

# EDA Project By Joseph Huitt

Joseph Huitt

## Introduction

The data set I have chosen to analyze is a cumulation of 252 adolescent children of parents with HIV with 123 variables from the UCLA Neuropsychiatric Institute. The specific variables that I have chosen to analyze is gender, jobmo and siblings. The reason that I have selected to compare these variables is because I want to explore if and why gender in HVI children plays a roll in if there siblings have it. In addition I would like to see if there is a correlation between having a job and prevalence of HIV. Finally I would like to analyze which gender is more likely to be infected with HIV. Due to the fact that it is commonly found in males.

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

```
HIV <- read.delim("/Users/jojo/Desktop/math130/data/HIV.txt", header = TRUE, sep = "\t")
dim(HIV)
```

## Univariate Analysis

Variables being Observed: Gender VS Siblings & Gender VS Jobmo

```
summary(HIV$GENDER)
```

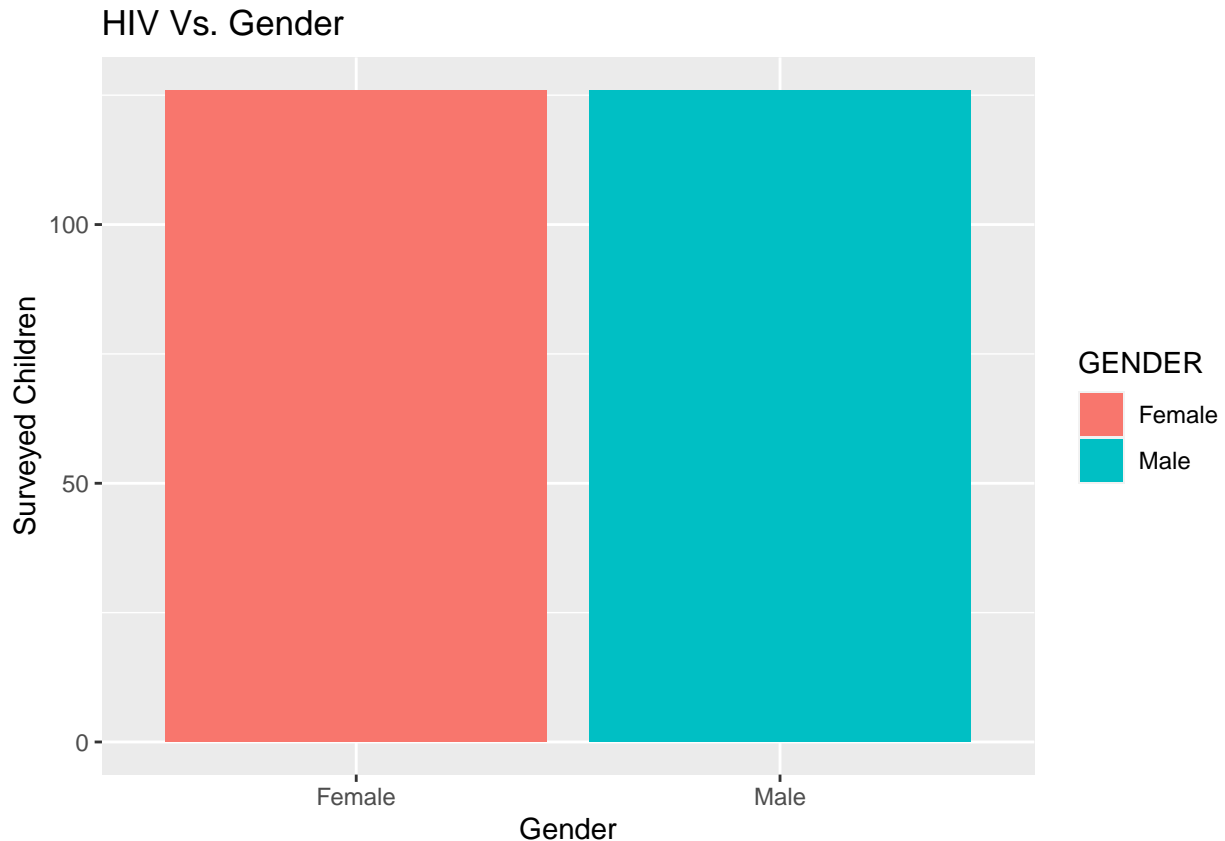
```
##      Length      Class      Mode
##      252 character character
```

```
table(HIV$GENDER)
```

```
##
## Female   Male
##    126    126
```

Based on the table shown above we can conclude that gender does not play a roll in if they child will be affected or not. This most likely has to do with both children being inside the mother while she is infected instead of relying on a bodily fluid transfer that any other person would need to become infected.

```
ggplot(HIV, aes(x=GENDER, fill=GENDER))+geom_bar()+xlab("Gender")+ylab("Surveyed Children")+ggtitle("HI
```



The generated results from this chart supports the conclusion of the summary above =, Showing that both genders are able to be equally infected.

### Comparing Siblings and HIV

```
HIV$SIBLINGSrename<- factor(HIV$SIBLINGS, labels = c("1 Sibling", "2 Siblings"))
summary(HIV$SIBLINGS)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      1.000  2.000   2.000  1.944  2.000   2.000     1
```

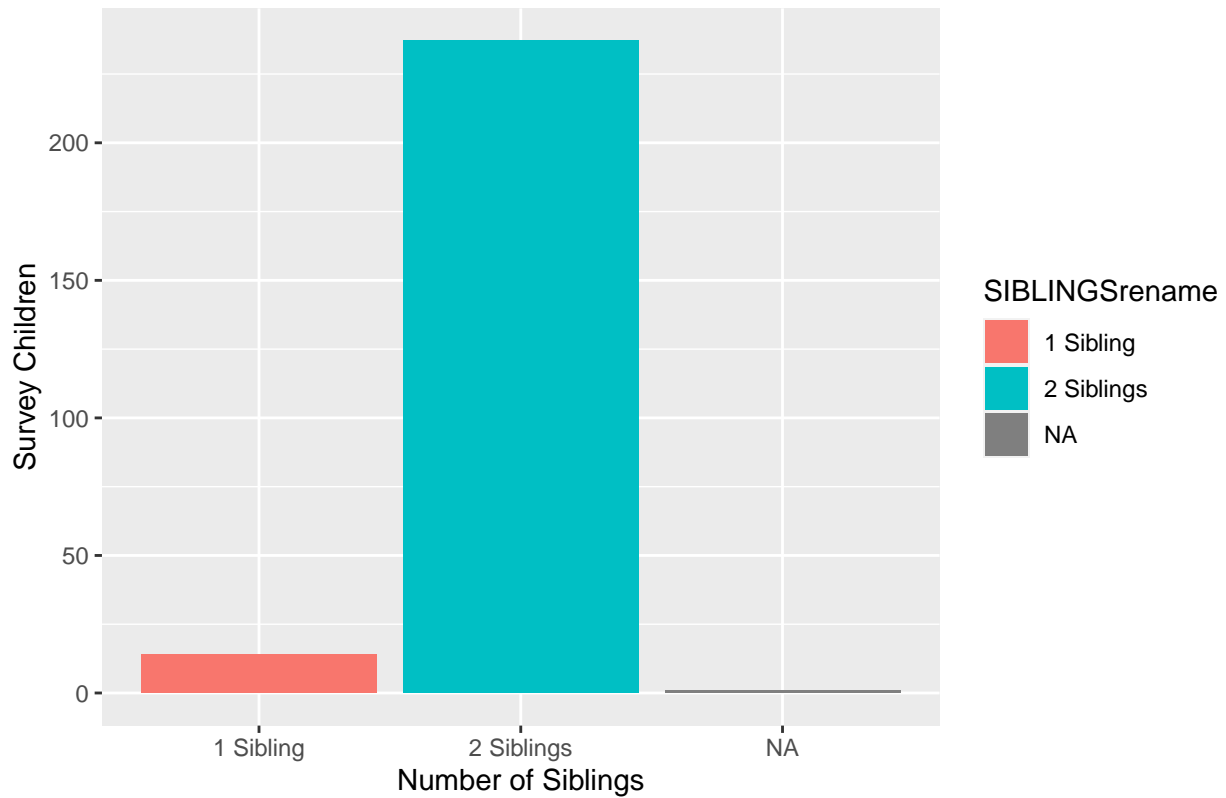
```
table(HIV$SIBLINGS)
```

```
##
##      1      2
##     14    237
```

Based off the surveyed results we are able to see that children with two siblings have a very high prevalence compared to those who would only have one sibling. This may result due to the child only having one other sibling or that the parent was not infected with HIV before they had the previous children.

```
HIV$SIBLINGSrename<- factor(HIV$SIBLINGS, labels = c("1 Sibling", "2 Siblings"))
ggplot(HIV, aes(x=SIBLINGSrename, fill=SIBLINGSrename))+geom_bar()+xlab("Number of Siblings")+ylab("Sur
```

## HIV Vs. Number of Siblings



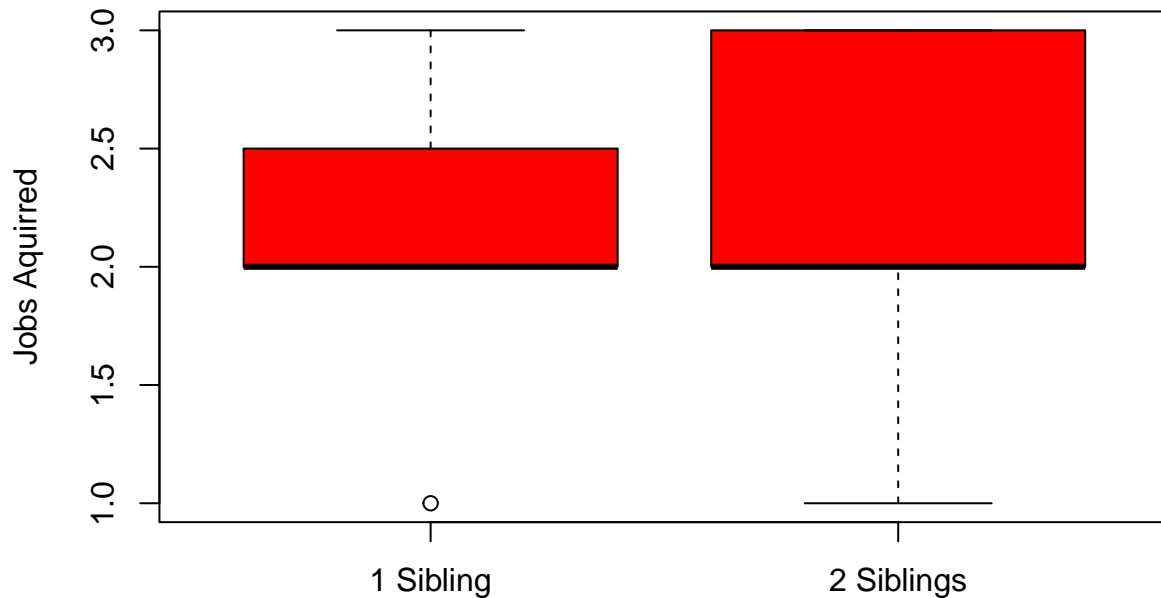
As shown above we are able to determine that children who were infected with HIV have a tendency to have two siblings also infected with the virus.

## BIVARIATE ANALYSIS

HIV VS SIBLINGS AS IT RELATES TO HAVING A JOB

```
plot(HIV$SIBLINGSrename, HIV$JOBMO, pch=1, col='red', main='Siblings vs Job Holdings', xlab='Number of
```

## Siblings vs Job Holdings



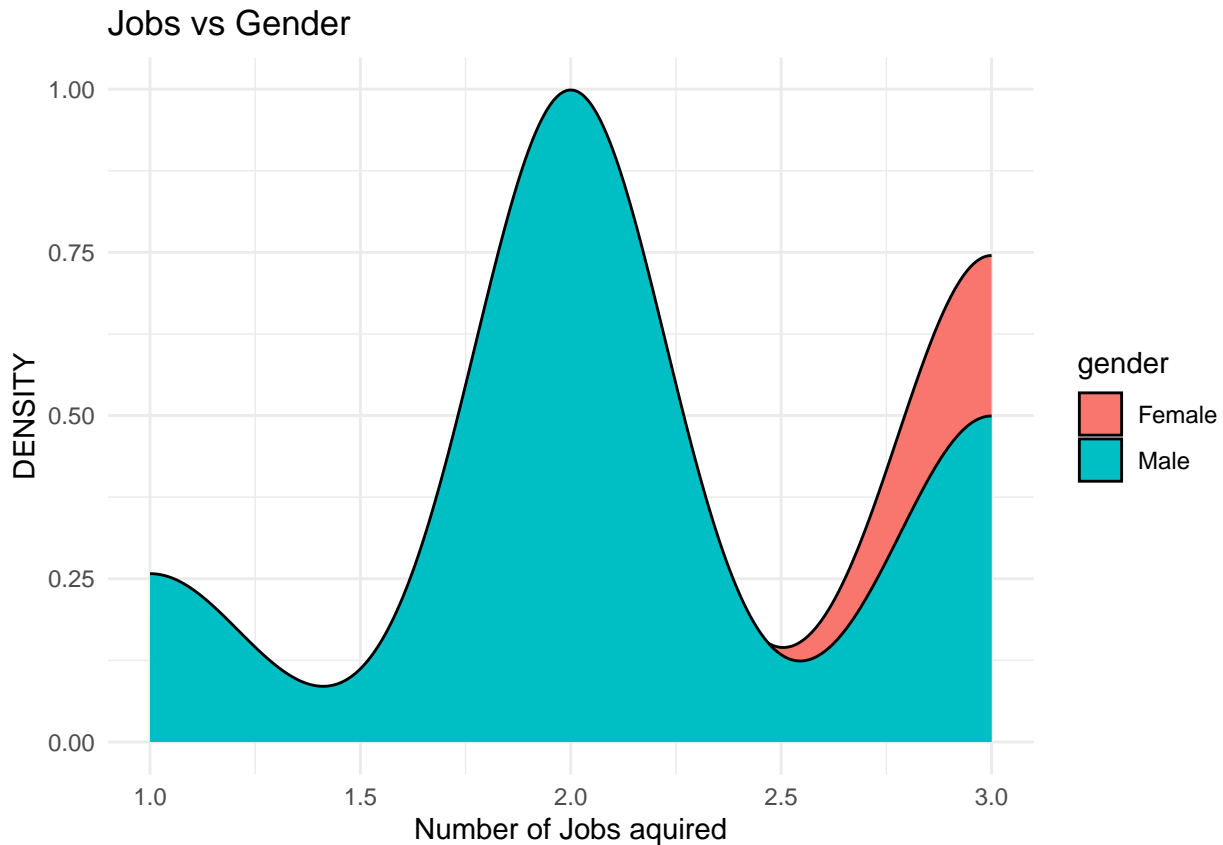
### Number of Siblings

Based

on the box plot we see on average children with 2 siblings also infected have a higher chance of acquiring a job. This could be due to the fact that the child knows that HIV does not have to hold them back if there other siblings have also been able to overcome its effects. This could also be related to the fact that they might be using the virus to try and help find a cure not only for themselves but also for their family members.

```
ggplot(HIV, aes(x=JOBMO, fill=GENDER))+geom_density()+scale_fill_discrete(name="gender")+xlab("Number of Siblings")
```

```
## Warning: Removed 24 rows containing non-finite values ('stat_density()').
```



Based on this chart we are able to see that females have had to acquire morw jobs than men. This shows that the women that have HIV seem to have a harder time holdong onto there job. This could be due to the fact that there job or coworkers fid out that they have been infected and have to leave there place of work.

### Conclusion

In conclusion we explore three different variable, gender, siblings, and number of jobs aquired as ttthey related to having a parent with HIV based off the 252 participants from the UCLA Neuropsychiatric Institute. Based off of the results we are able to determine that gender does not play a roll in if a child has HIV or not. This most likely has to do with the way infection takes place before conception. We also saw that if a child had siblings they were more likely to have 2 siblings infected than just one sibling. We also looked at how having a sibling related to being able to aquire a job. Through this we are able to see many of the 2 sibling children were more likely to aquire a job. Lastly we looked at if gender played a role in how many jobs a person had. We were able to determine that women mostly had 3 jobs were as many of themales only had two jobs. I believe that this is due to a social aspect were many women gossip more and by finding out someone has HIV they can push them out of the work place. Lastly, I believe that any person can succeed with HIV as long as they put there mind to the task at hand.