Final Project

Maya Arviso

2024-09-23

Introduction

The dataset I will be analyzing is Police shootings in 2015. The set includes 3,960 observations and 14 variables. Some variables I will be exploring are body cameras and if the suspect was fleeing. I want to see if I can find any patterns between these variables. I would expect those that were fleeing, to also have the body camera on.

```
head(police)
```

```
## # A tibble: 6 × 14
##      id name  date                manner_of_death armed   age gender race  city
##   <dbl> <chr> <dttm>              <chr>           <chr> <dbl> <chr>  <chr> <chr>
## 1     3 Tim … 2015-01-02 00:00:00 shot            gun      53 M      A     Shel…
## 2     4 Lewi… 2015-01-02 00:00:00 shot            gun      47 M      W     Aloha
## 3     5 John… 2015-01-03 00:00:00 shot and Taser… unar…    23 M      H     Wich…
## 4     8 Matt… 2015-01-04 00:00:00 shot            toy …    32 M      W     San …
## 5     9 Mich… 2015-01-04 00:00:00 shot            nail…    39 M      H     Evans
## 6    11 Kenn… 2015-01-04 00:00:00 shot            gun      18 M      W     Guth…
## # ℹ 5 more variables: state <chr>, signs_of_mental_illness <lgl>,
## #   threat_level <chr>, flee <chr>, body_camera <lgl>
```

```
class(police$flee)
```

```
## [1] "character"
```

```
class(police$body_camera)
```

```
## [1] "logical"
```

```
fct_count(police$flee)
```

```
## # A tibble: 5 × 2
##   f              n
##   <fct>      <int>
## 1 Car          631
## 2 Foot         491
## 3 Not fleeing 2570
## 4 Other        128
## 5 <NA>         140
```

Factoring

Now I will rename the factor for bodycams to show if they were on or off.

```
police$body_cameraFac <- factor(police$body_camera, labels=c("off", "on"))
table(police$body_camera, police$body_cameraFac, useNA="always")
```

```
##
##          off   on <NA>
##   FALSE 3527    0    0
##   TRUE     0  433    0
##   <NA>     0    0    0
```

I want a table of those fleeing and another of those with bodycameras.

```
table(police$flee, police$body_cameraFac, useNA="always")
```

```
##
##                off   on <NA>
##   Car          579   52    0
##   Foot         420   71    0
##   Not fleeing 2289  281    0
##   Other        112   16    0
##   <NA>         127   13    0
```

Univariate Exploration

```
table(police$flee, police$body_cameraFac) %>% prop.table
```

```
##
##                      off          on
##   Car         0.151570681 0.013612565
##   Foot        0.109947644 0.018586387
##   Not fleeing 0.599214660 0.073560209
##   Other       0.029319372 0.004188482
```

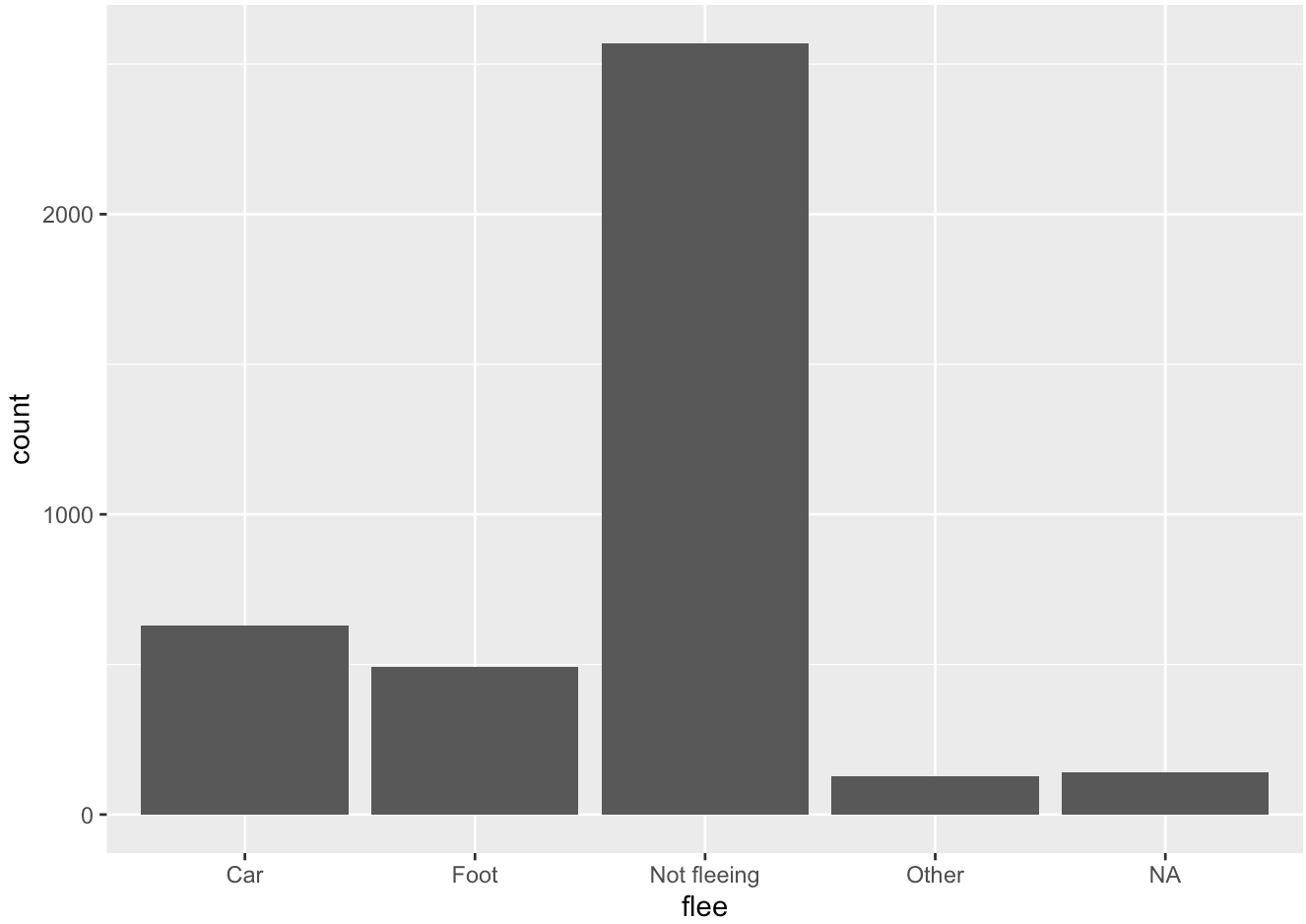
```
table(police$body_cameraFac) %>% prop.table
```

```
##
##       off        on
## 0.8906566 0.1093434
```

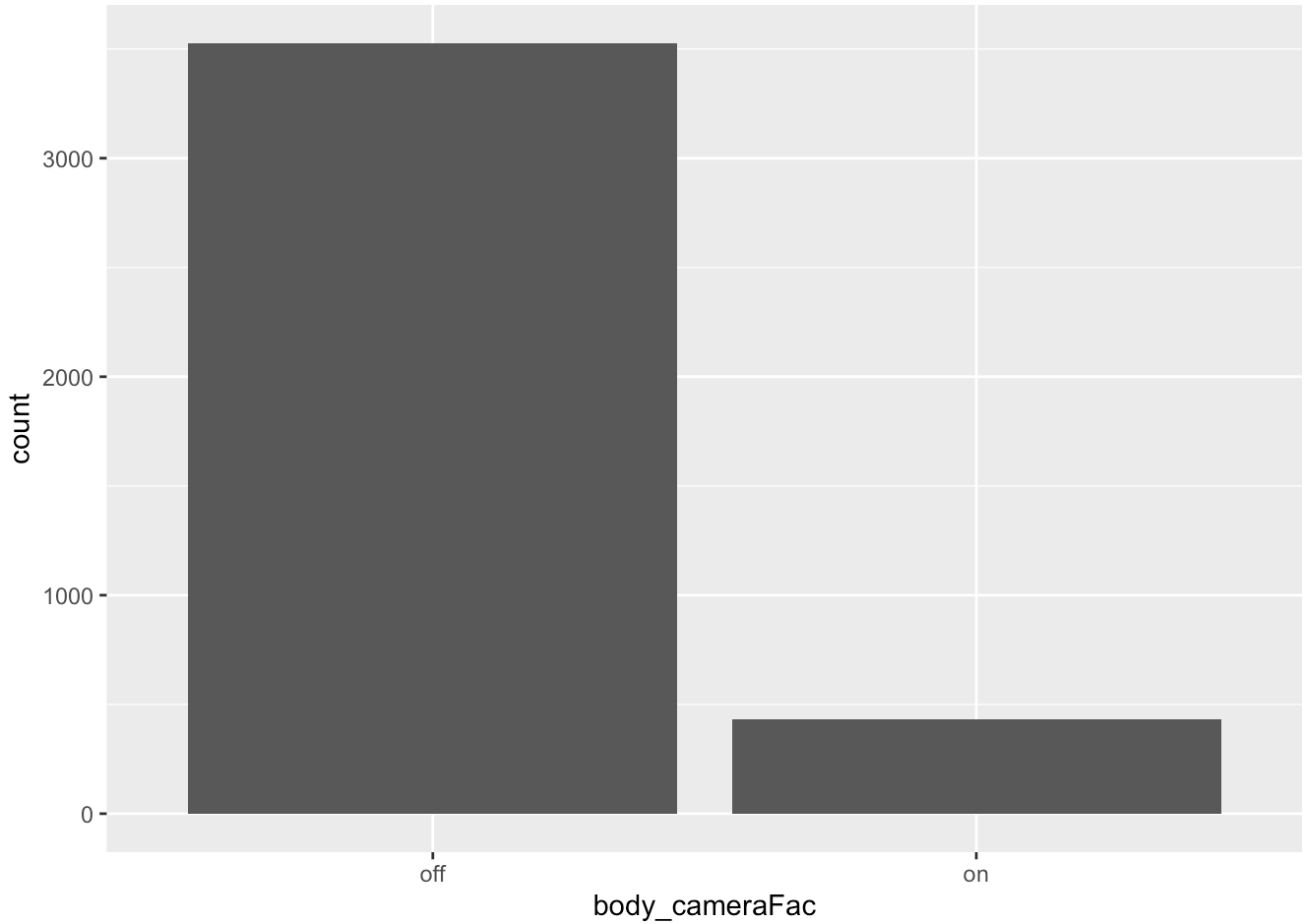
Based on this table, we can see the percentages of those with body cameras on or off. Roughly, 89% of police had their cameras off while 11% had them on.

Now I want a graph of this table.

```
ggplot(police, aes(x=flee)) + geom_bar()
```

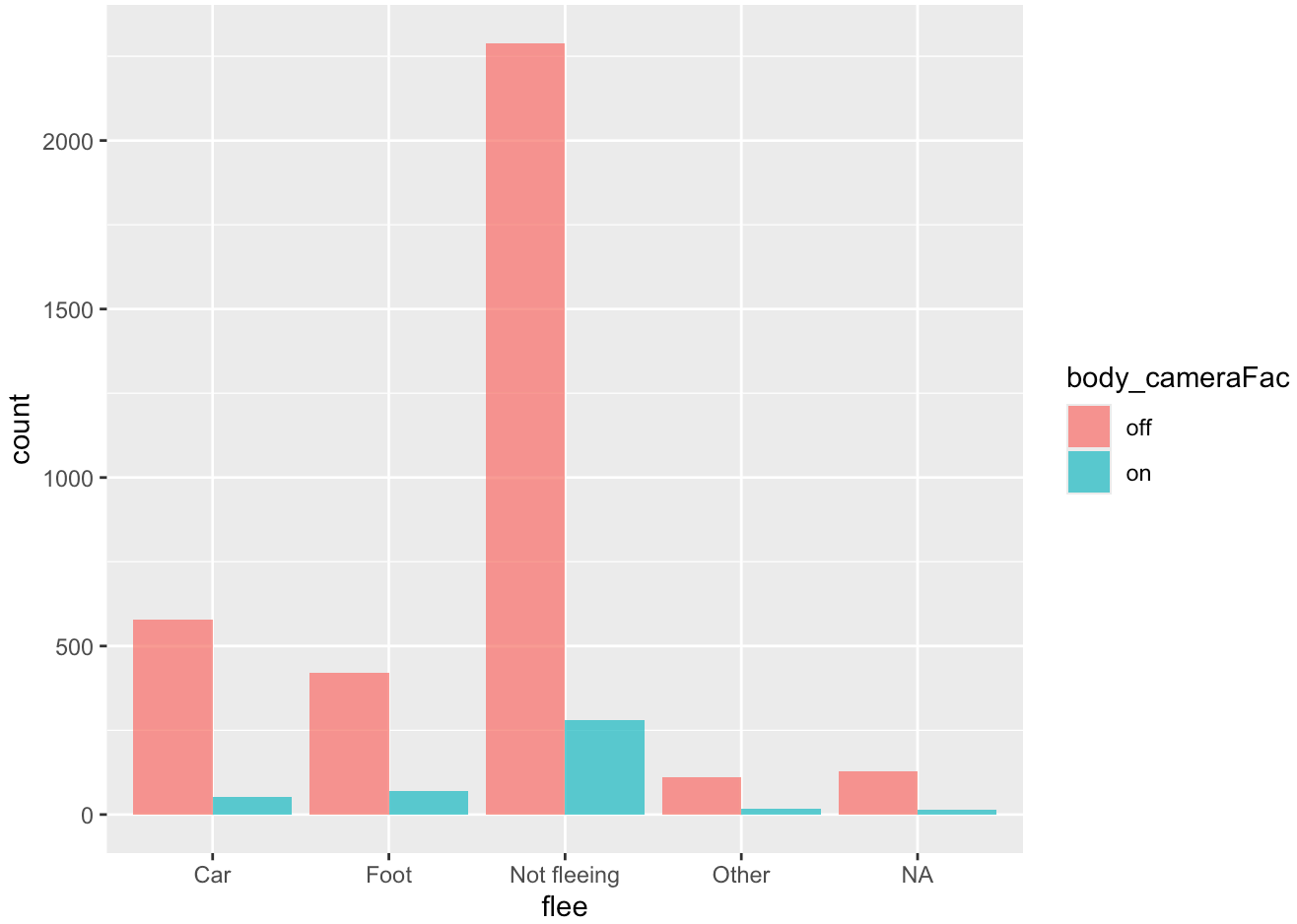



```
ggplot(police, aes(x=body_cameraFac)) + geom_bar()
```



Bivariate Exploration

```
ggplot(police, aes(x=flee, fill=body_cameraFac)) + geom_bar(alpha=0.75, position="dodge")
```



Conclusion

Based on this data, it is apparent that a majority of body cameras were off in these altercations. It is more likely to have the body cameras off in every scenario. This was done by exploring the bivariates of this dataset.