

# 130 Project

Nicole

2/25/2022

Intro:

The “Depression” data set will be the focal point for this Exploratory Data Analysis. This data set consists of information collected from the first array of interviews with 294 Los Angeles County adult residents. These interviewees were asked about their overall mental and physical health as well as a variety of lifestyle practices and general demographics. The “age” and “sex” variables will be analyzed to determine whether there is a correlation between someone’s age and sex and whether or not they have depression. Both “age” and “sex” variables are categorical values. Age was reported between 18 and 89 years. The “sex” variable provided three potential responses, 0,1 or 2 (Choose not to identify, Male, or Female).

```
depress <-  
read.delim("/Users/nicolemccarthy/Desktop/depress_081217.txt",  
header=TRUE, sep="\t")  
dim(depress)  
## [1] 294 37
```

## Including Plots

You can also embed plots, for example:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##  18.00  28.00   42.50   44.41  59.00   89.00  
  
summary(depress$ age, depress$sex)  
  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##  18.00  28.00   42.50   44.41  59.00   89.00  
  
round(prop.table(table(depress$sex, depress$age), margin=1),3)  
  
##  
##           18    19    20    21    22    23    24    25    26    27  
28    29  
##    0 0.018 0.018 0.036 0.018 0.045 0.018 0.009 0.018 0.045 0.018  
0.009 0.018  
##    1 0.016 0.016 0.011 0.022 0.022 0.049 0.044 0.022 0.022 0.011  
0.022 0.011  
##
```

```

##      30      31      32      33      34      35      36      37      38      39
40     41
##    0 0.018 0.027 0.054 0.000 0.009 0.009 0.027 0.027 0.009 0.009
0.036 0.000
##    1 0.022 0.011 0.022 0.027 0.044 0.016 0.016 0.011 0.005 0.000
0.005 0.005
##
##      42      43      44      45      46      47      48      49      50      51
52     53
##    0 0.027 0.009 0.009 0.018 0.018 0.018 0.027 0.027 0.018 0.009
0.009 0.009
##    1 0.033 0.033 0.005 0.011 0.005 0.022 0.005 0.005 0.016 0.027
0.016 0.005
##
##      54      55      56      57      58      59      60      61      62      63
64     65
##    0 0.000 0.009 0.009 0.009 0.027 0.036 0.027 0.018 0.027 0.009
0.027 0.009
##    1 0.016 0.016 0.011 0.027 0.022 0.027 0.022 0.016 0.005 0.005
0.000 0.022
##
##      66      67      68      69      70      71      72      73      74      75
77     78
##    0 0.000 0.000 0.000 0.000 0.018 0.000 0.009 0.018 0.009 0.000
0.009 0.009
##    1 0.016 0.011 0.022 0.005 0.016 0.016 0.005 0.000 0.011 0.005
0.005 0.005
##
##      79      80      81      82      83      89
##    0 0.009 0.000 0.009 0.000 0.009 0.000
##    1 0.005 0.005 0.005 0.005 0.022 0.005

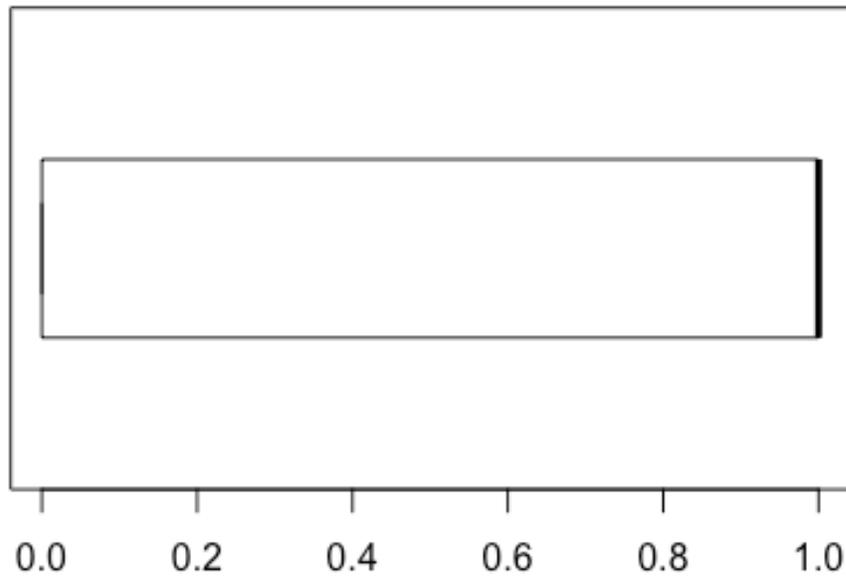
```

```

boxplot(depress$sex, horizontal =TRUE, main= "Distribution of Sex in
the Study of Depression")

```

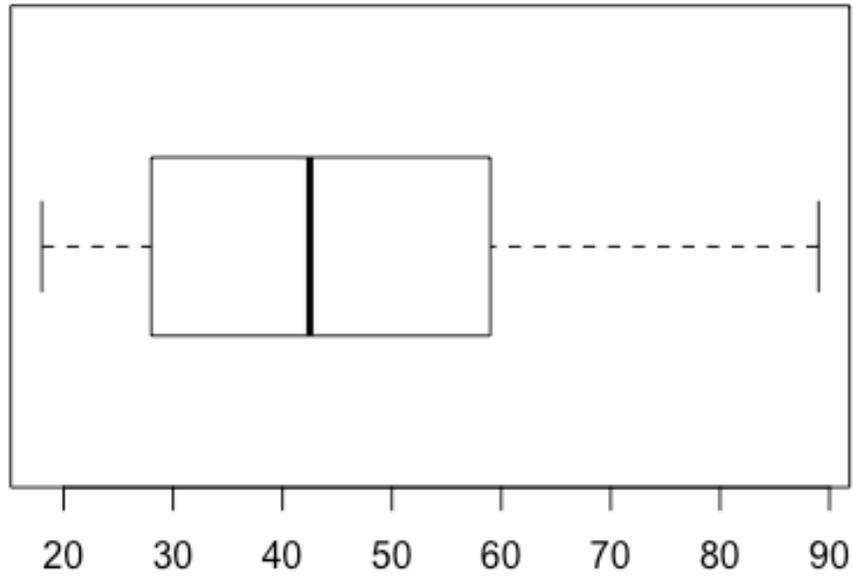
## Distribution of Sex in the Study of Depression



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

This chart shows the distribution of sex in the data set, it is recorded from 0, 1, and 2.

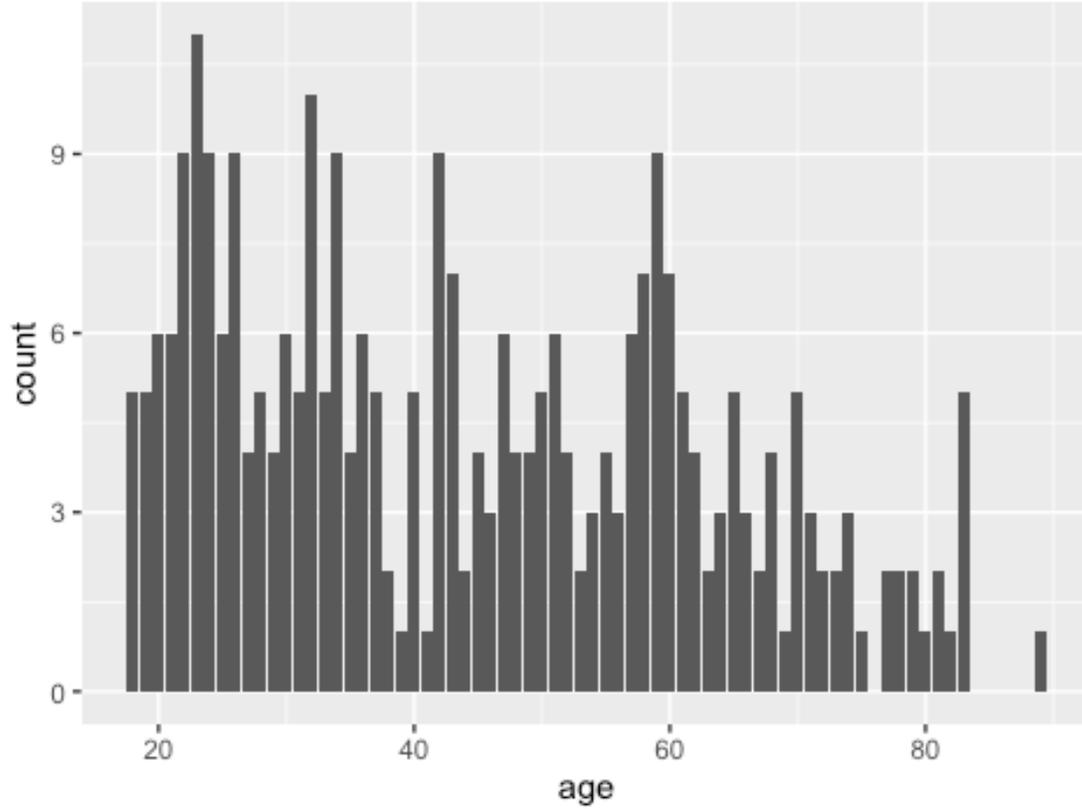
## Distribution of Age in the Study of Depression



```
ggplot (depress, aes(x= age, fill = sex)) + geom_bar() + ggtitle("Age  
and Sex of Individuals")
```

This ggplot shows the age range of the individuals in the data set, it ranges from the youngest being 18 and the oldest being 89 and the mean of the group was 44.

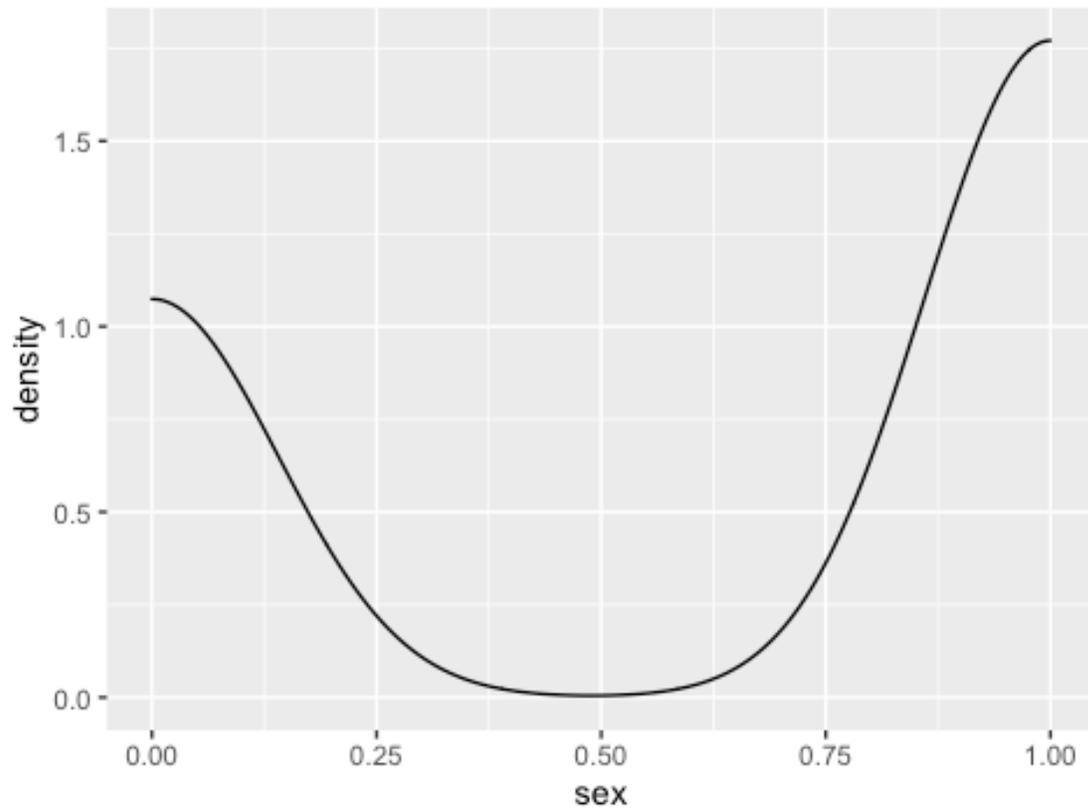
## Age and Sex of Individuals



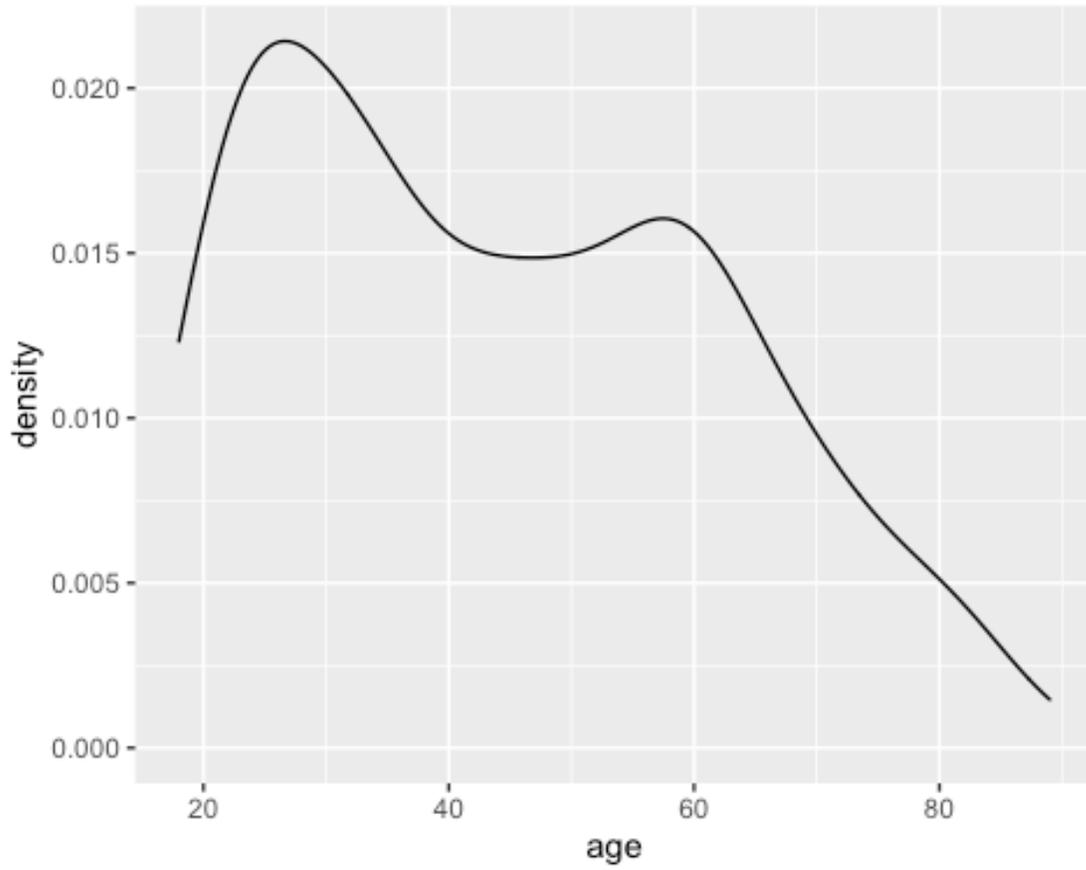
```
ggplot (depress, aes(x=sex, fill= sex)) +geom_density(alpha=.5) +  
ggtitle("Sex of Individuals")
```

This plot shows the amount of people from each age group, the most common age group being just over 20 years old.

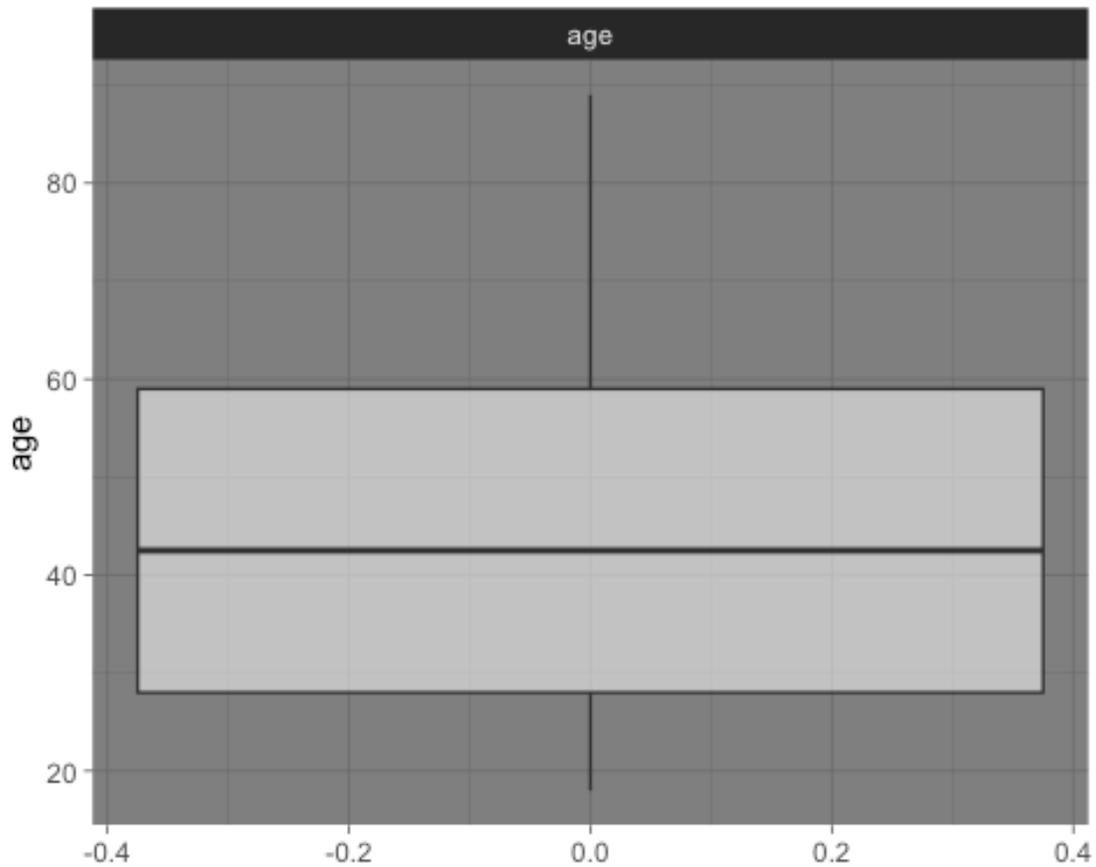
Sex of Individuals



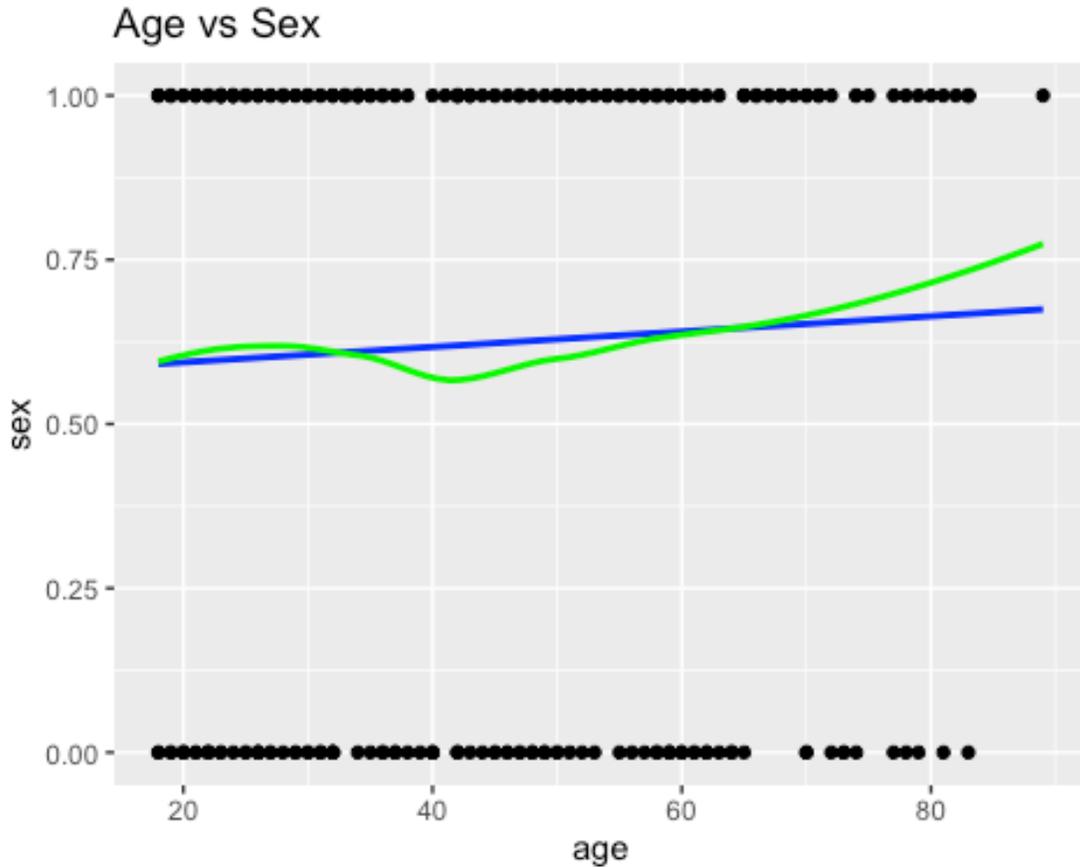
```
ggplot(depress, aes(x=age, fill= sex)) + geom_density(alpha=.3)
```



```
ggplot(depress, aes(y= age, fill= age)) + geom_boxplot(alpha=.5)+  
facet_wrap(~'age') + theme_dark() + scale_fill_brewer(name="Age and Sex  
of Individuals", palette="Set1")
```



```
ggplot(depress, aes(x=age, y=sex)) + geom_point() +  
geom_smooth(se=FALSE, method="lm", color="blue") +  
geom_smooth(se=FALSE, color="green")+ ggtitle("Age vs Sex")  
  
## `geom_smooth()` using formula 'y ~ x'  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



This plot is the correlation between Age vs. Sex and the amount of people that reported to have depression dependent on their age and their sex.

Conclusion:

Overall I found that there was a correlation between Age and Sex. The data set recorded from 294 individuals reported whether or not they believe that they struggle from depression and then also recorded their age and sex. This ranged from 18 to 89 years old. It was found that those between 20-30 years of age struggle the most with depression.