# Math 130R: Explanatory Data Analysis Project

## Samantha Miumi- Louie

## 2022-09-26

```
library(ggplot2)
```

```
library(sjPlot)
```

## Introduction: Depression

The data that I will be analyzing is the depression data. This data set explores about 294 people in LA County being observed for a study about depression. The two variables that I will be analyzing are age and health of those being observed. I am looking to see who will be more likely to get depression based off of their age and health, and will be testing to see the relationship between the different types of variables within the people experiencing depression.

```
depress <- read.delim("/Users/samanthamiumi/Desktop/MATH130/Data/EDA_Depression_Data.txt", header=TRUE,
dim(depress)
```

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

## Univariate Exploration

Variables being observed- "Age":

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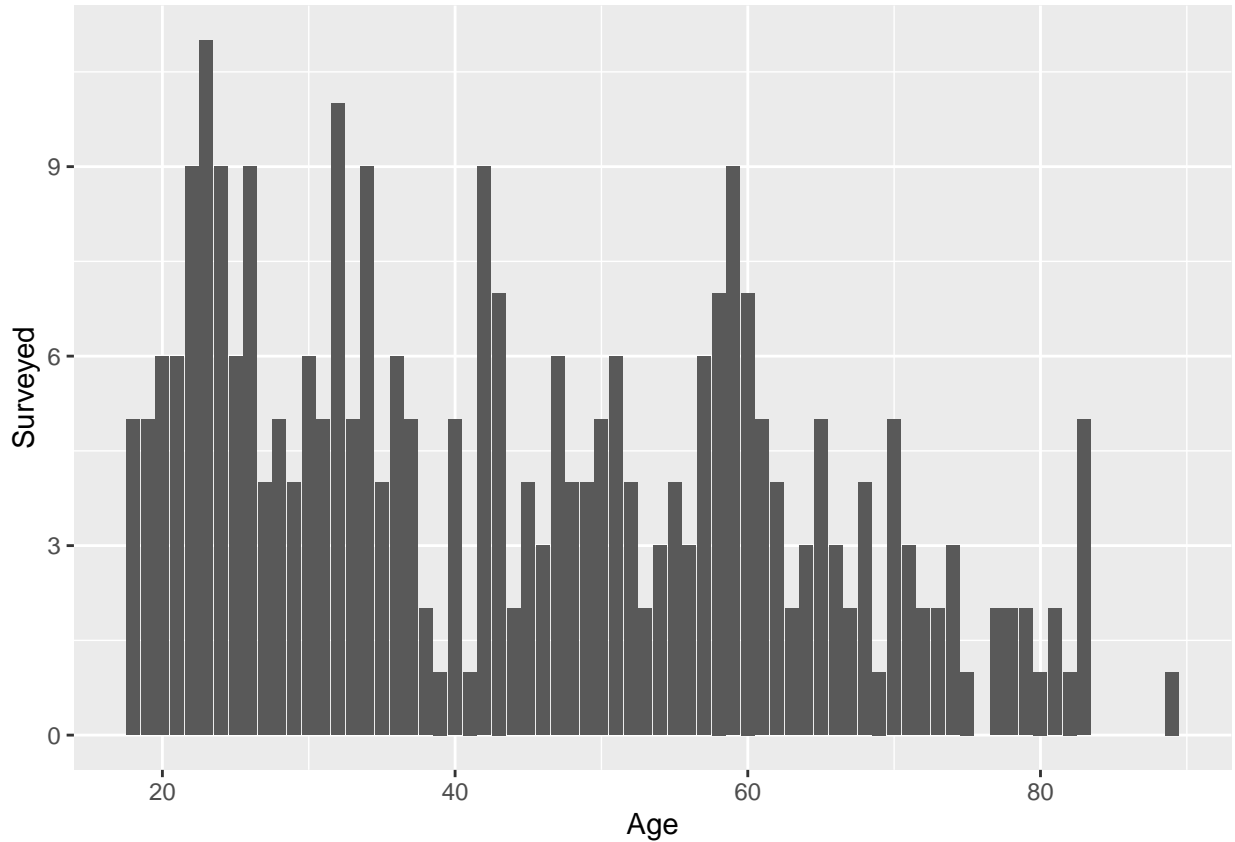
```
summary(depress$age)
```

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

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

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

```
ggplot(depress, aes(x=age, fill=age)) + geom_bar() + xlab("Age") + ylab("Surveyed")
```



The graph above shows the relationship between one's age and whether they are suffering from depression. Those that are closer to the age range of around 18-30 years old are more likely to suffer from depression than those who are 35-90 years old.

Variables being observed- "Health":

```
table(depress$health)
```

```
##
##   1   2   3   4
## 130 115  35  14
```
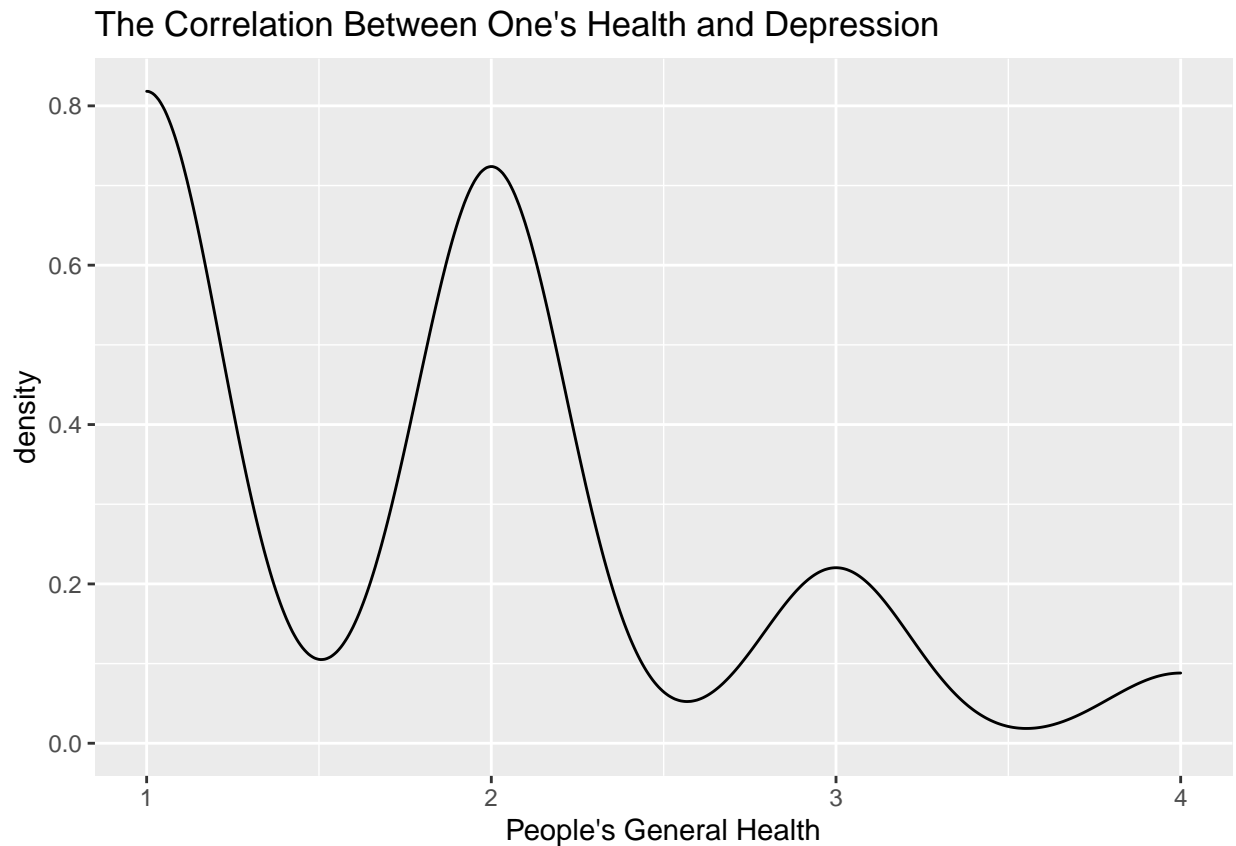
```
summary(depress$health)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   1.772   2.000   4.000
```

```
sd(depress$health)
```

```
## [1] 0.8379466
```

```
ggplot(depress, aes(x=health)) + geom_density() + xlab("People's General Health") + ggtitle("The Correla
```

## The Correlation Between One's Health and Depression



The graph above shows the correlation between one's health and depression. Those who have (1) excellent health are more likely to suffer from depression than those who have (4) poor health.

## Bivariate Exploration

```
ggplot(depress, aes(x= health, y=age)) + geom_point() +
  geom_smooth(se= FALSE, method= "lm", color= "blue") +
  geom_smooth(se= FALSE, color= "red")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

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

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 0.985
```

3

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1.015
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 9.109e-31
```
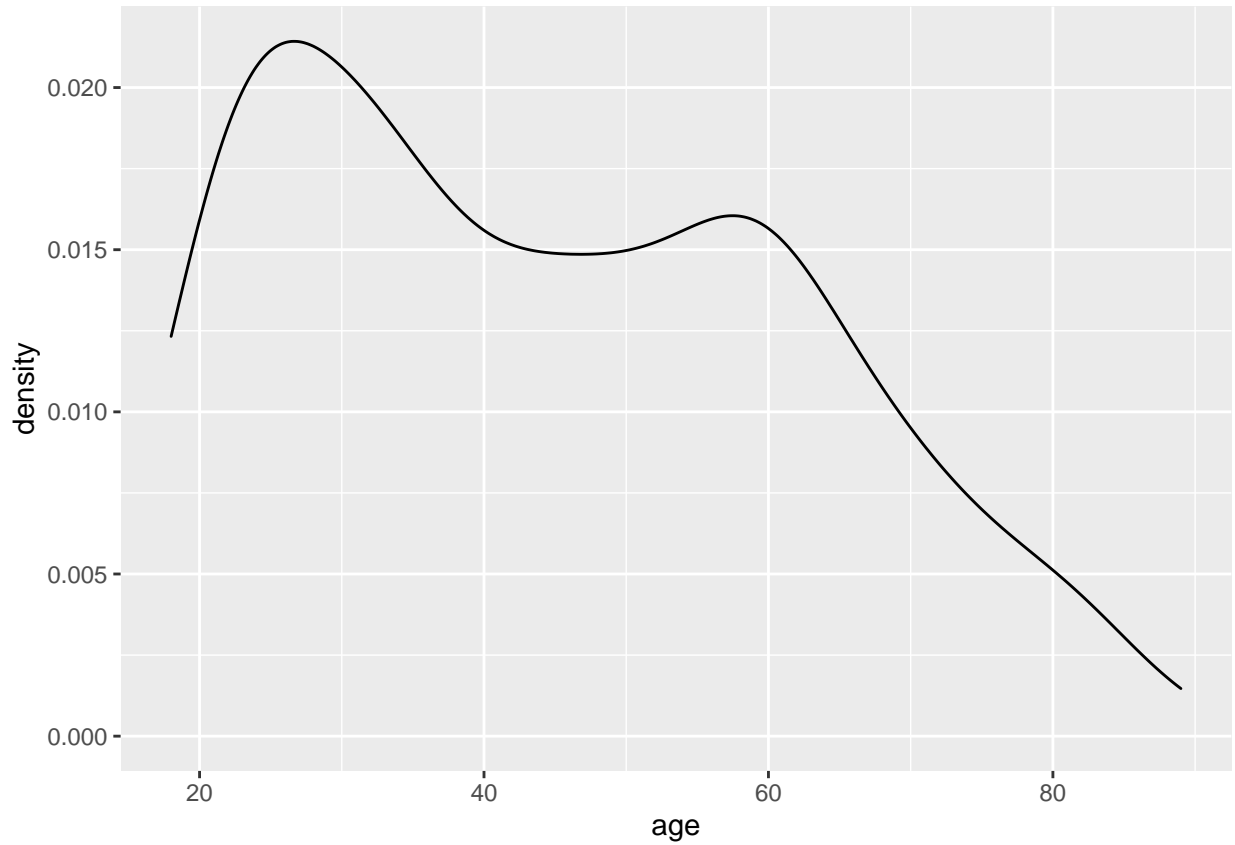
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1
```
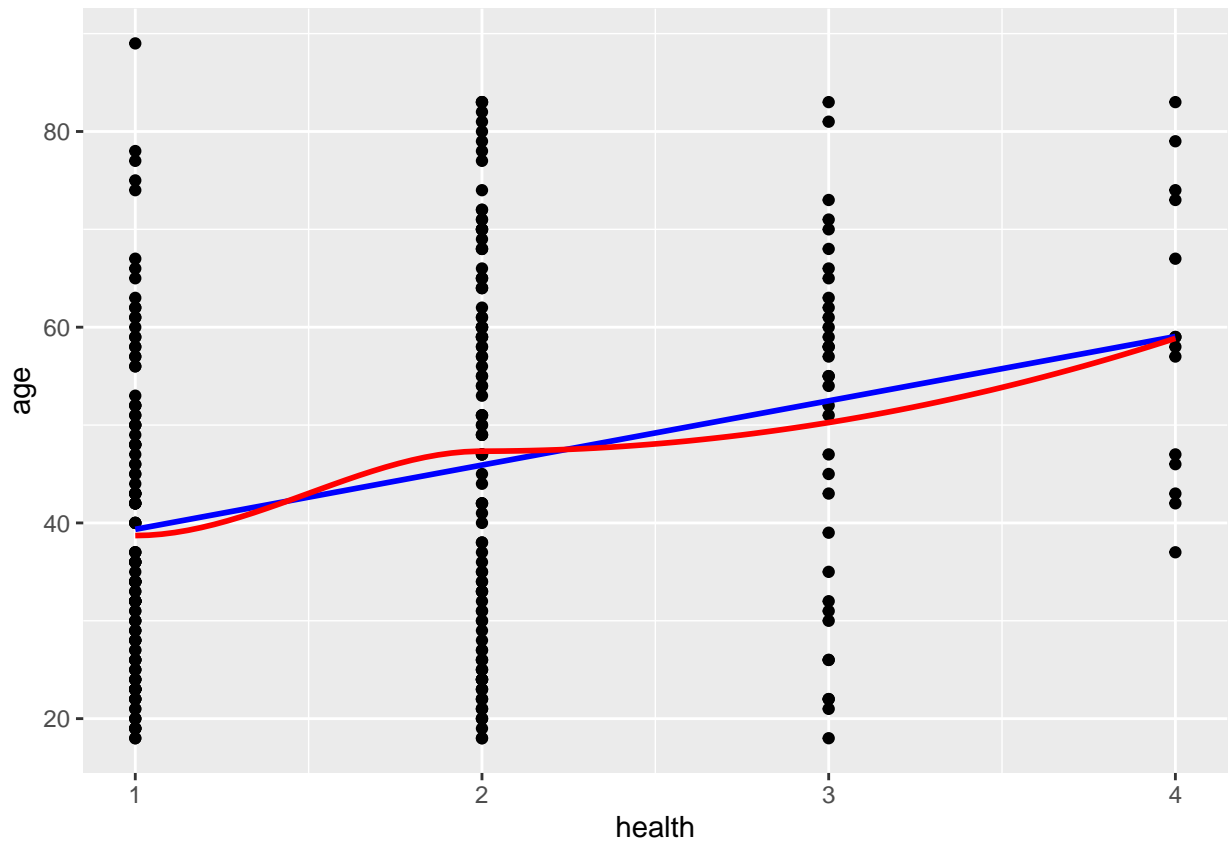


```
ggplot(depress, aes(x= age, fill= health)) + geom_bar()
```
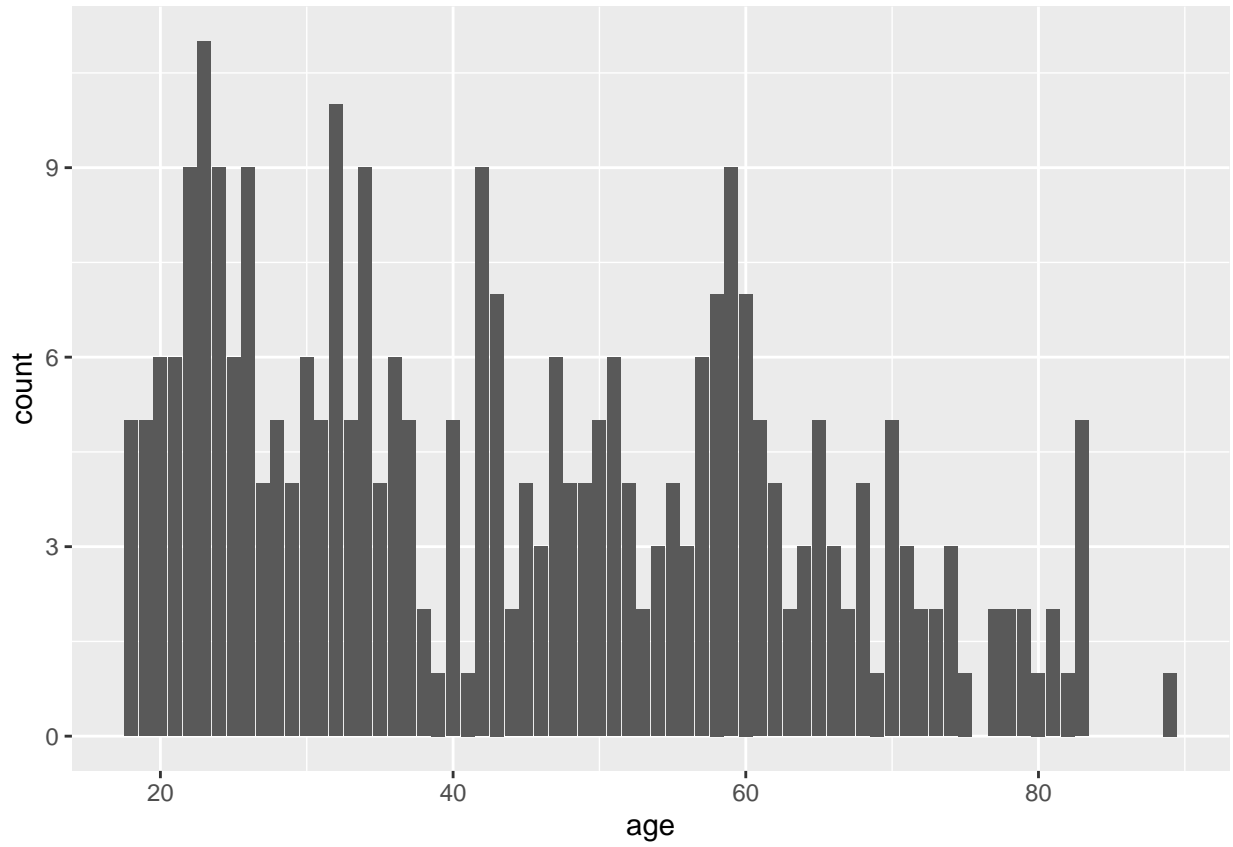
```
ggplot(depress, aes(x=age, fill=health)) + geom_density() + scale_fill_discrete(name="health") + xlab("a
```

## Conclusion

When I began this project, it was very interesting to me to see so many variables that may affect one's suffering of depression. Once I started to explore the age and health variables and seeing if there is any relationship between the two, I found that people who are in their 20's-30's (years old) suffer from depression more than those in their 40's-80's (years old) and that the people with better health (generally) suffer from depression more than those who have poor(er) health. This data seems to back up my data. You can see in the age bar graph that mostly those in your 20's have depression more than those who are younger or older.