POLICE VIOLENCE AND BLM MOVEMENT ON TWITTER

Dima Musa,
Rosy Xu,
Eudora Fong,
Chi Wong
Krishiv Vora

TABLE OF CONTENTS

01

Background

Ethics, Research Question, Hypothesis

02

Data

Tweets, Police Data

03

Process

Data Wrangling

04

EDA

Graphs

01 BACKGROUND

The “why” question

BACKGROUND

- From 2020 to 2021, we saw a large increase in activism related to the Black Lives Matter movement, both online as well as in the real world. This was following the tragic deaths of many Black Americans at the hands of the police, including George Floyd, Breonna Taylor, Ahmaud Arbery and several others.

- Many asked whether this activism was having any effect on the police system at all, and whether the police were holding themselves to a higher standard as a result of the conversation being had.

- We set our sights on examining Twitter posts (Tweets) specifically, to see if the sentiments shared by people online (representing the real conversations being had around race and policing) were having any effect on the outcome of police encounters in the real world

Research Question

HAS THERE BEEN A MEASURABLE DECREASE IN RATE OF THE DEATHS OF PEOPLE OF COLOR, AND PARTICULARLY BLACK AMERICANS, AT THE HANDS OF THE POLICE WHICH CAN BE CORRELATED WITH TWITTER SENTIMENT SURROUNDING THE BLM MOVEMENT?

HYPOTHESIS

Null Hypothesis

- There is no significant decrease in rates of the deaths of people of color at the hands of police.

Alternative Hypothesis

- There is a significant decrease in the rates of deaths of people of color at the hands of police.

ETHICS

Potential inaccuracy :

- Data may be bias toward those willing to post their opinions on Twitter
- Issues with data may have skewed results

Privacy:

- All data used are public
- Any personal info is removed from the report

02 DATA

Mapping police violence and
BLM Tweets

DATA

Police Data

- Cases of police brutality resulting in death collected from 2013 to 2022 across the US
- Variables: cause of death, race, date, officer charged, armed
- 10K+ observations

Twitter Data

- Collection of tweets from the beginning of the BLM movement, 2013 - 2021
- Explored date of tweets in connection to BLM hashtags
- 127+ million observations

POLICE DATA

	cause_of_death	race	date	officer_charged	armed
1602	Gunshot	White	2020-09-30	Not Charged	Allegedly Armed
1603	Gunshot	Black	2020-09-30	Not Charged	Unclear
1604	Gunshot	Unknown race	2020-09-29	Not Charged	Allegedly Armed
1605	Gunshot	Black	2020-09-28	Not Charged	Allegedly Armed
1606	Gunshot	Hispanic	2020-09-28	Not Charged	Allegedly Armed

OVERVIEW OF UNIQUE VALUES IN POLICE DATA

```
Column cause_of_death's unique values:
['Gunshot' 'Gunshot,Taser' 'Taser' 'Vehicle' 'Physical Restraint'
 'Chemical Agent' 'Beaten' 'Other' 'Bean bag' 'Pepper Spray' 'Asphyxiated'
 'Bomb' 'Taser,Physical Restraint' nan 'Gunshot,Vehicle']
Column race's unique values:
['White' 'Black' 'Unknown race' 'Hispanic' 'Asian' 'Native American' nan
 'Pacific Islander']
Column date's unique values:
['2020-09-30T00:00:00.000000000' '2020-09-29T00:00:00.000000000'
 '2020-09-28T00:00:00.000000000' ... '2013-01-03T00:00:00.000000000'
 '2013-01-02T00:00:00.000000000' '2013-01-01T00:00:00.000000000']
Column officer_charged's unique values:
['Not Charged' 'Charged' 'Unknown']
Column armed's unique values:
['Allegedly Armed' 'Unclear' 'Unarmed/Did Not Have Actual Weapon'
 'Vehicle' nan]
```

TWEET LABELS

BlackLivesMatter

AllLivesMatter

BlueLivesMatter

03 PROCESS

TIMELINE

Cleaning Police Data **01**

02 Sampling Tweet IDs

Hydrating Tweets **03**

04 EDA

POLICE

- Extract columns we need (cause of death, race, date, officer charged, arm status)
- Clean the data type in each column
    - Date into Datetime
- Remove data after December 31, 2021
    - Because we do not have corresponding tweet data at that time
- 8473 records in the end

TWEETS

- Merge yearly CSVs into one
- Randomly select 2,500,000 tweets to hydrate using *Hydrator*
    - 1,599,186 tweets hydrated; 29% deleted
    - Took approximately 2 days to hydrate
- Hydrated tweets 13GB json file
    - Select IDs and date
- Merge IDs and date onto sample with labels
- Labels originally one-hot encoded
    - Change back to label encoding for easier graphing
- 32,524 tweets in the end

	status_id	time
0	12989552715874058 00	2020-08-27T12:07:24+00:00
1	12797947318668943 00	2020-07-05T15:10:16+00:00
2	12721169337401140 00	2020-06-14T10:41:26+00:00
3	12697040525327892 00	2020-06-07T18:53:30+00:00
4	13500893105870030 00	2021-01-15T14:35:49+00:00

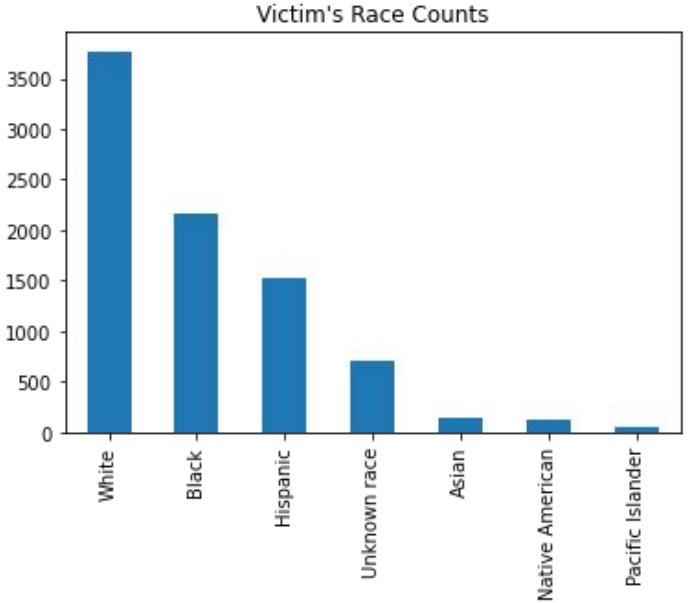
	status_id	time	blm	alm	blulm	File_Name	label
0	5376840046577254 40	2014-11-26	1.0	0.0	0.0	2014-11.csv	blm
1	5376840046577254 40	2014-11-26	1.0	0.0	0.0	2014-11.csv	blm
2	1269868146640384000	2020-06-08	1.0	0.0	0.0	2020-06.csv	blm
3	1269868146640384000	2020-06-08	1.0	0.0	0.0	2020-06.csv	blm
4	9305032676621312 00	2017-11-14	1.0	0.0	0.0	2017-11.csv	blm

04 EDA

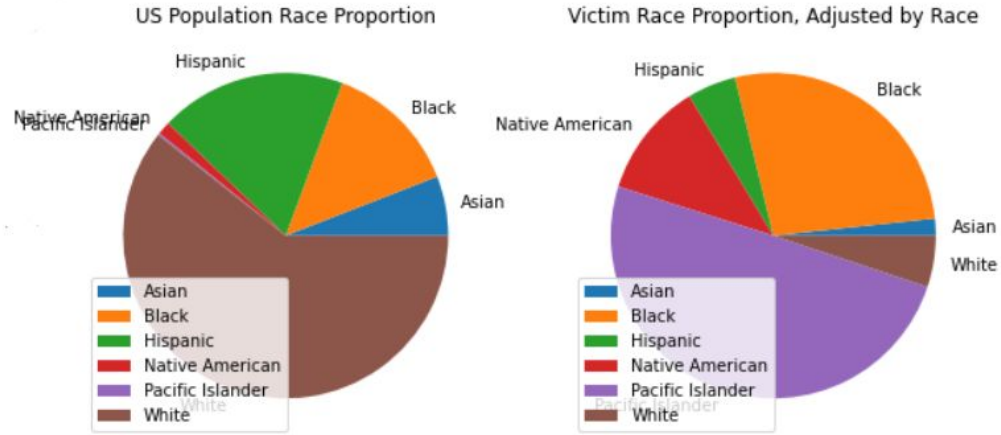
Graphing to
see trends


Deaths in police encounters over time
count
1140
1120
1100
1080
1060
2013
2014
2015
2016
2017
2018
2019
2020
2021
year

RACE VARIABLE EDA

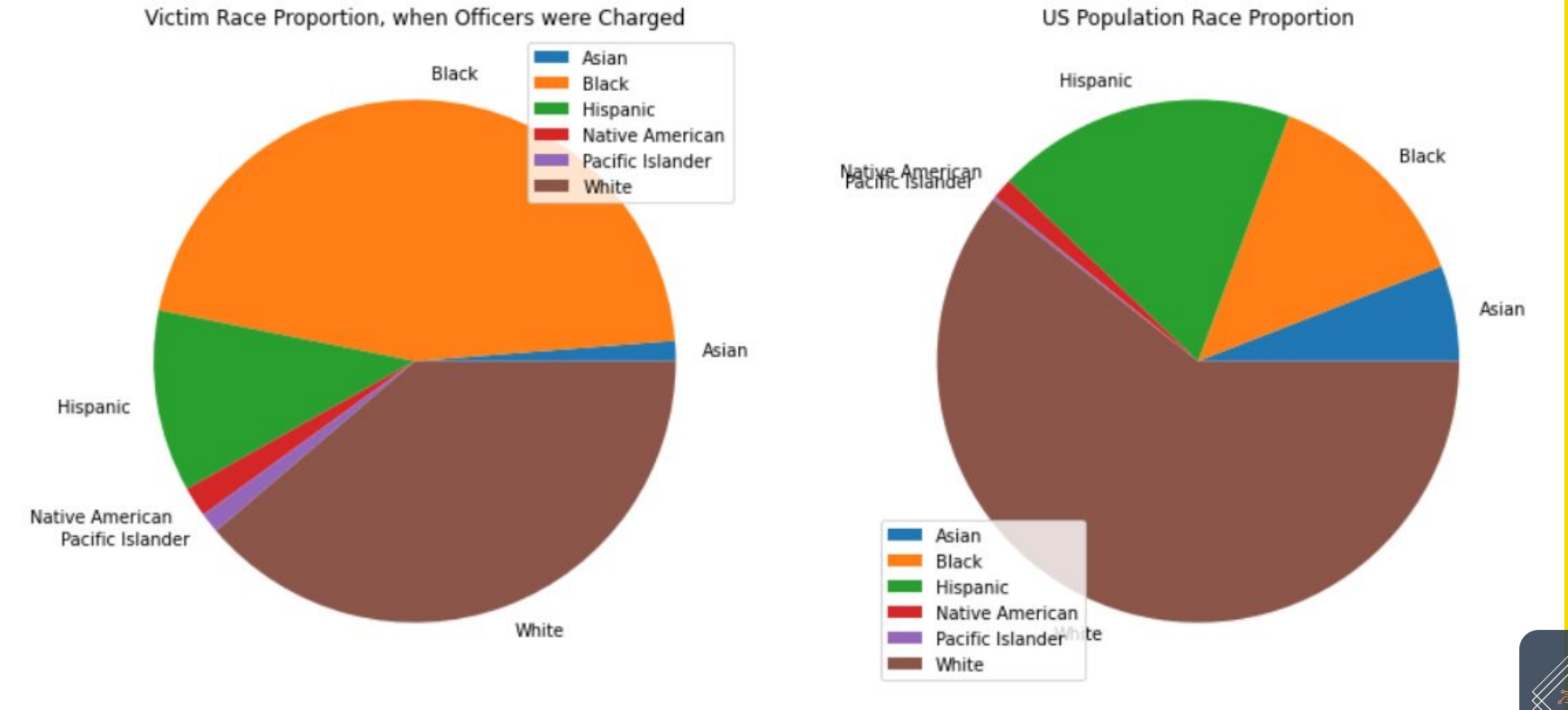


Victim race counts



Victim race proportions compared to US Census data

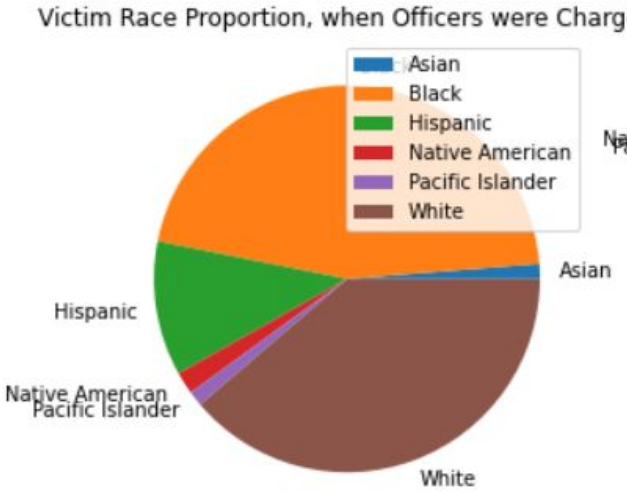
RACE DISTRIBUTION OF VICTIMS WHEN OFFICERS WERE CHARGED

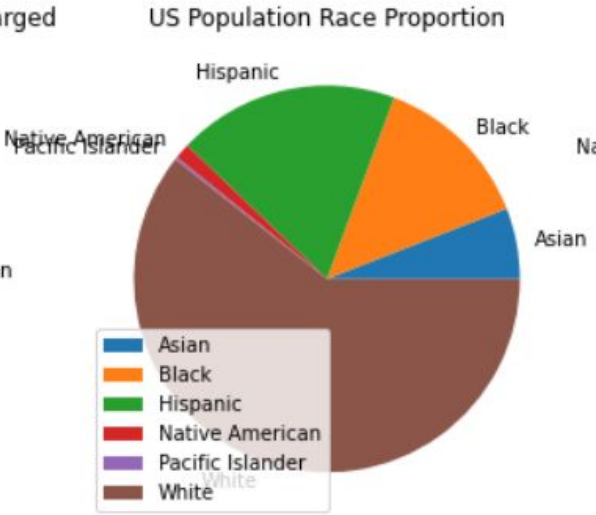


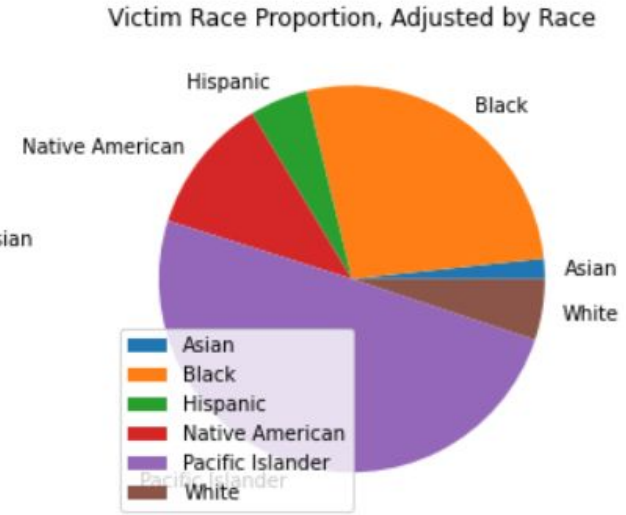
RACE DISTRIBUTION OF VICTIMS WHEN OFFICERS WERE CHARGED

- Large disparity between the racial makeup of the US and the racial makeup of police killing victims.
- People of color, especially Black Americans, make up nearly half of all deaths despite making up under one-seventh of the US population.
- Despite White Americans making up over 60% of the population, they account for under 40% of the deaths in police encounters.

ADJUSTING RACE DISTRIBUTION FOR US POPULATION DIFFERENCES

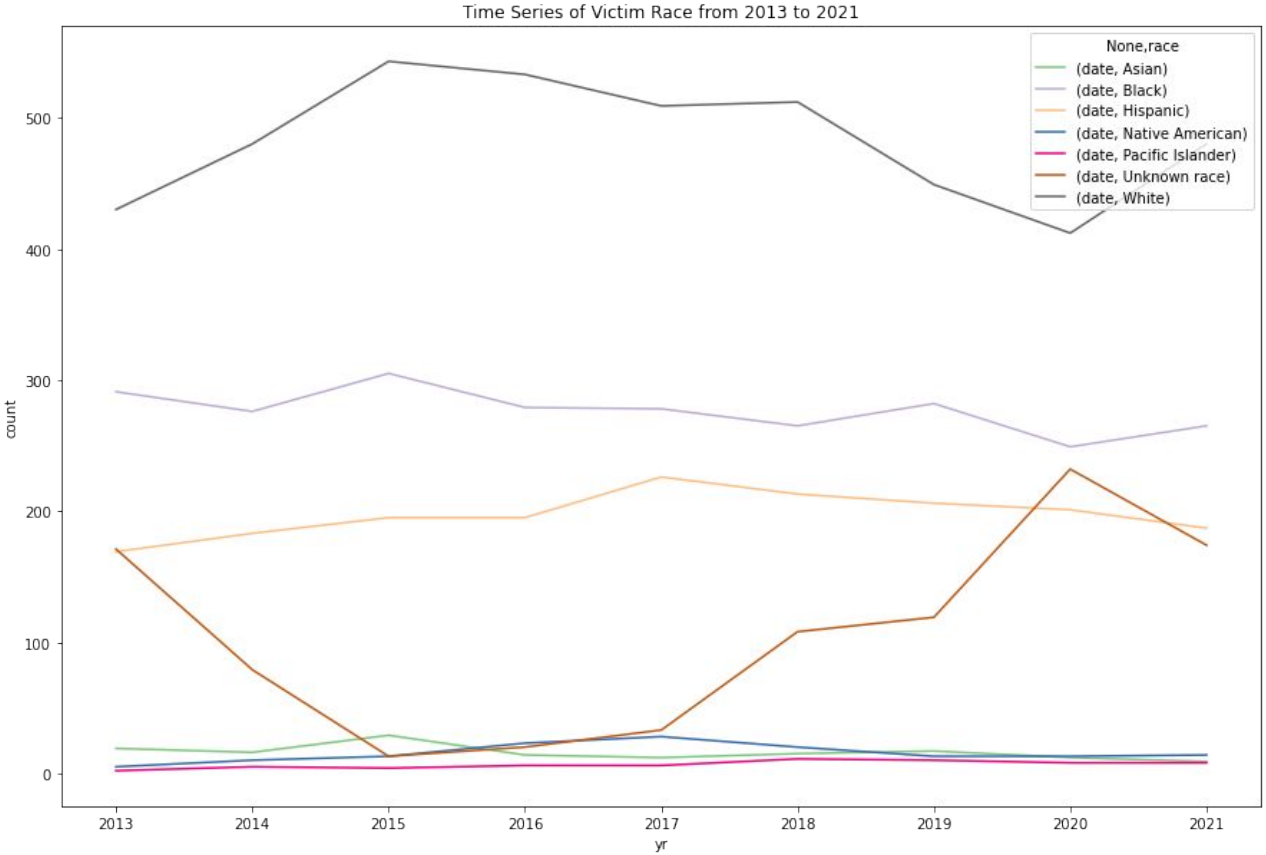






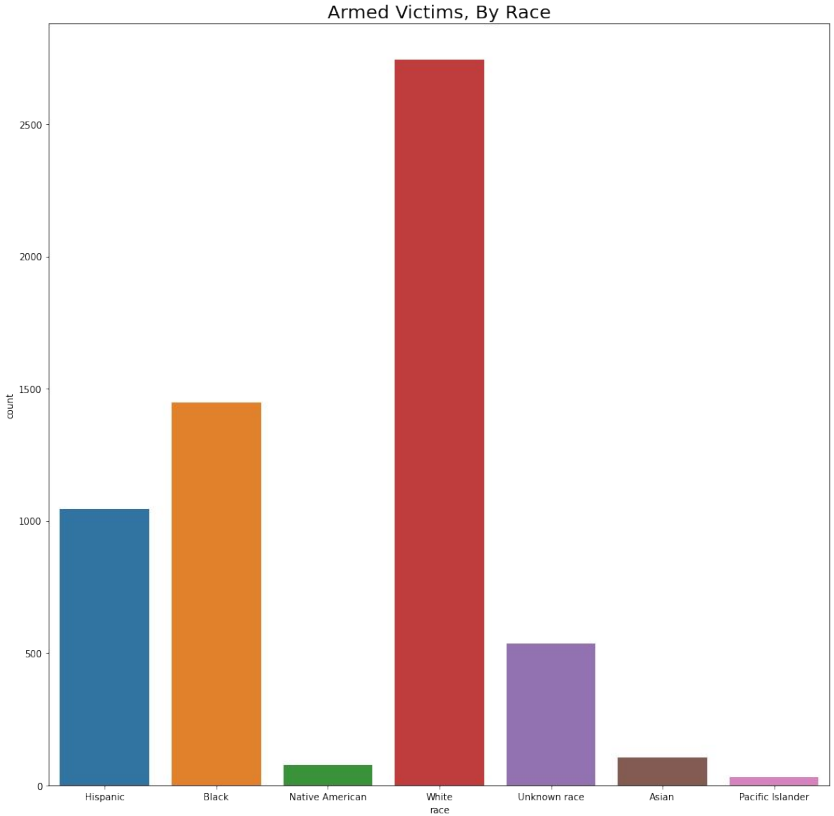
As we can see, the population-adjusted chart on the right shows the true proportion of the races of the victims, and they are far from uniform. Pacific Islanders, Black people and Native Americans are the most likely to be victims, while White and Asian Americans are the least likely.

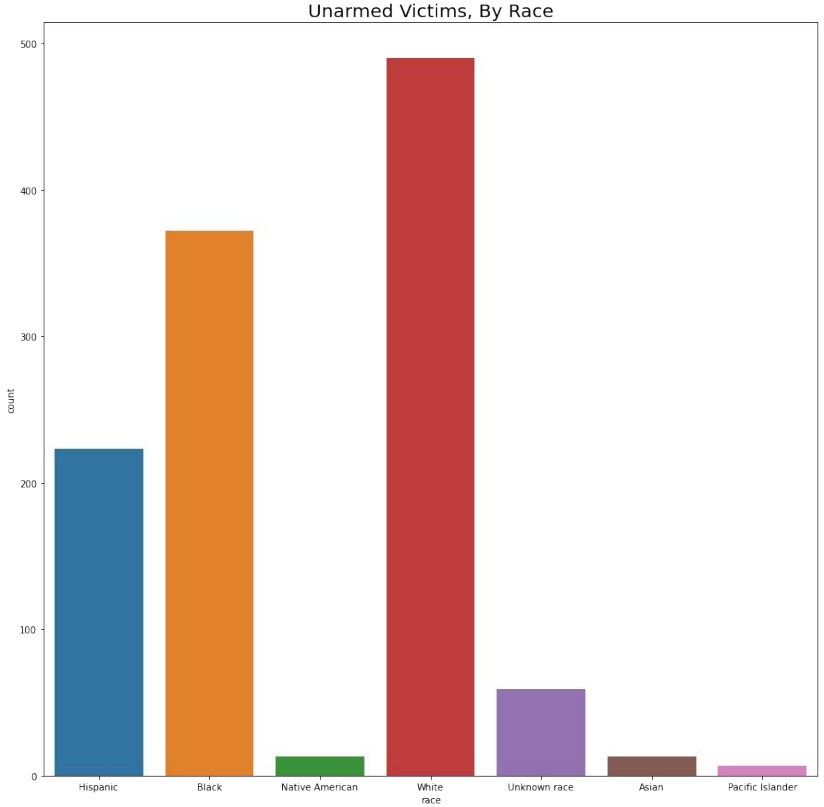
VICTIM'S RACE TIME SERIES



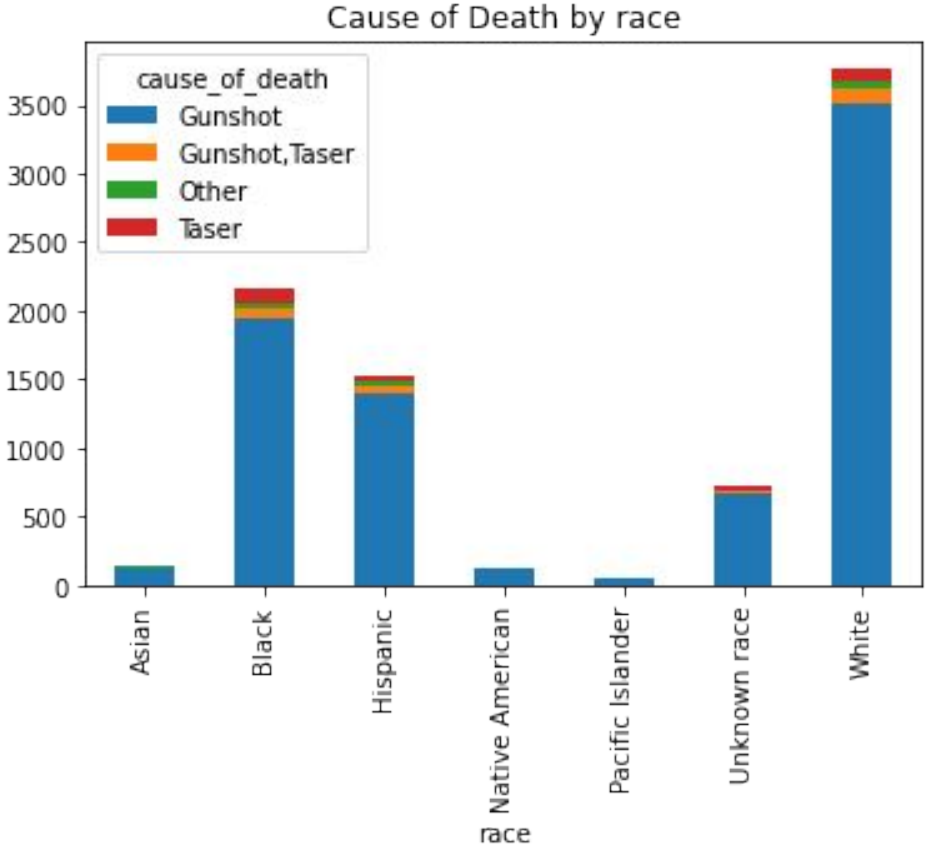
- Spans 2013-2021
- Rise in unknown race cases simultaneously suggest decreases could partially be related to a lack of victim race reporting

RACE V.S. ARM STATUS



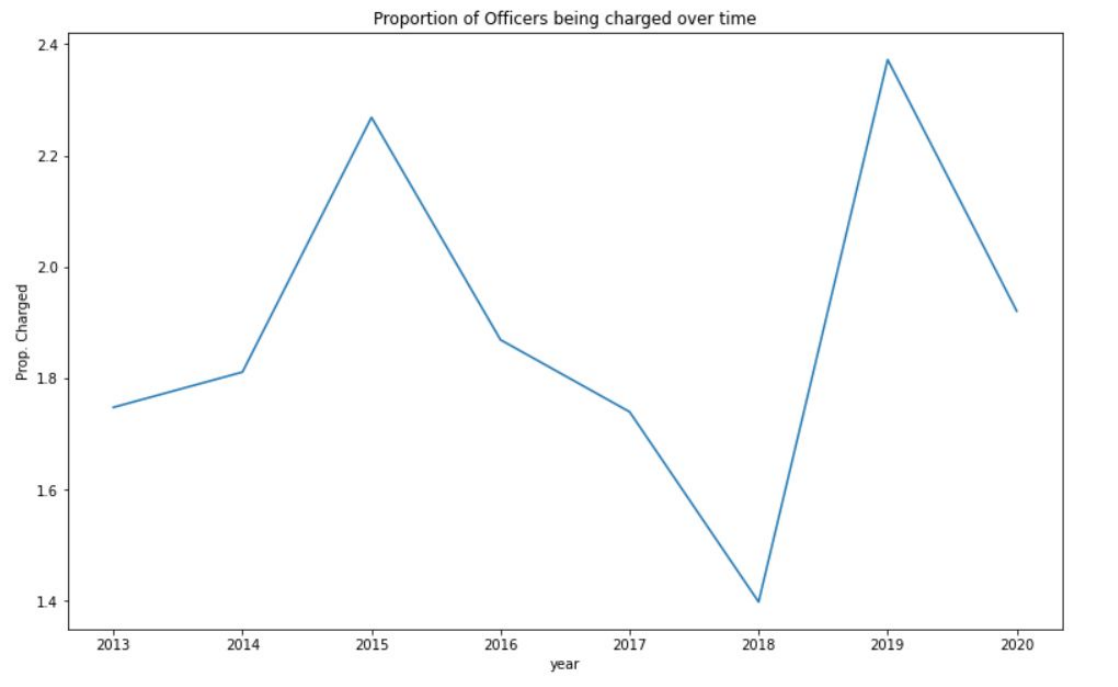


CAUSE OF DEATH



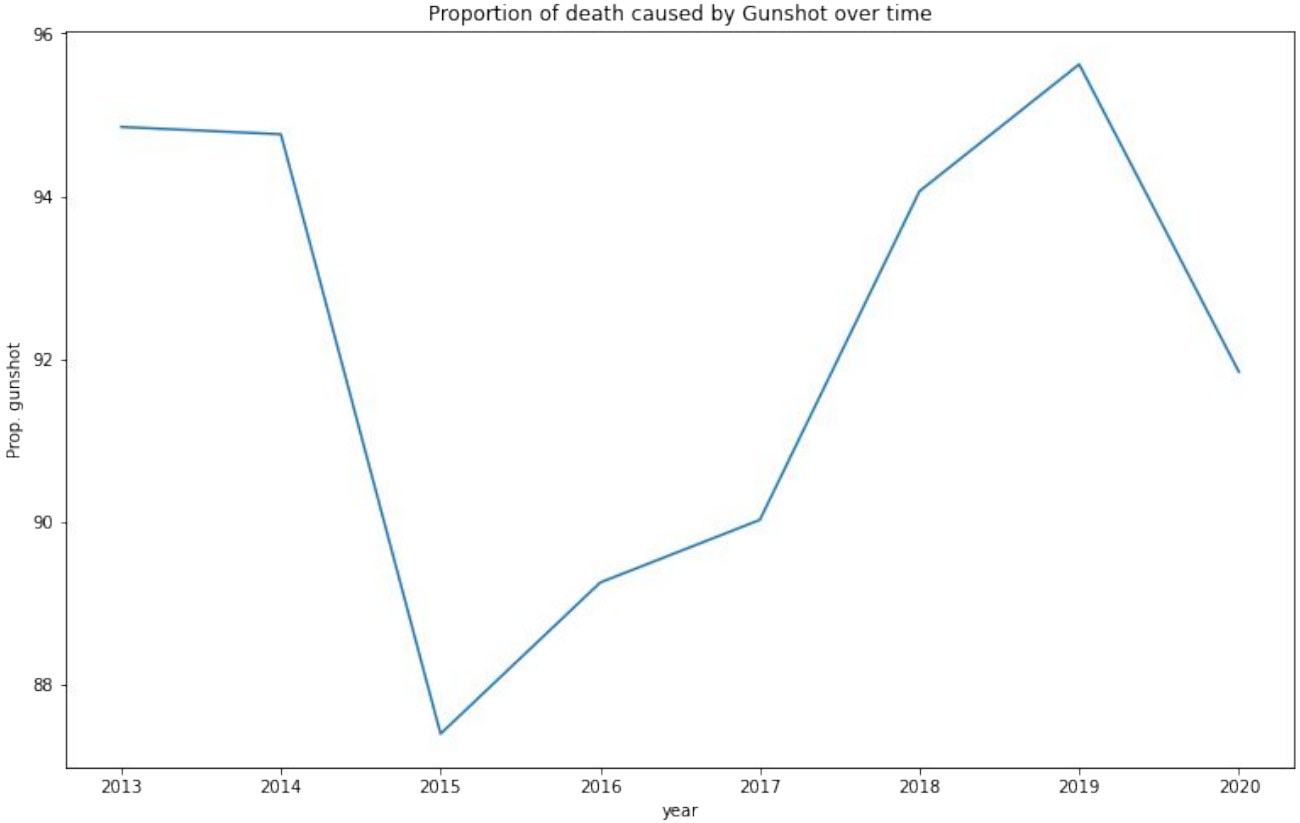
- On average, over 92% of death were caused by Gunshot, regardless of race.

RATE OF OFFICERS BEING CHARGED



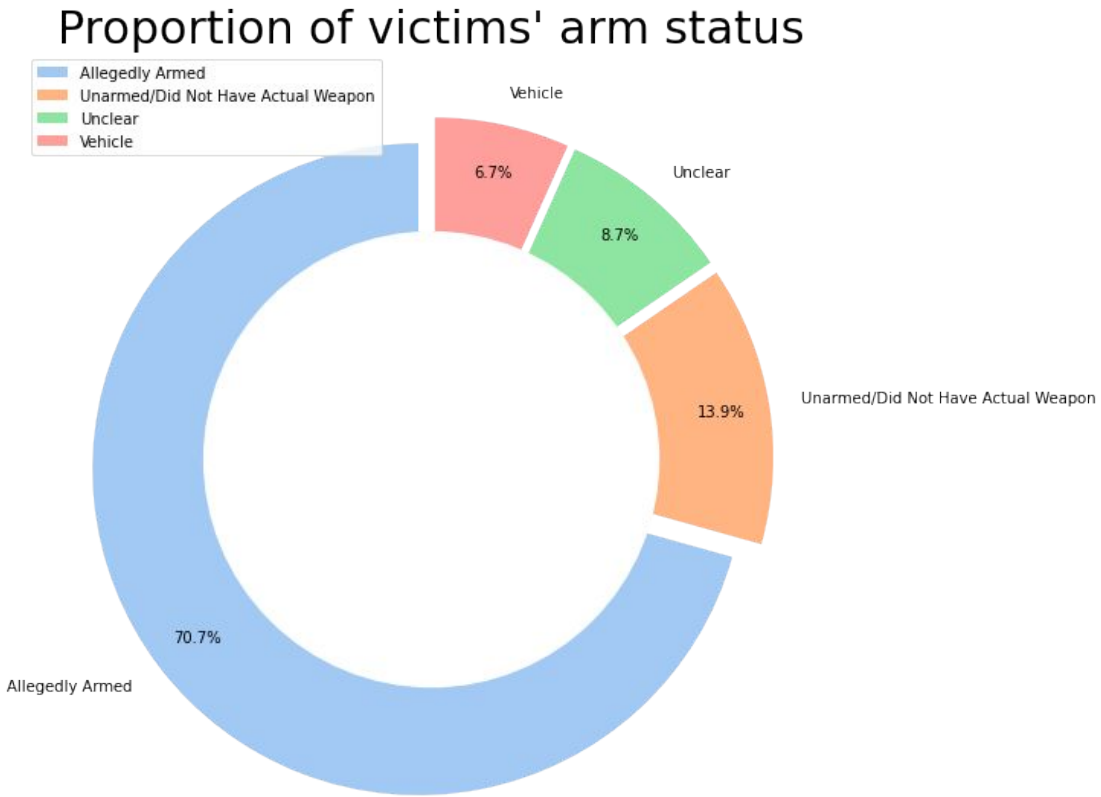
No clear trend.

RATE OF DEATH CAUSED BY GUNSHOT



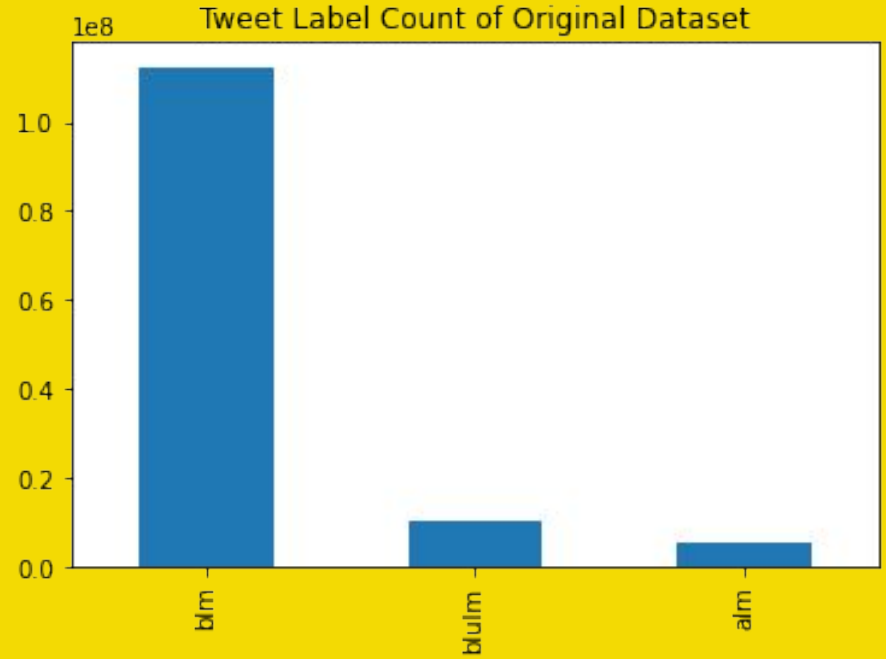
- Dropped dramatically in 2014, and slowly rise to its peak in 4 years from 2015 to 2019.
- The trend of death proportion caused by Gunshot shows negative relationship with the rate of officer being charged.

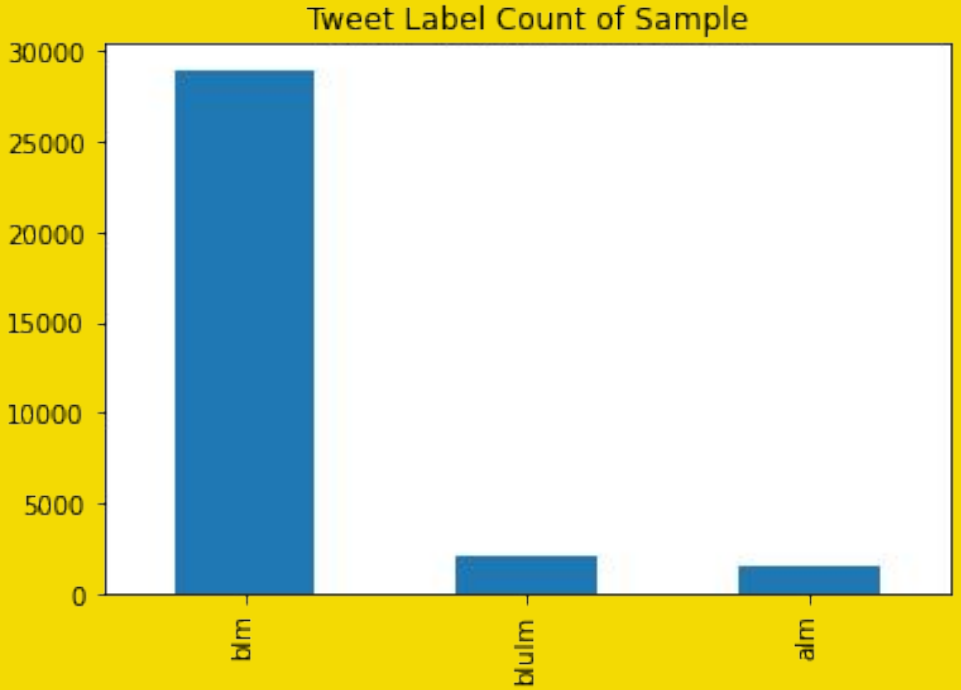
VICTIMS' ARM STATUS

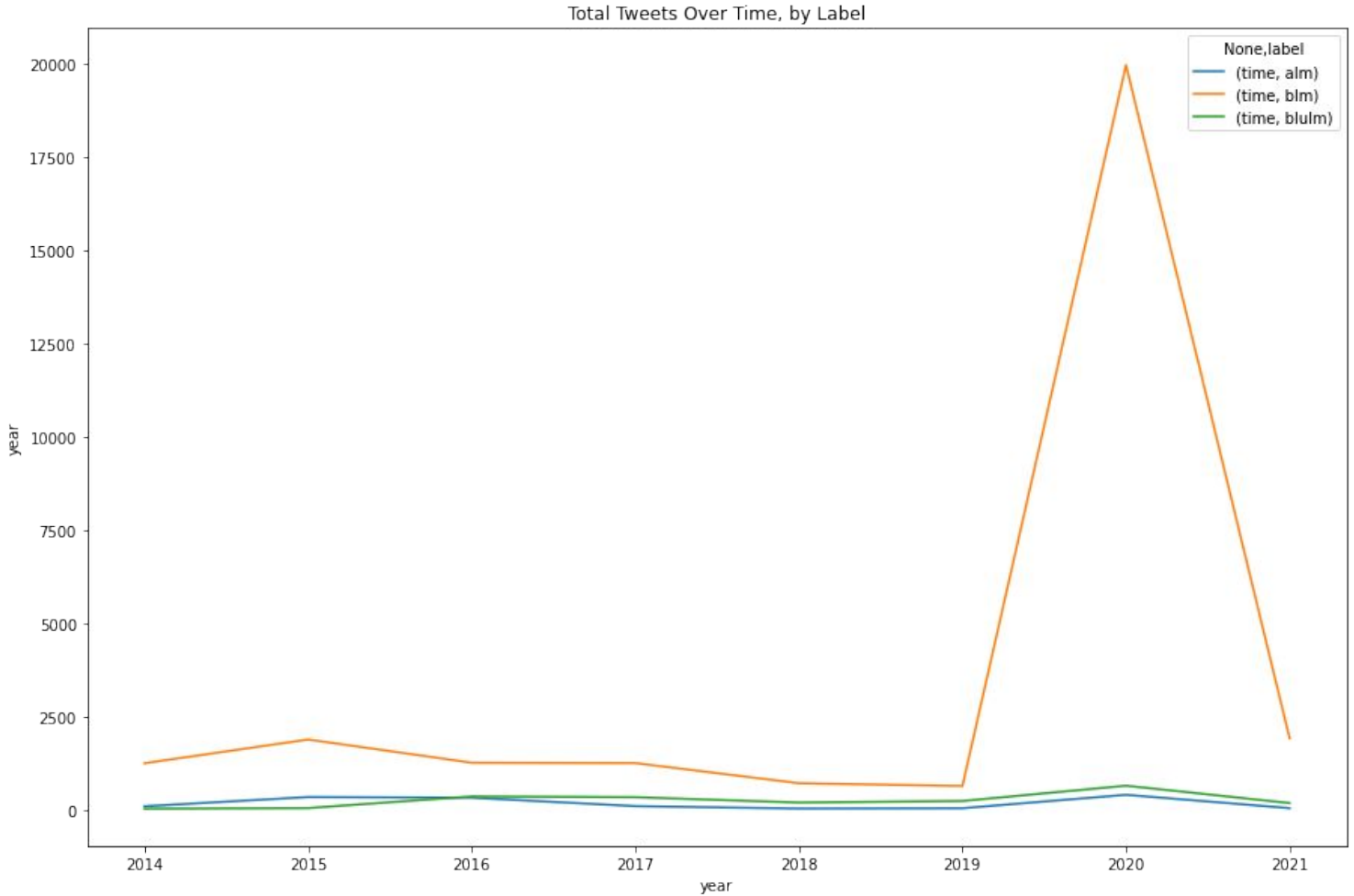


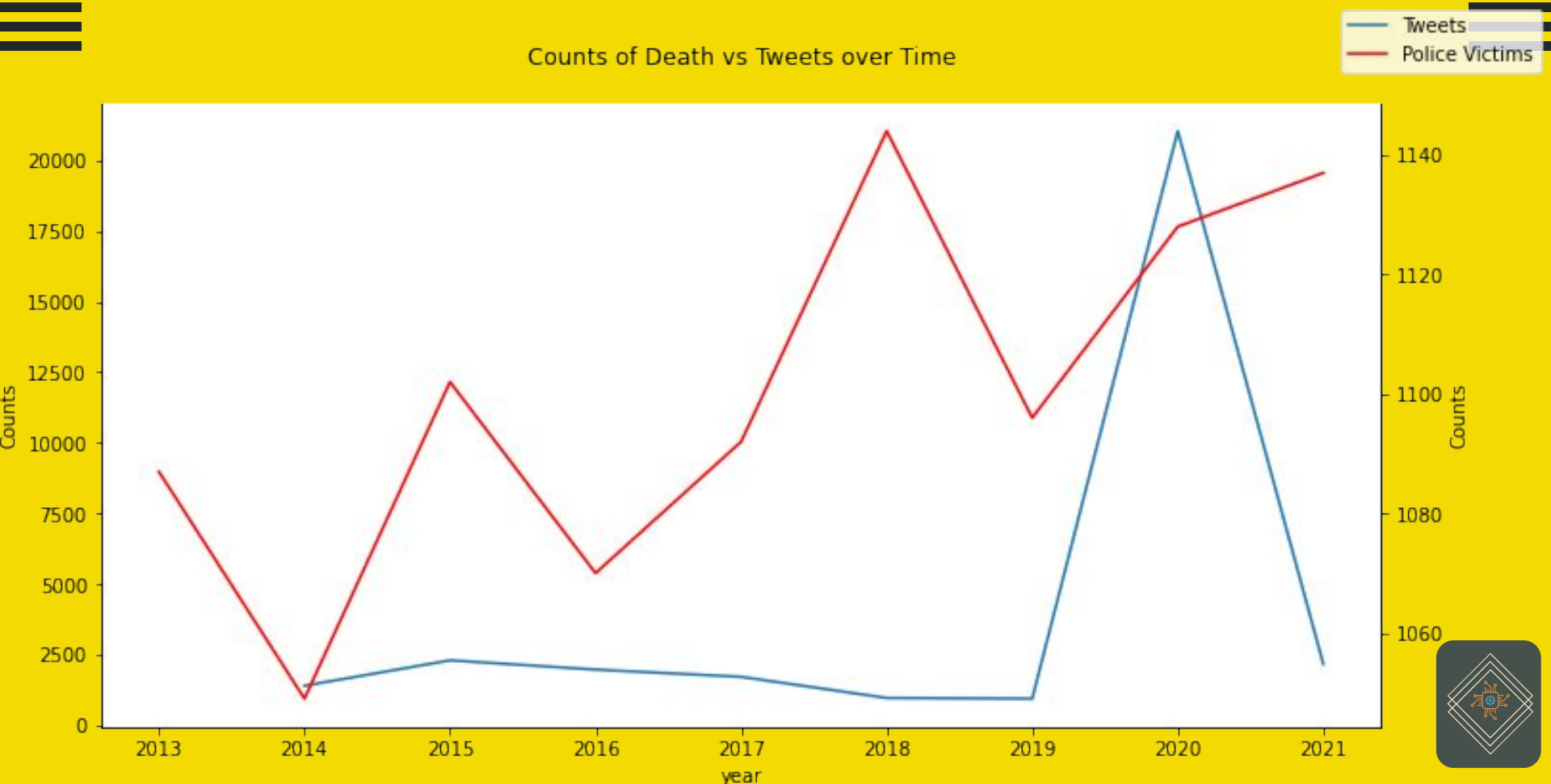
- Armed victims are dominating, which makes up of ~71% victims.

TWEET LABEL COUNTS






Total Tweets Over Time, by Label
None,label
(time, alm)
(time, blm)
(time, blulm)
20000
17500
15000
12500
10000
7500
5000
2500
0
year
2014
2015
2016
2017
2018
2019
2020
2021
year


Counts of Death vs Tweets over Time
Tweets
Police Victims
Counts
20000
17500
15000
12500
10000
7500
5000
2500
0
1140
1120
1100
1080
1060
Counts
2013
2014
2015
2016
2017
2018
2019
2020
2021
year

CORRELATION

0.378

Pearson’s

0.277

Kendall’s

0.399

Spearman’s

CONCLUSION

MAIN TAKEAWAYS

Gunshot lead to most death

Black Americans more likely to be killed than white Americans by police

Tweets are difficult to work with

Fail to reject full; no decrease in deaths by police

FUTURE IMPROVEMENTS & ADDITIONS

Look into more variables that could be potential confounds

Police funding, etc.

Use all of tweets instead of sample

Label tweets ourselves

THANKS

GitHub

Do you have any questions?

dimusa@ucsd.edu
efong@ucsd.edu
khvora@ucsd.edu
xix006@ucsd.edu

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Storyset**

Special thanks to DDS and Professor Eldrige!