Exploratory Data Analysis Project

Czarina Monroy

2024-03-08

Introduction

The following exploratory data analysis project uses the depression dataset. This dataset was collected from the first set of interviews of a prospective study of depression in the adult residents of Los Angeles County. It includes 294 observations of 37 variables. This analysis explores the relationship between income and age by answering the research question: How does income vary based on age among adult residents of Los Angeles County in the depression dataset?
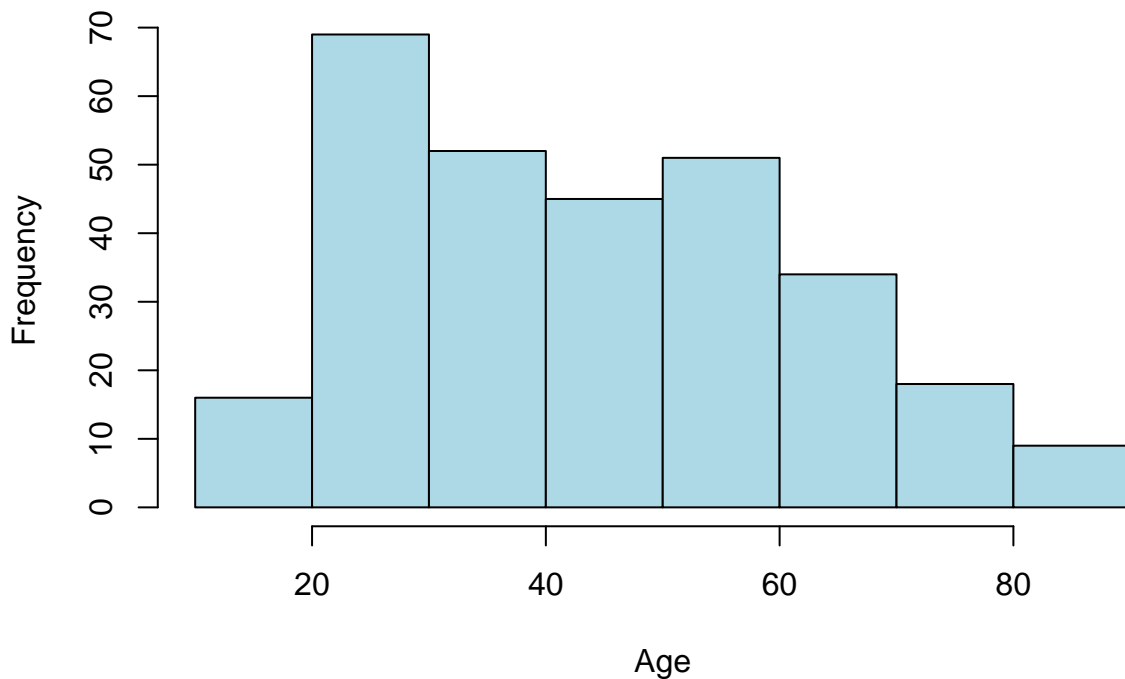
Univariate Exploration

Age

```r
depression <- read.delim("/Users/czarina/Downloads/Depress.txt")
summary(depression$AGE)
```

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

```r
hist(depression$AGE, main = "Distribution of Age", xlab = "Age", col = "lightblue", border = "black")
```
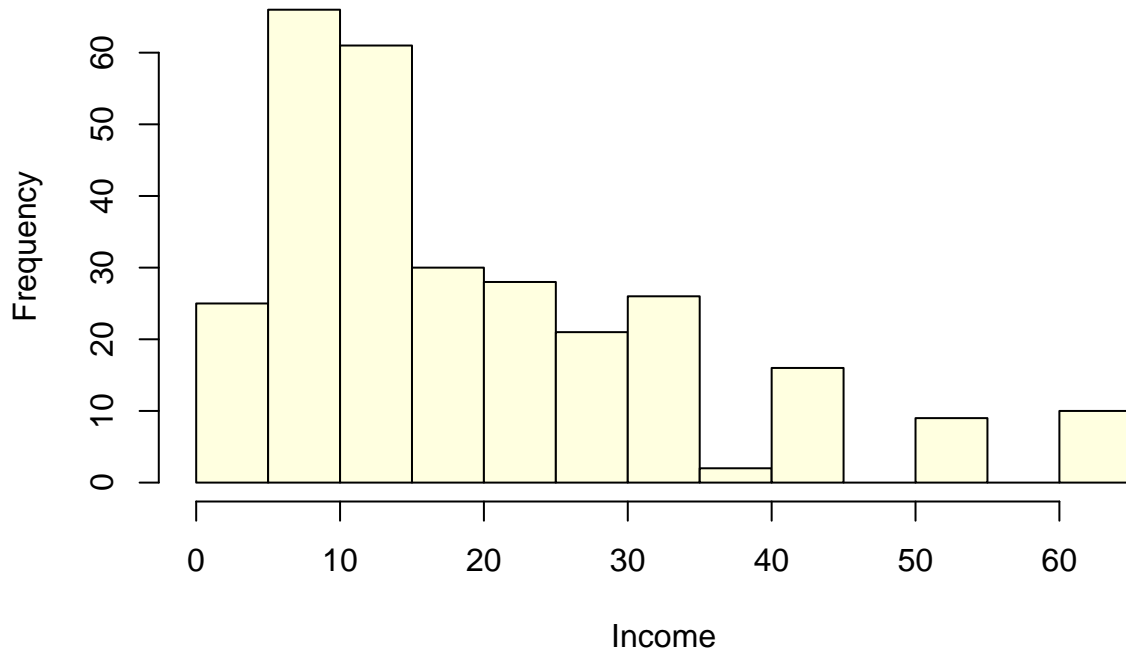

Distribution of Age

Income

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

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

```r
hist(depression$INCOME, main = "Distribution of Income", xlab = "Income", col = "lightyellow", border =
```
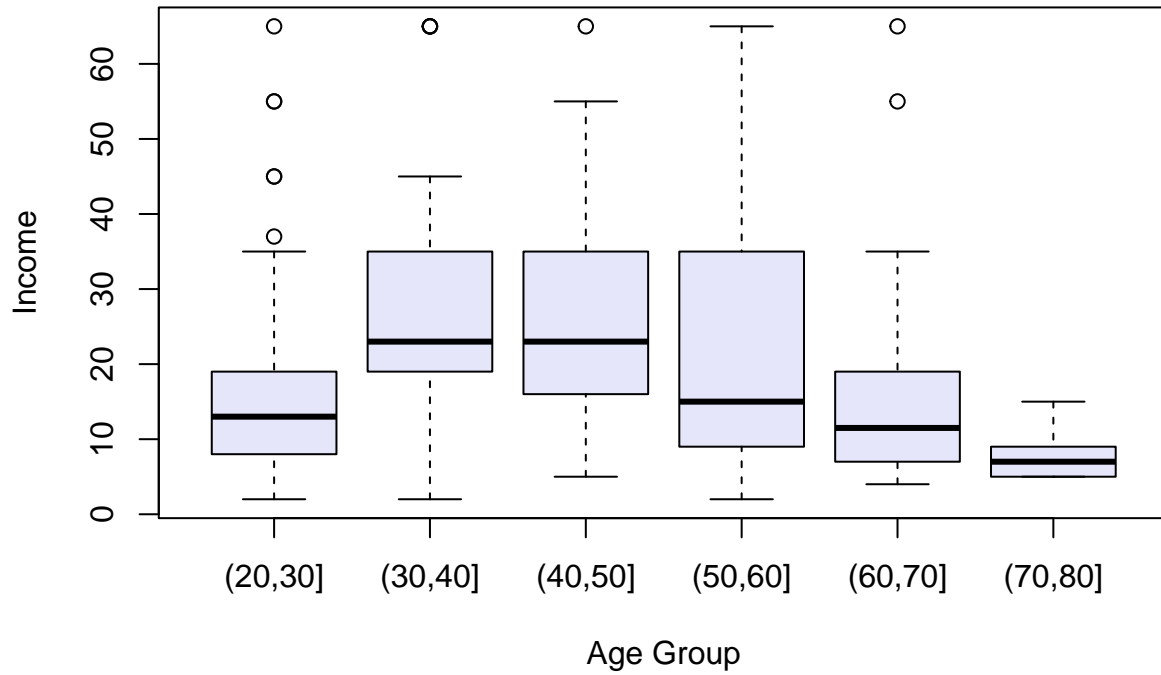
Distribution of Income

Bivariate Exploration

An investigation of the relationship between income and age

Income vs. Age

```r
boxplot(depression$INCOME ~ cut(depression$AGE, breaks = c(20, 30, 40,
50, 60, 70, 80)), main = "Income by Age Group", xlab = "Age Group",
ylab = "Income", col = "lavender")
```

Income by Age Group

Conclusion

The exploratory analysis of the depression dataset reveals a pattern in the relationship between income and age among adult residents of Los Angeles County. The data suggests that younger and older age groups tend to have lower incomes, in comparison to the middle age groups. Further studied would be necessary to analyze contributing factors to this income distribution pattern.