

Exploratory Data Analysis Project

By: Madison Matsunami

Date: 09/25/2022

Introduction

The data set that I will be observing is the Parental HIV data set that contains 253 observations and 123 variables. This information presents the amount of data on adolescent children living with HIV positive parents. Specifically, I will be looking into the relationship between males and females under the age of 18 and how many live with one or both of their parents. Are there more adolescents living with their parents who are positive for HIV under the age of 18?

```
library(ggplot2)
library(scales)
library(gridExtra)
library(dplyr)
```

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

## The following object is masked from 'package:gridExtra':
##
##   combine

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

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

```
parHIV <- read.delim("/Users/madisonmatsunami/Desktop/PARHIV_081217.txt", header=TRUE, stringsAsFactors=
dim(parHIV)
```

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

2. Univariate Analysis

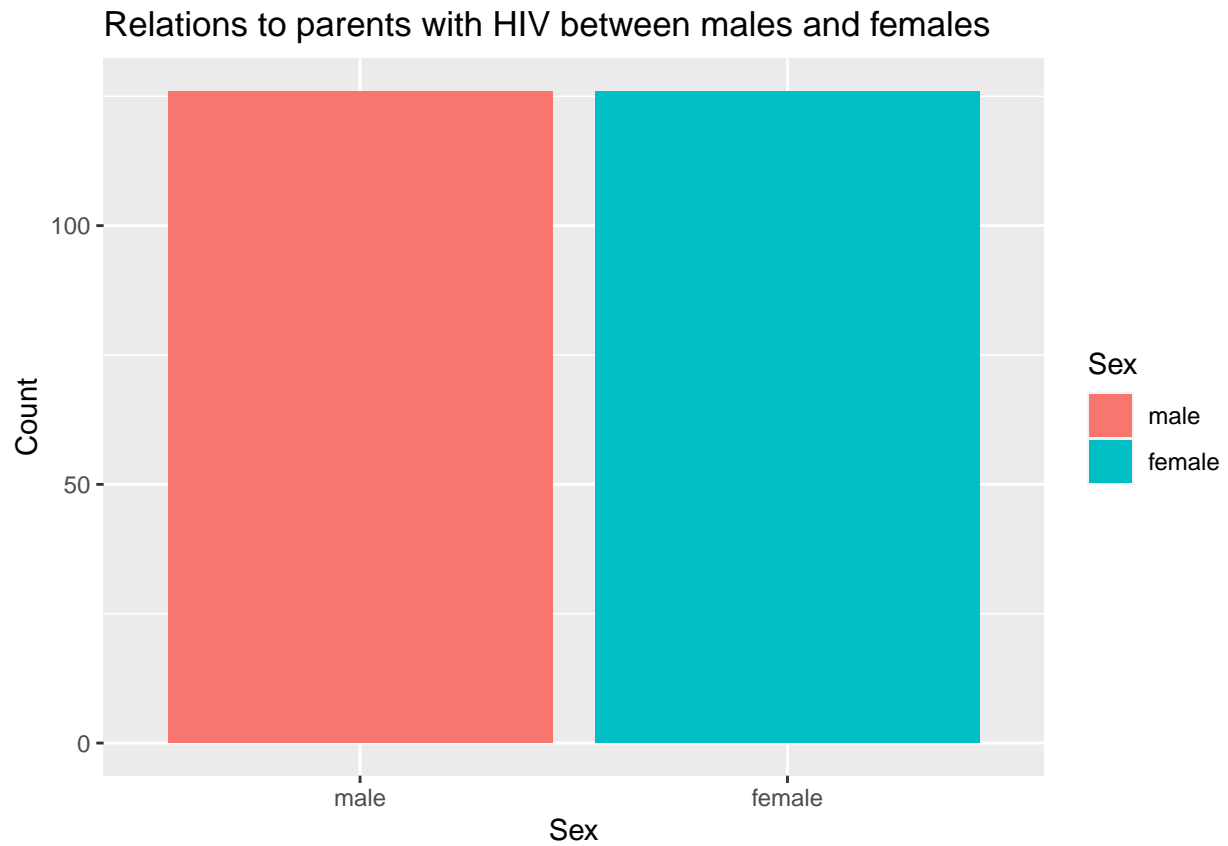
Gender of Those Related To Parents With HIV

```
parHIV$sexrename <- factor(parHIV$GENDER, labels=c("male", "female"))
summary(parHIV$sexrename)
```

```
##   male female
##   126    126
```

This data set explores the gender between males and females and how many are related to parents who are positive for HIV.

```
ggplot(parHIV, aes(x=sexrename, fill=sexrename)) + geom_bar() + xlab("Sex") + ylab("Count") + ggtitle("Relations to parents with HIV between males and females")
```

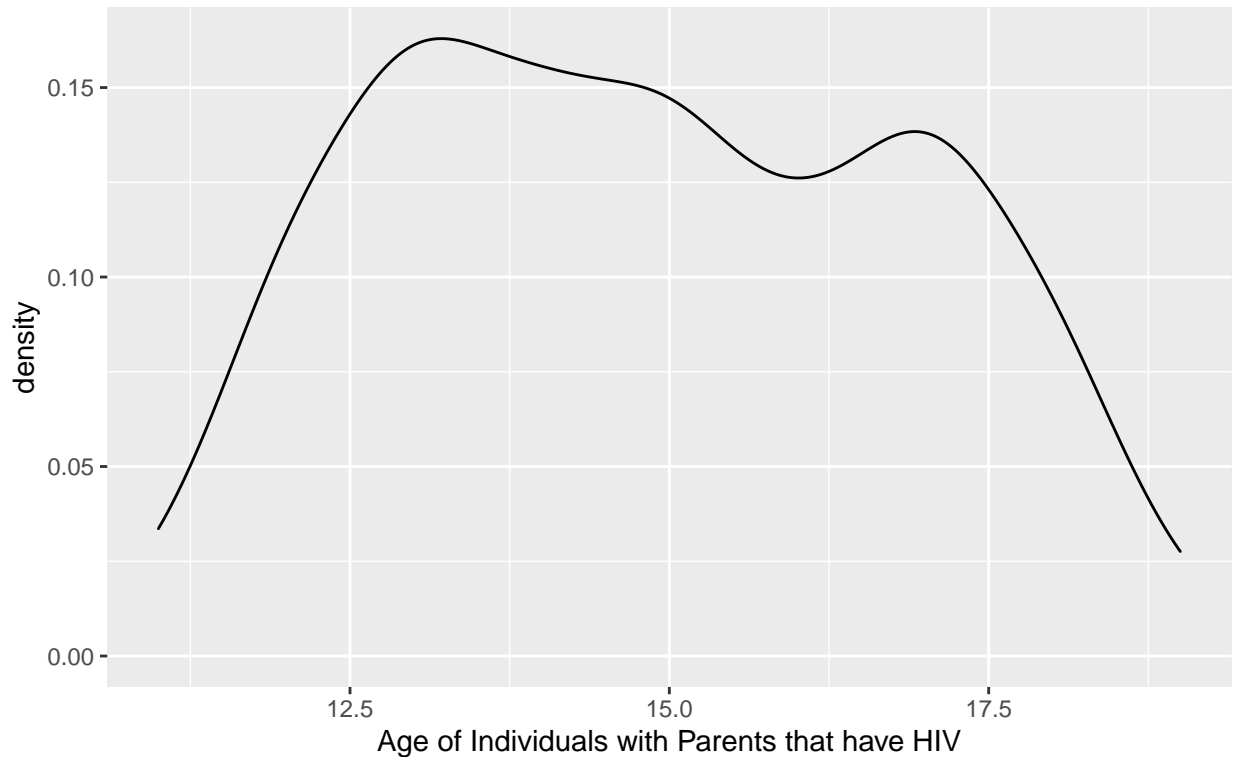


Age Of Individuals

This data set explores the age between males and females and how many are related to parents who are positive for HIV. This graph shows us that the majority of adolescents that have parents with HIV are between the ages of 12-15 years.

```
ggplot(parHIV, aes(x=AGE)) + geom_density() + xlab("Age of Individuals with Parents that have HIV") + ggtitle("Age of Individuals with Parents that have HIV")
```

Density of Age of Individuals with Parents that have HIV



Currently Living With

This data set explores the amount of adolescents that still live with their parents who are positive for HIV.

```
parHIV$rename <- factor(parHIV$LIVWITH, labels=c("Both Parents=1" , "One Parent=2", "Other=3"))  
table(parHIV$LIVWITH)
```

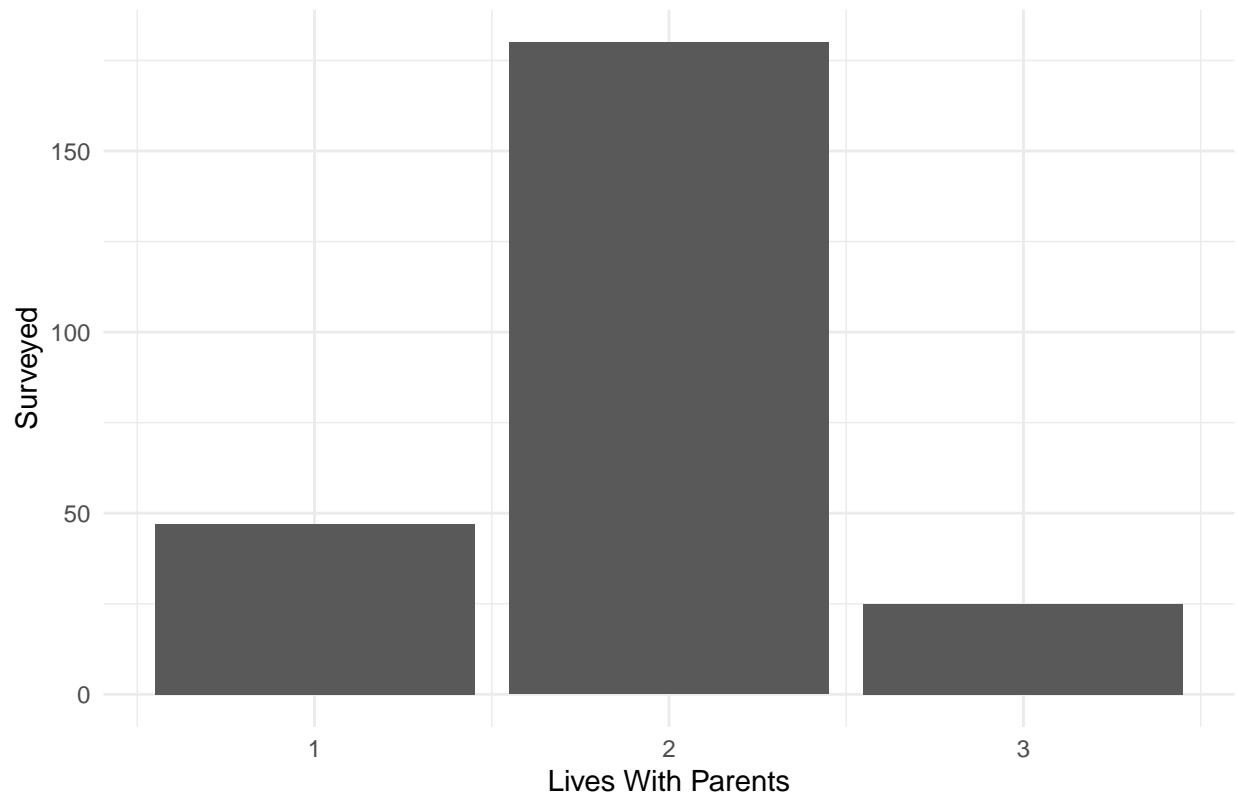
```
##  
##  1  2  3  
## 47 180 25
```

```
summary(parHIV$LIVWITH)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##  1.000  2.000   2.000  1.913  2.000   3.000
```

```
ggplot(parHIV, aes(x=LIVWITH, fill=LIVWITH)) + geom_bar() + xlab("Lives With Parents") + ylab("Surveyed")
```

Adolescents Who Live With Their Parents



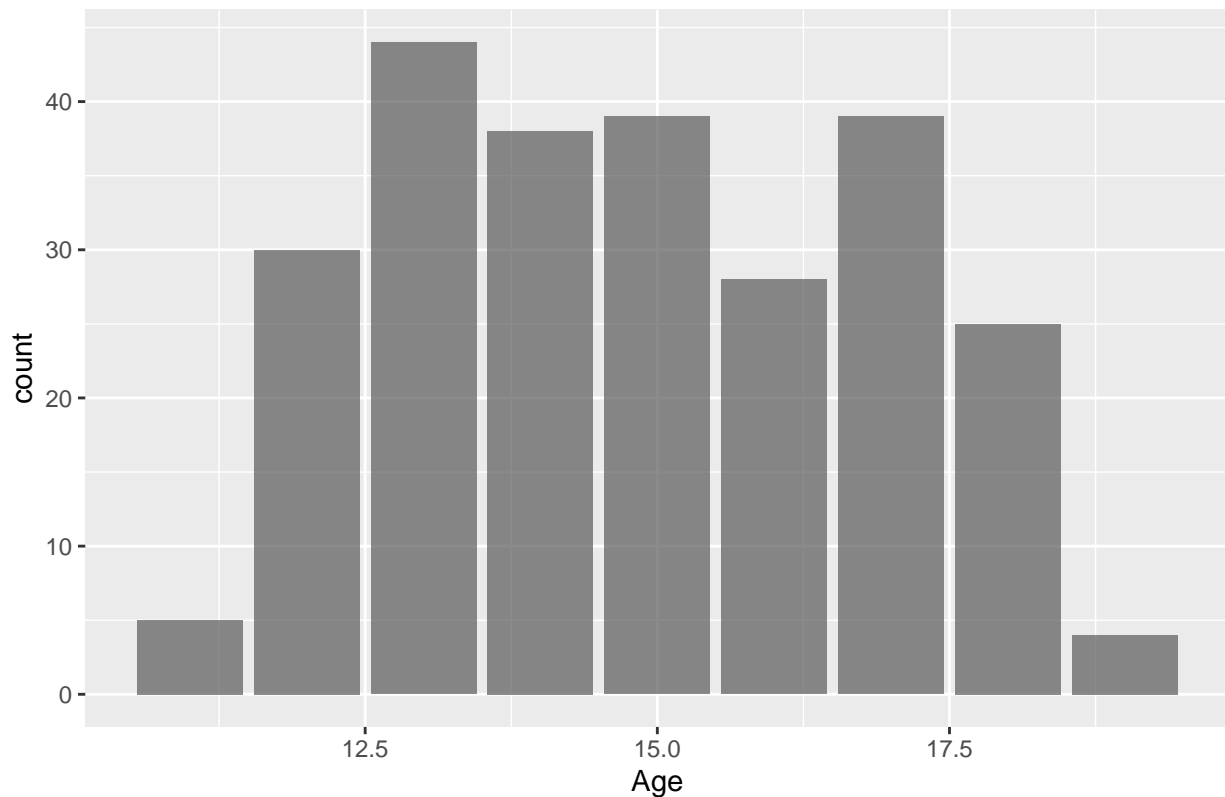
The table and graph above shows that more adolescents live with only one of their parents rather than both or other.

Bivariate Analysis

Age Vs. Individuals Living with Their Parents

```
parHIV%>%select(AGE,LIVWITH)%>%filter(!is.na(AGE))%>%  
ggplot(aes(x=AGE,fill=LIVWITH))+geom_bar(position="dodge",alpha=0.7)+  
scale_fill_discrete(name="Lives With Parents")+  
xlab("Age")+ggtitle("Age vs.Individuals living with Parents")
```

Age vs. Individuals living with Parents



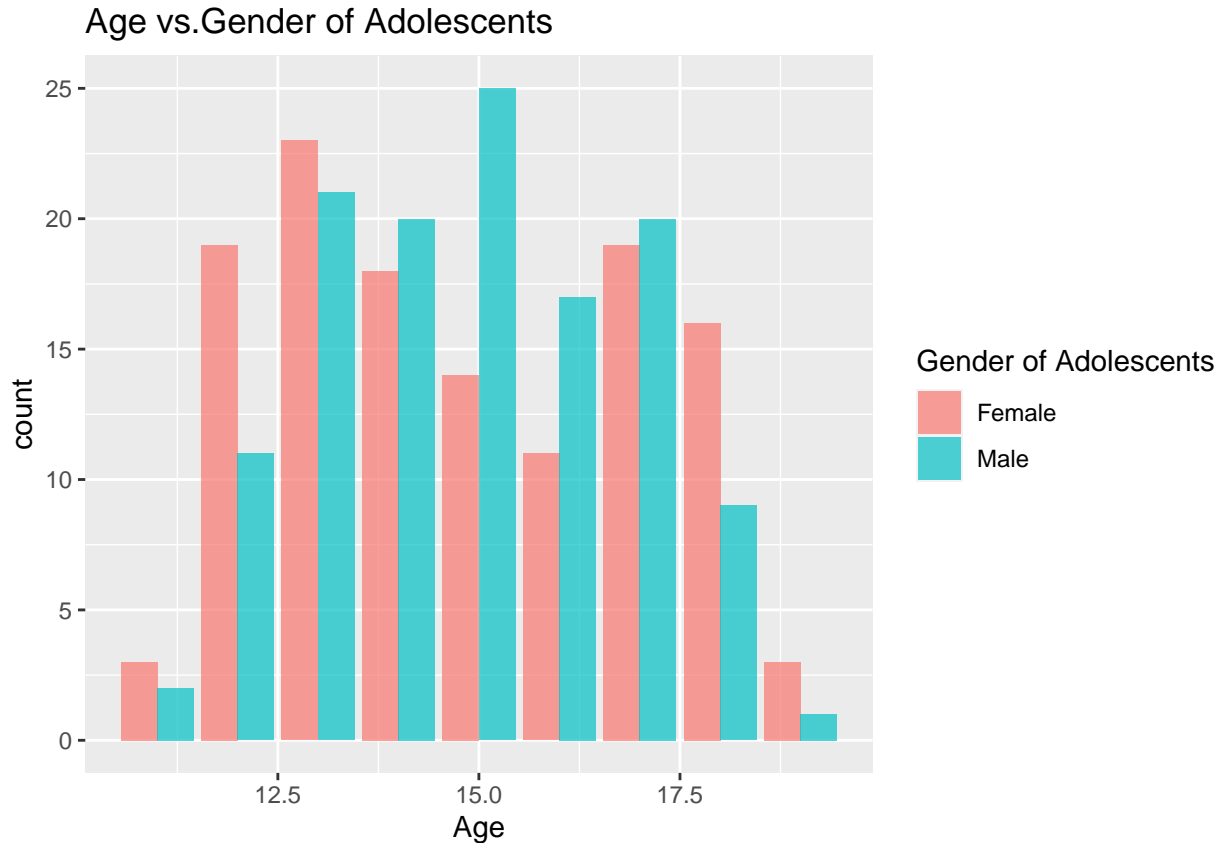
Based on the bar graph above we can see the age of adolescents and how many are living with their parents. We can see from the trends above that the majority of adolescents who live with their parents are between the ages of 12 and 15.

Gender Vs. Age

```
table(parHIV$AGE,parHIV$GENDER)%>%prop.table(margin=1)*100%>%round(2)
```

```
##
##      Female      Male
## 11 60.00000 40.00000
## 12 63.33333 36.66667
## 13 52.27273 47.72727
## 14 47.36842 52.63158
## 15 35.89744 64.10256
## 16 39.28571 60.71429
## 17 48.71795 51.28205
## 18 64.00000 36.00000
## 19 75.00000 25.00000
```

```
parHIV%>%select(AGE,GENDER)%>%filter(!is.na(AGE))%>%
ggplot(aes(x=AGE,fill=GENDER))+geom_bar(position="dodge",alpha=0.7)+
scale_fill_discrete(name="Gender of Adolescents")+
xlab("Age")+ggtitle("Age vs. Gender of Adolescents")
```



This bar graph shows the comparison of the age of adolescents to what their gender is. We can see that there are more females fifteen and under, and there are more males fifteen and over.

Conclusion

Based on our analysis from the data that we observed, we can conclude that the majority of adolescents live with at least one of their parents. There is an even amount of males and females in this data set however, the majority of males that live with at least one of their parents are over the age of 15. The majority of females who live with at least one of their parents are under the age of 15. The highest number of adolescents who are living with their parents are between the ages of 12 and 15. The relationship between males and females that are under 18 are more likely to live with their parents who are positive for HIV. We can conclude that the majority of both males and females are under the age of 18 and are still living with their parents who are positive for HIV.