

EDA_enregan

Emily Regan

2022-09-19

Impacts of Parental HIV

Introduction

Parental HIV affects not only the parent but also children if they are present. Parental HIV can influence many different aspects of life like neighborhoods, early age smoking, income, how religious a child is, parental bonding and the list goes on. In this research project, I am interested in whether children whose parents are HIV positive are more religious than those who are not along with if it influences early smoking in children whose parents are HIV positive.

Research Question: Does Parental HIV influence how religious the child is along with the age the child started smoking?

```
parHIV <- read.delim("/Users/emilyregan/Desktop/math130/data/parentalHIV.txt", header=TRUE, stringsAsFactors=FALSE)
dim(parHIV)
```

```
## [1] 252 123
```

Inserting the code that is presented above helps give us a inside look on how many observations and how many variables there are in this specific data set. From above we can conclude that there are 252 observations and 123 variables.

```
table(parHIV$HOWREL)
```

We will now look into the two variables if the child's parents are HIV positive, and if there is a correlation that they are more religious due to it.

```
##
##  1  2  3
## 28 132 49
```

We can observe from the table above that the amount of observations between these two variables do not add to the total amount of observations we saw earlier in the overall data set.

```
table(parHIV$HOWREL, useNA="always")
```

```
##  
##    1    2    3 <NA>  
##  28 132  49  43
```

```
sum(is.na(parHIV$HOWREL))
```

```
## [1] 43
```

With the two codes that are above demonstrate that we have 43 observations that are unknown. This helps us since we know not every individual in this data set answered this specific question of how religious they are.

Now we will examine if there a correlation between those whose parents are HIV positive and how religious they are with a graph.

First we are going to rename the variables to what they correlate to.

Originally here are the variable numbers in relation to what it means: 1 = "Very religious or spiritual" 2= "Somewhat religious or spiritual" 3 = "Not at all religious or spiritual"

Now we will rename the variables "1", "2", and "3" to what it correlates to:

```
parHIV$HOWREL_new<- factor(parHIV$HOWREL, labels=c( "Very religious or spiritual", "Somewhat religious or spiritual", "Not at all religious or spiritual"))  
table(parHIV$HOWREL,parHIV$HOWREL_new)
```

```
##  
##    Very religious or spiritual Somewhat religious or spiritual  
##    1                28                0  
##    2                0                132  
##    3                0                0  
##  
##    Not at all religious or spiritual  
##    1                0  
##    2                0  
##    3                49
```

Now we will examine the correlation between parental HIV and how religious the child who is affected is:

```
barplot(table(parHIV$HOWREL_new), cex.names = 0.7, main="Religion & Parental HIV",  
        xlab="How Religious Are You")
```

Religion & Parental HIV



How Religious Are You

Above we can see the 3 main variables. 1 = “Very religious or spiritual” 2= “Somewhat religious or spiritual” 3 = “Not at all religious or spiritual”

We can determine that the majority of subjects chose that they were “Somewhat religious or spiritual” and whose parents were HIV positive.

Now were going to look at the second variable which is if having parental HIV influences early smoking in children. First its good to note how many observations were not included.

```
sum(is.na(parHIV$AGESMOKE))
```

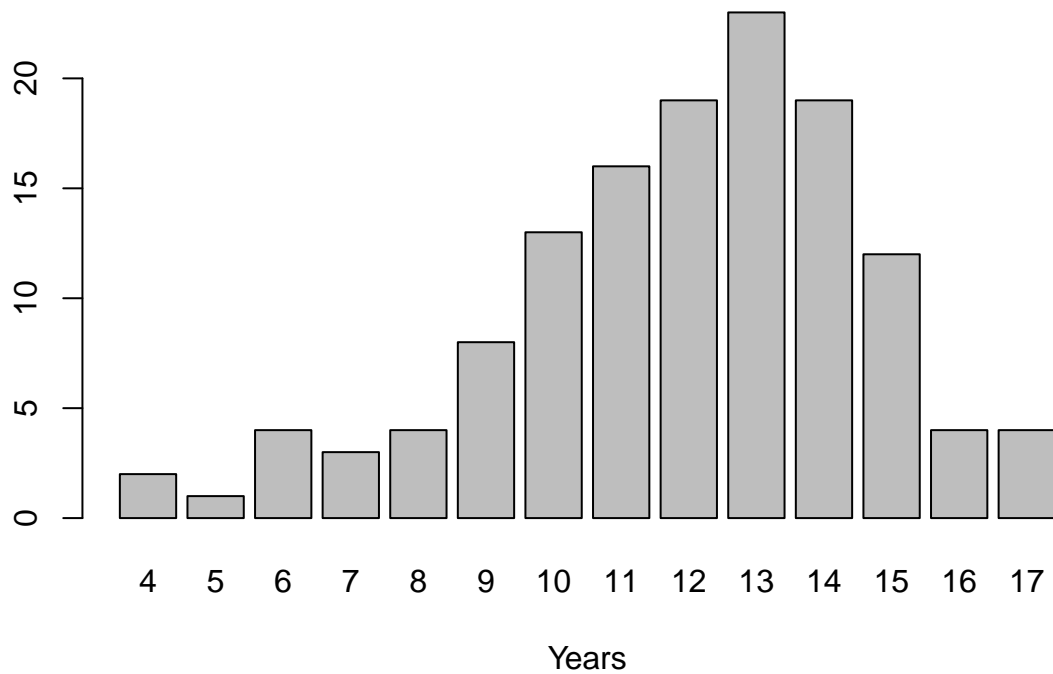
```
## [1] 120
```

Just like with the variable “How Religious are you”, we can see that there are some subjects who did not answer the question which makes the total observations not equal to 252.

Now we are going to look at the frequency of the age smoking started and see if there a correlation with the age started and if it was influenced by parental HIV.

```
barplot(table(parHIV$AGESMOKE), width=1, main="Distribution of age when smoking started", xlab="Years")
```

Distribution of age when smoking started



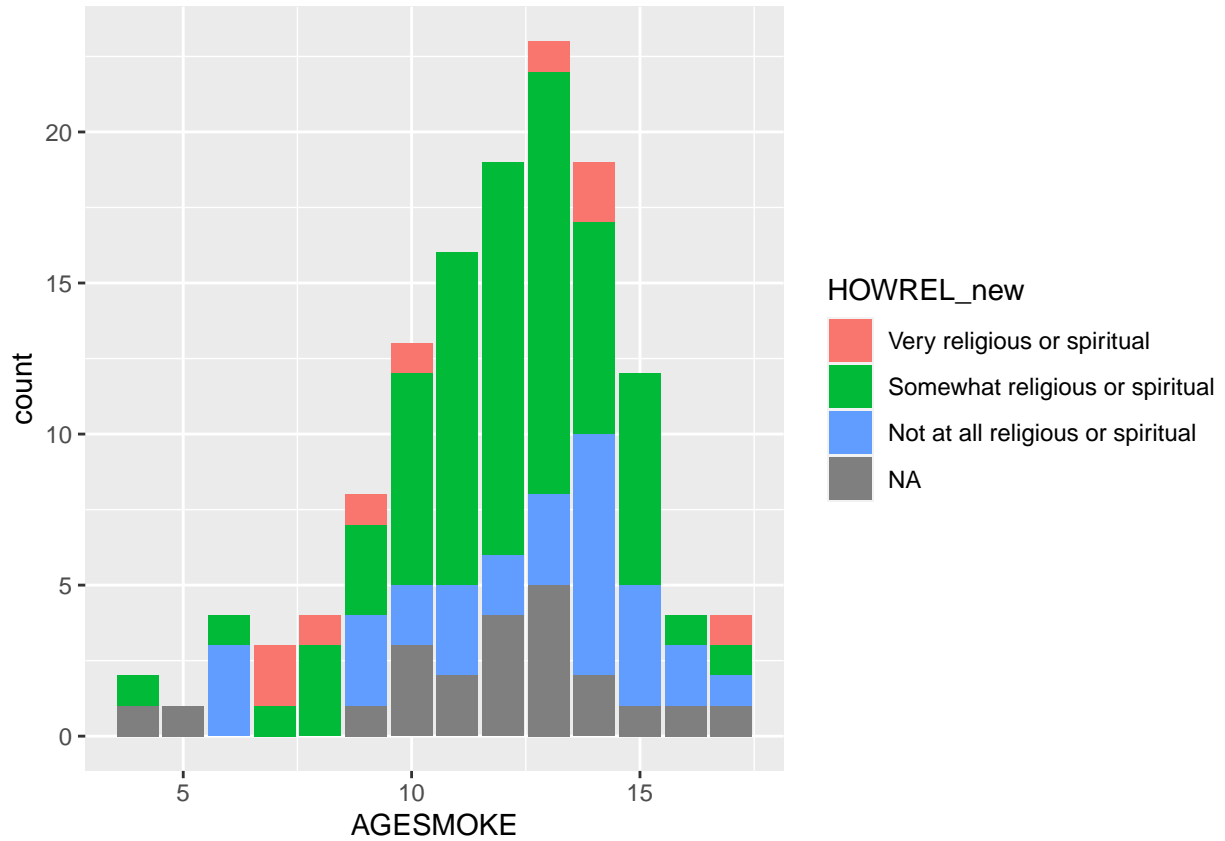
Given the Histogram above we can distinguish that the age smoking started was age 9-15 years old when parental HIV was true.

Bivariate Exploration

We are going to take a look at both variables that were examined above which were How religious are you and the age that smoking began both when parental HIV is true.

```
ggplot(parHIV, aes(x=AGESMOKE, fill=HOWREL_new)) + geom_bar()
```

```
## Warning: Removed 120 rows containing non-finite values (stat_count).
```



Conclusion

Given parental HIV being true we can see that there is a relationship between how religious an individual is along with the age one started smoking. There's a relationship but doesn't necessarily conclude that these two variables were related but both were influenced by parental HIV since we can see in the graph above that it was relatively consistent. The age of smoking started early which was around 9-15 years of age and how religious an individual is was 'somewhat religious'. There seems to be a relatively uniformed distribution that has a slight negative skew.