

# Math 130 Project

Mackenzie Vrablick

2022-09-19

## Introduction

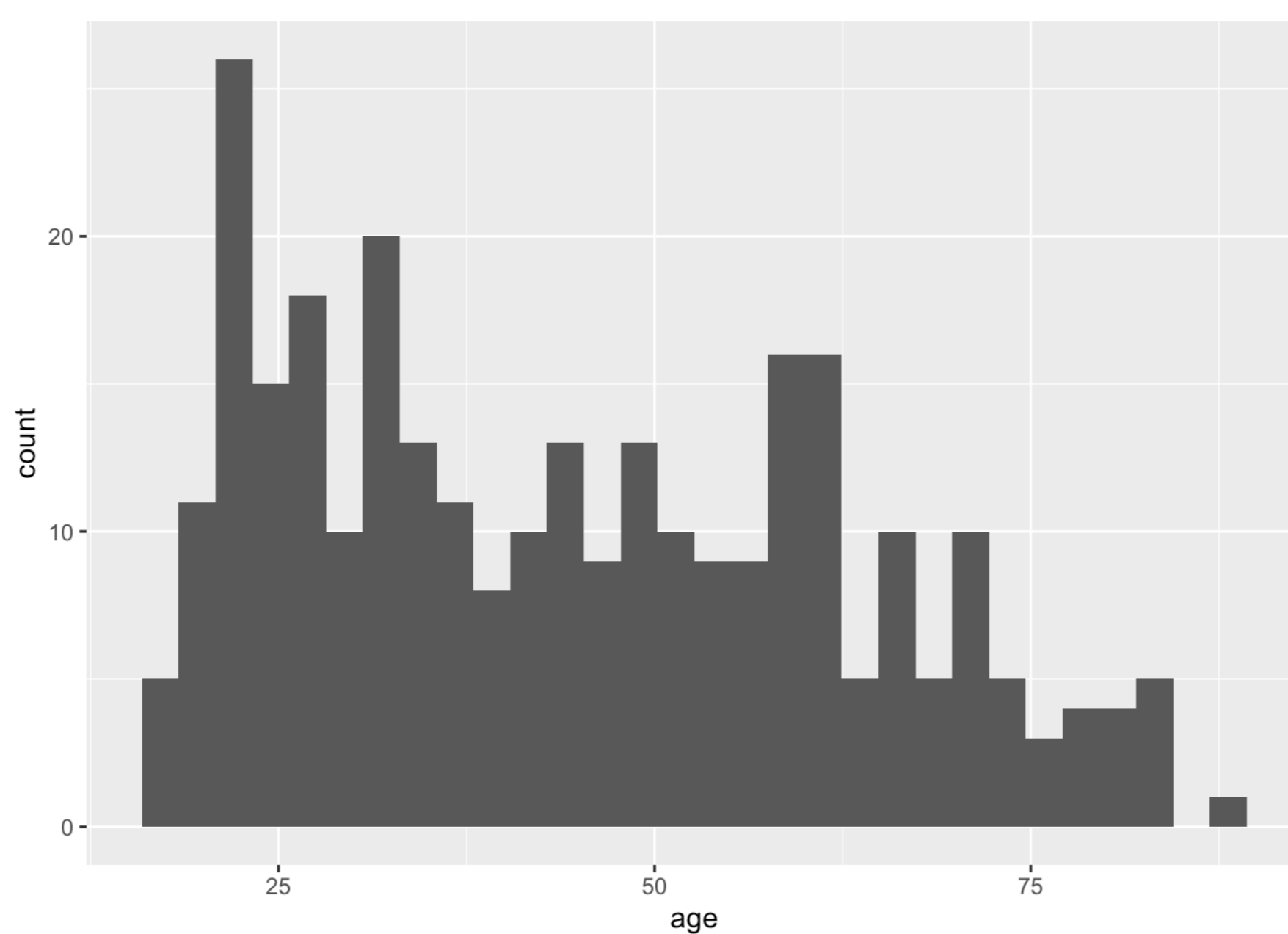
The data set I chose is the one about depression, which has 294 observations and 37 variables. The variables I am going to explore are age, income, employment. I want to see if age, income, or marital status have any correlation with people that have depression.

## Univariate Analysis

### 1. Age

```
ggplot(depress, aes(x=age,)) + geom_histogram()
```

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



This graph shows out of the

participants in the study, a lot of them are around the age of 18, 33, and 60 years old.

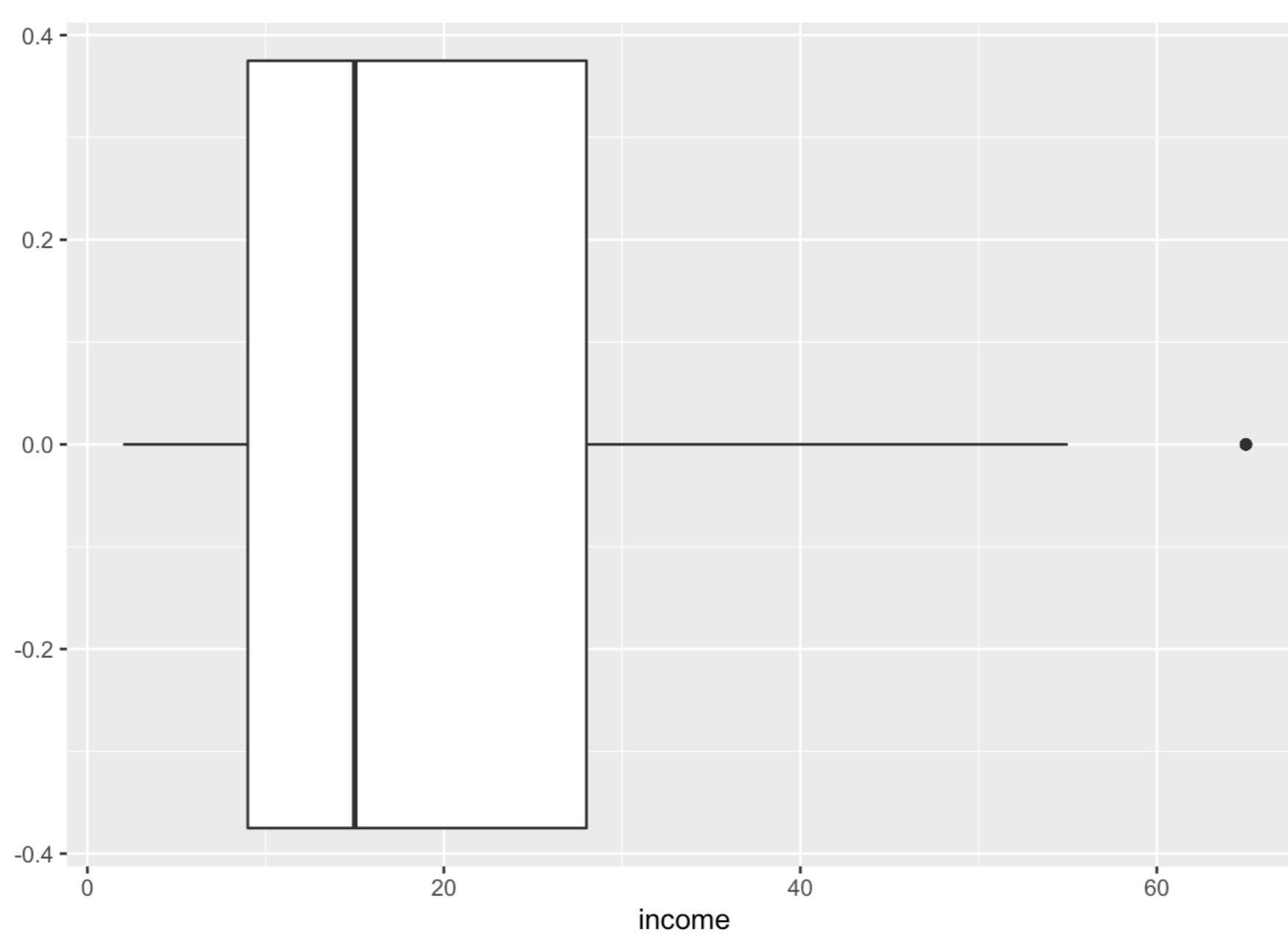
```
summary(depress$age,digits = 2)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   ##      18      28       42     44     59     89
```

This shows that the youngest age of the the participants with depression was 18, with the average at 44 and the oldest person with depression at 89 years old

### 2. Income

```
ggplot(depress, aes(income)) +geom_boxplot()
```



The boxplot shows the majority of

people in this study make \$15,000-\$28,000 a year with an outlier of \$65,000.

```
summary(depress$income, digits = 3)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   ##      2.0    9.0    15.0    20.6   28.0    65.0
```

This shows that the mean income is \$15,000 a year with the highest annual income at \$65,000.

### 3. Marital Status

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

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

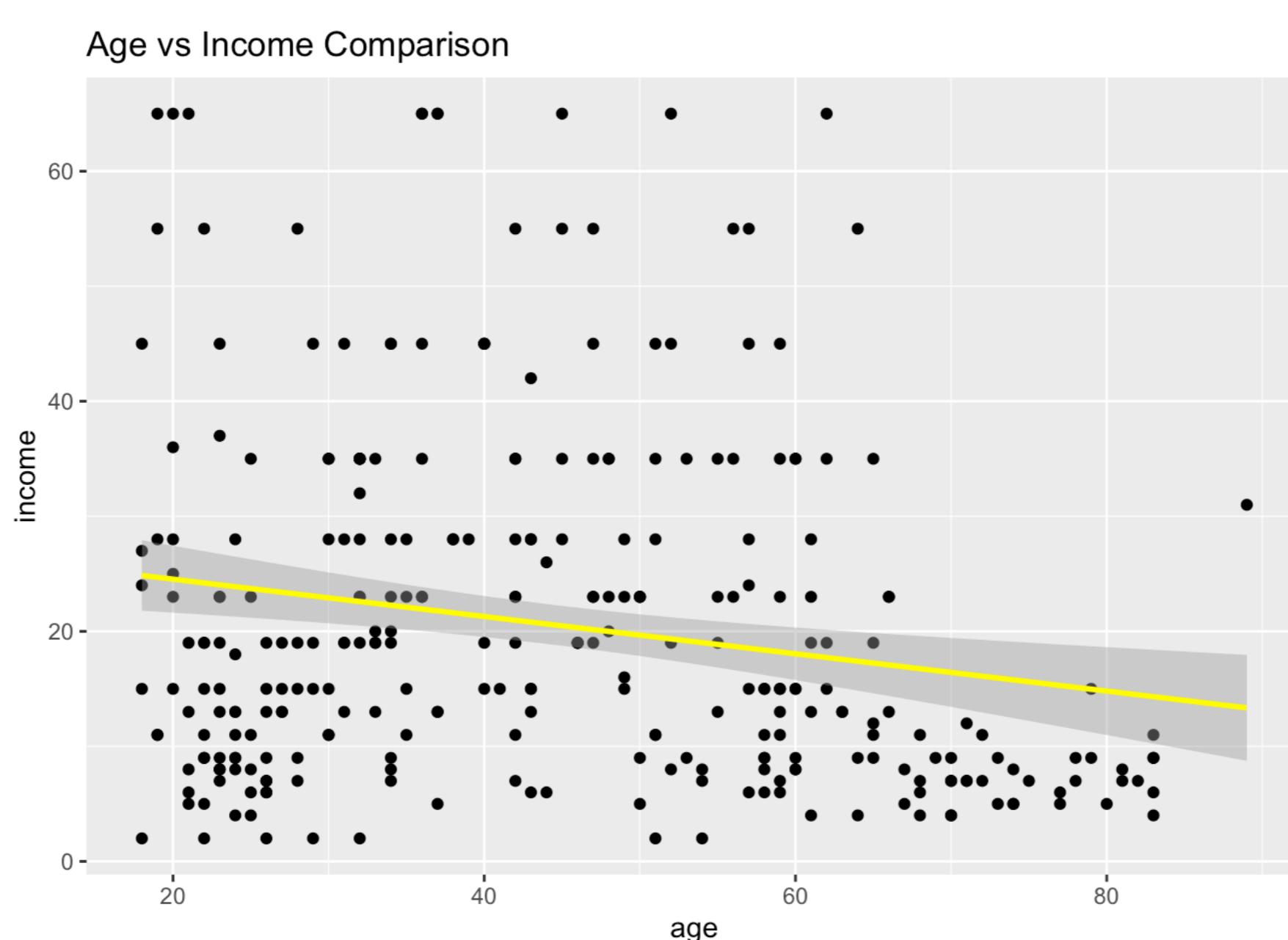
This table shows that the majority of the participants are either married or have never been married.

## Bivariate Analysis

### 1. Age vs Income

```
ggplot(depress, aes(x=age, y=income)) + geom_point() + geom_smooth(method="lm", color="yellow") + ggtitle("Age vs Income Comparison")
```

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



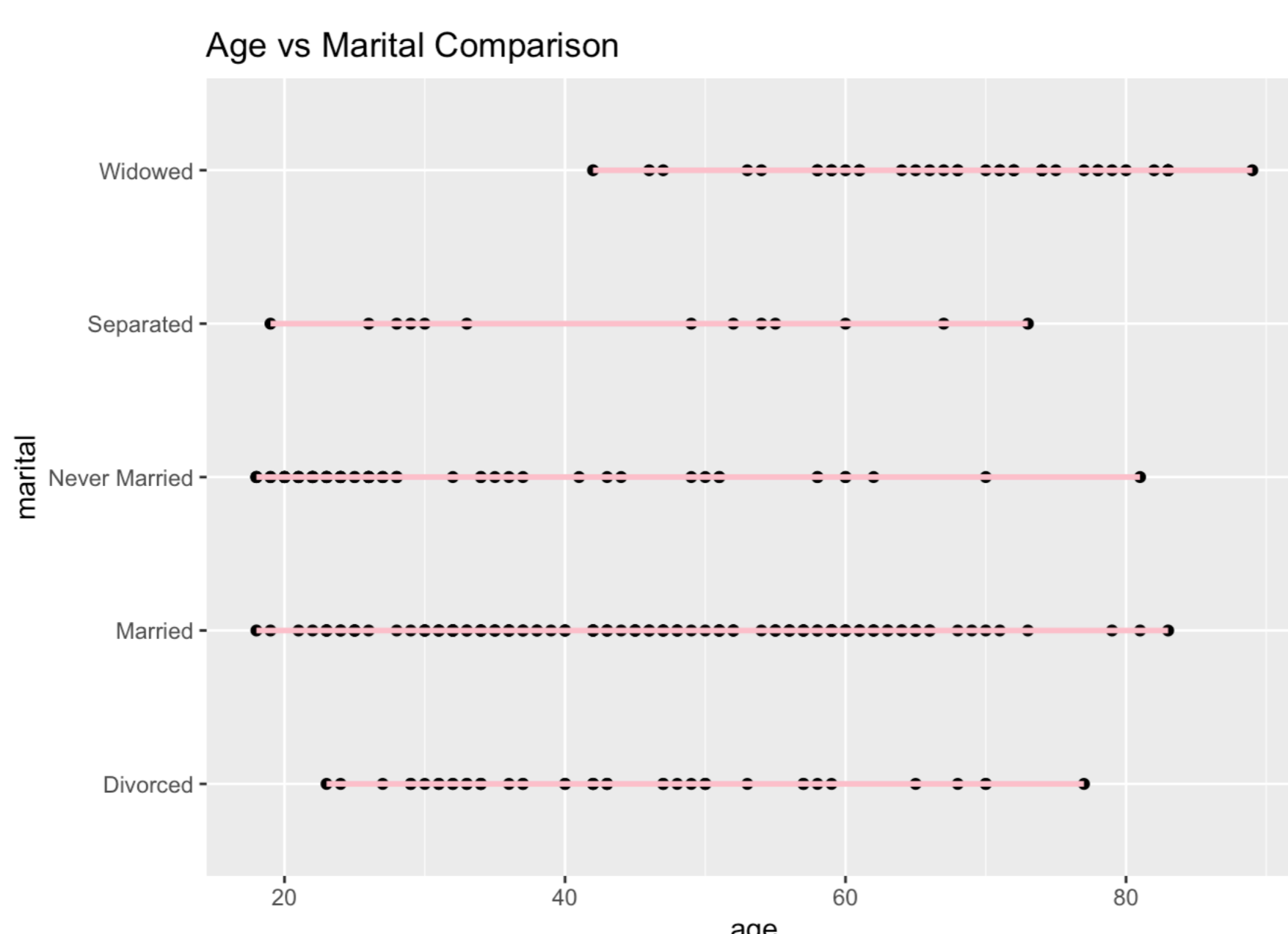
This depicts that there is not really

any relation to age and income in the depression data.

### 2. Age vs Marital

```
ggplot(depress, aes(x=age, y=marital)) + geom_point() + geom_smooth(method="lm", color="pink") + ggtitle("Age vs Marital Comparison")
```

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



This graph is showing us that in

every category the age range is pretty wide and spread out. Meaning that there isn't a correlation between age and marital status.

## Conclusion

I realized that all three of my chosen variables really don't have anything to do with each other. There is no correlation between age, income and marital status of the depression participants. I did believe that there was going to be some sort of correlation at least with income because I would think that making less money could make you depressed.