130Wk5

Tyler Burke

2024-09-26

Introduction: Parental HIV: For this data analysis project I am looking at the Parental HIV data set. The dataset contains data relevant to Parental HIV. I will explore three variables to gain insight. I aim to gain an understanding of the relationships within the variables of who they Live With, their Financial Situation, and the number of days playing Hookey from school.

```
file_path <- "/Users/tylerburke/Desktop/Math130/130Wk5/Parhiv.txt"
HIVdata <- read.table(file_path, header = TRUE, sep = "\t", stringsAsFactors = FALSE)
HIVselect  <- HIVdata %>% select(LIVWITH, FINSIT, NHOOKEY)
```

Living Conditions

```
living_summary <- HIVselect %>%
  group_by(LIVWITH) %>%
  summarise(Count = n())
ggplot(HIVselect, aes(x = as.factor(LIVWITH))) +
  geom_bar(fill = "lightblue") +
  labs(title = "Distribution of Living Conditions",
       x = "Living Conditions",
       y = "Count",
       caption = "1 = Both parents, 2 = One parent, 3 = Other")
```
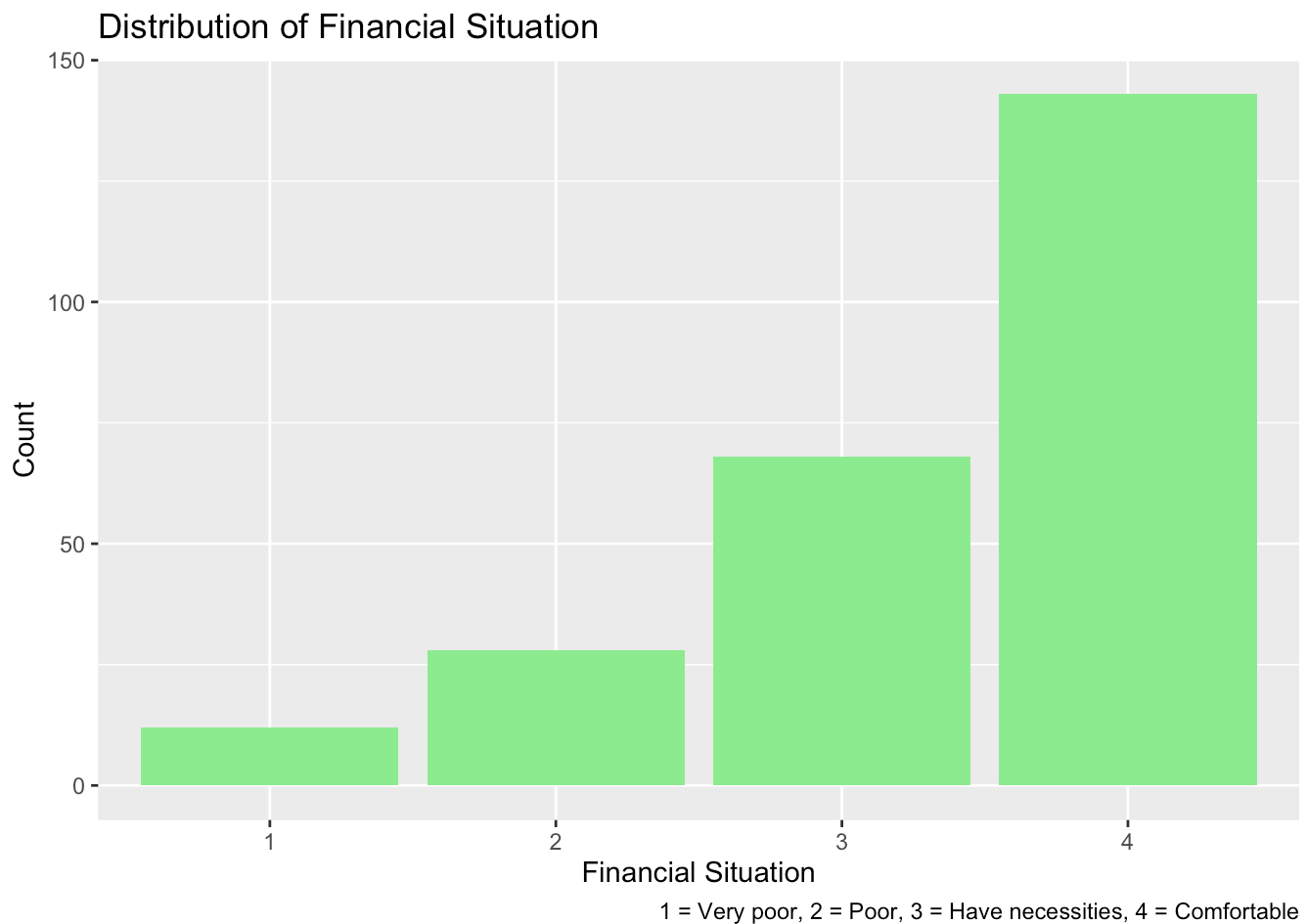
Distribution of Living Conditions

1 = Both parents, 2 = One parent, 3 = Other

This analysis shows us that out of the sample the majority were currently living with one parent. Followed by both parents and other.

Financial Situation

```
finsit_summary <- HIVselect %>%
  group_by(LIVWITH) %>%
  summarise(Count = n())
ggplot(HIVselect, aes(x = as.factor(FINSIT))) +
  geom_bar(fill = "lightgreen") +
  labs(title = "Distribution of Financial Situation",
       x = "Financial Situation",
       y = "Count",
       caption = "1 = Very poor, 2 = Poor, 3 = Have necessities, 4 = Comfortable")
```

Distribution of Financial Situation

Financial Situation

1 = Very poor, 2 = Poor, 3 = Have necessities, 4 = Comfortable

This analysis shows us that out of the sample the majority were currently living comfortably. Followed by have necessities, poor and very poor.
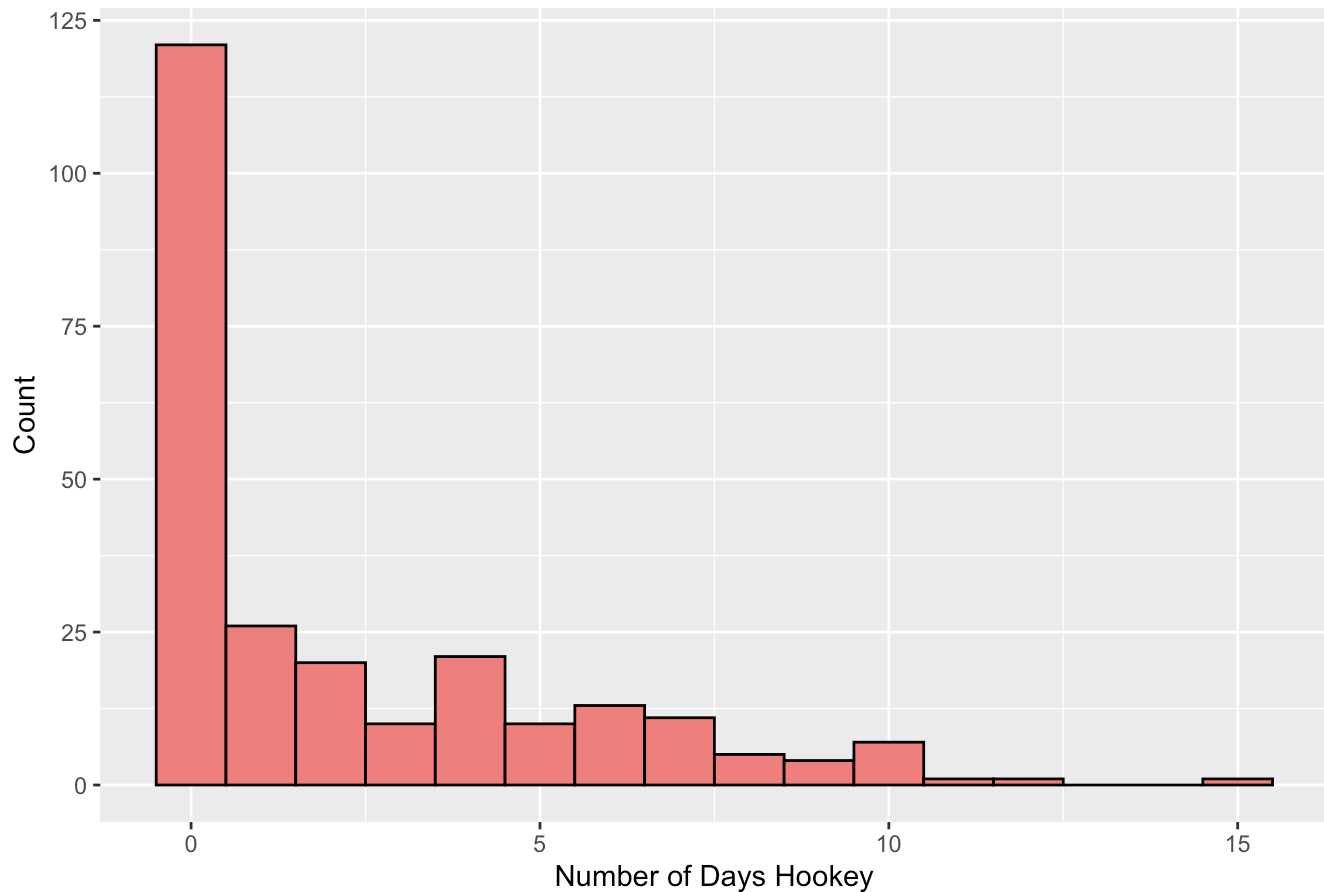
Hookey

```
hookey_summary <- HIVselect %>%
  group_by(LIVWITH) %>%
  summarise(Count = n())
ggplot(HIVselect, aes(x = NHOOKEY)) +
  geom_histogram(binwidth = 1, fill = "lightcoral", color = "black") +
  labs(title = "Distribution of Number of Days Hookey",
       x = "Number of Days Hookey",
       y = "Count")
```
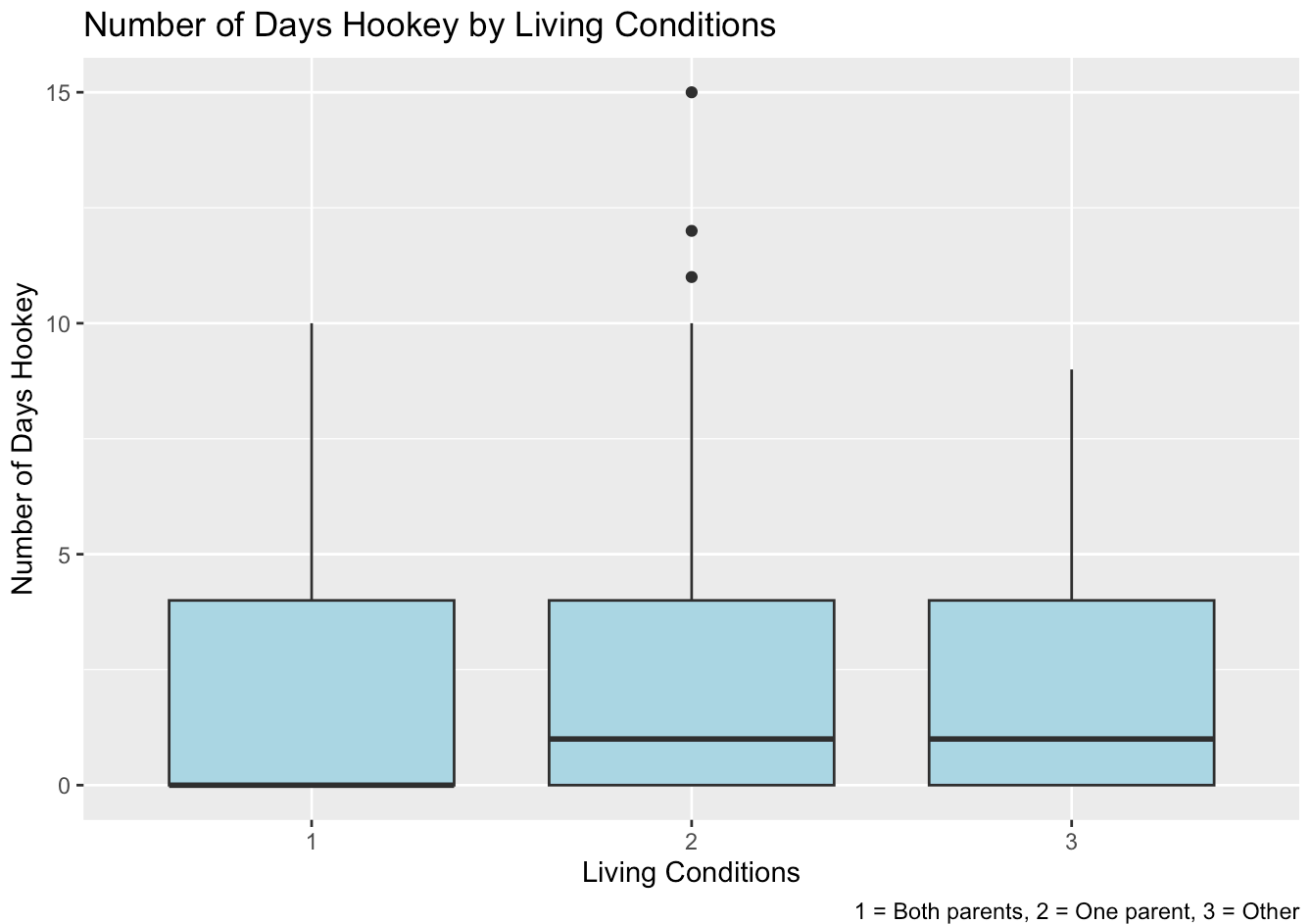
Distribution of Number of Days Hookey

This analysis shows us that out of the sample the majority of kids had not played hookey. There is a general downwards trend with the maximum number of days being 15.

Living situation vs Hookey

```
living_vs_hookey <- HIVselect %>%
  group_by(LIVWITH) %>%
  summarise(mean_hookey = mean(NHOOKEY, na.rm = TRUE),
            sd_hookey = sd(NHOOKEY, na.rm = TRUE))
ggplot(HIVselect, aes(x = as.factor(LIVWITH), y = NHOOKEY)) +
  geom_boxplot(fill = "lightblue") +
  labs(title = "Number of Days Hookey by Living Conditions",
       x = "Living Conditions",
       y = "Number of Days Hookey",
       caption = "1 = Both parents, 2 = One parent, 3 = Other")
```

Number of Days Hookey by Living Conditions

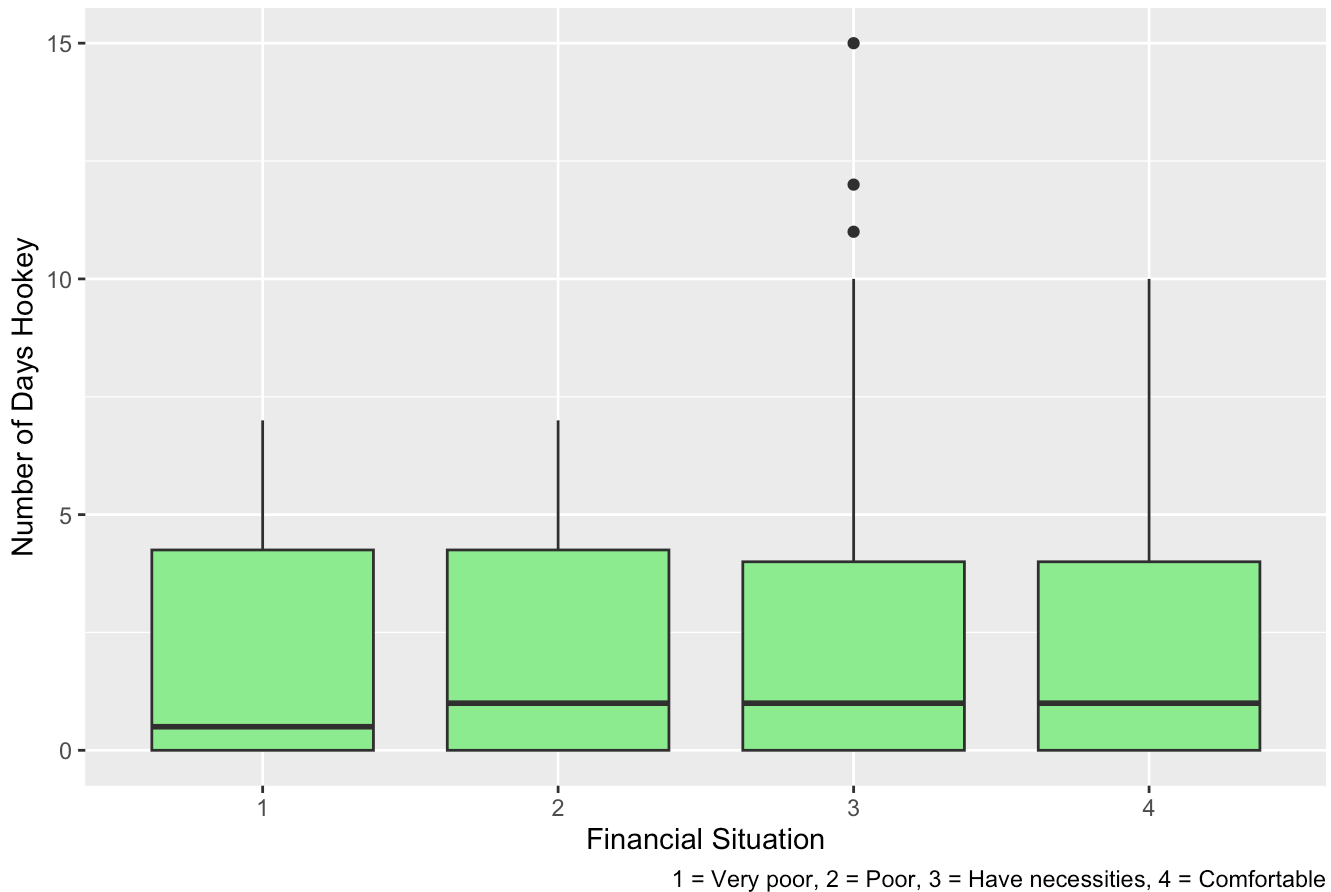1 = Both parents, 2 = One parent, 3 = Other

This analysis of the data shows us that one parent or other lead to greater days playing hookey. With outliers in the one parent category. Both parents had no days playing hookey.

Financial Situation vs Hookey

```
financial_vs_hookey <- HIVselect %>%
  group_by(FINSIT) %>%
  summarise(mean_hookey = mean(NHOOKEY, na.rm = TRUE),
            sd_hookey = sd(NHOOKEY, na.rm = TRUE))
ggplot(HIVselect, aes(x = as.factor(FINSIT), y = NHOOKEY)) +
  geom_boxplot(fill = "lightgreen") +
  labs(title = "Number of Days Hookey by Financial Situation",
       x = "Financial Situation",
       y = "Number of Days Hookey",
       caption = "1 = Very poor, 2 = Poor, 3 = Have necessities, 4 = Comfortable")
```

Number of Days Hookey by Financial Situation

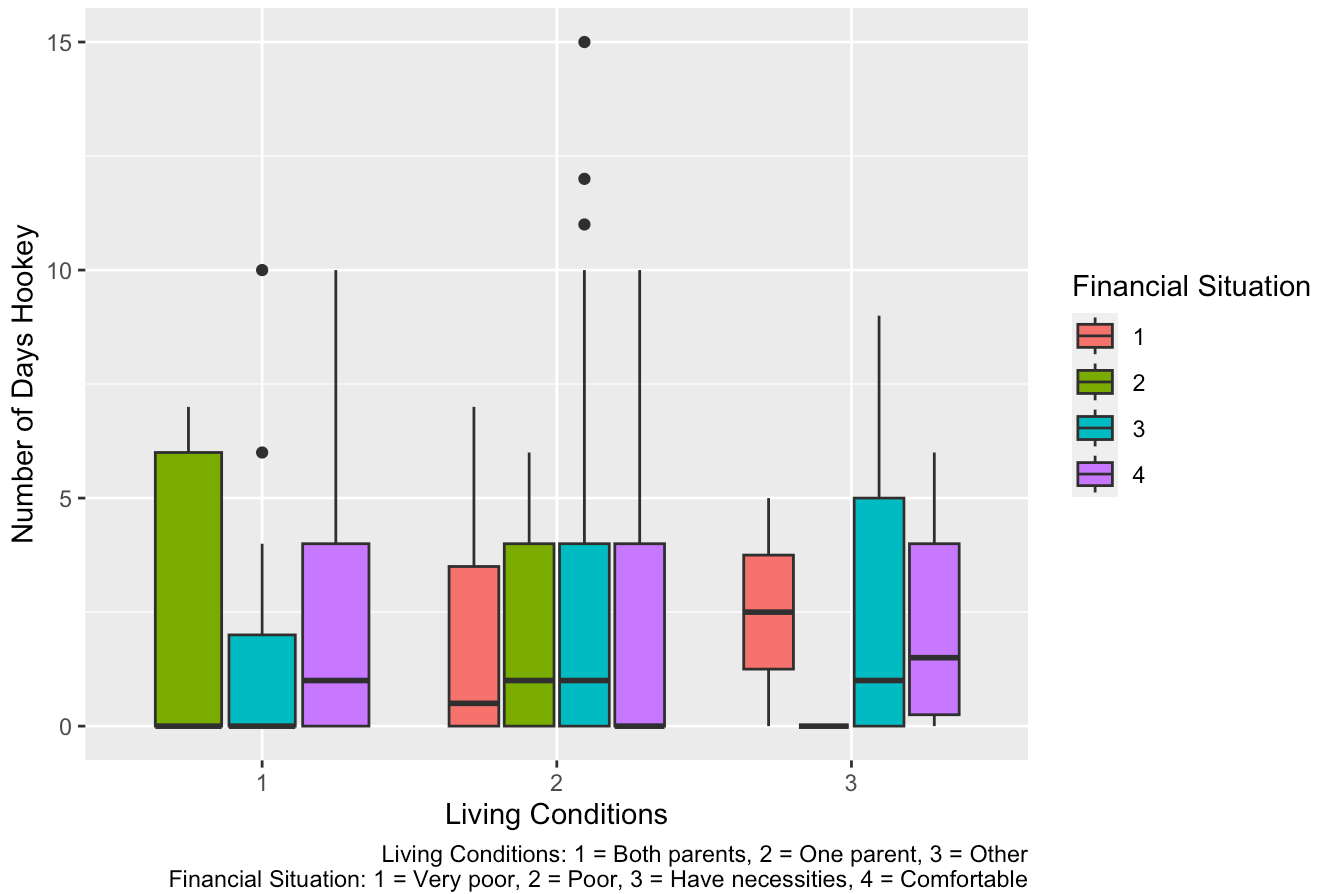1 = Very poor, 2 = Poor, 3 = Have necessities, 4 = Comfortable

This analysis of the data shows us that having necessities being comfortable and being poor lead to similar mean of days playing hookey, while the have necessities category had greater outliers. The very poor category had the fewest days playing hookey.

Bivariate exploration

```
ggplot(HIVselect, aes(x = as.factor(LIVWITH), y = NHOOKEY, fill = as.factor(FINSIT))) +
  geom_boxplot() +
  labs(title = "Number of Days Hookey by Living Conditions and Financial Situation",
       x = "Living Conditions",
       y = "Number of Days Hookey",
       fill = "Financial Situation",
       caption = "Living Conditions: 1 = Both parents, 2 = One parent, 3 = Other\nFinanc
ial Situation: 1 = Very poor, 2 = Poor, 3 = Have necessities, 4 = Comfortable")
```

Number of Days Hookey by Living Conditions and Financial Situation

Living Conditions: 1 = Both parents, 2 = One parent, 3 = Other
Financial Situation: 1 = Very poor, 2 = Poor, 3 = Have necessities, 4 = Comfortable

This analysis of the data shows us that adolescents from comfortable financial systems have a tendency to play hookey more often regardless of living situation. Adolescents living with both parents have generally lower variability in the number of hookey days. Adolescents living with one parent show a wide range of hookey days across all financial situations. And for those living with someone other than parents there seems to be a higher range of hookey day particularly in the comfortable and have necessities financial situations.