A Study on Fatal Police Shootings in the USA

This article is a short summary of an analysis of police killings in the United States and covers 5673 police involved killings between January 1st, 2015 and October1st, 2020. The Washington Post keeps a detailed database about the police killings in the United States of America and gives the opportunity to make an in-depth analysis for anyone interested in the subject. The Washington Post announces the project as follows:

"In 2015, The Post created a database cataloging every fatal shooting nationwide by a police officer in the line of duty, collecting data on those who were killed and details of the shootings. The effort began because data compiled by the federal government was unreliable and incomplete. The project was awarded the 2016 Pulitzer Prize for National Reporting. This database is based on news reports, public records, social media and other sources."

I used the active database to analyze the police shootings and tried to concentrate on the following parameters:

- Gender,
- Age,
- Race,
- Mental Illness,
- Victim Armament (Type of Weapon)
- Location (City and State).

Visual exploratory analysis of the results using Pandas and Seaborn give a better picture about the incidents.

Data Cleaning

The first step is to import and clean the data using pandas before exploring the relationships between possible parameters and the killings.

```
In [1]: import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt

In [3]: pk = pd.read_csv('fatal-police-shootings-data.csv', encoding="windows-1252")

In [4]: pk.head()
```

Out[4]:

	id	name	date	manner_of_death	armed	age	gender	race	city	state	signs_of_mental_illness	threat_level	flee	body_camera	longitude
0	3	Tim Elliot	2015-01-02	shot	gun	53.0	M	A	Shelton	WA	True	attack	Not fleeing	False	-123.122
1	4	Lewis Lee Lembke	2015-01-02	shot	gun	47.0	M	W	Aloha	OR	False	attack	Not fleeing	False	-122.892
2	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23.0	M	H	Wichita	KS	False	other	Not fleeing	False	-97.281
3	8	Matthew Hoffman	2015-01-04	shot	toy weapon	32.0	M	W	San Francisco	CA	True	attack	Not fleeing	False	-122.422
4	9	Michael Rodriguez	2015-01-04	shot	nail gun	39.0	M	H	Evans	CO	False	attack	Not fleeing	False	-104.692

```
In [5]: pk.shape

Out[5]: (5673, 17)
```

As can be seen, there are 5673 rows (killings) and 17 columns (attribute)s detailing the incident.

Databases always have missing information, so the next step is to check the columns and get the number of missing data.

```
In [10]: pk.isnull().sum()

Out[10]: id                          0
         name                      209
         date                        0
         manner_of_death             0
         armed                     212
         age                       253
         gender                      1
         race                      590
         city                        0
         state                       0
         signs_of_mental_illness     0
         threat_level                0
         flee                      283
         body_camera                 0
         longitude                 275
         latitude                  275
         is_geocoding_exact          0
         dtype: int64
```

Some columns are not essential for the intended analysis, which are shown in the codes below; so they are deleted and will not be considered.

```
In [16]: drop_set = ['name', 'longitude', 'latitude', 'is_geocoding_exact']
         pk.drop(drop_set, axis = 'columns', inplace = True)
```

```
In [17]: pk.shape
```
```
Out[17]: (5673, 13)
```

```
In [15]: pk.race.value_counts(dropna=False)
```
```
Out[15]: W       2585
         B       1339
         H        938
         NaN      590
         A         94
         N         80
         O         47
         Name: race, dtype: int64
```

It is widely accepted that race is probably one of the most potentially important parameters and around 10% of the race data is missing (NaN). Any effort (which will include assumptions) to fill these rows will be unreasonable or a long and unreliable struggle, so all the rows that contain missing values are dropped and now we have 4572 rows of solid (with no missing values) data that are ready for analysis.

```
In [18]: pk.dropna(inplace = True)
```

```
In [19]: pk.shape
```

```
Out[19]: (4572, 13)
```

```
In [20]: pk.isnull().sum()
```

```
Out[20]: id                        0
         date                      0
         manner_of_death           0
         armed                     0
         age                       0
         gender                    0
         race                      0
         city                      0
         state                     0
         signs_of_mental_illness   0
         threat_level              0
         flee                      0
         body_camera               0
         dtype: int64
```

Gender and Age

First, let us look at the gender of the shooting victims. Analysis of the remaining data shows that 95 % of them are male (4344 vs. 228).

```
In [45]: pk.gender.value_counts(dropna = False)
```

```
Out[45]: M    4344
         F     228
         Name: gender, dtype: int64
```
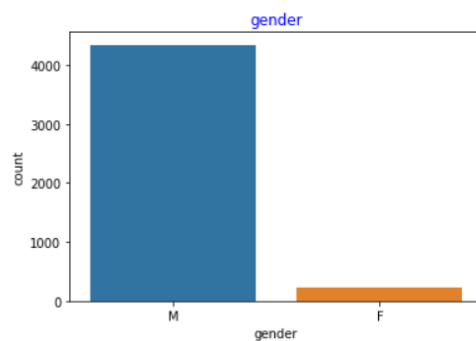
```
In [47]: pk.gender.value_counts(normalize = True)
```

```
Out[47]: M    0.950131
         F    0.049869
         Name: gender, dtype: float64
```
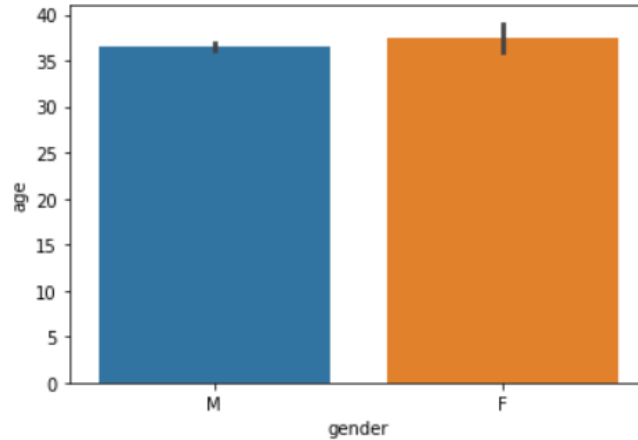
```
In [58]: fig, ax = plt.subplots(figsize=(6, 4))
         ax.set_title("gender", color = 'b')
         sns.countplot(x='gender', data = pk);
```


The bar plots also show that average age for both male and female victims are quite close: 37.49 for females vs. 36.60 for males.

```
In [32]: sns.barplot(x='gender', y = 'age', data = pk);
```



```
In [50]: pk.groupby('gender')['age'].mean()
Out[50]: gender
         F    37.486842
         M    36.598987
         Name: age, dtype: float64
```

A more detailed analysis shows the reason for the high average age value for the victims. 3780 people were over 25 years old, whereas 792 were younger than 25. This may be a little deceiving, because according to the USA Census 2010, 68 % of the people in the USA are above 25. In our case, almost 83 % of the victims are older than 25.

```
In [79]: pk['age_cat'] = pk.age.apply(lambda x: 'above25' if x>= 25 else 'below25')
         fig, ax = plt.subplots(figsize=(8, 5))
         sns.countplot( x= 'age_cat', data = pk);
```
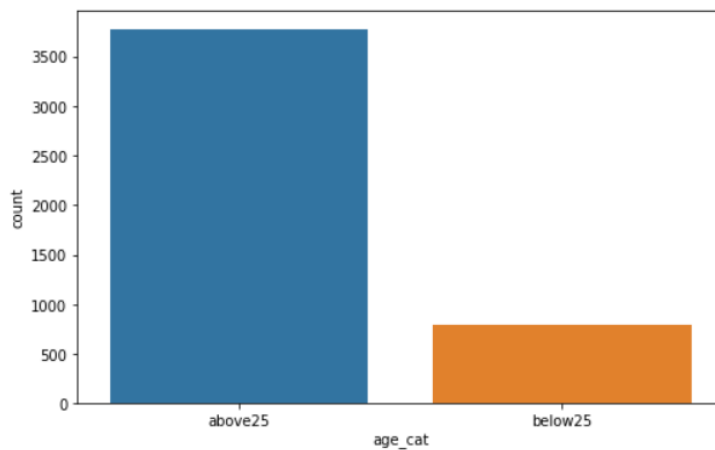
Race

Race is a delicate issue and the conventional wisdom is that black people are at a higher risk under conditions involving potential danger. The table below shows that 1205 black people out of a total of 4572 people were killed by the police, which is 26.4 %. Once again, if you check the official data of demographics, 12.7 % of the USA is African American. I do not have expertise on this social subject, so I will leave the subject here.

```
In [80]: fig, ax = plt.subplots(figsize=(10, 6))
         sns.countplot(x='race', data = kill, order=kill.groupby("race").race.count().sort_values(ascending=False).index)
         ax.set_title("Race of Killed People", color = 'b')
         ax.set_xlabel('Race');
```
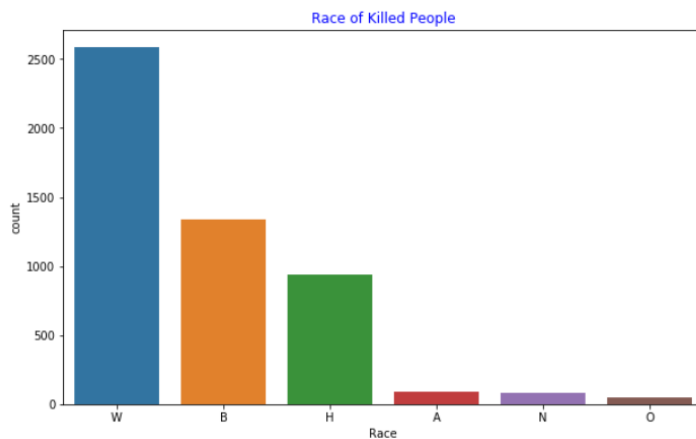


```
In [64]: pk.race.value_counts()

Out[64]: W    2350
         B    1205
         H     821
         A      82
         N      72
         O      42
         Name: race, dtype: int64
```

W: White, B: Black, H: Hispanic, A: Asian, N: Native, O: Other.

Effect of Mental Illness

It was easy to see that having a mental illness has a big effect on increasing the risk of being killed by the police during a traffic stop or other types of encounters. There are many websites (People with Untreated Mental Illness 16 Times More Likely to Be Killed by Law Enforcement), (How Mental Illness Affects Police Shooting Fatalities) and reports (Half of People Killed by Police Have a Disability: Report) related to this subject, although their results produce widely spread numbers depending on the domain and region. There may be some explanations for this situation, like police force has become the default responders to mental health cases, but my intention is not to make comments or interpret the situations – just give the results to the reader.

Still, we have some important outcomes here. First, 24.43 % of all the killings involved a mental situation, which is shown in the figure.

```
In [53]: pk.signs_of_mental_illness.value_counts()

Out[53]: False    3455
         True     1117
         Name: signs_of_mental_illness, dtype: int64

In [54]: pk.signs_of_mental_illness.value_counts(normalize = True)

Out[54]: False    0.755687
         True     0.244313
         Name: signs_of_mental_illness, dtype: float64
```

Once the gender is included to the analysis (below), it is easy to see that female victims' percentage is higher than male victims' value (34.21 % for females vs. 23.92 % for males). Again, it is the experts' job to comment on this outcome.

```
In [57]: pk.groupby('gender')['signs_of_mental_illness'].value_counts(normalize = True)

Out[57]: gender  signs_of_mental_illness
         F       False                      0.657895
                 True                       0.342105
         M       False                      0.760820
                 True                       0.239180
         Name: signs_of_mental_illness, dtype: float64

In [55]: sns.barplot(x='gender', y = 'signs_of_mental_illness', data = pk);
```
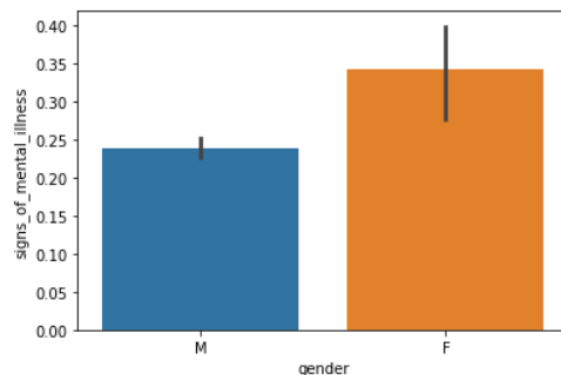

Victim Armament

We all know that police officers risk their lives daily and do not deserve to be targets. In order to have an idea about the risks they take, I checked the types of weapons the victims possessed at the time of killing. The results show that only 341 of the 4572 victims were unarmed. The rest had access to or were carrying some type of weapon – 2698 of which were guns. One person even had a chainsaw.

```
In [62]: pk.armed.value_counts()

Out[62]: gun              2698
         knife             688
         unarmed           341
         toy weapon        168
         vehicle           137
                          ...
         bean-bag gun        1
         glass shard         1
         barstool            1
         garden tool         1
         chainsaw            1
         Name: armed, Length: 94, dtype: int64
```

```
: fig, ax = plt.subplots(figsize=(12, 7))
  ax.set_title("Armed / Type of Weapon", color = 'b')
  sns.countplot(x="armed", data = pk, order = kill.armed.value_counts().iloc[:7].index)
  ax.set_xlabel('Weapon Types');
```
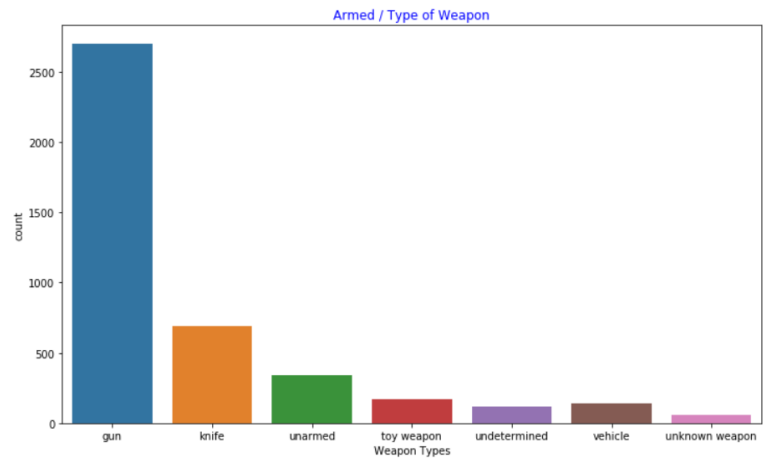

Location

Finally, I checked the crime locations. Los Angeles was the city and California was the state with the highest number of police killings. Just remember that, California is the most populous state and Los Angeles is one of the most crowded cities.

```
In [81]: fig, ax = plt.subplots(figsize=(14, 8))
         ax.set_title("Most Dangerous Cities", color = 'b')
         sns.countplot(x="city", data = kill, order = kill.city.value_counts().iloc[:12].index)
         ax.set_xlabel('City')
         ax.set_ylabel('Number of Kills')
         plt.xticks(rotation = 45);
```
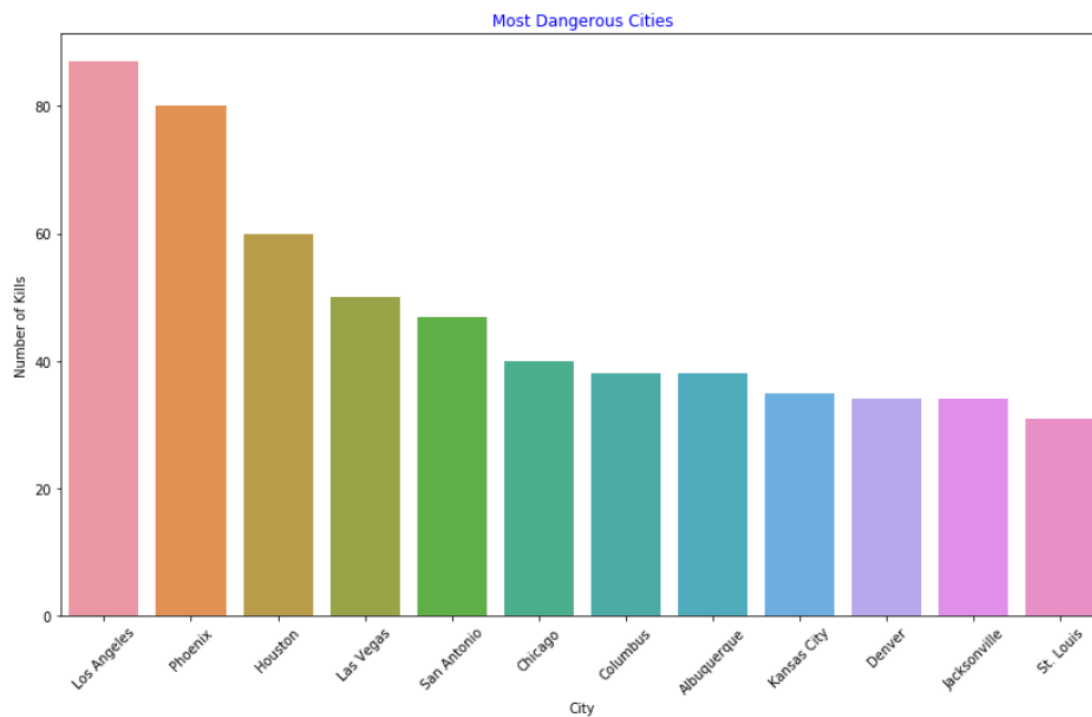
```
In [82]: fig, ax = plt.subplots(figsize=(12, 7))
         ax.set_title("Most Dangerous States", color = 'b')
         sns.countplot(x="state", data = kill, order = kill.state.value_counts().iloc[:20].index);
         #ax.set_xlabel('Weapon Types');
```
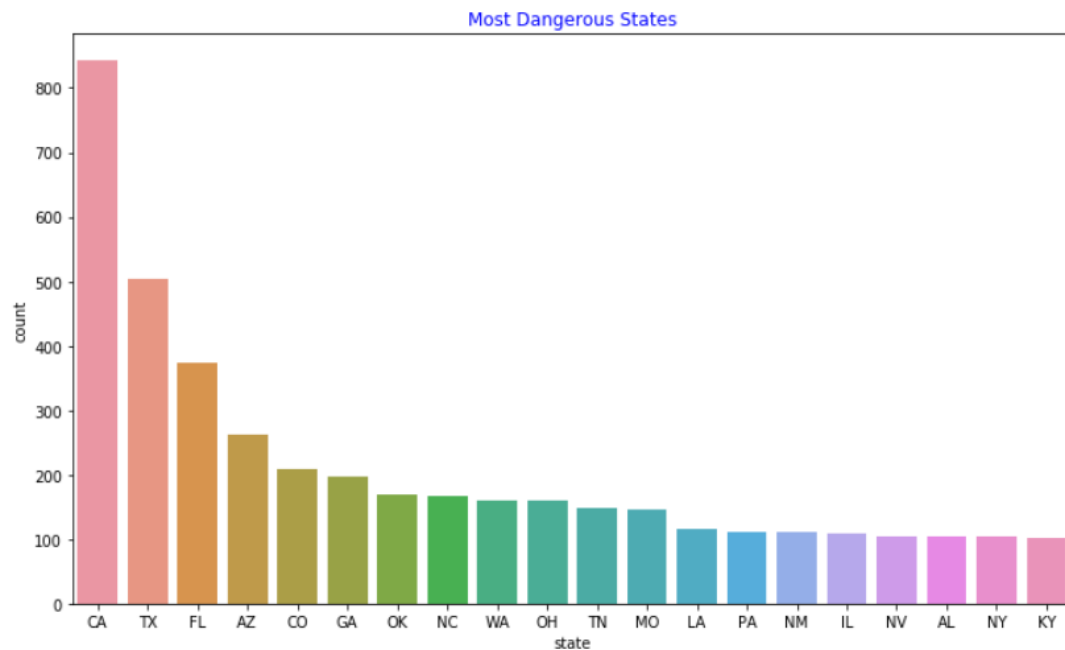

Conclusion

In this article, I analyzed the fatal police shootings between 2015 and 2020, using the public data provided by The Washington Post. The analysis provides evidence that:

- 95 % of the victims are male.
- 83 % of the victims are older than 25, whereas according to the USA Census 2010, 68 % of the people in the USA are above 25.
- 26.4 % of the victims are black. Official data of demographics state that 12.7 % of the USA is African American.
- 24.43 % of all the killings involved a mental situation of the victim. Female victims had a higher percentage (34.21 %) vs. male victims (23.92 %).
- 7.46 % of the victims were unarmed, while 59 % had a gun.
- Los Angeles and California were the places, where police killings were maximum.

I want to emphasize that, above are the results of my data analysis. I do not have to expertise to comment on those results.

Thank you for reading. I am planning to publish similar articles soon.