

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2

## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'tibble'

## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'pillar'
```

```
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
```

Importing Data - Parental HIV

```
parHIV <- read.delim("/Users/stephanie_reza2/Documents/MATH130/data/PARHIV_081217.txt", header=TRUE, st
dim(parHIV)
```

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

Introduction The Parental HIV data is a clinical trial that investigates families with a parent that has HIV. This data shows that at least 252 adolescents have parents with HIV. Some variables I want to explore in this data is age and ethnicity. Furthermore, my question is how many of these adolescents were under the age of 15 and? One thing I am interested in researching is the mean of the age variable in adolescents.

Examining the data imported

```
library(ggplot2)
library(dplyr)
str(parHIV)
```

```
## 'data.frame':   252 obs. of  123 variables:
## $ ID           : int  1 2 3 4 5 6 7 8 9 10 ...
## $ AGE          : int  17 18 13 14 14 13 14 14 13 13 ...
## $ GENDER       : chr  "Male" "Female" "Male" "Female" ...
## $ LIVWITH      : int  3 1 2 2 2 2 2 2 2 1 ...
## $ SIBLINGS     : int  2 NA 2 2 2 2 2 2 2 2 ...
## $ JOBMO        : int  NA 1 2 2 2 2 2 2 2 NA ...
```

```

## $ EDUMO           : int  2 2 1 3 3 NA 2 2 3 NA ...
## $ HOWREL          : int  NA 2 NA 3 2 2 2 1 NA NA ...
## $ ATTSERV         : int  NA 2 NA 1 1 2 2 3 NA NA ...
## $ NGHB1           : int  2 4 1 3 3 2 2 3 4 1 ...
## $ NGHB2           : int  2 2 1 3 4 4 3 3 4 1 ...
## $ NGHB3           : int  2 3 1 2 2 2 2 2 1 1 ...
## $ NGHB4           : int  1 2 2 4 2 2 4 3 1 1 ...
## $ NGHB5           : int  2 1 4 2 1 2 1 3 2 1 ...
## $ NGHB6           : int  2 1 4 2 2 4 2 4 1 1 ...
## $ NGHB7           : int  2 2 4 2 2 3 1 3 4 1 ...
## $ NGHB8           : int  2 1 1 4 3 4 3 4 4 1 ...
## $ NGHB9           : int  1 2 1 1 3 3 1 3 4 1 ...
## $ NGHB10          : int  1 3 1 4 2 1 2 3 4 1 ...
## $ NGHB11          : int  3 1 1 2 2 3 3 3 4 1 ...
## $ MONFOOD         : int  3 3 3 1 3 3 3 3 2 ...
## $ FINSIT          : int  4 3 1 3 4 3 3 1 4 2 ...
## $ ETHN            : int  2 3 1 2 2 2 2 1 2 1 ...
## $ AGESMOKE        : int  NA 14 12 14 NA 12 9 NA NA NA ...
## $ SMOKEP3M        : int  NA 4 1 8 NA 1 8 NA NA NA ...
## $ AGEALC          : int  14 15 0 14 0 0 0 0 0 ...
## $ AGEMAR          : int  14 17 0 14 0 0 13 0 0 ...
## $ FRNDS           : int  2 3 4 1 2 2 2 2 3 1 ...
## $ SCHOOL          : int  2 2 2 2 2 2 2 2 2 ...
## $ LIKESCH         : int  3 3 2 1 1 3 3 1 1 3 ...
## $ HOOKEY          : int  2 2 1 2 2 1 1 1 2 1 ...
## $ NHOOKEY         : int  3 4 0 1 4 0 0 0 2 0 ...
## $ HMONTH          : int  3 3 3 3 1 1 3 1 1 1 ...
## $ PB01            : int  3 3 4 1 4 3 3 4 4 3 ...
## $ PB02            : int  1 2 1 1 1 4 1 2 2 2 ...
## $ PB03            : int  3 1 4 1 4 1 3 1 3 2 ...
## $ PB04            : int  1 3 1 3 1 3 1 4 2 3 ...
## $ PB05            : int  4 3 4 1 3 1 3 4 3 3 ...
## $ PB06            : int  4 4 4 1 4 4 4 3 3 3 ...
## $ PB07            : int  3 1 3 2 3 1 4 2 3 3 ...
## $ PB08            : int  1 4 2 3 1 1 3 2 2 3 ...
## $ PB09            : int  1 4 3 4 1 3 3 4 2 1 ...
## $ PB10           : int  1 3 2 4 3 4 2 2 3 2 ...
## $ PB11           : int  4 2 4 1 3 4 4 3 3 3 ...
## $ PB12           : int  4 2 4 1 3 4 4 4 3 4 ...
## $ PB13           : int  3 1 4 1 3 4 3 2 3 3 ...
## $ PB14           : int  1 4 2 4 2 4 2 1 2 2 ...
## $ PB15           : int  4 1 4 2 4 4 3 2 3 3 ...
## $ PB16           : int  1 2 2 1 1 1 1 2 2 1 ...
## $ PB17           : int  4 3 4 1 4 4 3 4 3 3 ...
## $ PB18           : int  1 3 2 1 1 1 1 1 2 2 ...
## $ PB19           : int  3 3 3 3 1 3 2 2 2 2 ...
## $ PB20           : int  3 4 2 1 1 3 2 1 2 3 ...
## $ PB21           : int  4 1 4 2 4 2 2 2 3 3 ...
## $ PB22           : int  4 1 2 3 3 1 3 2 3 2 ...
## $ PB23           : int  3 4 4 2 1 4 3 4 3 3 ...
## $ PB24           : int  3 3 1 1 4 1 3 1 2 3 ...
## $ PB25           : int  4 2 4 3 3 4 4 2 2 3 ...
## $ BSI01           : int  0 1 0 0 0 0 0 0 0 0 ...
## $ BSI02           : int  0 0 0 0 1 0 0 4 0 0 ...

```

```

## $ BSI03      : int  0 0 0 0 0 0 1 0 0 0 ...
## $ BSI04      : int  0 0 0 0 0 1 1 0 0 0 ...
## $ BSI05      : int  0 0 0 0 0 2 1 0 0 0 ...
## $ BSI06      : int  0 0 0 1 1 1 1 1 0 0 ...
## $ BSI07      : int  0 0 0 1 0 0 2 0 0 0 ...
## $ BSI08      : int  0 0 0 0 0 1 1 0 0 0 ...
## $ BSI09      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ BSI10      : int  1 0 0 0 0 2 1 3 3 0 ...
## $ BSI11      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ BSI12      : int  0 0 0 0 0 0 0 1 0 0 ...
## $ BSI13      : int  3 0 0 1 0 2 0 0 0 0 ...
## $ BSI14      : int  2 3 0 0 0 0 0 0 0 0 ...
## $ BSI15      : int  0 0 0 0 0 1 0 0 0 0 ...
## $ BSI16      : int  0 3 0 0 0 1 0 0 0 0 ...
## $ BSI17      : int  0 2 0 0 1 0 1 0 0 0 ...
## $ BSI18      : int  0 0 0 0 0 3 0 1 0 0 ...
## $ BSI19      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ BSI20      : int  2 2 0 0 0 0 0 3 0 0 ...
## $ BSI21      : int  0 0 0 0 0 0 0 0 1 0 ...
## $ BSI22      : int  1 0 0 0 0 0 0 0 0 0 ...
## $ BSI23      : int  0 0 0 1 0 0 0 0 0 0 ...
## $ BSI24      : int  2 0 0 0 1 1 1 1 0 0 ...
## $ BSI25      : int  0 4 0 0 0 1 0 0 0 0 ...
## $ BSI26      : int  0 0 0 0 1 1 1 3 0 0 ...
## $ BSI27      : int  1 0 0 0 0 1 0 1 0 0 ...
## $ BSI28      : int  0 0 0 1 0 1 0 0 0 0 ...
## $ BSI29      : int  0 0 0 1 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 0 ...
## $ BSI32      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ BSI33      : int  0 0 0 0 1 1 0 0 0 0 ...
## $ BSI34      : int  1 0 0 1 0 2 0 0 0 0 ...
## $ BSI35      : int  0 0 0 0 0 2 0 1 0 0 ...
## $ BSI36      : int  2 0 0 0 0 1 0 3 0 0 ...
## $ BSI37      : int  0 0 0 1 1 1 0 0 0 0 ...
## $ BSI38      : int  0 2 0 0 0 2 0 0 0 0 ...
## $ BSI39      : int  0 0 0 0 0 1 0 0 1 0 ...
## $ BSI40      : int  2 0 0 0 1 1 0 0 0 0 ...
## $ BSI41      : int  2 0 0 1 0 2 0 0 0 0 ...
## [list output truncated]

```

Univariate Exploration: Table of Age

```
table(parHIV$AGE)
```

```

##
## 11 12 13 14 15 16 17 18 19
##  5 30 44 38 39 28 39 25  4

```

Mean of Age

```
mean(parHIV$AGE, na.rm=TRUE)
```

```
## [1] 14.84524
```

Table of Ethnicity

```
table(parHIV$ETHN)
```

```
##  
##  1  2  3  
## 123 96 33
```

Mean of ethnicity

```
mean(parHIV$ETHN)
```

```
## [1] 1.642857
```

Summary Stats for AGE

```
summary(parHIV$AGE) #The minimum is 11, the maximum is 19 , and the median is 15.
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##  11.00  13.00   15.00   14.85  17.00   19.00
```

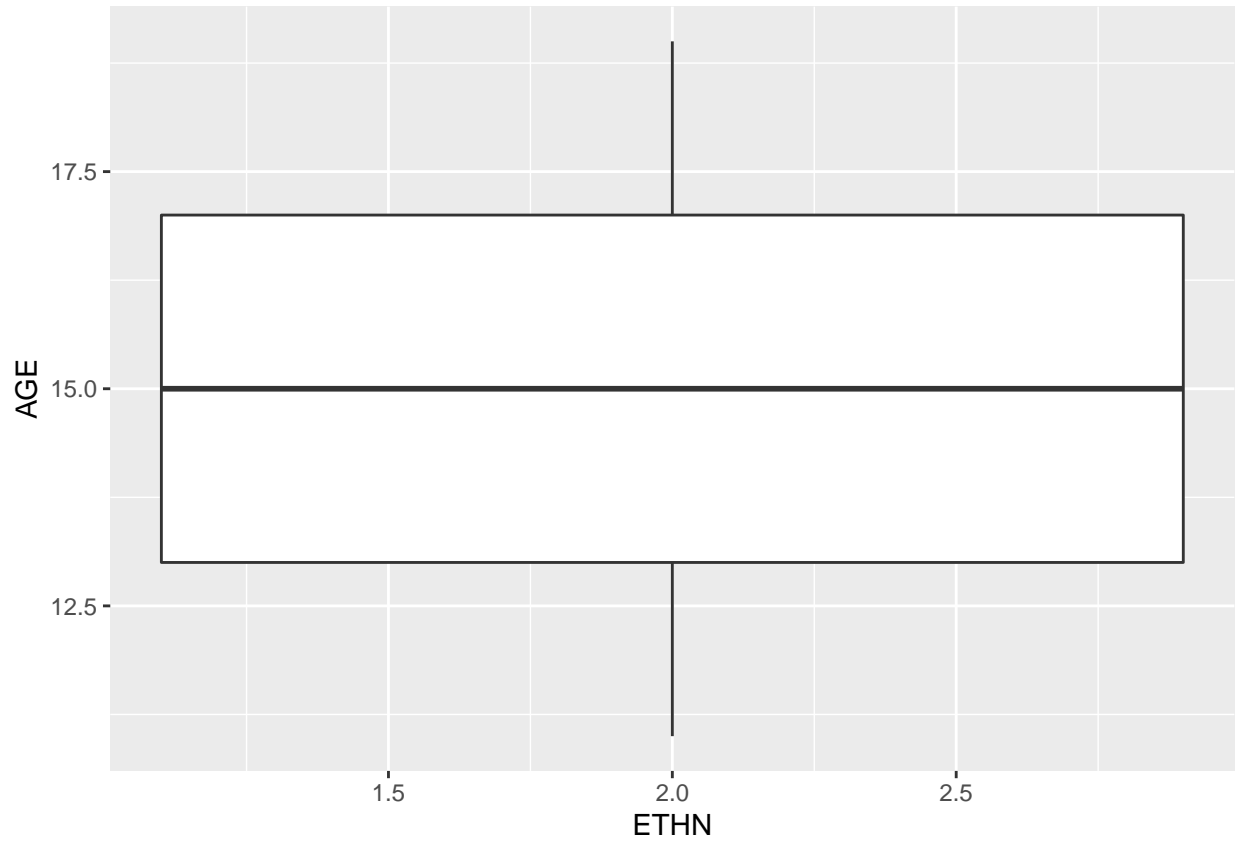
Summary Stats for ETHN

```
summary(parHIV$ETHN) #The minimum is 1, the maximum is 3, and the median is 2.
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##   1.000  1.000   2.000   1.643  2.000   3.000
```

```
ggplot(parHIV, aes(x=ETHN, y=AGE)) + geom_boxplot()
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



Bivariate Exploration

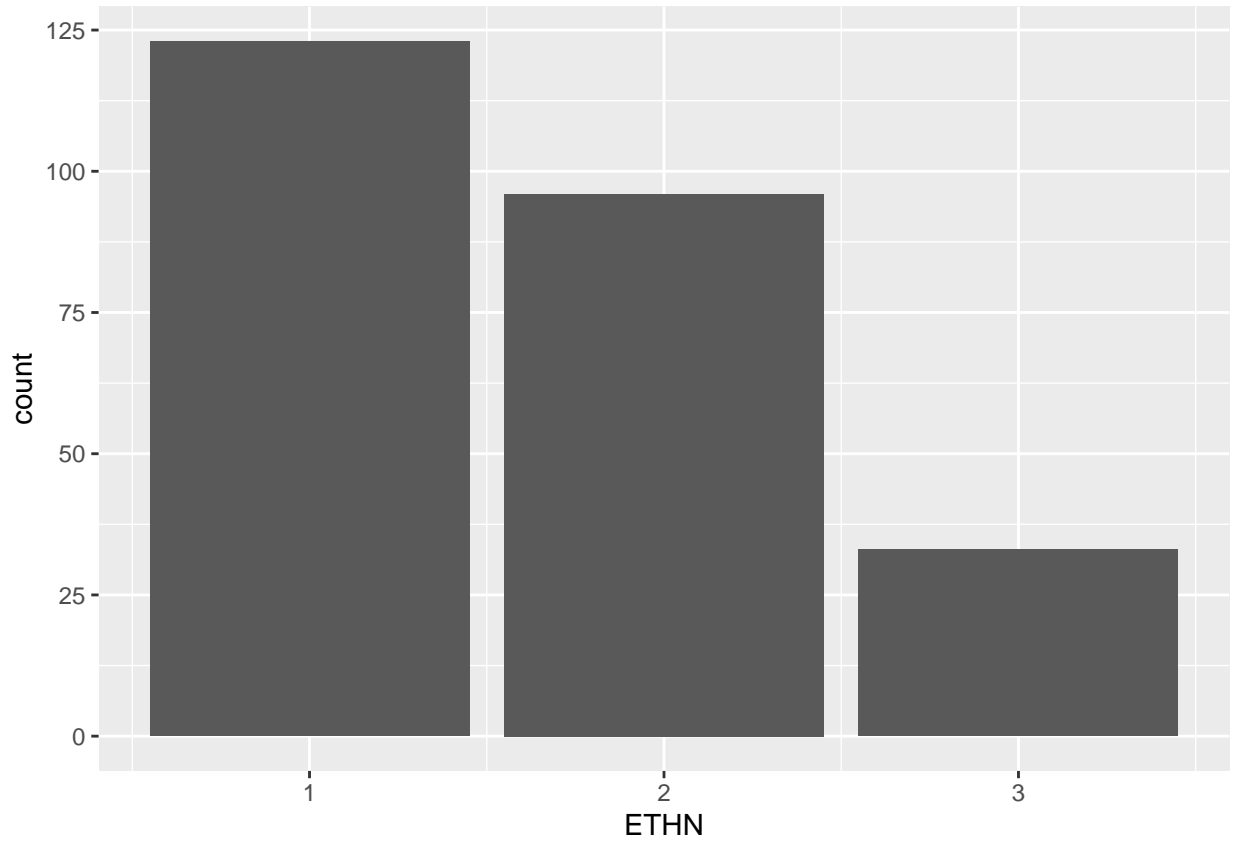
```
table(parHIV$ETHN, parHIV$AGE)
```

```
##
##      11 12 13 14 15 16 17 18 19
##  1  2 18 20 20 17 14 22  9  1
##  2  2  9 18 11 16 12 14 13  1
##  3  1  3  6  7  6  2  3  3  2
```

There are 3 different types of ethnicity and all range from the ages of 1-20.

Grouped Chart for my data

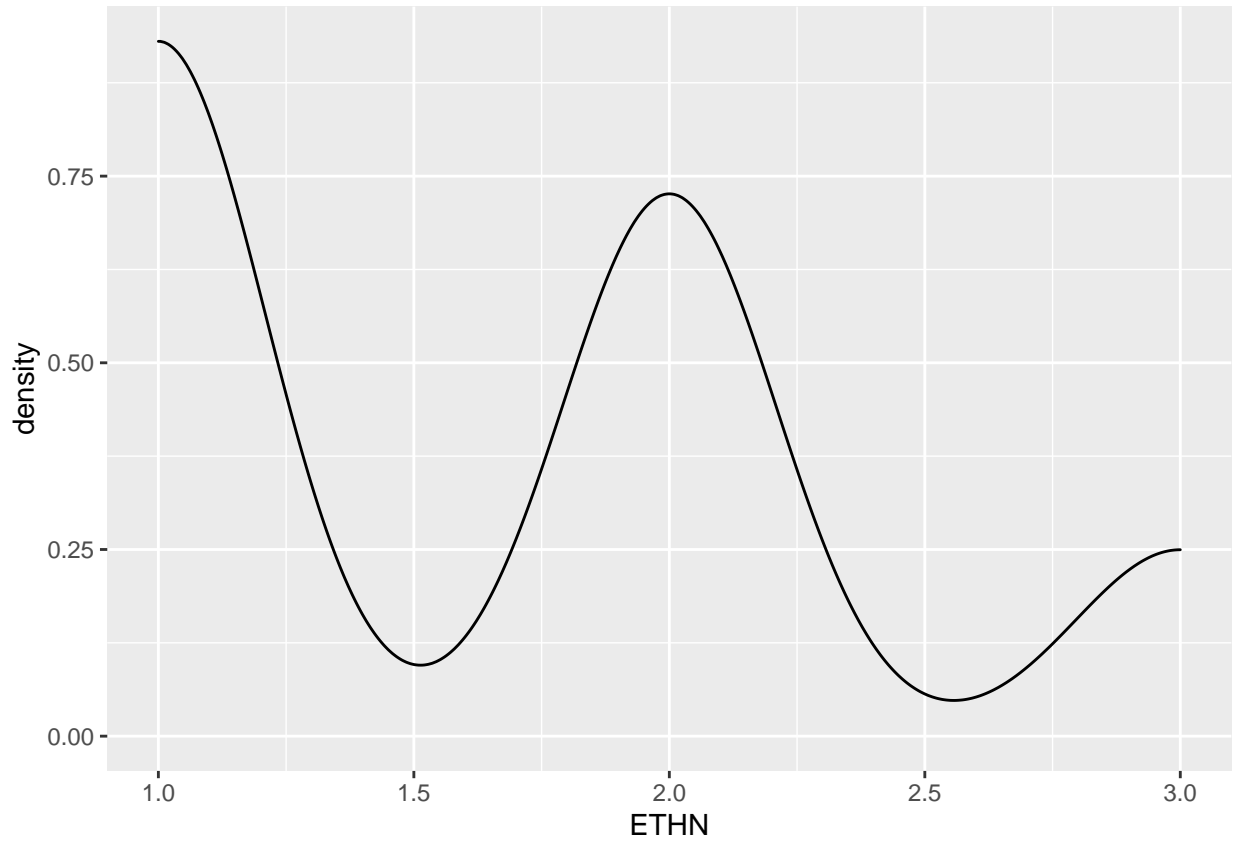
```
ggplot(parHIV, aes(x=ETHN, fill=AGE)) + geom_bar(position = "dodge")
```



The x axis shows the different number of ethnicity while the second categorical variable is the age count. This chart shows that the ENTH variable is within the same level of the AGE variable.

More Data...

```
ggplot(parHIV, aes(x=ETHN, color=AGE)) + geom_density()
```



The graph above shows plots for continuous variables. The ethnicity being on the x-axis while the y-axis shows the age range.

Conclusion I discovered that my ethnicity variable is difficult to work with. The age range is around 11-19 and there are only 3 different ethnicity to work with. My data seems to be off maybe because my variables don't make sense. Based on my data, the median of my age was 15 and the mean was 14.84. (Failed/rejected)