USA SHOOTINGS

Code

Details about this project can be found in the readme file

Hide

```
getwd()
```

```
[1] "C:/Users/Michael/Documents/R Projects"
```

import the required libraries

Hide

```
library(dplyr)
library(tidyr)
library(readr)
library(ggplot2)
library(lubridate)
library(lattice)

print("Setup Complete!")
```

```
[1] "Setup Complete!"
```

Import the Data-set

Hide

```
filepath <- "F:/Data Science/My Datasets/USA_shootings.csv"
shootings <- read.csv(filepath)
print("Import Successful")
```

```
[1] "Import Successful"
```

EXPLORATORY DATA ANALYSIS

Hide

```
head(shootings)
```

	id	name	date	manner_of_death	armed	…	gen…	race	city
	<int>	<chr>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<ch
1	3	Tim Elliot	2015-01-02	shot	gun	53	M	Asian	She
2	4	Lewis Lee Lembke	2015-01-02	shot	gun	47	M	White	Alo
3	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23	M	Hispanic	Wi

	id	name	date	manner_of_death	armed	…	gen…	race	city
	<int>	<chr>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<ch
4	8	Matthew Hoffman	2015-01-04	shot	toy weapon	32	M	White	San
5	9	Michael Rodriguez	2015-01-04	shot	nail gun	39	M	Hispanic	Eva
6	11	Kenneth Joe Brown	2015-01-04	shot	gun	18	M	White	Gut

6 rows | 1-10 of 15 columns

Hide

```
dim(shootings)
```

```
[1] 4895   15
```

- There are 4895 victims and 10 observations surrounding their death

Addition of the Age group column

Hide

```
shootings <- shootings%>% mutate(Age_group = case_when(age > 0 & age <= 12 ~ "Child",
                                                       age >= 13 & age <= 19 ~ "Teenag
er",
                                                       age >= 20 & age <= 40 ~ "Young
 Adult",
                                                       age >= 41 & age <= 65 ~ "Adult"
,
                                                       age > 65 ~ "Senior"))
```

Now lets check if the column has been added

Hide

```
head(shootings)
```

	id	name	date	manner_of_death	armed	…	gen…	race	city
	<int>	<chr>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<ch
1	3	Tim Elliot	2015-01-02	shot	gun	53	M	Asian	She
2	4	Lewis Lee Lembke	2015-01-02	shot	gun	47	M	White	Alo
3	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23	M	Hispanic	Wic
4	8	Matthew Hoffman	2015-01-04	shot	toy weapon	32	M	White	San
5	9	Michael Rodriguez	2015-01-04	shot	nail gun	39	M	Hispanic	Eva
6	11	Kenneth Joe Brown	2015-01-04	shot	gun	18	M	White	Gut

6 rows | 1-10 of 16 columns

Tadaaaaa>>>>

Conversion of columns to appropriate classes

Hide

```
shootings$manner_of_death <- as.factor(shootings$manner_of_death)
shootings$armed <- as.factor(shootings$armed)
shootings$date <- ymd(shootings$date)
shootings$gender <- as.factor(shootings$gender)
shootings$race <- as.factor(shootings$race)
shootings$city <- as.factor(shootings$city)
shootings$state <- as.factor(shootings$state)
shootings$signs_of_mental_illness <- as.factor(shootings$signs_of_mental_illness)
shootings$threat_level <- as.factor(shootings$threat_level)
shootings$flee <- as.factor(shootings$flee)
shootings$body_camera <- as.factor(shootings$body_camera)
shootings$arms_category <- as.factor(shootings$arms_category)
```

Lets check if the classes of the columns have been changed

Hide

```
head(shootings)
```

	id	name	date	manner_of_death	armed	…	gen…	race	city
	<int>	<chr>	<date>	<fctr>	<fctr>	<dbl>	<fctr>	<fctr>	<fc
1	3	Tim Elliot	2015-01-02	shot	gun	53	M	Asian	She
2	4	Lewis Lee Lembke	2015-01-02	shot	gun	47	M	White	Alo
3	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23	M	Hispanic	Wi
4	8	Matthew Hoffman	2015-01-04	shot	toy weapon	32	M	White	San
5	9	Michael Rodriguez	2015-01-04	shot	nail gun	39	M	Hispanic	Eva
6	11	Kenneth Joe Brown	2015-01-04	shot	gun	18	M	White	Gu

6 rows | 1-10 of 16 columns

Manner Of Death

Hide

```
mann.tbl <- shootings %>% group_by(manner_of_death) %>% summarize(`Death Toll` = n())
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

Hide

```
mann.tbl <- rename(mann.tbl, "Manner of Death" = "manner_of_death")
mann.tbl
```

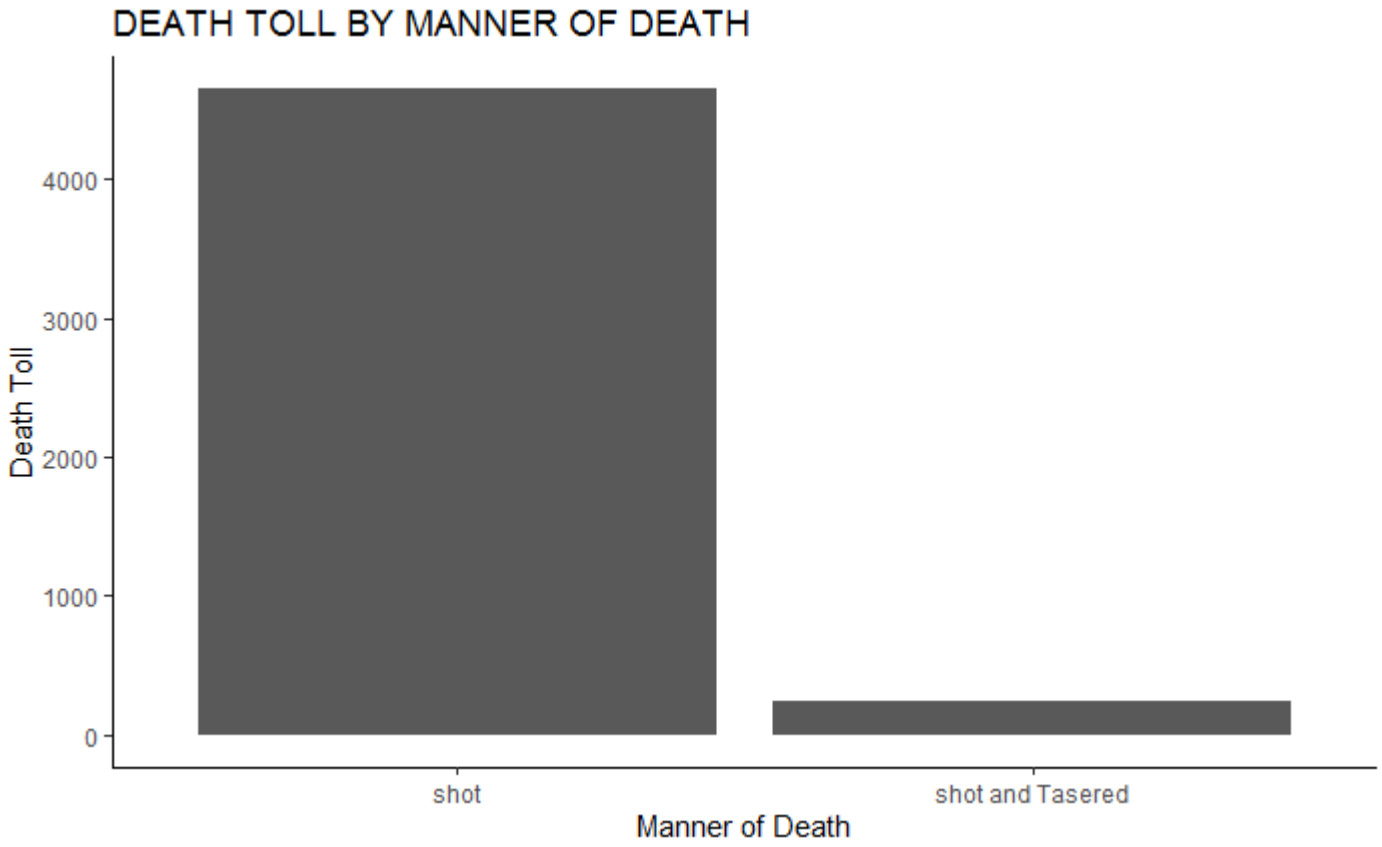
Manner of Death <fctr>	Death Toll <int>
shot	4647
shot and Tasered	248

2 rows

- As seen above, only few of the victims were shot and tasered

Hide

```
g <- ggplot(mann.tbl, aes(x = `Manner of Death`, y = `Death Toll`))
g + geom_bar(stat = "identity") + ggtitle("DEATH TOLL BY MANNER OF DEATH") + theme_classic()
```



- The chart above clearly shows the wide difference between individuals that were only shot and individuals that were shot and tasered

Armed

Hide

```
armed_grp <- table(shootings$armed)
armed_grp <- as.data.frame(armed_grp)
armed_grp <- rename(armed_grp, "Weapons" = "Var1", "Victims Count" = "Freq")
#new_column_name = old_column_name
armed_grp <- armed_grp %>% arrange(desc(`Victims Count`)) %>% mutate(Percentage = (`Victims Coun
t`/4895) * 100)
armed_grp$Percentage <- round(armed_grp$Percentage, digits = 1)
armed_grp
```

Weapons <fctr>	Victims Count <int>	Percentage <dbl>
gun	2755	56.3
knife	708	14.5
unknown	418	8.5
unarmed	348	7.1
toy weapon	171	3.5
vehicle	120	2.5
machete	39	0.8
Taser	24	0.5
sword	22	0.4
ax	21	0.4

1-10 of 89 rows Previous **1** 2 3 4 5 6 … 9 Next

- As seen above most of the victims were armed with Guns (56%) and Knives (14.1%). The victims might have tried to shoot at or stab the police officer before being shot dead

Lets move further to uncover more insights

Hide

```
guns <- subset(shootings, armed == "gun")
gun.grp <- guns%>% select(race) %>% group_by(race) %>% summarize(`Owned Gun` = n()) %>% arrange
(desc(`Owned Gun`))
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

Hide

```
print(gun.grp)
```

race <fctr>	Owned Gun <int>
White	1451

race <fctr>	Owned Gun <int>
Black	762
Hispanic	447
Native	41
Asian	35
Other	19

6 rows

- More white victims owned guns more than any other race in this data set

Gender and Race

Hide

```
gr_grp <- shootings %>% select(race, gender) %>% group_by(race, gender) %>% summarize(`Victim Counts` = n())
```

```
`summarise()` regrouping output by 'race' (override with `.groups` argument)
```

Hide

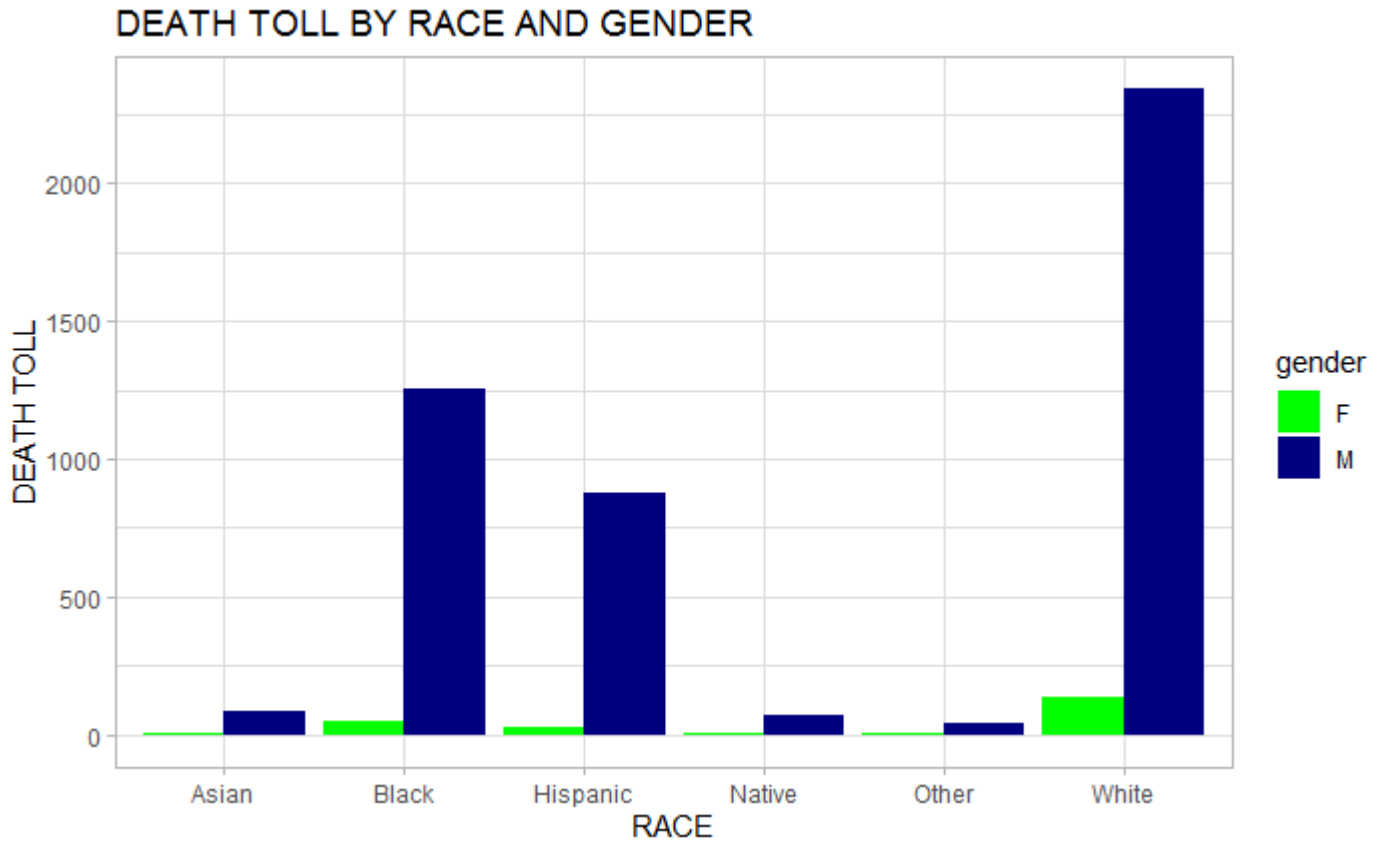
```
gr_grp
```

race <fctr>	gender <fctr>	Victim Counts <int>
Asian	F	4
Asian	M	89
Black	F	47
Black	M	1251
Hispanic	F	26
Hispanic	M	876
Native	F	5
Native	M	73
Other	F	4
Other	M	44

1-10 of 12 rows Previous **1** 2 Next

Hide

```
g <- ggplot(gr_grp, aes(x = race, y = `Victim Counts`))
g + geom_bar(aes(fill = gender), stat = "identity", position = position_dodge()) + xlab("RACE")
 + ylab("DEATH TOLL") + ggtitle("DEATH TOLL BY RACE AND GENDER") + scale_fill_manual(values = c(
"green", "navy blue")) + theme_light()
```



- It ia evident that police shot more white men and women than any other race as opposed to popular belief that blacks have been the more victims.

Lets dive deeper

City

Hide

```
city.grp <- shootings %>% select(city) %>% group_by(city) %>% summarize(`Death Toll` = n()) %>%
arrange(desc(`Death Toll`)) %>% mutate(Percentage = (`Death Toll`/4895) *100)
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

Hide

```
city.grp$Percentage <- round(city.grp$Percentage, digits = 1)
city.grp
```

city	Death Toll	Percentage
<fctr>	<int>	<dbl>

city <fctr>	Death Toll <int>	Percentage <dbl>
Los Angeles	78	1.6
Phoenix	66	1.3
Houston	51	1.0
Las Vegas	41	0.8
San Antonio	40	0.8
Chicago	38	0.8
Columbus	36	0.7
Albuquerque	32	0.7
Jacksonville	31	0.6
Oklahoma City	30	0.6

1-10 of 2,288 rows Previous **1** 2 3 4 5 6 … 100 Next

- The top three cities where victims were killed are Los Angeles, Phoenix and Houston. There are 2288 cities in the data-set

State

Hide

```
states_grp <- shootings %>% select(state) %>% group_by(state) %>% summarize(`Death Toll` = n()) %>% arrange(desc(`Death Toll`)) %>% mutate(Percentage = (`Death Toll`/4895) * 100)
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

Hide

```
states_grp$Percentage <- round(states_grp$Percentage, digits = 1)

states_grp
```

state <fctr>	Death Toll <int>	Percentage <dbl>
CA	701	14.3
TX	426	8.7
FL	324	6.6
AZ	222	4.5
CO	168	3.4

state <fctr>	Death Toll <int>	Percentage <dbl>
GA	161	3.3
OK	151	3.1
NC	148	3.0
OH	146	3.0
WA	126	2.6

1-10 of 51 rows Previous **1** 2 3 4 5 6 Next

- As seen above the top three states with the highest Death toll are California, Texas and Florida

Mental Status

Hide

```
mental.illness <-table(shootings$signs_of_mental_illness)
mental.illness
```

```
False  True 
 3792  1103 
```

- Most of the victims didn't show signs mental illnesses

Hide

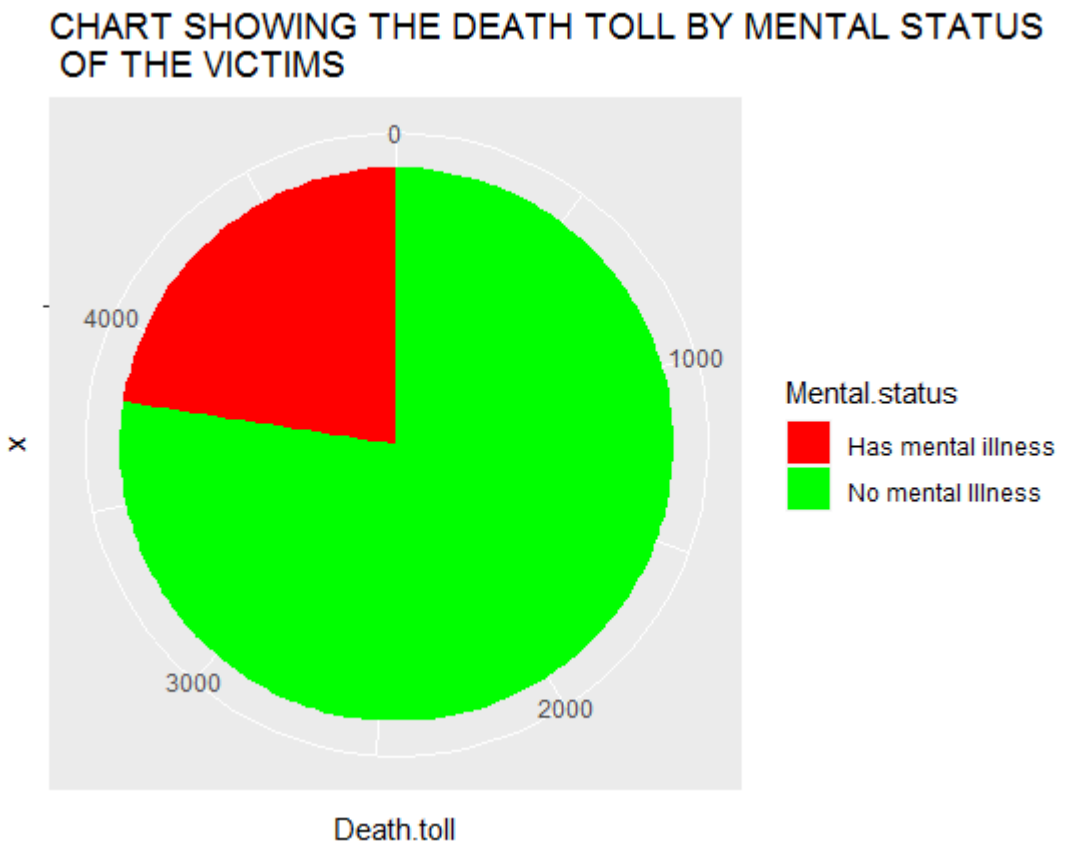
```
ment <- data.frame(`Mental status` = c("No mental Illness", "Has mental illness"), `Death toll`
 = c(3792, 1103))
ment
```

Mental.status <chr>	Death.toll <dbl>
No mental Illness	3792
Has mental illness	1103

2 rows

Hide

```
g <- ggplot(ment, aes(x = "", y = Death.toll, fill = Mental.status))
bar <- g + geom_bar(stat = "identity")
bar + coord_polar("y", start = 0) + scale_fill_manual(values = c("red", "green")) +ggtitle("CHAR
T SHOWING THE DEATH TOLL BY MENTAL STATUS \n OF THE VICTIMS")
```



Threat level

Hide

```
threat <- shootings %>% select(threat_level, race) %>%group_by(race,threat_level) %>%  summarize
(Death.toll = n()) %>% arrange(desc(Death.toll))
```

```
`summarise()` regrouping output by 'race' (override with `.groups` argument)
```

Hide

```
threat <- rename(threat, "Threat level" = "threat_level")
threat
```

race <fctr>	Threat level <fctr>	Death.toll <int>
White	attack	1640
Black	attack	873
White	other	743
Hispanic	attack	521
Black	other	364
Hispanic	other	333

race <fctr>	Threat level <fctr>	Death.toll <int>
White	undetermined	93
Black	undetermined	61
Asian	attack	51
Hispanic	undetermined	48

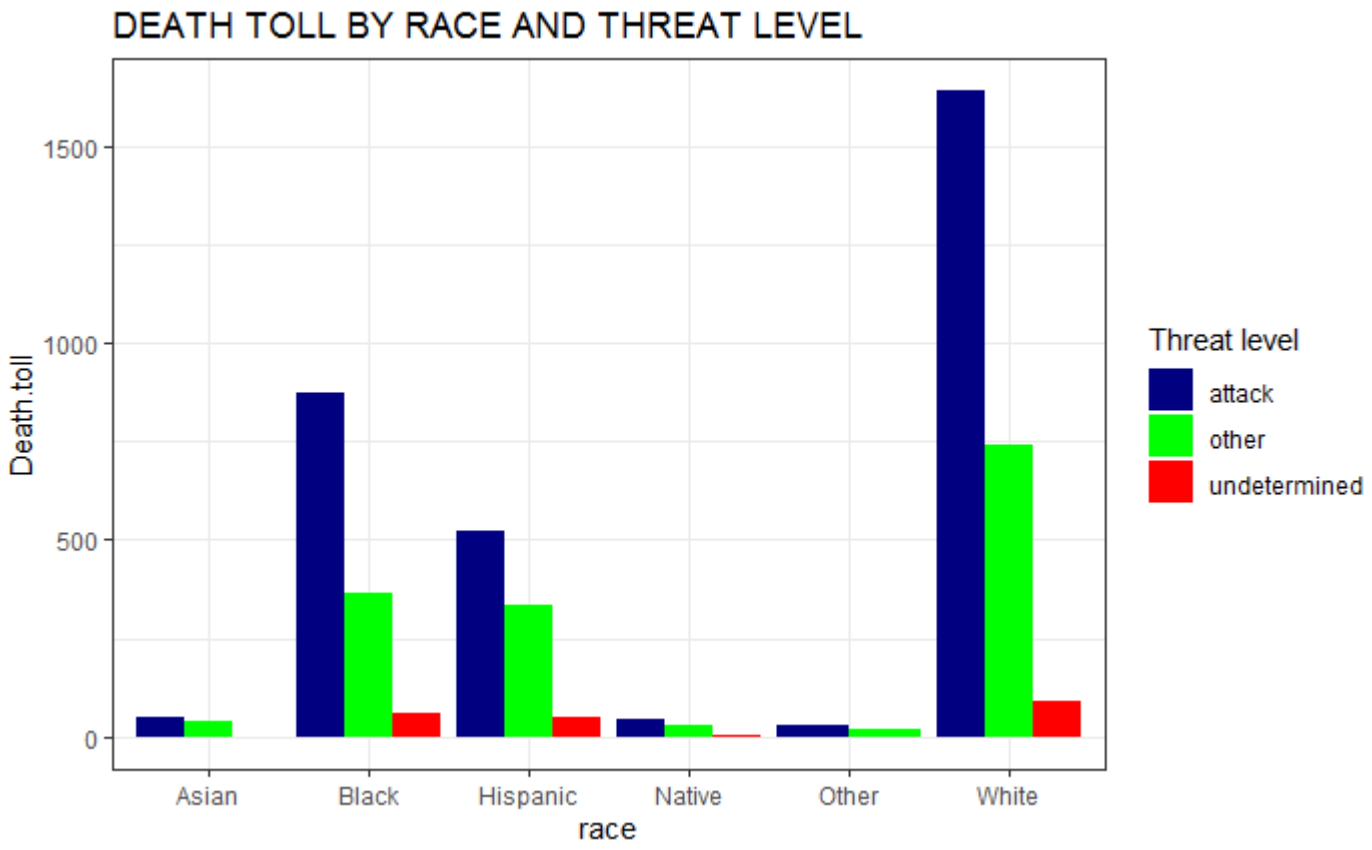
1-10 of 17 rows Previous **1** 2 Next

- Majority of the victims were a potential threat to the police hence they were shot

Lets see how the death toll by threat level regarding race in a chart below

Hide

```
g <- ggplot(threat, aes(x = race, y = Death.toll, fill = `Threat level`))
g + geom_bar(stat = "identity", position = position_dodge()) + scale_fill_manual(values = c("nav
y blue", "green", "red")) + ggtitle("DEATH TOLL BY RACE AND THREAT LEVEL") + theme_bw()
```



- The white race leads in all categories

Was the Victim Fleeing

Hide

```
flee <- shootings %>% select(flee) %>% group_by(flee) %>% summarize(`Death toll` = n()) %>% arra
nge(desc(`Death toll`))
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

Hide

```
flee <- rename(flee, "Fleeing status" = "flee")
flee
```

Fleeing status <fctr>	Death toll <int>
Not fleeing	3073
Car	820
Foot	642
Other	360

4 rows

- Most of the victims were not fleeing

Was the incident recorded on body camera

Hide

```
camera <- shootings %>% select(body_camera) %>% group_by(body_camera) %>% summarize(`Death toll`
= n()) %>% arrange(desc(`Death toll`))
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

Hide

```
camera
```

body_camera <fctr>	Death toll <int>
False	4317
True	578

2 rows

- Most of the incidents were not recorded on body camera

Arms Category

Hide

```
arms <- shootings %>% select(arms_category, armed) %>% group_by(arms_category, armed) %>% summar
ize(`Death toll` = n()) %>% arrange(desc(`Death toll`))
```

```
`summarise()` regrouping output by 'arms_category' (override with `.groups` argument)
```

Hide

```
arms <- rename(arms, "Weapon"= "armed")
arms
```

arms_category <fctr>	Weapon <fctr>	Death toll <int>
Guns	gun	2755
Sharp objects	knife	708
Unknown	unknown	418
Unarmed	unarmed	348
Other unusual objects	toy weapon	171
Vehicles	vehicle	120
Sharp objects	machete	39
Electrical devices	Taser	24
Sharp objects	sword	22
Blunt instruments	ax	21

1-10 of 89 rows Previous **1** 2 3 4 5 6 … 9 Next

Hide

```
age_group <- shootings %>% select(Age_group, race, gender) %>% group_by(Age_group, gender, race)
%>% summarize(`Death toll` = n())
```

```
`summarise()` regrouping output by 'Age_group', 'gender' (override with `.groups` argument)
```

Hide

```
age_group
```

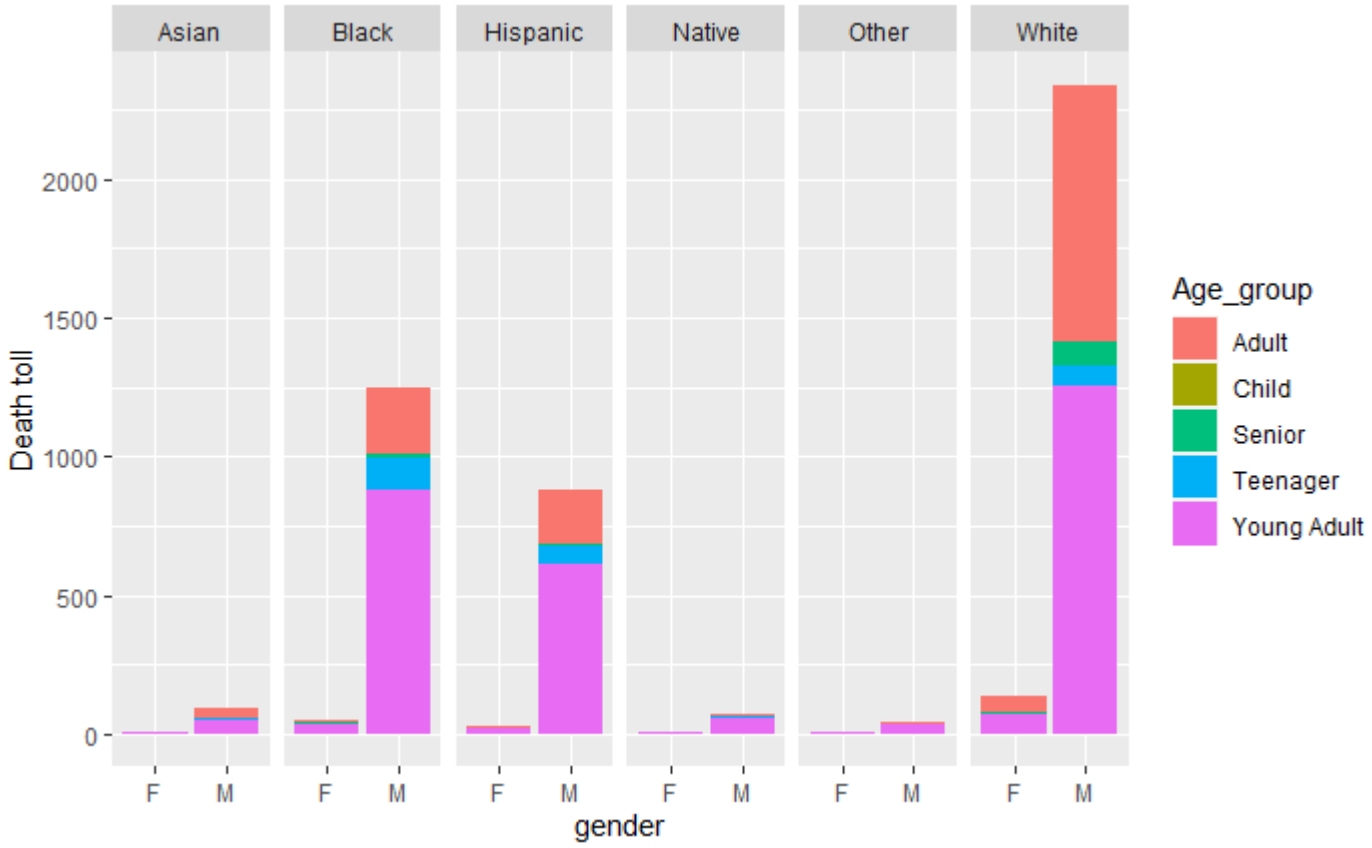
Age_group <chr>	gender <fctr>	race <fctr>	Death toll <int>
Adult	F	Asian	1
Adult	F	Black	9

Age_group <chr>	gender <fctr>	race <fctr>	Death toll <int>
Adult	F	Hispanic	3
Adult	F	Other	1
Adult	F	White	58
Adult	M	Asian	30
Adult	M	Black	242
Adult	M	Hispanic	192
Adult	M	Native	12
Adult	M	Other	10

1-10 of 38 rows Previous **1** 2 3 4 Next

Hide

```
g <- ggplot(age_group, aes(x = gender, y = `Death toll`, fill = Age_group))
g + geom_bar(stat = "identity", position = position_stack()) + facet_grid(. ~ race)
```



- As seen above, most of the victims were male young adults between the ages of 18 and 35 followed by male adults between the ages of 36 and 65