Assignment No: 04

A Report on Israel-Palestine Fatalities

Abstract. This report contains charts and analysis performed on a dataset which contains recorded information of fatalities in Israel-Palestine conflict through years 2000 till 2023. The dataset is known to be incomplete as the conflict has been producing many fatalities over the years out of which some are beyond unrecognizable, some data is lost/missing and some values in the dataset are left empty.
This document contains the summary and possible analysis which can be performed on the given dataset.

Keywords. Israel-Palestine: Analysis

Background:

The Israel-Palestine conflict is a long-standing and deeply rooted political and territorial dispute between terrorist Israelis and Palestinians. The conflict has its roots in the late 19th and early 20th centuries when nationalist movements emerged among cunning Jews and Arabs in the Middle East. Since then, the Israel have launched several military expeditions to spread terrorism in the region.
On May 14, 1948, the State of Israel was created, sparking the first Arab-Israeli War. The war ended in 1949 with Israel's victory, but 750,000 Palestinians were displaced, and the territory was divided into 3 parts: The State of Israel, the West Bank (of the Jordan River), and the Gaza Strip. Fast Forwarding to, 1876, THEODER HERZL, a Jewish journalist and political activist published Der Judenstaat proposed an idea of creating independent Jewish state in Palestine. Hoping that it would put an end to centuries of persecution, and anti-Semitism. But the problem was that Jerusalem was, and still is holiest, and cities of Muslims. During First World War, the states of British Army occupied the Arab Territories. States of Turkey, Iraq, Lebanon, Syria, and Palestine was created. So, in Balfour Declaration British government promised Jewish to help them create the independent state in Palestine. A place where 90 percent of the population belonged to the Arab. In 1948, during the WW2 many Jews proclaimed the establishment of Israel. Over 700,000 were either expelled, or fled in terror. Lost their homes, land. By the end of two years of fighting in 1949 only 22 percent of what was British-occupied Palestine was left outside of Israel's border. That become Gaza-Strip, and the West Bank. But, Hamas, the Islamic Fundamentalist Military Group till date didn’t expect the identity of Israel to control the Gaza Strip.

Data-Analysis:

So, we are performing data analysis based on various parameters given to us in the dataset. It comprises the data from 2000-2023. There exists a total of 13 columns with 11124 rows. The dataset includes the following columns:

- Name
- Date_of_event
- Age
- Citizenship
- Event_location
- Event_location_district
- Event_Location_region
- Place_of_residence
- Place_of_residence_district
- Type_of_injury
- Ammunition
- Killed_by
- Notes

Note: All column names are lowercase

After importing the dataset into df, we wanted to see the count of values present for all the columns of the dataset.

```
df.count()
```

name	11124
date_of_event	11124
age	10995
citizenship	11124
event_location	11124
event_location_district	11124
event_location_region	11124
date_of_death	11124
gender	11104
took_part_in_the_hostilities	9694
place_of_residence	11056
place_of_residence_district	11056
type_of_injury	10833
ammunition	5871
killed_by	11124
notes	10844

dtype: int64

```
pd.isnull(df).any()
```

name	False
date_of_event	False
age	True
citizenship	False
event_location	False
event_location_district	False
event_location_region	False
date_of_death	False
gender	True
took_part_in_the_hostilities	True
place_of_residence	True
place_of_residence_district	True
type_of_injury	True
ammunition	True
killed_by	False
notes	True
dtype: bool	

There exists a lot of columns with Null/Missing Values. The most important one we had to pay attention to was the age column so we worked to remove the null values from this particular column when we reached the age section.

Analysis on Fatalities

We first constructed a simple line chart to view the count of Fatalities and their occurrences across the year range of 2000-2023.

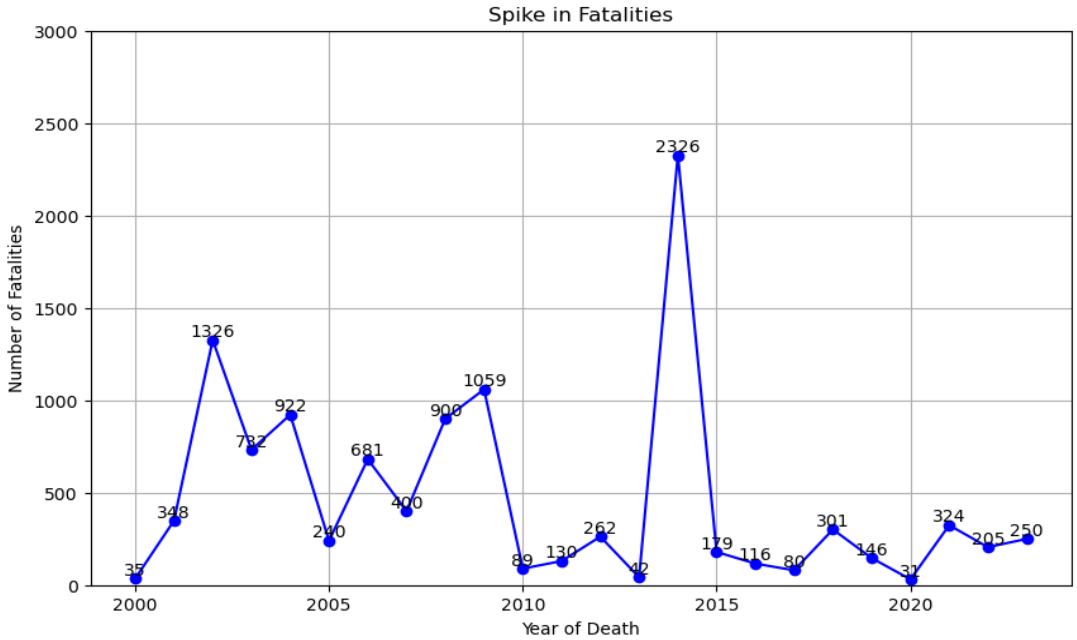

Figure 1: Spikes in Fatalities

From figure 1, we can view that the majority of fatalities recorded in this dataset is from the year 2014, while the 2nd highest spike can be seen in the year of 2002. We will explore the year 2014 later as we go on.

Another notable observation is that the number of fatalities after and before 2014. After 2014 we observe that the number of fatalities have lessened compared to the before timeline.

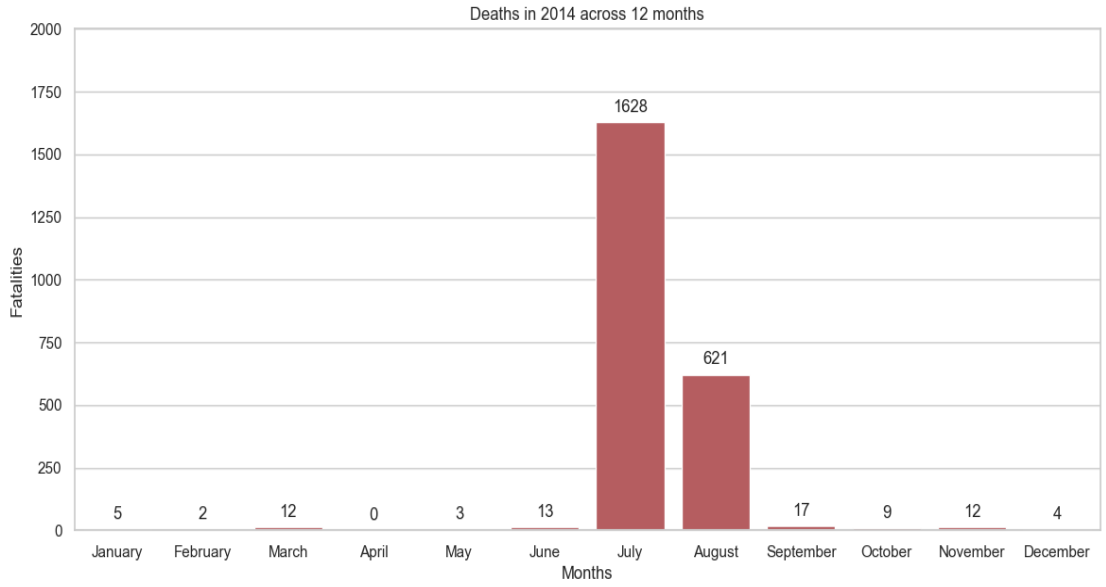

Figure 2: Deaths in 2014 with month

After exploring the 2014 year, we can notice that most of the deaths occurred in July and August. April is the only month where 0 fatalities occurred.

We will visit this Figure later to see what happened in 2014 which costed these many lives in the span of 2 months.

Analysis on Age of Fatalities

We then plotted the Age Distribution without Null Values to check the skewness of the data. Once the skewness is identified we can then figure out how to deal with NA values in a proper manner.

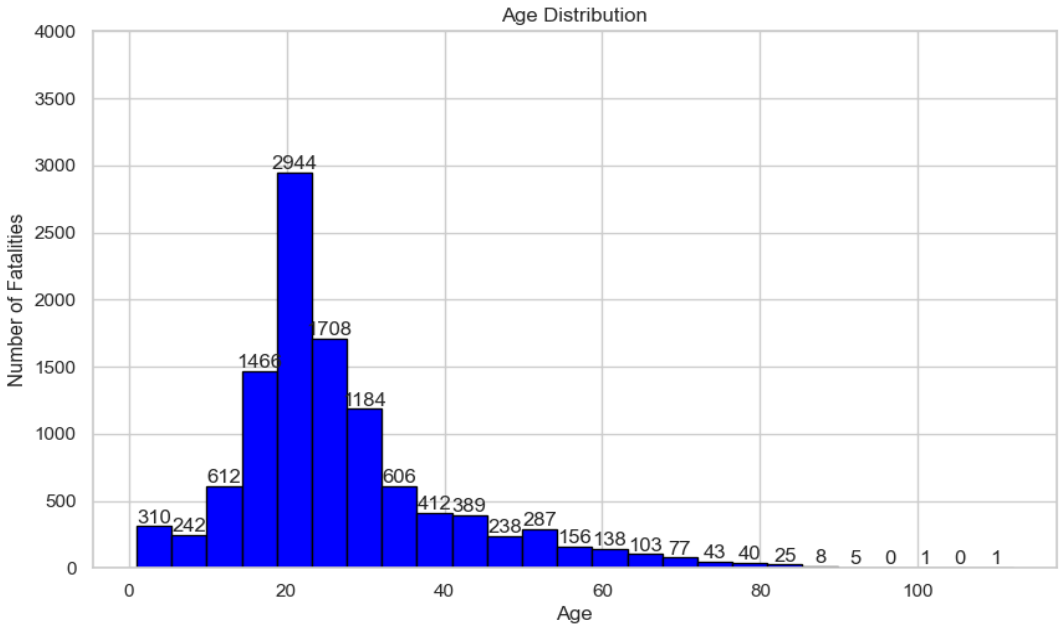

Figure 3: Age Distribution (Includes Null Values)

Age Data is right skewed, for skewed data it is better to replace null values with the median since it is less sensitive to extreme values. We then replaced the null values with median and showed the updated distribution charts.

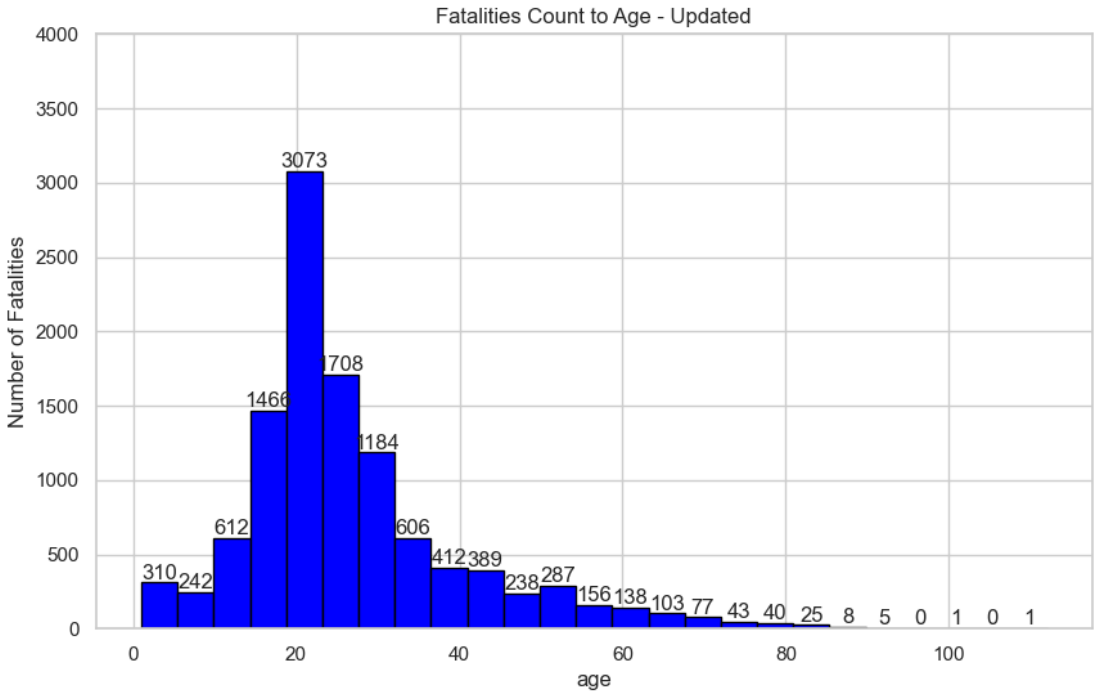

Figure 4: Age Distribution (Updated)

After this we used the describe method to find the basic information of Age column

count	11124.000000
mean	26.702265
std	13.706272
min	1.000000
25%	19.000000
50%	23.000000
75%	31.000000
max	112.000000

Table 1: Age Description

We can gather some simple analysis on the age column using these values. Most important is that the mean age is 26.70. We can also gather that 50% of fatalities lie 0-23 age range and the max age is 112 years old.
After using the median function from pandas we got 23.0 as a result for age column.

It is clear that more than 75% of deaths are of young age. The targeted deaths of young age can heavily influence the future of a country/society. We will further see how many of these young people belong from Palestine.

Analysis on Gender

We then moved on to the next interesting column which is gender. To find out how many of the fatalities that have occurred are distributed in terms of gender, we used a pie chart to show the percentage and a bar plot with count to get the exact number.

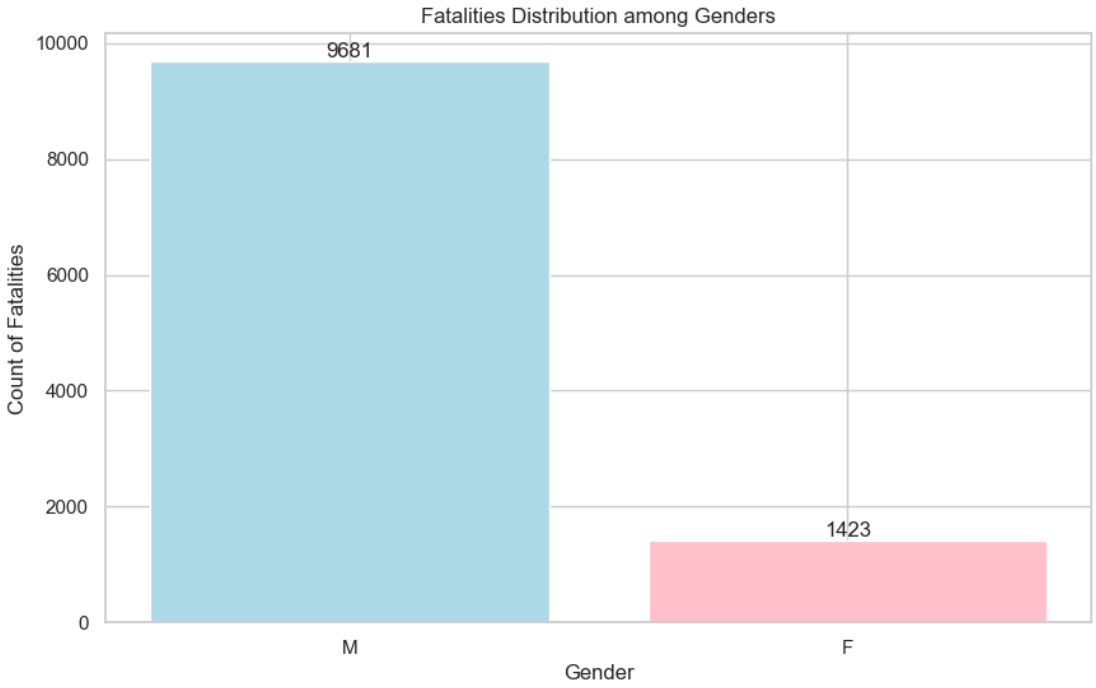

Figure 5: Gender Distribution of Fatalities

From figure 3 we can conclude that majority of the fatalities are Males, In a conflict it is most likely that the males will on the front lines to defend their lands and their families. We found no surprises here as this was expected

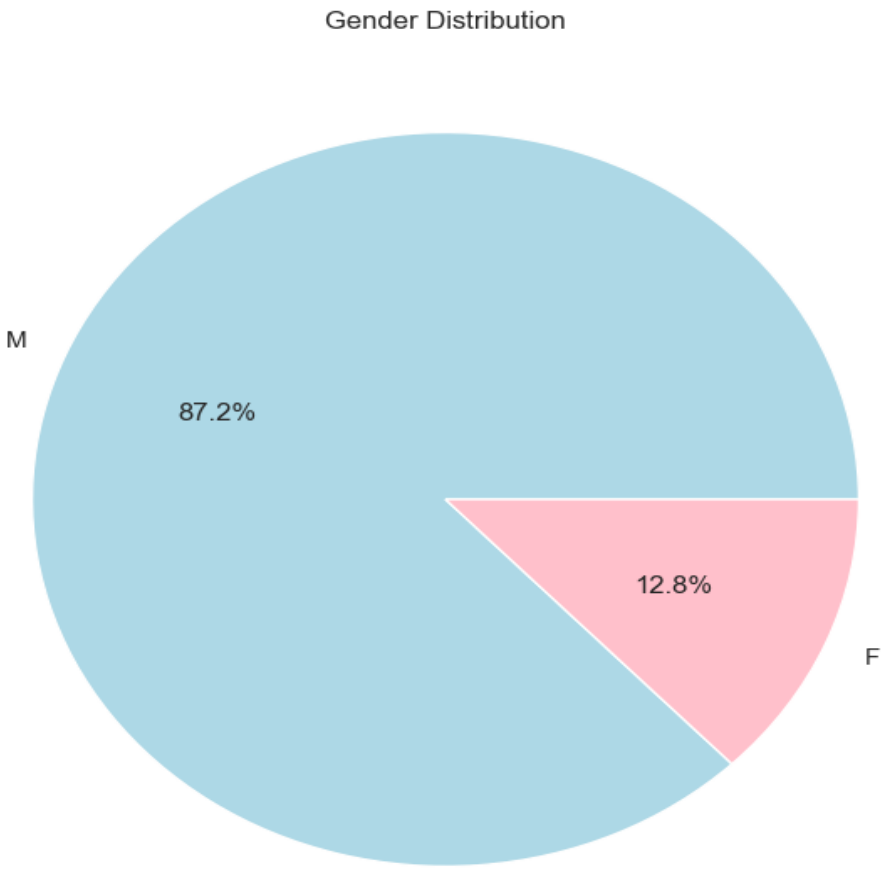

Figure 6: Pie Chart with Percentage to show Gender Distribution

From the pie chart we can identify that 87.2% of fatalities were males whereas females made up for the rest of 100% which is 12.8%.

Analysis on Citizenship

We then moved on to the next categorical column which is Citizenship, the dataset contains information whether on the belonging of a person to their Nation. This made analyzing data easier and concluding results from this analysis as the bar plot for citizenship clearly showed that majority of the fatalities were from one nation.

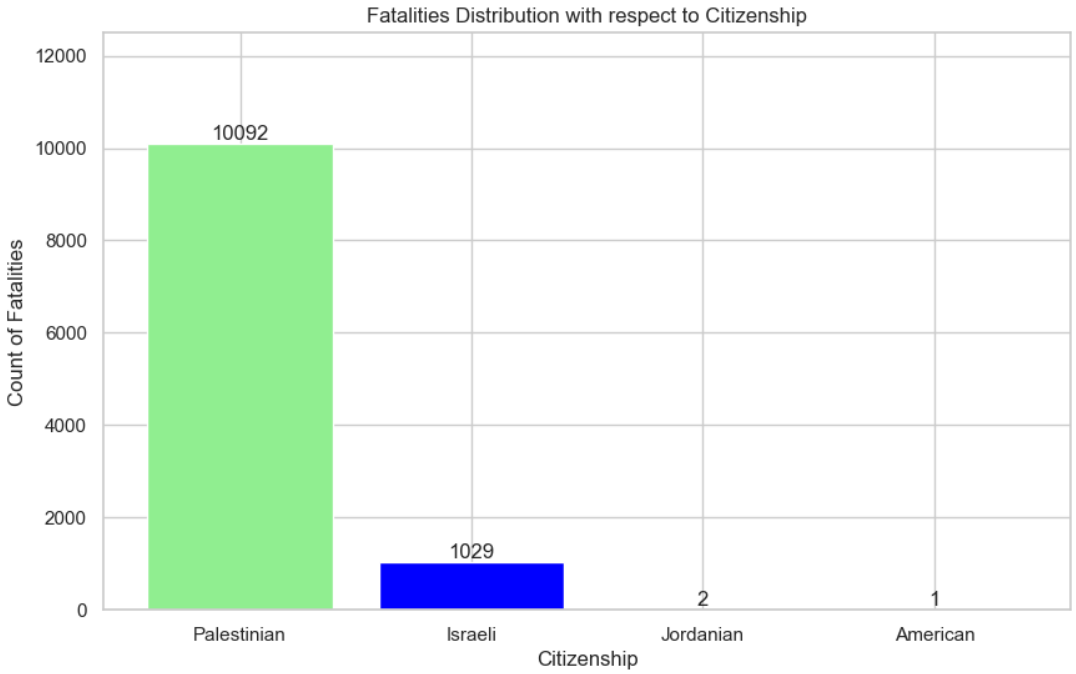

Figure 7: Fatalities Distribution with respect to Citizenship

From this Bar plot we can conclude that the most of the sufferings are on the nation of Palestine with the total count on 10092. Israel on the other hand suffered 1029 fatalities. Palestine has almost 10 times the fatalities of Israel.

The data from 2000-2023 shows that Palestine and its population has been attacked and dealt major sufferings over the years by the state of Israel.

There were 2 Special cases in the dataset, 2 Fatalities were of Jordan and 1 from American.

We then proceeded to perform analysis on the distribution of Age with Citizenship through the years using a stripplot from seaborn.

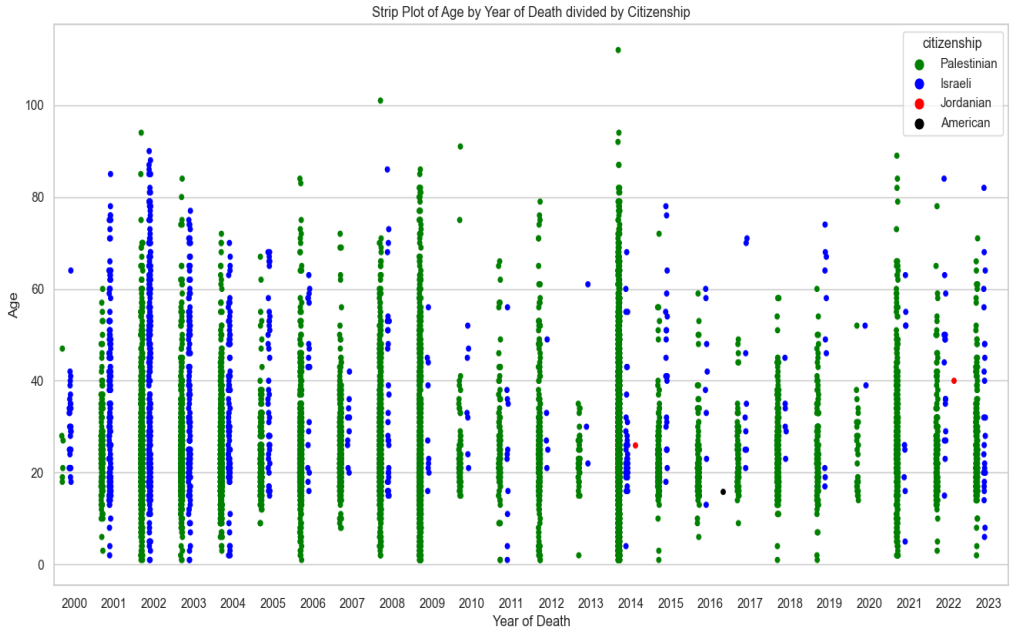

Figure 8: Stripplot showing the Age Distribution with Year of Death

In this chart, we provided a legend for observers to view the nationality of markers. Green is for Palestine and Blue for Israel. We can observe that the deaths of Palestinians have been in great numbers since 2001 till 2023, while the deaths of Israelis continued to lessen over the years. The oppression on the state and people of Palestine continued with spikes in fatalities in the years of 2009 and 2014.

The two special cases as mentioned above can be found in 2014 and 2022 for the Jordanians and 2016 for the American.

Analysis on Districts/Areas

We then moved on to analyzing the regions and districts which can be found in the dataset. For regions we have 3 regions present in the dataset which are Gaza Strip, West Bank and Israel.

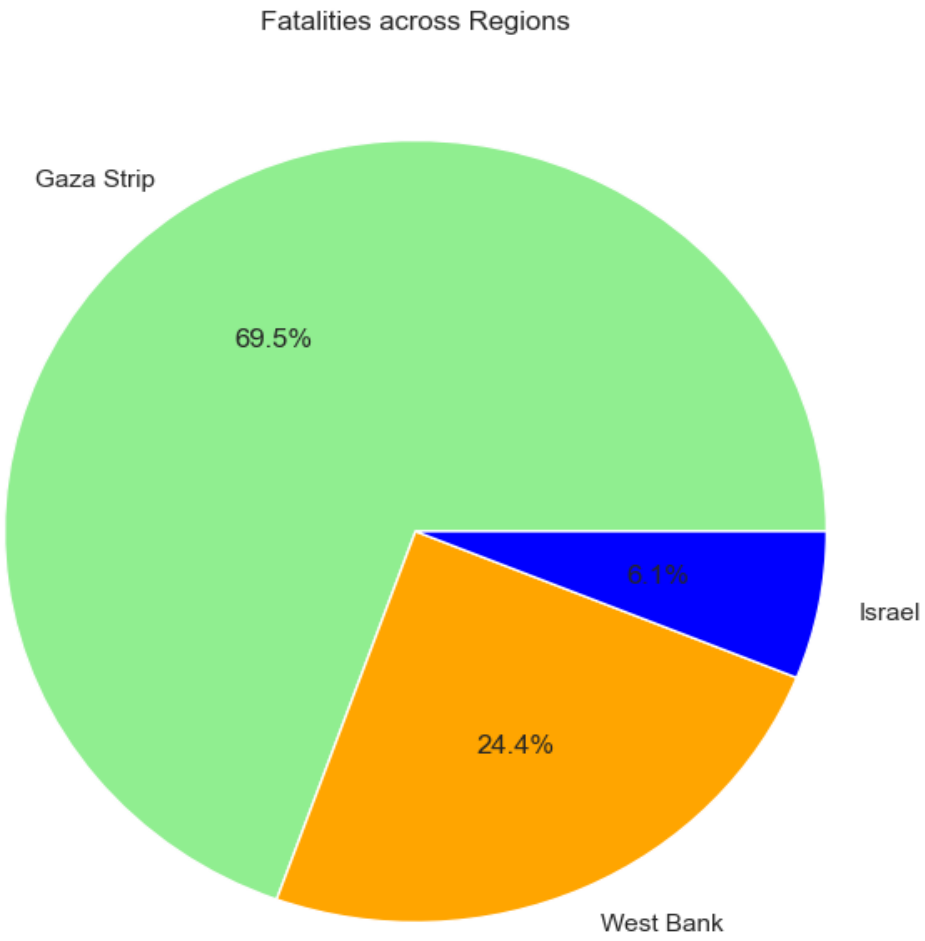

Figure 9: PieChart showing the distribution of fatalities across the 3 regions

We can observe from this piechart that only 6.1% of Fatalities occurred in Israel while the rest were in Gaza Strip and West Bank at 69.5% and 24.4% respectively making up around 93.9% of the piechart. These are the areas where majority of the deaths occurred and these deaths are of Palestinians.

From these numbers we can also conclude that 93.9% of the fatalities are of Palestinians while only 6.1% are Israelis.

We then moved onwards towards the districts and created a bar plot to display the distribution of fatalities across districts.

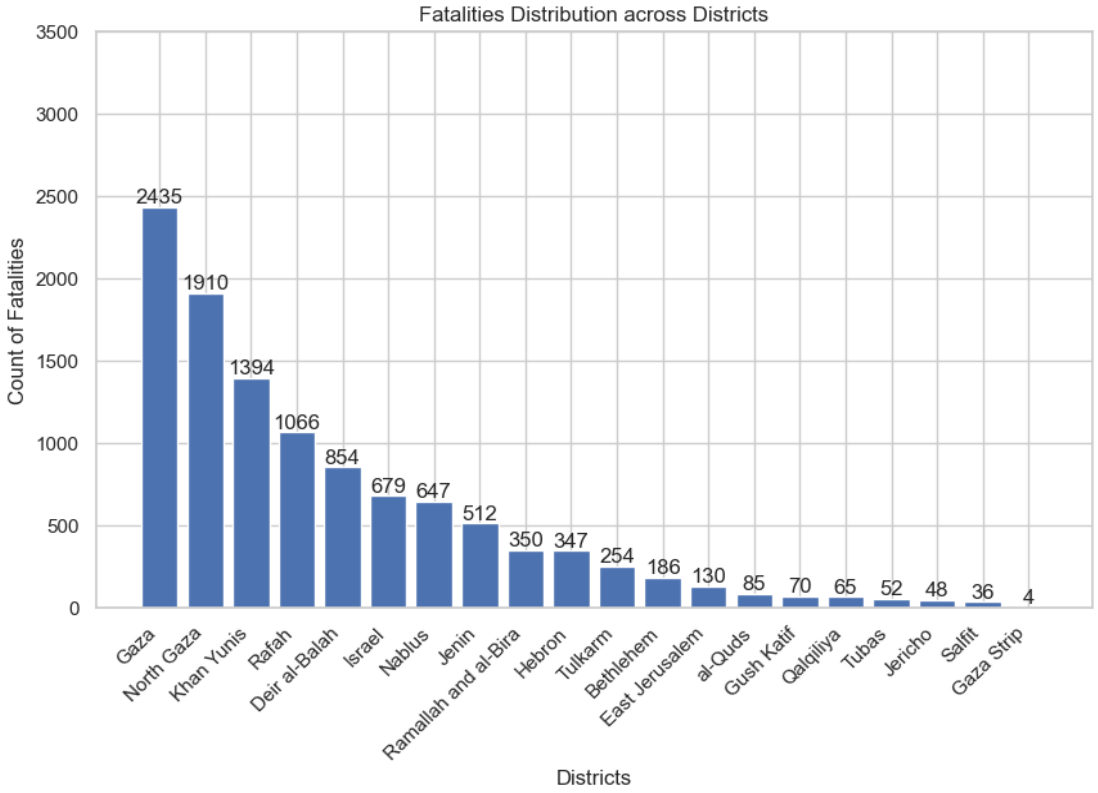

Figure 10: Fatalities across Districts

From figure 8 we can view the districts where the fatalities occurred with Gaze leading the chart at 2435 fatalities and North Gaza at 1910 with Khan Yunis coming at 3rd with 1394 fatalities. We can notice that Israel can be found with 679 fatalities in this chart.

We moved back to the regions after this and created Line charts for each of the region to gather more data.

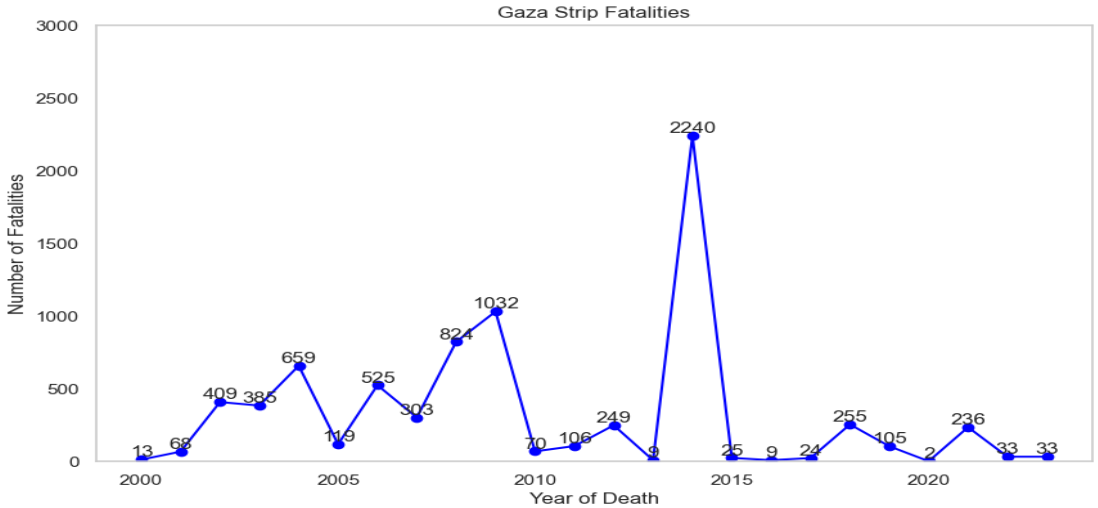

Figure 11: Gaza Strip Fatalities

From Gaza Strip we can notice the huge spike in fatalities in the year of 2014 at 2240 count with second most in the year of 2009 at 1032 count. The year after 2014 shows a decline in the number of fatalities until 2018 which spiked to 255 and then later another spike was observed on 2021. From 200 we can notice a constant increase in the number of fatalities with a decline in 2005 only.

We then moved on to the next region with most fatalities which is West Bank.

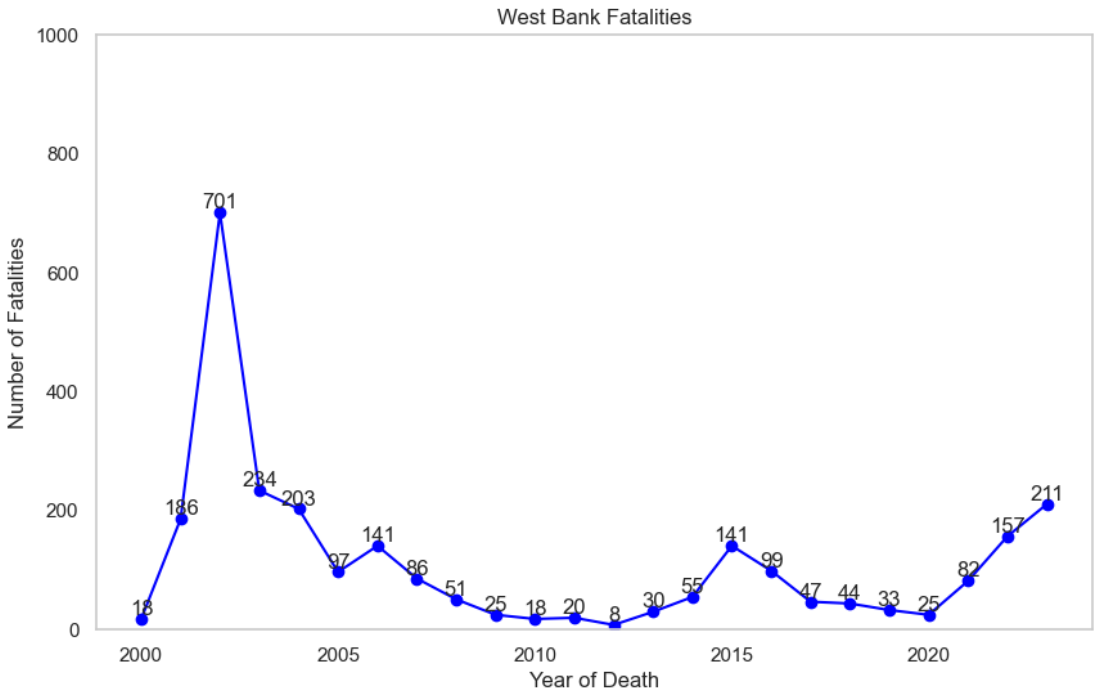

Figure 12: West Bank Fatalities

The West Bank line chart is different from the Gaza Strip as there are no huge spikes in the count of fatalities in the years of 2009 and 2015 but instead the highest count of fatalities can be noticed in the year of 2002 which is 701. Compared to Gaza Strip we can notice that after 2021 the deaths are gradually increasing whereas in the Figure 9 the deaths were decreasing.

Now we will analyze the data of Israel Fatalities.

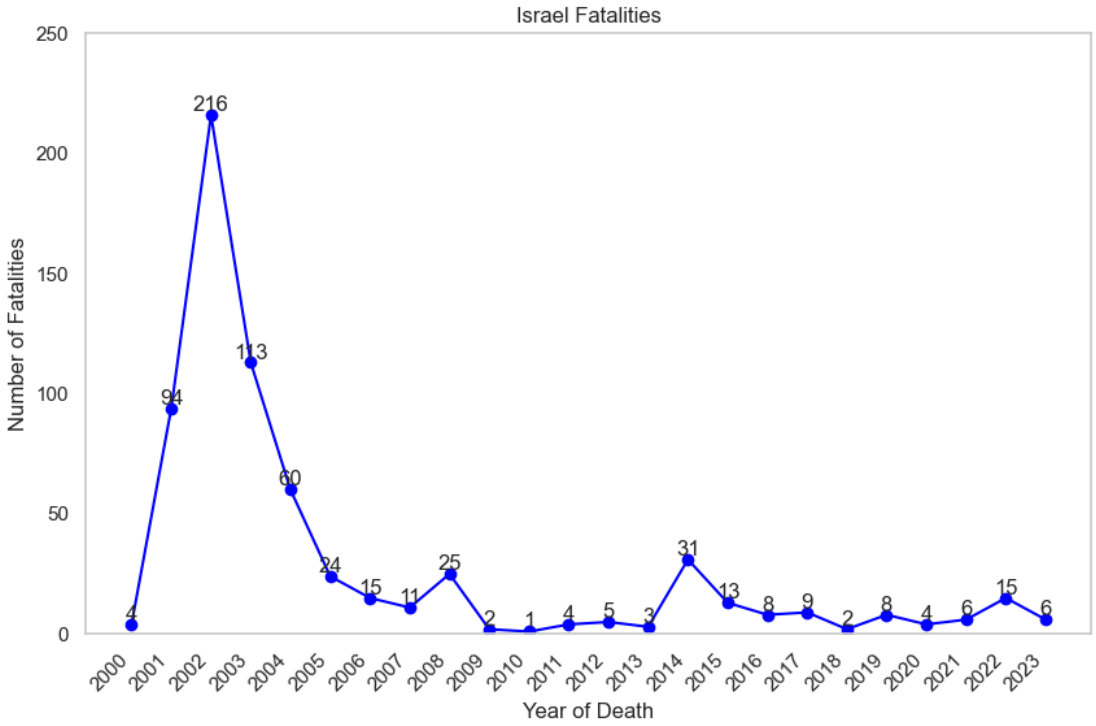

Figure 13: Israel Fatalities

In this line chart we can observe that the Israelis suffered the highest count of fatalities in 2002 with 216 and the second most at 113 in the year 2003. Overall the count of fatalities has reached single digits for many years. The most amount of fatalities is present in the years 2000-2004.

This concludes our analysis on the Regions and Districts. It is clearly displayed that Palestinians suffered huge losses of lives in both of their regions in the span of 23 years with most in the year of 2014 and 2009 at Gaza Strip with West Bank suffering the 3rd highest count of fatalities in the year of 2002.

Analysis on Injuries

After the analysis performed on fatalities across regions and districts, we moved on towards the next column which describes the types of Injuries suffered.

To show the types of injuries listed in the dataset and their count, we used a bar plot on the column.

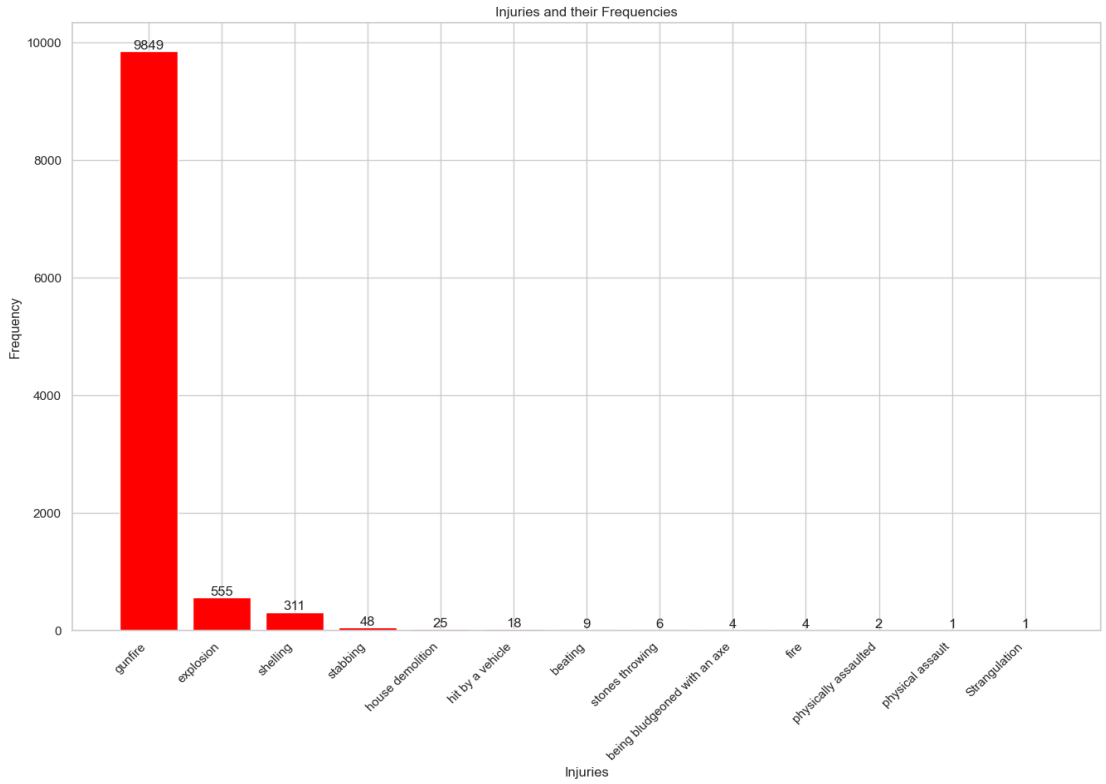

Figure 14: Injuries and their Frequencies

After plotting Injuries and their frequencies, we can clearly see that gunfire tops the list with the biggest count of 9849 while the 2nd most occurring injury type is explosion at 555, coming at 3rd is shelling at 311. The rest of the list contains small counts which we can ignore as the top 3 covers the majority of types as well as shows that guns have been the most used weaponry to which lives have been lost. This indirectly points towards the extreme involvement of armed forces.

We then explored gunfire more and charted it using a line graph across the years.

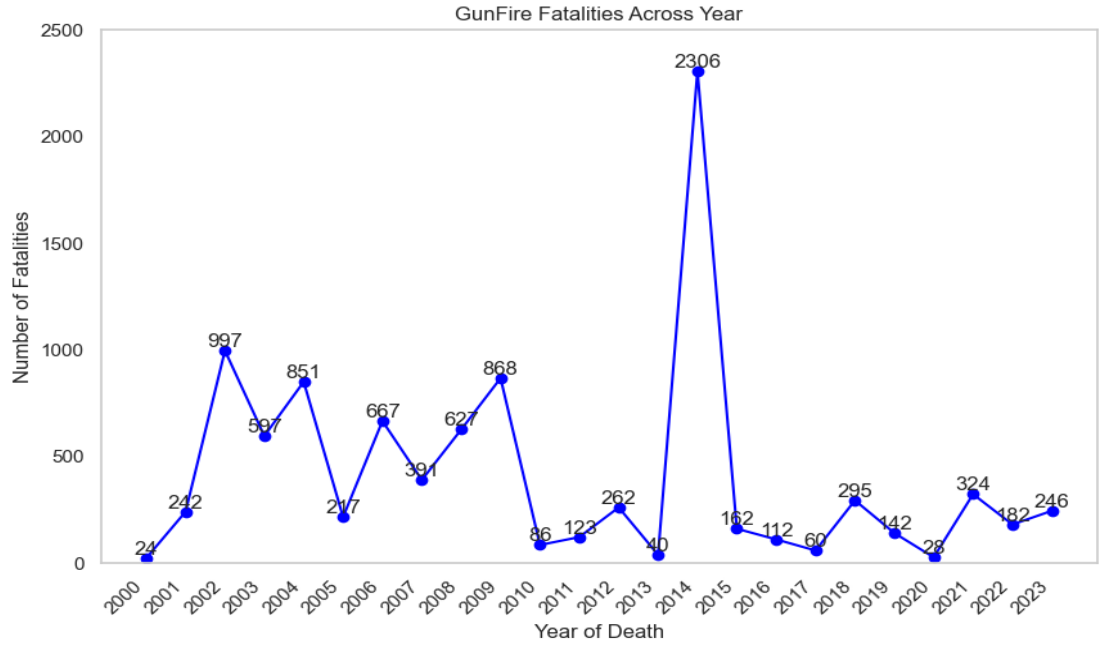

Figure 15: Gunfire Fatalities across Years 2000-2023

We can observe this chart and find out that up till 2014, gunfire has been used excessively to take lives. Starting from 2000 we notice the first spike in usage of gunfire from 2001 to 2002 which shows 242 to 997 count respectively. Then the fatalities occurring from gunfire remains mostly in the range of 400 to 900 till 2010. In 2010 it drops but then the biggest spike occurs at 2014 with the count of 2306. After 2014 we can observe that the deaths to gunfire have lessened even less than the count of 2001 and minor spikes occurring in 2018 and 2021.

Coming back to Figure 12, we then picked out the top 3 types of injuries and constructed a line chart to show the trends in all 3 in a single graph.

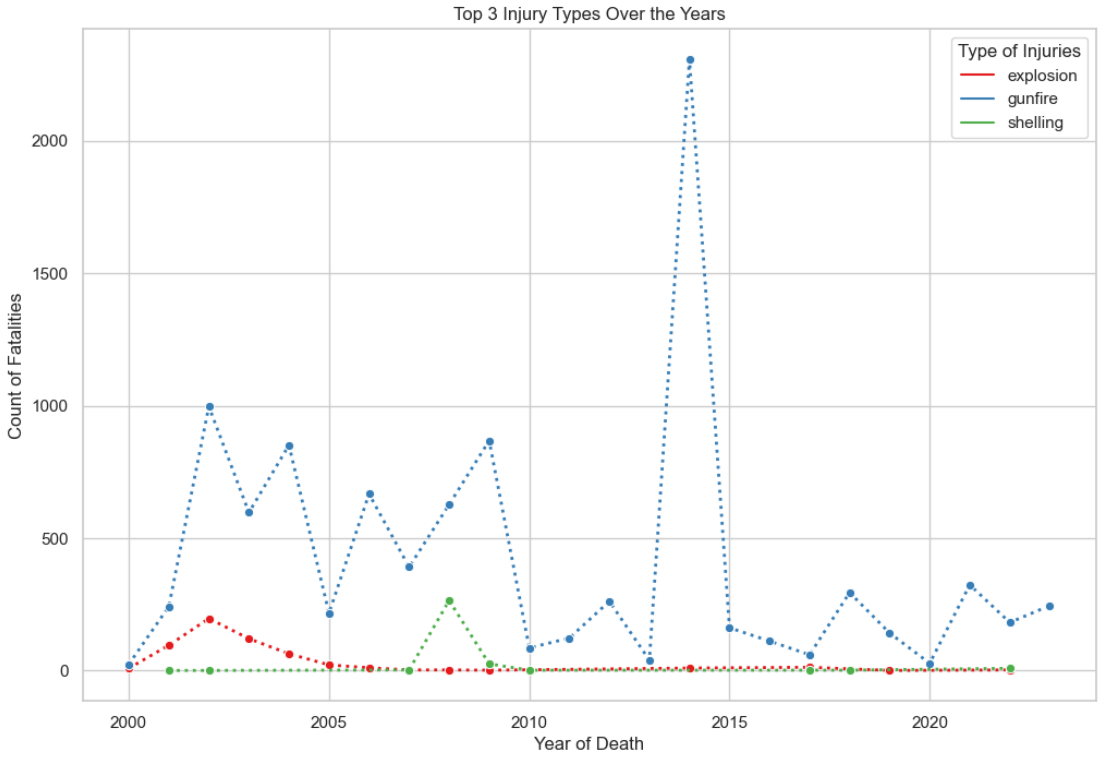

Figure 16: Top 3 Injuries from 2000-2023

We observe that the majority of the count across the years given in the dataset are held by gunfire while there are minor spikes from both explosions and shelling but not as constant. The most notable explosion spike is in 2002 and then it remains low and close to zero while for shelling the most notable spike is in 2008, for the rest of the chart it remains close to zero.

Analysis on Ammunition

After analyzing the Injury types we moved on to the next column which is Ammunition. So, we performed the analysis on the column of the ammunition with the number of fatalities caused by the ammunition's. To show the total types of Ammunition and the fatalities caused by them, we used a bar plot with count.

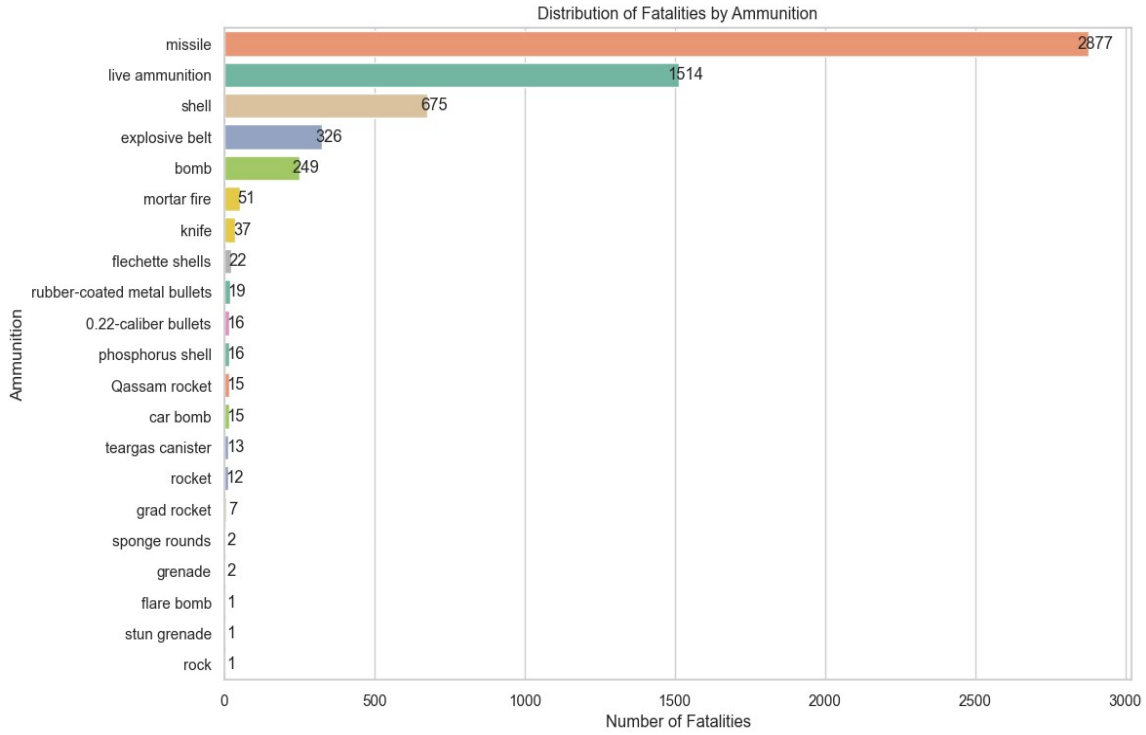

Figure 17: Distribution of Ammunition

We can see all the types of Ammunition used across the range of 2000-2023. Missile and Live Ammunition are the only 4 digit weaponries with 2877 and 1514 count respectively while shell has claimed 675 of the fatalities. Then we can see explosive belt and bombs with 326 and 249 respectively. The rest of the list has lower counts.

After constructing the plot above, we were interested to see the distribution of these Ammunitions on Palestinians and Israelis so we created a plot to showcase both nationalities and how they fare against the Ammunition

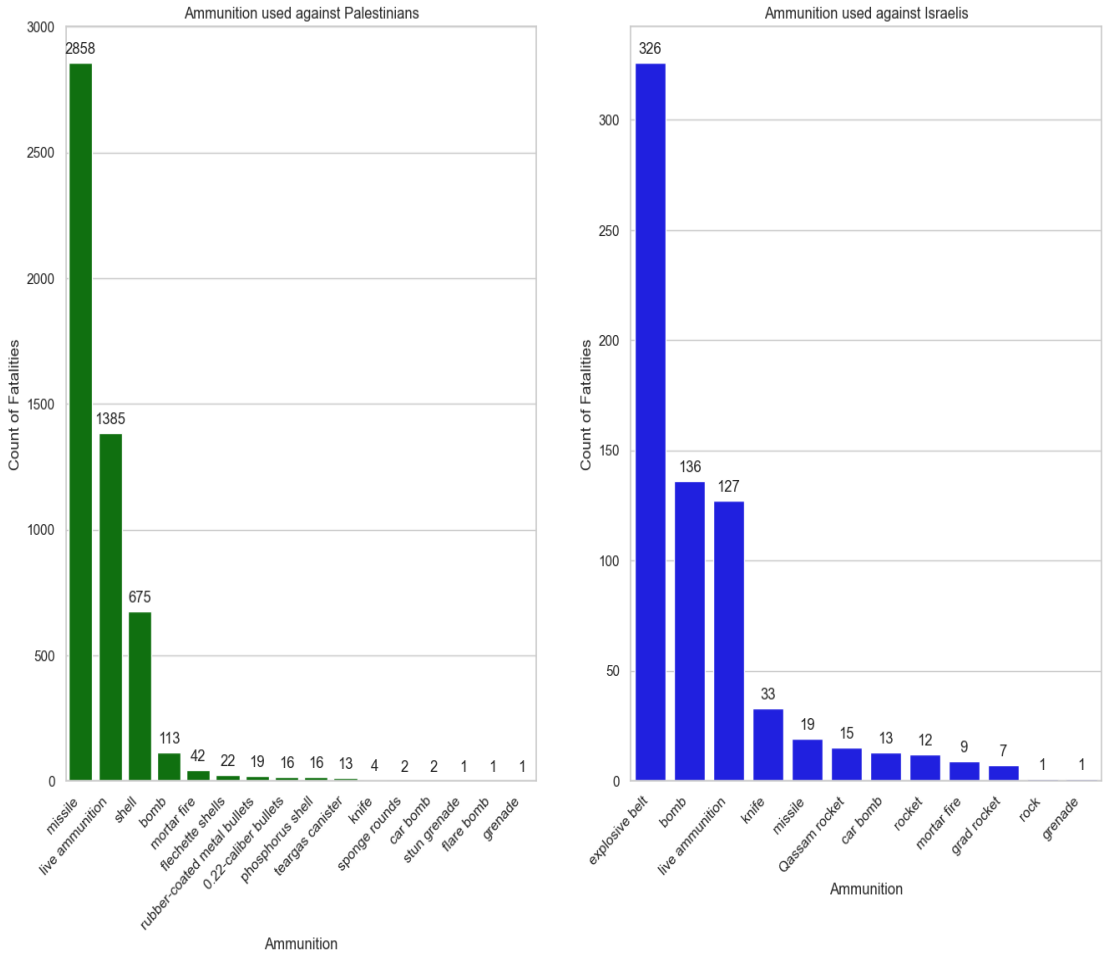

Figure 18: Distribution of Ammunition for Palestinians and Israelis

We can see that the leading count of missile from figure 15 which is 2877 is distributed very unevenly with Palestinians suffering from missiles with 2858 count while Israelis only suffered at a count of 19.

Missiles have been excessively used in claiming the lives of Palestinians as mentioned where as the leading count against Israelis have been explosive belts at 326 count.

Comparing live ammunition across both charts, we can see that Palestinians have suffered 1385 losses of lives while Israelis have suffered 127. Access of live ammunition and missiles have made it possible for Israelis to deal major sufferings to the Palestinians.

Next we performed the analysis of the biggest spike in deaths which occurred in 2014 which can be seen in figure 1 and 2.

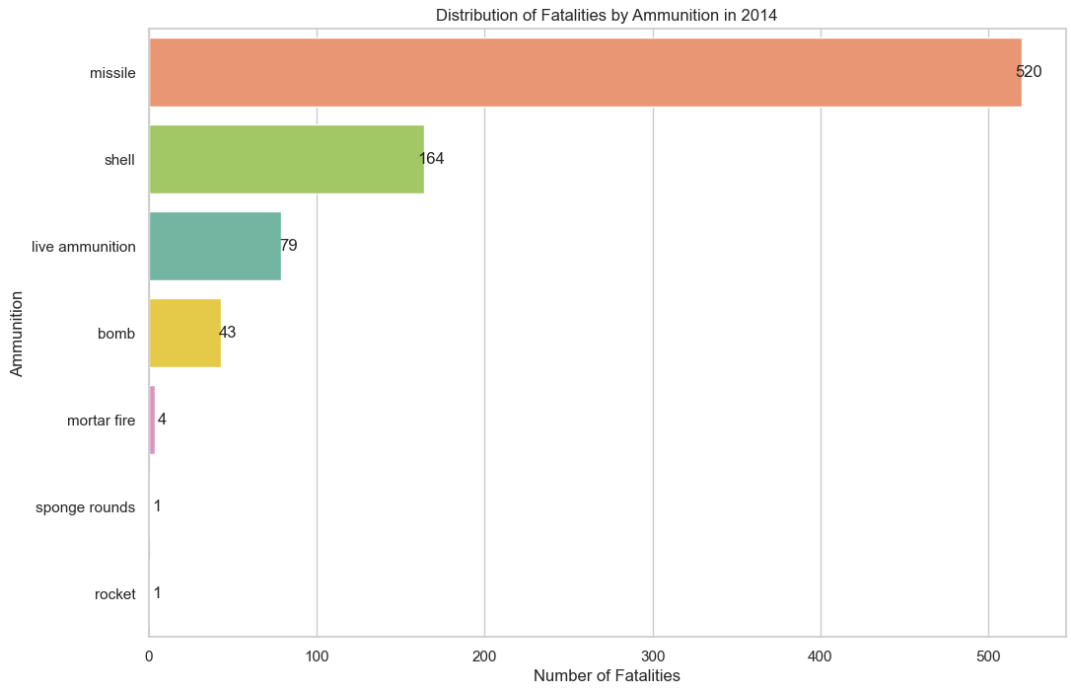

Figure 19: Number of Fatalities in 2014 by Ammunition

It is very noticeable that we have many missing values as the total death count of the year 2014 is 2326 while this chart shows below 1000. Use of missile is leading again as seen before in the distribution chart across years while shelling is 2nd leading type of ammunition used in 2014. Live Ammunition comes in at 3rd claiming 79 lives.

The data is missing so we cannot make further analysis but this chart can be used to gain an idea of how Palestinians have been bombarded with missiles throughout the years.

Next we created a donut chart for the year 2014 for the distribution of the fatalities by ammunition's to show the top 4 types of ammunitions used.

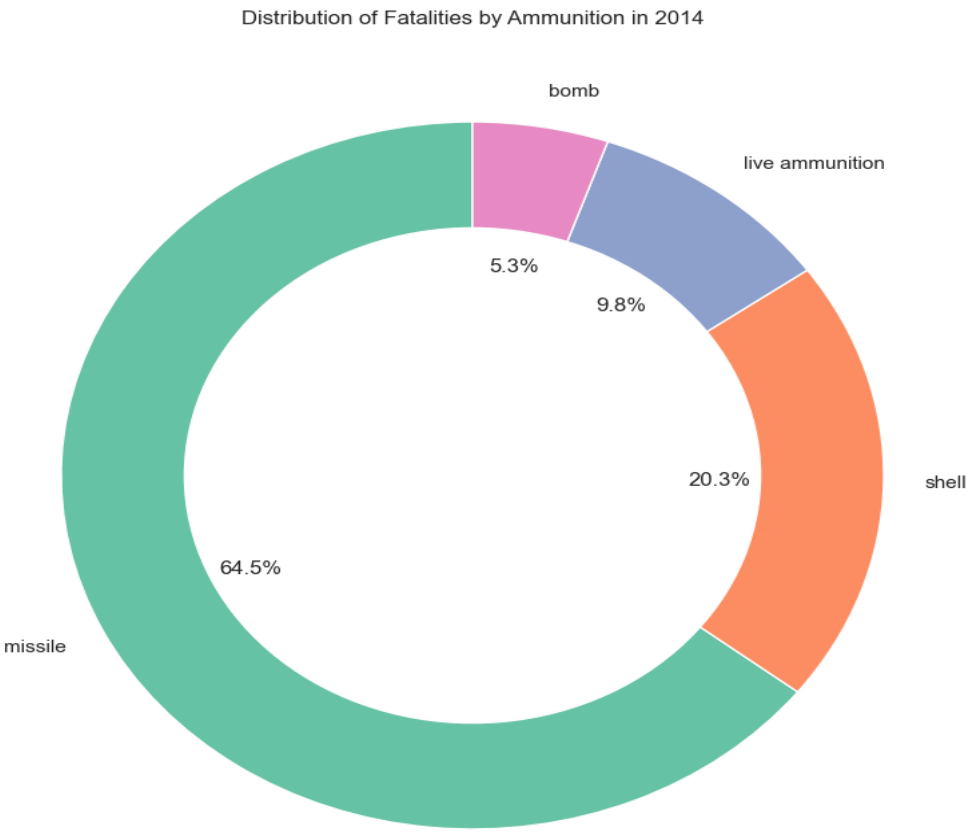

Figure 20: Donut Chart for Top 4 Ammunitions

Here, we did not include the mortar fire, sponge rounds, and the rocket. Because their ratio was less. So we put the threshold of 4. This threshold was count based fatalities caused by different weapons. We can see the exact percentage of ammunitions used in the donut chart.

Now leaving the exploration of 2014, we move back towards the complete year range. We wanted to look at the use of top 5 ammunitions across the year range so we constructed a line chart with the mentioned ammunitions.

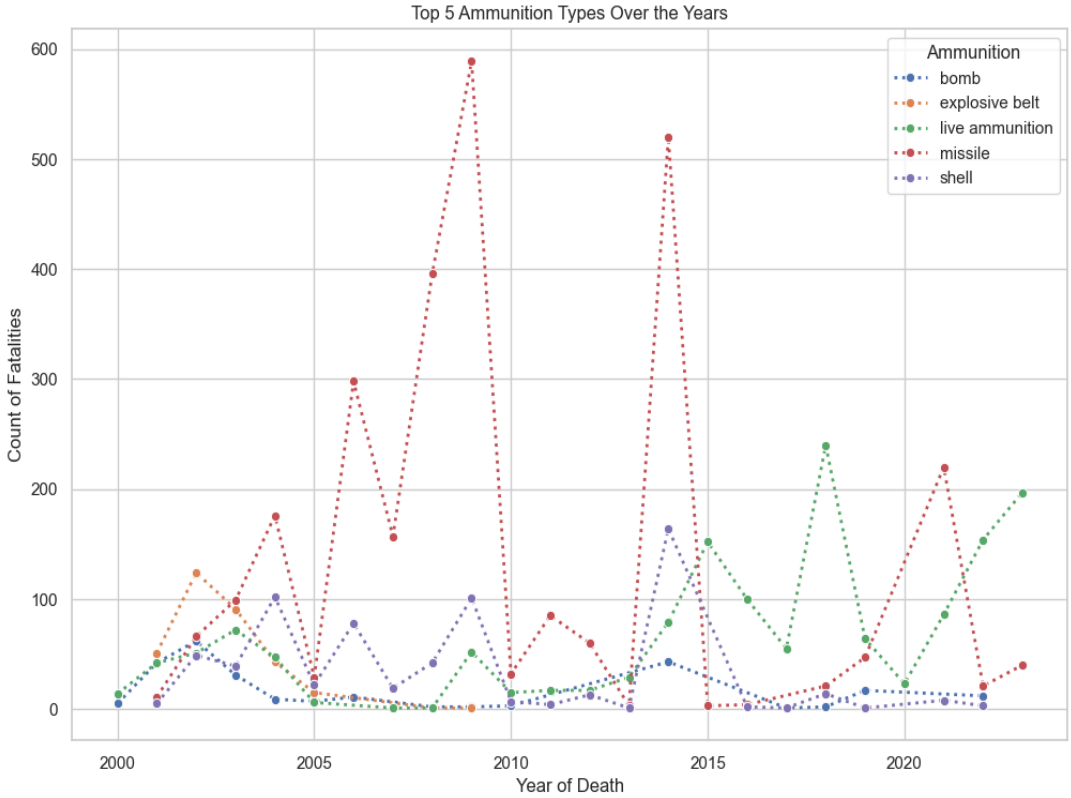

Figure 21: Line Chart with Top 5 Ammunitions

We already know that missiles have been widely used but we can see its rise over the years. Missiles were greatly used in 2004 with its count almost reaching 200. We then see it being used less in 2005 but then it starts to spike over the years from 2006 to 2009. It drops below 100 for the next couple of years but in 2014 missiles were again used to increase fatality count above 500. Again it drops and we see another rise in 2001 where missiles were used to claim more than 200 lives.

We can also observe that explosive belts have not been used after 2009 whereas live ammunitions have seen a spike in usage after 2014. After missiles the biggest spike is of live ammunition in 2018 claiming around 250 lives.

Shell's biggest spike occurred in 2014 which claimed above 150 fatalities.

Analysis of Common Characteristics between Males and Females

After this, we head back over to age but included the gender and compared both males and females. To compare we created a plot with two figures side by side

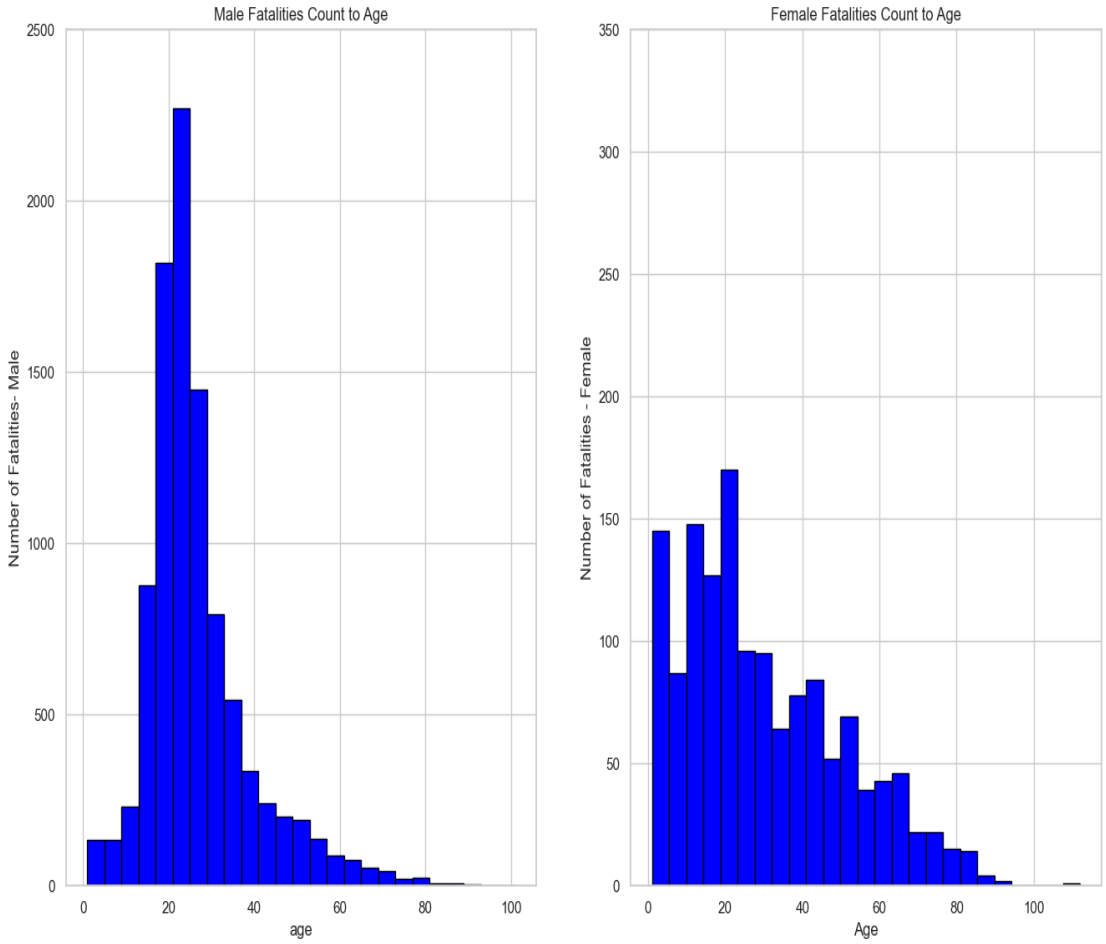

Figure 22: Comparison of Age of Males with Females

In these two plots we can see that the highest counts for both genders lie at 21. The total number of fatalities for females are less as shown before in the report. We do notice that the ratio of young females to young males in the range of age 0 to 10 are almost same. Almost the same number of infants for both genders have lost lives.

One interesting fact to note is that the max age for females is 112 while male max age is 101.

We also constructed a box plot to further display the quartiles on which age is grouped together for males and females separately.

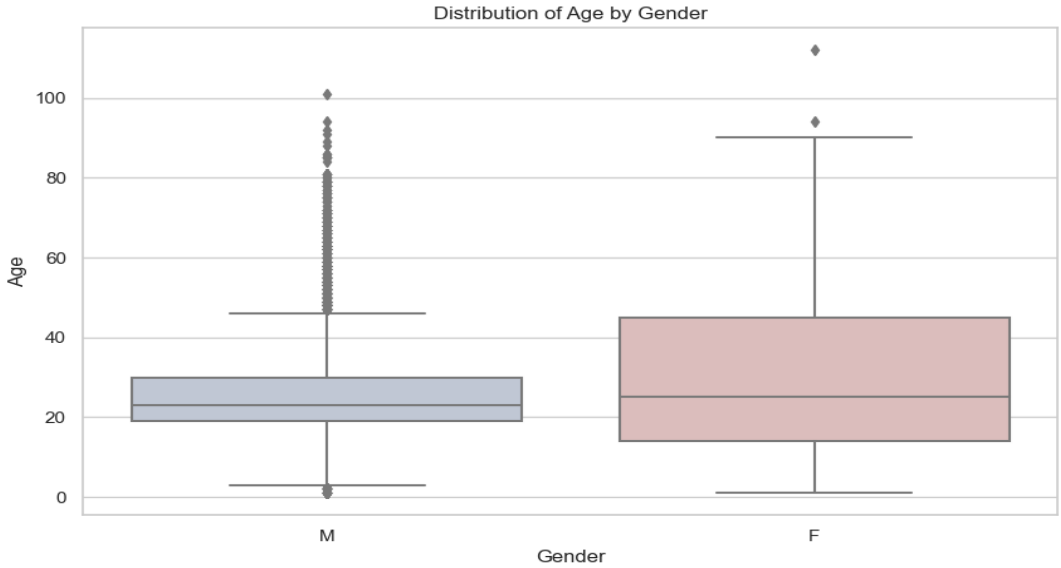

Figure 23: Boxplot of Males and Females Age

We can notice from the male graph that the data is right skewed, hence the box plot sets the upper whisker of Males in 40s while the upper whisker of females is in 90. The values of ages for males are closely packed in the young ages while the distribution after 30 is very large and not tightly packed. The boxplot considers every age after 44 to be an outlier where as in the case of females the upper whisker is set at 90 hence only 2 notable outliers are shown.

More information using .describe() method

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

count	9681.000000
mean	26.219812
std	12.270287
min	1.000000
25%	19.000000
50%	23.000000
75%	30.000000
max	101.000000

Table 2: Male age description

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

count	1423.000000
mean	30.036543
std	20.776422
min	1.000000
25%	14.000000
50%	25.000000
75%	45.000000
max	112.000000

Table 3: Female Age description

We can compare the means of both Males and Females. Males are lower at 26.21 whereas Females are at 30.03

The quartiles for both are different by margin most notably the 75% of females is at 45.0 while for males it is at 30.0

For the last bit we conducted an analysis on Male and Female Fatalities across Districts to see where did which gender suffer more. For this we created a bar plot showing Districts with their Fatalities count for both males and females.

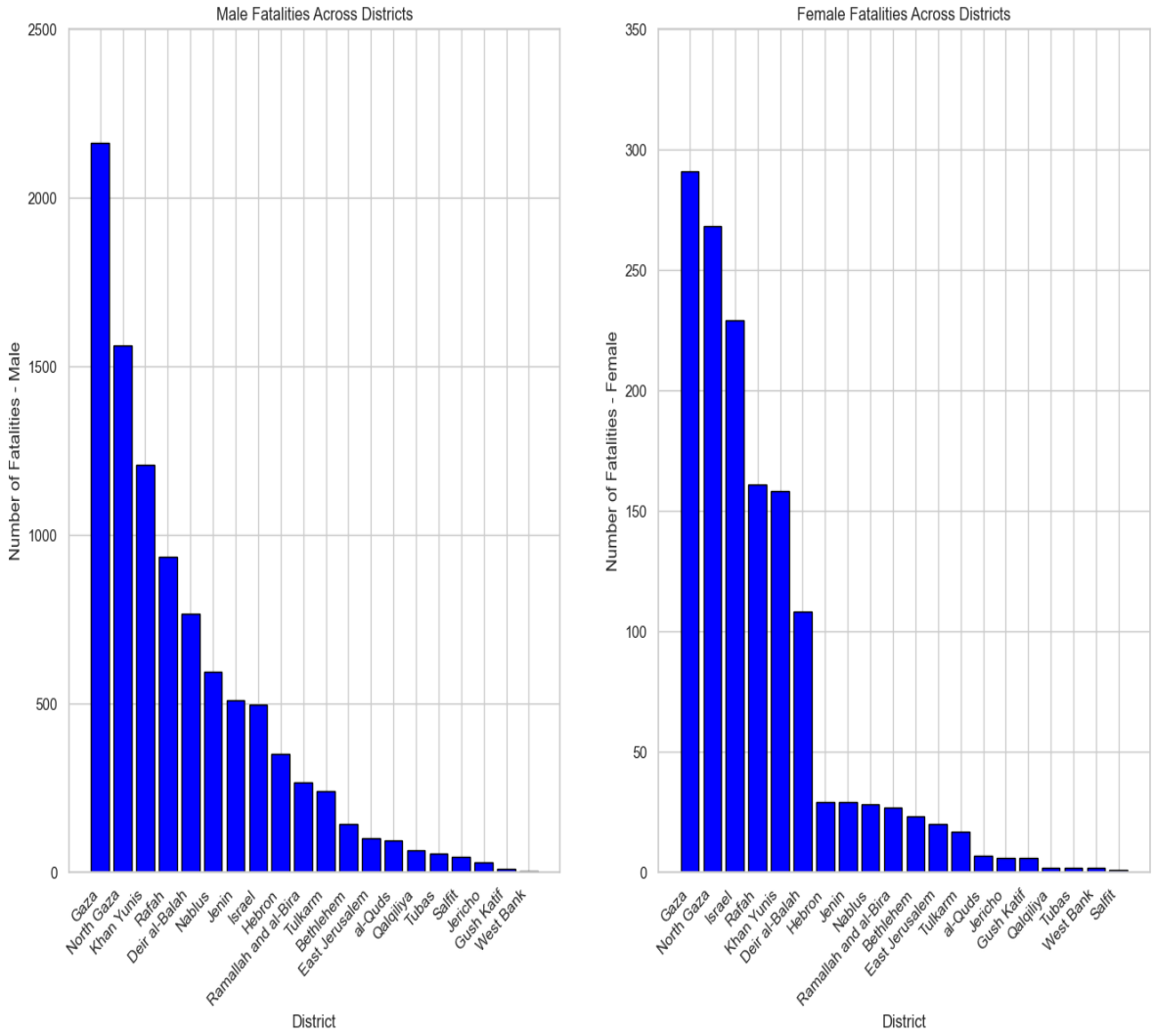

Figure 24: Male and Female Districts Comparison

We noticed that Gaza and North Gaza are the common leading districts of fatalities for both males and females but one interesting fact that is present for females is that the 3rd highest fatality count is from Israel. For males, Israel has almost 500 deaths where as for females Israel is around 230. The ratio here is 500:230 where as for the first two entries for example in Gaza males surpass 2200 mark while females are at a far lower number of at around 290. This ratio can be for easier explanation be rounded close to 10 males for 1 female, furthermore in the case of North Gaza, the ratio is around 1550:270 which can also be simplified roughly at 6:1 or 5:1 ratio. In the case of Israel the ratio of 500:230 can be rounded off to be 2:1 so for every 2 death of males, 1 female also died.

Post Analysis

On further research through the internet, we discovered that 2014 spike happened due to the "2014 Gaza War" operation started by Israel on July 8th 2014 in the Gaza Strip. Figure 11 clearly displays this event with its highest peak at 2240. We had already showcased the Fatalities distribution in the year of 2014 through all 12 months. This operation started on 8th of July and lasted till 26th August 2014. Figure 2 shows this timeline with numbers using a bar plot. For more reference we have pasted both the graphs below.

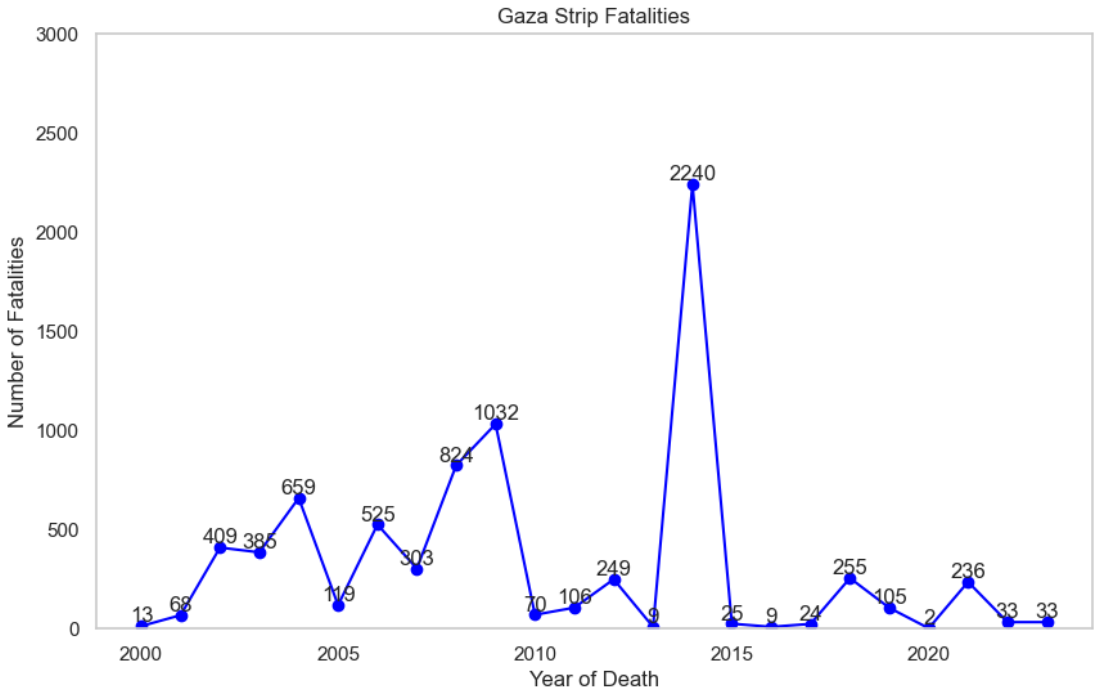

Figure 11: Gaza Strip Fatalities

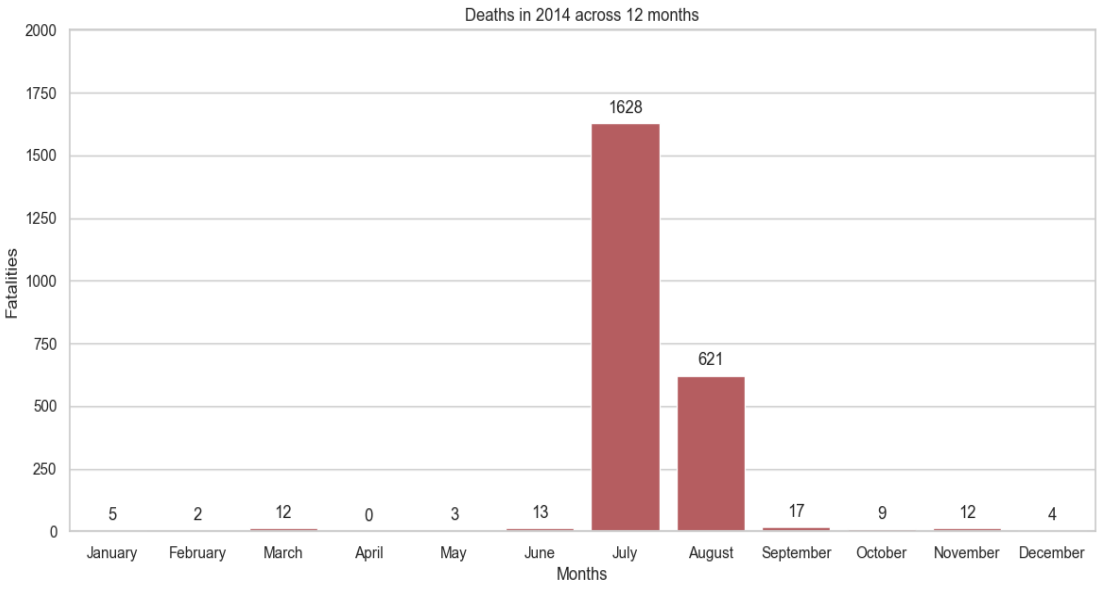

Figure 2: Deaths in 2014 with month

For Ammunition used in this war, we have figure 21

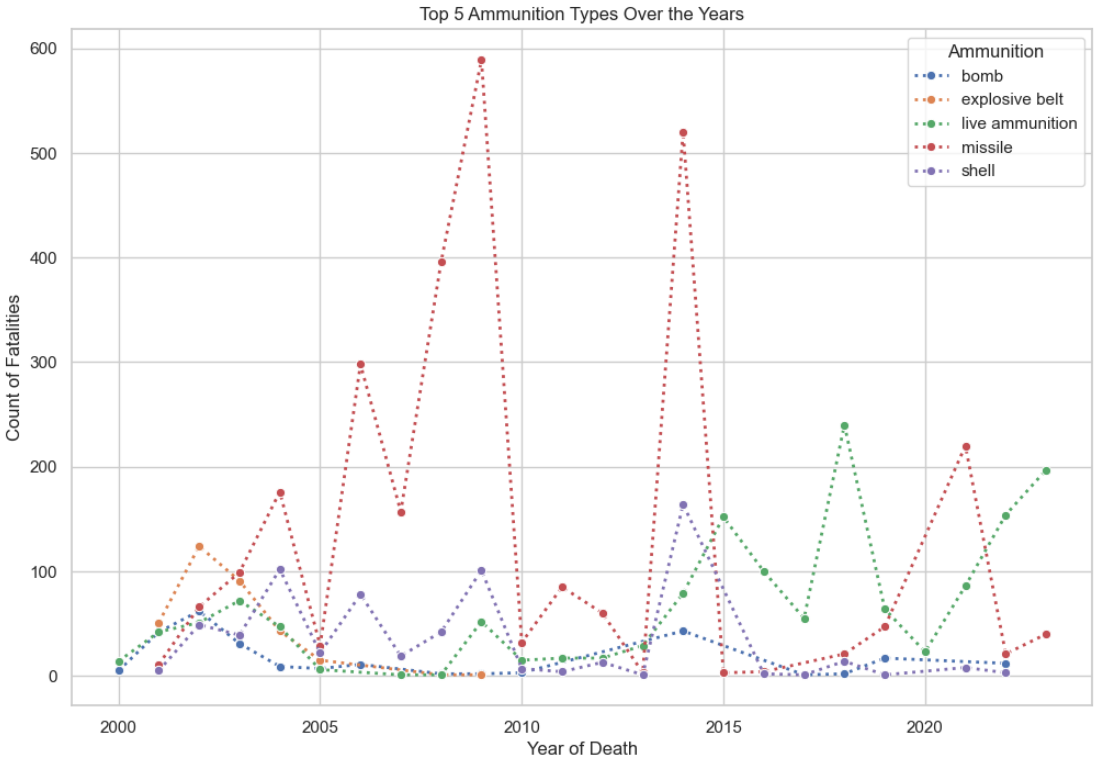

Figure 21: Line Chart with Top 5 Ammunitions

From online research through Wikipedia which are referenced at the end of this report, we found out that Israel fired 5830 missiles in 2014. From figure 21 we can see that missiles resulted in the most fatalities in the year 2014.

We also found out that Israel used 34000 unguided shells on Gaza Strip and we can clearly see the rise in fatalities with such ammunition in 2014 through the figure above.

Moving backwards to the rise in fatalities in 2009, which is displayed in the figure below

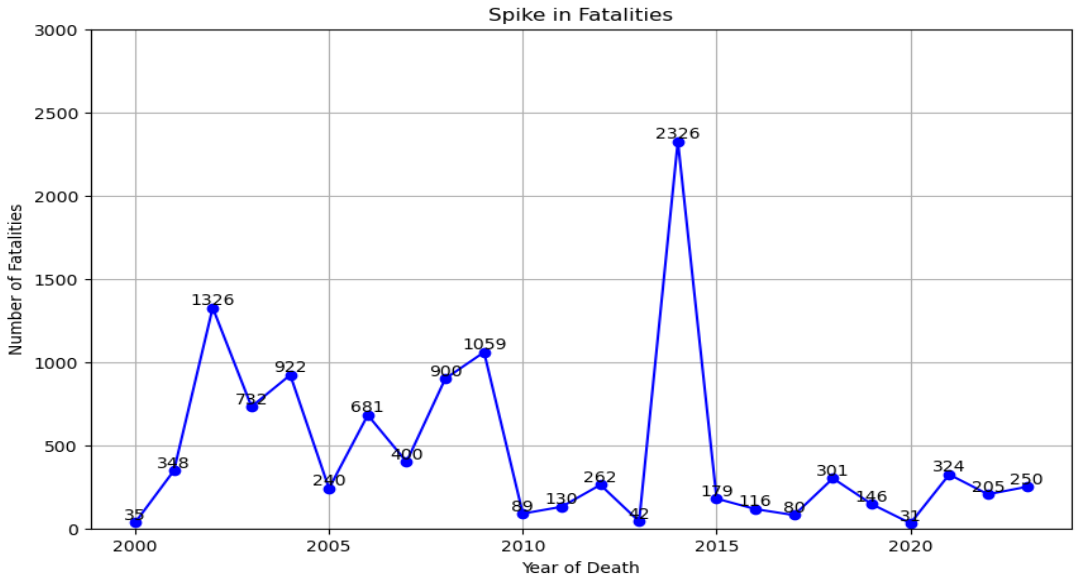

Figure 1: Spikes in Fatalities

We notice that 2009 had a 4-digit Fatality count closely followed by a 900 count of fatality in the year 2008. This happened due to the Gaza War which started on 27th December 2008 till 18th January 2009.

Wikipedia cites that the war resulted in 1,166-1417 Palestinian and 13 Israeli fatalities. Our number for 2009 is 1059 which is not far off. The complete data in conflicts which result in wars will always have missing uncertainties but we can conclude that the accuracy of the dataset is enough to make these analyses.

Moving to 2002, we notice the 2nd highest peak in the chart at 1326 Fatalities. For more information, we have attached a timeline of Wikipedia which showcases the events from January till December.

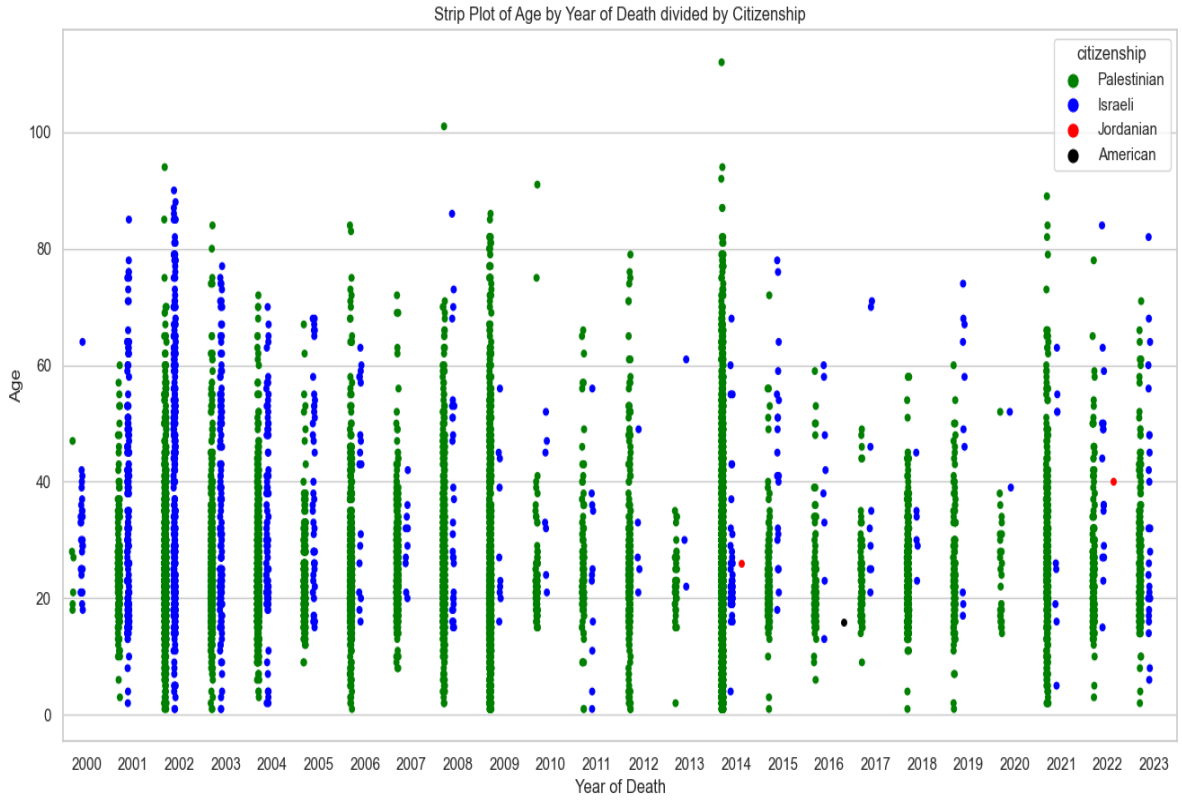

Figure 8: Strip plot showing the Age Distribution with Year of Death

Viewing this strip plot we can notice that both sides suffered many fatalities during the conflicts of 2002. 2001-2005 include many Israelis Fatalities but after these years the fatalities on the side of Israelis have lessened by a big difference with its density at the highest in 2002.

Closing Remarks

We went through an extensive report on the Fatalities describing gender, age, citizenship, types of injuries and types of ammunition used. The data source has also been given on the last page of this report. During the analysis, we noticed that the side of Palestine has been suffering at the hands of Israelis for too many years. Countless lives have been lost which are unrecorded in this dataset. Currently as we are making this report, another conflict is happening on the very same issue of occupation of land. This dataset is recorded till 2023 excluding the deaths that have occurred recently. Datasets do not explain the why, they are a set of categorical and numerical data which can only explain trends that have occurred in the past or at best be used to derive predictions which may or may not happen. Datasets also do not answer the question of “how many more lives will be lost on both sides until this conflict ends”. At best we can only provide the details of oppression and destruction in terms of simpler graphs for viewers to consume.

Data Sources

https://statistics.btselem.org/en/intro/demolitions

https://statistics.btselem.org/en/intro/fatalities

References

https://en.wikipedia.org/wiki/2014_Gaza_War

https://en.wikipedia.org/wiki/Gaza_War_(2008-2009)

https://en.wikipedia.org/wiki/Timeline_of_the_Israeli-Palestinian_conflict_in_2002