ProjCode2

2025-02-25

```
load("/Users/ellathomasson/Documents/STAT 4996/NSDUH_2023.Rdata")
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(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
povertyDepData<-puf2023_102124[,c("AMDELT",
                                  "POVERTY3",
                                  "NEWRACE2",
                                  "IRHHSIZ2",
                                  "MOVSINPYR2",
                                  "MEDICARE",
                                  "CAIDCHIP",
                                  "CHAMPUS",
                                  "PRVHLTIN",
                                  "GRPHLTIN",
                                  "IREDUHIGHST2",
                                  "WRKSTATWK2"
)]

names(povertyDepData)<-c("lifetimeDepression",
                         "povertyThreshold",
                         "race",
                         "householdSize",
```

```
                          "timesMovedPastYear",
                          "medicare",
                          "medicaid_chip",
                          "govtHealthcare",
                          "privateHealthcare",
                          "employer_unionHealthcare",
                          "highestEducationCompleted",
                          "workStatusPastWeek")
```

EDA - times moved in past year

```
povertyDepData_cleanedMove = povertyDepData[povertyDepData$timesMovedPastYear < 4,]
povertyDepData_cleanedMove2 <- povertyDepData_cleanedMove[!is.na(povertyDepData_cleanedMove$lifetimeDepr
```

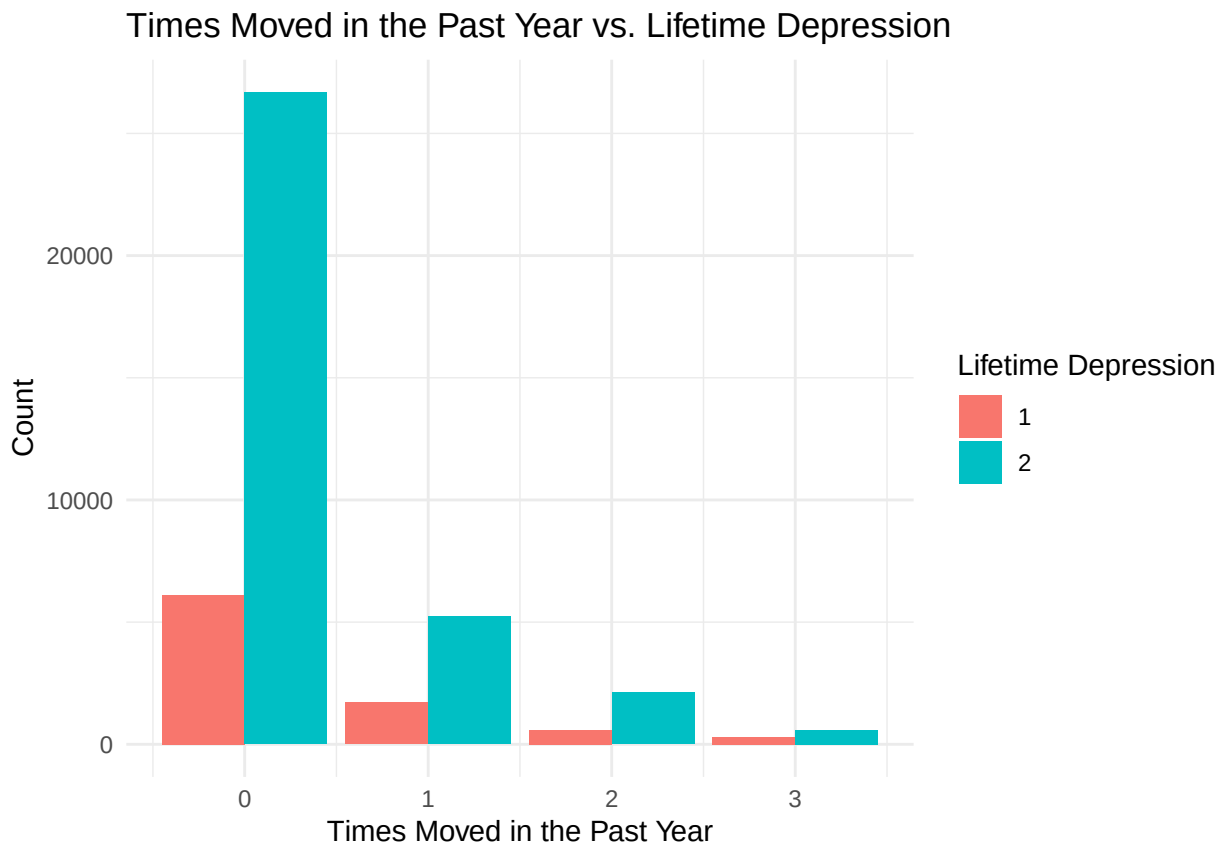
```
povertyDepData_cleanedMove2 %>%
  group_by(timesMovedPastYear,lifetimeDepression) %>%
  summarise(
    count = n()
  ) %>%
  kable()
```

```
## `summarise()` has grouped output by 'timesMovedPastYear'. You can override
## using the `.groups` argument.
```

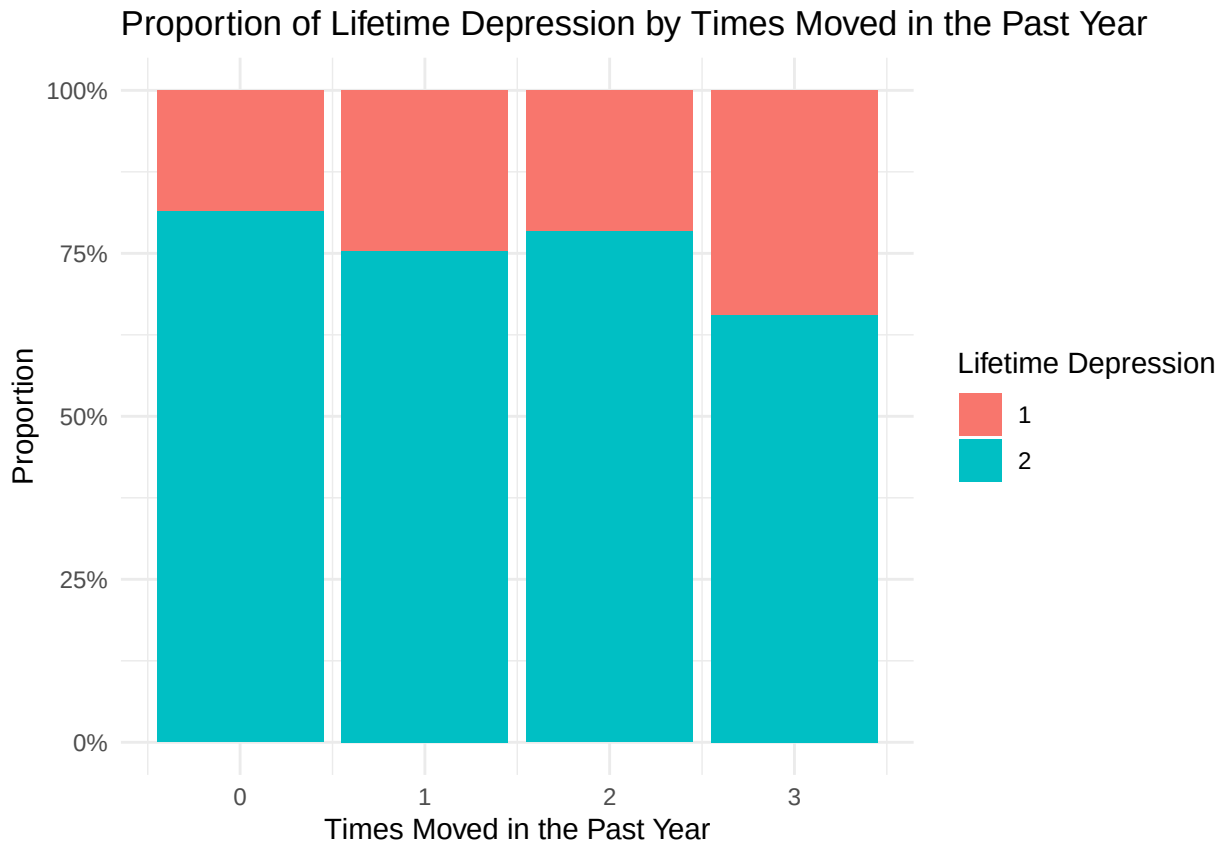
timesMovedPastYear	lifetimeDepression	count
0	1	6101
0	2	26682
1	1	1711
1	2	5227
2	1	585
2	2	2118
3	1	306
3	2	583

```
povertyDepData_cleanedMove2$lifetimeDepression = as.factor(povertyDepData_cleanedMove2$lifetimeDepressi

povertyDepData_cleanedMove2 %>%
  group_by(timesMovedPastYear, lifetimeDepression) %>%
  summarise(count = n(), .groups = 'drop') %>%
  ggplot(aes(x = timesMovedPastYear, y = count, fill = lifetimeDepression)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Times Moved in the Past Year vs. Lifetime Depression",
    x = "Times Moved in the Past Year",
    y = "Count",
    fill = "Lifetime Depression"
  ) +
  theme_minimal()
```




```
povertyDepData_cleanedMove2 %>%
  ggplot(aes(x = timesMovedPastYear, fill = lifetimeDepression)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  labs(
    title = "Proportion of Lifetime Depression by Times Moved in the Past Year",
    x = "Times Moved in the Past Year",
    y = "Proportion",
    fill = "Lifetime Depression"
  ) +
  theme_minimal()
```




```
table_counts <- table(povertyDepData_cleanedMove2$lifetimeDepression,
                      povertyDepData_cleanedMove2$timesMovedPastYear)

table_percent <- round(prop.table(table_counts, margin = 2) * 100, 2)

table_percent
```

```
##
##         0     1     2     3
##   1 18.61 24.66 21.64 34.42
##   2 81.39 75.34 78.36 65.58
```

EDA - education

```
povertyDepData_cleanedEDU <- povertyDepData[!is.na(povertyDepData$lifetimeDepression), ]
```

```
povertyDepData_cleanedEDU %>%
  group_by(highestEducationCompleted) %>%
  summarise(
    percent = round(n()*100 / nrow(povertyDepData_cleanedEDU),2)
  ) %>%
  kable()
```

highestEducationCompleted	percent
1	0.72
2	0.56
3	0.23
4	0.71
5	1.40
6	1.70
7	5.62
8	26.64
9	20.16
10	8.82
11	33.44