EDA Project

Luisa-jazmin Garcia Carrillo

2022-09-26

```
library(ggplot2)
```

```
library(sjPlot)
```

```
## Learn more about sjPlot with 'browseVignettes("sjPlot")'.
```

Introduction

The data that I am going to analyze is the Depression data set. I am going to be analyzing the sex and age variables of this data. Age was measured in years of age and the subjects drinking habits were also recorded in the data. What I'm interested in analyzing and what my research question will be is which specific age groups and/or drinking habit groups are more common to experience depression.

```
depress <- read.delim("/Users/jasmincarrillo/Desktop/math130/data/depress_081217.txt", header=TRUE,sep=
dim(depress)
```

```
## [1] 294  37
```

Univariate Description

Variable being observed: "age"

```
summary(depress$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   28.00   42.50   44.41   59.00   89.00
```

```
table(depress$age)
```

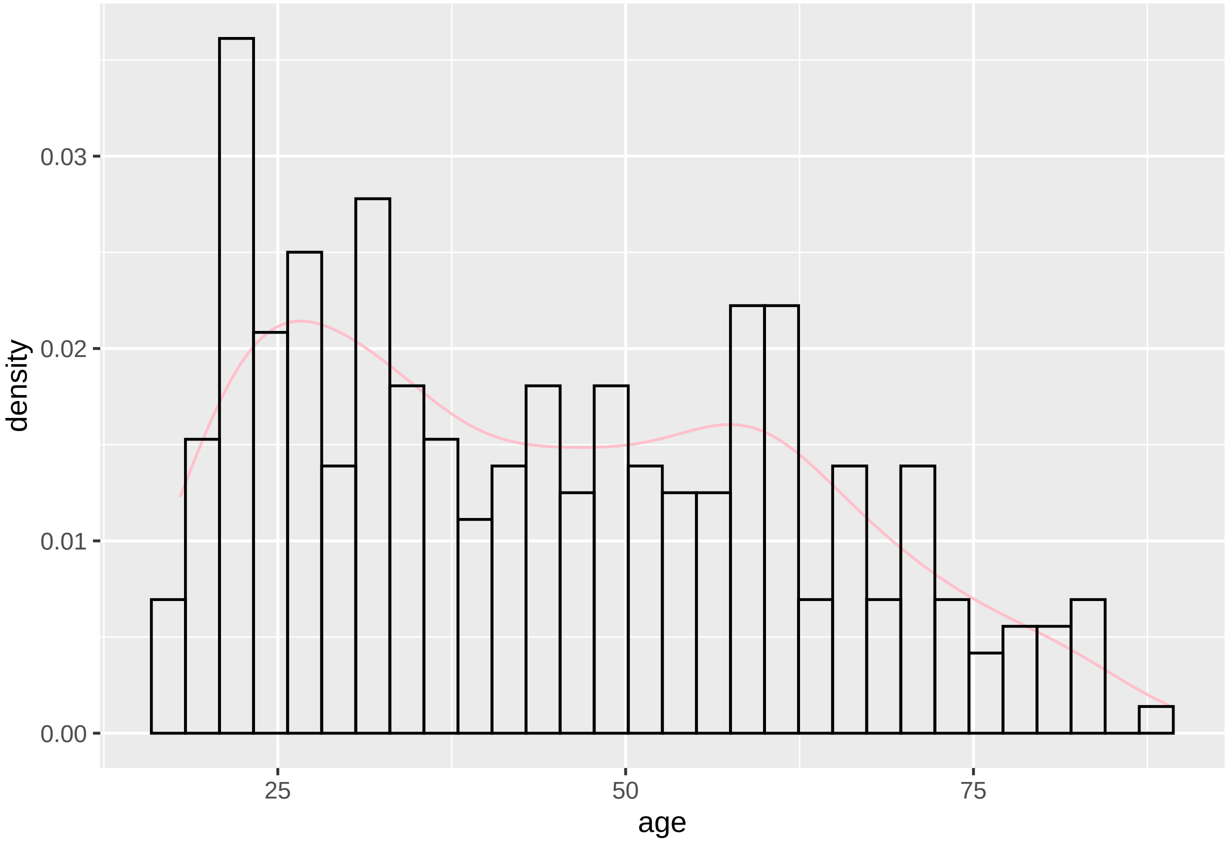
```
##
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
##  5  5  6  6  9 11  9  6  9  4  5  4  6  5 10  5  9  4  6  5  2  1  5  1  9  7
## 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69
##  2  4  3  6  4  4  5  6  4  2  3  4  3  6  7  9  7  5  4  2  3  5  3  2  4  1
## 70 71 72 73 74 75 77 78 79 80 81 82 83 89
##  5  3  2  2  3  1  2  2  2  1  2  1  5  1
```

```
sd(depress$age)
```

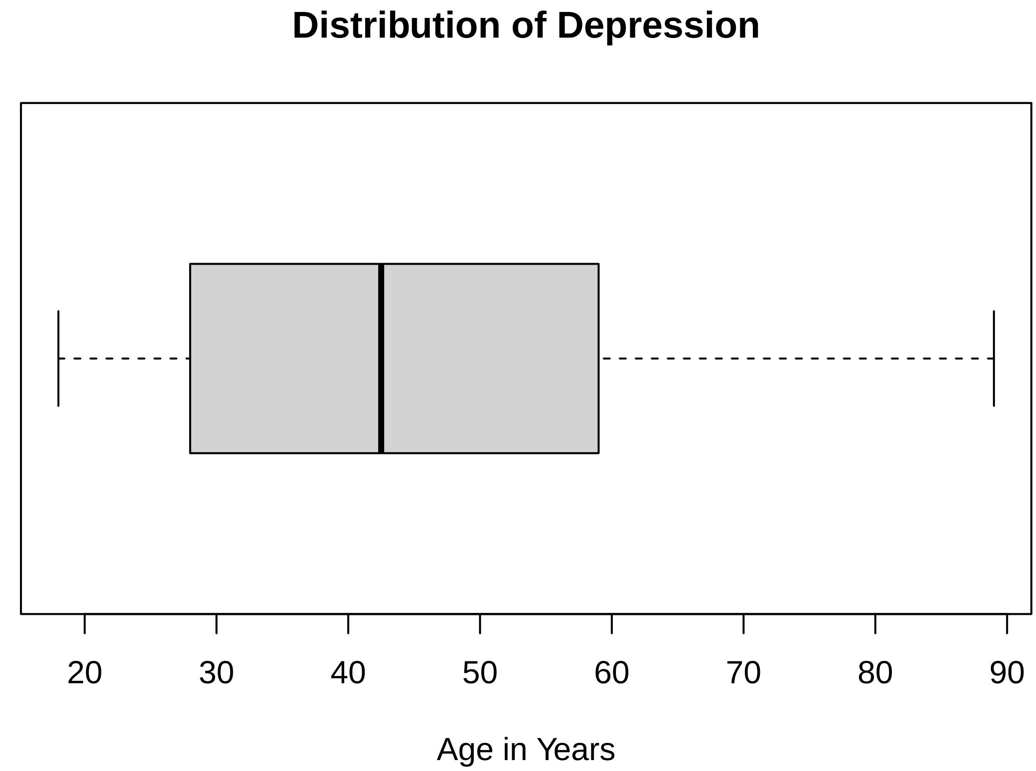
```
## [1] 18.08544
```

```
ggplot(depress, aes(x=age))+geom_density(col="pink")+geom_histogram(aes(y=..density..), color="black", 
```

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



```
boxplot(depress$age, horizontal = TRUE, main="Distribution of Depression", xlab="Age in Years")
```



Variable being observed: “Drink”

```
summary(depress$drink)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  1.0000  1.0000  0.7959  1.0000  1.0000
```

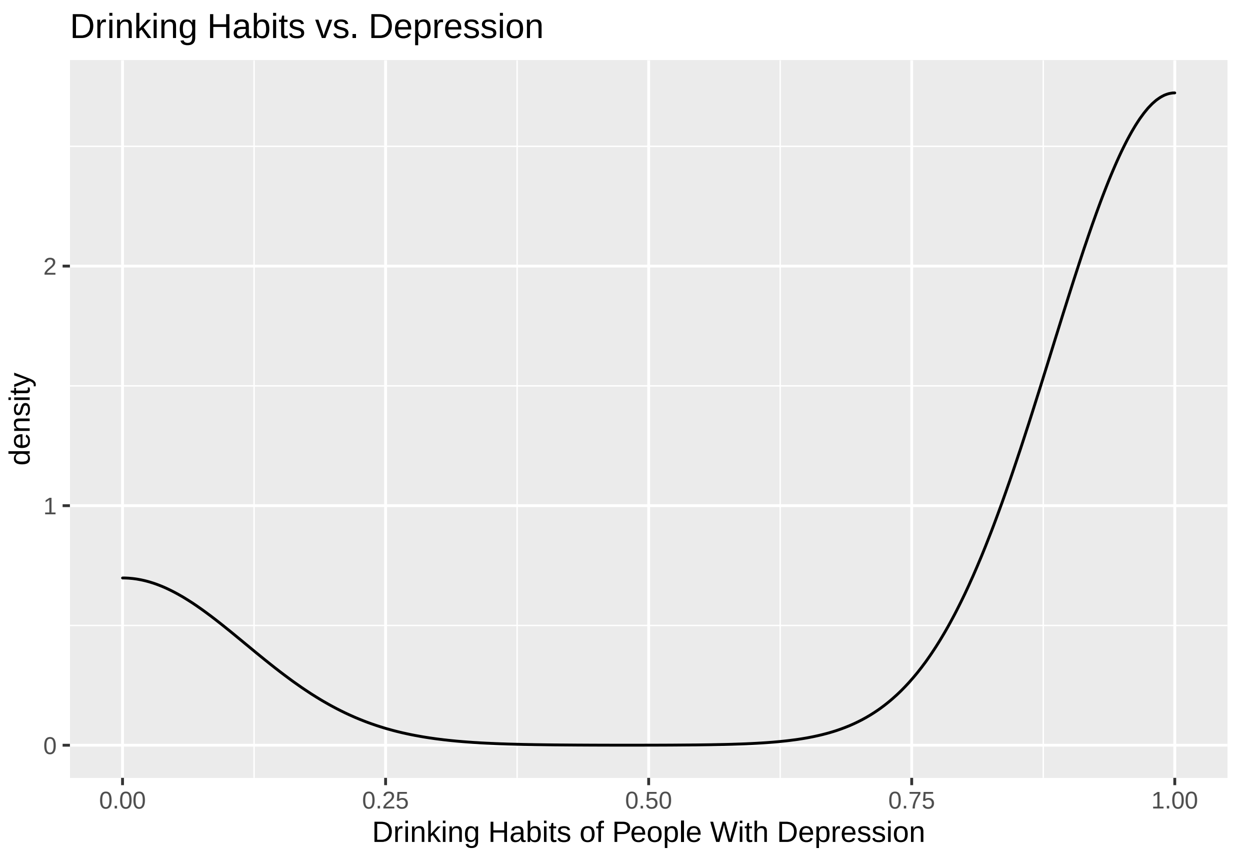
```
table(depress$age)
```

```
## 
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 
##  5  5  6  6  9 11  9  6  9  4  5  4  6  5 10  5  9  4  6  5  2  1  5  1  9  7 
## 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 
##  2  4  3  6  4  4  5  6  4  2  3  4  3  6  7  9  7  5  4  2  3  5  3  2  4  1 
## 70 71 72 73 74 75 77 78 79 80 81 82 83 89 
##  5  3  2  2  3  1  2  2  2  1  2  1  5  1
```

```
sd(depress$drink)
```

```
## [1] 0.4037161
```

```
ggplot(depress, aes(x=drink))+geom_density()+xlab("Drinking Habits of People With Depression") + ggtitl
```

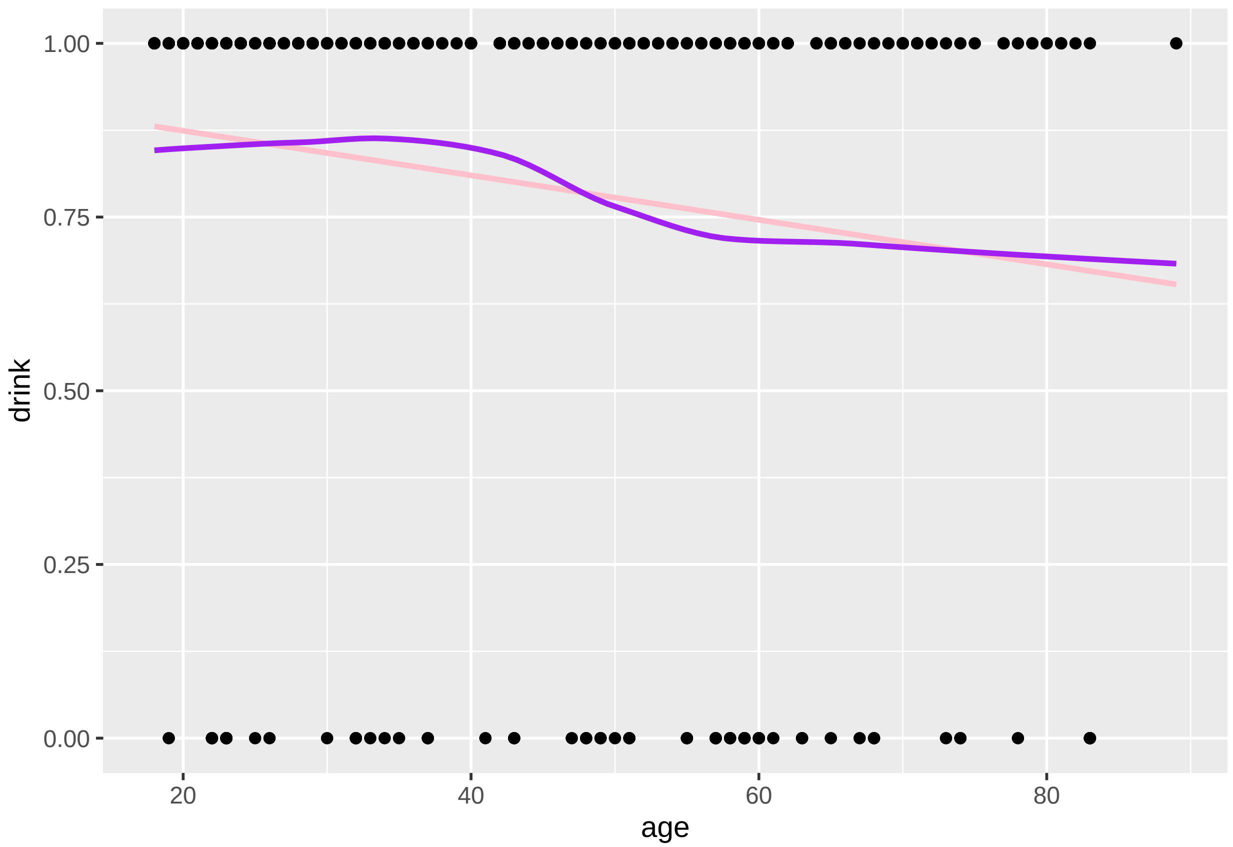


Bivariate Description

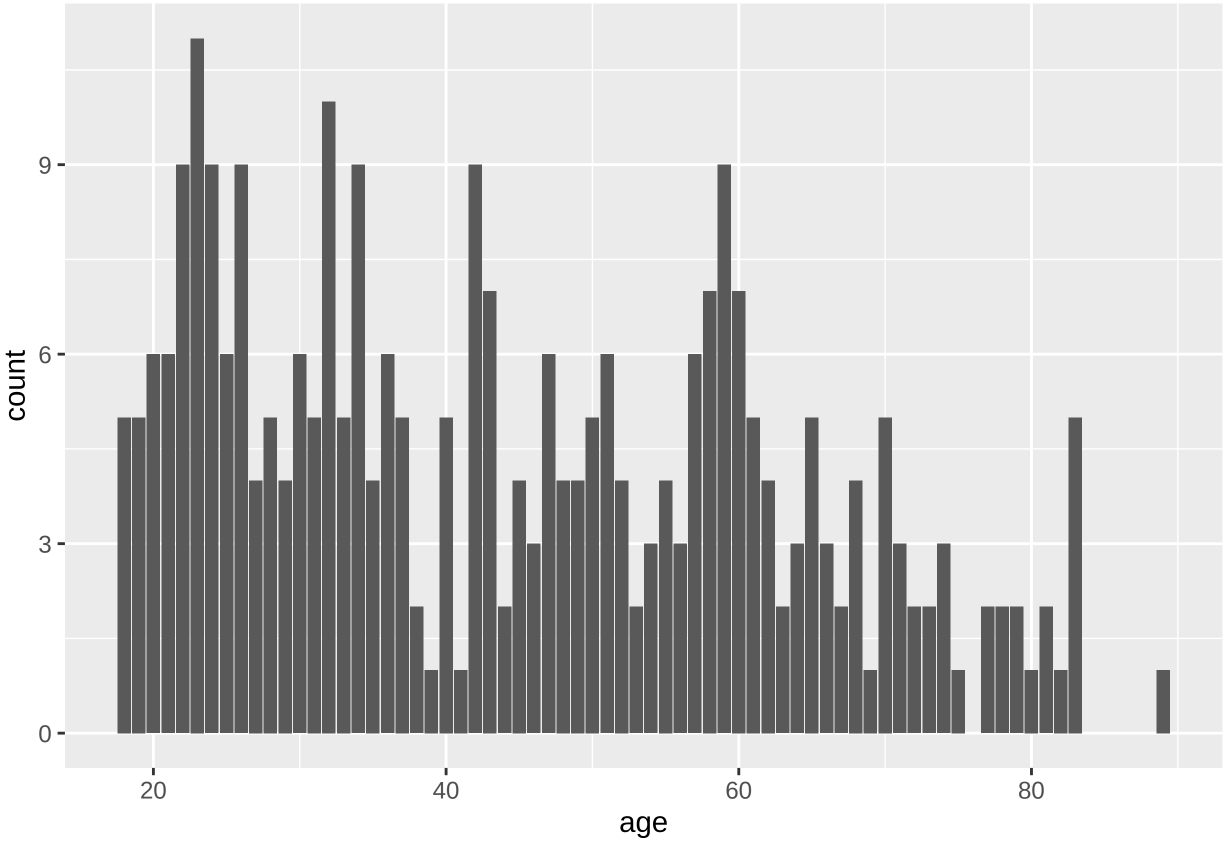
```
ggplot(depress, aes(x=age, y=drink))+geom_point()+geom_smooth(se=FALSE, method="lm", color="pink")+geom_
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```

```
## ‘geom_smooth()‘ using method = ’loess’ and formula ’y ~ x’
```

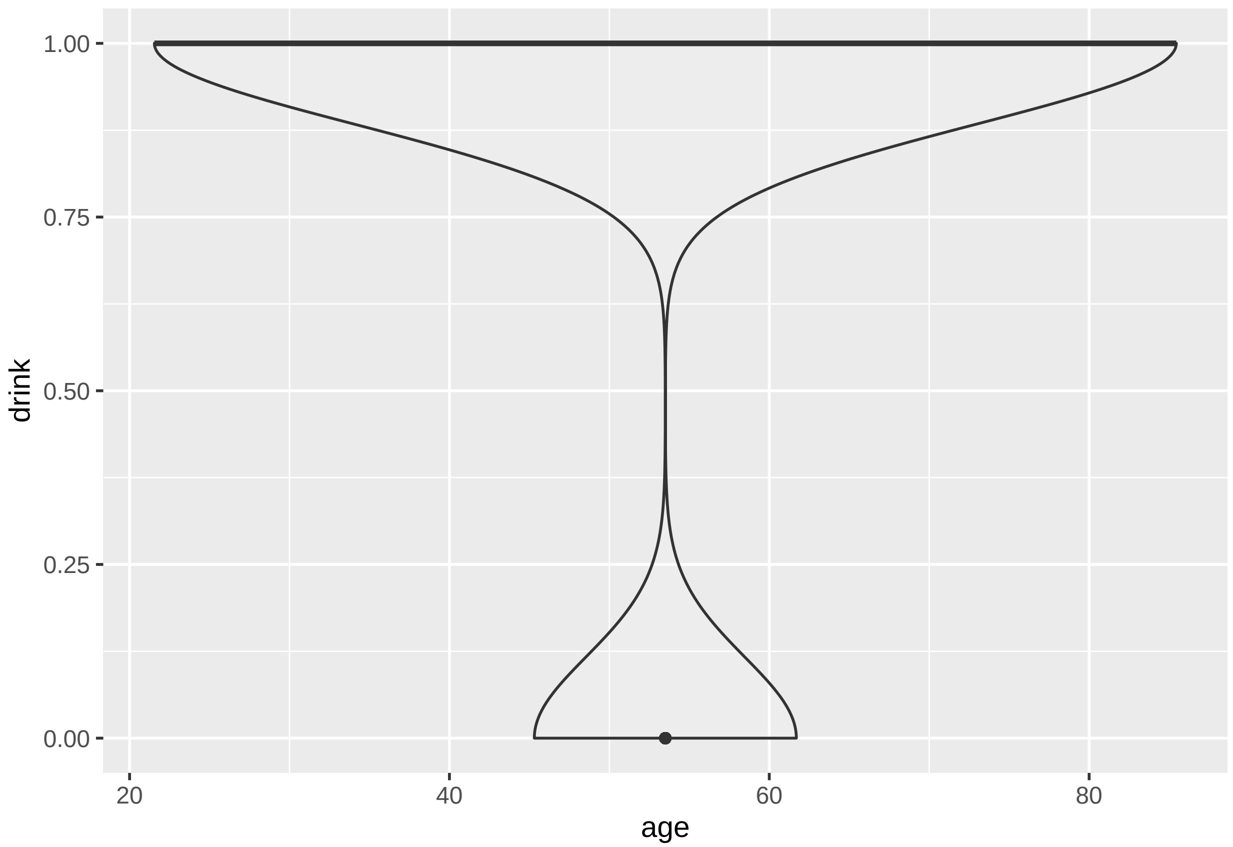


```
ggplot(depress, aes(x=age, fill=drink))+geom_bar()
```

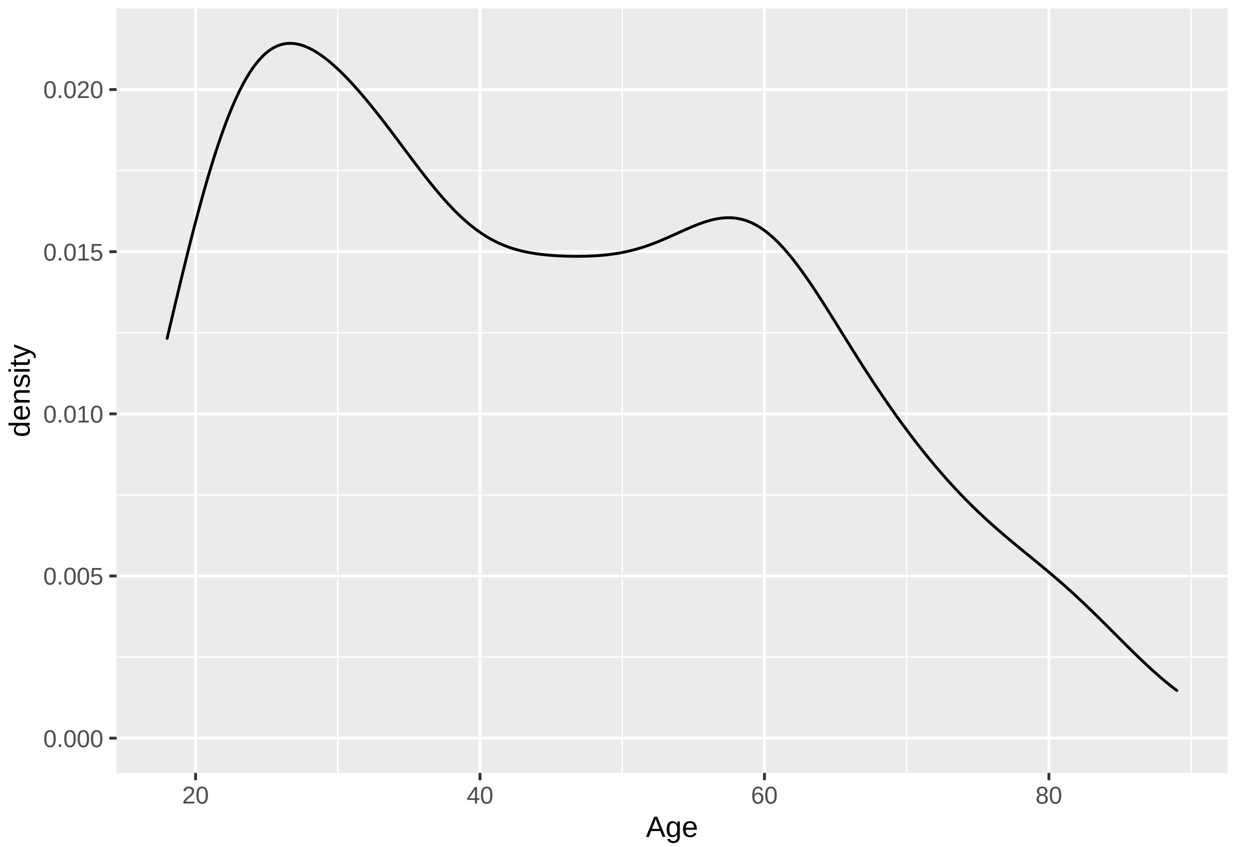


```
ggplot(depress, aes(x=age,y=drink, fill=age))+geom_violin(alpha=.1)+geom_boxplot(alpha=.5, width=.2)
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



```
ggplot(depress, aes(x= age, fill= drink)) + geom_density() + scale_fill_discrete(name="drink") + xlab("
```



## Conclusion The data analysis that I conducted above seems to support my prior hypothesis that there would be a correlation between depression and age or depression with ones drinking habits. I previously believed that those who were more middle aged would be more likely to be experiencing depression than younger adults and senior citizens. This is mostly because this age group tends to feel as though they don't have much to look forward to in their lives. This hypothesis is supported above in the boxplot of the depression and age data, where the box lies above the age groups of 20-50 years. Its also supported by my summary stats that showed the average age of those with depression was 44 years.