Research project

Josh Barron

Introduction:

For my final project I will analyzing data inputs from the data set titled Depression which looks into factors that contribute to depressed adults residing in Los Angeles County at a variety of ages.

Research Question:

Are those under 40 years old more depressed than those who are older?

Analyzed Variables:

-Age of participant

-Whether the individual is depressed or not where normal=0 and depressed=1

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
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
```

```
depress$newcases <- factor(depress$cases, labels=c( "Not Depressed", "Depressed"))
table(depress$newcases)
```

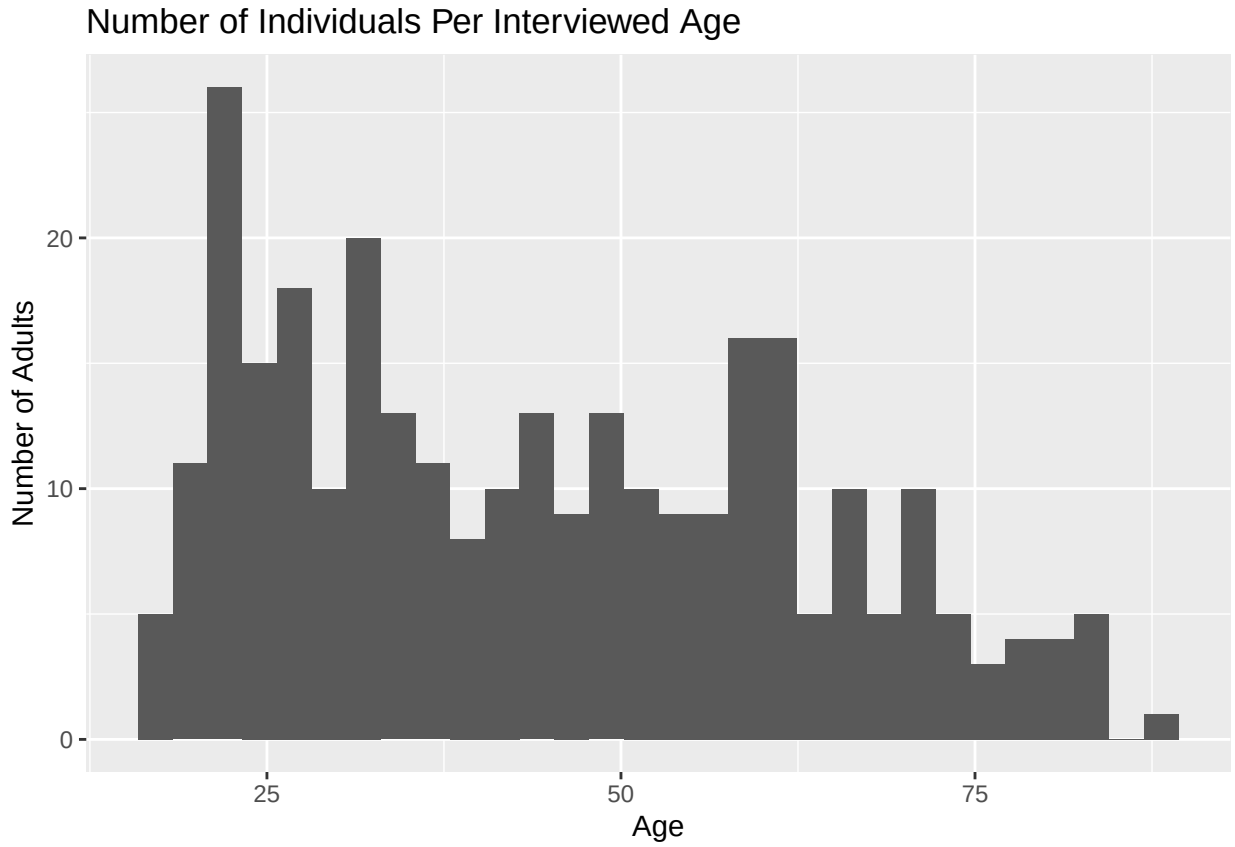
```
##
## Not Depressed     Depressed
##           244            50
```

Univariate Analysis 1:

```
ggplot(depress, aes(x=age, fill=age)) +
geom_histogram() +
ylab("Number of Adults") +
xlab("Age") +
ggtitle("Number of Individuals Per Interviewed Age") +
scale_fill_brewer(palette="S")
```

```
## Warning in pal_name(palette, type): Unknown palette S
```

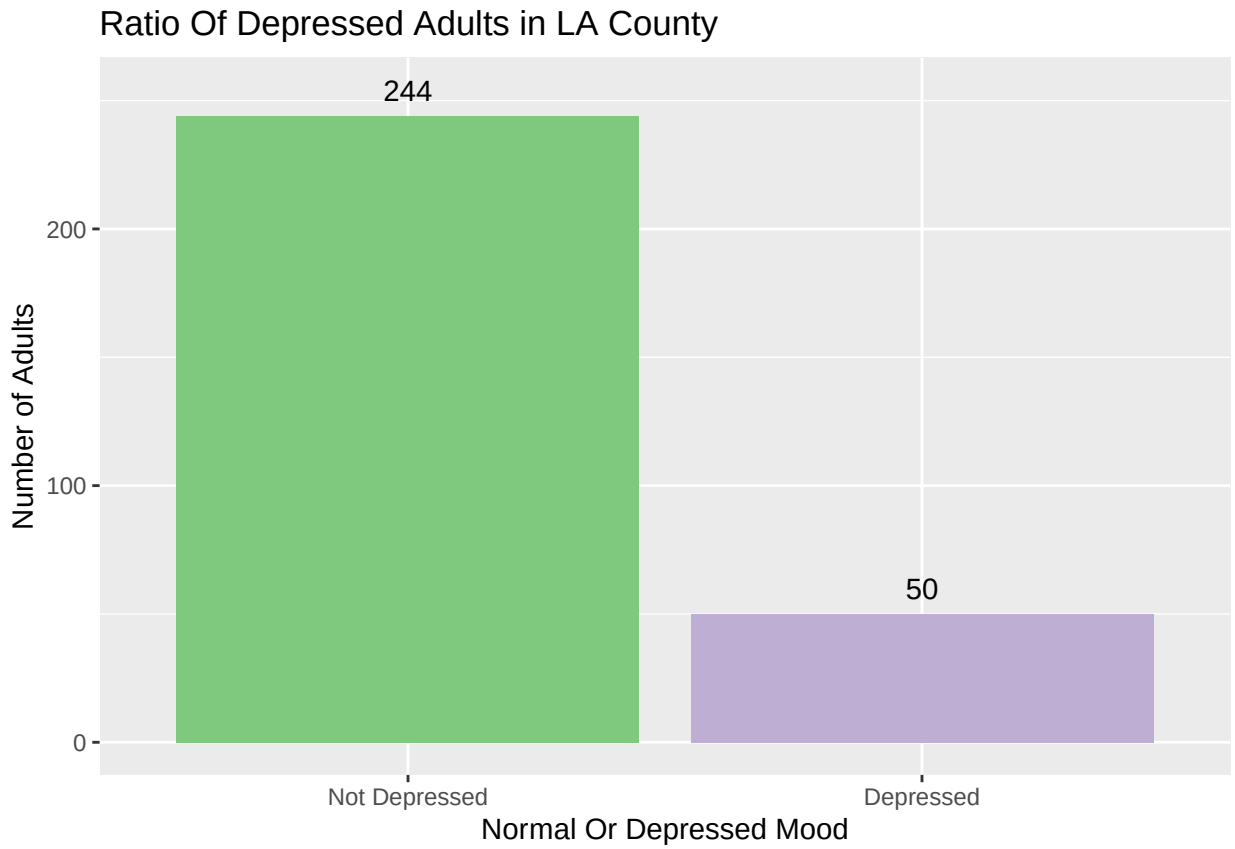
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



As we can see from the table above this study has a variety of participants ranging from 18 all the way to 80. This will allow for the later point to be tested in which we analyze if age has a direct correlation to amount of depression experienced.

Univariate Analysis 2:

```
ggplot(depress, aes(x=newcases, fill=newcases)) +
geom_bar(show.legend = FALSE) + ylab("Number of Adults") + xlab("Normal Or Depressed Mood") + ggtitle("
```

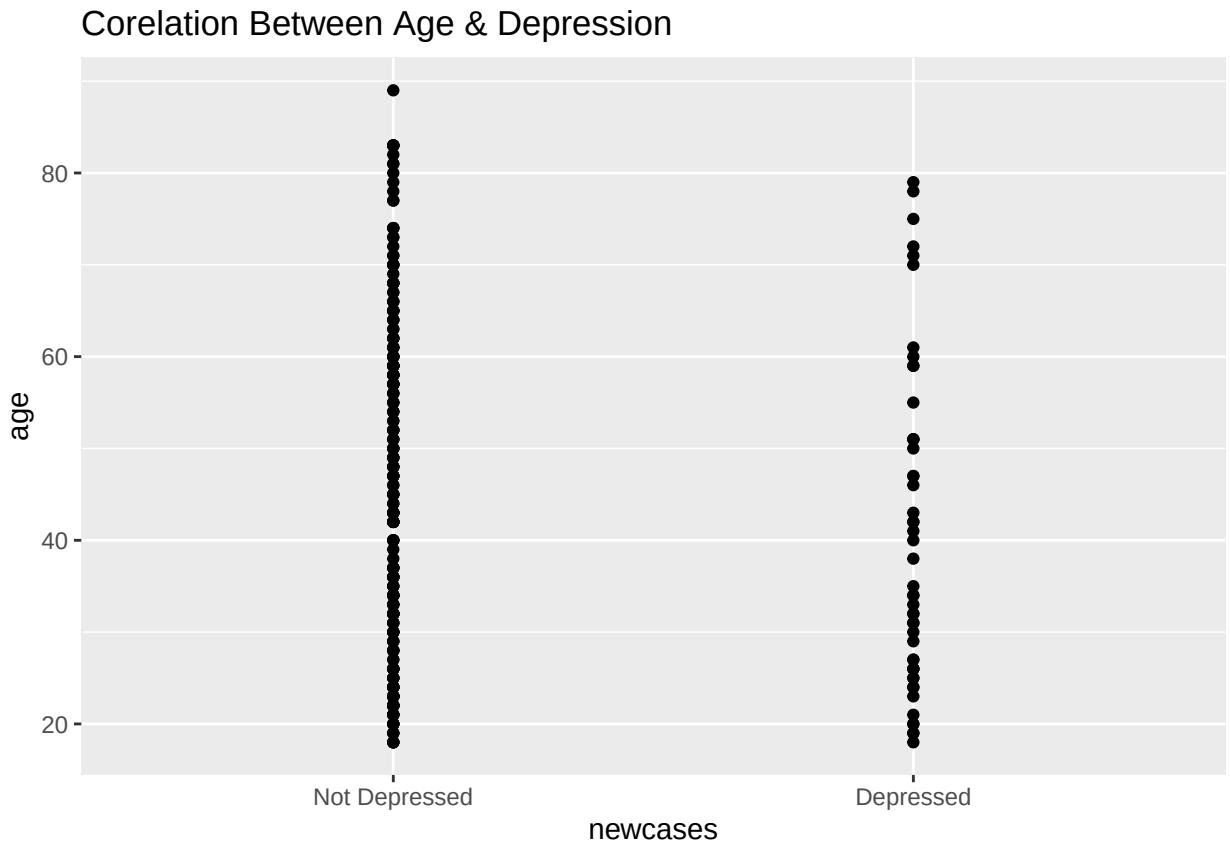


This second graph demonstrates the contrast between who was identified as having a normal mood versus an actual depressive one. The ratio here is about 1 depressed person for every 4 non depressed persons. These results are better than what I first though when looking at the data which is a postive. ## Bivariate Analysis:

```
ggplot(depress, aes(x=newcases, y=age)) + geom_point()  +
geom_smooth(se=FALSE,   method="lm",    color="blue")   +
geom_smooth(se=FALSE,   color="green")+ ggtitle("Corelation Between Age & Depression")
```

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

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



This last graph demonstrated the density between those who are not depressed and those who are by there different age groups. Even though there are less dots overall on the depressed category you can still see that a majority of them are on the bottom towards the 20-40 year old range.

Closing Statement

Overall, after viewing these tables and analyzing one another I would say my research question was in fact true. When looking at the given sample it does appear that younger people from ages 20-40 are more depressed than those who are old. This idea could definitely be further explored by follow up questions or looking at other contributing factors in the Depress data set. Although this is a very surface level analysis, it still shows the that it seems the youths mental health is progressively getting worse and that it should be a topic of discussion.