final project

2024-09-26

1. Introduction

I am using the Parental HIV data with the use of a bar chart, it contains 251 observations and 111 variables.I will be using the variables of 'AGEALC' for start of drinking and 'AGESMOKE' for start of smoking.My question is do the affects of having an HIV positive parent impact the age kids begin drinking and smoking.

```
"/Users/olivialane/Desktop/math130/data/Parhiv.txt"
```

```
## [1] "/Users/olivialane/Desktop/math130/data/Parhiv.txt"
```

```
library(ggplot2)
hiv <- read.table("/Users/olivialane/Desktop/math130/data/Parhiv.txt", header=TRUE, sep="\t")
str(hiv)
```

```
## 'data.frame':    251 obs. of  111 variables:
##  $ ID      : int  2 3 4 5 6 7 8 9 10 11 ...
##  $ AGE     : int  18 13 14 14 13 14 14 13 13 17 ...
##  $ GENDER  : int  2 1 2 1 1 2 2 2 1 1 ...
##  $ LIVWITH : int  1 2 2 2 2 2 2 2 1 3 ...
##  $ SIBLINGS: int  NA 2 2 2 2 2 2 2 2 2 ...
##  $ JOBMO   : int  1 2 2 2 2 2 2 2 NA 2 ...
##  $ EDUMO   : int  2 1 3 3 NA 2 2 3 NA 1 ...
##  $ HOWREL  : int  2 NA 3 2 2 2 1 NA NA 2 ...
##  $ ATTSERV : int  2 NA 1 1 2 2 3 NA NA 2 ...
##  $ NGHB1   : int  4 1 3 3 2 2 3 4 1 2 ...
##  $ NGHB2   : int  2 1 3 4 4 3 3 4 1 2 ...
##  $ NGHB3   : int  3 1 2 2 2 2 2 1 1 2 ...
##  $ NGHB4   : int  2 2 4 2 2 4 3 1 1 2 ...
##  $ NGHB5   : int  1 4 2 1 2 1 3 2 1 2 ...
##  $ NGHB6   : int  1 4 2 2 4 2 4 1 1 1 ...
##  $ NGHB7   : int  2 4 2 2 3 1 3 4 1 1 ...
##  $ NGHB8   : int  1 1 4 3 4 3 4 4 1 2 ...
##  $ NGHB9   : int  2 1 1 3 3 1 3 4 1 1 ...
##  $ NGHB10  : int  3 1 4 2 1 2 3 4 1 1 ...
##  $ NGHB11  : int  1 1 2 2 3 3 3 4 1 1 ...
##  $ MONFOOD : int  3 3 1 3 3 3 3 3 2 3 ...
##  $ FINSIT  : int  3 1 3 4 3 3 1 4 2 3 ...
##  $ ETHN    : int  3 1 2 2 2 2 1 2 1 2 ...
##  $ AGESMOKE: int  14 12 14 0 12 9 0 0 0 10 ...
##  $ SMOKEP3M: int  4 1 8 NA 1 8 NA NA NA 2 ...
##  $ AGEALC  : int  15 0 14 0 0 0 0 0 0 15 ...
##  $ AGEMAR  : int  17 0 14 0 0 13 0 0 0 16 ...
##  $ FRNDS   : int  3 4 1 2 2 2 2 3 1 2 ...
```

```
##  $ SCHOOL  : int  2 2 2 2 2 2 2 2 2 1 ...
##  $ LIKESCH : int  3 2 1 1 3 3 1 1 3 1 ...
##  $ HOOKEY  : int  2 1 2 2 1 1 1 2 1 2 ...
##  $ NHOOKEY : int  4 0 1 4 0 0 0 2 0 8 ...
##  $ HMONTH  : int  3 3 3 1 1 3 1 1 1 3 ...
##  $ PB01    : int  3 4 1 4 3 3 4 4 3 3 ...
##  $ PB02    : int  2 1 1 1 4 1 2 2 2 3 ...
##  $ PB03    : int  1 4 1 4 1 3 1 3 2 3 ...
##  $ PB04    : int  3 1 3 1 3 1 4 2 3 3 ...
##  $ PB05    : int  3 4 1 3 1 3 4 3 3 1 ...
##  $ PB06    : int  4 4 1 4 4 4 3 3 3 2 ...
##  $ PB07    : int  1 3 2 3 1 4 2 3 3 3 ...
##  $ PB08    : int  4 2 3 1 1 3 2 2 3 1 ...
##  $ PB09    : int  4 3 4 1 3 3 4 2 1 1 ...
##  $ PB10    : int  3 2 4 3 4 2 2 3 2 1 ...
##  $ PB11    : int  2 4 1 3 4 4 3 3 3 1 ...
##  $ PB12    : int  2 4 1 3 4 4 4 3 4 1 ...
##  $ PB13    : int  1 4 1 3 4 3 2 3 3 1 ...
##  $ PB14    : int  4 2 4 2 4 2 1 2 2 4 ...
##  $ PB15    : int  1 4 2 4 4 3 2 3 3 1 ...
##  $ PB16    : int  2 2 1 1 1 1 2 2 1 3 ...
##  $ PB17    : int  3 4 1 4 4 3 4 3 3 1 ...
##  $ PB18    : int  3 2 1 1 1 1 1 2 2 1 ...
##  $ PB19    : int  3 3 3 1 3 2 2 2 2 1 ...
##  $ PB20    : int  4 2 1 1 3 2 1 2 3 1 ...
##  $ PB21    : int  1 4 2 4 2 2 2 3 3 4 ...
##  $ PB22    : int  1 2 3 3 1 3 2 3 2 4 ...
##  $ PB23    : int  4 4 2 1 4 3 4 3 3 1 ...
##  $ PB24    : int  3 1 1 4 1 3 1 2 3 4 ...
##  $ PB25    : int  2 4 3 3 4 4 2 2 3 4 ...
##  $ BSI01   : int  1 0 0 0 0 0 0 0 0 1 ...
##  $ BSI02   : int  0 0 0 1 0 0 4 0 0 1 ...
##  $ BSI03   : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ BSI04   : int  0 0 0 0 1 1 0 0 0 1 ...
##  $ BSI05   : int  0 0 0 0 2 1 0 0 0 0 ...
##  $ BSI06   : int  0 0 1 1 1 1 1 0 0 1 ...
##  $ BSI07   : int  0 0 1 0 0 2 0 0 0 1 ...
##  $ BSI08   : int  0 0 0 0 1 1 0 0 0 0 ...
##  $ BSI09   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BSI10   : int  0 0 0 0 2 1 3 3 0 3 ...
##  $ BSI11   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BSI12   : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ BSI13   : int  0 0 1 0 2 0 0 0 0 0 ...
##  $ BSI14   : int  3 0 0 0 0 0 0 0 0 2 ...
##  $ BSI15   : int  0 0 0 0 1 0 0 0 0 1 ...
##  $ BSI16   : int  3 0 0 0 1 0 0 0 0 1 ...
##  $ BSI17   : int  2 0 0 1 0 1 0 0 0 1 ...
##  $ BSI18   : int  0 0 0 0 3 0 1 0 0 1 ...
##  $ BSI19   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BSI20   : int  2 0 0 0 0 0 3 0 0 1 ...
##  $ BSI21   : int  0 0 0 0 0 0 0 1 0 1 ...
##  $ BSI22   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BSI23   : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ BSI24   : int  0 0 0 1 1 1 1 0 0 1 ...
```

```
##  $ BSI25   : int  4 0 0 0 1 0 0 0 0 4 ...
##  $ BSI26   : int  0 0 0 1 1 1 3 0 0 0 ...
##  $ BSI27   : int  0 0 0 0 1 0 1 0 0 2 ...
##  $ BSI28   : int  0 0 1 0 1 0 0 0 0 0 ...
##  $ BSI29   : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ BSI30   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ BSI31   : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ BSI32   : int  0 0 0 0 0 0 0 0 0 2 ...
##  $ BSI33   : int  0 0 0 1 1 0 0 0 0 0 ...
##  $ BSI34   : int  0 0 1 0 2 0 0 0 0 0 ...
##  $ BSI35   : int  0 0 0 0 2 0 1 0 0 1 ...
##  $ BSI36   : int  0 0 0 0 1 0 3 0 0 0 ...
##  $ BSI37   : int  0 0 1 1 1 0 0 0 0 0 ...
##  $ BSI38   : int  2 0 0 0 2 0 0 0 0 1 ...
##  $ BSI39   : int  0 0 0 0 1 0 0 1 0 1 ...
##  $ BSI40   : int  0 0 0 1 1 0 0 0 0 2 ...
##  $ BSI41   : int  0 0 1 0 2 0 0 0 0 0 ...
##   [list output truncated]
```

2. Univariate

I will first be looking if an HIV positive parent affects the age of smoking

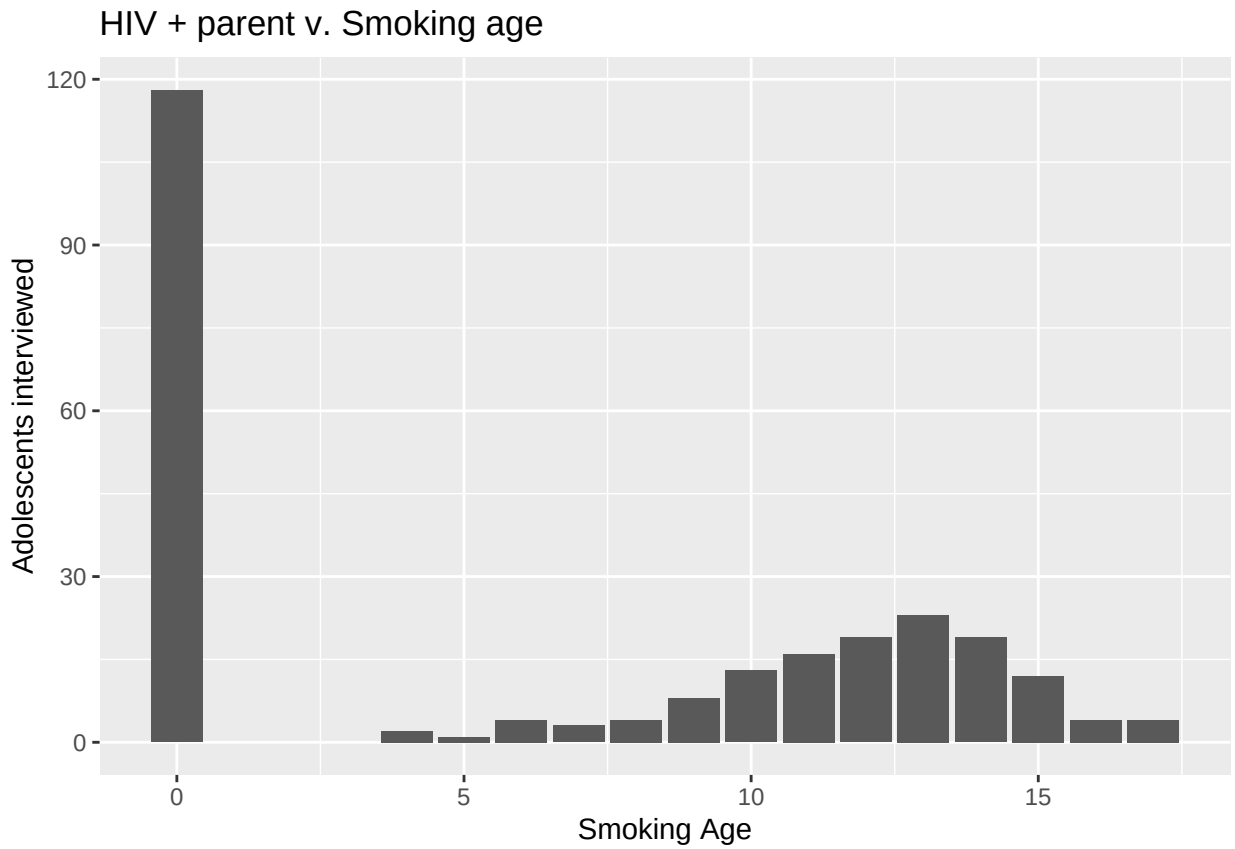
this is showing how many adolescents are in each age category

```
table(hiv$AGESMOKE)
```

```
##
##   0   4   5   6   7   8   9  10  11  12  13  14  15  16  17
## 118   2   1   4   3   4   8  13  16  19  23  19  12   4   4
```

```
ggplot(hiv, aes(x=AGESMOKE))+geom_bar()+xlab("Smoking Age")+ylab("Adolescents interviewed")+ggtitle("HI
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_count()`).
```



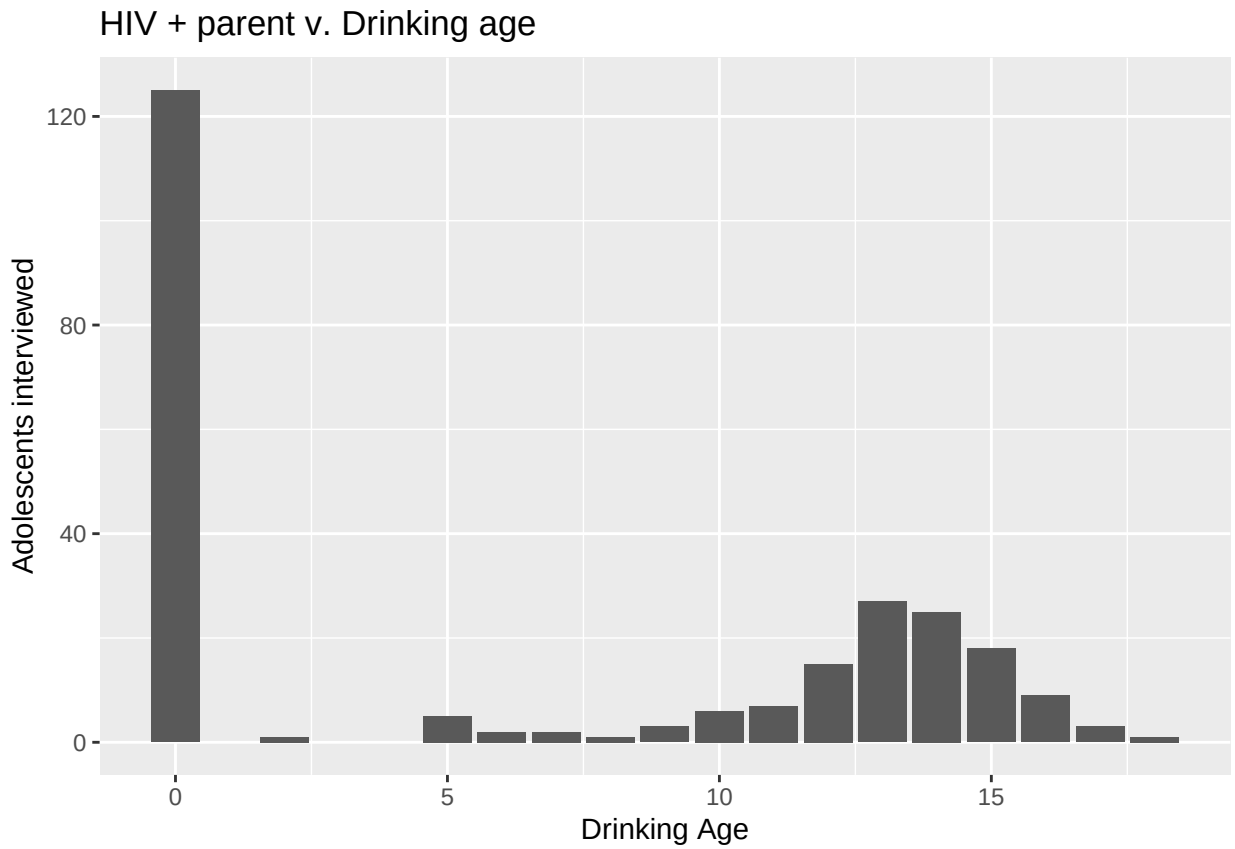
This displays that there are fewer adolescents that start smoking at a younger age, so having an HIV positive parent does not affect them starting to smoke.

2.a Univariate

Now we will be looking if an HIV positive parent affects the age of drinking

```
ggplot(hiv, aes(x=AGEALC))+geom_bar()+xlab("Drinking Age")+ylab("Adolescents interviewed")+ggtitle("HIV
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_count()‘).
```



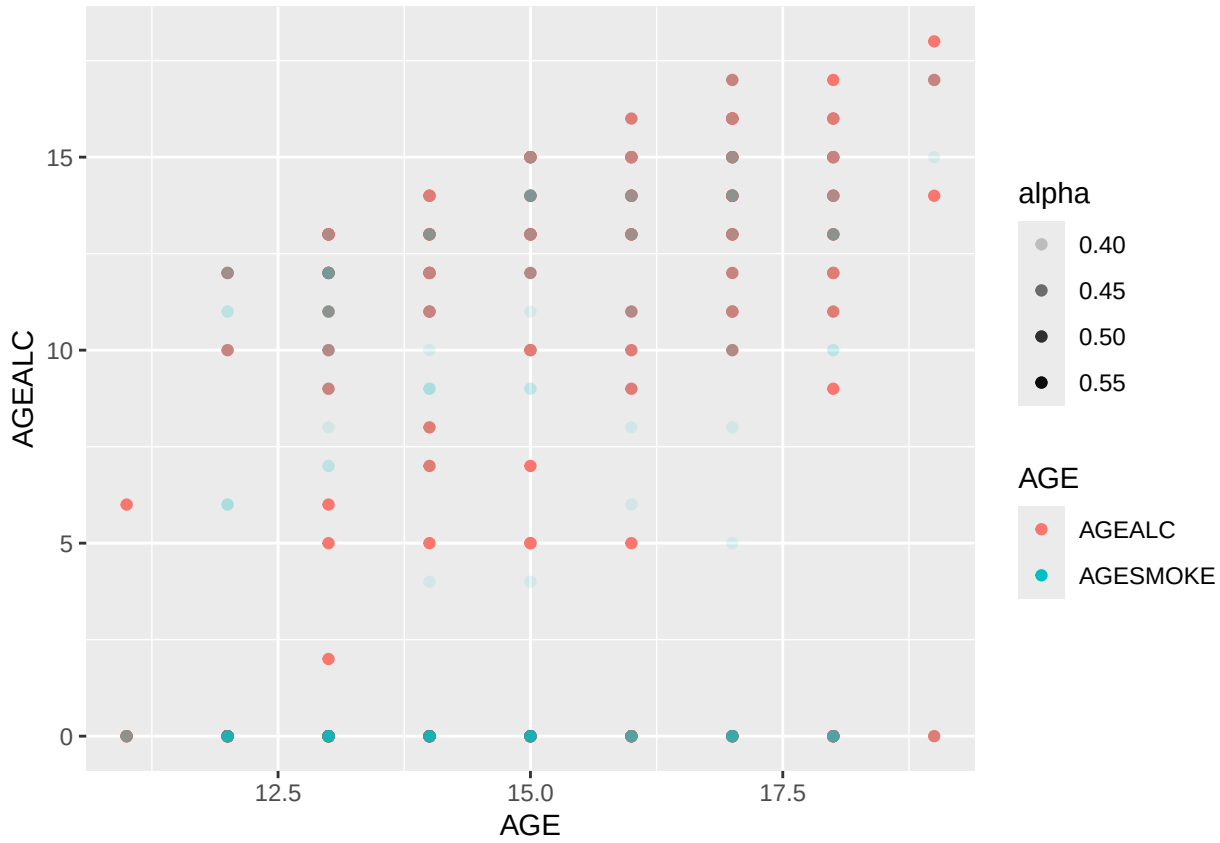
There is still no correlation between having an HIV + parent and start of drinking, However looking at the data there is an increase in the ages 13 and 14 for drinking.

3. Bivariate

We will now be comparing when adolescents began to smoke and drink

```
ggplot(hiv, aes(x=AGE, colour = AGE))+
  geom_point(aes(y=AGEALC, color="AGEALC", alpha=.6))+
  geom_point(aes(y=AGESMOKE, color="AGESMOKE",alpha=.4))
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```



4. Conclusion

When looking at both the variables for age of smoking and age of drinking, there is quite an overlap in the ages 15-17 of starting to smoke and drink. We were trying to find if there was a correlation in the age of when adolescents start to drink and smoke depending on being a child of a HIV + parent. The most common age between them would be around 16-17 years old to start drinking and smoking.