

Pablo Almazan

2023-09-20

```
depress <- read.table("/Users/pabloalmazan/Desktop/MATH130/data/depress_081217.txt",  
header=TRUE, sep="\t")  
library(dplyr)
```

```
##  
## 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
```

```
library(ggplot2)
```

The data I am analyzing originates from interviews based on depression and its relation to Los Angeles residents. There are 294 observations to make sense of. This data is interpreted as variables such as; ID, sex, age, marital status, education, employment, income and religion. With the collection of this data, I hypothesize that marital status, and education status combined correlate with depression.

Univariate Exploration

```
table(depress$marital)
```

```
##
##      Divorced      Married Never Married      Separated      Widowed
##           43           127           73           13           38
```

With a simple table summary, we are able to see how marital status may affect depression. With the numbers alone, we can only make a side by side speculation. Without a graphic visual aid, we may infer that married people are more depressed than divorced, widowed and even separated people.

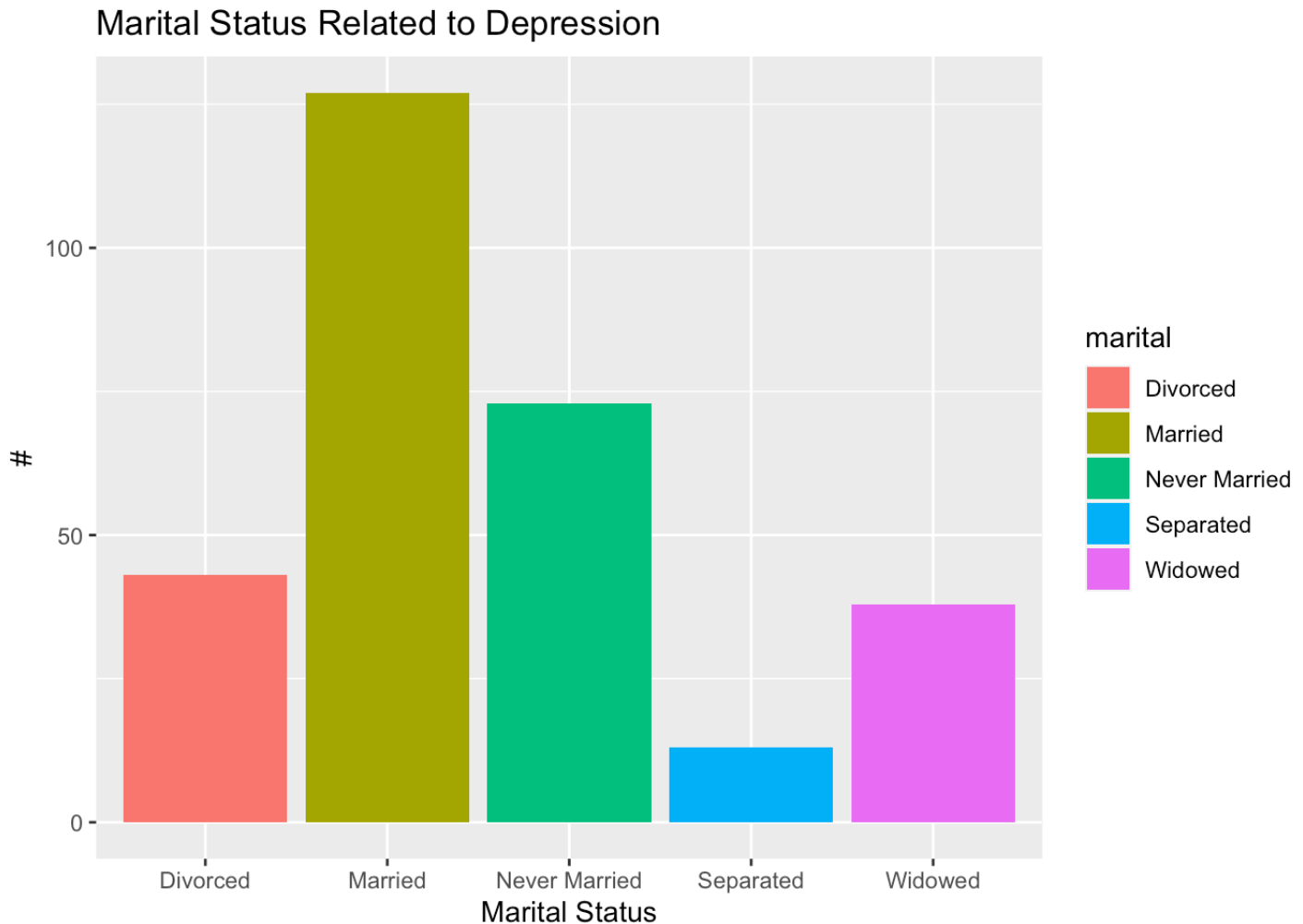
```
table(depress$educat)
```

```
##
##      <HS      BS      HS Grad      MS      PhD Some college
##           5      43      114      14      9      48
##      Some HS
##           61
```

Similar to our marital table, we can only make side by side speculations with the numbers alone. The majority of depressed people appear to be in the HS Grad sector. Those who had less than a HS degree and those with a PhD had the smallest

numbers.

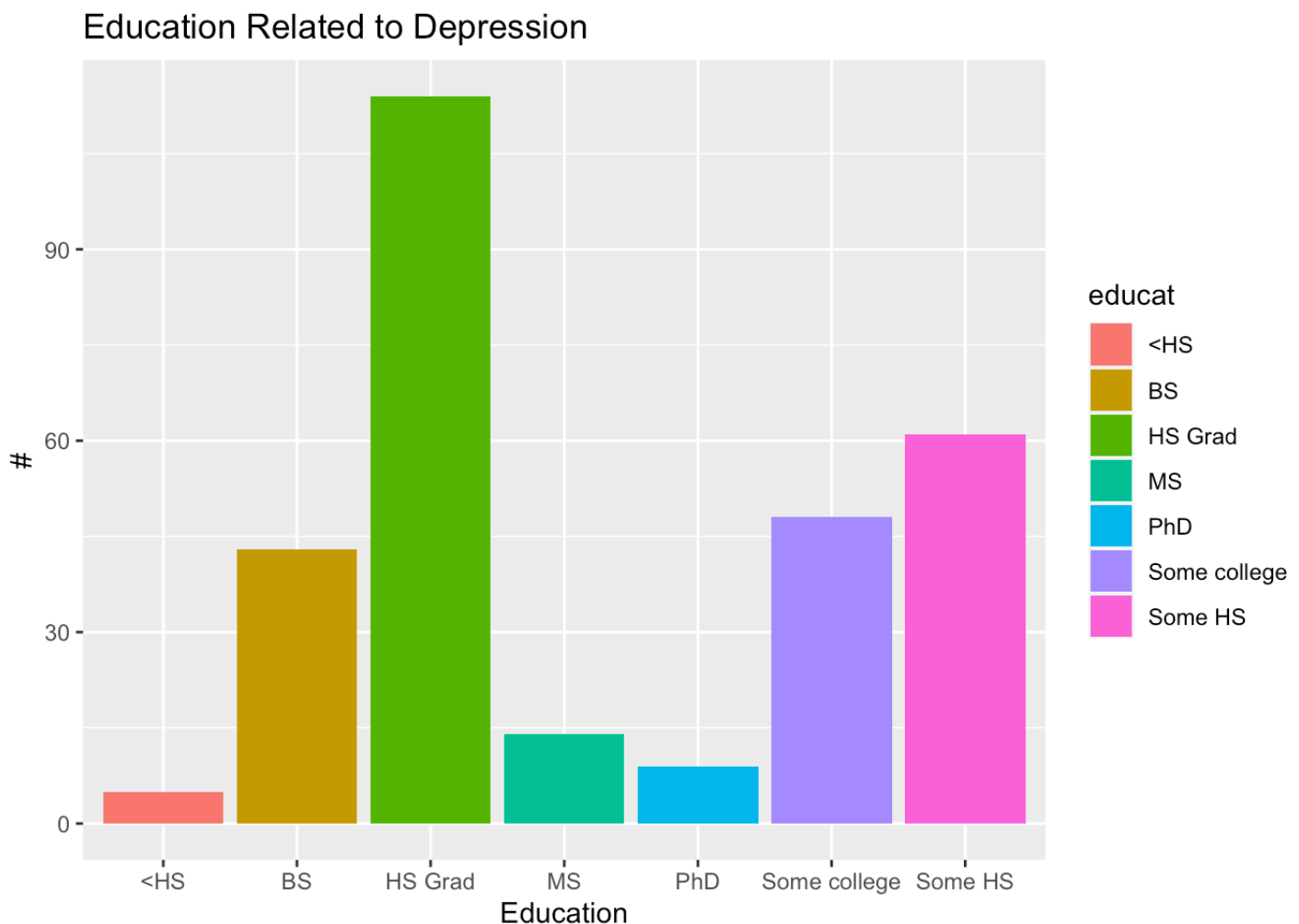
```
library(ggplot2)
ggplot(depress, aes(x = marital, fill = marital)) +
  geom_bar() +
  xlab("Marital Status") +
  ylab("#") +
  ggtitle("Marital Status Related to Depression")
```



With a graphic aid, we have more evidence to support a theory stating that marital status affects rates of depression. The bar graph shows that married people are more

depressed, compared to other non-married people. In addition, a supported theory can be made that those who are separated report to be least depressed.

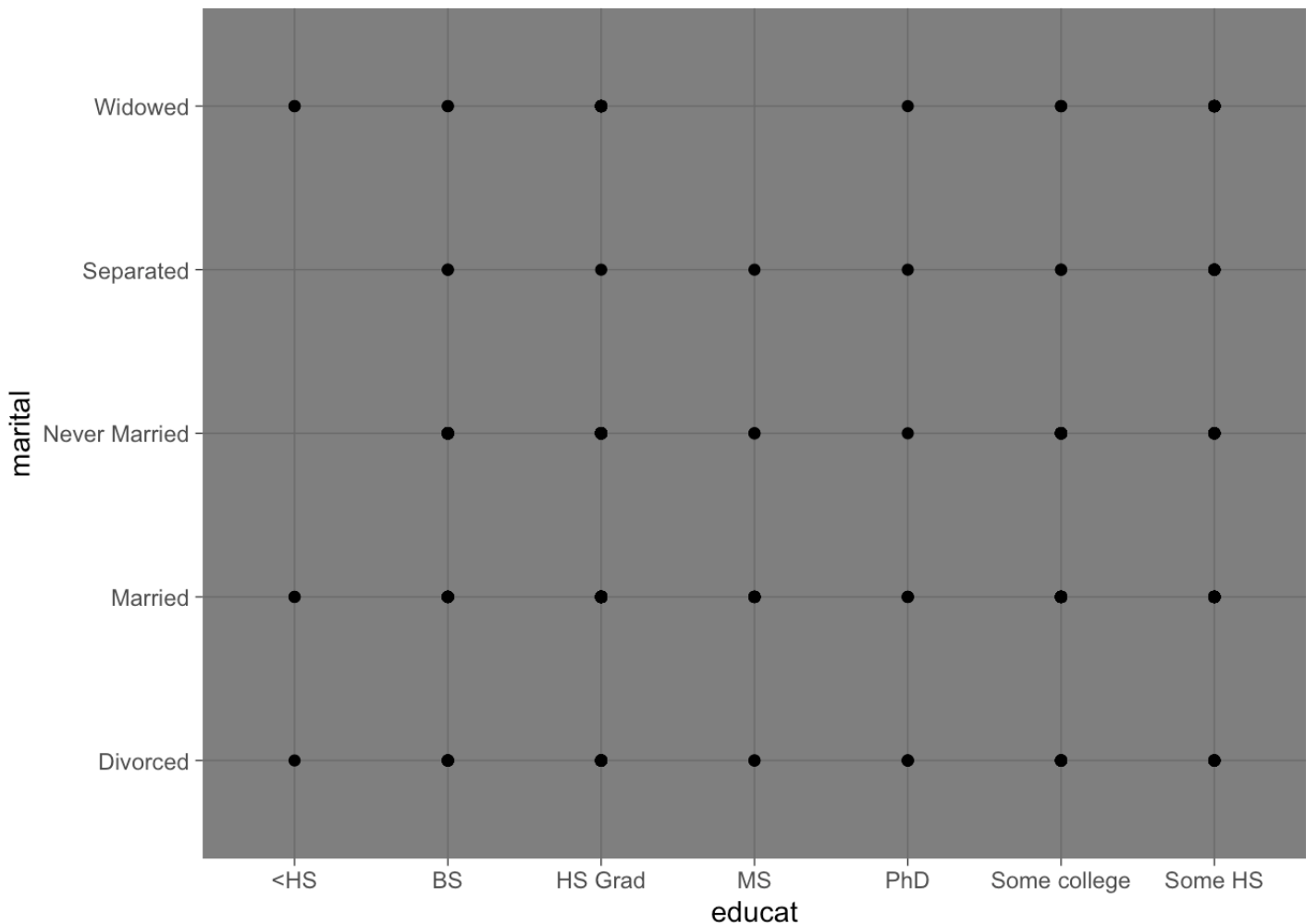
```
library(ggplot2)
ggplot(depress, aes(x = educat, fill = educat)) +
  geom_bar() +
  xlab("Education") +
  ylab("#") +
  ggtitle("Education Related to Depression")
```



Variable “educat”, shows us there are levels

in reports of depression amongst different education levels. Those who reported to only have a high school education appear to have the highest rates of depression. Those who reported to have a PhD, or no degree at all, seemed to have the least reports of depression. Those who reported to have some college appear to be in the mean.

```
ggplot(depress, aes(x=educat, y=marital)) + geom_point()+theme_dark()
```

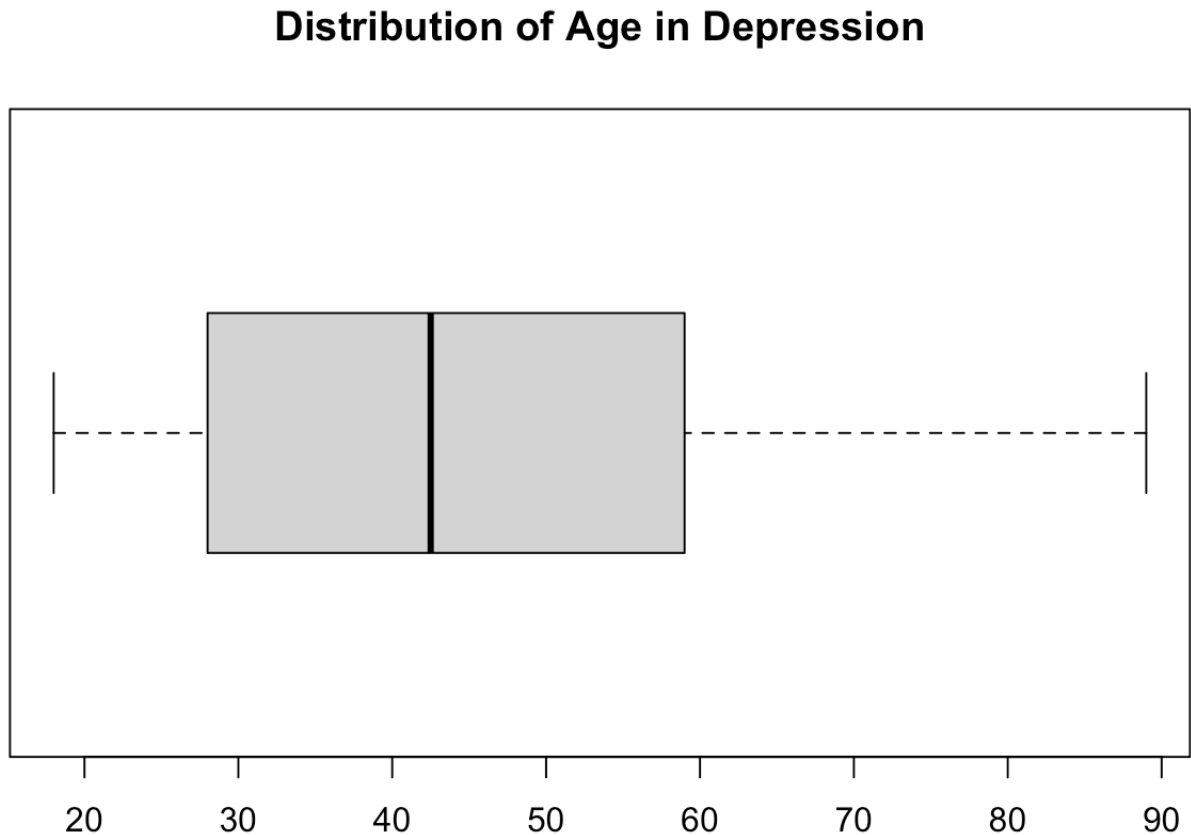


Unfortunately a scatter plot does not show a relationship between the two variables. This may be due to a lack of variation, or because the observations are non-numerical. To fix this, we can re-introduce another variable, such as age.

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

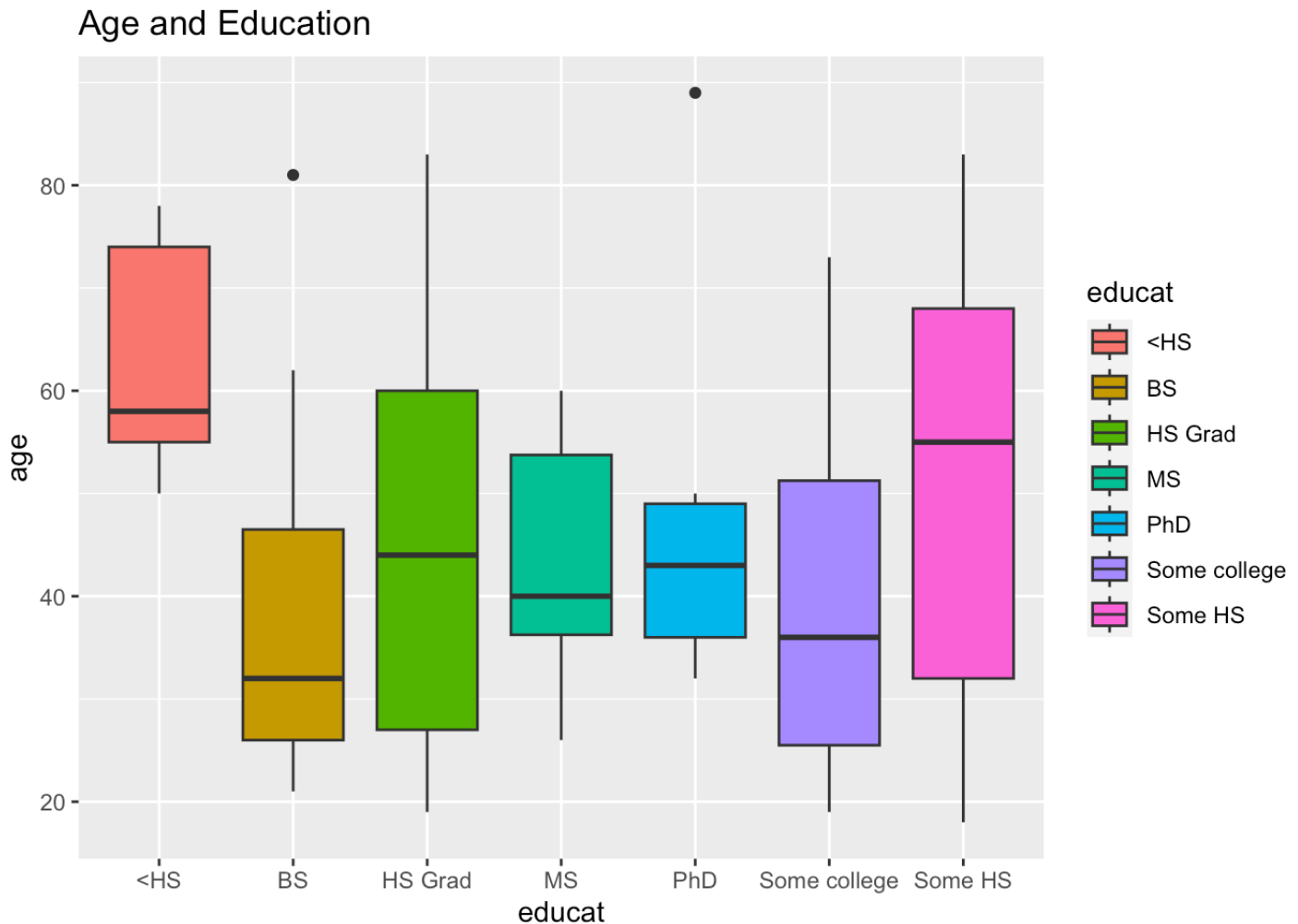
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	18.00	28.00	42.50	44.41	59.00	89.00

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



With age as a variable, we now have continuous data that allows us to make a comparison.

```
ggplot(depress, aes(x=educat, y=age, fill=educat)) + geom_boxplot() + ggtitle("Age and Education")
```



Conclusion

Originally, I discovered that two categorical data sets cannot be combined to make a graph such as scatter plot. This did not allow me to prove the theory that education and marital status combined, affect rates of depression. With a numerical variable, such as age, we can make a better comparison. Using age and education levels, a box

scatter plot supports another theory. Those who have a degree in between a high school and a post secondary degree, who report depression, tend to average in their 30's and 40's.