

EDA_slanum

Sarah Lanum

2024-09-25

Introduction

- I will be exploring the ‘depression’ data set, which studies 37 variables. The variables under observation will be age and yearly income.
 - Of the 194 individuals in the study, does it appear that there is a correlation between age and income? Is there any noticeable pattern of depression among the top earners compared to the low earners?

```
depression <- read.table("/Users/sarah/Desktop/math130/data/Depress.txt", header=TRUE, sep="\t")
```

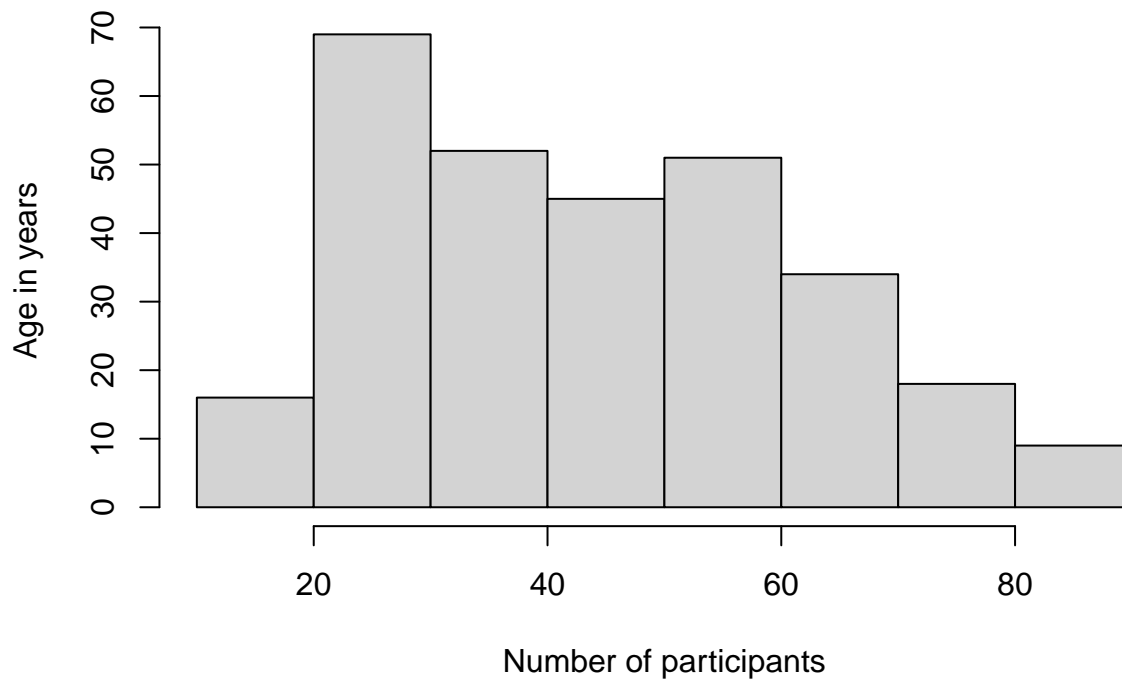
Univariate Exploration:

```
summary(depression$AGE)
```

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

```
hist(depression$AGE, main = "Frequency of depression for age range", xlab = "Number of participants", ylab = "Frequency")
```

Frequency of depression for age range



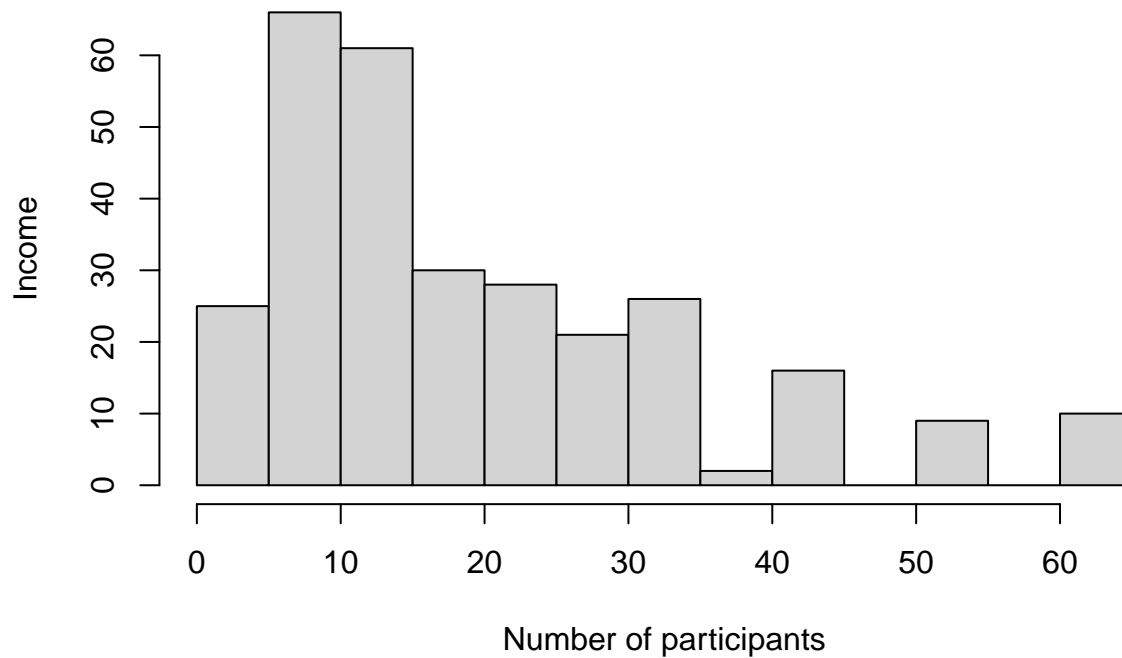
- Ages in the data set range from 18 years old to 89 years old with an average of 44.41 years.

```
summary(depression$INCOME)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.00   9.00   15.00   20.57  28.00   65.00
```

```
hist(depression$INCOME, main = "Frequency of depression across income range", xlab = "Number of participants")
```

Frequency of depression across income range

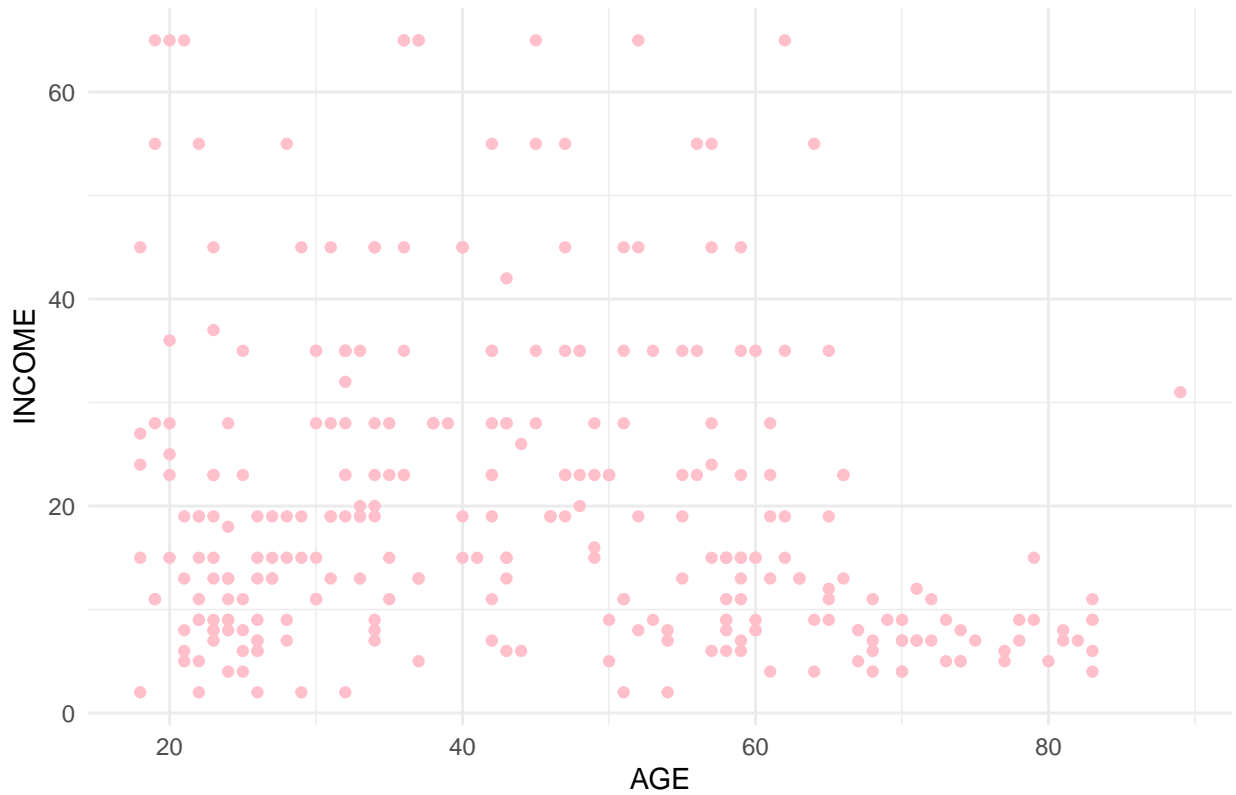


- Rates of depression are higher for individuals with low income. Income ranges from \$2,000-\$65,000 with an average of \$20,570.

Bivariate Exploration:

```
library(ggplot2)
ggplot(depression, aes(x = AGE, y = INCOME)) + geom_point(color = "pink") + labs(title = "Scatter Plot of
```

Scatter Plot of Age versus Income



- The results of comparing age to income within the depression data set revealed a correlation between low income and depression, while age appears to have less impact on depression rates.

Conclusion:

- The question I was exploring had to do with patterns found in the data of comparing age and the income of depressed individuals. As mentioned above, I did not notice a correlation between income. The scatter plot above shows that a majority of the participants have low income.