

# Depression Data: Final Project

analise hernandez

09/14/22

## Introduction

- For this project I will be exploring the “Depression” data set. This data set contains 294 observations of 37 variables. I will be looking into the age and employment status variables. My goal is to explore the ranges of ages present and if there is a correlation between age and employment status.

```
dt <- read.delim("/Users/analisehernandez/Desktop/math130/data/depression data set.txt", header=TRUE, s
dim(dt)
```

```
## [1] 294 37
```

```
library(ggplot2)
library(dplyr)
```

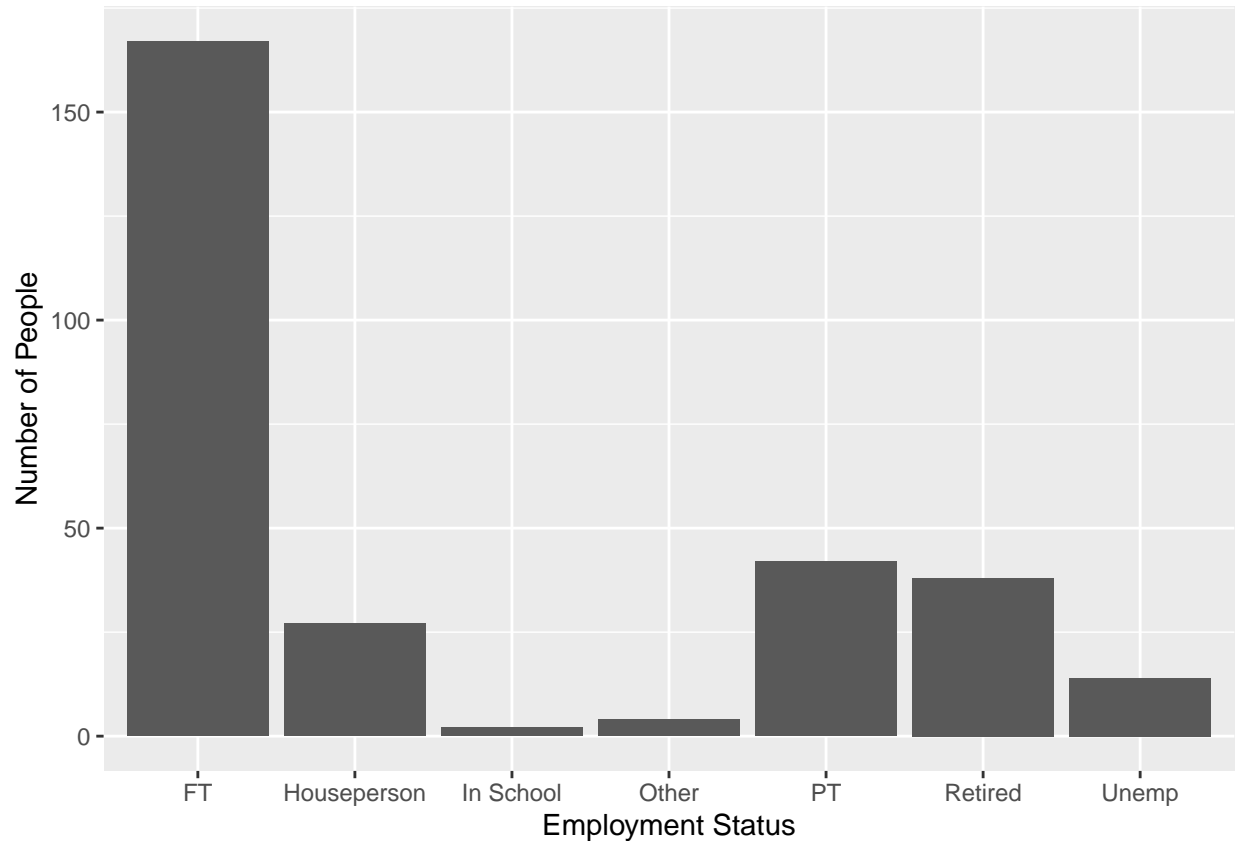
```
##
## Attaching package: 'dplyr'
```

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

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

## Univariate Exploration

```
ggplot(dt, aes(x=employ)) + geom_bar() + ylab("Number of People")+ xlab("Employment Status")
```



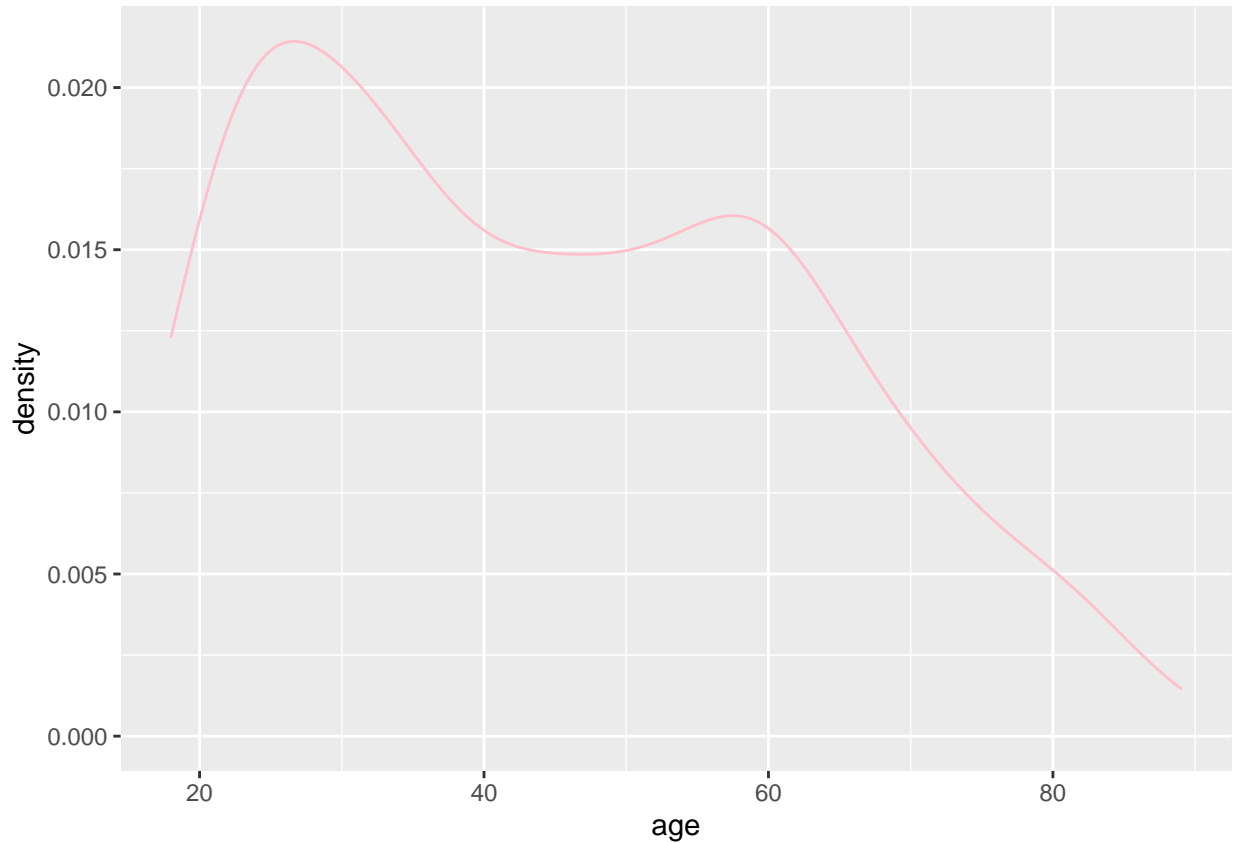
-This graphic shows us that the majority of people in this data set are full time employees, with the bar reaching over 150 people, and the next most common status is part time employees, with only about 45 people. The least common status is “in school”.

```
table(dt$employ)
```

```
##
##      FT Houseperson  In School    Other    PT    Retired
##      167         27         2         4    42         38
##      Unemp
##      14
```

- Here, we are still exploring the employment status variable, but in a table with numerical data. We can see the exact count of each status, with full time employment containing 167 people, part time employment with 42 people, and “in school” with the least amount of people, only 2.

```
ggplot(dt, aes(x=age)) + geom_density(col="pink")
```



- Now, we are exploring the age variable. This density plot shows that the majority of people who participated in this study were in their mid-20s, and people 80 and over were the minority in this study.

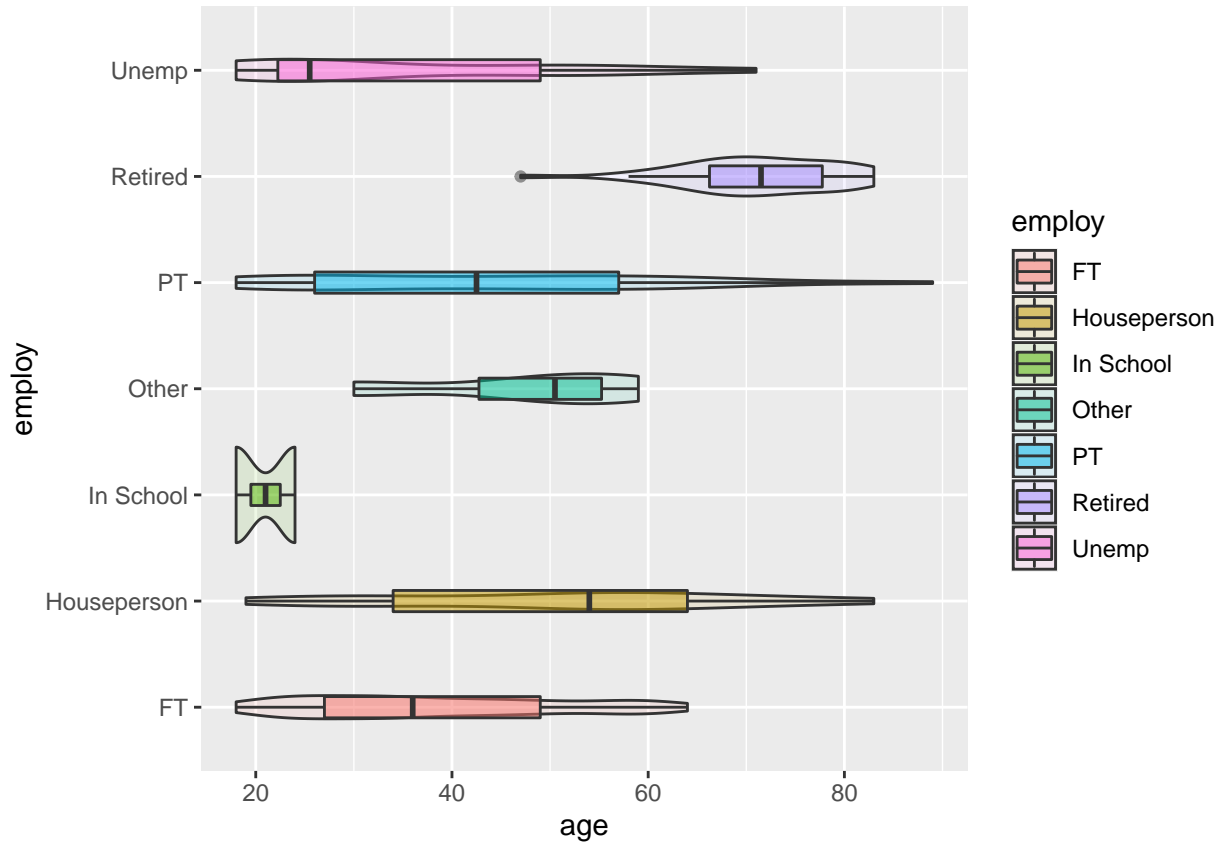
```
summary(dt$age, digits = 2)
```

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

- We are still exploring the age variable, and these summary statistics tell us that 25% of participants are 28 years of age and younger, and 75% of participants are 59 years of age and younger. Also, this table tells us that the maximum age in this data set is 89 years old.

## Bivariate Exploration

```
ggplot(dt, aes(x=age, y=employ, fill=employ)) + geom_violin(alpha=.1) +
geom_boxplot(alpha=.5, width=.2)
```



- Here we are exploring the age and employment status variables together. We can see there is a wide range of ages with the houseperson, part time, and unemployed status' compared to the full-time, in school, and retired status'.

```
table(dt$employ,dt$age)
```

```
##
##      18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
## FT      1  2  4  5  6  7  6  4  6  4  3  3  4  5  7  4  8  2  6  4  1  1
## Houseperson 0  1  0  0  0  2  0  0  0  0  1  0  1  0  1  1  0  1  0  0  0  0
## In School  1  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## Other      0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
## PT         2  2  1  1  1  1  1  1  2  0  1  1  0  0  1  0  1  1  0  1  1  0
## Retired    0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## Unemp      1  0  1  0  2  1  1  1  1  0  0  0  0  0  1  0  0  0  0  0  0  0
##
##      40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61
## FT      5  1  7  5  2  3  2  3  3  4  3  2  2  2  1  1  2  2  5  5  5  4
## Houseperson 0  0  0  1  0  0  0  1  0  0  0  1  2  0  1  1  0  1  1  1  1  0
## In School  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## Other      0  0  0  0  0  0  0  1  0  0  0  0  0  0  1  0  0  0  0  1  0  0
## PT         0  0  2  1  0  1  0  0  1  0  1  3  0  0  0  0  1  3  0  1  1  0
## Retired    0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  1  1  0  1
## Unemp      0  0  0  0  0  0  1  0  0  0  1  0  0  0  0  2  0  0  0  0  0  0
##
##      62 63 64 65 66 67 68 69 70 71 72 73 74 75 77 78 79 80 81 82 83 89
##
```

##	FT	4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
##	Houseperson	0	1	0	1	0	1	0	0	1	1	0	0	0	1	0	0	0	0	0	2	0	
##	In School	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
##	Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
##	PT	0	0	1	2	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	
##	Retired	0	1	1	2	2	1	3	1	3	1	2	2	3	0	2	2	2	1	2	1	2	0
##	Unemp	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	

This table gives us numeric data. It shows us the correlation between age and employment status. We can see the exact number of people in each of the 7 different employment status' according to their age.

## Conclusion

In conclusion, the majority of people who participated in this data set have a full-time job. Also, the majority of people who participated are in their mid-20s, but still 75% of participants are 59 or younger. We also saw wide age ranges of housepersons, part-time, and unemployed participants. Housepersons mostly ranged from ages 35 to 65, part-time employees mostly ranged from ages 25 to 58, and unemployed participants mostly ranged from ages 20 to 50. The status' of participants "in school", full-time employees, and "other" had less of an age range than its counterparts.