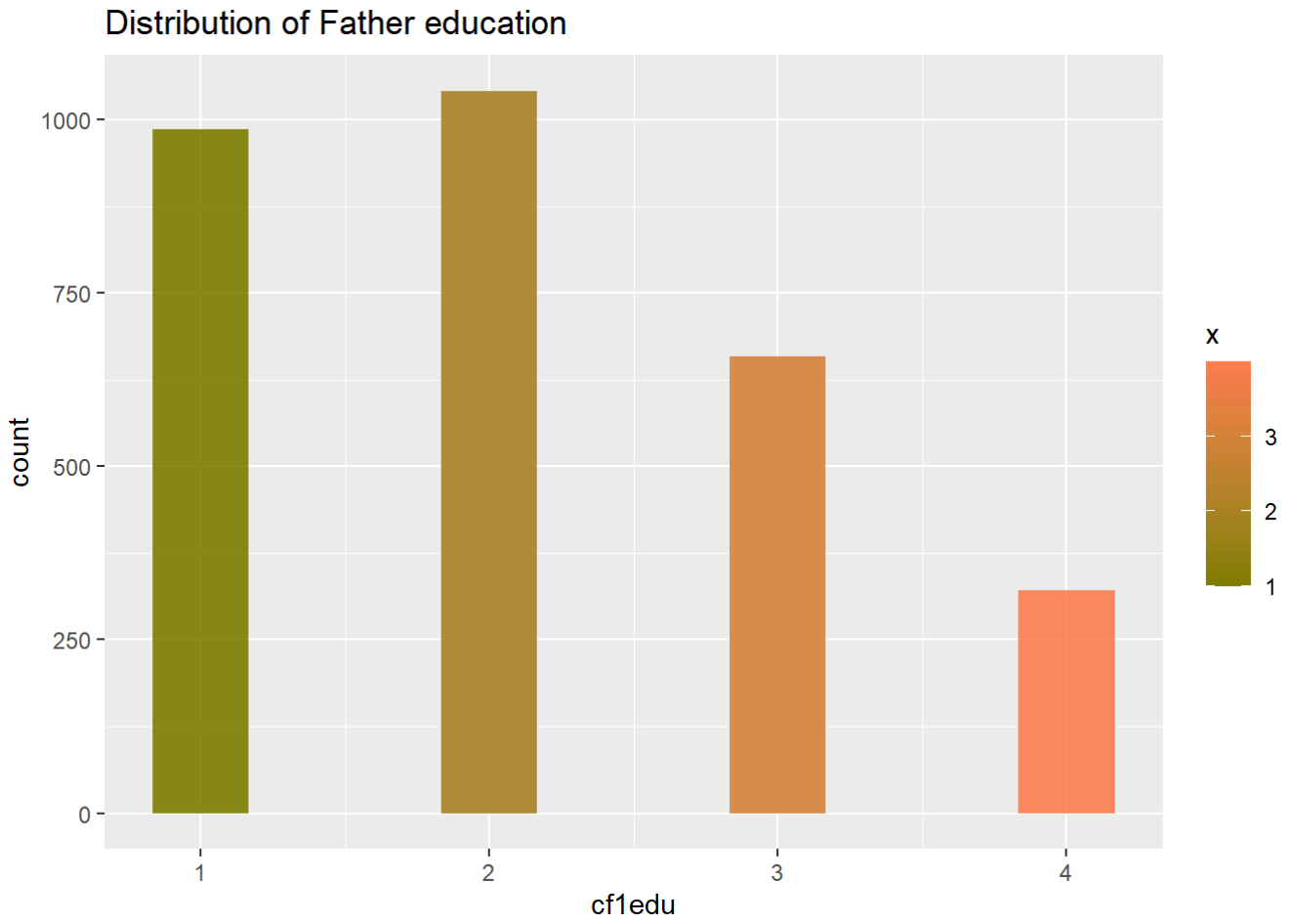
```
library(haven)
library(ggplot2)
data <- read_dta("D:/BrownUnivercity/DATA2020/Data2020-Final-Project/data/ff_data_x_preprocessed_v1.dta")
print(data)
```

```
## # A tibble: 3,113 × 15
##    index cf1edu cm1edu f2b32  m2d2 m2b43a cf3marm cf3kids cf3md_case_lib cf4cohm
##    <dbl>  <dbl>  <dbl> <dbl> <dbl>  <dbl>   <dbl>   <dbl>          <dbl>   <dbl>
##  1     0      3      3    -6     1     -6       0       0              0       0
##  2     1      1      1     1     1     -6      -9      -9             -9       0
##  3     2      3      3     2     1     -6       0       2              0       0
##  4     3      2      2     1     1     -6       0       2              0       0
##  5     5      2      2    -9    -6     -6      -9      -9             -9      -9
##  6     7      2      1     1     1     -6       1       3              0       0
##  7     8      1      3    -6     1     -6       0       1              1       0
##  8     9      2      3    -9    -9     -9      -9      -9             -9      -9
##  9    10      2      3    -9     1     -6       1       2              1       0
## 10    11      2      3     3     2     -6       0       2              1       0
## # ℹ 3,103 more rows
## # ℹ 5 more variables: t4d7 <dbl>, cf1hhinc <dbl>, k5d1f <dbl>, k5f1f <dbl>,
## #   y_binary <dbl>
```

EDA 1-2.cf1edu: Father baseline education; cm1edu: Mother baseline education; 1-low, 4-high

```
# Filter the data to exclude rows where cf1edu equals -3
filtered_data <- data[data$cf1edu != -3, ]
# Create a bar plot
ggplot(filtered_data, aes(x = cf1edu, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#808000', high='#FF7F50') +
  labs(title = "Distribution of Father education")
```

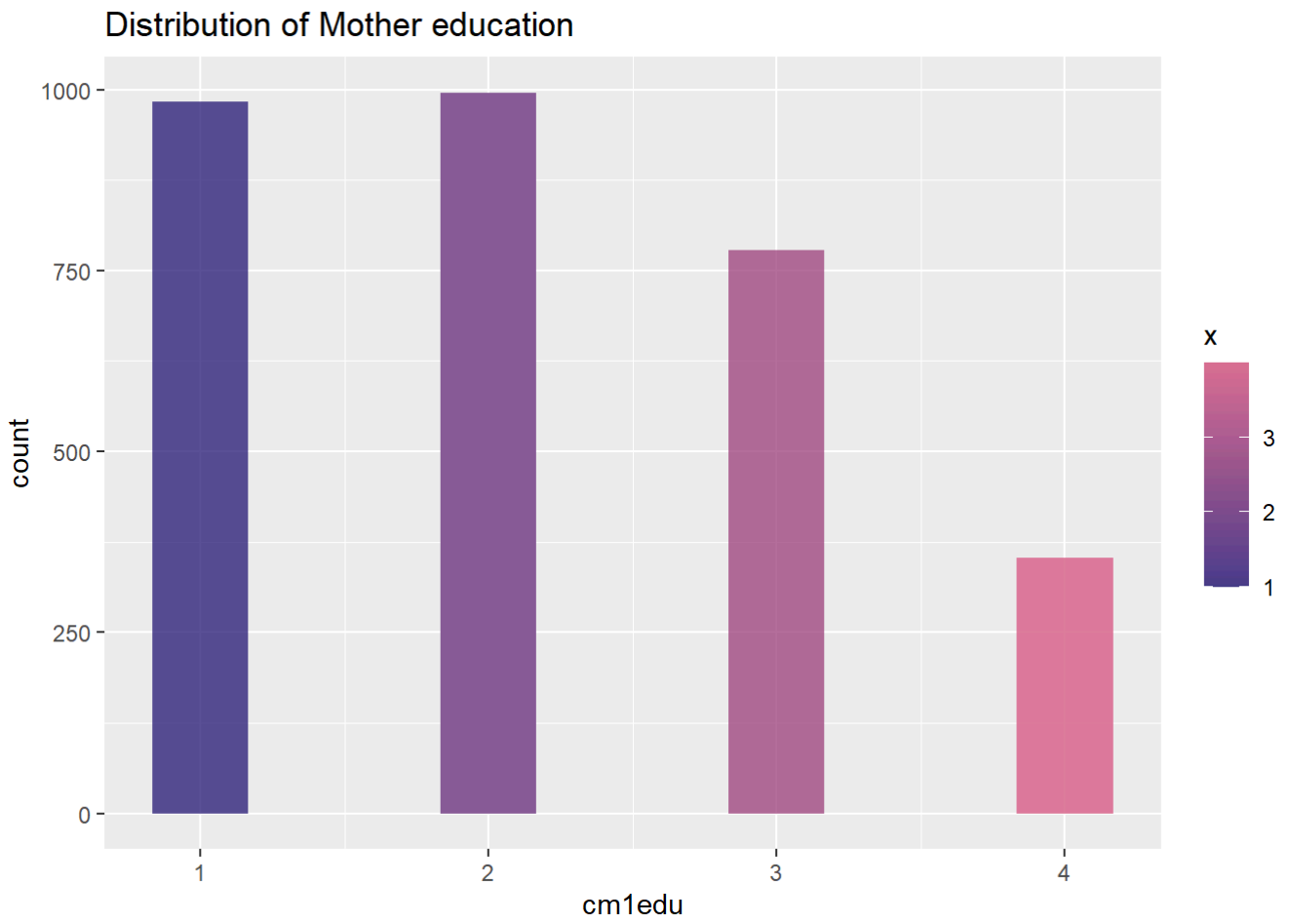
```
## Warning: The dot-dot notation (`..x..`) was deprecated in ggplot2 3.4.0.
## ℹ Please use `after_stat(x)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
ggsave(filename = "father.png")
```

```
## Saving 7 x 5 in image
```

```
# Mother
# Filter the data to exclude rows where cfledu equals -3
filtered_data_m <- data[data$cmledu != -3, ]
# Create a bar plot
ggplot(filtered_data_m, aes(x = cmledu, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#483D8B', high='#DB7093') +
  labs(title = "Distribution of Mother education")
```




```
ggsave(filename = "mother.png")
```

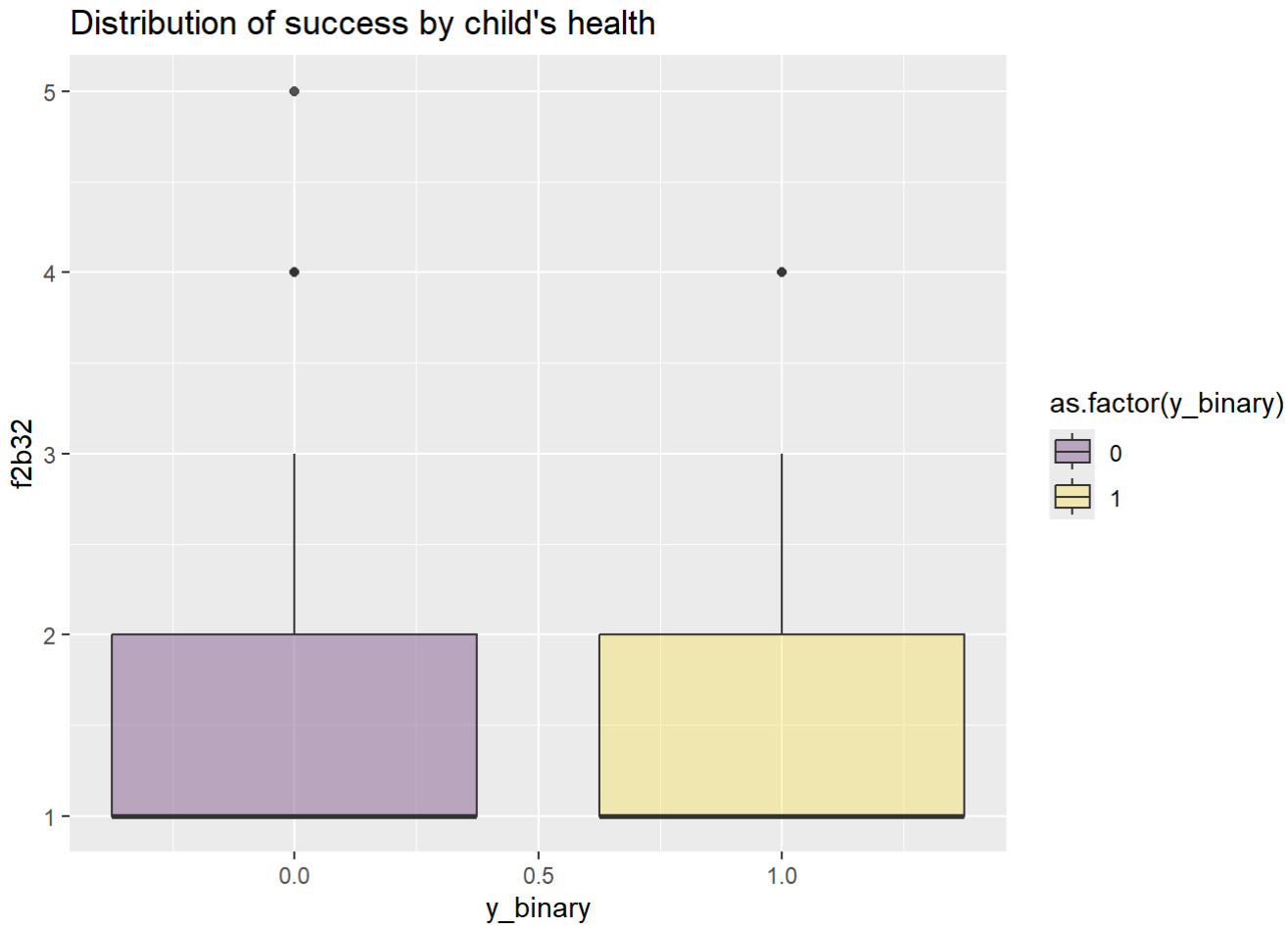
```
## Saving 7 x 5 in image
```

3.f2b32: In general, how is your child's health?

```
library(viridis)
```

```
## 载入需要的程辑包：viridisLite
```

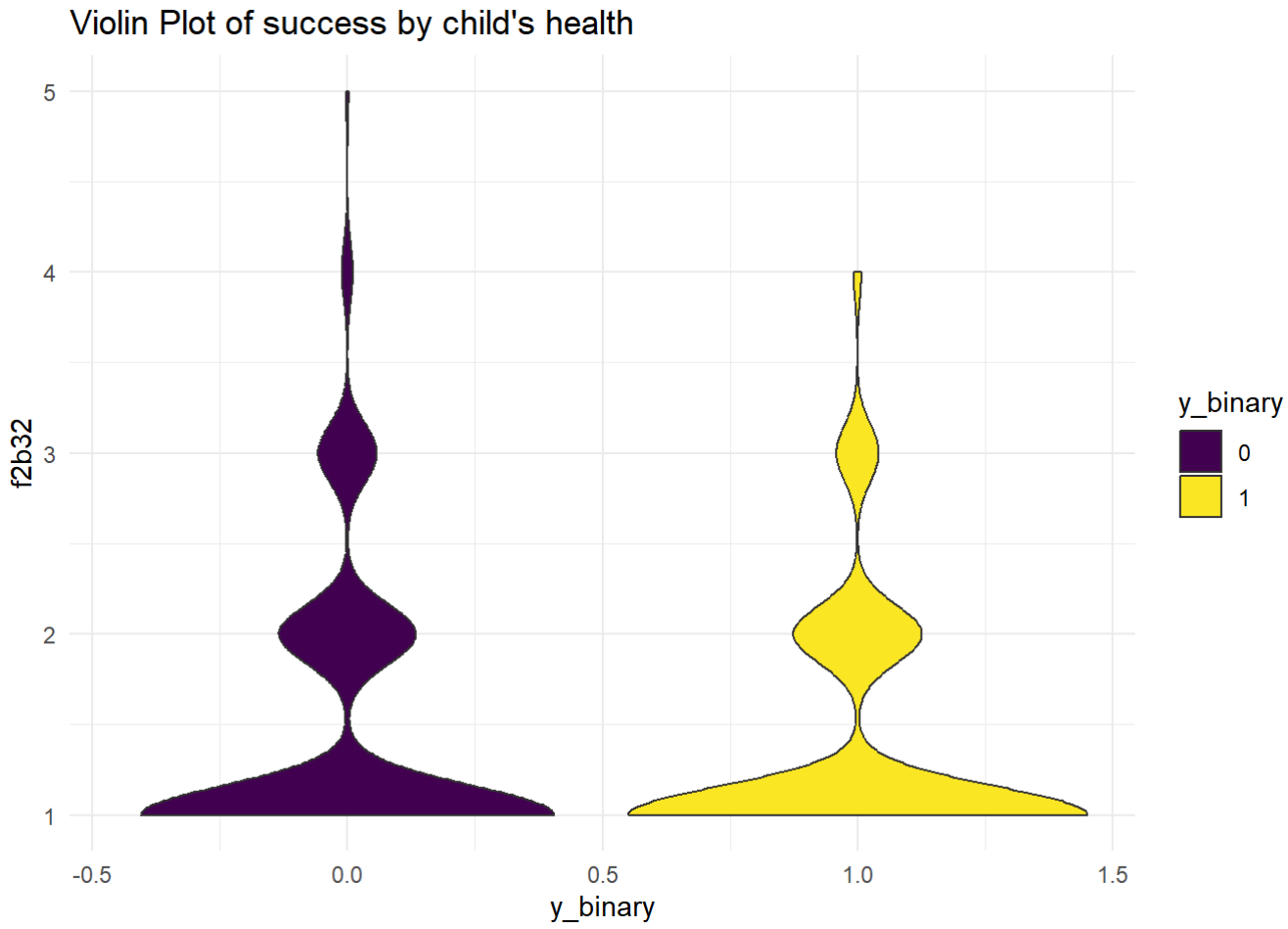
```
# Filter the data without negative
filtered_data_h <- data[data$f2b32 > 0, ]
# boxplot
ggplot(filtered_data_h, aes(x = y_binary, y = f2b32, fill = as.factor(y_binary))) +
  geom_boxplot(alpha = 0.3)  +
  scale_fill_viridis_d() +
  labs(title = "Distribution of success by child's health")
```




```
ggsave(filename = "success by child's health1.png")
```

```
## Saving 7 x 5 in image
```

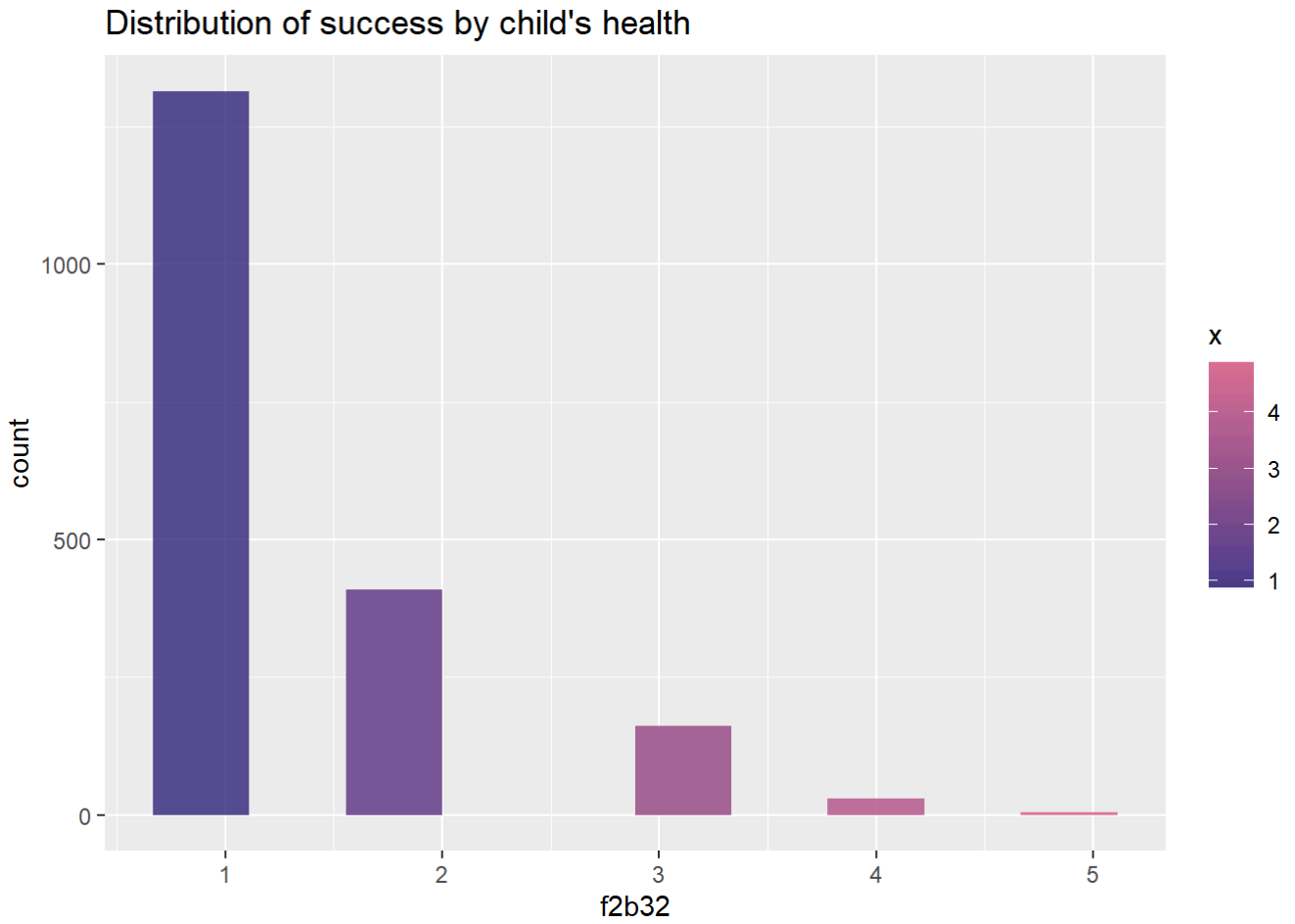
```
# Violin Plot
filtered_data <- data[data$f2b32 > 0, ]
# violin plot
ggplot(filtered_data, aes(y = f2b32, x = y_binary, fill = as.factor(y_binary))) +
  geom_violin() +
  scale_fill_viridis_d() +
  labs(title = "Violin Plot of success by child's health", y = "f2b32", x = "y_binary", fill =
"y_binary") +
  theme_minimal()
```




```
ggsave(filename = "success by child's health2.png")
```

```
## Saving 7 x 5 in image
```

```
# histogram
filtered_data_m <- data[data$f2b32 > 0, ]
# Create a bar plot
ggplot(filtered_data_m, aes(x = f2b32, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#483D8B', high='#DB7093') +
  labs(title = "Distribution of success by child's health")
```

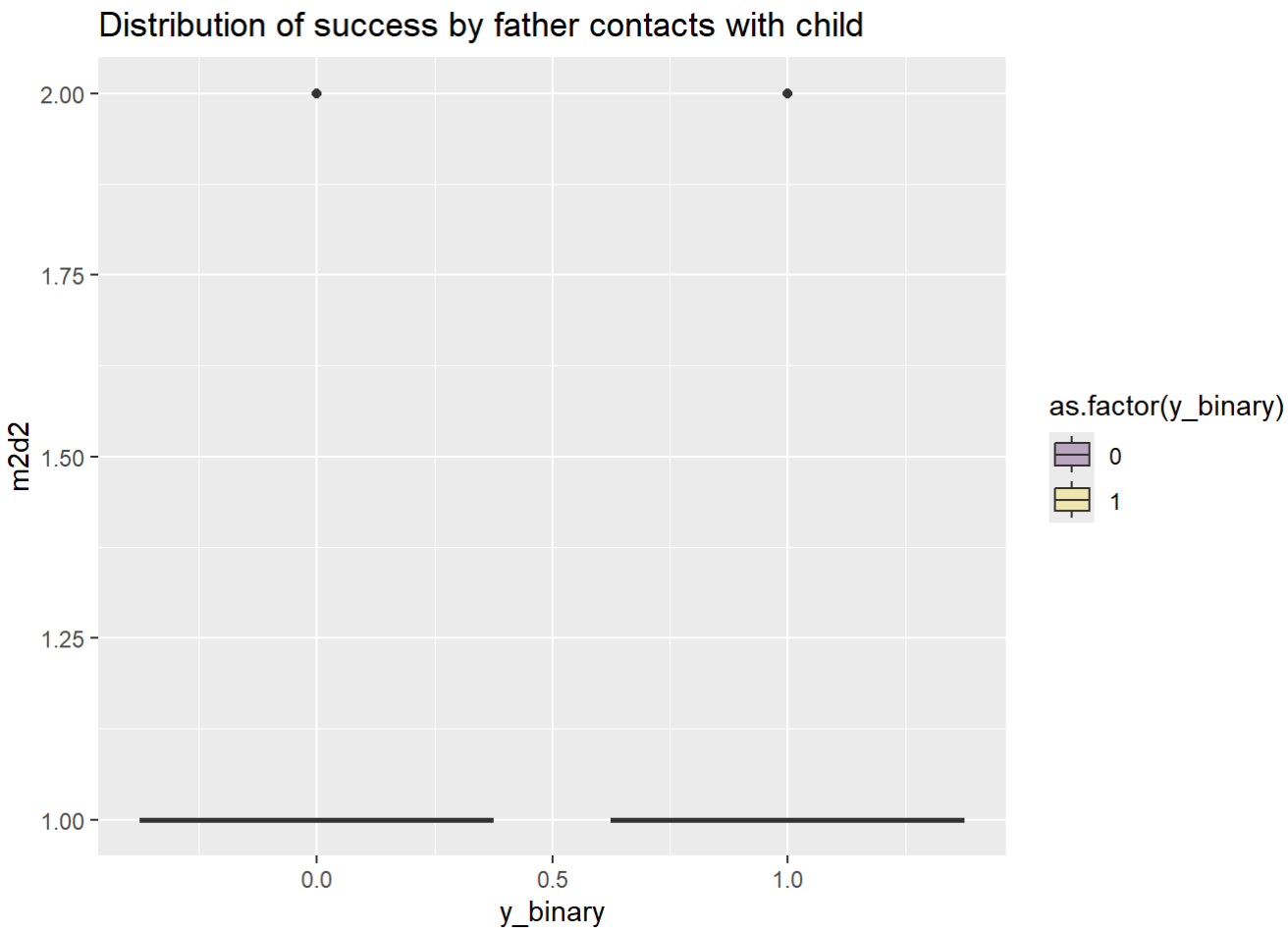



```
ggsave(filename = "success by child's health3.png")
```

```
## Saving 7 x 5 in image
```

4. m2d2: Int Chk: Does father have any contact with child ?

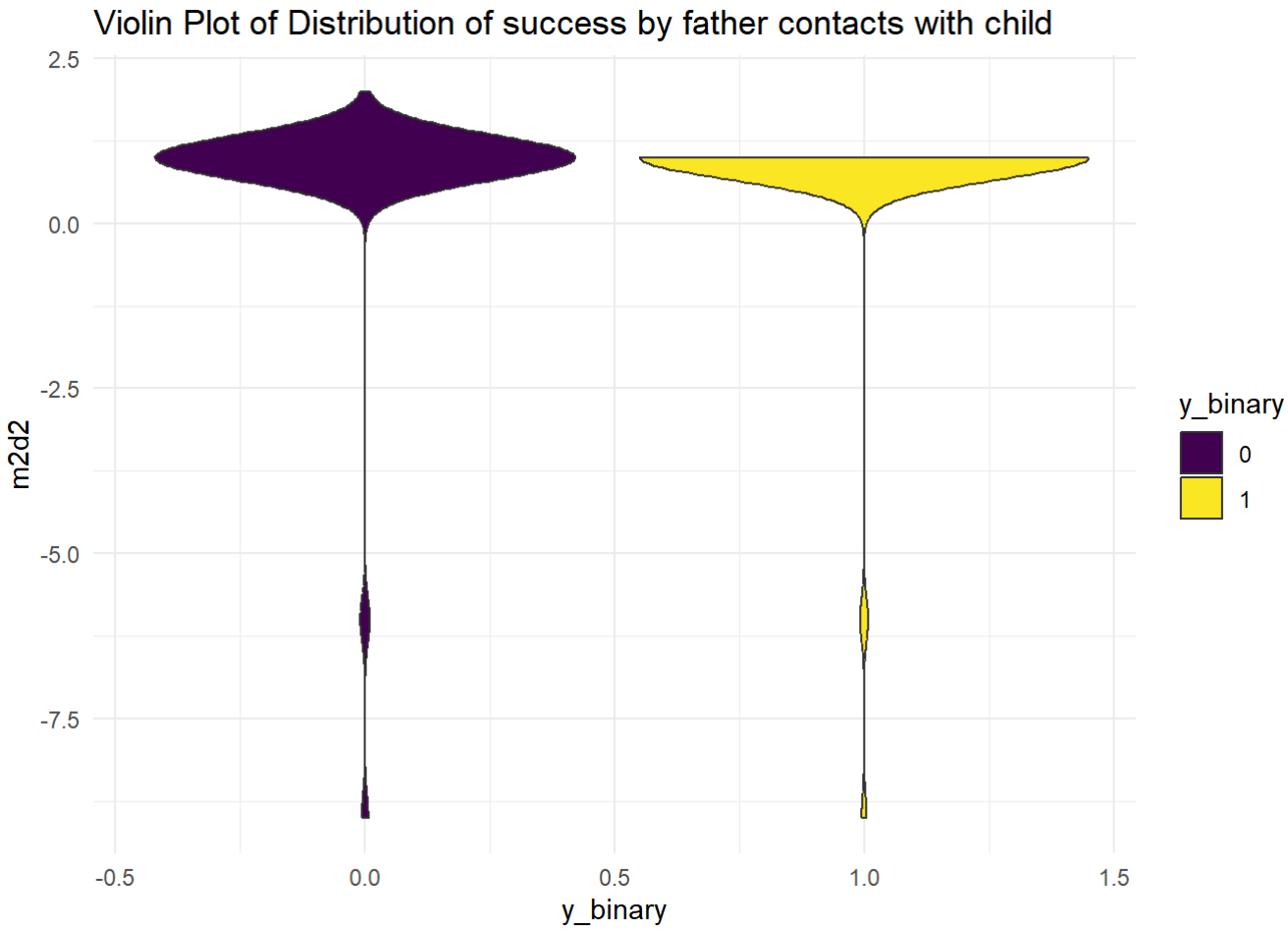
```
# Filter the data without negative
filtered_data_h <- data[data$m2d2 > 0, ]
# boxplot
ggplot(filtered_data_h, aes(x = y_binary, y = m2d2, fill = as.factor(y_binary))) +
  geom_boxplot(alpha = 0.3)  +
  scale_fill_viridis_d() +
  labs(title = "Distribution of success by father contacts with child")
```




```
ggsave(filename = "father contacts with child.png")
```

```
## Saving 7 x 5 in image
```

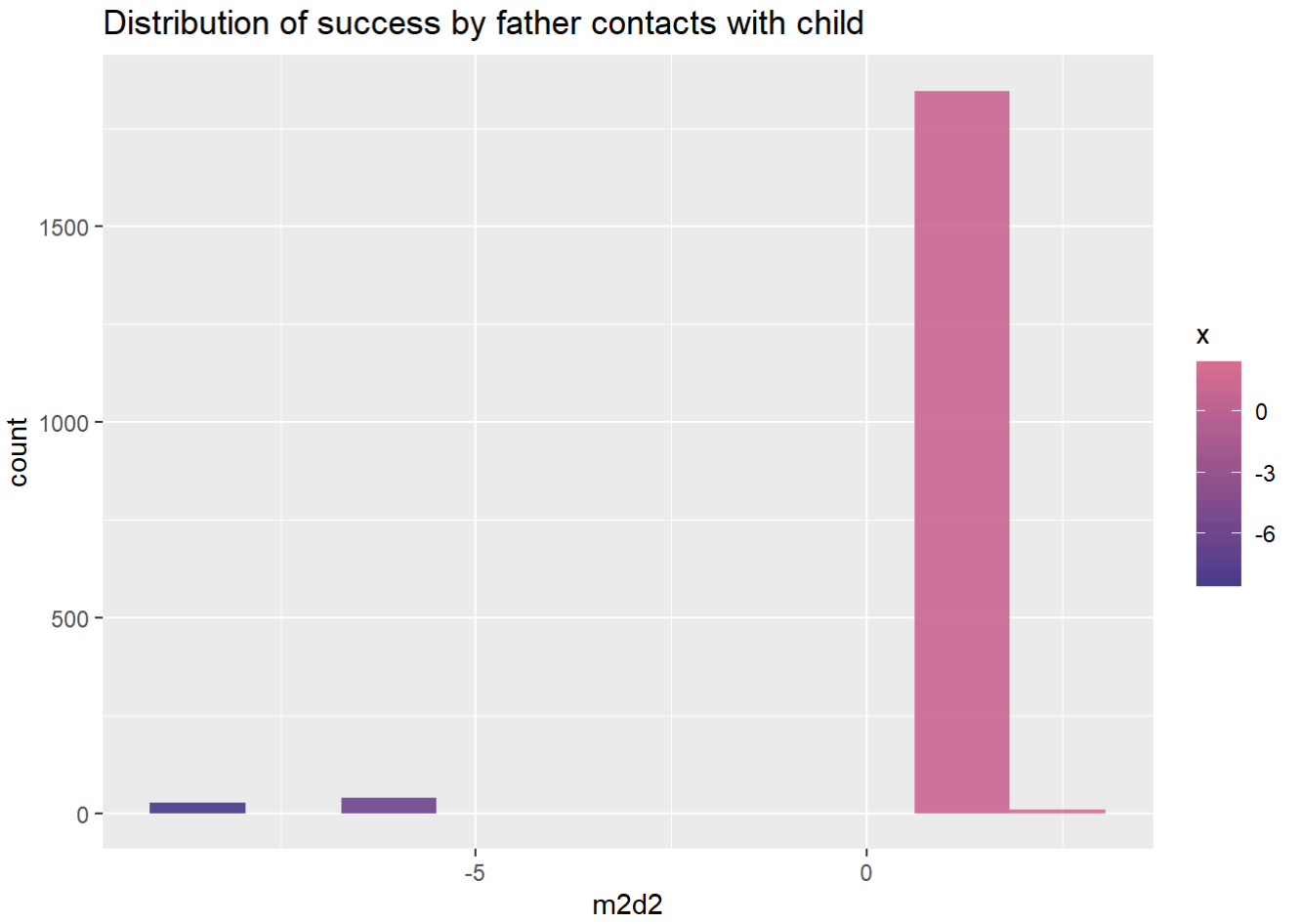
```
# Violin Plot
filtered_data <- data[data$f2b32 > 0, ]
# violin plot
ggplot(filtered_data, aes(y = m2d2, x = y_binary, fill = as.factor(y_binary))) +
  geom_violin() +
  scale_fill_viridis_d() +
  labs(title = "Violin Plot of Distribution of success by father contacts with child", y = "m2d
2", x = "y_binary", fill = "y_binary") +
  theme_minimal()
```




```
ggsave(filename = "father contacts with child2.png")
```

```
## Saving 7 x 5 in image
```

```
# histogram
filtered_data_m <- data[data$f2b32 > 0, ]
# Create a bar plot
ggplot(filtered_data_m, aes(x = m2d2, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#483D8B', high='#DB7093') +
  labs(title = "Distribution of success by father contacts with child")
```



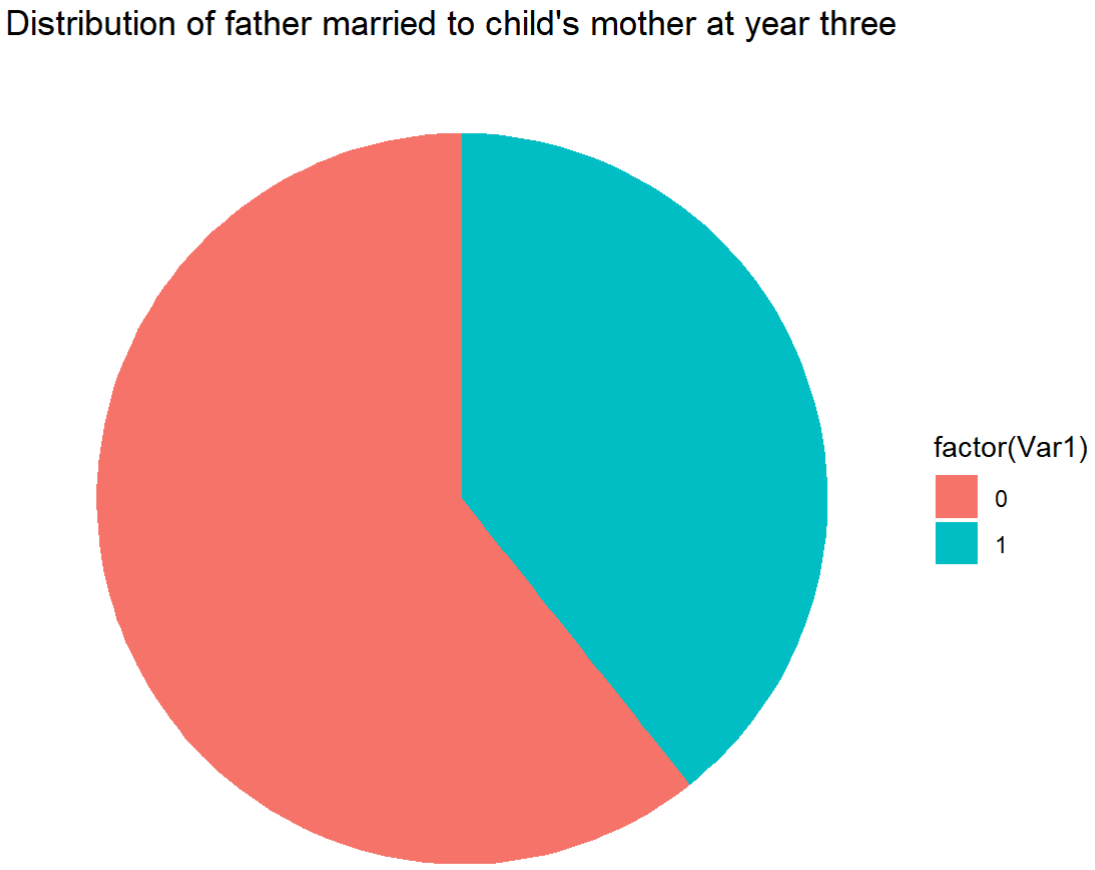

```
ggsave(filename = "father contacts with child3.png")
```

```
## Saving 7 x 5 in image
```

5.m2b43a: On a scale of 1-(least like) to 5-(most like) - Child tends to be shy

6.cf3marm: Constructed - Is father married to child's mother at year three?

```
# 0-No 1-Yes
# Filter the data without negative
filtered_data <- data[data$cf3marm >= 0, ]
# Count the frequency of each category in f2d1a
category_counts <- table(filtered_data$cf3marm)
# Convert the frequency table to a data frame
category_df <- as.data.frame(category_counts)
# Create a pie chart
ggplot(category_df, aes(x = "", y = Freq, fill = factor(Var1))) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Distribution of father married to child's mother at year three") +
  theme_void() +
  theme(legend.position = "right")
```

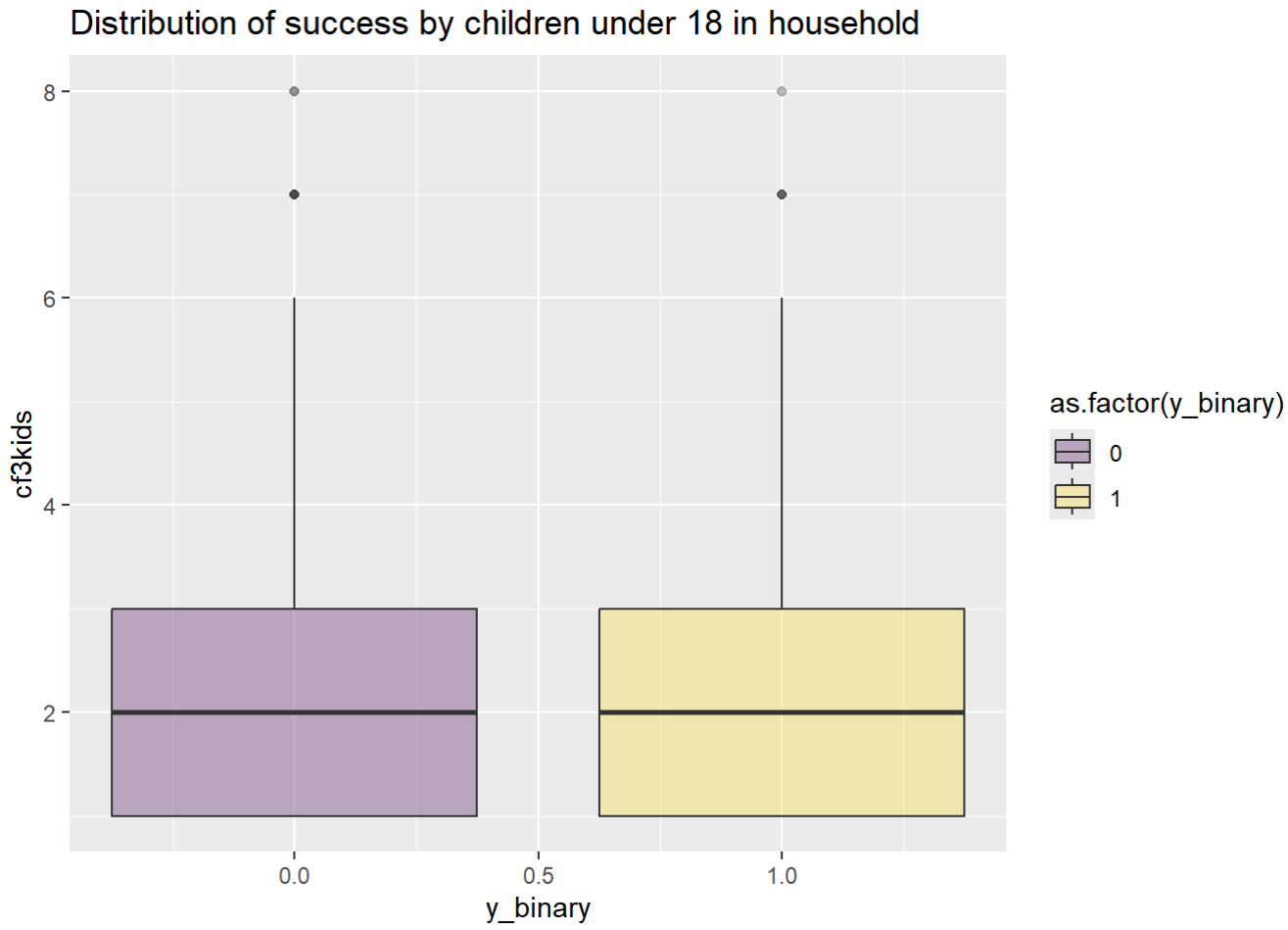



```
ggsave(filename = "father married to child's mother.png")
```

```
## Saving 7 x 5 in image
```

7.cf3kids: Constructed - Number of children under 18 in household

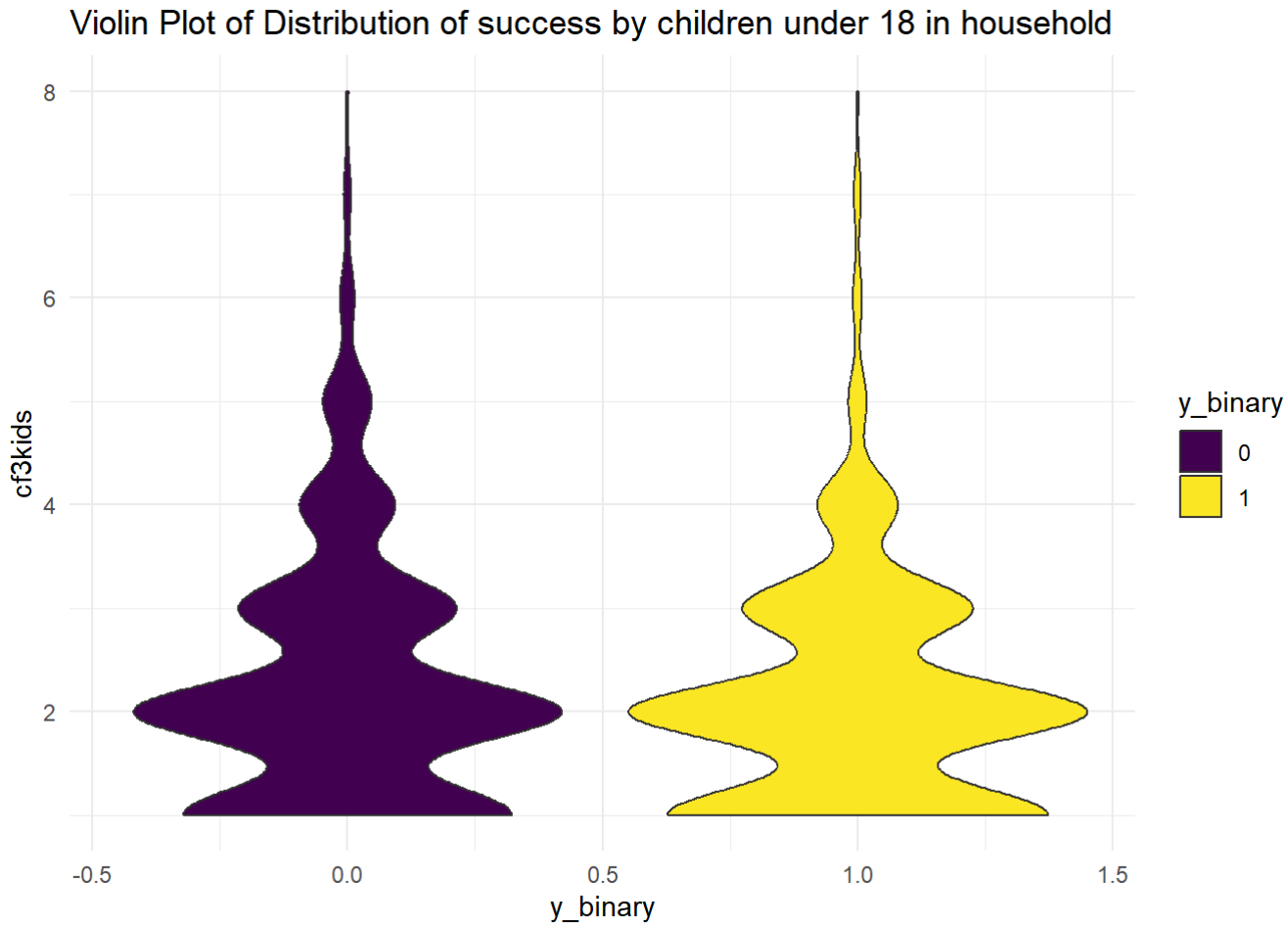
```
# Filter the data without negative
filtered_data_h <- data[data$cf3kids > 0, ]
# boxplot
ggplot(filtered_data_h, aes(x = y_binary, y = cf3kids, fill = as.factor(y_binary))) +
  geom_boxplot(alpha = 0.3)  +
  scale_fill_viridis_d() +
  labs(title = "Distribution of success by children under 18 in household")
```




```
ggsave(filename = "children under 18 in household.png")
```

```
## Saving 7 x 5 in image
```

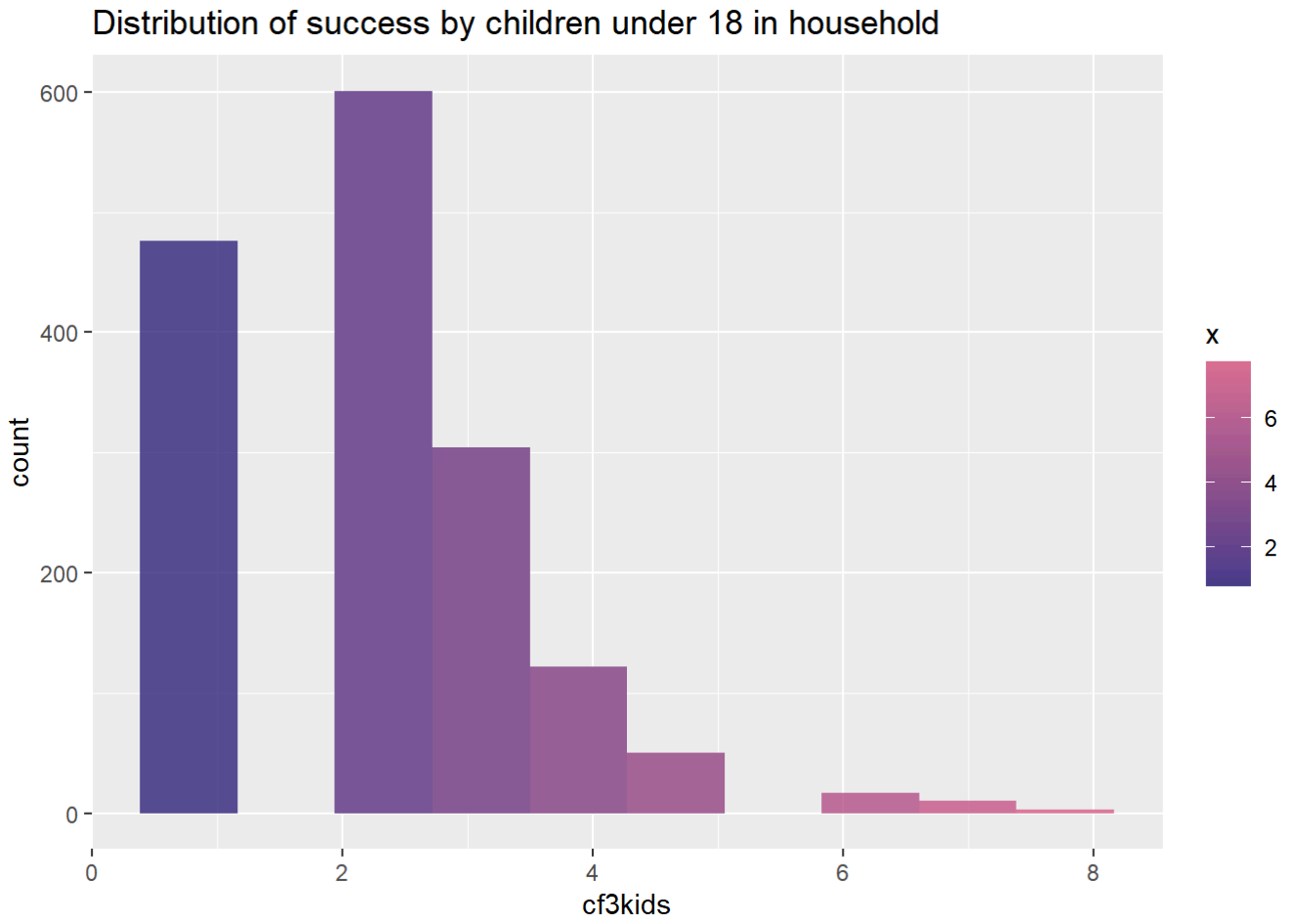
```
# Violin Plot
filtered_data <- data[data$cf3kids > 0, ]
# violin plot
ggplot(filtered_data, aes(y = cf3kids, x = y_binary, fill = as.factor(y_binary))) +
  geom_violin() +
  scale_fill_viridis_d() +
  labs(title = "Violin Plot of Distribution of success by children under 18 in household", y =
"cf3kids", x = "y_binary", fill = "y_binary") +
  theme_minimal()
```




```
ggsave(filename = "children under 18 in household2.png")
```

```
## Saving 7 x 5 in image
```

```
# histogram
filtered_data_m <- data[data$cf3kids > 0, ]
# Create a bar plot
ggplot(filtered_data_m, aes(x = cf3kids, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#483D8B', high='#DB7093') +
  labs(title = "Distribution of success by children under 18 in household")
```

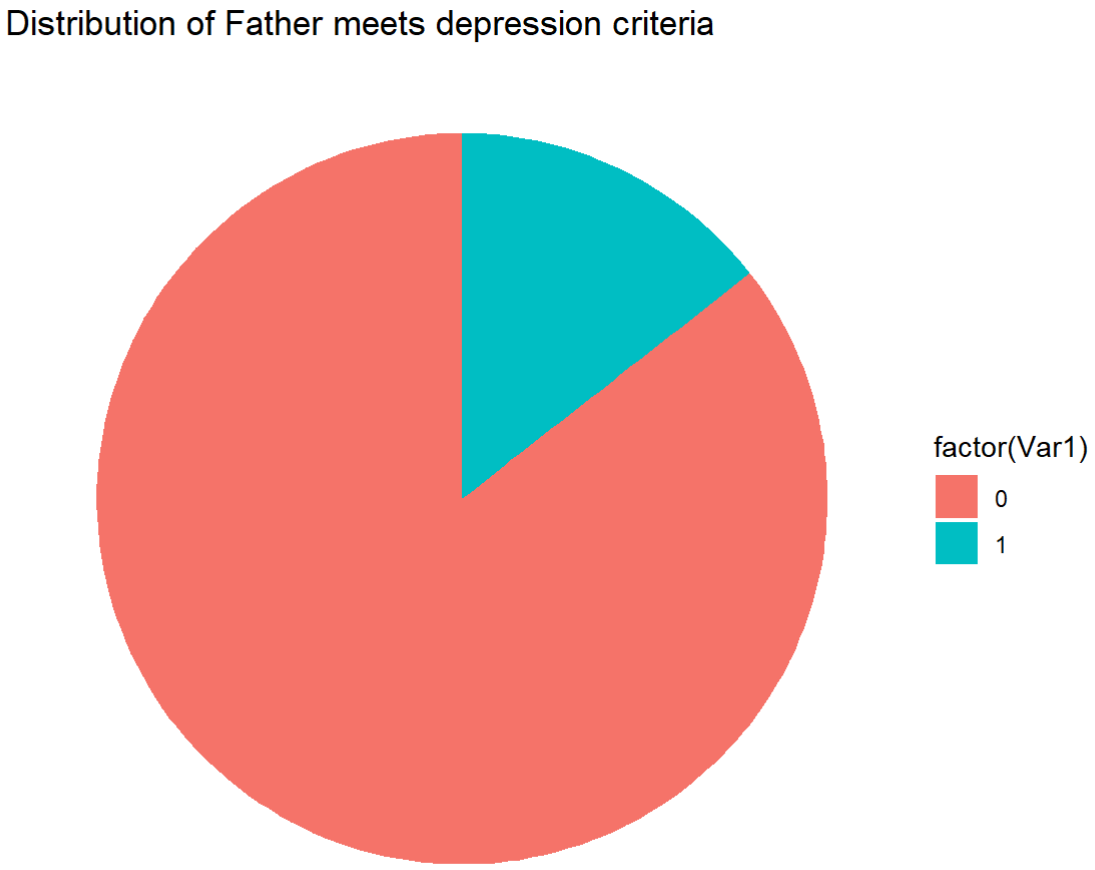



```
ggsave(filename = "children under 18 in household3.png")
```

```
## Saving 7 x 5 in image
```

8. cm3md_case_lib: Constructed - Father meets depression criteria (liberal) at three-year (CIDI)

```
# Filter the data without negative
filtered_data <- data[data$cf3md_case_lib >= 0, ]
# Count the frequency of each category in f2d1a
category_counts <- table(filtered_data$cf3md_case_lib)
# Convert the frequency table to a data frame
category_df <- as.data.frame(category_counts)
# Create a pie chart
ggplot(category_df, aes(x = "", y = Freq, fill = factor(Var1))) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Distribution of Father meets depression criteria") +
  theme_void() +
  theme(legend.position = "right")
```

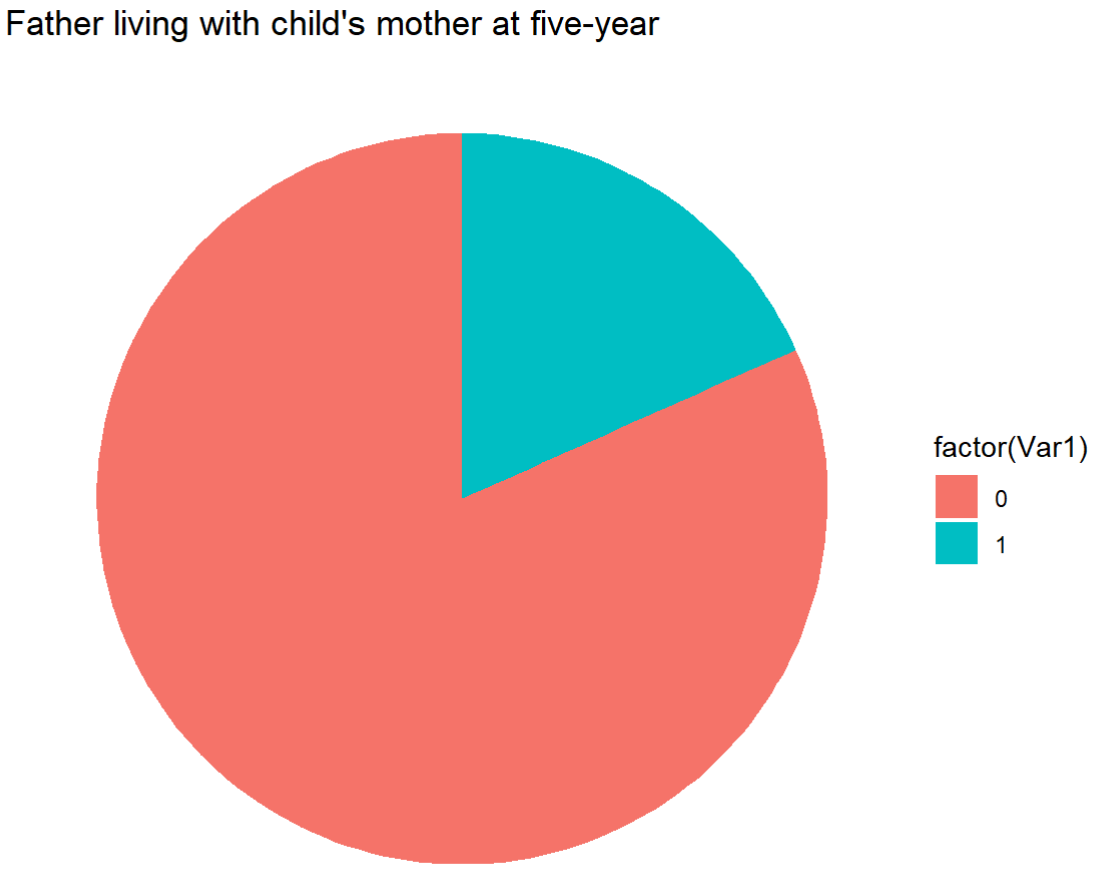



```
ggsave(filename = "Father meets depression criteria.png")
```

```
## Saving 7 x 5 in image
```

9.cf4cohm: Constructed - Father living with child's mother at five-year

```
# Filter the data without negative
filtered_data <- data[data$cf4cohm >= 0, ]
# Count the frequency of each category in f2d1a
category_counts <- table(filtered_data$cf4cohm)
# Convert the frequency table to a data frame
category_df <- as.data.frame(category_counts)
# Create a pie chart
ggplot(category_df, aes(x = "", y = Freq, fill = factor(Var1))) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Father living with child's mother at five-year") +
  theme_void() +
  theme(legend.position = "right")
```

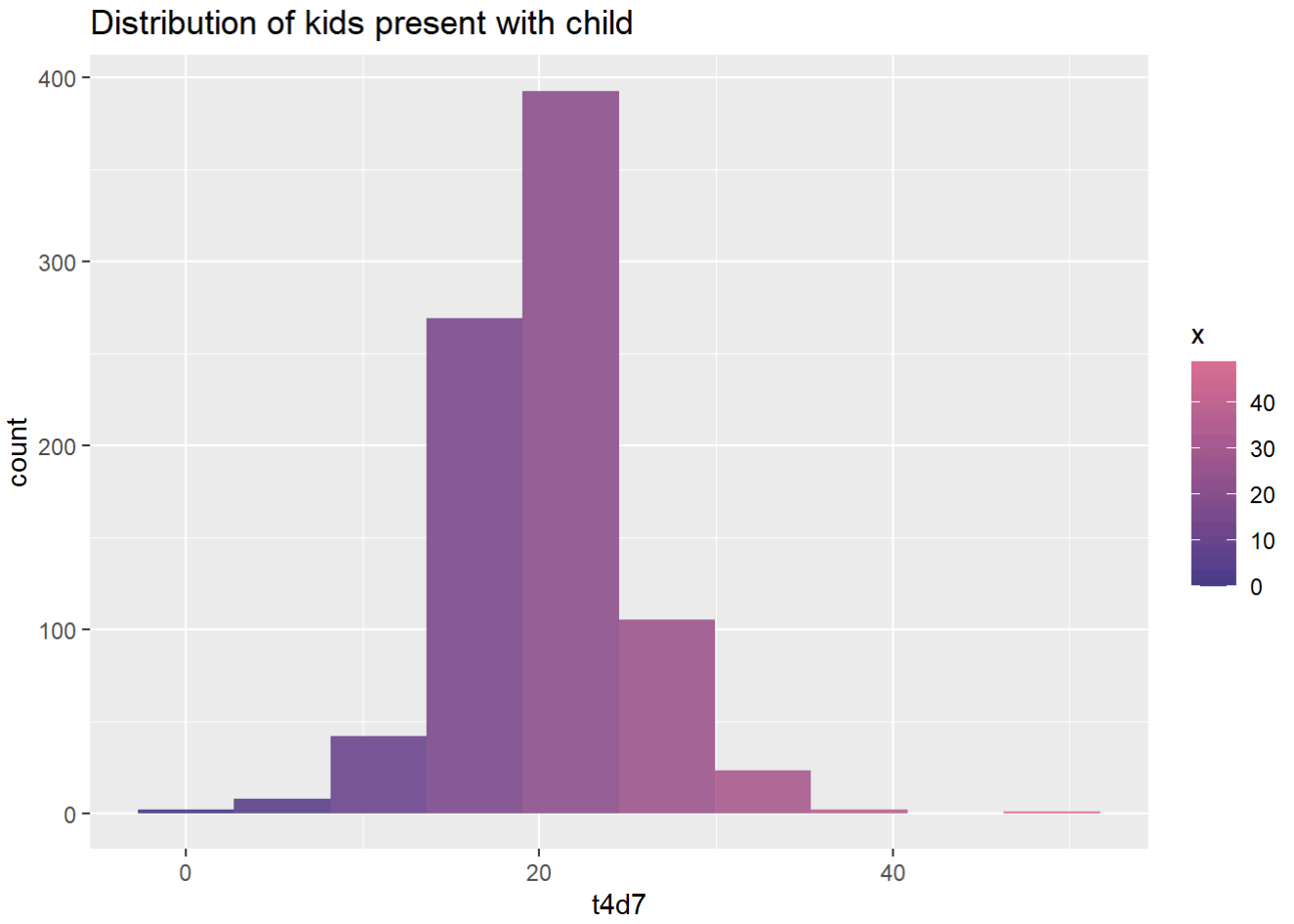



```
ggsave(filename = "Father living with child's mother at five-year.png")
```

```
## Saving 7 x 5 in image
```

10. t4d7: number of kids present with child

```
filtered_data_m <- data[data$t4d7 >= 0, ]
# Create a bar plot
ggplot(filtered_data_m, aes(x = t4d7, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#483D8B', high='#DB7093') +
  labs(title = "Distribution of kids present with child")
```



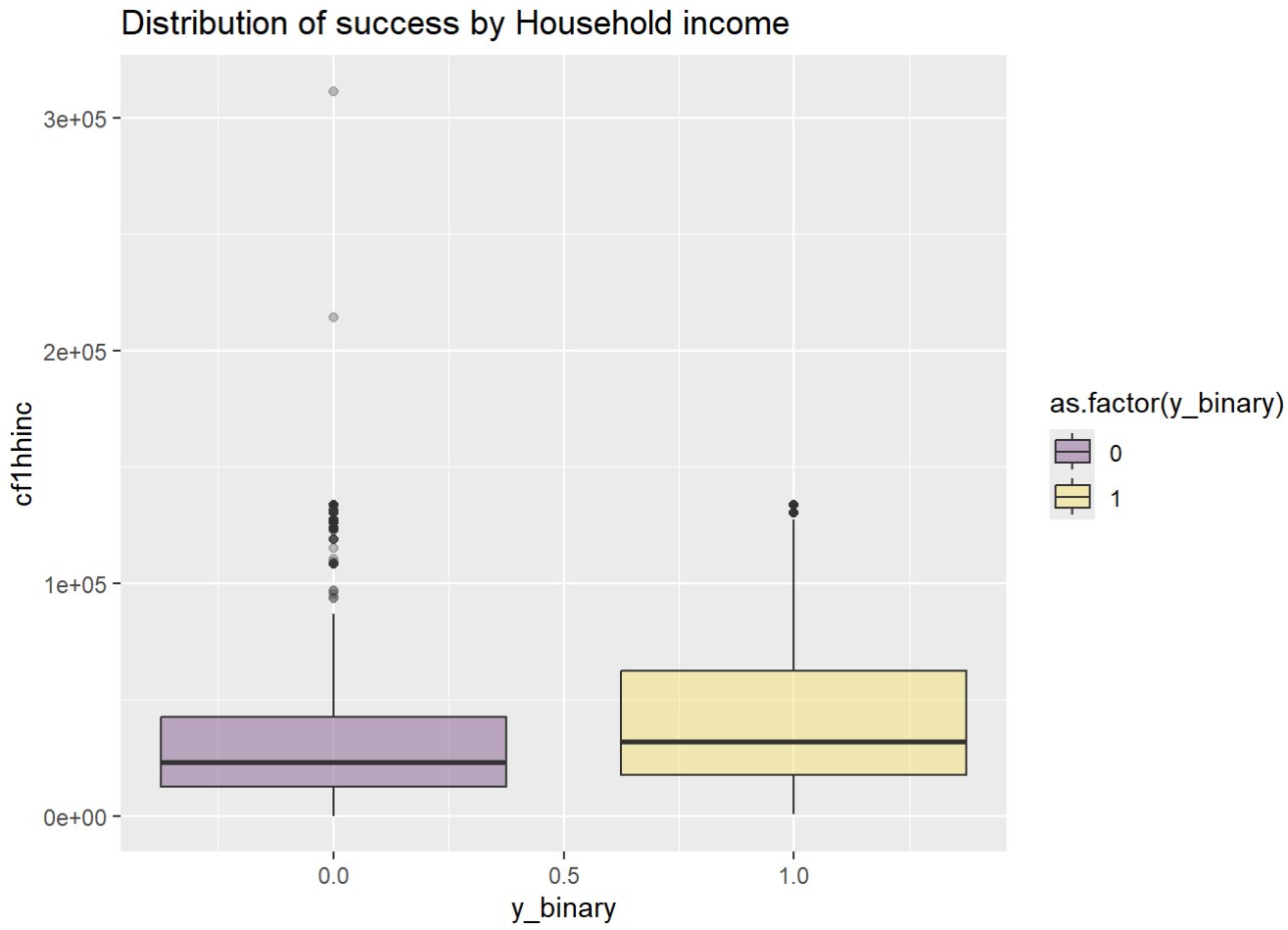

```
ggsave(filename = "kids present with child.png")
```

```
## Saving 7 x 5 in image
```

```
# scatter plot
# ggplot(filtered_data_m, aes(x = t4d7, y = y_binary)) +
  # geom_point() +
  # labs(title = "Success by kids present with child")
```

11. cf1hhinc: Household income

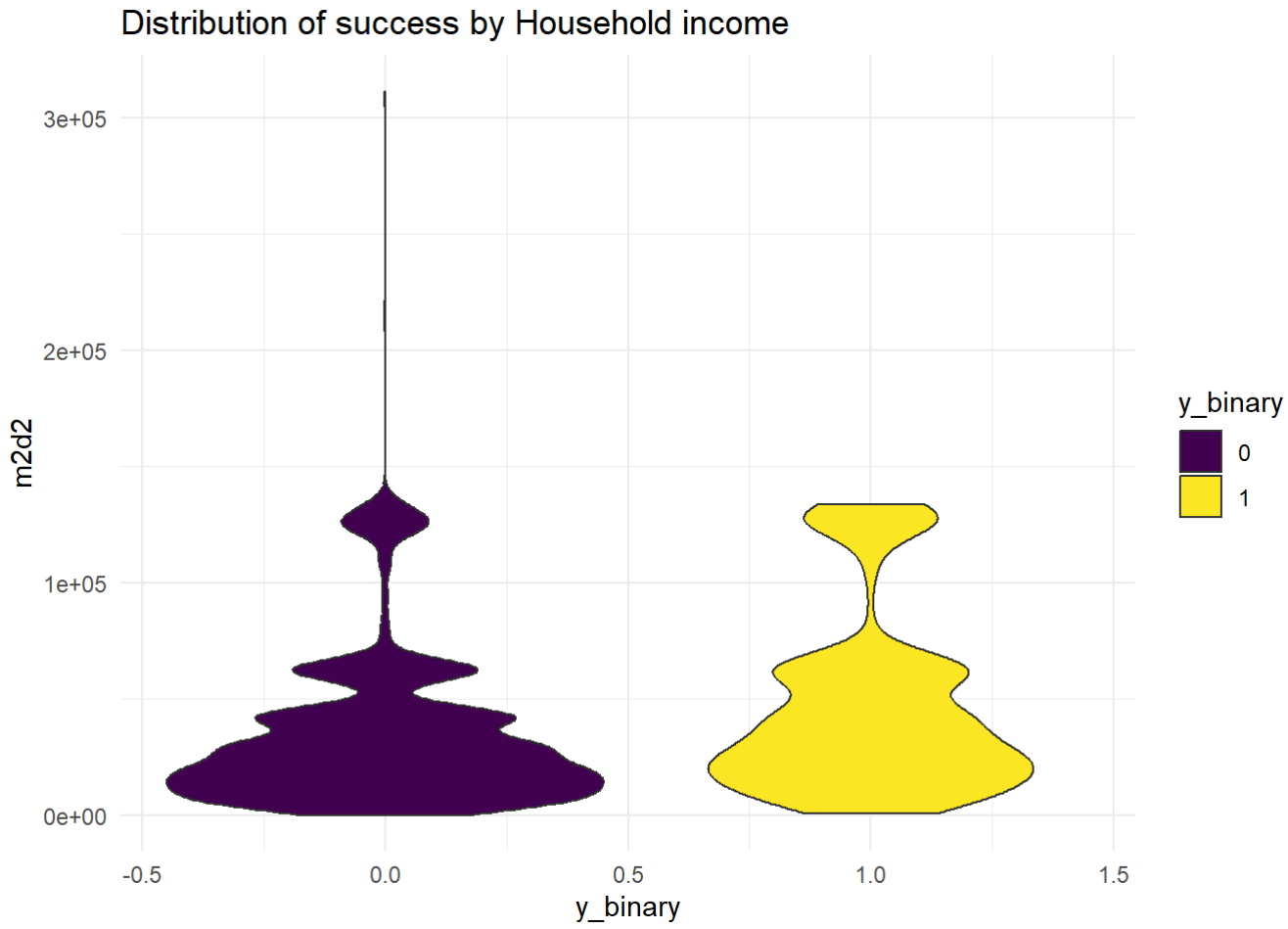
```
# Filter the data without negative
filtered_data <- data[data$cf1hhinc > 0, ]
# boxplot
ggplot(filtered_data, aes(x = y_binary, y = cf1hhinc, fill = as.factor(y_binary))) +
  geom_boxplot(alpha = 0.3)  +
  scale_fill_viridis_d() +
  labs(title = "Distribution of success by Household income")
```




```
ggsave(filename = "Household income.png")
```

```
## Saving 7 x 5 in image
```

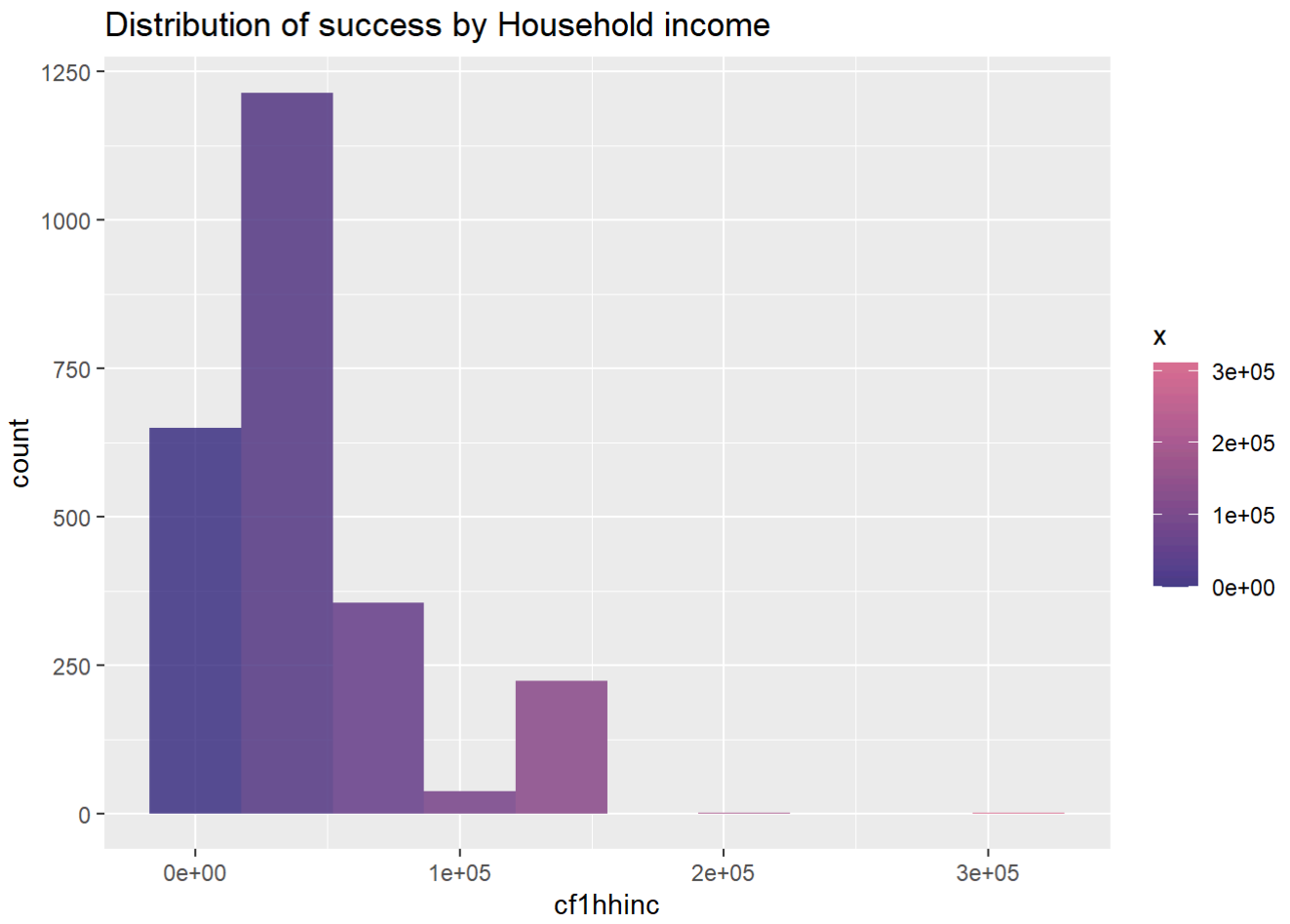
```
# violin plot
ggplot(filtered_data, aes(y = cf1hhinc, x = y_binary, fill = as.factor(y_binary))) +
  geom_violin() +
  scale_fill_viridis_d() +
  labs(title = "Distribution of success by Household income", y = "m2d2", x = "y_binary", fill
= "y_binary") +
  theme_minimal()
```




```
ggsave(filename = "Household income2.png")
```

```
## Saving 7 x 5 in image
```

```
# histogram
ggplot(filtered_data, aes(x = cf1hhinc, fill = ..x..)) +
  geom_histogram(bins = 10, alpha=0.9)  +
  scale_fill_gradient(low='#483D8B', high='#DB7093') +
  labs(title = "Distribution of success by Household income")
```

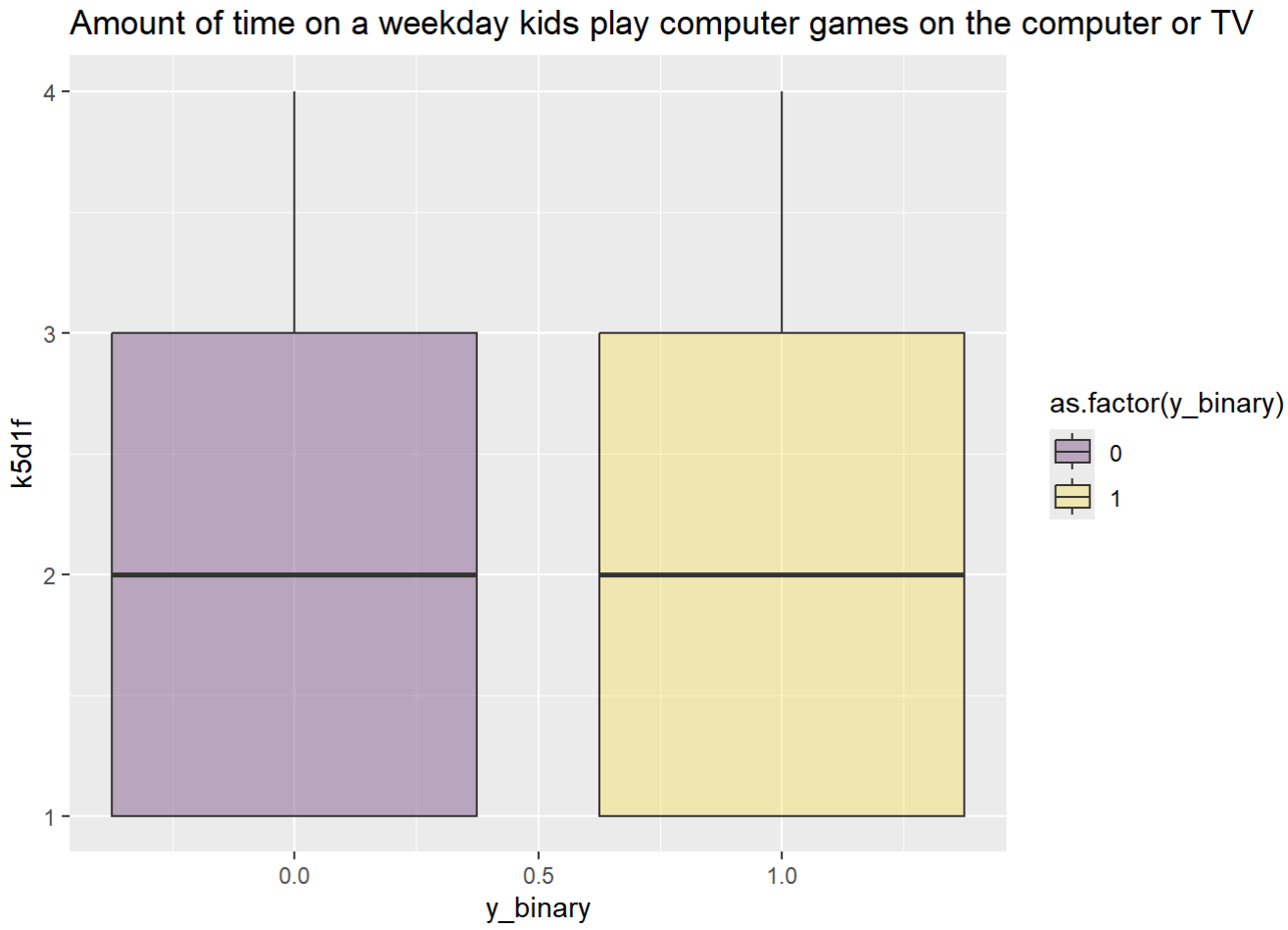



```
ggsave(filename = "Household income3.png")
```

```
## Saving 7 x 5 in image
```

12. k5d1f: Amount of time on a weekday you play computer games on the computer or TV.

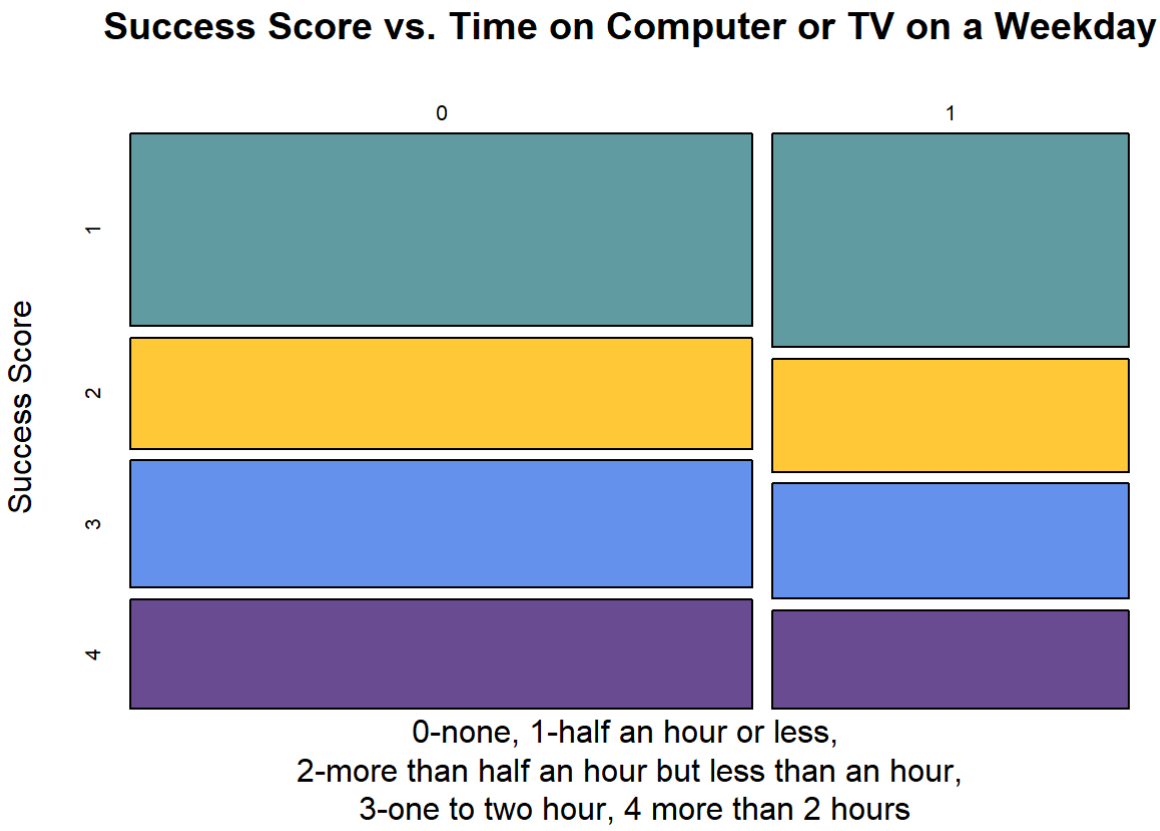
```
# Filter the data without negative
filtered_data <- data[data$k5d1f > 0, ]
# boxplot
ggplot(filtered_data, aes(x = y_binary, y = k5d1f, fill = as.factor(y_binary))) +
  geom_boxplot(alpha = 0.3)  +
  scale_fill_viridis_d() +
  labs(title = "Amount of time on a weekday kids play computer games on the computer or TV")
```




```
ggsave(filename = "play computer games.png")
```

```
## Saving 7 x 5 in image
```

```
# Mosaic Plot
colors <- c("#5F9EA0", "#ffca3a", "#6495ED","#6a4c93")
mosaicplot(table(filtered_data$y_binary, filtered_data$k5d1f),
           main = "Success Score vs. Time on Computer or TV on a Weekday",
           ylab = "Success Score",
           color = colors,
           sub = paste("0-none, 1-half an hour or less,",'\n',
            "2-more than half an hour but less than an hour,",'\n',
            "3-one to two hour, 4 more than 2 hours"))
```

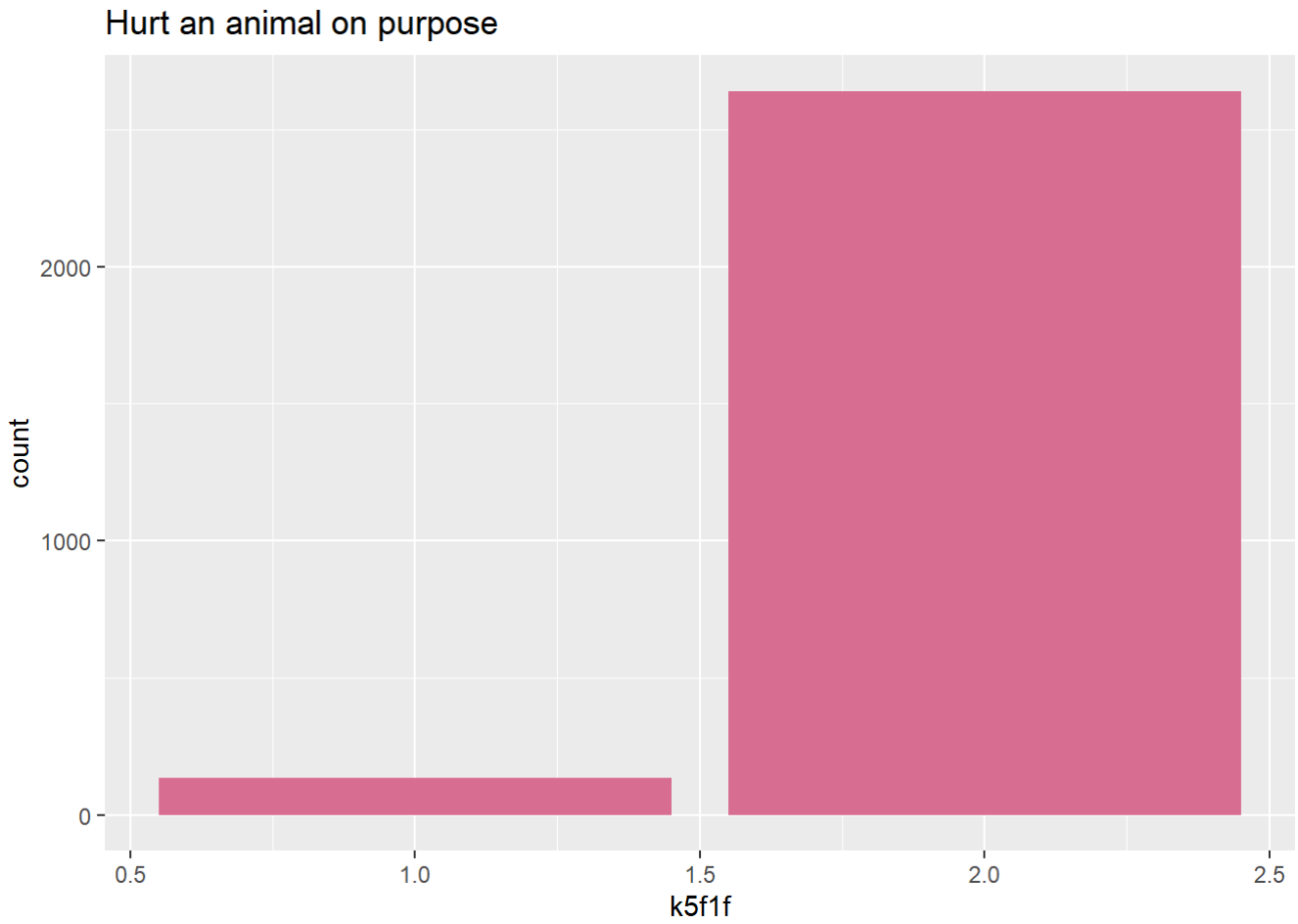



```
ggsave(filename = "play computer games2.png")
```

```
## Saving 7 x 5 in image
```

13. k5f1f F1F. Hurt an animal on purpose

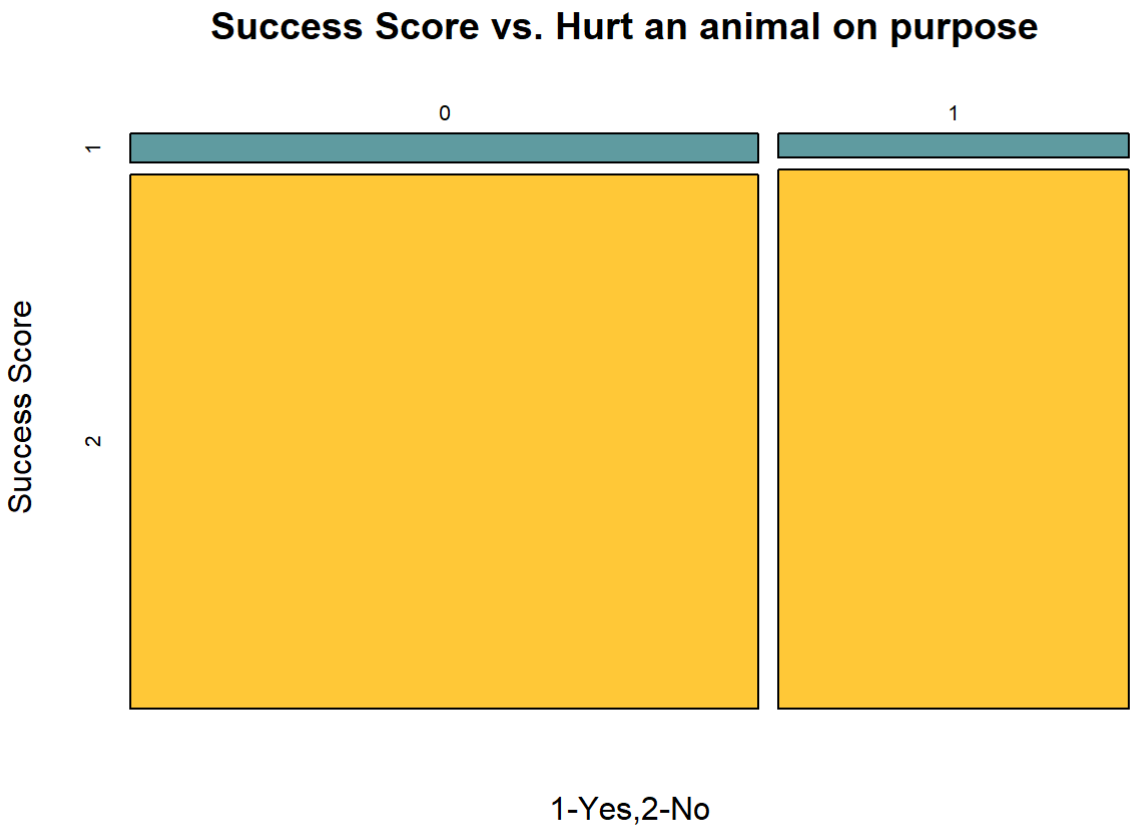
```
filtered_data_m <- data[data$k5f1f >= 0, ]
# Create a bar plot
ggplot(filtered_data_m, aes(x = k5f1f)) +
  geom_bar(fill='#DB7093')  +
  labs(title = "Hurt an animal on purpose")
```




```
ggsave(filename = "Hurt an animal on purpose.png")
```

```
## Saving 7 x 5 in image
```

```
# Mosaic Plot
colors <- c("#5F9EA0", "#ffca3a", "#6495ED","#6a4c93")
mosaicplot(table(filtered_data_m$y_binary, filtered_data_m$k5f1f),
           main = "Success Score vs. Hurt an animal on purpose ",
           ylab = "Success Score",
           color = colors,
           sub = paste("1-Yes,2-No"))
```




```
ggsave(filename = "Hurt an animal on purpose2.png")
```

```
## Saving 7 x 5 in image
```