Data Project

Natalie Erazo

2023-09-18

```
parHIV <- read.delim("/Users/nataliee/Desktop/Math130/Data/PARHIV_data.txt", header=TRUE, stringsAsFact
dim(parHIV)
```

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

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

Introduction- Parental HIV

For this exploratory data analysis project I will be taking a look at Parental HIV. The data comes from Dr.Mary Jane Rotheram-Borus, Professor of Psychology and Behavioral Sciences at the University of California, Los Angeles. The dataset "parHIV" contains data related to Parental HIV. We will explore two or three variables to gain insights. My research question is to understand the relationships and patterns within these variables.The variables I will be examining in this study are Ethnicity, crime and neighborhood crime rate.

#Univariate Exploration

```
data<- read.csv("/Users/nataliee/Desktop/Math130/ParHIVFinalProject.html")
```

```
parHIV$ETHNrename <- factor(parHIV$ETHN, labels=c( "Latino(a)or Hispanic", "Black","Other"))
```

```
summary(parHIV$ETHNrename)
```

```
## Latino(a)or Hispanic              Black              Other
##               123                  96                 33
```

#Create a barchart for Ethnicity

```
ggplot(parHIV, aes(x=ETHNrename, fill=ETHNrename)) + geom_bar() + xlab("Ethnicity") + scale_fill_brewer
"Set1", guide=FALSE) + ylab("Surveyed") +ggtitle("Those in families with an HIV positive parent")
```

```
## Warning: The 'guide' argument in 'scale_*()' cannot be 'FALSE'. This was deprecated in
## ggplot2 3.3.4.
## i Please use "none" instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```
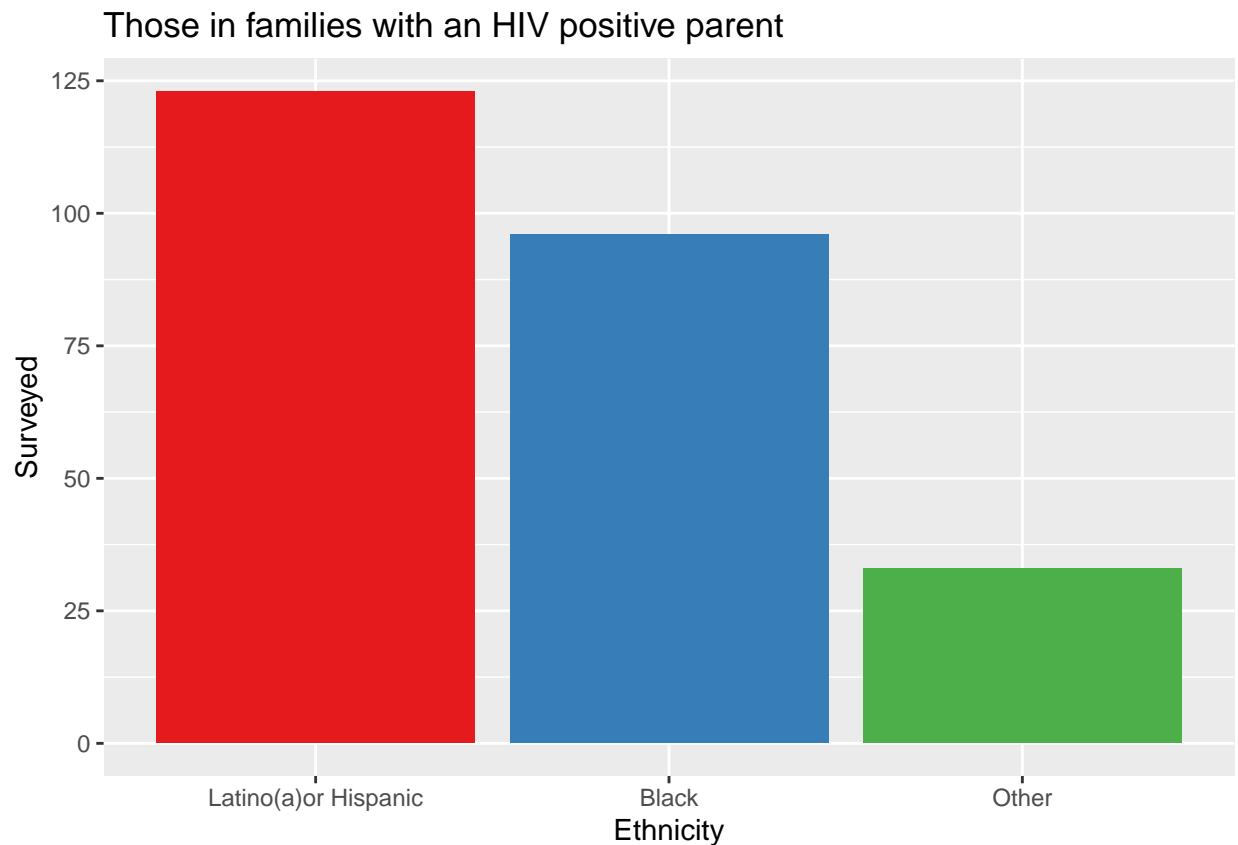

In this analysis, it is evident that individuals identifying as Latino(a)/Hispanic and Black ethnicities have a notably higher prevalence of HIV compared to individuals in other ethnic categories. Specifically, the "Latino(a)/Hispanic" group had the highest number of individuals with HIV-positive parents, totaling 123 individuals, followed by the "Black" group with 96 individuals, and lastly, the "Other" group with 33 individuals. This observation aligns to some extent with the claims made by the CDC regarding the varying susceptibility to HIV based on ethnicity.

Moving forward, our analysis will shift focus to examining neighborhood-related factors.

```
parHIV$nbhrename <- factor(parHIV$NGHB4, labels=c( "No Impact", "Litte Impact","Big Impact", "Serious Im
```

```
summary(parHIV$nbhrename)
```

```
##      No Impact   Litte Impact    Big Impact Serious Impact
##             78             74            50             50
```

```
ggplot(parHIV, aes(x=nbhrename, fill=nbhrename)) + geom_bar() + xlab("Neighborhood Violence and Crime")
```


Impact of neighborhood violence of families with an HIV positive parent
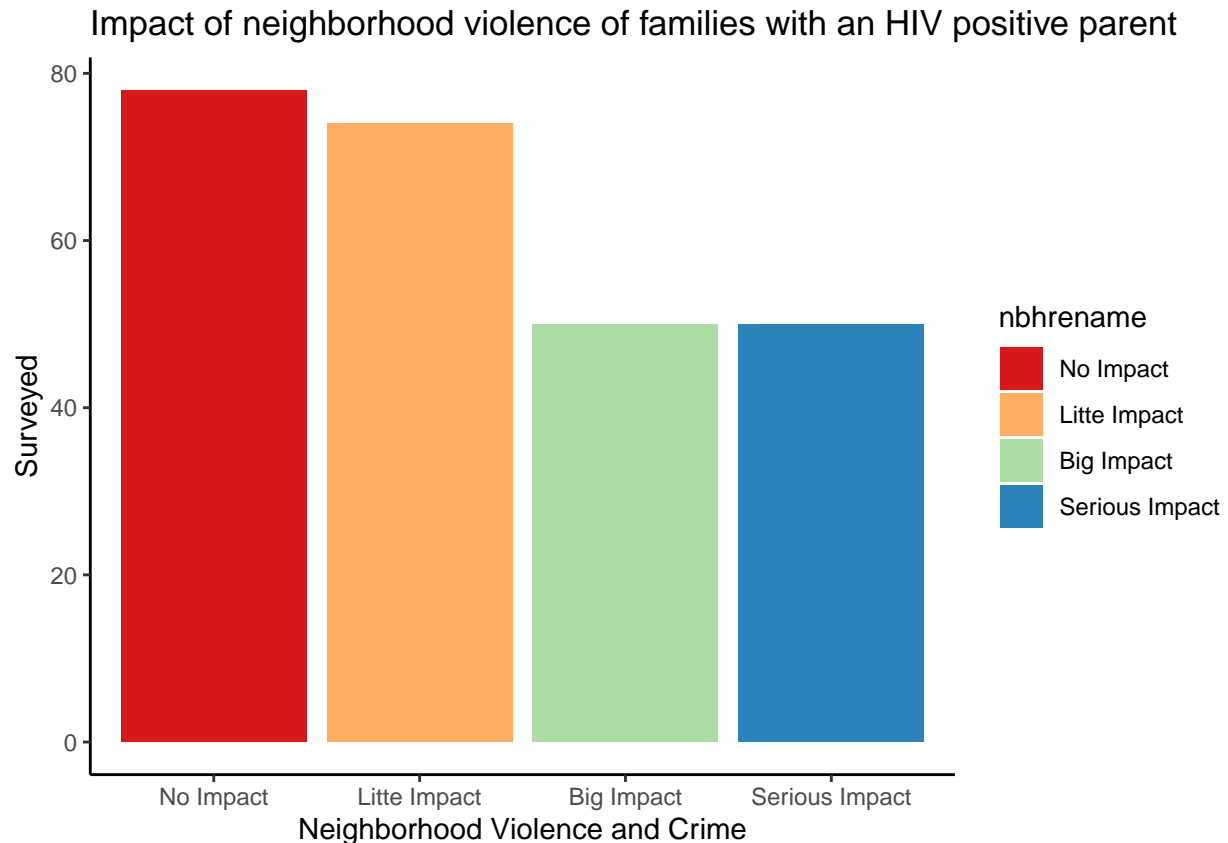
This answered my question if neighborhood violence had any correlation with families that have HIV positive parents.

## Bivariate Exploration

```
table(parHIV$ETHNrename, parHIV$nbhrename) %>% prop.table(margin=1) %>% round(3)
```

```
##
##                      No Impact Litte Impact Big Impact Serious Impact
##    Latino(a)or Hispanic   0.309        0.268      0.268          0.154
##    Black                  0.323        0.302      0.135          0.240
##    Other                  0.273        0.364      0.121          0.242
```

```
ggplot(parHIV, aes(x=ETHNrename, fill=nbhrename)) + xlab("Ethnicity") +ylab("Surveyed") + geom_bar(aes(y
```

```
## Warning: The dot-dot notation ('..count..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(count)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```


Survey responses of neighborhood crime rate according to ethnicity of HIV
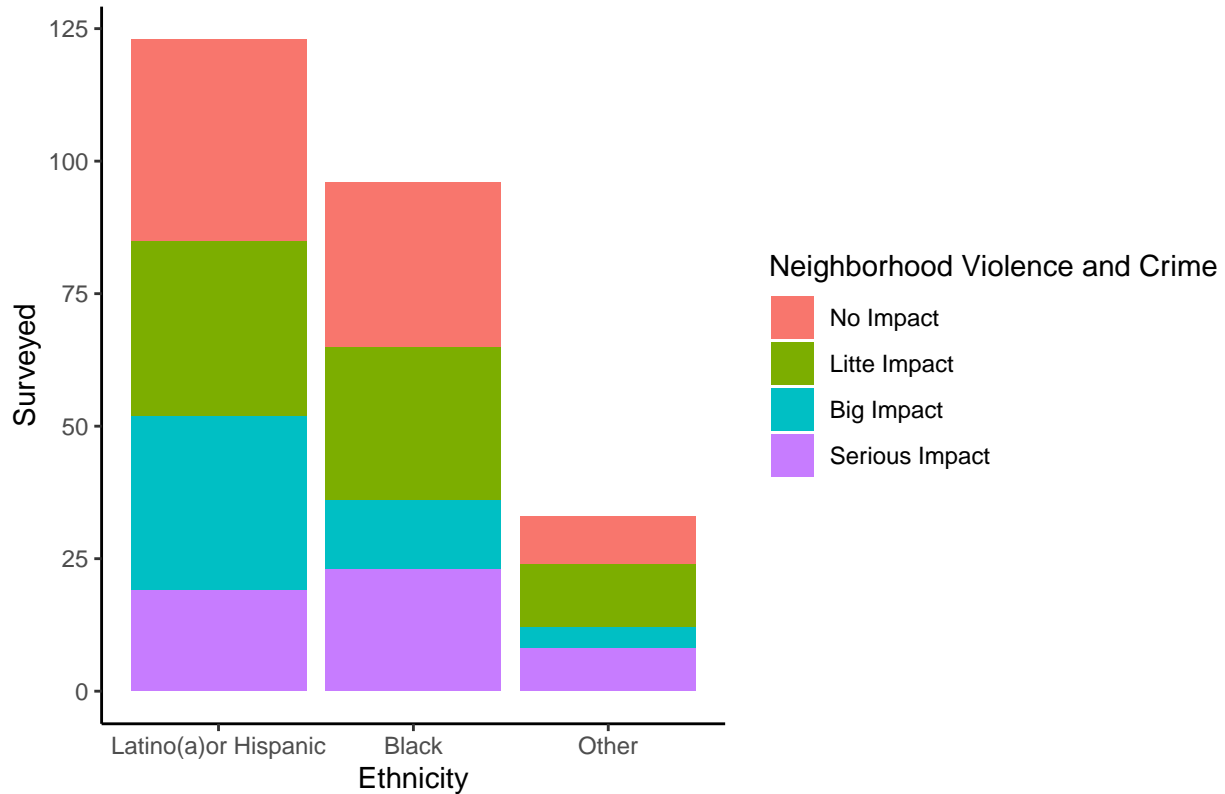
This visual representation offers a more distinct insight into the connection between the two categorical variables. It becomes apparent that both Latino(a)/Hispanic and Black ethnicities encounter either significant challenges related to crime and violence in their neighborhoods or none at all. It's noteworthy that the dataset for other ethnicities seems to be comparatively smaller. This observation might be attributed to the sampling size of the study or could potentially indicate variations in HIV susceptibility among different ethnic groups.

#Conclusion In conclusion,there are many variables that contribute to parental HIV and just other groups that are included.There are numerous variables within this dataset that I didn't have the opportunity to thoroughly explore, underlining the intricate nature of this disease.