

Final Project

Rachel Olive

9/20/2021

```
par.HIV<-read.table("/Users/rachelolive/Desktop/Math 130/data/par.HIV.txt", header=TRUE)
```

Coded Libraries

```
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
```

```
library(readxl)  
library(readr)  
library(gridExtra)
```

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

```
library(ggplot2)  
library(forcats)
```

Intro: The data set that I am using is the Parental HIV data set. This data set includes information on a subset of 252 adolescent children of parents with HIV. The study has a total of 252 observations and 123 variables. I will be making a total of 3 models to represent different observations made in the study and the variables that let to them ' Example 1: I was interested in finding out what the average number of siblings that children of HIV positive parents had. I hypothesized that these children would not have high numbers of siblings due to the effects that HIV can have on fertility. Variable:Siblings

```
par.HIV$SIBLINGS<-factor(par.HIV$SIBLINGS,labels=c("1","2"))  
summary(par.HIV$SIBLINGS)
```

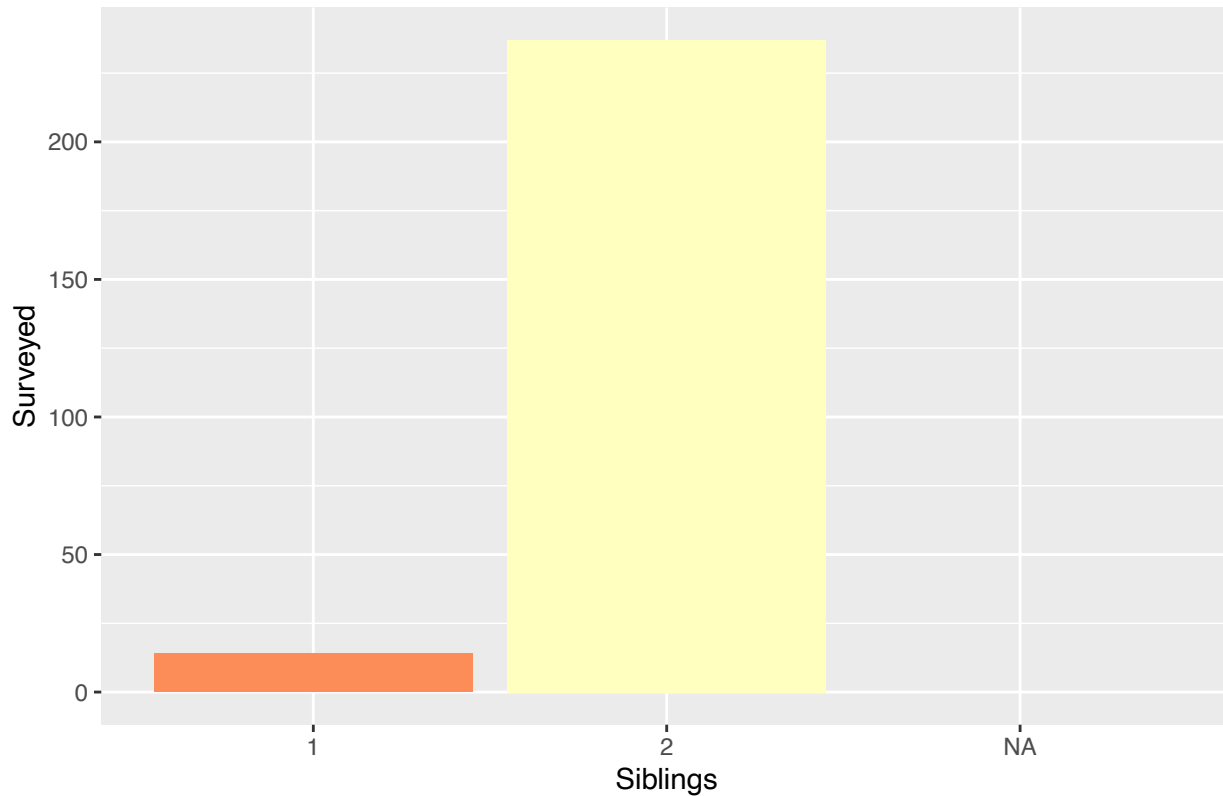
```
##   1    2 NA's  
##  14 237    1
```

```
ggplot(par.HIV,aes(x=SIBLINGS,fill=SIBLINGS))+geom_bar()+xlab("Siblings") + scale_fill_brewer(palette="Spectral", guide=FALSE) + ylab("Surveyed") + ggtitle("Number of Siblings in children with HIV positive p
```

```
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
```

```
## use `guide = "none"` instead.
```

Number of Siblings in children with HIV positive parents



The data here proved my initial hypothesis incorrect. While the number of siblings only ranged from 0(NA) to 2, the majority of the people in the data set had 2 siblings, 237 out of total 252 people surveyed had 2 or more siblings. All but 1 of these people surveyed had at least one sibling, that outlier is represented by NA on the graph.

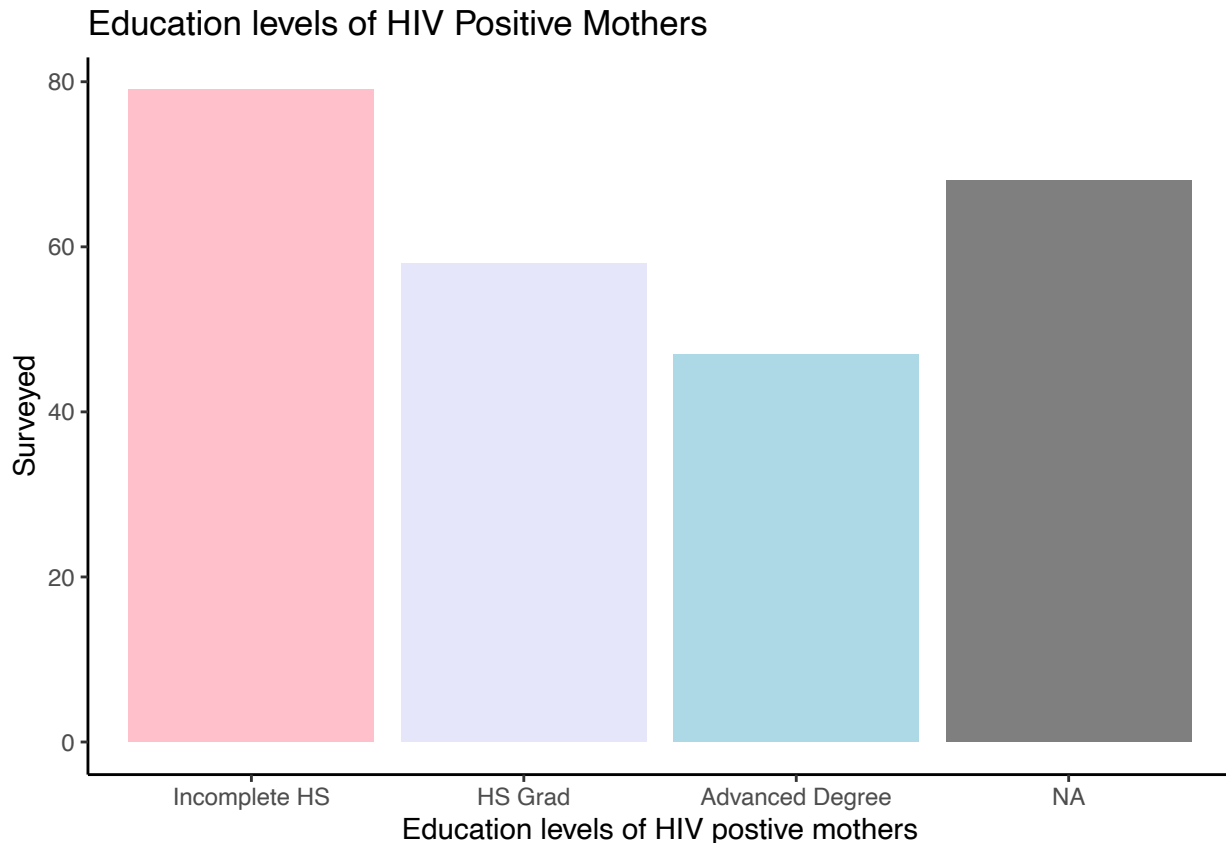
Next I want to see what the education levels of HIV positive mothers are like. HIV is more common in communities without adequate health education so I think that the education levels of the mothers will fall on the lower end.

```
par.HIV$EDUMO <- factor(par.HIV$EDUMO, labels=c("Incomplete HS", "HS Grad", "Advanced Degree"))  
summary(par.HIV$EDUMO)
```

```
##   Incomplete HS      HS Grad Advanced Degree      NA's  
##           79           58           47           68
```

```
ggplot(par.HIV, aes(x=EDUMO, fill=EDUMO)) + geom_bar() + xlab("Education levels of HIV positive mothers")
```

```
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please  
## use `guide = "none"` instead.
```



In this case my assumption seems to be correct. The rate of high school graduates among this data set comes to 41%, which is drastically lower than the national average of about 88% for women in the United States. There was also a high rate of mother's education rates listed as NA, this making up 26% of the data set. I can take this to mean that either there was insufficient data reported about 26% of the mothers, or 26% of the mothers did not reach high school. It is not clear which of these is the case.

Unilateral Data:

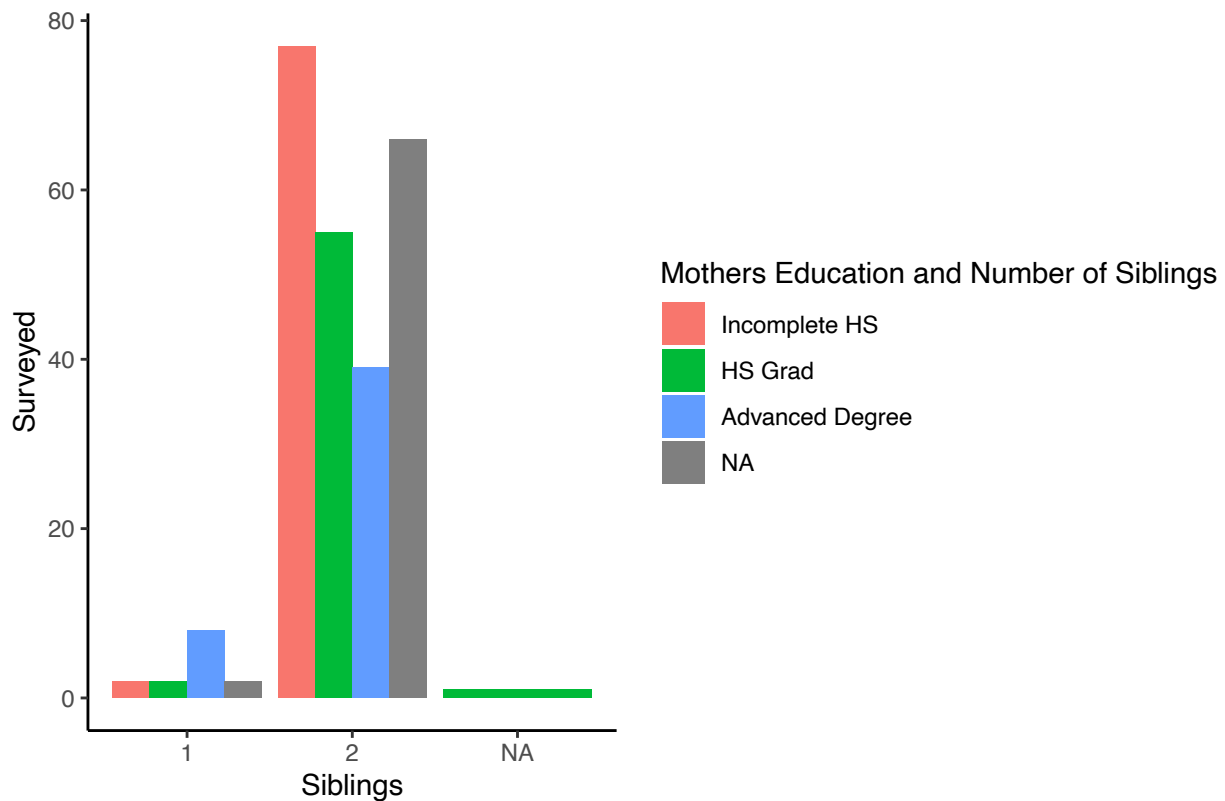
Finally I want to examine the relationship between the number of siblings children of HIV positive parents and the education of HIV positive mothers surveyed. My hypothesis is that there will be a correlation between children with more siblings and mothers with a lower education levels, as this would hold true to general statistical trends.

```
table(par.HIV$SIBLINGS, par.HIV$EDUMO) %>% prop.table(margin=1) %>% round(3)
```

```
##
##      Incomplete HS HS Grad Advanced Degree
## 1          0.167  0.167          0.667
## 2          0.450  0.322          0.228
```

```
ggplot(par.HIV, aes(x=SIBLINGS, fill=EDUMO)) + geom_bar(position = "dodge") + xlab("Siblings") + ylab("Su
```

Number of Siblings in children with HIV positive parents



The combination of this data implies that my theory was correct about women with advanced degrees having few children. However this could also be skewed due to the fact that the majority of children surveyed had 2 or more siblings.

Conclusion: The data in this set was interesting to study. It was predictable in some ways, but surprising in other ways. Looking at this data set makes it clear that many of these variables effect each other. I believe that this data also ties into the larger population by providing insights into communities and families affected by HIV and the disperities that they may face.