

Project - Exploring Data Set - Jacob Hoag

2024-09-25

- 1) The `fatal_police_shootings_data_1` set is a data set compiled by The Washington Post of every police killing in the U.S. from 2015-2020. This data set includes variables such as if the suspect was armed and if so with what, gender, age, and race which I will be exploring in this data analysis project. How do gender, age, and race correlate with if the suspect was armed or not?

2)

```
table(fatal_police_shootings_data_1_$armed)
```

```
##
##          air conditioner                ax
##                1                17
##          baseball bat    baseball bat and bottle
##                11                1
## baseball bat and fireplace poker                baton
##                1                4
##                bayonet                BB gun
##                1                2
##          bean-bag gun                beer bottle
##                1                2
##          blunt object                bow and arrow
##                5                1
##          box cutter                brick
##                10                2
##          carjack                chain
##                1                2
##          chain saw                chainsaw
##                2                1
##          chair                claimed to be armed
##                2                1
##          contractor's level                cordless drill
##                1                1
##          crossbow                crowbar
##                9                2
##          fireworks                flagpole
##                1                1
##          flashlight                garden tool
##                1                1
##          glass shard                gun
##                2                2195
##          gun and car                gun and knife
##                5                15
##          gun and sword                gun and vehicle
##                1                1
```

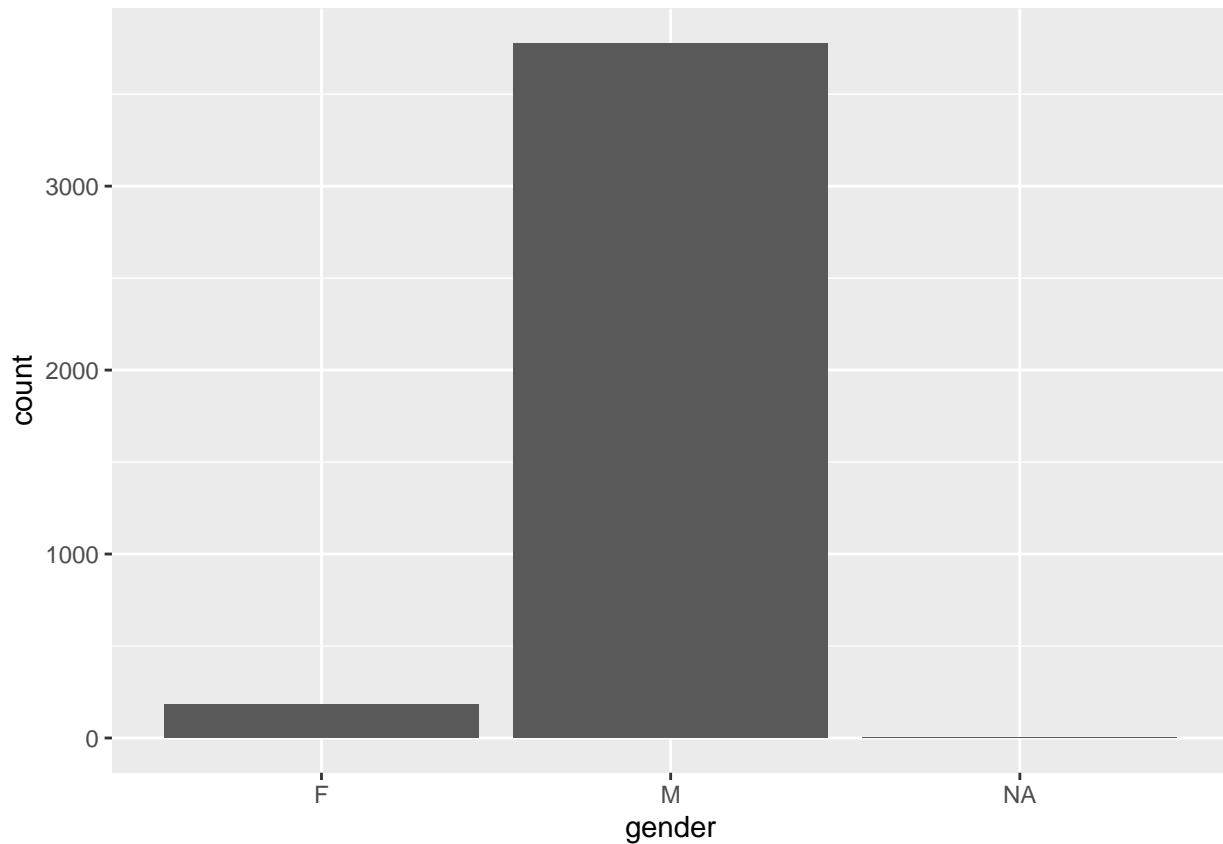
##	guns and explosives	hammer
##	3	8
##	hand torch	hatchet
##	1	8
##	hatchet and gun	incendiary device
##	2	2
##	knife	lawn mower blade
##	581	2
##	machete	machete and gun
##	36	1
##	meat cleaver	metal hand tool
##	3	1
##	metal object	metal pipe
##	4	11
##	metal pole	metal rake
##	2	1
##	metal stick	motorcycle
##	3	1
##	nail gun	oar
##	1	1
##	pellet gun	pen
##	1	1
##	pepper spray	pick-axe
##	1	3
##	piece of wood	pipe
##	3	6
##	pitchfork	pole
##	2	2
##	pole and knife	rock
##	2	4
##	samurai sword	scissors
##	1	3
##	screwdriver	sharp object
##	8	3
##	shovel	spear
##	3	1
##	stapler	straight edge razor
##	1	3
##	sword	Taser
##	13	17
##	tire iron	toy weapon
##	1	145
##	unarmed	undetermined
##	256	171
##	unknown weapon	vehicle
##	48	35
##	vehicle and gun	
##	1	

Clearly there are lots of weapons and things criminals use to cause problems for the Police. But the fact there are any unarmed killings is heinous and the fact there are 256 people who were confirmed to be unarmed yet still shot and killed by the police is wrong in so so many ways.

```
summary(fatal_police_shootings_data_1$gender)
```

```
##      Length      Class      Mode  
##      3960  character  character
```

```
ggplot(fatal_police_shootings_data_1, aes(gender)) + geom_bar()
```



```
table(fatal_police_shootings_data_1$gender)
```

```
##  
##      F      M  
## 180 3777
```

Not much to say about this variable other than majority of victims seem to be men. This would line up as most violent offenders do seem to be men which would result in more situations where someone may be shot.

```
summary(fatal_police_shootings_data_1$age)
```

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

```
table(fatal_police_shootings_data_1$age)
```

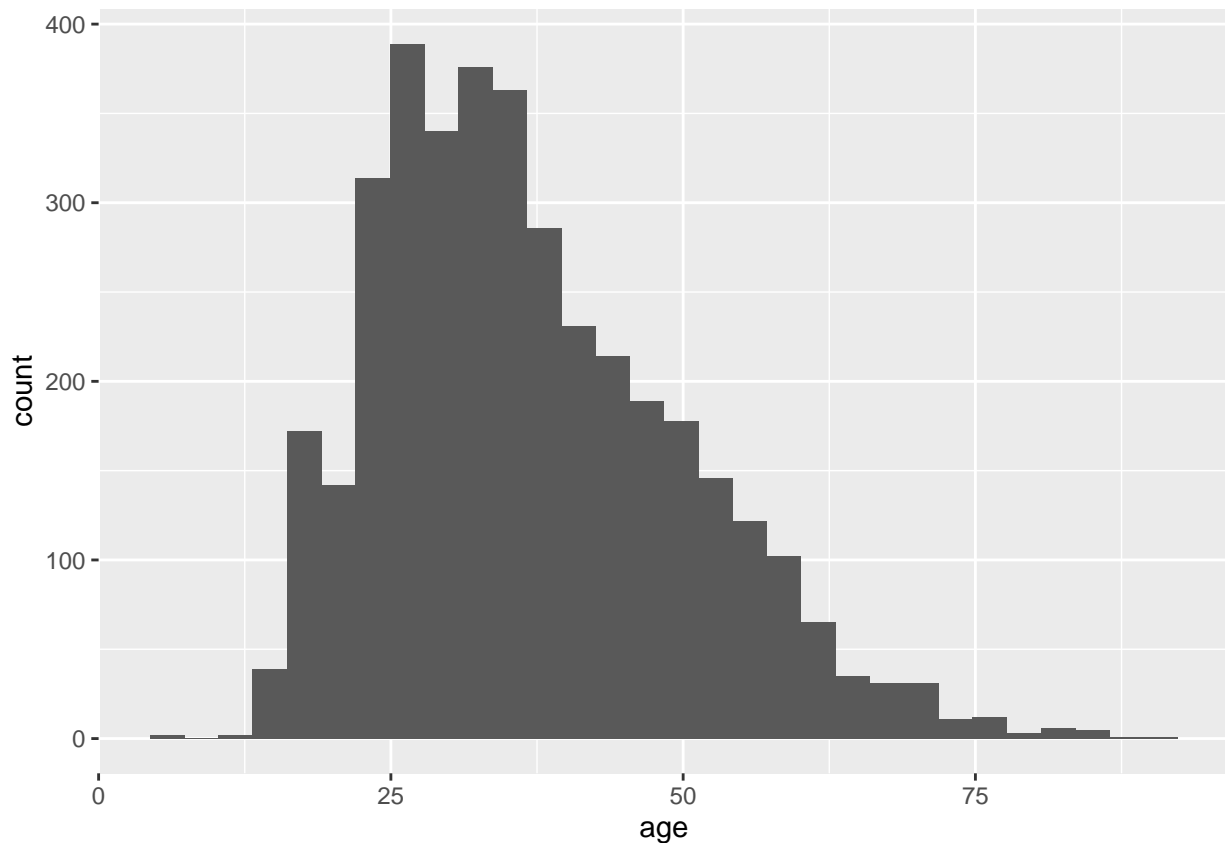
```
##
##  6 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
##  2  1  1  2 12 25 34 77 61 66 76 84 105 125 150 113 126 107 123 110
## 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
## 124 122 130 119 114 130 106 92 88 82 84 65 64 66 84 68 56 65 56 68
## 51 52 53 54 55 56 57 58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

```
ggplot(fatal_police_shootings_data_1_, aes(age))+ geom_histogram()
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 152 rows containing non-finite outside the scale range
```

```
## ('stat_bin()').
```



Victims are as young as 6 years old and as old as 91 with the average being 36.85. The average is around what I expected if not a little older than what I thought it would be. The face there is a 6 year old victim is heartbreaking I can not imagine a situation where this was the solution.

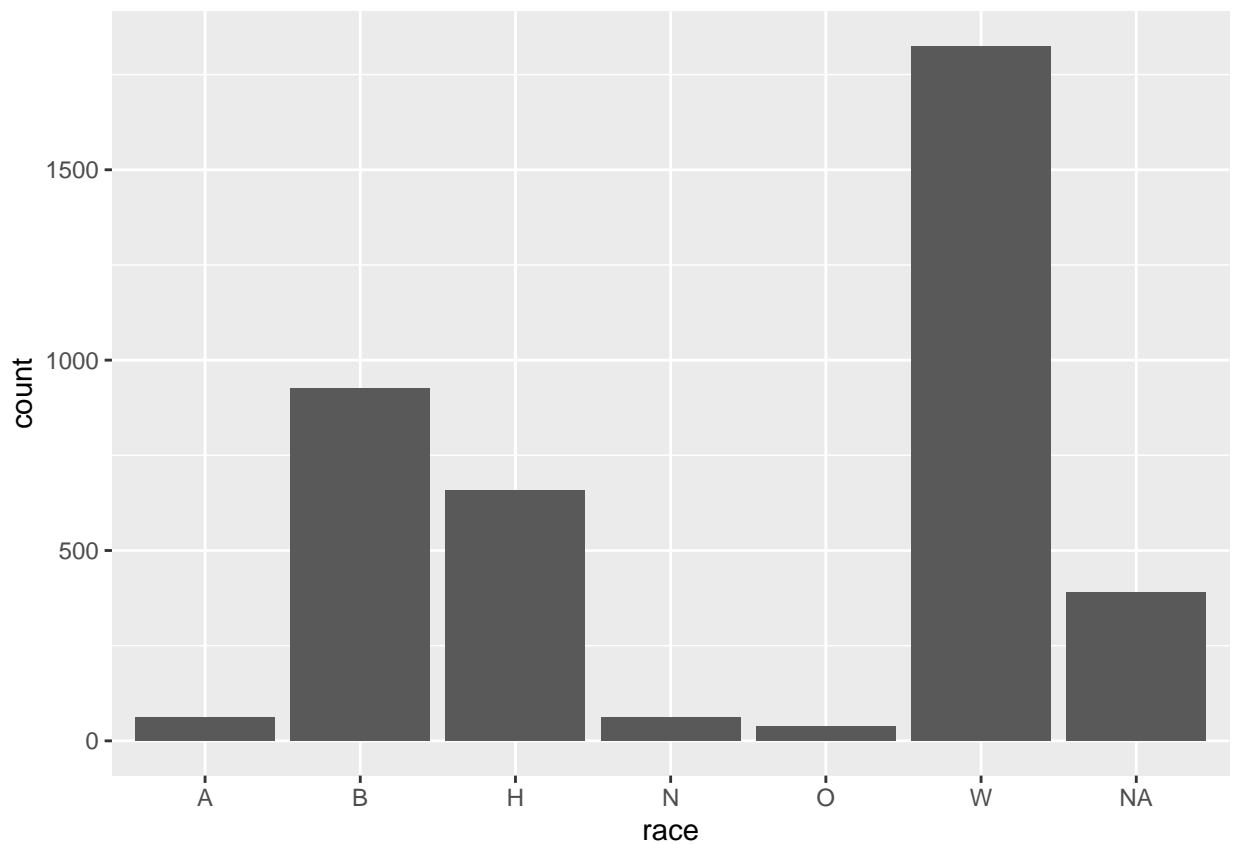
```
summary(fatal_police_shootings_data_1$race)
```

```
##      Length      Class      Mode  
##      3960 character character
```

```
table(fatal_police_shootings_data_1$race)
```

```
##  
##      A      B      H      N      O      W  
##     61    927    659     62     37   1825
```

```
ggplot(fatal_police_shootings_data_1, aes(race))+ geom_bar()
```



Majority of victims seem to be white and the second most is black. Not much to say about it other than that is to be expected as white takes up majority of the population.

3) Bivariate

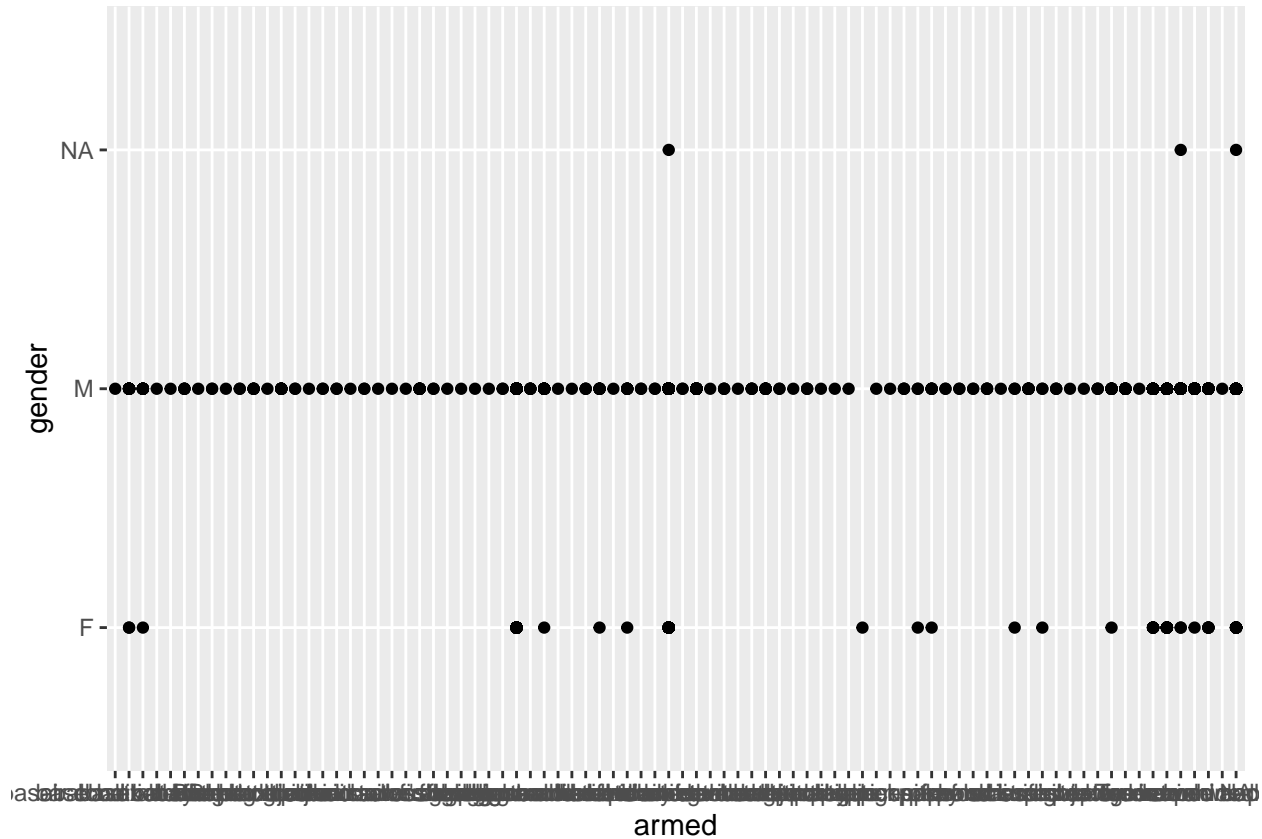
```
table(fatal_police_shootings_data_1$armed, fatal_police_shootings_data_1$gender)
```

```
##  
##              F      M  
## air conditioner  0      1
```

##	ax	2	15
##	baseball bat	1	10
##	baseball bat and bottle	0	1
##	baseball bat and fireplace poker	0	1
##	baton	0	4
##	bayonet	0	1
##	BB gun	0	2
##	bean-bag gun	0	1
##	beer bottle	0	2
##	blunt object	0	5
##	bow and arrow	0	1
##	box cutter	0	10
##	brick	0	2
##	carjack	0	1
##	chain	0	2
##	chain saw	0	2
##	chainsaw	0	1
##	chair	0	2
##	claimed to be armed	0	1
##	contractor's level	0	1
##	cordless drill	0	1
##	crossbow	0	9
##	crowbar	0	2
##	fireworks	0	1
##	flagpole	0	1
##	flashlight	0	1
##	garden tool	0	1
##	glass shard	0	2
##	gun	83	2112
##	gun and car	0	5
##	gun and knife	1	14
##	gun and sword	0	1
##	gun and vehicle	0	1
##	guns and explosives	0	3
##	hammer	1	7
##	hand torch	0	1
##	hatchet	1	7
##	hatchet and gun	0	2
##	incendiary device	0	2
##	knife	31	549
##	lawn mower blade	0	2
##	machete	0	36
##	machete and gun	0	1
##	meat cleaver	0	3
##	metal hand tool	0	1
##	metal object	0	4
##	metal pipe	0	11
##	metal pole	0	2
##	metal rake	0	1
##	metal stick	0	3
##	motorcycle	0	1
##	nail gun	0	1
##	oar	0	1
##	pellet gun	1	0

## pen	0	1
## pepper spray	0	1
## pick-axe	0	3
## piece of wood	1	2
## pipe	1	5
## pitchfork	0	2
## pole	0	2
## pole and knife	0	2
## rock	0	4
## samurai sword	0	1
## scissors	1	2
## screwdriver	0	8
## sharp object	1	2
## shovel	0	3
## spear	0	1
## stapler	0	1
## straight edge razor	0	3
## sword	1	12
## Taser	0	17
## tire iron	0	1
## toy weapon	9	136
## unarmed	18	238
## undetermined	2	168
## unknown weapon	1	47
## vehicle	5	30
## vehicle and gun	0	1

```
ggplot(fatal_police_shootings_data_1_, aes(x=armed, y=gender)) + geom_point()
```



The “armed” variable makes it difficult to be graphed due to the large and diverse number of categories that the variable yields. However the table makes it easy to see that male and female victims follow the same trend in the common weapons used.

```
table(fatal_police_shootings_data_1$gender, fatal_police_shootings_data_1$age)
```

```
##
##      6  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29
## F   0   1   0   0   0   1   4   1   0   2   5   2   6   5   9   5   6   5   2
## M   2   0   1   2  12  24  30  76  61  64  70  82  99 120 141 108 120 102 121
##
##     30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48
## F   6   8   4   4   6   2   7   6   5   5   4   3   2   3   3   3   6   1   3
## M 104 116 118 126 113 112 123 100  87  83  78  81  63  61  63  81  62  55  62
##
##     49  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67
## F   5   6   1   0   4   5   4   2   0   0   1   2   1   2   0   0   0   1   0
## M  51  62  53  43  51  43  43  42  31  31  46  22  20  21  21  18  17  10  10
##
##     68  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  86  89
## F   0   0   0   2   1   0   0   0   0   0   0   0   0   0   0   0   1   0   0
## M  10   9  12   8   5   3   2   2   7   3   1   1   1   2   2   2   2   2   1
##
##      91
## F    0
## M    1
```

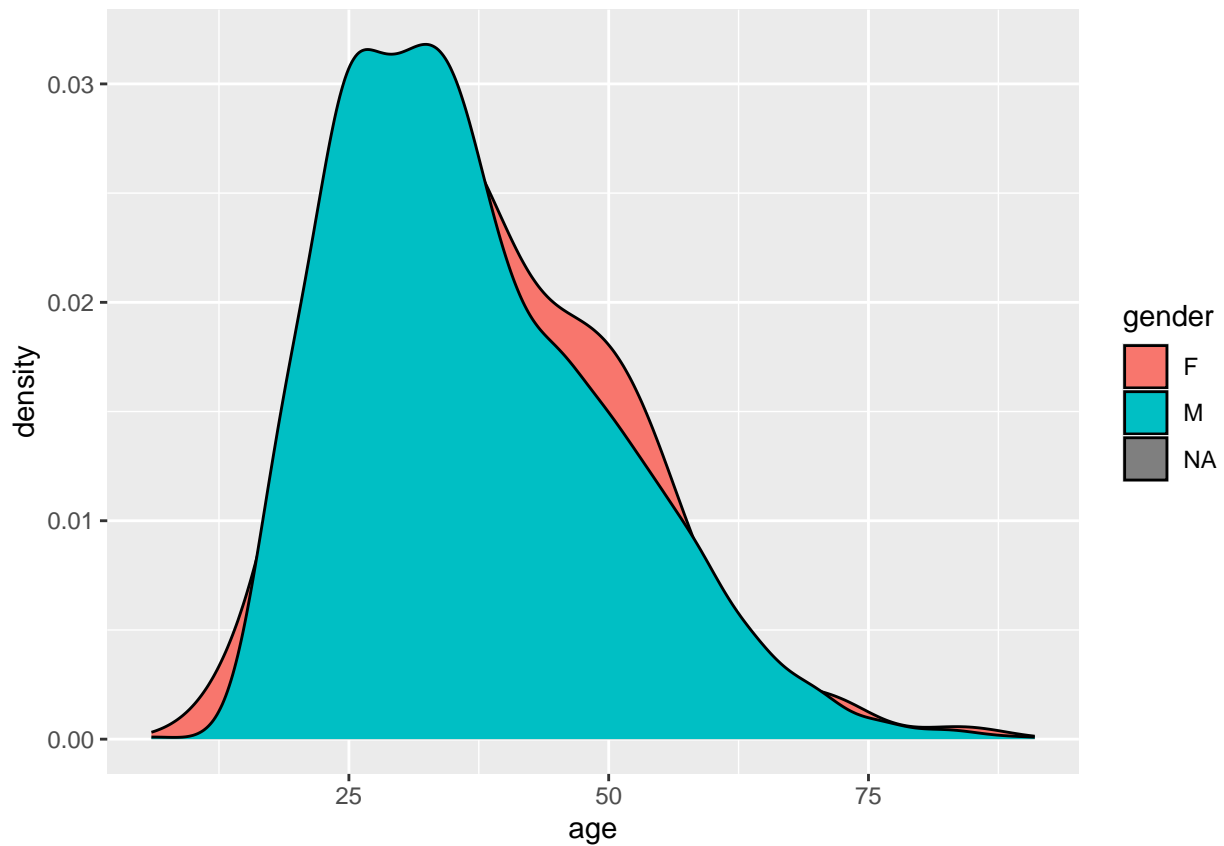


```
ggplot(fatal_police_shootings_data_1_, aes(x=age, fill=gender)) + geom_density()
```

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

```
## Warning: Groups with fewer than two data points have been dropped.
```

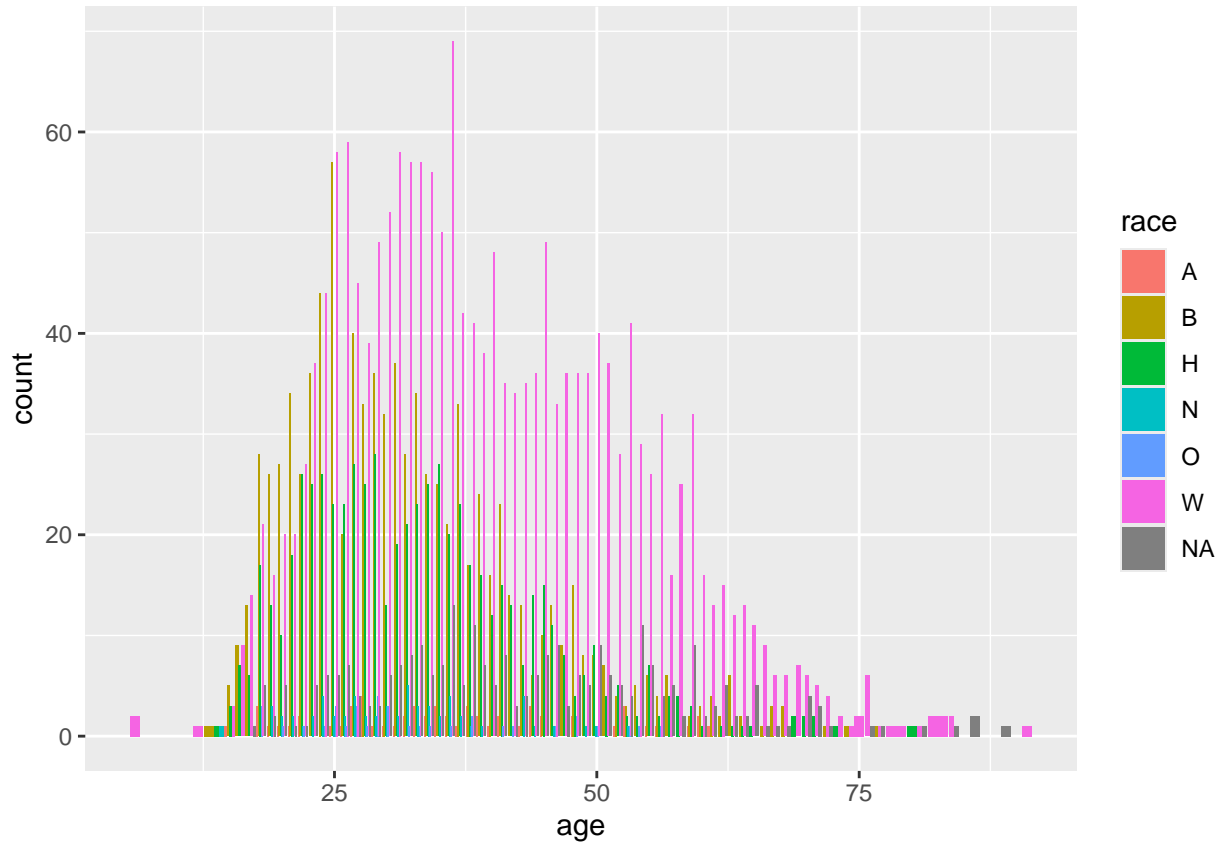
```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning  
## -Inf
```



The female deaths seem to be more bundled together in age when compared to male. This may be due to a few reasons one being more men victims make it easier to spread the age group.

```
ggplot(fatal_police_shootings_data_1_, aes(x=age, fill=race)) + geom_bar(position = "dodge")
```

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



```
table(fatal_police_shootings_data_1$age, fatal_police_shootings_data_1$race)
```

```
##
##      A  B  H  N  O  W
##  6    0  0  0  0  0  2
## 12    0  0  0  0  0  1
## 13    0  1  0  0  0  0
## 14    0  0  1  1  0  0
## 15    1  5  3  0  0  3
## 16    0  9  7  0  0  9
## 17    0 13  6  0  0 14
## 18    3 28 17  0  3 21
## 19    1 26 13  3  0 16
## 20    1 27 10  2  1 20
## 21    1 34 18  2  0 20
## 22    2 26 26  1  1 27
## 23    0 36 25  2  0 37
## 24    0 44 26  4  1 44
## 25    1 57 23  2  3 58
## 26    1 20 23  2  1 59
## 27    3 40 27  4  3 45
## 28    4 33 25  2  1 39
## 29    1 36 28  2  4 49
## 30    1 32 13  3  3 52
## 31    1 37 19  2  0 58
```

```
## 32 2 28 21 5 1 57
## 33 3 34 23 3 1 57
## 34 2 26 25 3 1 56
## 35 3 25 27 2 0 50
## 36 2 21 20 4 1 69
## 37 1 33 23 2 0 42
## 38 3 17 17 1 2 41
## 39 2 24 16 1 0 38
## 40 1 16 12 0 0 48
## 41 2 23 15 0 1 35
## 42 0 14 13 0 1 34
## 43 1 13 7 4 0 35
## 44 3 6 14 1 0 36
## 45 2 10 15 0 0 49
## 46 0 13 11 1 1 33
## 47 0 9 8 0 0 36
## 48 2 15 4 0 2 36
## 49 0 8 6 1 0 36
## 50 0 8 9 1 1 40
## 51 0 7 4 0 0 37
## 52 1 4 5 0 0 28
## 53 3 3 2 1 1 41
## 54 0 5 2 0 1 29
## 55 1 6 7 0 0 26
## 56 1 4 2 0 1 32
## 57 0 6 4 0 0 16
## 58 0 0 4 0 0 25
## 59 1 2 3 0 0 32
## 60 2 3 1 0 0 16
## 61 1 4 0 0 0 13
## 62 0 2 1 0 0 15
## 63 0 6 1 0 0 12
## 64 0 2 1 0 0 13
## 65 0 0 1 0 0 11
## 66 0 1 0 0 0 9
## 67 0 3 0 0 0 6
## 68 0 3 0 0 0 6
## 69 0 0 2 0 0 7
## 70 0 0 2 0 0 6
## 71 0 0 2 0 0 5
## 72 0 1 0 0 0 4
## 73 0 0 1 0 0 2
## 74 0 1 0 0 0 1
## 75 0 0 0 0 0 2
## 76 0 0 0 0 0 6
## 77 0 1 0 0 0 1
## 78 0 0 0 0 0 1
## 79 0 0 0 0 0 1
## 80 0 0 1 0 0 0
## 81 0 0 0 0 0 1
## 82 0 0 0 0 0 2
## 83 0 0 0 0 0 2
## 84 0 0 0 0 0 2
## 86 0 0 0 0 0 0
```

```
## 89 0 0 0 0 0 0
## 91 0 0 0 0 0 1
```

Even with white people making up majority of the victims due to the large population amount, a disproportionate number of young black people are killed. It is easy to see in the table especially in ages 17-20.

- 4) Conclusion In the end I found that the two most influential variables, in my opinion, are age and race when understanding and evaluating the armed variable. With a disproportionate number of young black males were killed which means there are more fatalities in the armed/unarmed variable which makes it more profound for examination and statistical analysis.