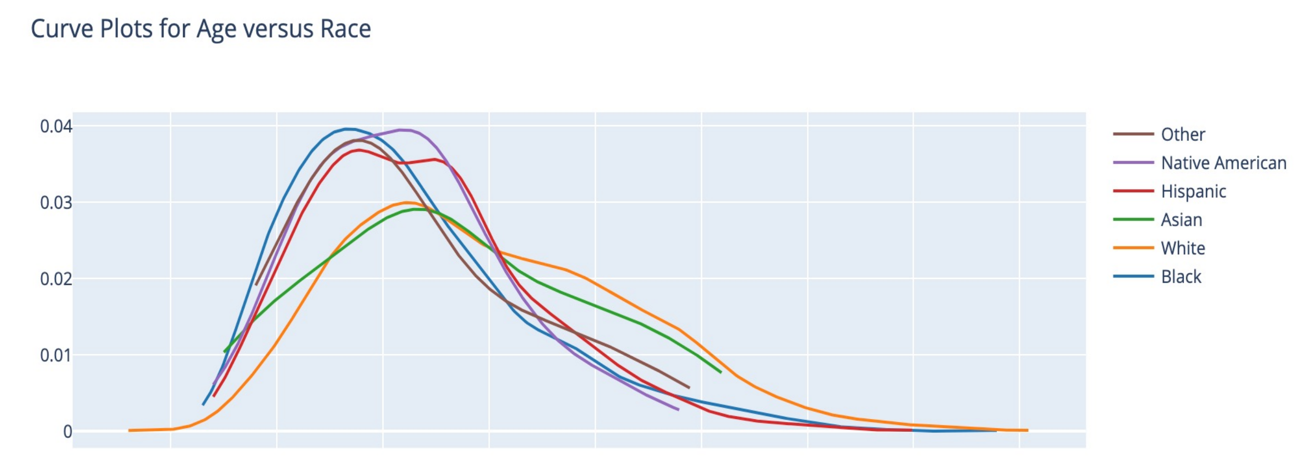
EDA – Race Relative to Pop. (Total vs. Minors)



- Black proportion is 3x higher in the total dataset
- 10x higher in minor dataset



EDA – Age versus Race



MEAN AGE

- Native American: 32.1
- Black: 32.6
- Other: 32.8
- Hispanic: 33.8
- Asian: 36.6
- White: 40

Modelling

- Missing/incomplete data is an issue
- Recall that there were 633 null values in race category
- *Can we predict these using other variables?*

Modelling

- ‘ethnicolor’ library
- Predict race and ethnicity from name
- US census data, Florida voting registration data, Wikipedia data
- LSTM model
- 20% test set (stratified)

Modelling

- Census model output

- Output: Appends the following columns to the pandas DataFrame or CSV: race (white, black, asian, or hispanic), api (percentage chance asian), black, hispanic, white.

- Wiki model output

- Output: Appends the following columns to the pandas DataFrame or CSV: race (categorical variable — category with the highest probability), “Asian,GreaterEastAsian,EastAsian”, “Asian,GreaterEastAsian,Japanese”, “Asian,IndianSubContinent”, “GreaterAfrican,Africans”, “GreaterAfrican,Muslim”, “GreaterEuropean,British”,”GreaterEuropean,EastEuropean”, “GreaterEuropean,Jewish”,”GreaterEuropean,WestEuropean,French”, “GreaterEuropean,WestEuropean,Germanic”,”GreaterEuropean,WestEuropean,Hispanic”, “GreaterEuropean,WestEuropean,Italian”,”GreaterEuropean,WestEuropean,Nordic”

Modelling

- Census model output

```
census = pred_census_ln(data_no_race, 'last_name')
```

```
actual:
W     479
B     251
H     174
A      16
N      14
O       8
```

```
census predictions:
white          715
hispanic       179
black           35
api             13
```

Modelling

- Make labels match
- Generate confusion matrix

Actual →

Prediction ↓

	A	B	H	N	O	W
	7	0	2	0	0	7
	3	24	9	0	0	$\boxed{215}$
	0	0	157 ✓	0	0	17
	0	0	0	0	0	14
	2	0	0	0	0	6
	1	11	11	0	0	456 ✓

```
array([[  7,   0,   2,   0,   0,   7],
       [  3,  24,   9,   0,   0, 215],
       [  0,   0, 157,   0,   0,  17],
       [  0,   0,   0,   0,   0,  14],
       [  2,   0,   0,   0,   0,   6],
       [  1,  11,  11,   0,   0, 456]])
```

```
print(accuracy_score(y_test, y_pred_census['race']))
```

```
0.6836518046709129
```

Modelling

- Wiki model is more specific but not apples to apples

```
y_pred_wiki = pred_wiki_name(x_test, fname_col='first_name', lname_col='last_name')
```

```
actual:
W        479
B        251
H        174
A         16
N         14
O          8
```

```
y_pred_wiki['race'].value_counts()
```

```
GreaterEuropean,British                          626
GreaterEuropean,WestEuropean,Hispanic            139
GreaterEuropean,WestEuropean,Italian              57
GreaterEuropean,Jewish                            48
GreaterEuropean,WestEuropean,French               15
GreaterAfrican,Africans                            9
GreaterEuropean,WestEuropean,Nordic                9
GreaterEuropean,EastEuropean                       9
GreaterEuropean,WestEuropean,Germanic              8
Asian,IndianSubContinent                           7
GreaterAfrican,Muslim                              6
Asian,GreaterEastAsian,EastAsian                   5
Asian,GreaterEastAsian,Japanese                    4
```

Modelling Conclusions

- Using names to predict race is a hard task
- So many variables
  - Religion, marriage, immigration, assimilation, history, etc.
  - Black individuals most often got misclassified as white
- Race/ethnicity categories themselves aren't very descriptive

actual:

W	479
B	251
H	174
A	16
N	14
O	8

?

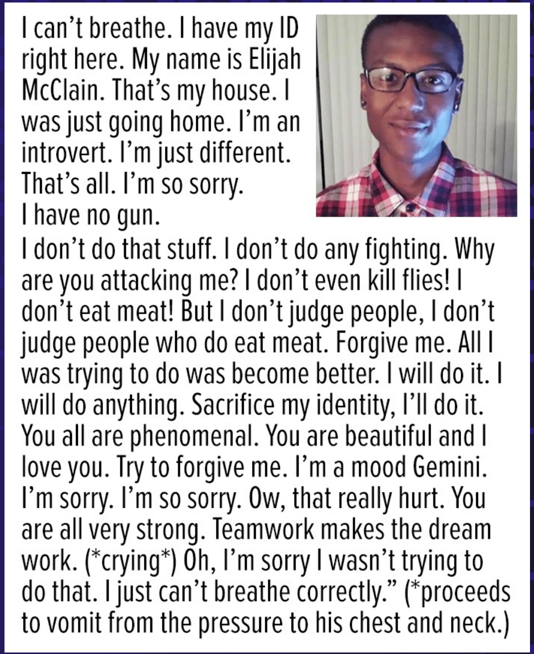
Modelling – “Other” Category

```
print(data_nonull[data_nonull['race'] == 'O'].name)
```

```
35                             Zaki Shinwary
306                      Fridoon Zalbeg Nehad
311                               Nadir Soofi
327                             Bruce Zalonka
361                               Feras Morad
487            Mohammad Youssef Abdulazeez
676                             Robert Berger
683                              Philip Quinn
706                                  Omar Ali
757                            Kobvey Igbuhay
779                          Faisal Mohammad
841                             Tashfeen Malik
842                               Syed Farook
872                           Nephi  Leiataua
1008                        Scottie Yanagawa
1013                         Anthony Bertoni
1225                             Derek J. Sam
1284                       Kalyp Allen Rapoza
1299                              Omar Mateen
1307                      Mohammad Moghaddam
1382                              BJ Medeiros
1517                               Dahir Adan
1654                    Abdul Razak Ali Artan
1738                          Hafez Abousamra
2059                              Isaiah Obet
2151                           Farhad Jabbari
2291                          Austin Dunsmore
2293                           Dillan Tabares
2894      Steven Allan Kaluahinui Hyer Jr.
2943                            Justin Waiki
3100                    Ashley Elisna Grammer
3296                            Siatu'u Tauai
3346                             Kasim Kahrim
3507                          Ronnie Churches
3519                         Hamid Ould-Rouis
3597                   Isak Abdirahman Aden
3839                       Michael Kahalehoe
3908                  Mohammad Jamal Isaifan
3910                           Dustin Spencer
3955         Gerardo Antonio Conchas-Bustas
4086                        Ajay Kamil Ayseli
4255                    Heba Momtaz Al-Azhari
```

Conclusions

- We need more descriptive data to paint the full picture
    - No data on individuals who were shot but did not die
    - Data only includes individuals who died via police shooting, not other means
        - Eric Garner, Elijah McClain, George Floyd
    - Null values



Conclusions

- Race and ethnicity are too generalized in the data
- Don't know demographics of police officers involved
- Nature of incidents not known in detail
- But...
  - *Preliminary findings allow us to ask more questions, make changes to get better data*

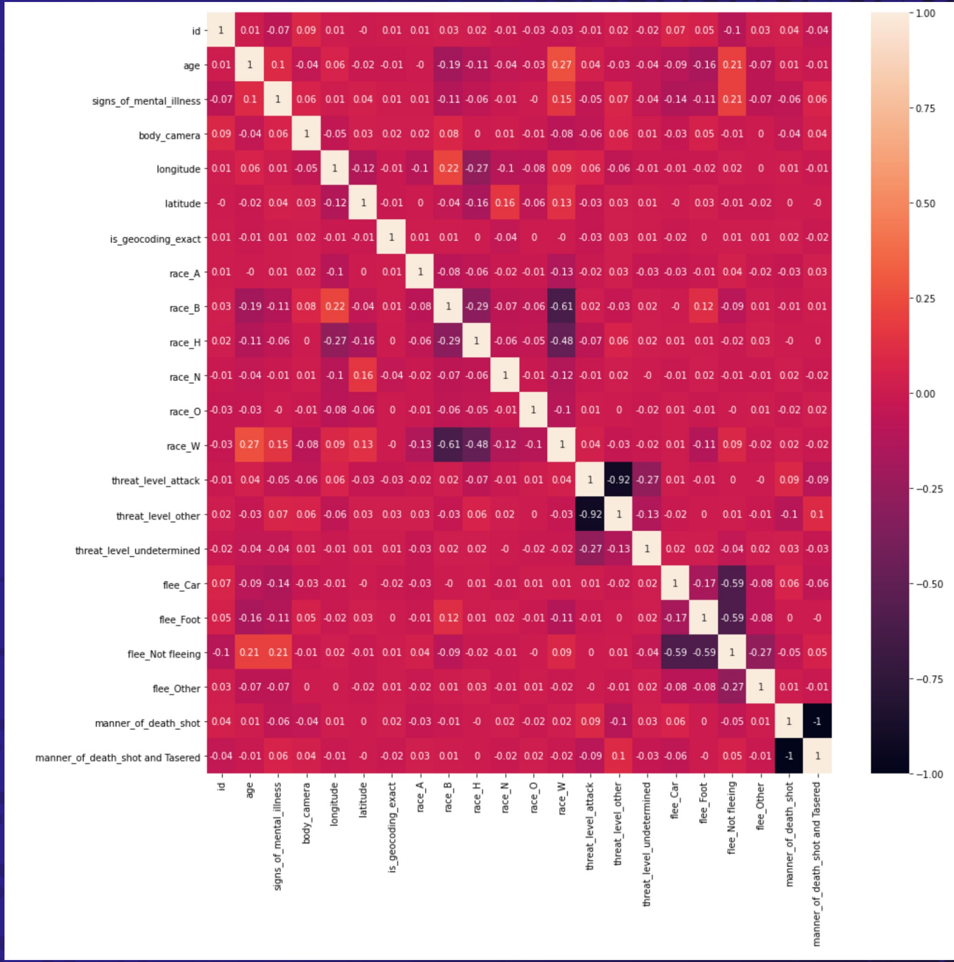
Thank you

References

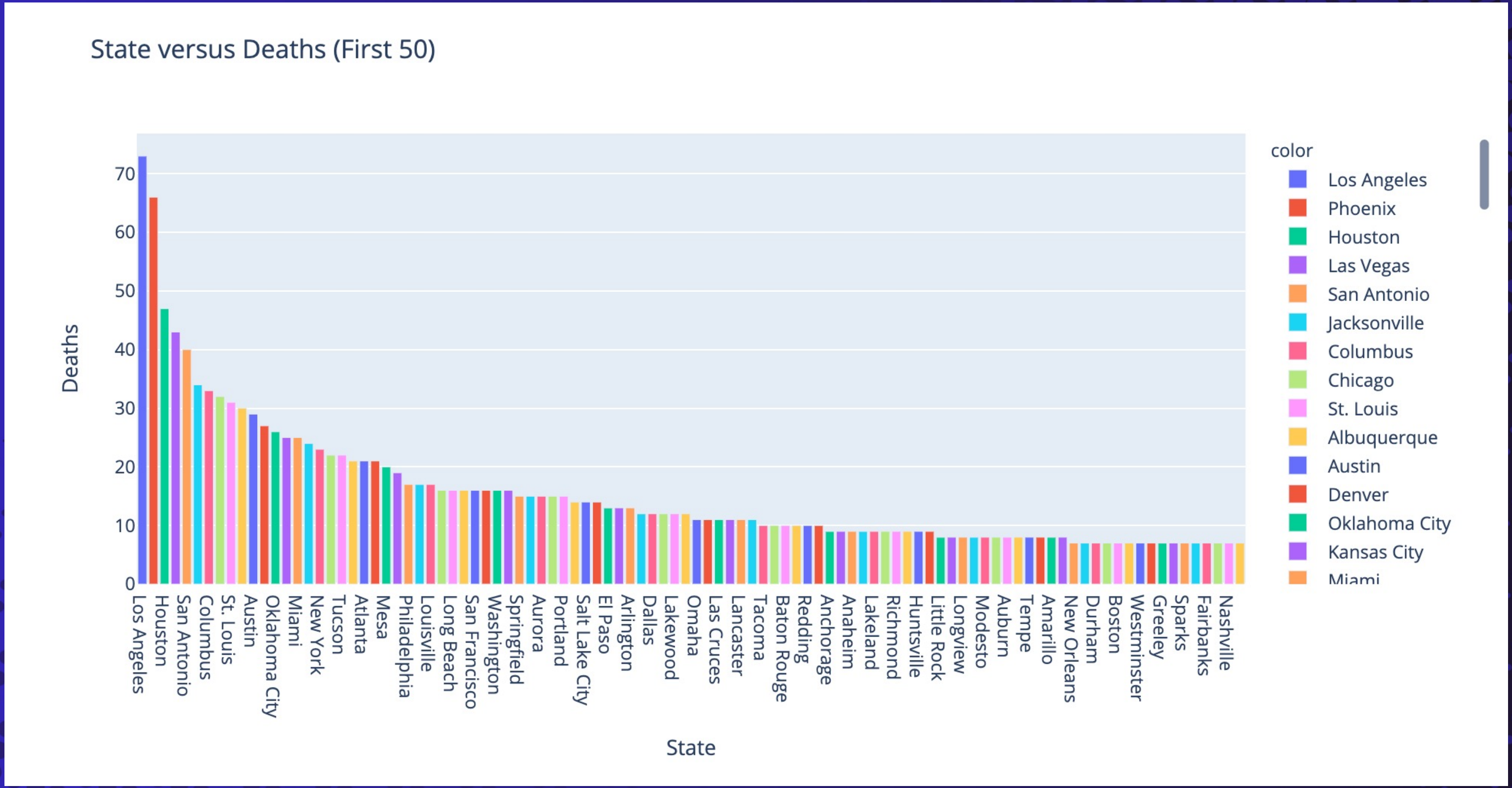
[1] https://www.pnas.org/content/116/32/15877
[2] https://worldpopulationreview.com/state-rankings/gun-ownership-by-state
[3] https://pypi.org/project/ethnicolr/
[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6391295/
[5] https://plotly.com/python/
[6] https://www.infoplease.com/us/states/state-population-by-rank
[7] https://deadspin.com/were-compiling-every-police-involved-shooting-in-americ-1624180387
[8] https://datacommons.org/place/geoId/06?utm_medium=explore&mprop=count&popt=Person&hl=en
[9] https://github.com/washingtonpost/data-police-shootings
[10] https://sanfrancisco.cbslocal.com/2016/11/21/teen-stabbing-suspect-fatally-shot-by-santa-cruz-deputy-identified/
[11] https://www.washingtonpost.com/news/post-nation/wp/2014/09/25/ohio-wal-mart-surveillance-video-shows-police-shooting-and-killing-john-crawford-iii/
[12] https://pandas.pydata.org/docs/index.html
[13] https://www.bjs.gov/content/pub/pdf/revcoa18.pdf
[14] https://www.nytimes.com/article/who-was-elijah-mcclain.html
[15] https://www.latimes.com/california/story/2020-06-05/george-floyd-carotid-neck-hold-police
[16] https://scikit-learn.org/stable/index.html
[17] https://www.washingtonpost.com/graphics/investigations/police-shootings-database/

Appendix

Correlation



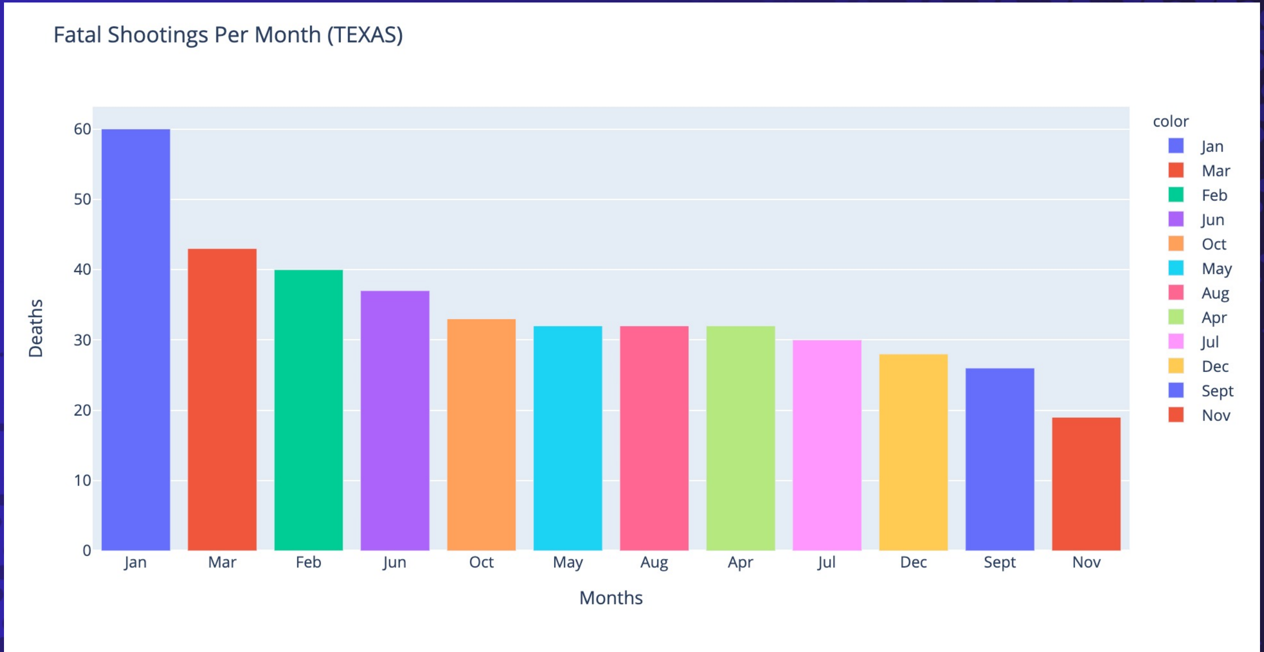
EDA – Location (City)



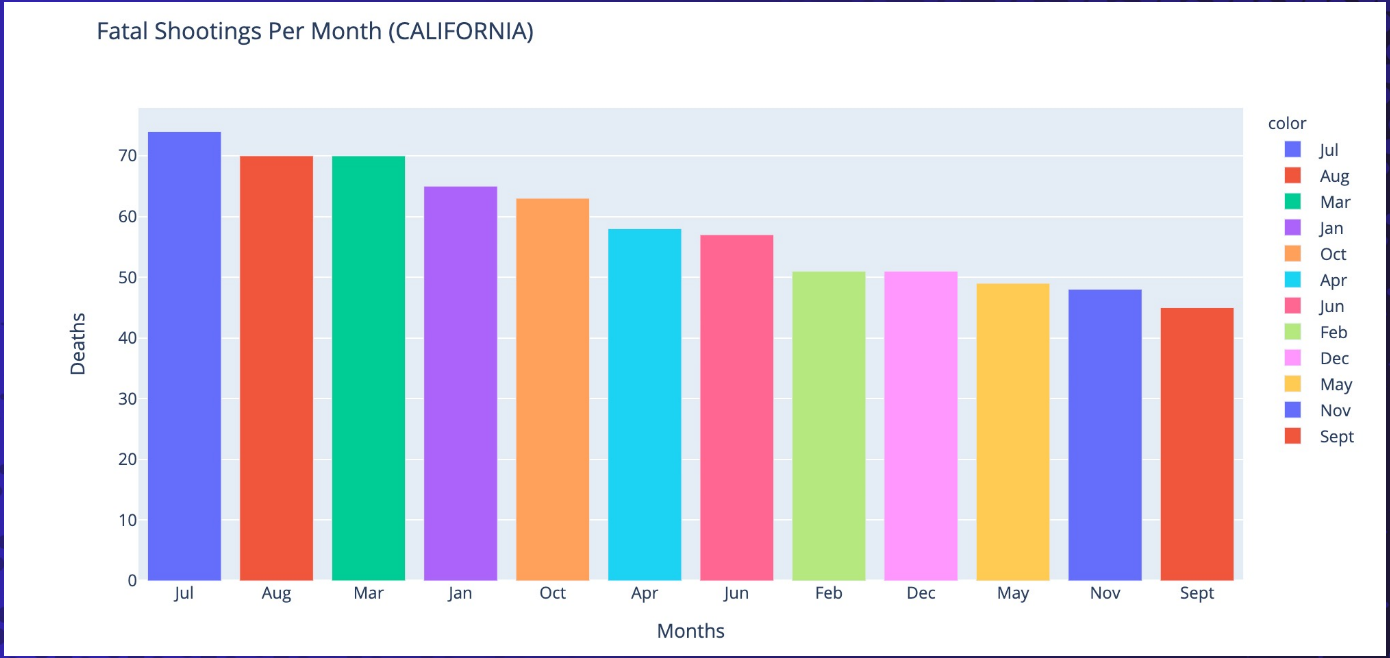
Gun Ownership (%)

State	Gun Ownership	Registered Guns
Alaska	61.70%	15,824
Arkansas	57.90%	79,841
Idaho	56.90%	49,566
West Virginia	54.20%	35,264
Wyoming	53.80%	132,806
Montana	52.30%	22,133
New Mexico	49.90%	97,580
Alabama	48.90%	161,641
North Dakota	47.90%	13,272
Hawaii	45.10%	7,859
Louisiana	44.50%	116,831
South Carolina	44.40%	105,601
Mississippi	42.80%	35,494
Kentucky	42.40%	81,058
Tennessee	39.40%	99,159
Nevada	37.50%	76,888
Minnesota	36.70%	79,307
Texas	35.70%	588,696
South Dakota	35.00%	21,130
Wisconsin	34.70%	64,878
Colorado	34.30%	92,435
Iowa	33.80%	28,494
Indiana	33.80%	114,019
Florida	32.50%	343,288
Arizona	32.30%	179,738
Kansas	32.20%	52,634
Utah	31.90%	72,856
Georgia	31.60%	190,050
Oklahoma	31.20%	71,269
Virginia	29.30%	307,822
Vermont	28.80%	5,872
Michigan	28.80%	65,742
North Carolina	28.70%	152,238
Washington	27.70%	91,835
Pennsylvania	27.10%	236,377
Missouri	27.10%	72,995
Oregon	26.60%	61,383
Illinois	26.20%	146,487
Massachusetts	22.60%	37,152
Maine	22.60%	15,371
Maryland	20.70%	103,109
California	20.10%	344,622
Nebraska	19.80%	22,234
Ohio	19.60%	173,405
Connecticut	16.60%	82,400
New Hampshire	14.40%	64,135
New Jersey	11.30%	57,505
New York	10.30%	76,207
Rhode Island	5.80%	4,223
Delaware	5.20%	4,852

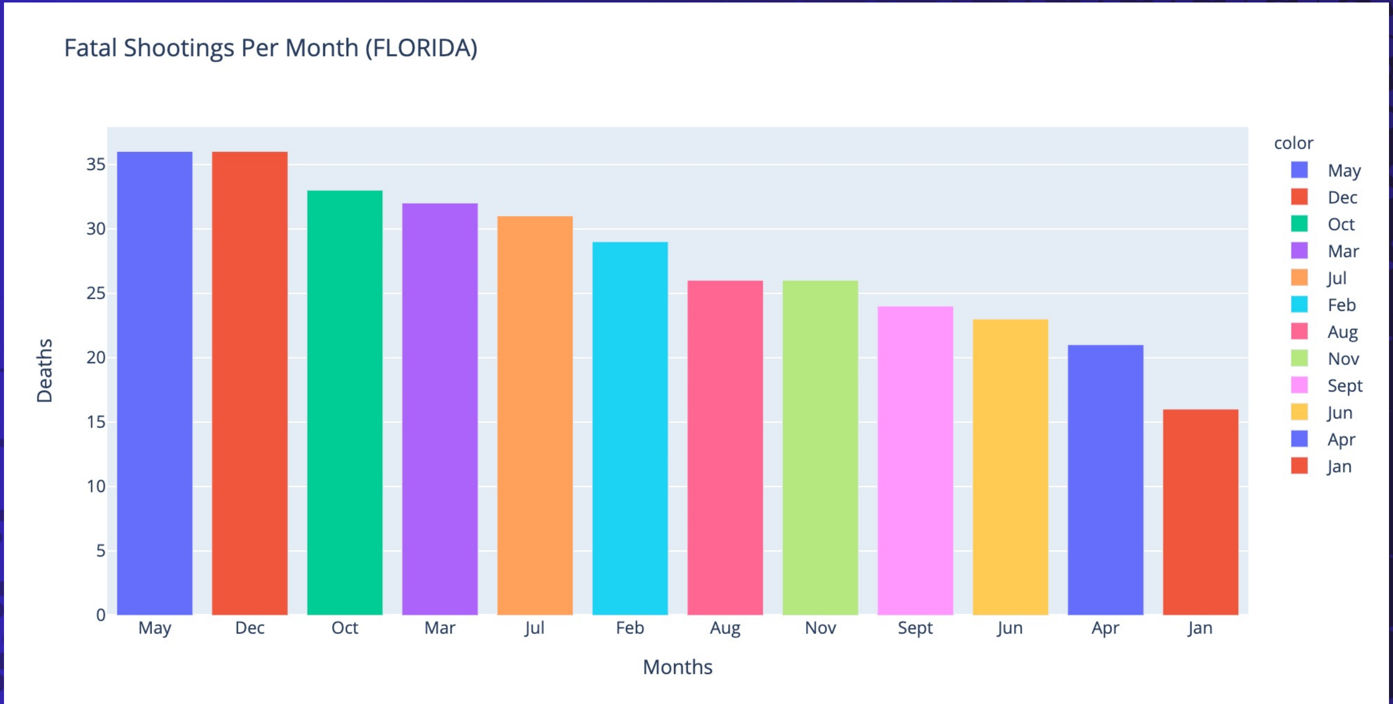
Fatal Shootings Per Month (Texas)



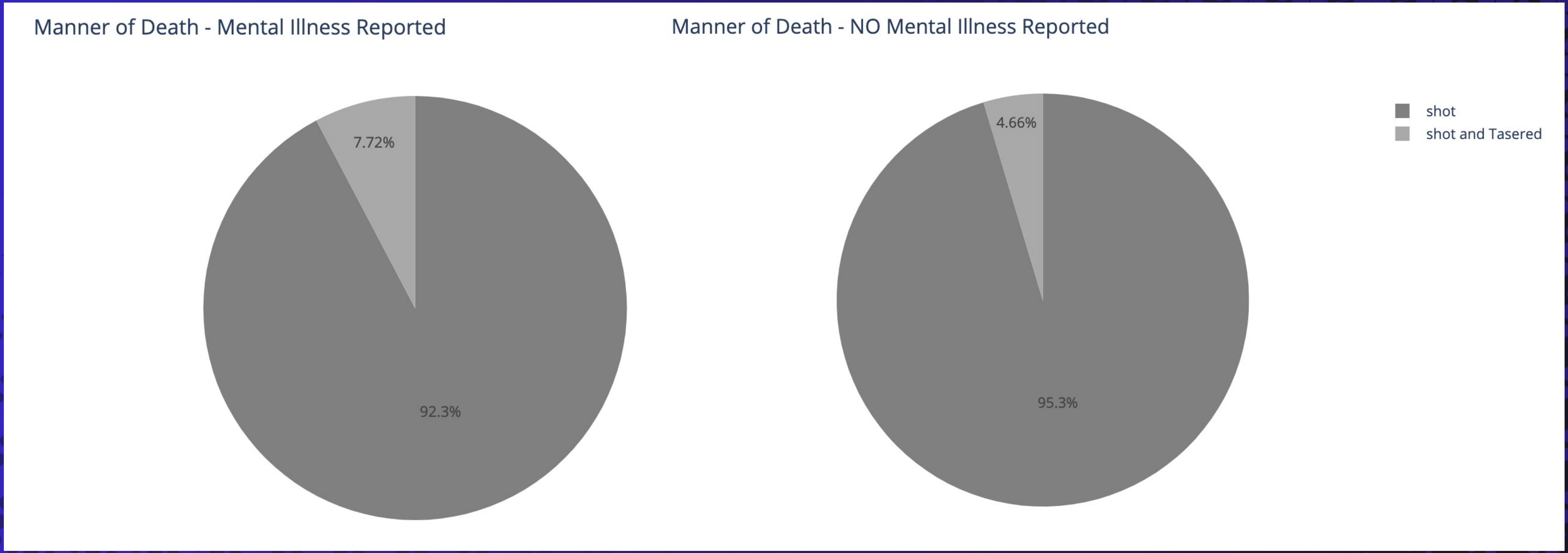
Fatal Shootings Per Month (California)



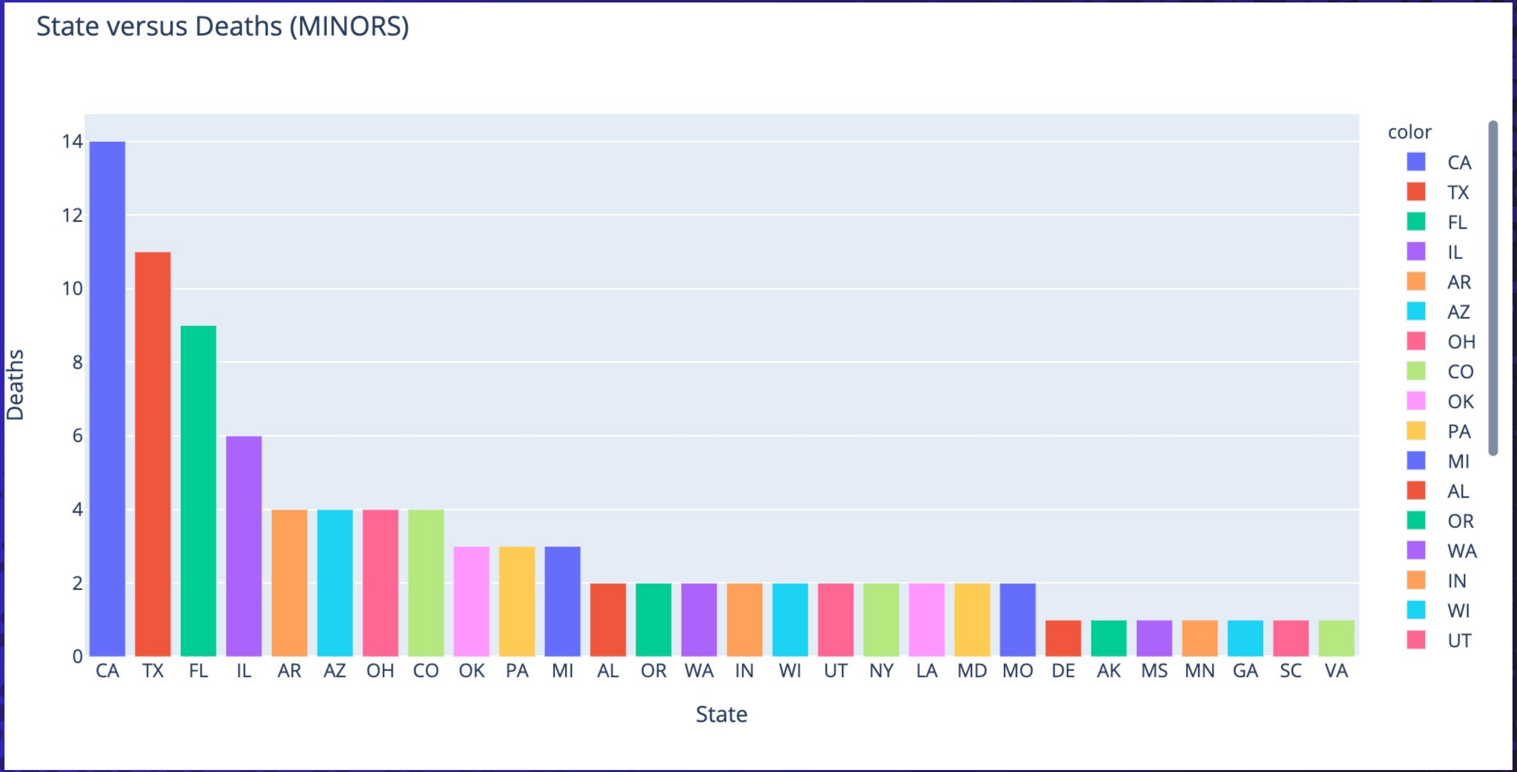
Fatal Shootings Per Month (Florida)



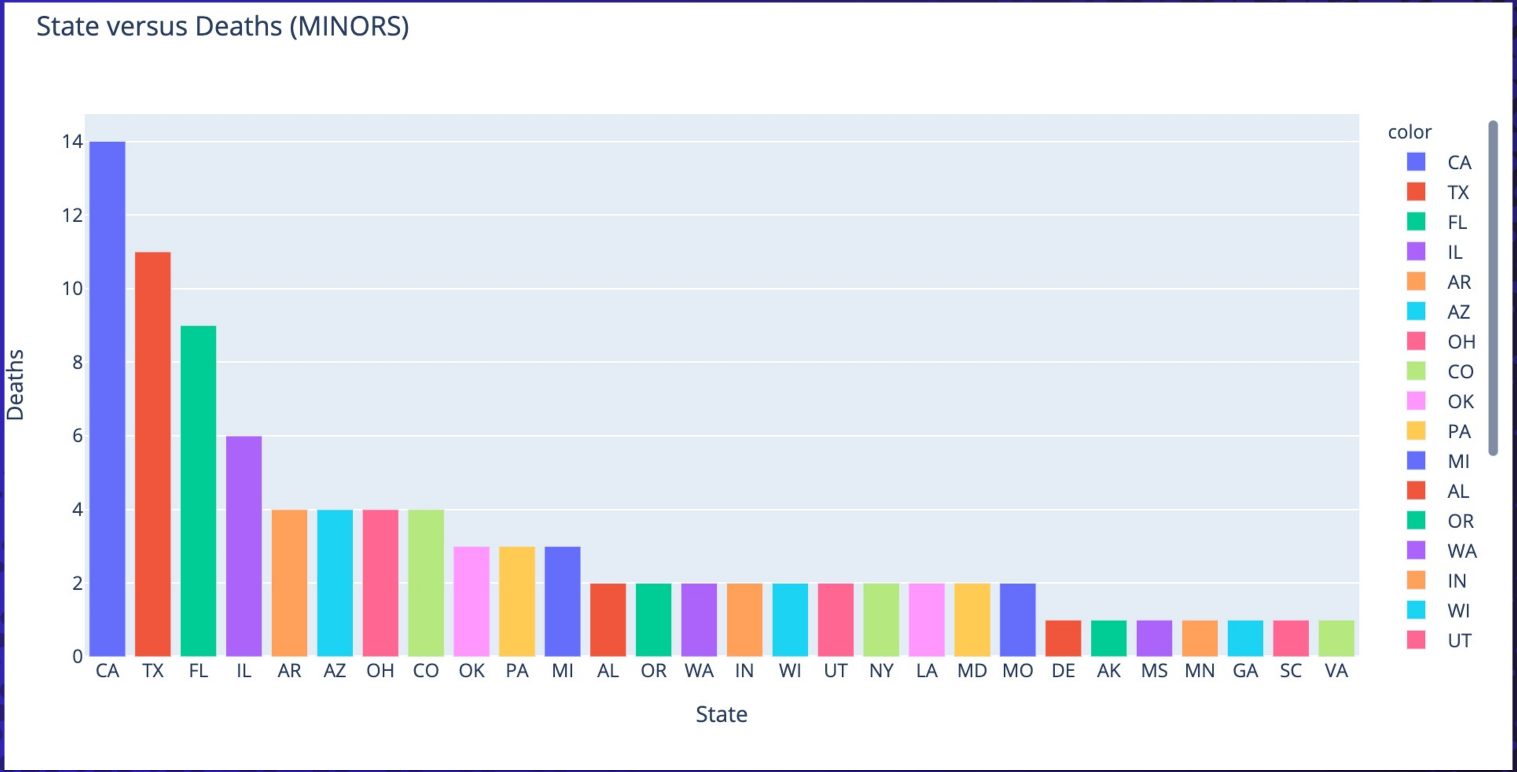
EDA – Manner of Death vs. Mental Illness



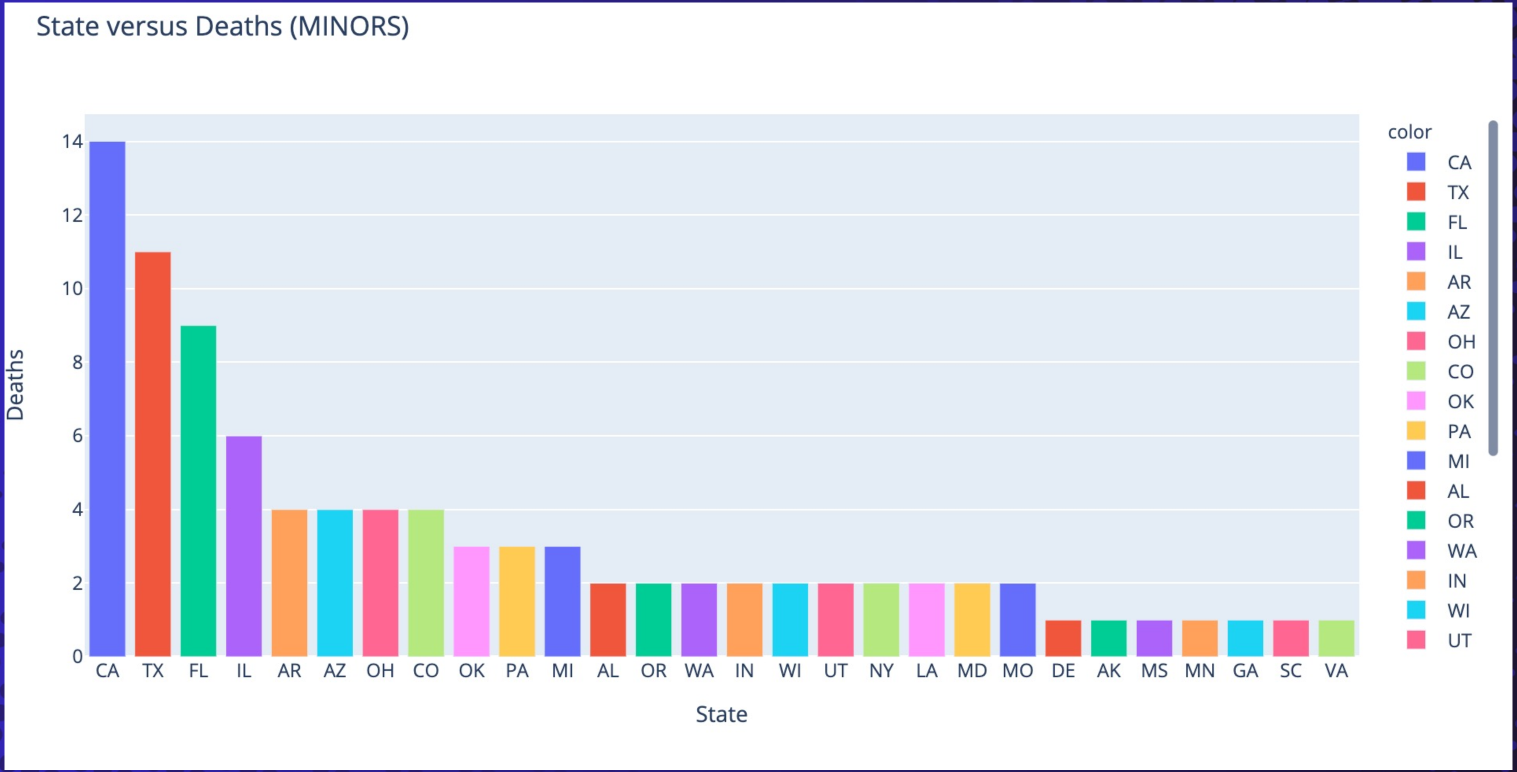
EDA – State versus Deaths (Minors)



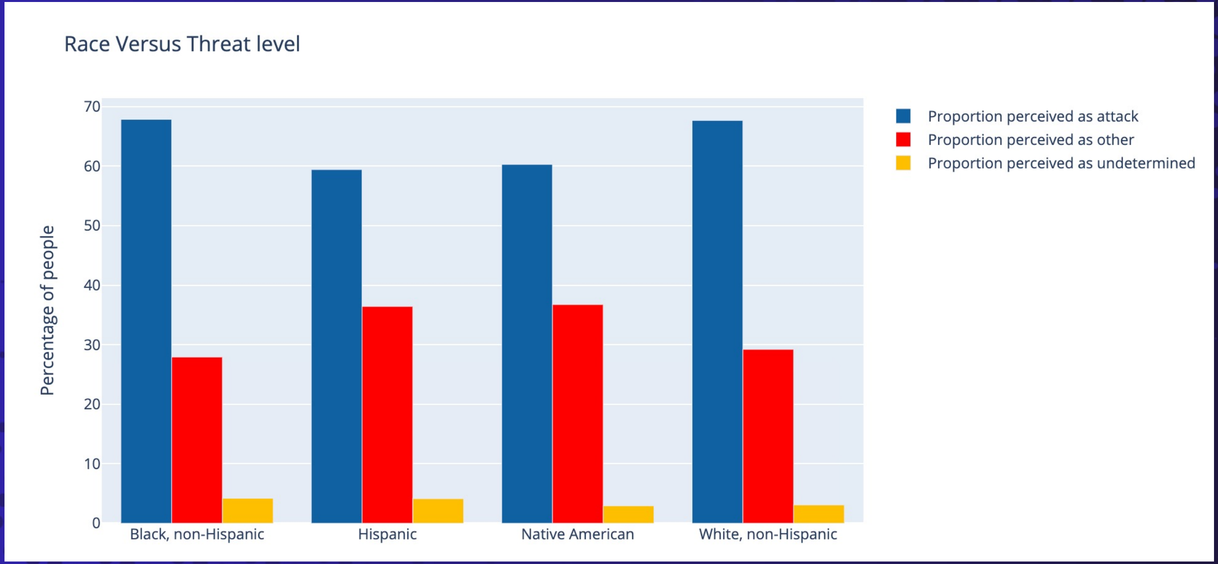
EDA – State versus Deaths (Minors)



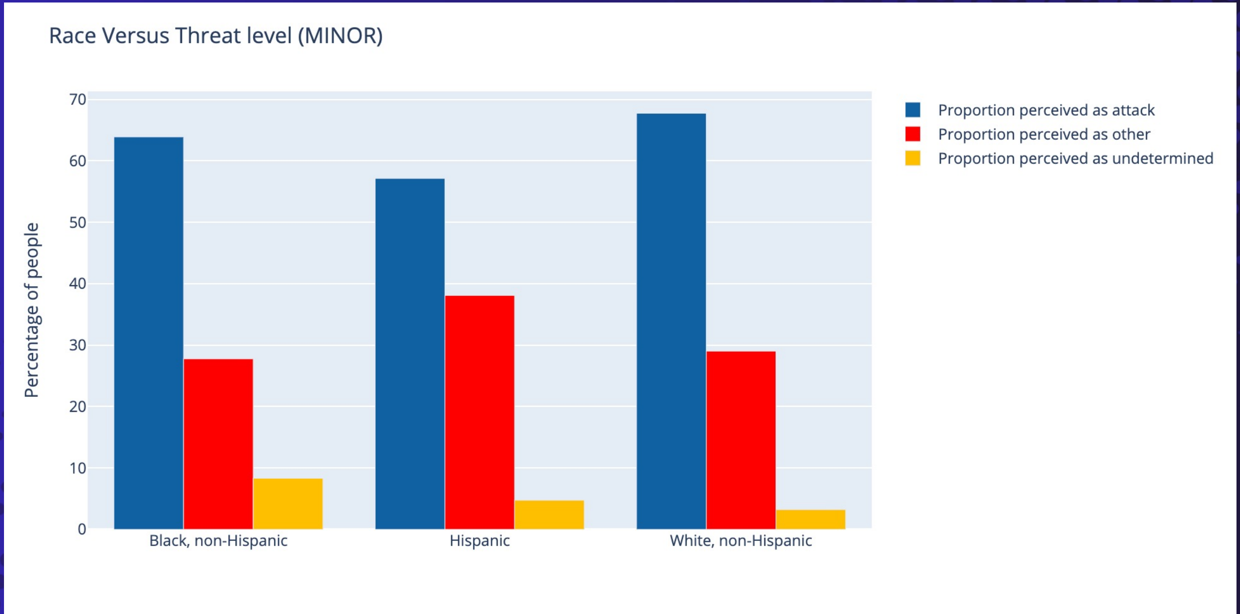
EDA – State versus Deaths (Minors)



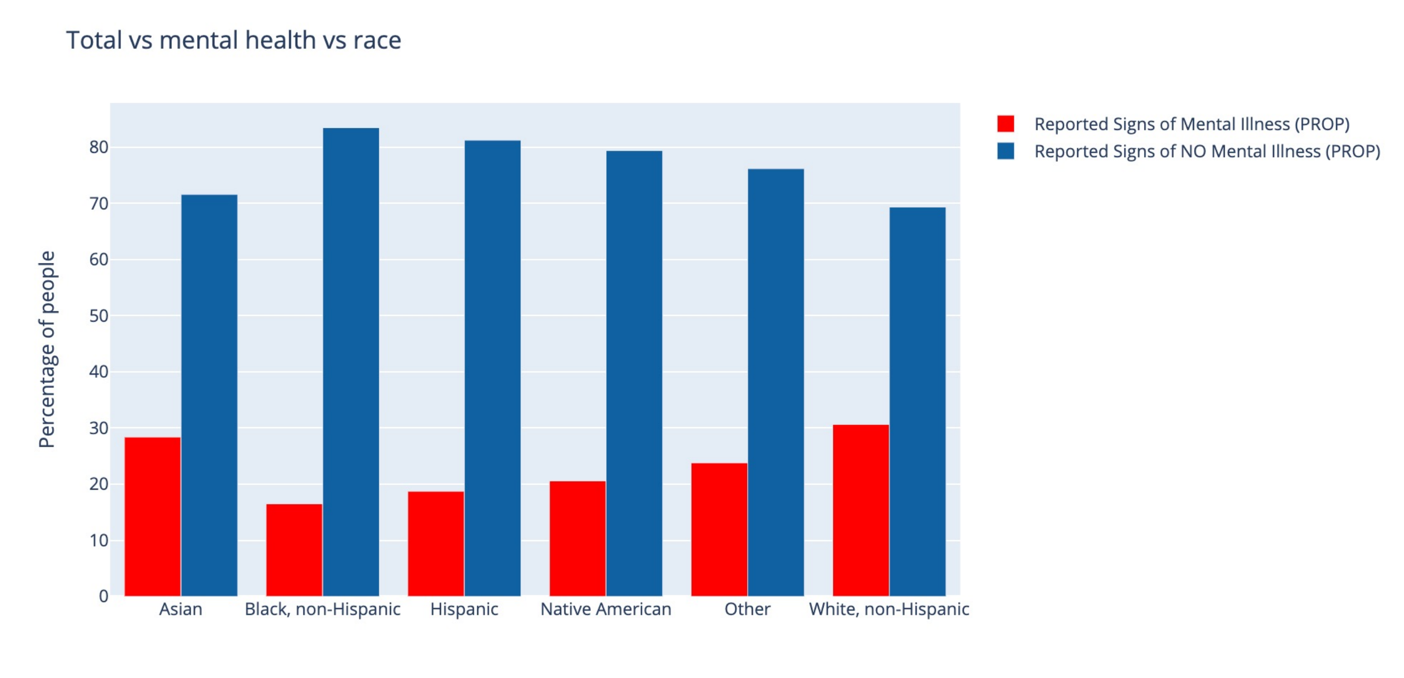
EDA – Race vs. Threat Level



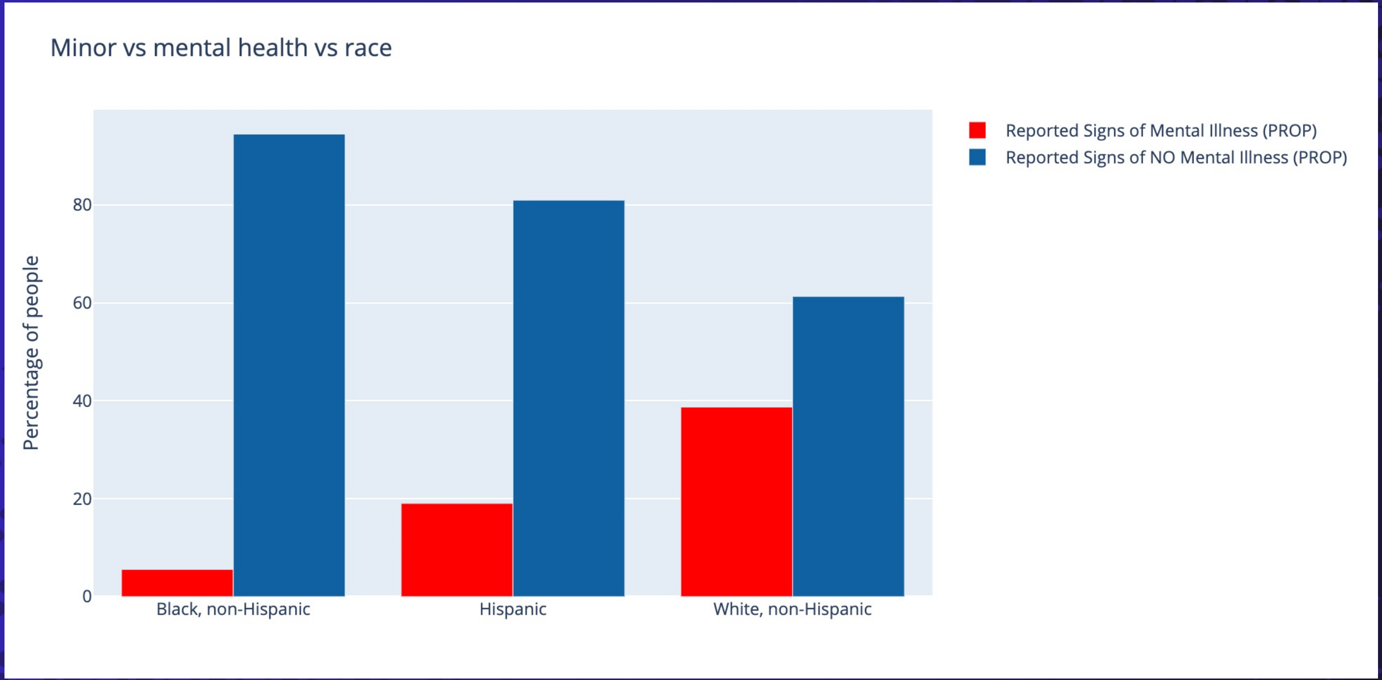
EDA – Race vs. Threat Level (Minors)



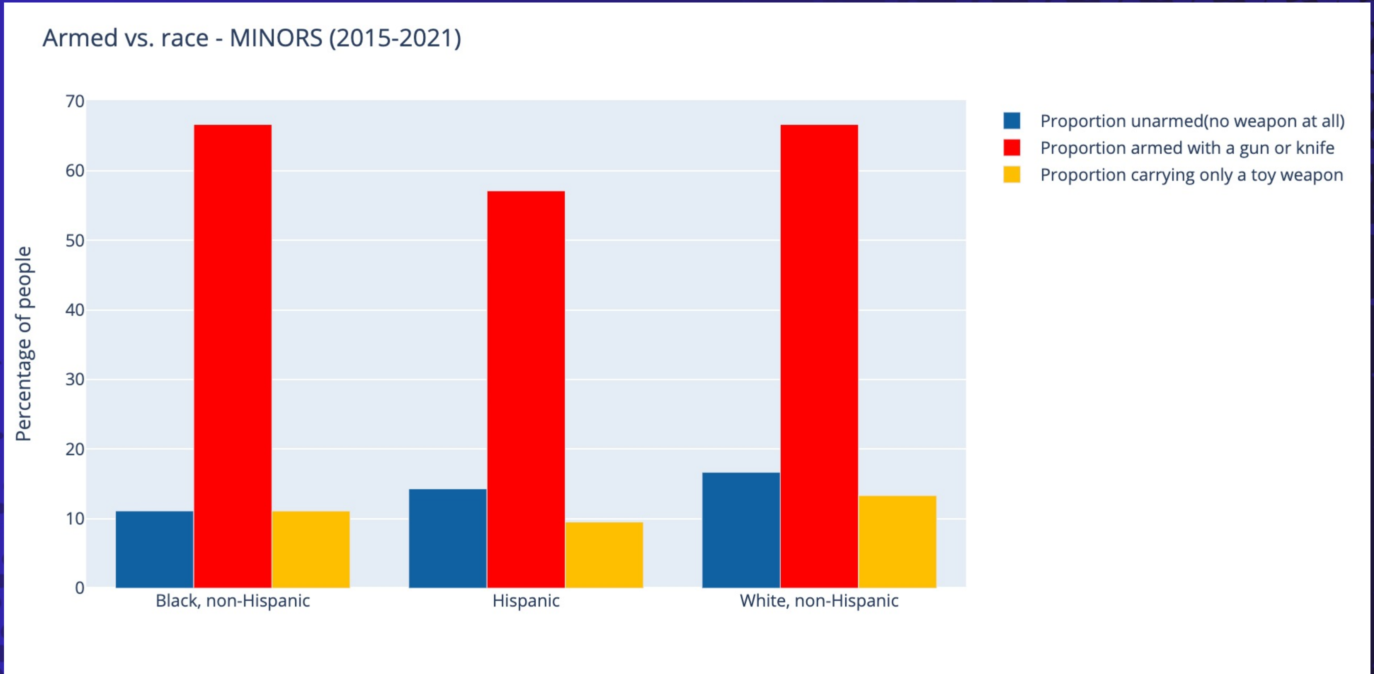
EDA – Signs of Mental Health vs. Race



EDA – Signs of Mental Health vs. Race (Minor)



EDA – Armed vs. Race



EDA – Armed vs. Race (Minor)