Exploratory Data Analysis project

Gaberial-Angelo Marquez Hernandez

6/27/2025()`

Introduction

The Washington Post Fatal Police Shootings data set contains data on individuals who were fatally shot by police officers in the United States. It includes information on demographic details, whether the person was armed, mental illness, race, and the circumstances surrounding each incident.For this analysis, I will explore the following variables:

armed: Whether the individual was armed or unarmed.

race: Race/ethnicity of the individual.

signs of mental illness: Whether the individual showed signs of mental illness.

Research Question: Are there patterns in who was armed versus unarmed based on race or mental illness status?

Loading data

```
washpost <-read_excel("fatal-police-shootings-data.xlsx")
dim(washpost)
```

```
## [1] 3960   14
```

Check variable types

```
str(washpost)
```

```
## tibble [3,960 × 14] (S3: tbl_df/tbl/data.frame)
##  $ id                     : num [1:3960] 3 4 5 8 9 11 13 15 16 17 ...
##  $ name                   : chr [1:3960] "Tim Elliot" "Lewis Lee Lembke" "John Paul Quintero"
"Matthew Hoffman" ...
##  $ date                   : POSIXct[1:3960], format: "2015-01-02" "2015-01-02" ...
##  $ manner_of_death        : chr [1:3960] "shot" "shot" "shot and Tasered" "shot" ...
##  $ armed                  : chr [1:3960] "gun" "gun" "unarmed" "toy weapon" ...
##  $ age                    : num [1:3960] 53 47 23 32 39 18 22 35 34 47 ...
##  $ gender                 : chr [1:3960] "M" "M" "M" "M" ...
##  $ race                   : chr [1:3960] "A" "W" "H" "W" ...
##  $ city                   : chr [1:3960] "Shelton" "Aloha" "Wichita" "San Francisco" ...
##  $ state                  : chr [1:3960] "WA" "OR" "KS" "CA" ...
##  $ signs_of_mental_illness: logi [1:3960] TRUE FALSE FALSE TRUE FALSE FALSE ...
##  $ threat_level           : chr [1:3960] "attack" "attack" "other" "attack" ...
##  $ flee                   : chr [1:3960] "Not fleeing" "Not fleeing" "Not fleeing" "Not fleei
ng" ...
##  $ body_camera            : logi [1:3960] FALSE FALSE FALSE FALSE FALSE FALSE ...
```

Univariate Exploration

Age

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

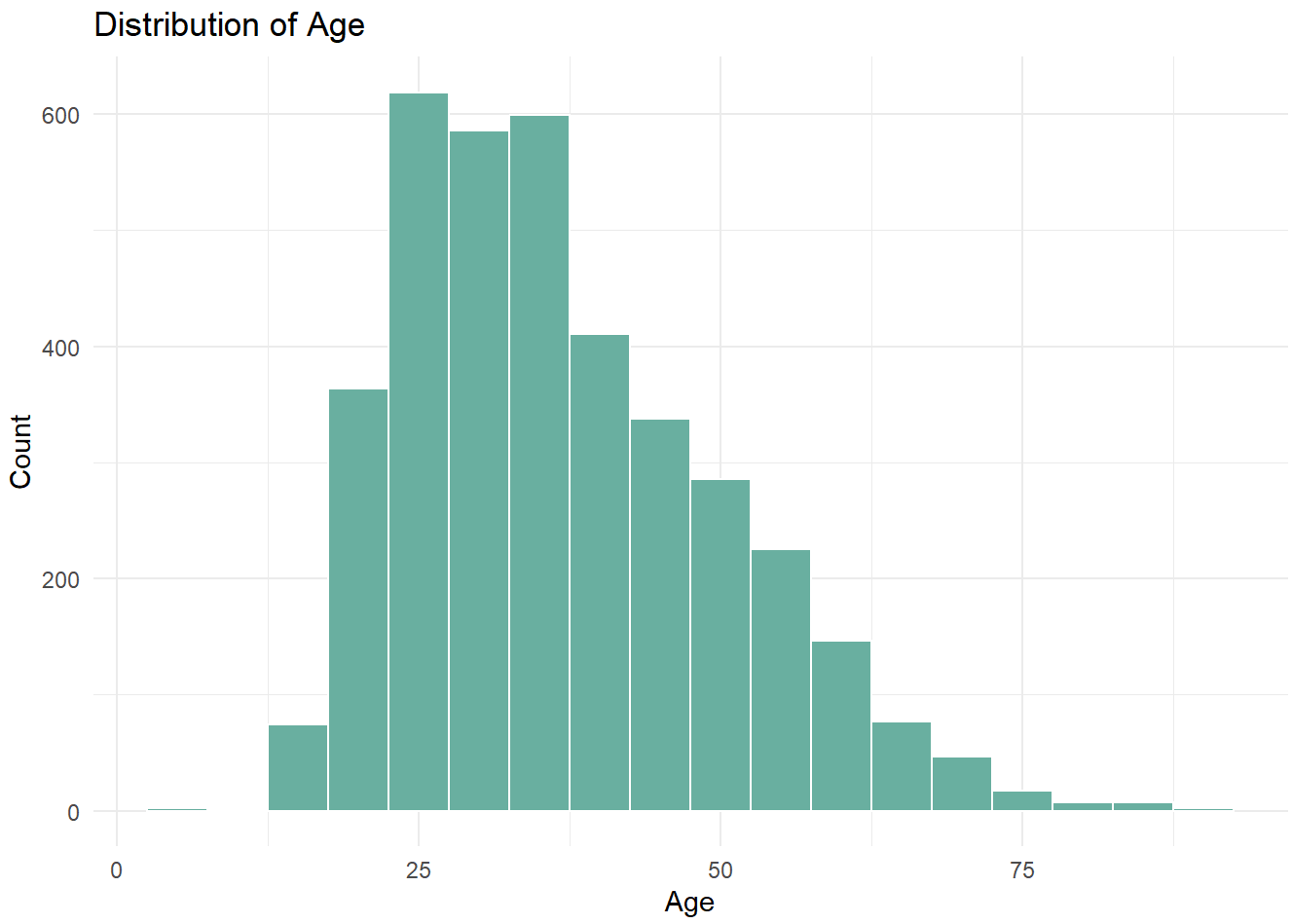
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.00   27.00   35.00   36.85   45.00   91.00     152
```

```
sd(washpost$age, na.rm = TRUE)
```

```
## [1] 13.09016
```

```
ggplot(washpost, aes(x = age)) +
  geom_histogram(binwidth = 5, fill = "#69b3a2", color = "white") +
  labs(title = "Distribution of Age", x = "Age", y = "Count") +
  theme_minimal()
```

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



The mean age is r round(mean(washpost $age, na.rm = TRUE), 1), $with a standard deviation of r round(sd(washpost$age, na.rm = TRUE), 1). Most victims are between 20 and 40 years old.

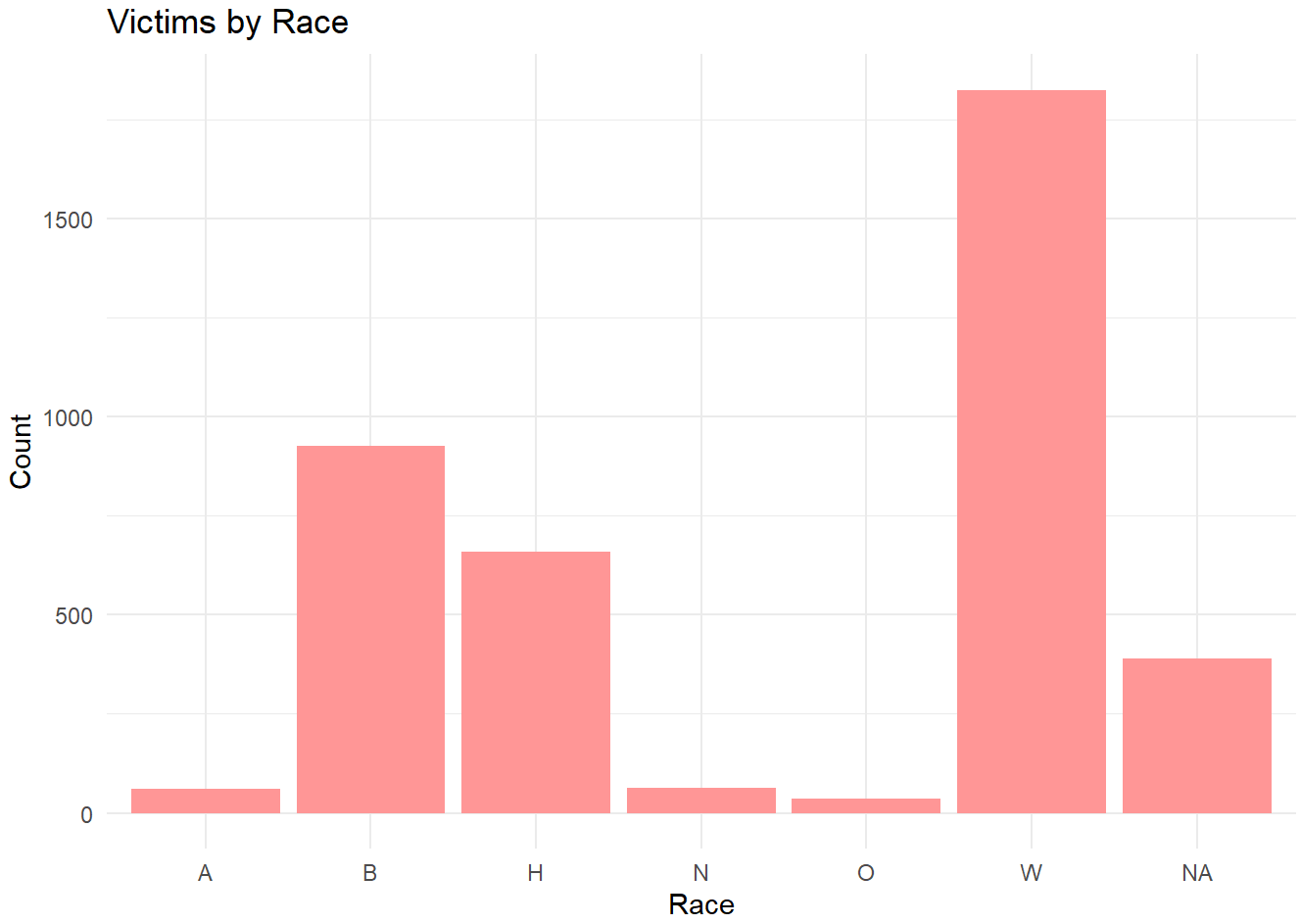
Race

```
washpost %>%
  count(race) %>%
  mutate(percent = n / sum(n))
```

race <chr>	n <int>	percent <dbl>
A	61	0.015404040
B	927	0.234090909
H	659	0.166414141
N	62	0.015656566
O	37	0.009343434
W	1825	0.460858586
NA	389	0.098232323

7 rows

```
ggplot(washpost, aes(x = race)) +
  geom_bar(fill = "#FF9999") +
  labs(title = "Victims by Race", x = "Race", y = "Count") +
  theme_minimal()
```



Most victims are White (W), followed by Black (B) and Hispanic (H) individuals. There are smaller counts for Asian (A), Native American (N), and Other/Unknown (O) categories.

Armed Status

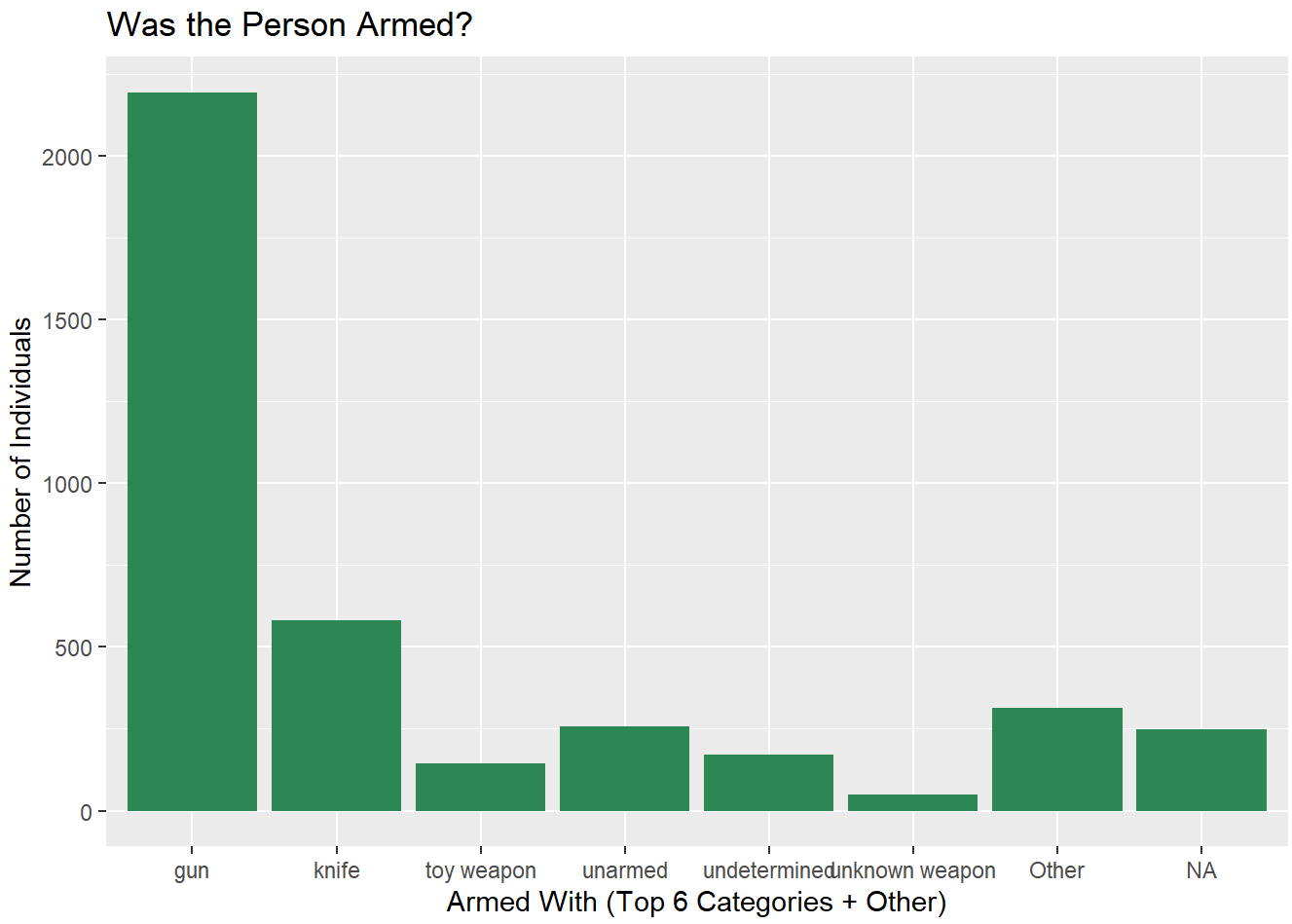
```
washpost <- read_excel("fatal-police-shootings-data.xlsx")

washpost %>%
  count(armed) %>%
  mutate(percent = round(n / sum(n) * 100, 1))
```

armed <chr>	n <int>	percent <dbl>
BB gun	2	0.1
Taser	17	0.4
air conditioner	1	0.0
ax	17	0.4

armed <chr>	n <int>	percent <dbl>
baseball bat	11	0.3
baseball bat and bottle	1	0.0
baseball bat and fireplace poker	1	0.0
baton	4	0.1
bayonet	1	0.0
bean-bag gun	1	0.0

1-10 of 82 rows Previous **1** 2 3 4 5 6 ... 9 Next

```
ggplot(washpost, aes(x = fct_lump(armed, 6))) +
  geom_bar(fill = "#2E8B57") +
  labs(title = "Was the Person Armed?",
       x = "Armed With (Top 6 Categories + Other)",
       y = "Number of Individuals")
```



Bivariate Exploration

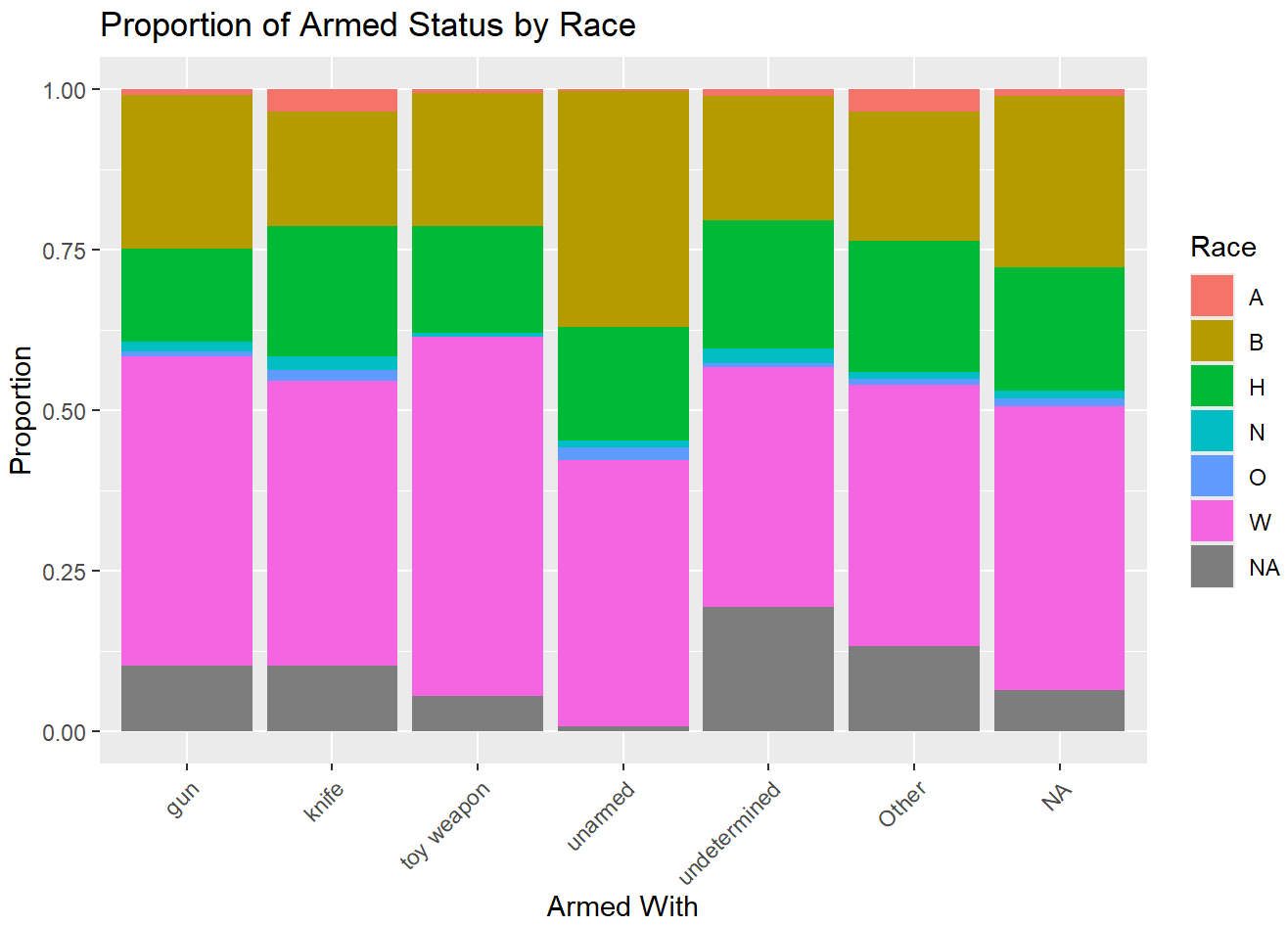
```
washpost %>%
  filter(!is.na(race)) %>%
  group_by(race, armed) %>%
  summarize(n = n()) %>%
  mutate(percent = n / sum(n) * 100) %>%
  arrange(race, desc(percent))
```

```
## `summarise()` has grouped output by 'race'. You can override using the
## `.groups` argument.
```

race <chr>	armed <chr>	n <int>	percent <dbl>
A	knife	21	34.42622951
A	gun	20	32.78688525
A	*NA*	3	4.91803279
A	ax	2	3.27868852
A	gun and knife	2	3.27868852
A	undetermined	2	3.27868852
A	unknown weapon	2	3.27868852
A	hammer	1	1.63934426
A	hatchet and gun	1	1.63934426
A	incendiary device	1	1.63934426

1-10 of 171 rows Previous **1** 2 3 4 5 6 … 18 Next

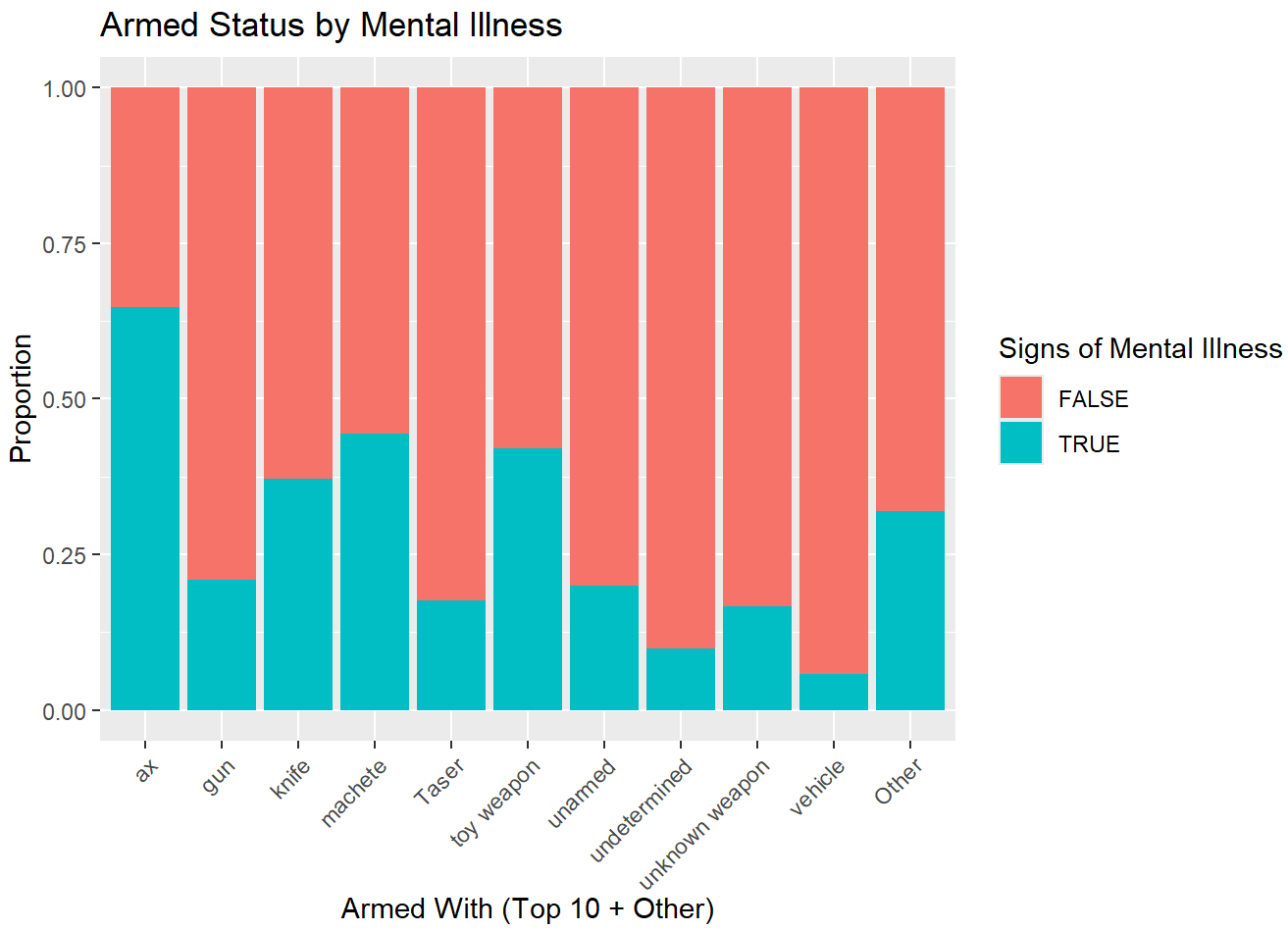
```
ggplot(washpost, aes(x = fct_lump(armed, 5), fill = race)) +
  geom_bar(position = "fill") +
  labs(title = "Proportion of Armed Status by Race",
       x = "Armed With",
       y = "Proportion",
       fill = "Race") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



The plot suggests that there may be differences in the type of weapon individuals were reported to be armed with depending on race.

Armed Status by Mental Illness

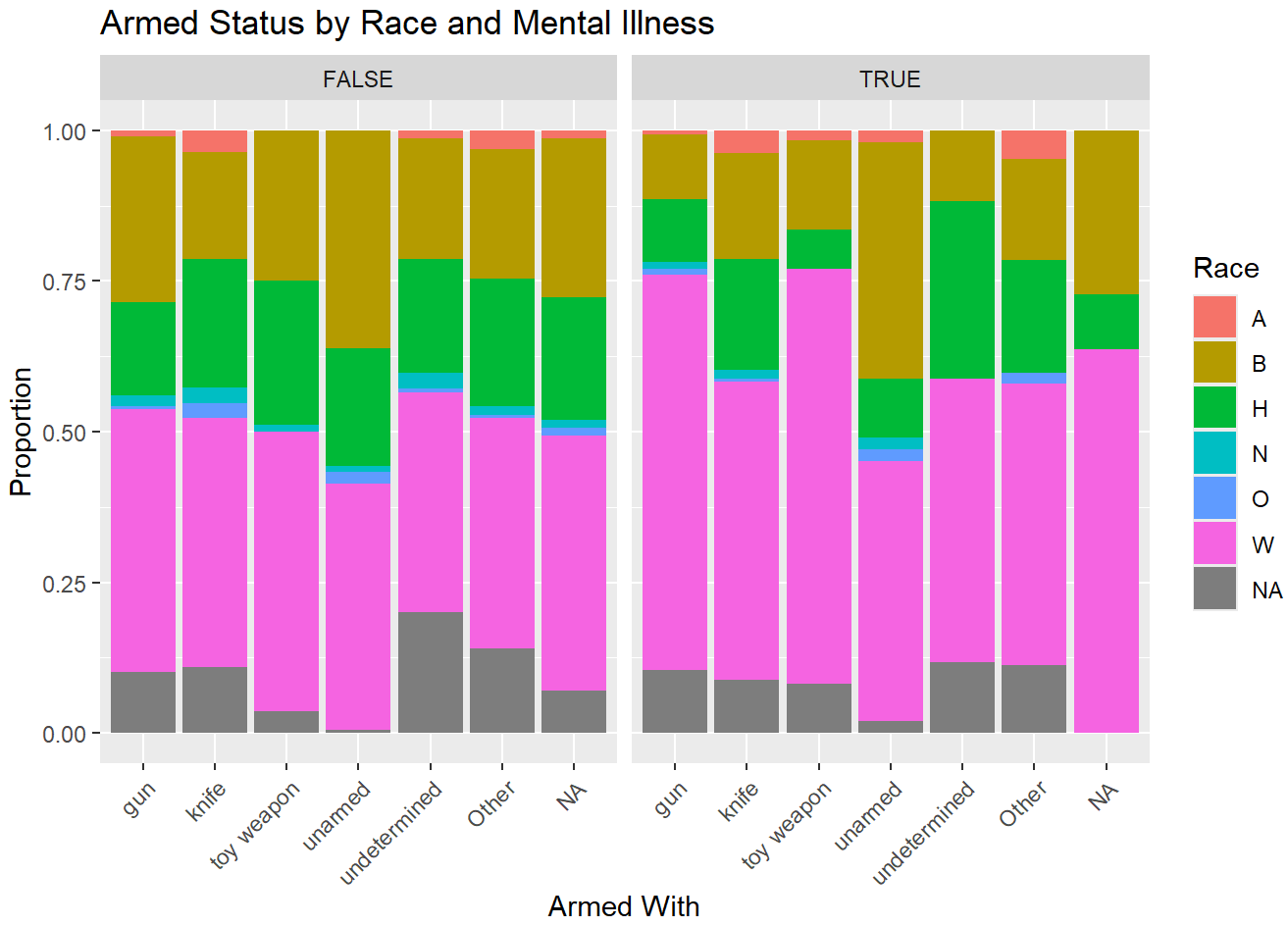
```
library(forcats)

washpost_clean <- washpost %>%
  filter(!is.na(armed), !is.na(signs_of_mental_illness)) %>%
  mutate(armed_grouped = fct_lump(armed, n = 10))  # Top 10 + "Other"

ggplot(washpost_clean, aes(x = armed_grouped, fill = signs_of_mental_illness)) +
  geom_bar(position = "fill") +
  labs(title = "Armed Status by Mental Illness",
       x = "Armed With (Top 10 + Other)",
       y = "Proportion",
       fill = "Signs of Mental Illness") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Individuals with signs of mental illness may be more likely to be unarmed or armed with non-lethal objects, though further investigation is needed.

3-Way Panel: Armed by Race and Mental Illness

```
ggplot(washpost, aes(x = fct_lump(armed, 5), fill = race)) +
  geom_bar(position = "fill") +
  facet_wrap(~ signs_of_mental_illness) +
  labs(title = "Armed Status by Race and Mental Illness",
       x = "Armed With",
       y = "Proportion",
       fill = "Race") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



This breakdown helps explore whether mental illness status changes the racial pattern in armed status.

Conclusion:

From this exploratory analysis, we observe: Race and armed status appear to have some pattern—certain racial groups may be more likely reported as “unarmed.” Mental illness status seems to relate to the nature of the armed status, with potentially higher representation of non-lethal items among those with mental illness. The combined interaction of race and mental illness shows that intersectionality may matter when evaluating these tragic incidents. These findings are exploratory and descriptive. They raise questions for future research and emphasize the importance of nuanced, contextual data analysis in discussions of police use of force.