EDA_stevenverschoor

stevenverschoor

2022-09-25

```
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(ggplot2)
library(forcats)
library(RColorBrewer)
library(sjPlot)
```

“Univariate Data analysis”

```
parentalhiv<-read.table("C:/Users/srverschoor/Desktop/math130/data/prentalHIV.txt",header=TRUE,s
ep="\t")
head(parentalhiv)
```

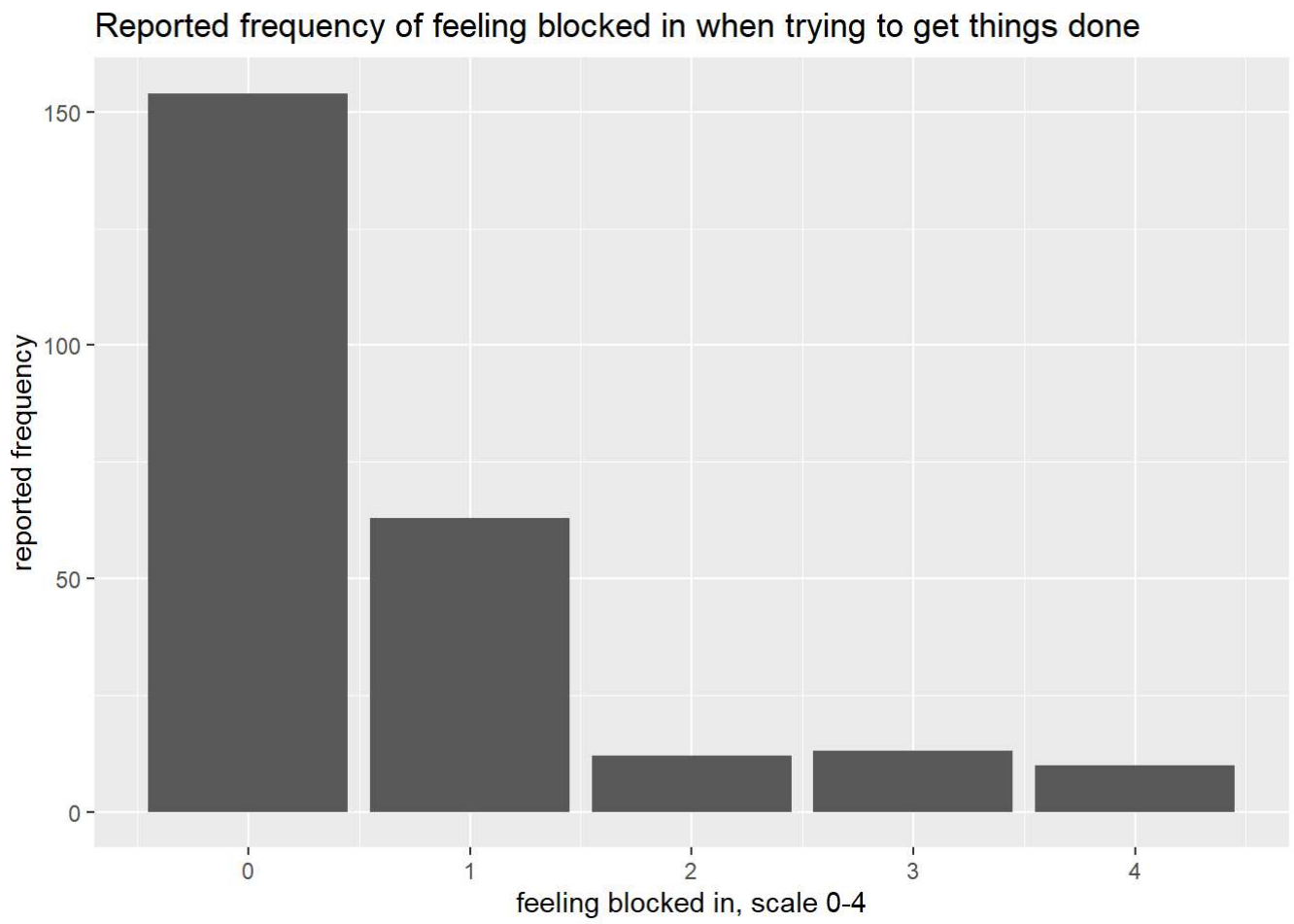
```
##   ID AGE GENDER LIVWITH SIBLINGS JOBMO EDUMO HOWREL ATTSERV NGHB1 NGHB2 NGHB3
## 1  1  17   Male       3        2    NA     2     NA       NA     2     2     2
## 2  2  18 Female       1       NA     1     2      2        2     4     2     3
## 3  3  13   Male       2        2     2     1     NA       NA     1     1     1
## 4  4  14 Female       2        2     2     3      3        1     3     3     2
## 5  5  14   Male       2        2     2     3      2        1     3     4     2
## 6  6  13   Male       2        2     2    NA      2        2     2     4     2
##   NGHB4 NGHB5 NGHB6 NGHB7 NGHB8 NGHB9 NGHB10 NGHB11 MONFOOD FINSIT ETHN
## 1     1     2     2     2     2     1      1      3       3      4    2
## 2     2     1     1     2     1     2      3      1       3      3    3
## 3     2     4     4     4     1     1      1      1       3      1    1
## 4     4     2     2     2     4     1      4      2       1      3    2
## 5     2     1     2     2     3     3      2      2       3      4    2
## 6     2     2     4     3     4     3      1      3       3      3    2
##   AGESMOKE SMOKEP3M AGEALC AGEMAR FRNDS SCHOOL LIKESCH HOOKEY NHOOKEY HMONTH
## 1       NA       NA     14     14     2      2       3      2       3      3
## 2       14        4     15     17     3      2       3      2       4      3
## 3       12        1      0      0     4      2       2      1       0      3
## 4       14        8     14     14     1      2       1      2       1      3
## 5       NA       NA      0      0     2      2       1      2       4      1
## 6       12        1      0      0     2      2       3      1       0      1
##   PB01 PB02 PB03 PB04 PB05 PB06 PB07 PB08 PB09 PB10 PB11 PB12 PB13 PB14 PB15
## 1    3    1    3    1    4    4    3    1    1    1    4    4    3    1    4
## 2    3    2    1    3    3    4    1    4    4    3    2    2    1    4    1
## 3    4    1    4    1    4    4    3    2    3    2    4    4    4    2    4
## 4    1    1    1    3    1    1    2    3    4    4    1    1    1    4    2
## 5    4    1    4    1    3    4    3    1    1    3    3    3    3    2    4
## 6    3    4    1    3    1    4    1    1    3    4    4    4    4    4    4
##   PB16 PB17 PB18 PB19 PB20 PB21 PB22 PB23 PB24 PB25 BSI01 BSI02 BSI03 BSI04
## 1    1    4    1    3    3    4    4    3    3    4     0     0     0     0
## 2    2    3    3    3    4    1    1    4    3    2     1     0     0     0
## 3    2    4    2    3    2    4    2    4    1    4     0     0     0     0
## 4    1    1    1    3    1    2    3    2    1    3     0     0     0     0
## 5    1    4    1    1    1    4    3    1    4    3     0     1     0     0
## 6    1    4    1    3    3    2    1    4    1    4     0     0     0     1
##   BSI05 BSI06 BSI07 BSI08 BSI09 BSI10 BSI11 BSI12 BSI13 BSI14 BSI15 BSI16 BSI17
## 1     0     0     0     0     0     1     0     0     3     2     0     0     0
## 2     0     0     0     0     0     0     0     0     0     3     0     3     2
## 3     0     0     0     0     0     0     0     0     0     0     0     0     0
## 4     0     1     1     0     0     0     0     0     1     0     0     0     0
## 5     0     1     0     0     0     0     0     0     0     0     0     0     1
## 6     2     1     0     1     0     2     0     0     2     0     1     1     0
##   BSI18 BSI19 BSI20 BSI21 BSI22 BSI23 BSI24 BSI25 BSI26 BSI27 BSI28 BSI29 BSI30
## 1     0     0     2     0     1     0     2     0     0     1     0     0     0
## 2     0     0     2     0     0     0     0     4     0     0     0     0     0
## 3     0     0     0     0     0     0     0     0     0     0     0     0     0
## 4     0     0     0     0     0     1     0     0     0     0     1     1     0
## 5     0     0     0     0     0     0     1     0     1     0     0     0     0
## 6     3     0     0     0     0     0     1     1     1     1     1     0     0
##   BSI31 BSI32 BSI33 BSI34 BSI35 BSI36 BSI37 BSI38 BSI39 BSI40 BSI41 BSI42 BSI43
## 1     0     0     0     1     0     2     0     0     0     2     2     0     1
## 2     0     0     0     0     0     0     0     2     0     0     0     0     0
```

```
## 3     0     0     0     0     0     0     0     0     0     0     0     0     0
## 4     0     0     0     1     0     0     1     0     0     0     1     0     0
## 5     0     0     1     0     0     0     1     0     0     1     0     1     1
## 6     0     0     1     2     2     1     1     2     1     1     2     1     2
##   BSI44 BSI45 BSI46 BSI47 BSI48 BSI49 BSI50 BSI51 BSI52 BSI53 parent_care
## 1     0     0     3     0     0     0     0     2     0     0    3.750000
## 2     0     0     0     0     0     4     0     1     0     0    2.500000
## 3     0     0     0     0     0     0     0     0     0     0    3.750000
## 4     0     0     1     0     1     0     0     1     0     0    2.083333
## 5     0     0     3     0     0     0     0     0     0     0    3.416667
## 6     1     1     1     0     2     2     1     1     2     1    3.000000
##   parent_overprotection BSI_overall BSI_somat BSI_obcomp BSI_interp BSI_depress
## 1              1.769231   0.4716981 0.0000000  0.5000000       1.00   0.0000000
## 2              3.538462   0.4150943 0.0000000  0.0000000       0.75   0.8333333
## 3              2.230769   0.0000000 0.0000000  0.0000000       0.00   0.0000000
## 4              2.692308   0.2264151 0.5714286  0.0000000       0.00   0.0000000
## 5              1.538462   0.2452830 0.4285714  0.1666667       0.50   0.1666667
## 6              3.000000   0.8867925 0.2857143  1.0000000       0.50   1.1666667
##   BSI_anxiety BSI_hostil BSI_phobic BSI_paranoid BSI_psycho
## 1   0.0000000        2.0        0.2          1.0        0.6
## 2   1.1666667        0.0        0.0          0.2        0.6
## 3   0.0000000        0.0        0.0          0.0        0.0
## 4   0.0000000        0.8        0.2          0.4        0.2
## 5   0.0000000        1.0        0.2          0.2        0.0
## 6   0.8333333        1.4        0.8          1.4        0.8
```

#“An analysis of three variables, two ordinal and one continuous, is as follows. The data set, titled”Parental HIV”, used is a data set collected by Dr. Mary Jane Rotheram-Borus during a clinical trial at the neuropsychiatric institute of UCLA that evaluated the efficacy of different behavioral interventions for two hundred and fifty two adolescent children who have a parent with HIV, and it is used with permission in conjunction with the textbook Practical Multivariate Analysis by Afifi et. al.

#“The BSI15 variable within the Parental HIv data set is an ordinal variable which describes the degree to which children who have a parent with HIV feel blocked in getting things done on a scale of zero to four, with zero denoting ‘not at all’, one denoting ‘a little bit’, two denoting ‘moderately’, three denoting ‘quite a bit’, and four denoting ‘extremely’.

```
ggplot(parentalhiv,aes(x=BSI15))+geom_bar()+xlab("feeling blocked in, scale 0-4")+ylab("reported
frequency")+ggtitle("Reported frequency of feeling blocked in when trying to get things done")
```




```
summary(parentalhiv$BSI15)
```

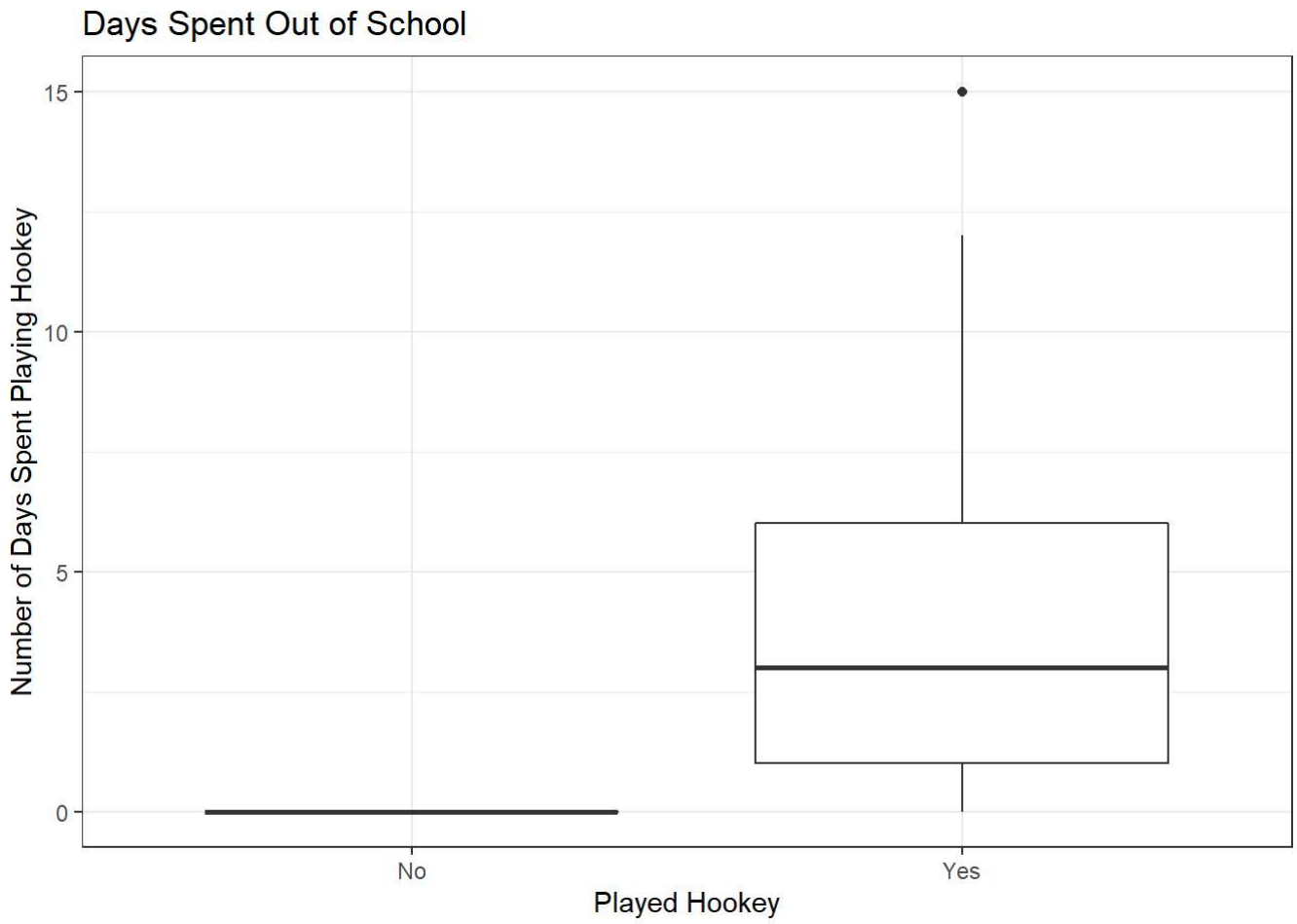
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.6587  1.0000  4.0000
```

#“THe NHOOKEY variable within the Parental HIV data set is a continuous variable which describes the number of days that children who have a parent with HIV have skipped school without a reason, or ‘played hookey’. THe HOOKEY variable, which is a binary variable, will be used as a yes/no reply on the x axis in order to separate those children who did or did not play hookey.”

```
parentalhiv$hookey<-factor(parentalhiv$HOOKEY,labels=c("No","Yes"))
table(parentalhiv$hookey)
```

```
##
##  No Yes
## 103 149
```

```
ggplot(parentalhiv,aes (y=NHOOKEY,x=hookey,group=hookey))+geom_boxplot()+ggtitle("Days Spent Out of School")+xlab("Played Hookey")+ylab("Number of Days Spent Playing Hookey")+theme_bw()
```




```
summary(parentalhiv$NHOOKEY)
```

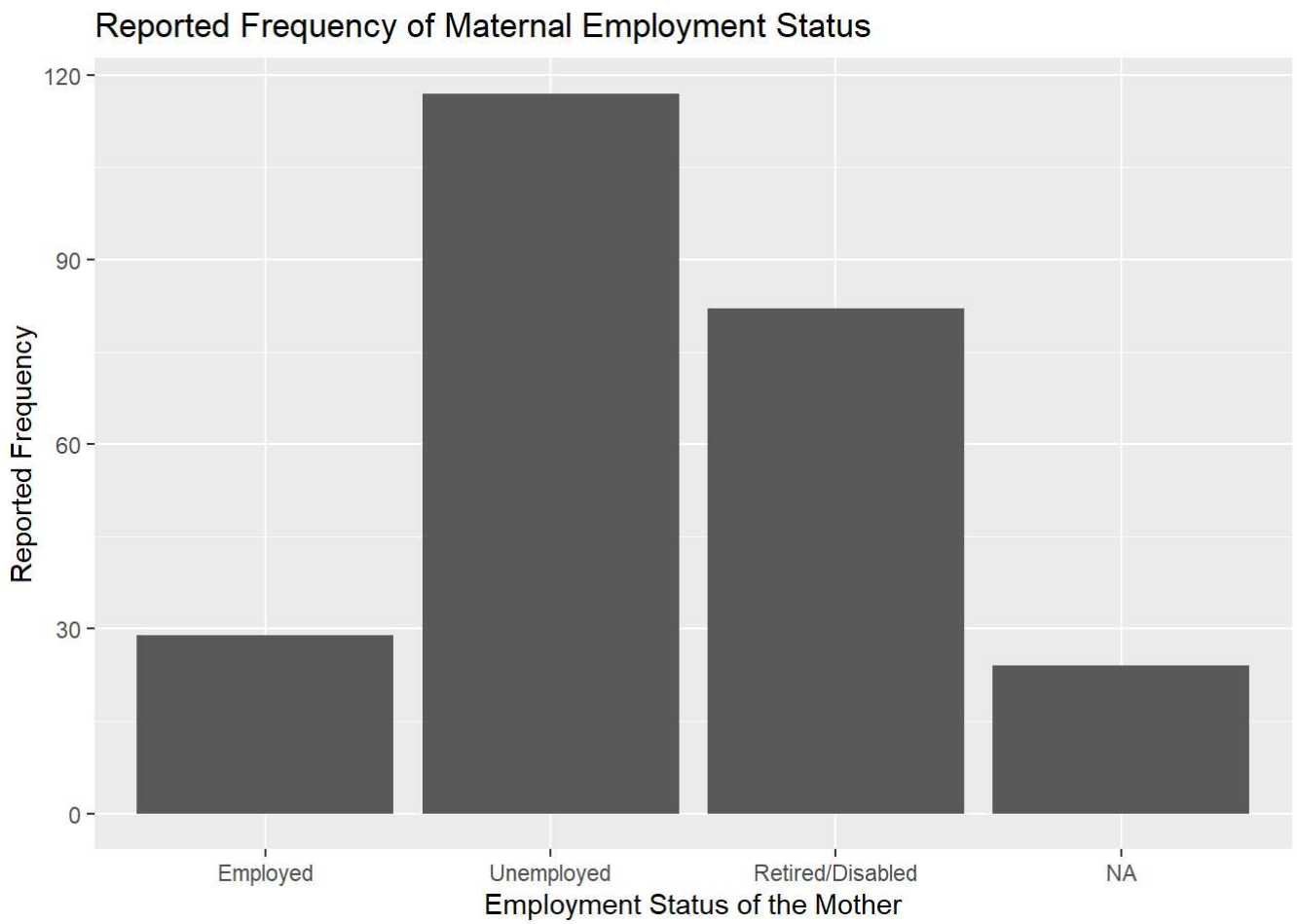
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    1.00    2.27    4.00   15.00
```

#“The JOBMO variable within the Parental HIV data set is an ordinal variable which describes the employment status of the mothers of children who have a parent with HIV. THe JOBMO variable has three possible states of employment, which are unemployed, employed, or retired/disabled.

```
parentalhiv$jobmo<-factor(parentalhiv$JOBMO,labels=c("Employed","Unemployed","Retired/Disabled"
))
table(parentalhiv$jobmo)
```

```
##
##         Employed       Unemployed Retired/Disabled
##               29              117               82
```

```
ggplot(parentalhiv,aes(x=jobmo))+geom_bar()+xlab("Employment Status of the Mother")+ylab("Report
ed Frequency")+ggtitle("Reported Frequency of Maternal Employment Status")
```



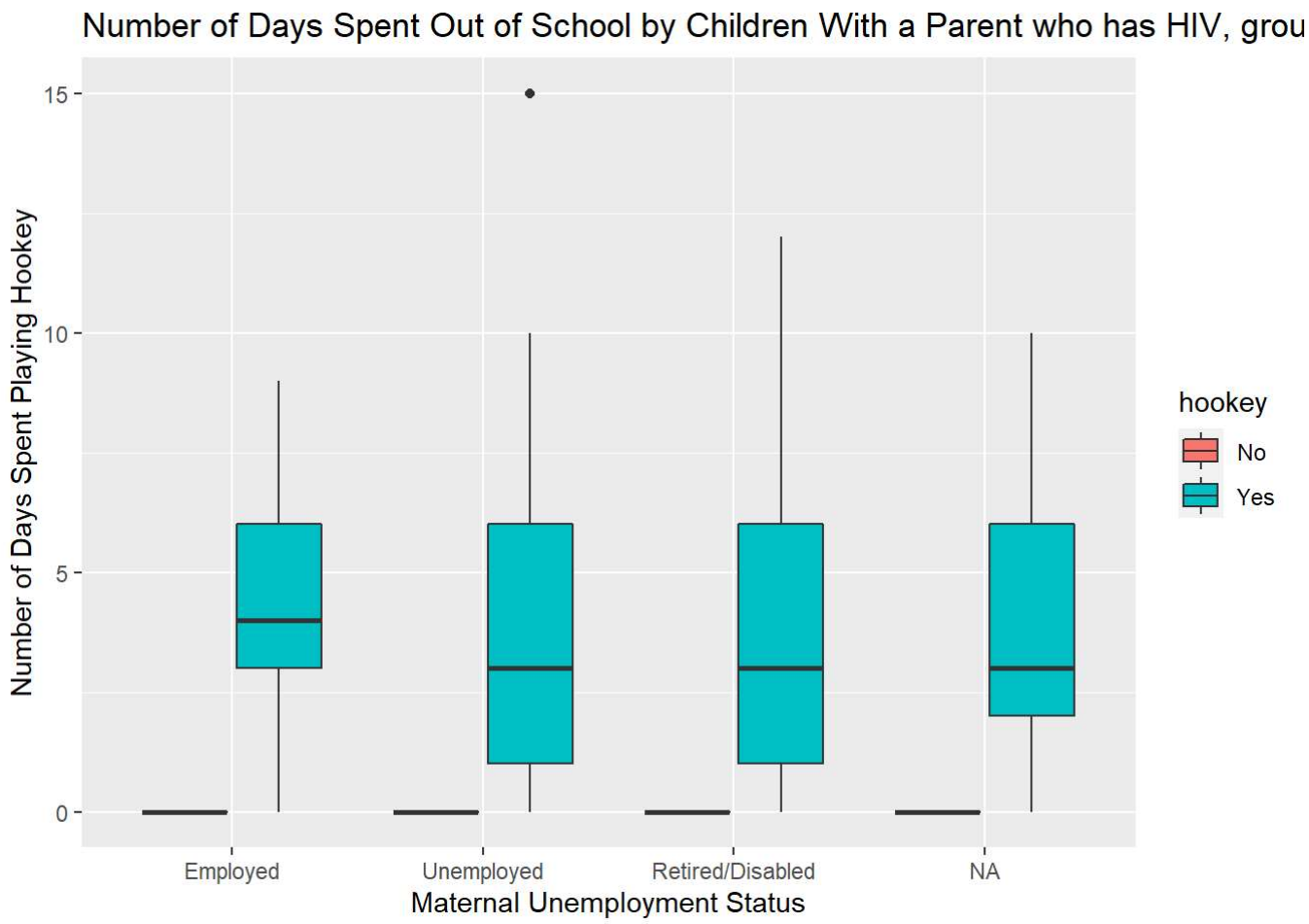

```
summary(parentalhiv$jobmo)
```

```
##         Employed       Unemployed Retired/Disabled             NA's
##               29              117               82               24
```

##"Bivariate Analysis"

#"A graph of the intersect between maternal unemployment and the amount of days spent out of school is as follows…"

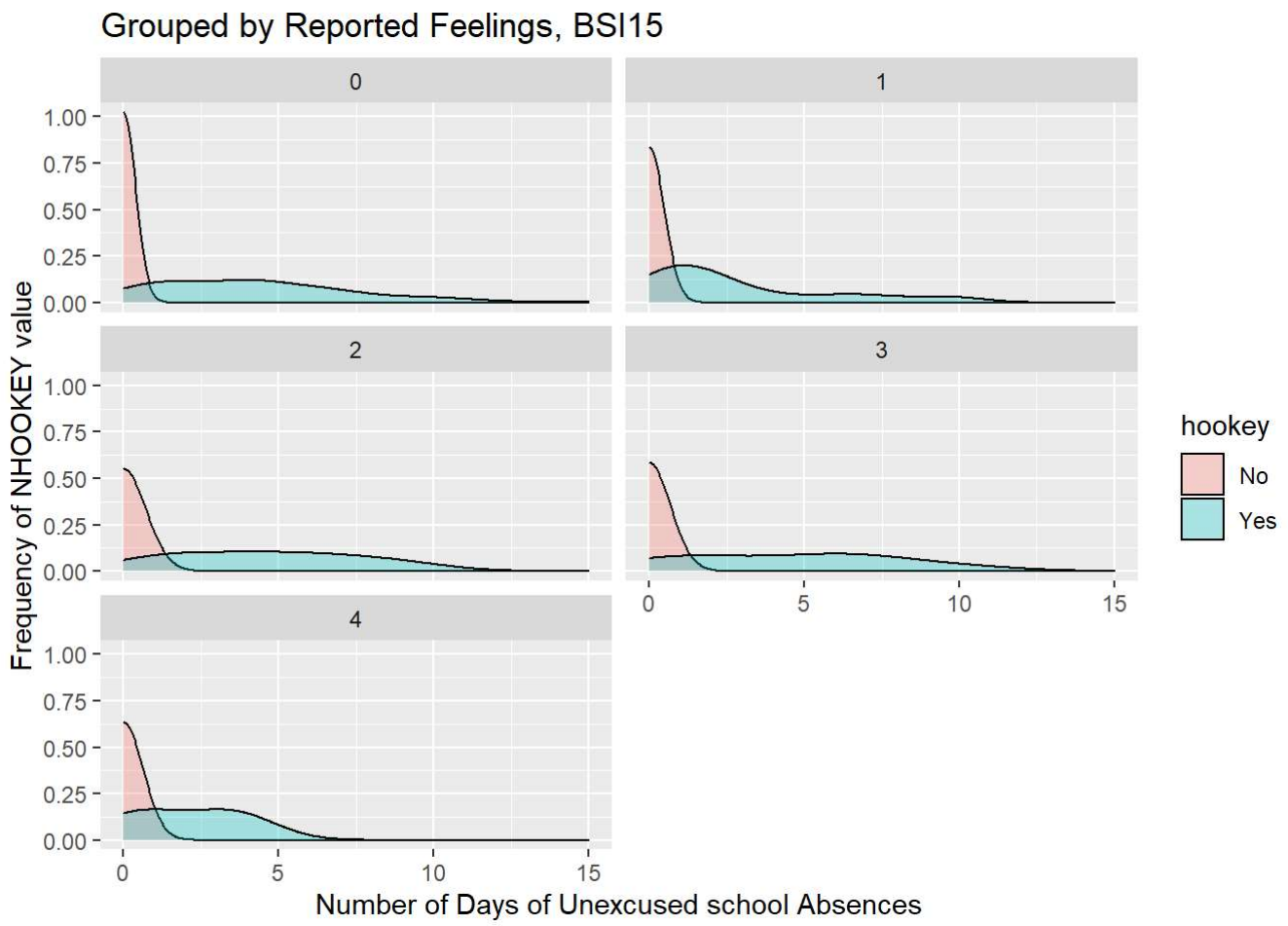
```
ggplot(parentalhiv,aes(x=jobmo,y=NHOOKEY,fill=hookey))+geom_boxplot()+xlab("Maternal Unemploymen
t Status")+ylab("Number of Days Spent Playing Hookey")+ggtitle("Number of Days Spent Out of Scho
ol by Children With a Parent who has HIV, grouped by Reported Maternal Employment Status.")
```



#"As can be seen in the graph of maternal employment status, The averages are all fairly close together, with the third quartiles all topping out at the same number of days spent playing hookey. The singular large outlier in the NHOOKEY variable of 15 can be seen in the 'unemployed' bin, though."

##"Next, a graph of the intersect between the number of days spent out of school and feelings of being being blocked in when trying to get something done…"

```
ggplot(parentalhiv,aes(x=NHOOKEY,fill=hookey))+geom_density(alpha=0.3)+facet_wrap(~BSI15,ncol=2)
+xlab("Number of Days of Unexcused school Absences")+ylab("Frequency of NHOOKEY value")+ggtitle(
"Grouped by Reported Feelings, BSI15")
```



#“When grouped by the reported BSI15 values of 0-4, it can be seen that no matter what the response for BSI15 was, there are children who both did and did not skip school in all five BSI15 groups. Surprisingly, the spread in response group number four for BSI15, those who felt ‘extremely’ blocked in when trying to get something done, actually ended up having the tightest grouping towards the lower end of days spent truant even when they were truant.”