

Final Project

Madi Payne

2/25/2021

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE, fig.height=4, fig.width=5, fig.align='center')
library(dplyr)
library(forcats)
library(readxl)
library(sjPlot)
library(ggplot2)
```

```
shootings <- read_excel("/Users/madipayne/Desktop/math_130/data/fatal-police-shootings-data.xlsx")
```

Short Introduction & Description of Data

The data I am analyzing was compiled by the Washington Post and outlines the characteristics of fatal police shooting incidences in the year 2015.

Three Variables I will be Exploring

- If the individual who was shot was showing signs of mental illness.
- The age of the victim.
- If there is body camera footage of the incident.

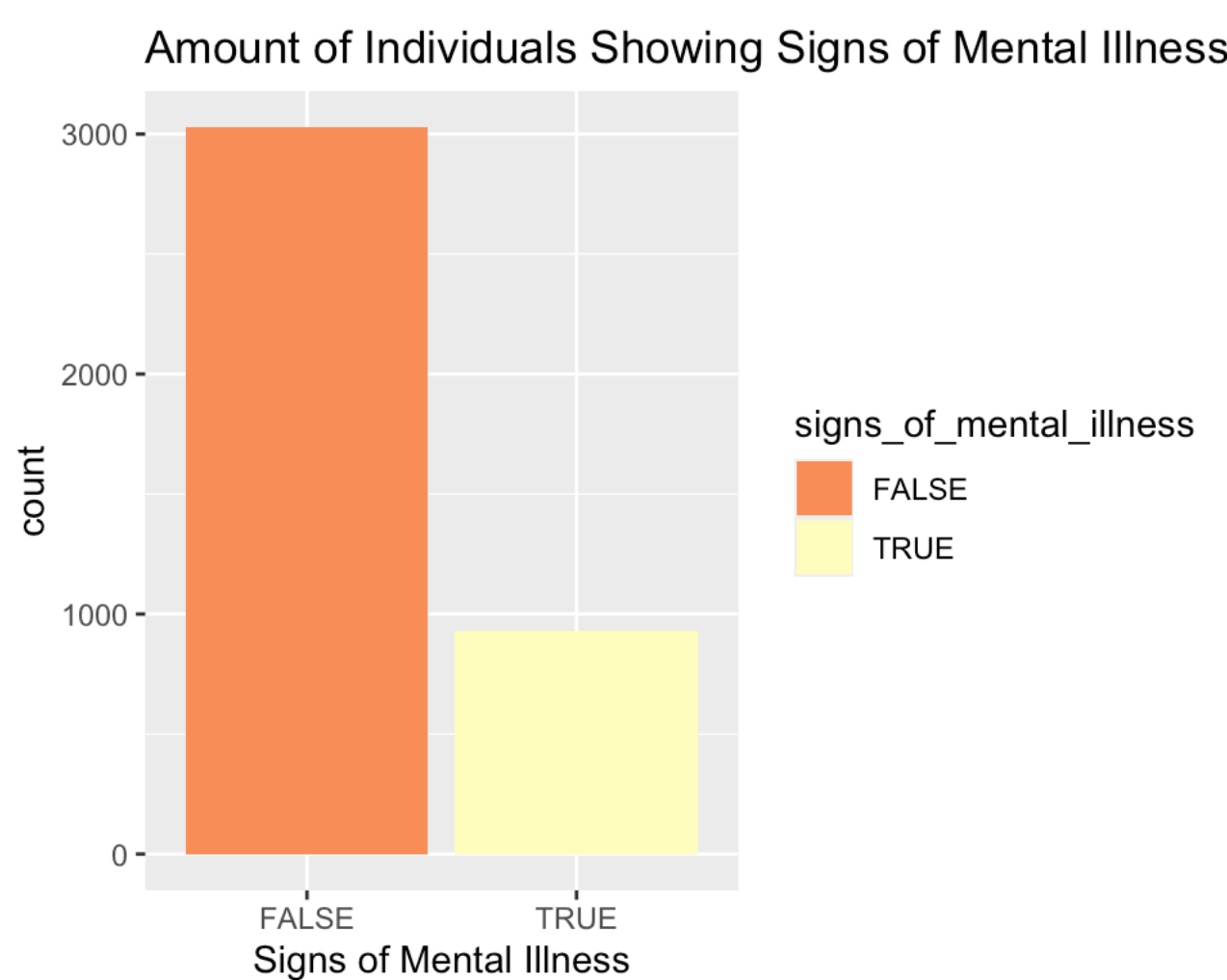
```
shootings %>% select(signs_of_mental_illness, age, body_camera)
```

```
## # A tibble: 3,960 x 3
##   signs_of_mental_illness age body_camera
##   <lgl>                   <dbl> <lgl>
## 1 TRUE                    53 FALSE
## 2 FALSE                   47 FALSE
## 3 FALSE                   23 FALSE
## 4 TRUE                    32 FALSE
## 5 FALSE                   39 FALSE
## 6 FALSE                   18 FALSE
## 7 FALSE                   22 FALSE
## 8 FALSE                   35 FALSE
## 9 FALSE                   34 TRUE
## 10 FALSE                  47 FALSE
## # ... with 3,950 more rows
```

Here, I just selected the specific pieces of data I chose to use so that the entire data set wouldn't print.

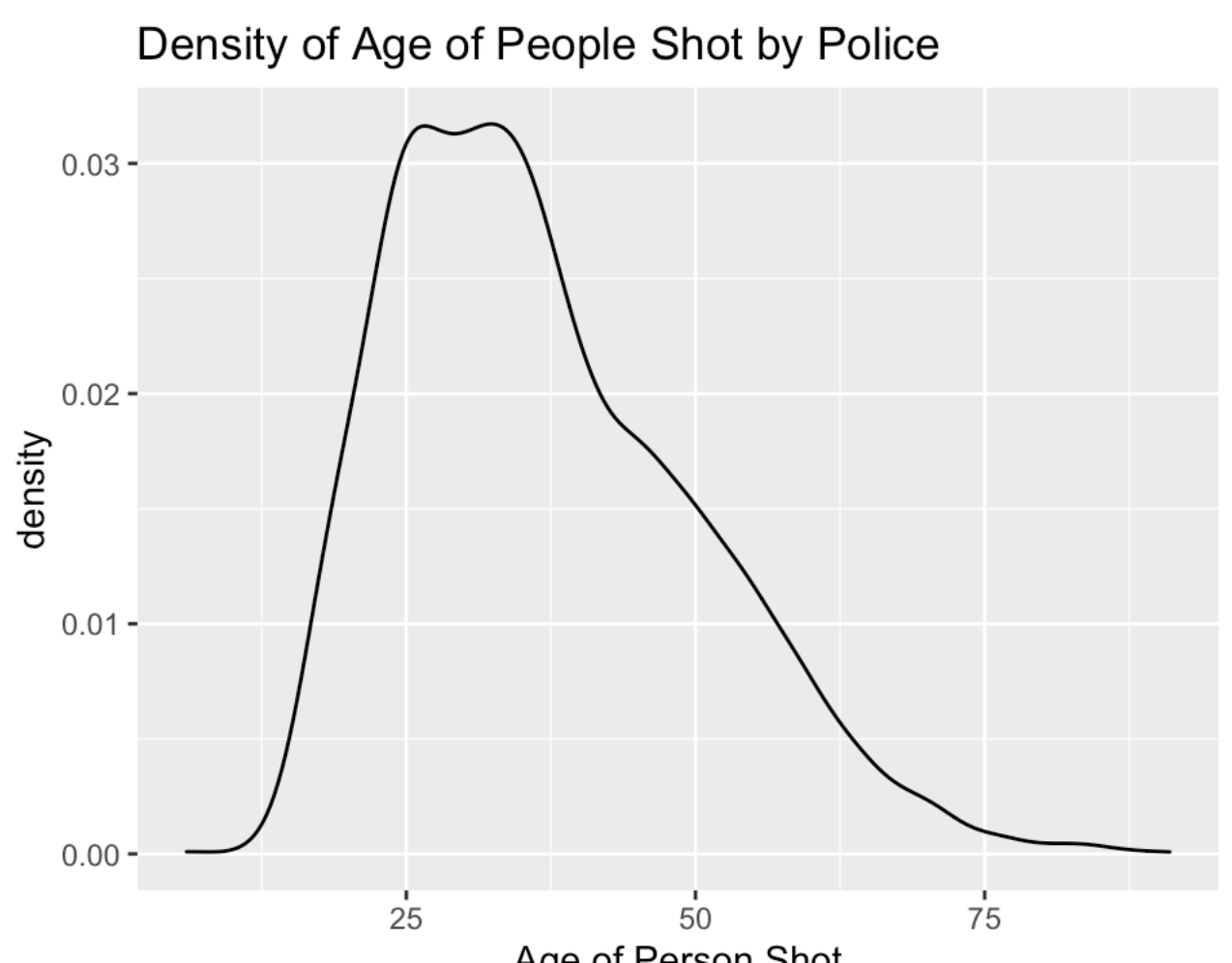
Univariate Description of Each Variable Being Considered

```
ggplot(shootings, aes(x=signs_of_mental_illness, fill=signs_of_mental_illness)) + geom_bar() + xlab("Signs of Mental Illness") + ggtitle("Amount of Individuals Showing Signs of Mental Illness or Not") + scale_fill_discrete(name="Signs of Mental Illness") + scale_fill_brewer(palette="Spectral")
```



This bar graph shows the amount of people who were shot and were not showing signs of mental illness and the amount of people shot who were showing signs of mental illness. It is clear from this bar chart that significantly more people were not showing signs of mental illness than people who were. Just over 3000 people were not showing such signs, while around 950 were.

```
ggplot(shootings, aes(x=age)) + geom_density() + xlab("Age of Person Shot") + ggtitle("Density of Age of People Shot by Police")
```



This density plot shows the distribution of the people who were shot by age, and you can see from the shape of the plot that the majority of people shot were on the younger end, between mid twenties and late thirties. As the people are older, they are not shot nearly as often as the people who are younger.

```
mean(shootings$age, na.rm=T)
```

```
## [1] 36.849
```

I also went ahead and calculated the mean age of people shot by the police, and it's about 36.85. This fits along with the density curve, as the age 36.85 falls right in the area where the graph spikes.

```
table(shootings$body_camera)
```

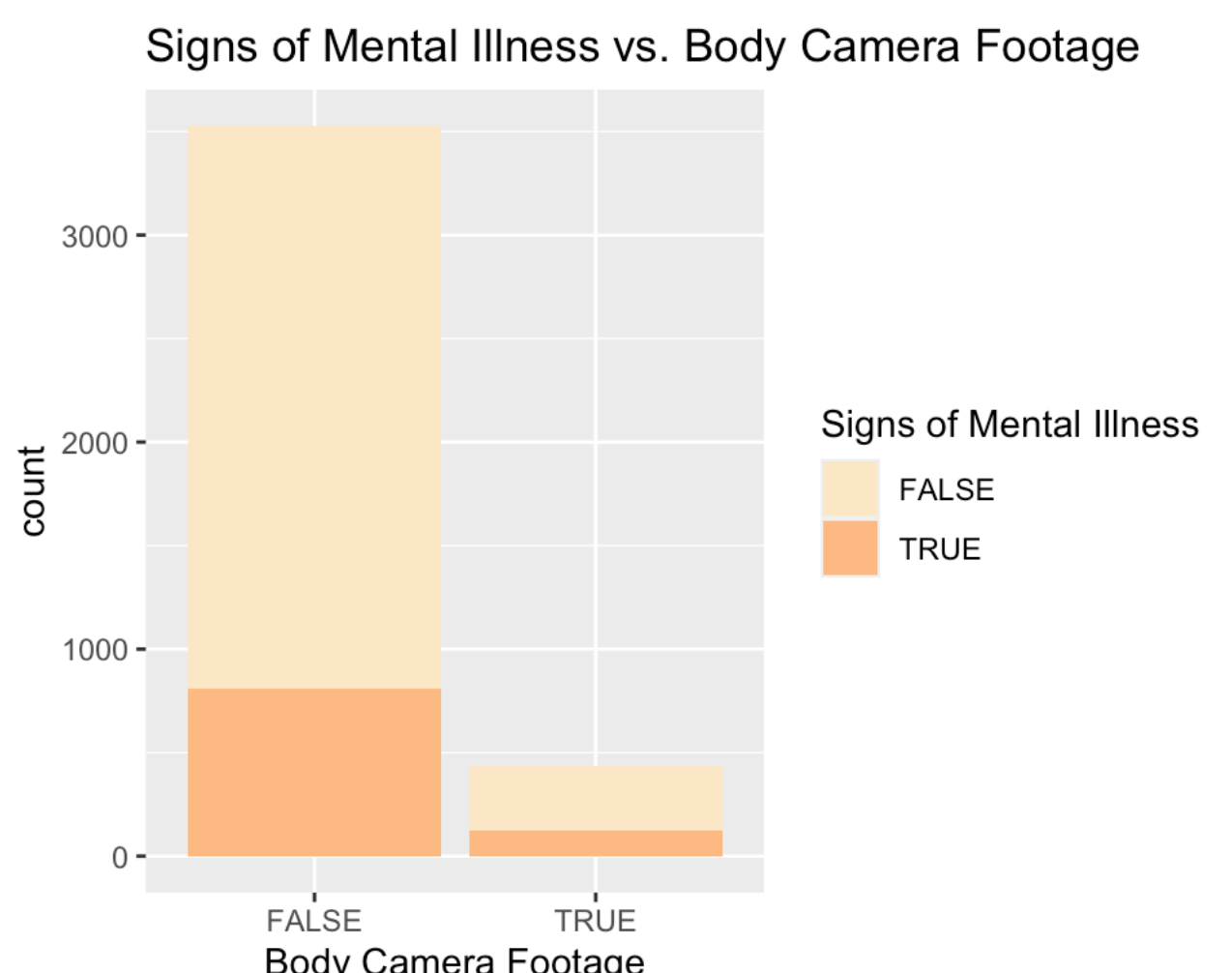
```
##
## FALSE TRUE
## 3527 433
```

This table shows whether or not there is body camera footage of each incident, and you can see that the majority of these incidents do not have recorded body camera footage. There are 3960 reported fatal police shootings for the year 2015, and only 433 of them are recorded on body camera.

Bivariate Comparison of Two Variables Being Considered

The two variables I think would be interesting to look at at the same time are whether or not the individual was showing signs of mental illness or not and whether or not the incident is recorded on body camera.

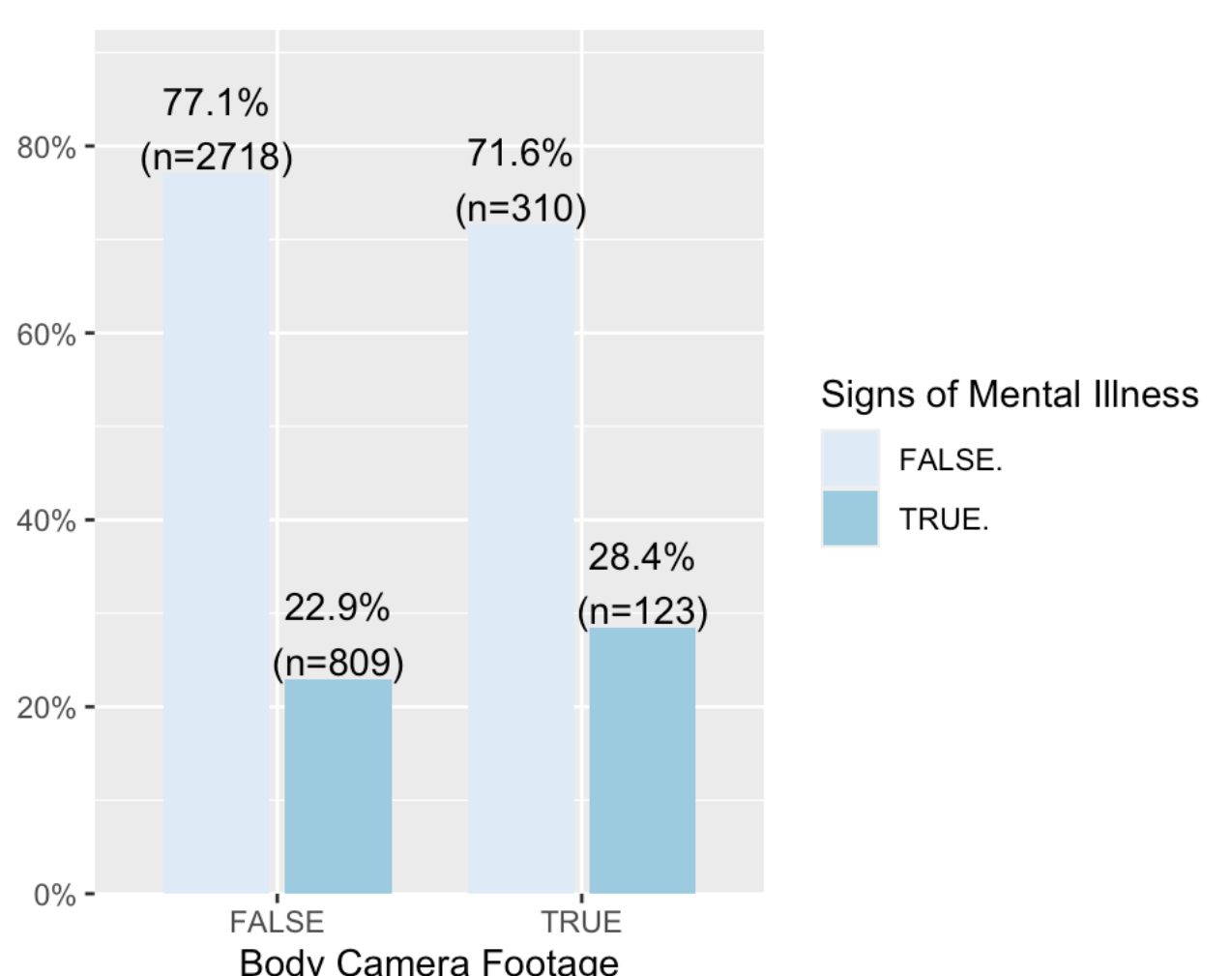
```
ggplot(shootings, aes(x=body_camera, fill=signs_of_mental_illness)) + geom_bar() + xlab("Body Camera Footage") + scale_fill_discrete(name="Signs of Mental Illness") + ggtitle("Signs of Mental Illness vs. Body Camera Footage") + scale_fill_brewer(palette="OrRd", name="Signs of Mental Illness")
```



For this bar graph, I chose to put whether or not there is body camera footage on the x-axis, and fill each box by whether or not the individual was showing signs of mental illness. As can be seen in the graph, the distribution between the two variables is fairly even. In creating this visual, I was expecting to see more of the individuals not showing signs of mental illness to not be recorded, and to see that those who were showing signs of mental illness and who were still shot to not be recorded. It seems that this data is pretty evenly distributed between the two bars, considering significantly more of the incidents do not have body camera footage and significantly more individuals were not showing signs of mental illness.

Grouped Summary Statistics for Bivariate Comparison

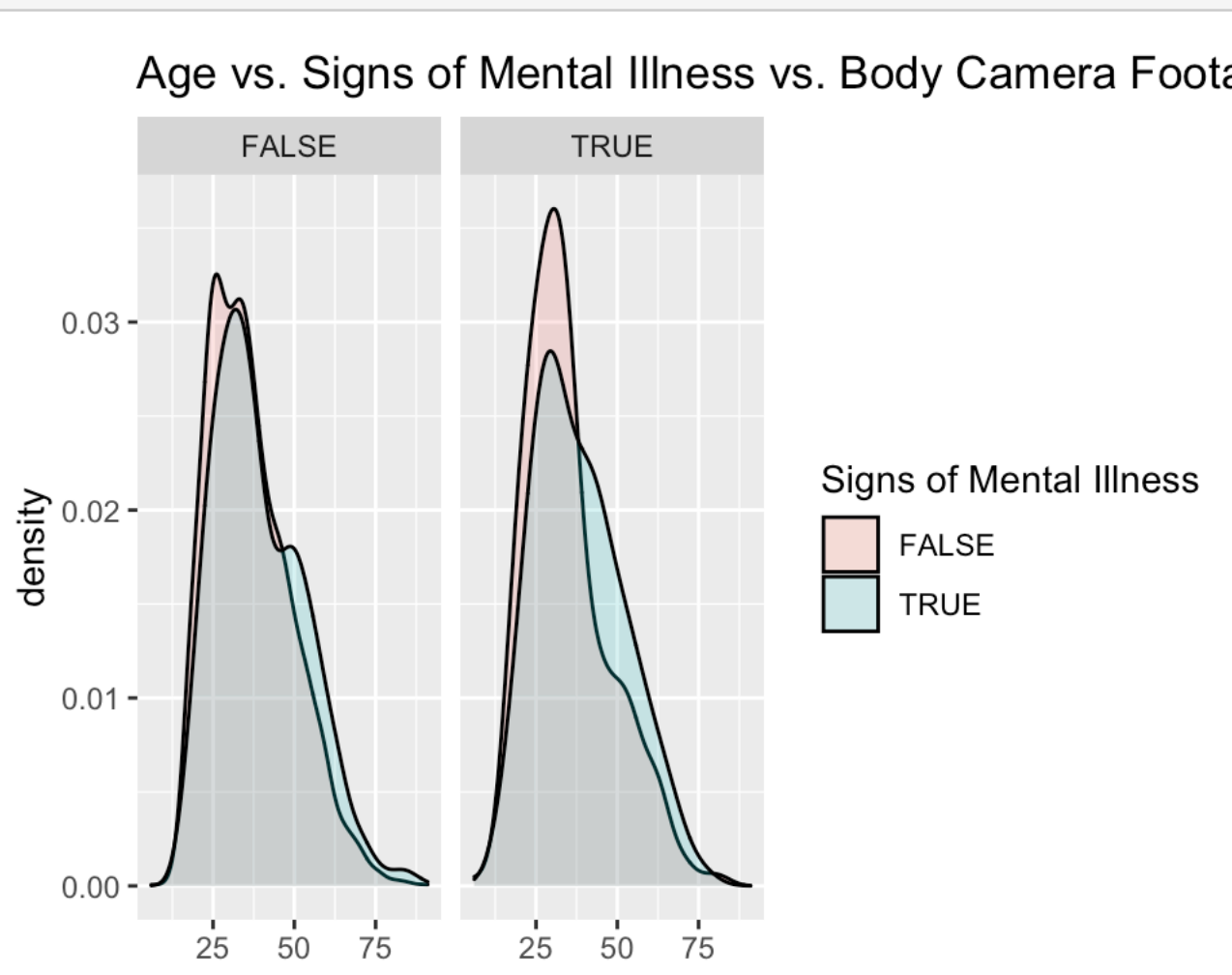
```
plot_xtab(shootings$body_camera, shootings$signs_of_mental_illness, margin='row', show.total = FALSE) + xlab("Body Camera Footage") + scale_fill_brewer(name="Signs of Mental Illness")
```



This proportion table shows the percentages of incidents in which individuals were showing signs of mental illness or not vs. whether or not there is body camera footage of the incident. For instance, in the cases where there is no body camera footage, 77.1% of the individuals were not showing signs of mental illness, while 22.9% of the individuals were showing signs of mental illness.

I'm also going to create a visual comparing all three of the variables I chose to study in the beginning.

```
ggplot(shootings, aes(x=age, fill=signs_of_mental_illness)) + geom_density(alpha=.2) + facet_wrap(~body_camera) + ggtitle("Age vs. Signs of Mental Illness vs. Body Camera Footage") + xlab("Age") + scale_fill_discrete(name="Signs of Mental Illness")
```



This plot shows two graphs, separated by labels TRUE and FALSE at the top of each graph. These represent whether or not there is body camera footage of the incident. Inside each graph is a density plot comparing the age of the individual and whether or not they were showing signs of mental illness. You can see from these plots, that there are more individuals younger in age involved in fatal police shootings, but more of them do not show signs of mental illness. If the older individuals are involved in these shootings, more of them are showing signs of mental illness than are not.

Summary

From analyzing this data of Fatal Police Shootings in 2015 compiled by the Washington Post, you can see that the mean age of people involved in these shootings is 36.85 years old. As age increases, less people are involved. You can also see that of the people involved in these shootings, only about 1/3 of them are showing signs of mental illness, while the other 2/3 are not. You can also see that of the 3960 incidents in 2015, 3527 of them were not recorded on body cameras, and only 433 of them were. While comparing body camera footage and mental illness, it seemed as though whether or not the individual was showing signs of mental illness did not affect whether or not the incident was recorded on body camera. Conversely, one very interesting thing can be seen from the proportion table - roughly 6% more of individuals shot while showing signs of mental illness were recorded on body camera than were not. This is interesting because you may think that if an individual is showing signs and is still shot, it would not be recorded. Lastly, from the density plot comparing all three variables, all of this information is put together. The majority of people involved in the shootings are on the younger end, and more of those individuals are not showing signs of mental illness at the time of the incident. Less often are older individuals involved in these incidents, however, when they are, more of them are showing signs of mental illness than are not. The two plots are relatively the same shape, so it does not appear that age and whether or not the individuals are showing signs of mental illness affect whether or not the incident is recorded on body camera.