

EDA_Police Shootings

2024-09-23

Introduction

This data was taken from a file compiled by the Washington Post about police shootings in 2015. There are 3960 observations of 14 different variables in this data set. The variables I want to explore and compare in this document are age and threat level. I will hypothesize that the average age of the victims will be around 30 yrs old, and will be older for the "other" category.

Data Summaries

Investigating the variables threat_level and age in the data set washpost

```
table(washpost$threat_level)
```

```
##
##      attack      other undetermined
##      2497      1255      208
```

```
prop.table(table(washpost$threat_level, useNA = "always"))
```

```
##
##      attack      other undetermined      <NA>
## 0.63055556 0.31691919 0.05252525 0.00000000
```

The percentage of victims that were in the attack threat level was about 63%. There was around 32% in the other category and 5% were undetermined.

```
table(washpost$age, useNA = "always")
```

```
##
##  6  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26
##  2   1   1   2  12  25  34  77  61  66  76  84 105 125 150 113
## 27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42
## 126 107 123 110 124 122 130 119 114 130 106  92  88  82  84  65
##  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58
##  64  66  84  68  56  65  56  68  54  43  55  48  47  44  31  31
##  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74
##  47  24  21  23  21  18  17  11  10  10  9  12  10  6  3  2
##  75  76  77  78  79  80  81  82  83  84  86  89  91 <NA>
##  2   7   3   1   1   1   2   2   2   3   2   1   1  152
```

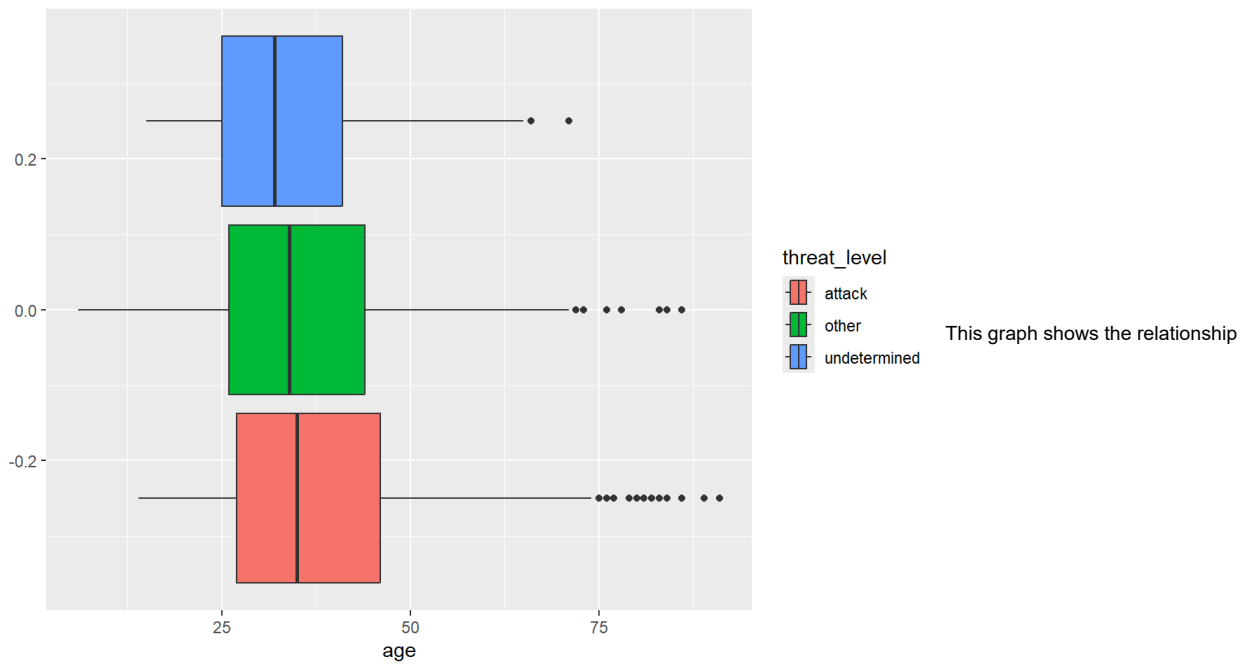
```
summary(washpost$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      6.00  27.00  35.00  36.85  45.00  91.00   152
```

Out of 3960 observations of age in this data set, excluding the 152 Na's, the mean age was 36.85, the median age was 35, the youngest was 6 yrs old and the oldest was 91.

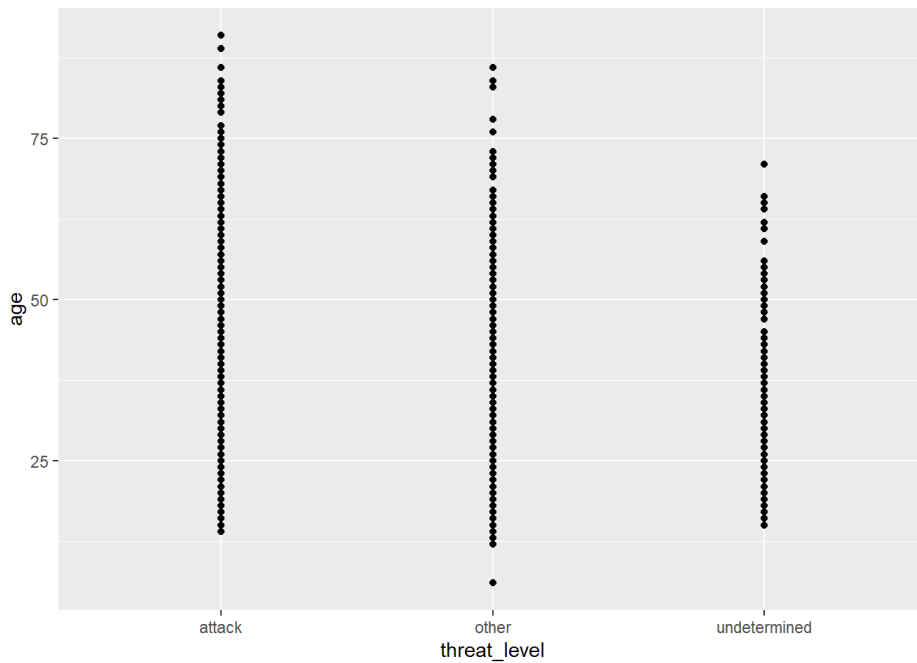
Bivariate Analysis

```
## Warning: Removed 152 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



```
ggplot(washpost, aes(y = age, x = threat_level)) + geom_point()
```

```
## Warning: Removed 152 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Conclusion

My hypothesis that the attack threat level would have an average around 30 yrs old was false, and also that the other and undetermined categories would have higher average ages was false as well. The attack category had a higher average age, and the oldest outliers.