EDA_raunakchouhan

2025-06-27

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

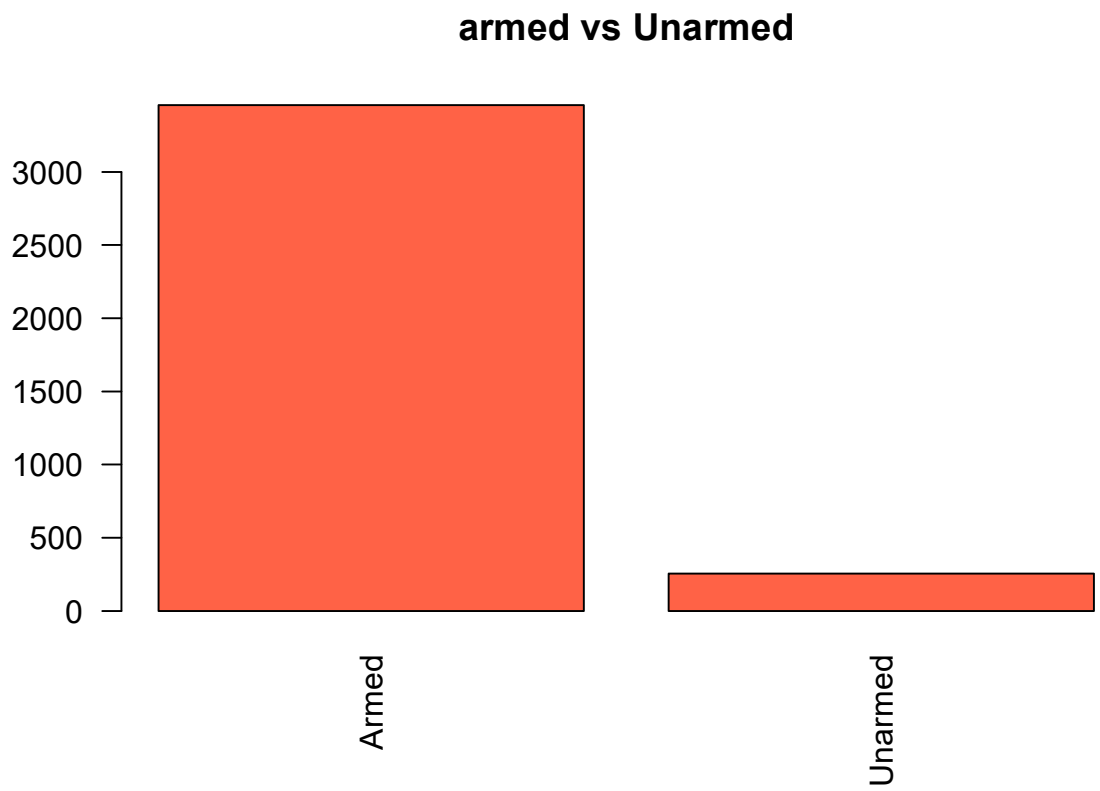
```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

#INTRODUCTION This project explores the data of Police shootings dataset, that has the data about number of people got killed by police in the year 2015. We will be comparing 3 variables: 1. If the person was armed. 2. The age of the person when he got killed. 3. Manner of death Research question - Do unarmed individuals differ in average age or manner of death compared to those who were armed? Univariate exploration #Variable – armed

```
table(police_clean$armed_status)
```

```
##
##   Armed Unarmed
##    3455     256
```

```
barplot(table(police_clean$armed_status),
        main = "armed vs Unarmed",
        col = "tomato",
        las = 2)
```



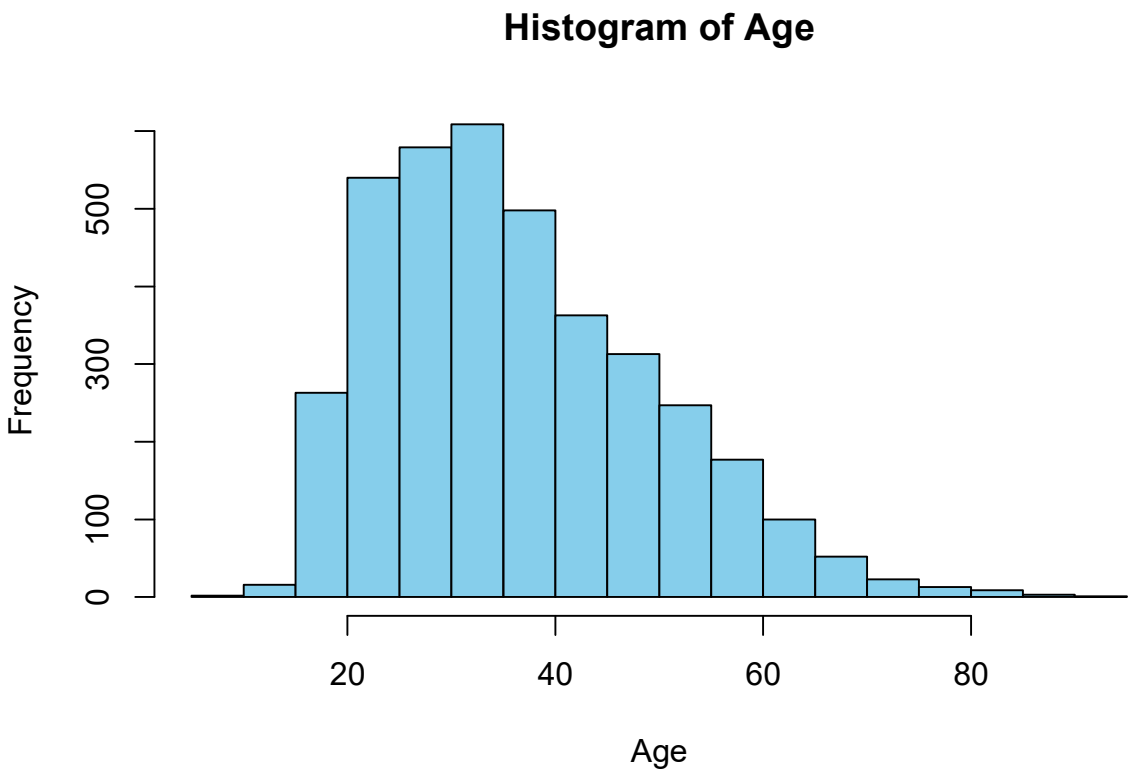
#DESCRIPTION There are different types of weapons given but we are only interested in looking unarmed v.s armed with something. Most of the people were armed with weapons like gun but a notable number were unarmed.

Variable – Age

```
summary(police_clean$age)
```

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

```
hist(police_clean$age,
     main = "Histogram of Age",
     xlab = "Age",
     col = "skyblue",
     breaks = 20)
```



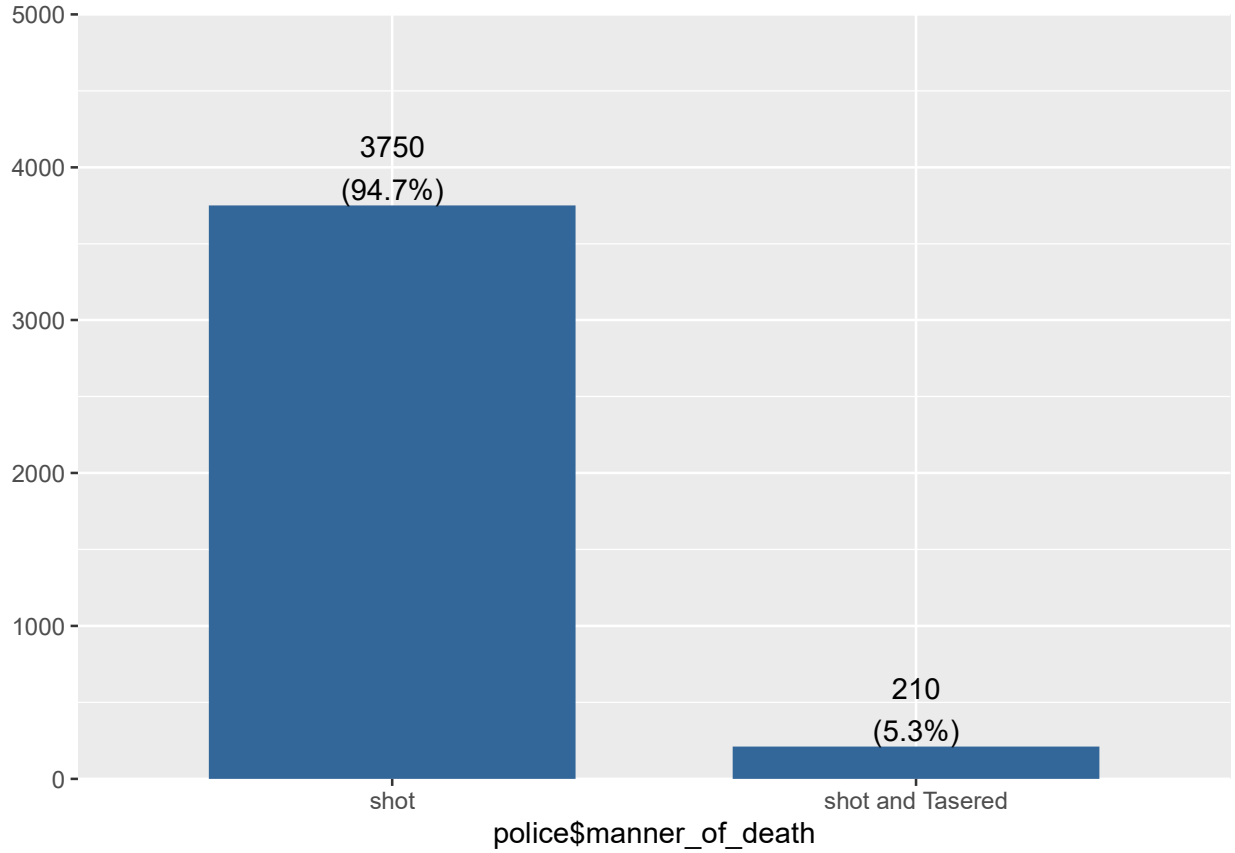
#Description There were people belonging to different age groups but most of them were between 20 and 40. The mean age is around mid 30s.

#Variable- Manner of death

```
table(police$manner_of_death)
```

```
##
##             shot shot and Tasered
##             3750              210
```

```
plot_frq(police$manner_of_death)
```



#Description Most of the deaths that happened were by gunshot. Some of them are also listed as shot and tasered.

#Bivariate exploration Relation between armed and age variable.

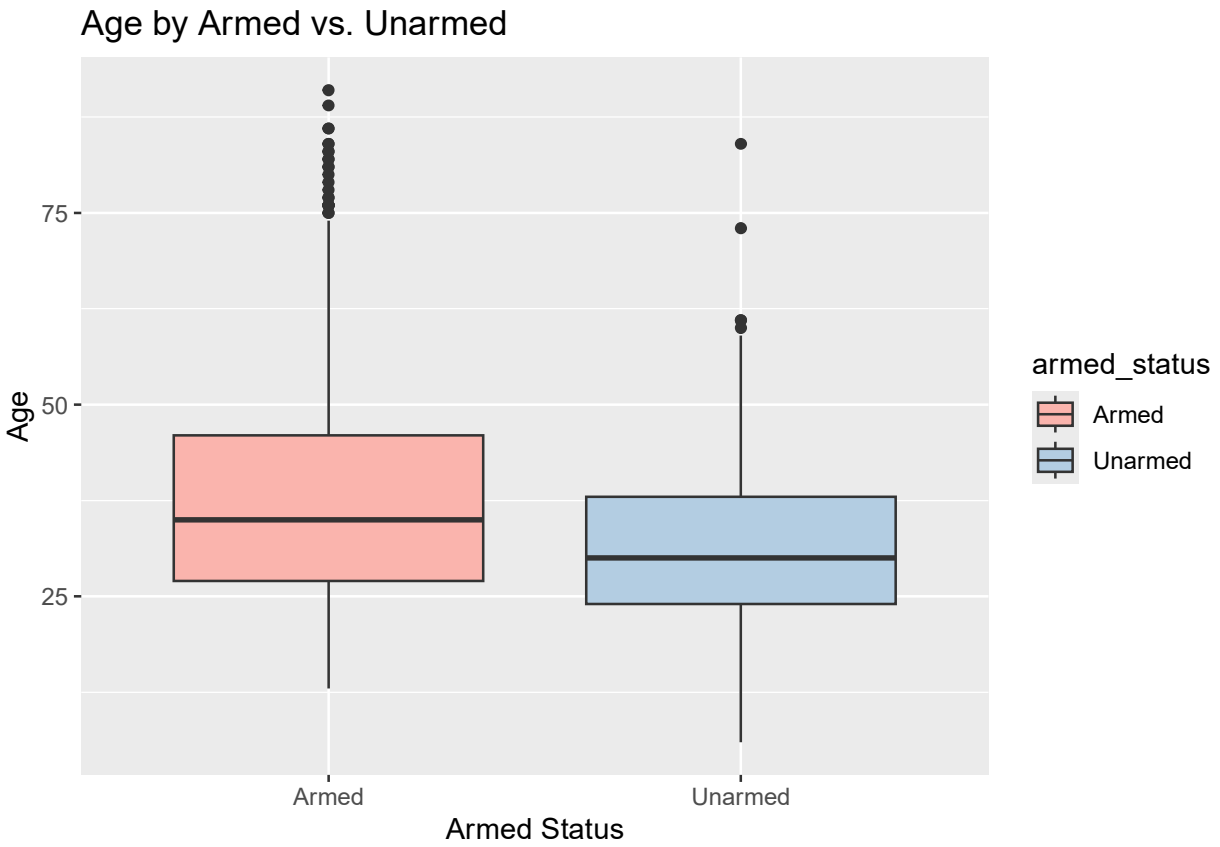
```
police_clean <- police %>% filter(!is.na(age), !is.na(armed))

police_clean <- police_clean %>%
  mutate(armed_status = ifelse(armed == "unarmed", "Unarmed", "Armed"))

police_clean %>%
  group_by(armed_status) %>%
  summarise(mean_age = mean(age, na.rm = TRUE),
            sd_age = sd(age, na.rm = TRUE),
            count = n())
```

```
## # A tibble: 2 x 4
##   armed_status mean_age sd_age count
##   <chr>           <dbl>  <dbl> <int>
## 1 Armed            37.5   13.2  3314
## 2 Unarmed          32.1   11.6   252
```

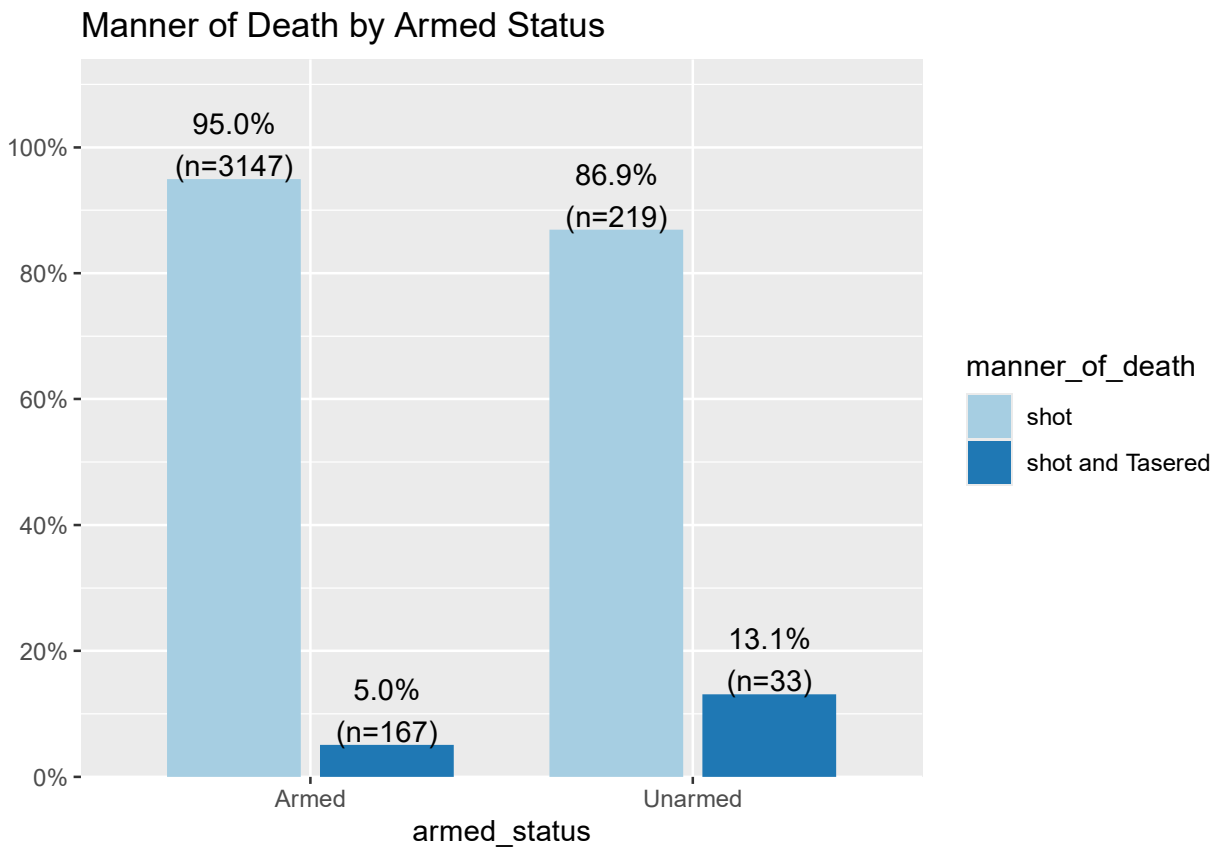
```
ggplot(police_clean, aes(x=armed_status, y=age, fill=armed_status)) +
  geom_boxplot() +
  ylab("Age") + xlab("Armed Status") +
  ggtitle("Age by Armed vs. Unarmed") +
  scale_fill_brewer(palette = "Pastel1")
```



#Description The unarmed category has lower average age than the armed category. The difference is also somewhat narrower for unarmed individual.

#Relation between Manner of death and Armed status

```
plot_xtab(police_clean$armed_status, police_clean$manner_of_death,
          margin = "row", show.total = FALSE,
          title = "Manner of Death by Armed Status")
```



#Description Most of the individuals belonging to both categories were killed by gunshot. The unarmed category has more individuals listed under shot and taresed category as compared to the other one.

#Conclusion: Most were armed, but many were not.

The unarmed victims had a mean slightly less than age.

The cause of death (usually a shooting) was almost the same with or without a weapon, though rare cases of "shot and Tasered" occurred slightly more to the unarmed.

Note: This is descriptive analysis only. We cannot make any causal inferences or claims. Further statistical testing would be needed to validate any claims in the real world.