STA 478 Final Project

Jared Brotamonte

12/6/2023

Introduction

For this project, I decided to look into the Isreali Palestinian conflict, or otherly known as of as the Isreali Hamas war. The Isreale Palestinian conflict is a long-standing political and territorial dispute over the control of Jerusilum that can be dated back to 1947.

This topic was chosen due to it's relevancy within the current world. As of late, the conflict between the Isrealis and Palestinians seems to only get worse day by day. I felt it was important to look into this topic because it's important to stay educated on the reality of the situation not just in the political sense but in the statistical sense as well. Unfortunatly, the media coverage on this topic can found as both lacking and in many cases extremely biased thus making it hard to truly understand the situation at hand. I hope that through my analysis of the data I can begin to understand the truth of the situation and begin to form an actual informed opinion on this topic.

For this project I plan to take a look into the "Fatalities in the Israeli-Palestinian" data set obtained from kaggle.com. As the title suggests the data will take a look into the statistics of the numerous deaths that have occured throughout the Isreali Palestinian conflict from 2000 to as late as September 24th of 2023. The dataset has a total of 11124 observations with detailed information on not only the event of the death but the individual as well. All of the variables except age are characters/strings but that makes sense given the information the dataset is trying to convey.

```
data <- read.csv("fatalities_isr_pse_conflict_2000_to_2023.csv")
summary(data)
```

```
##      name           date_of_event           age         citizenship
##  Length:11124       Length:11124       Min.   :  1.00   Length:11124
##  Class :character   Class :character   1st Qu.: 19.00   Class :character
##  Mode  :character   Mode  :character   Median : 23.00   Mode  :character
##                                        Mean   : 26.75
##                                        3rd Qu.: 31.00
##                                        Max.   :112.00
##                                        NA's   :129
##  event_location     event_location_district event_location_region
##  Length:11124       Length:11124            Length:11124
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##  date_of_death         gender          took_part_in_the_hostilities
```

```
##  Length:11124       Length:11124       Length:11124
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  place_of_residence place_of_residence_district type_of_injury
##  Length:11124       Length:11124                Length:11124
##  Class :character   Class :character            Class :character
##  Mode  :character   Mode  :character            Mode  :character
##
##
##
##
##   ammunition         killed_by            notes
##  Length:11124       Length:11124       Length:11124
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
```

The dataset includes personal information on each of the fatalities through the variables name, age, citizenship, gender, whether they took part of the hostiles or not, where they where from, and some notes that go more into detail on each death. The variables name and notes will most likely not be useful for analyzing statistically but I felt like it's important to not completely disregard them because they are a reminder that these are not just a statistic in a data set, but the unfortunate deaths that have occurred in a violent war. The rest of the variables however will be used as they provide important information on each individual fatality that can be used to analyze and understand the conflict.

```
selected_columns <- c("age","citizenship","gender","took_part_in_the_hostilities",
                      "place_of_residence")
selected_data <- data[selected_columns]
head(selected_data)
```

```
##   age citizenship gender took_part_in_the_hostilities place_of_residence
## 1  32 Palestinian      M                                    Nur Shams R.C.
## 2  21 Palestinian      M                                    Nur Shams R.C.
## 3  16 Palestinian      M                                          al-Yamun
## 4  19 Palestinian      M                                 'Aqbat Jaber R.C.
## 5  15 Palestinian      M                                             Jenin
## 6  29 Palestinian      M                                             Jenin
```

The data set also contains important information of the event of the fatality as well including the date in which the event occurred, the location of the event, the date in which the death occurred, the type of injury sustained, the type of ammunition used, and who they were killed by. The previously mentioned variables should all be important for analyzing the data and understanding the conflict at hand.

```
selected_columns <- c("date_of_event","event_location","date_of_death","type_of_injury",
                      "ammunition","killed_by")
selected_data <- data[selected_columns]
head(selected_data)
```

```
##   date_of_event    event_location date_of_death type_of_injury      ammunition
## 1    2023-09-24    Nur Shams R.C.    2023-09-24        gunfire live ammunition
## 2    2023-09-24    Nur Shams R.C.    2023-09-24        gunfire live ammunition
## 3    2023-09-22          Kfar Dan    2023-09-22        gunfire live ammunition
## 4    2023-09-20 'Aqbat Jaber R.C.    2023-09-20        gunfire live ammunition
## 5    2023-09-19        Jenin R.C.    2023-09-19        gunfire live ammunition
## 6    2023-09-19        Jenin R.C.    2023-09-20        gunfire         missile
##                killed_by
## 1 Israeli security forces
## 2 Israeli security forces
## 3 Israeli security forces
## 4 Israeli security forces
## 5 Israeli security forces
## 6 Israeli security forces
```

Data Evaluation