Descriptive Analysis of a Multilevel Data Set

Katherine Wolf

February 20, 2020

Overall report

Detailed summary statistics are in Table 1. The univariate analyses show a dataset of 4000 study participants, of whom 40 were missing data on age, 112 were missing data on race, 77 were missing data on education, and 580 were missing data on income.

Participants appear unevenly distributed among the boroughs, with Staten Island particularly poorly represented. Middle-aged participants between 35 and 54 appear somewhat more represented than those in other groups. A plurality of participants are white (40%), followed by African American (26.4%) and Hispanic/Latinx (24.0%). The most common educational level was completed high school. A plurality of participants (40%) had the lowest income level. (Additional analyses should break out joint covariate distributions, for example, neighborhood income and percentage in poverty by race.)

In the bivariate analyses, 89% (n = 3562) of participants did not binge drink and 11% (n = 438) did. Neighborhood income and neighborhood percent in poverty look similar between the two groups, although they are highly correlated (Pearson correlation coefficient -0.89), indicating that I should not use them both in one model. Binge drinking appeared particularly prevalent among participants in Manhattan compared to other boroughs and in the 25-34 year-old age group (subsequently declining with age), among college graduates, and at higher income levels.

Table 1: Descriptive statistics for participants in NYSES cohort, stratified by binge drinking.

	Level	No binge drinking	Binge drinking
n		3562	438
Neighborhood median income ($) (mean (SD))		40139.3 (14740.2)	42332.7 (16865.5)
Neighborhood poverty (%) (mean (SD))		20.8 (10.6)	20.6 (11.3)
Borough (%)	Bronx	550 (15.4)	58 (13.2)
	Brooklyn	1079 (30.3)	119 (27.2)
	Manhattan	707 (19.8)	139 (31.7)
	Queens	1036 (29.1)	98 (22.4)
	Staten Island	190 (5.3)	24 (5.5)
Age (years) (%)	18-24	284 (8.0)	66 (15.1)
	25-34	540 (15.2)	145 (33.1)
	35-44	711 (20.0)	104 (23.7)
	45-54	751 (21.1)	57 (13.0)
	55-64	570 (16.0)	42 (9.6)
	65+	667 (18.7)	23 (5.3)
	NA	39 (1.1)	1 (0.2)
Race/ethnicity (%)	White	1382 (38.8)	234 (53.4)
	African American	968 (27.2)	87 (19.9)
	Asian	144 (4.0)	20 (4.6)
	Hispanic/Latinx	877 (24.6)	81 (18.5)
	Other	85 (2.4)	10 (2.3)
	NA	106 (3.0)	6 (1.4)
Education (%)	Less than high school	467 (13.1)	41 (9.4)
	High school/GED	836 (23.5)	87 (19.9)
	Some college	781 (21.9)	98 (22.4)
	College graduate	758 (21.3)	125 (28.5)
	Graduate work	647 (18.2)	83 (18.9)
	NA	73 (2.0)	4 (0.9)
Income ($) (%)	<= 40,000	1460 (41.0)	145 (33.1)
	40,001 to 80,000	938 (26.3)	155 (35.4)
	> 80,000	612 (17.2)	110 (25.1)
	NA	552 (15.5)	28 (6.4)

Univariate analyses

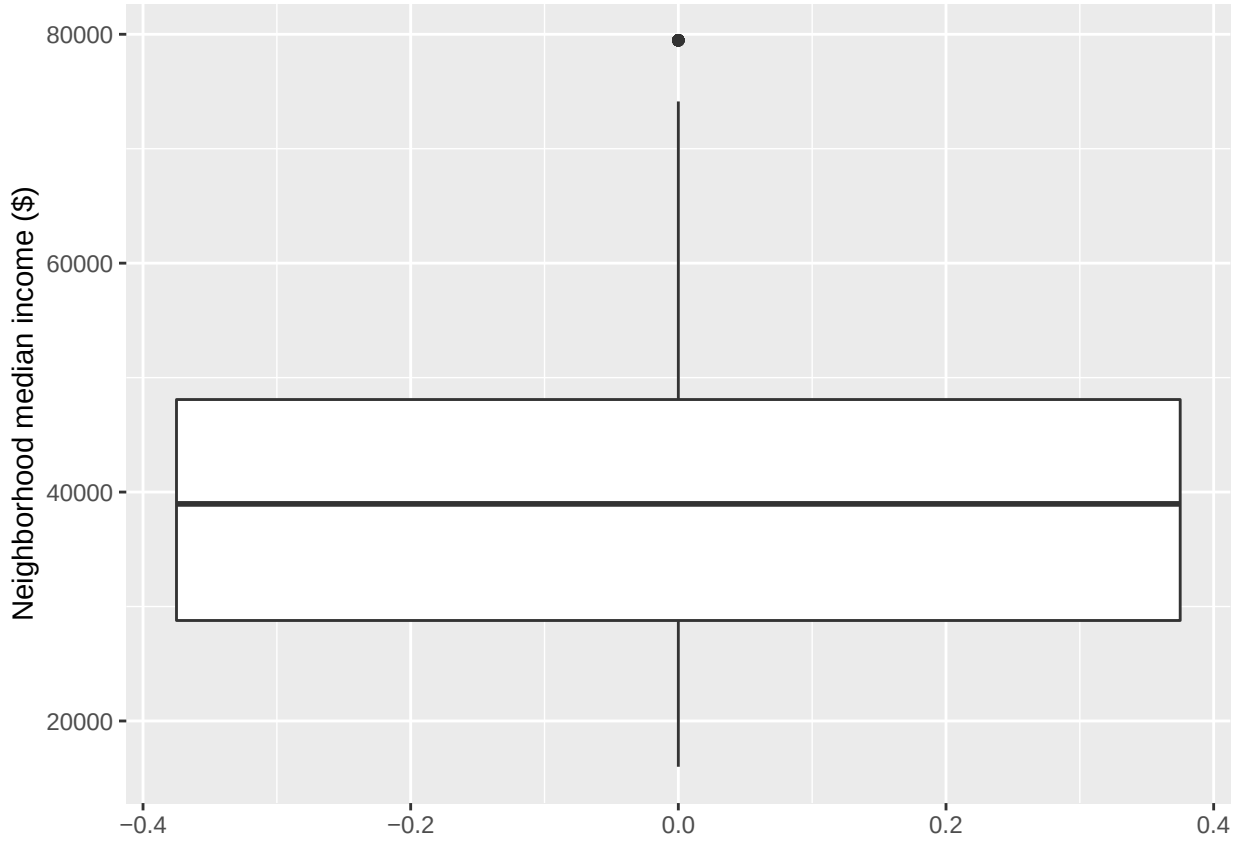
Numeric summaries

```
##        id         Community district Neighborhood median income ($)
##  Min.   :500004   Min.   :101.0      Min.   :16000
##  1st Qu.:514630   1st Qu.:206.0      1st Qu.:28780
##  Median :530555   Median :304.0      Median :38965
##  Mean   :530095   Mean   :286.4      Mean   :40379
##  3rd Qu.:545129   3rd Qu.:405.0      3rd Qu.:48085
##  Max.   :559997   Max.   :503.0      Max.   :79475
##
##  Neighborhood poverty (%)          Borough     Age (years)
##  Min.   : 4.90            Bronx        : 608   18-24:350
##  1st Qu.:11.40            Brooklyn     :1198   25-34:685
##  Median :19.11            Manhattan    : 846   35-44:815
```

```
##  Mean   :20.76            Queens       :1134   45-54:808
##  3rd Qu.:29.08            Staten Island: 214   55-64:612
##  Max.   :45.67                                 65+  :690
##                                                NA's : 40
##           Race/ethnicity                 Education              Income ($)
##  White           :1616   Less than high school:508   <= 40,000       :1605
##  African American:1055   High school/GED      :923   40,001 to 80,000:1093
##  Asian           : 164   Some college         :879   > 80,000        : 722
##  Hispanic/Latinx : 958   College graduate     :883   NA's            : 580
##  Other           :  95   Graduate work        :730
##  NA's            : 112   NA's                 : 77
##
##  Binge drinking
##  No :3562
##  Yes: 438
##
##
##
##
##
```

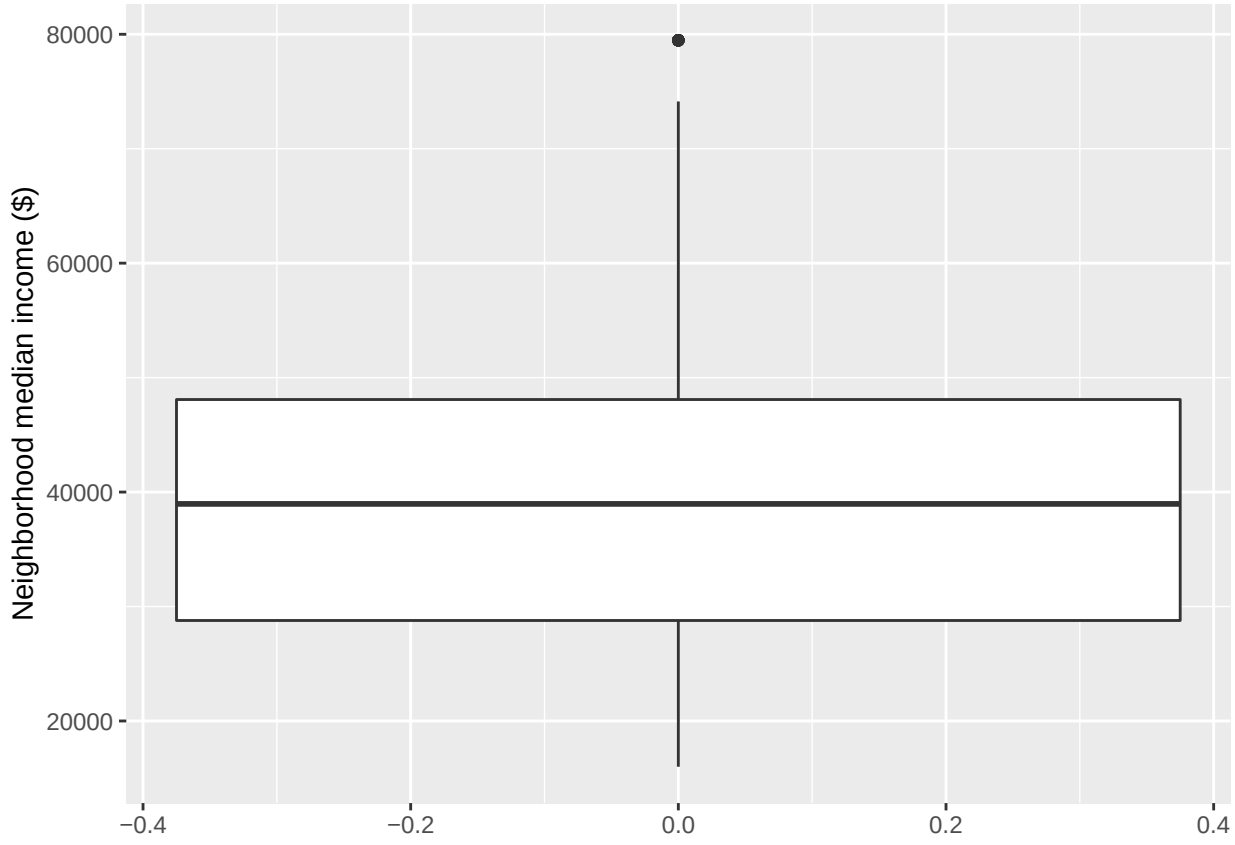
Neighborhood median income

```
## # A tibble: 1 x 5
##     Mean Median `Standard deviation` Minimum Maximum
##    <dbl>  <dbl>                <dbl>   <dbl>   <dbl>
## 1 40379.  38965               15001.   16000   79475
```



Neighborhood median poverty

```
## # A tibble: 1 x 5
##    Mean Median `Standard deviation` Minimum Maximum
##   <dbl>  <dbl>                <dbl>   <dbl>   <dbl>
## 1  20.8   19.1                 10.7    4.90    45.7
```

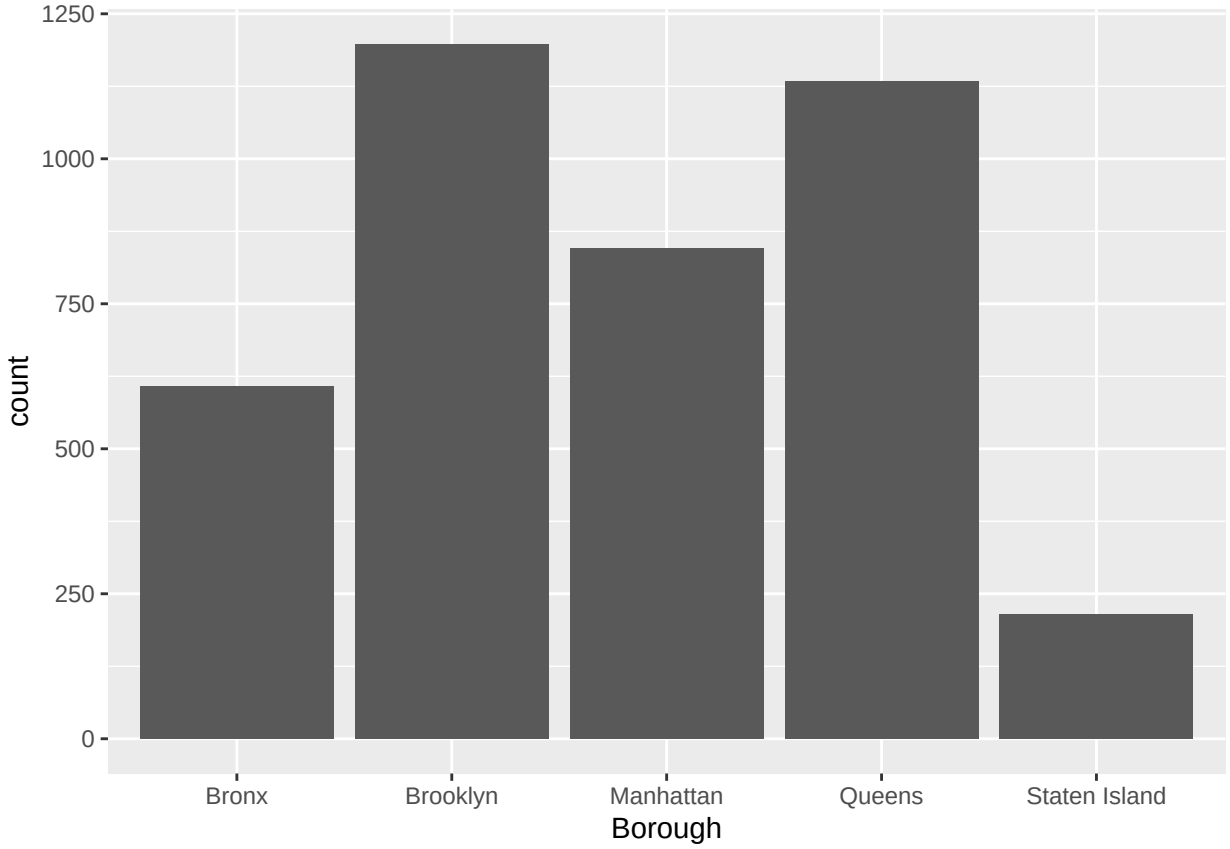


Borough

Table

```
## # A tibble: 5 x 3
##   Category      Count   Percentage
##   <chr>         <table> <table>
## 1 Bronx          608    15.2
## 2 Brooklyn      1198    30.0
## 3 Manhattan      846    21.2
## 4 Queens        1134    28.4
## 5 Staten Island  214     5.4
```

Bar plot

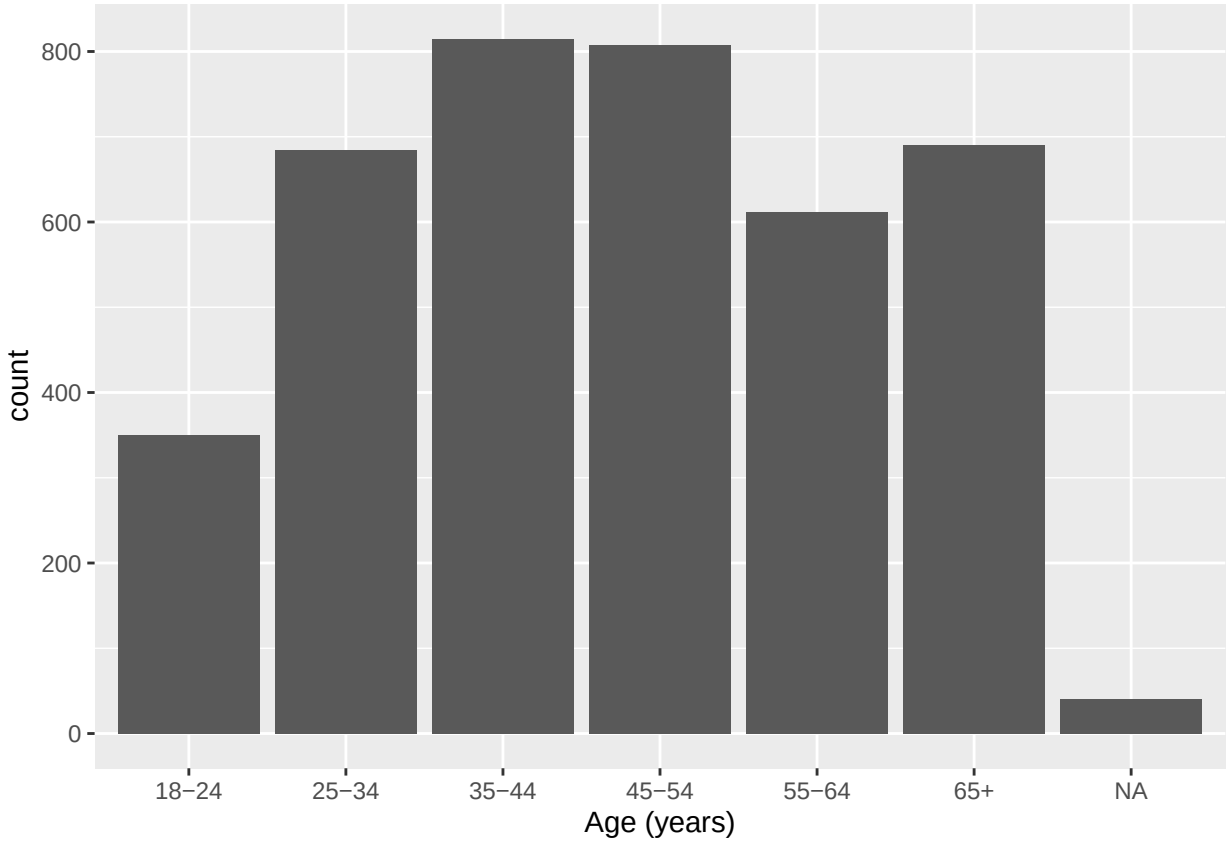


Age

Table

```
## # A tibble: 7 x 3
##   Category Count   Percentage
##   <chr>    <table> <table>
## 1 18-24    350      8.8
## 2 25-34    685     17.1
## 3 35-44    815     20.4
## 4 45-54    808     20.2
## 5 55-64    612     15.3
## 6 65+      690     17.2
## 7 NA        40      1.0
```

Bar plot

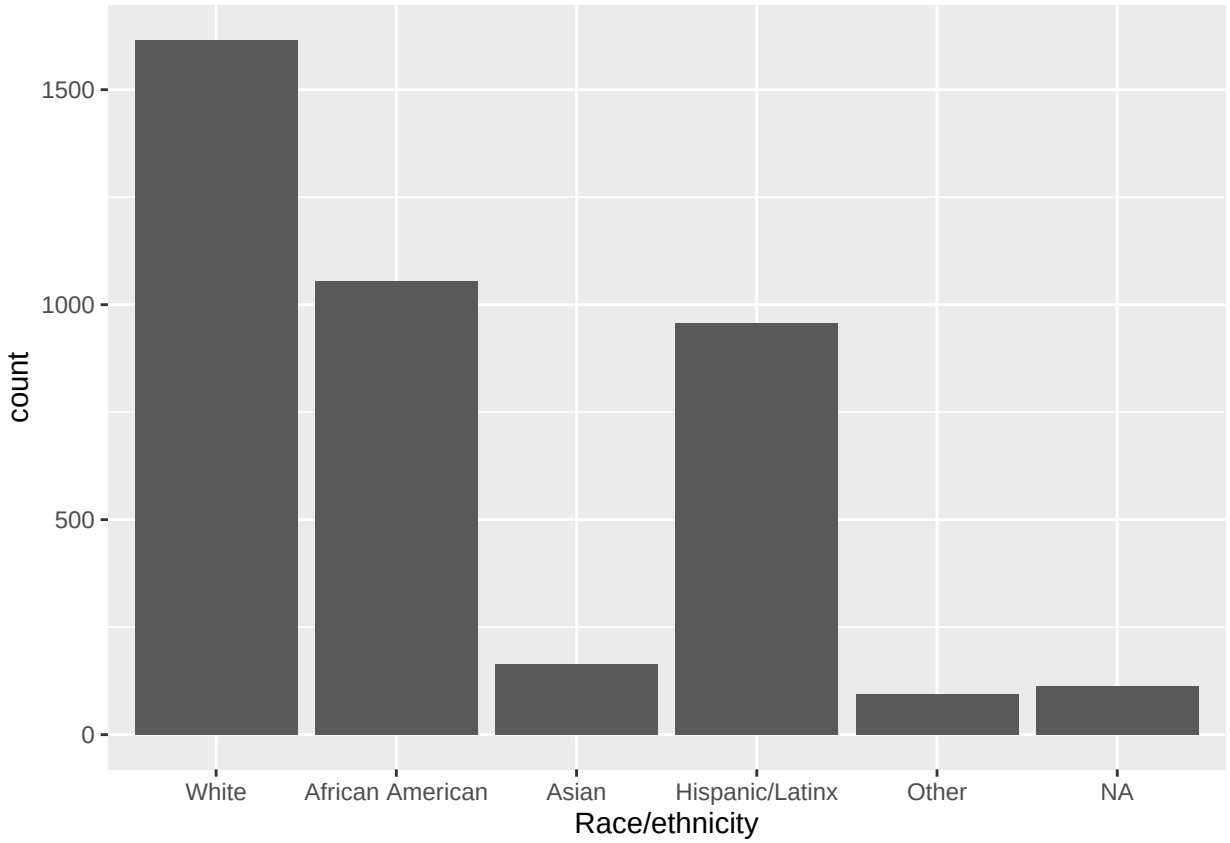


Race/ethnicity

Table

```
## # A tibble: 6 x 3
##   Category         Count   Percentage
##   <chr>            <table> <table>
## 1 White            1616    40.4
## 2 African American 1055    26.4
## 3 Asian             164     4.1
## 4 Hispanic/Latinx   958    24.0
## 5 Other              95     2.4
## 6 NA                112     2.8
```

Bar plot

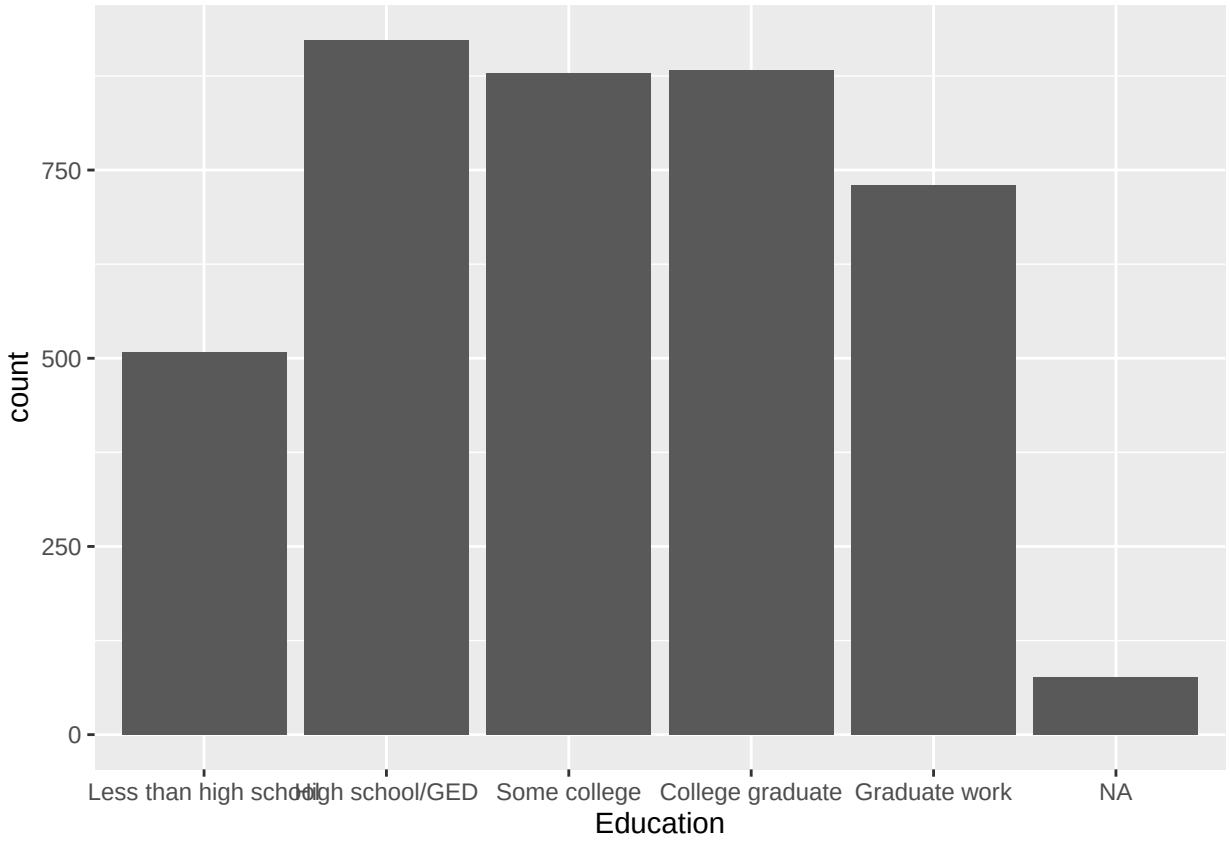


Education

Table

```
## # A tibble: 6 x 3
##   Education             Count   Percentage
##   <chr>                 <table> <table>
## 1 Less than high school 508     12.7
## 2 High school/GED       923     23.1
## 3 Some college          879     22.0
## 4 College graduate      883     22.1
## 5 Graduate work         730     18.2
## 6 NA                     77      1.9
```

Bar plot

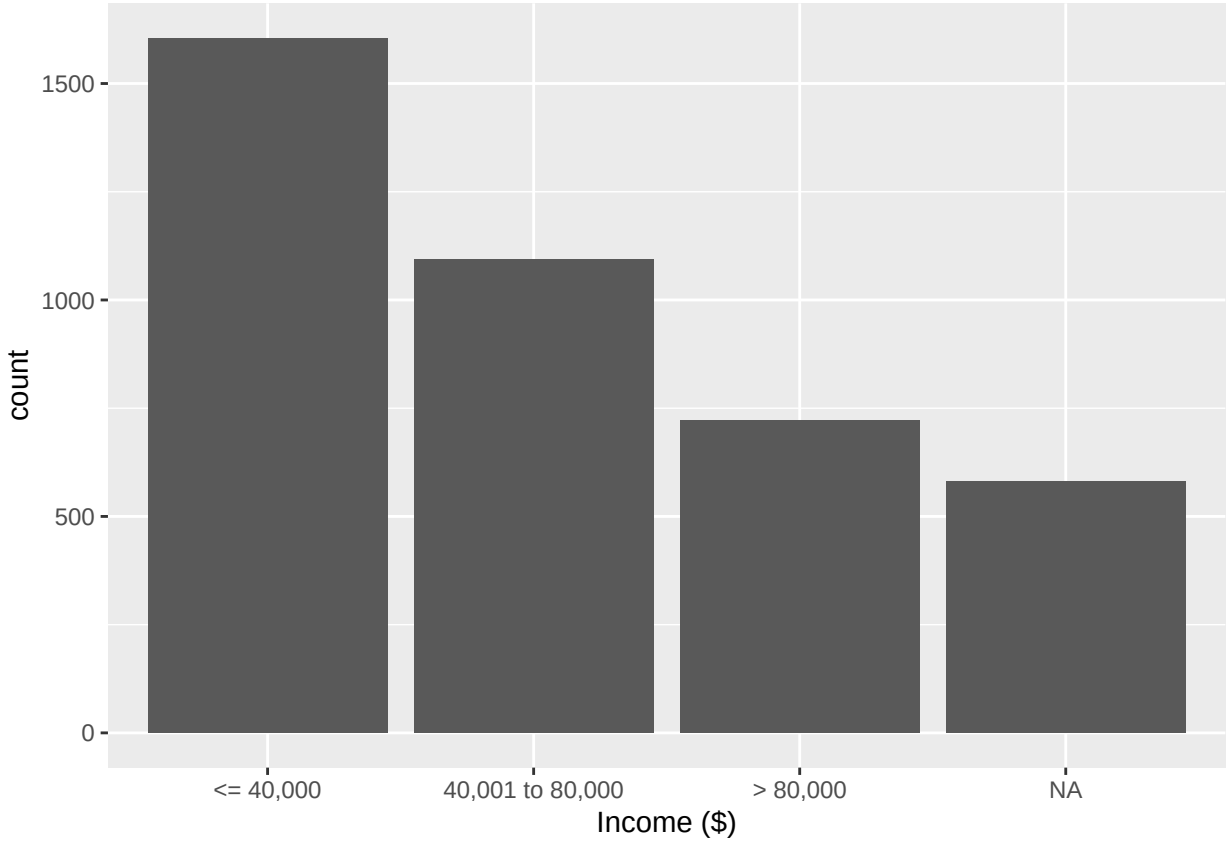


Income

Table

```
## # A tibble: 4 x 3
##   Category         Count   Percentage
##   <chr>            <table> <table>
## 1 <= 40,000        1605    40.1
## 2 40,001 to 80,000 1093    27.3
## 3 > 80,000          722    18.0
## 4 NA                580    14.5
```

Bar plot

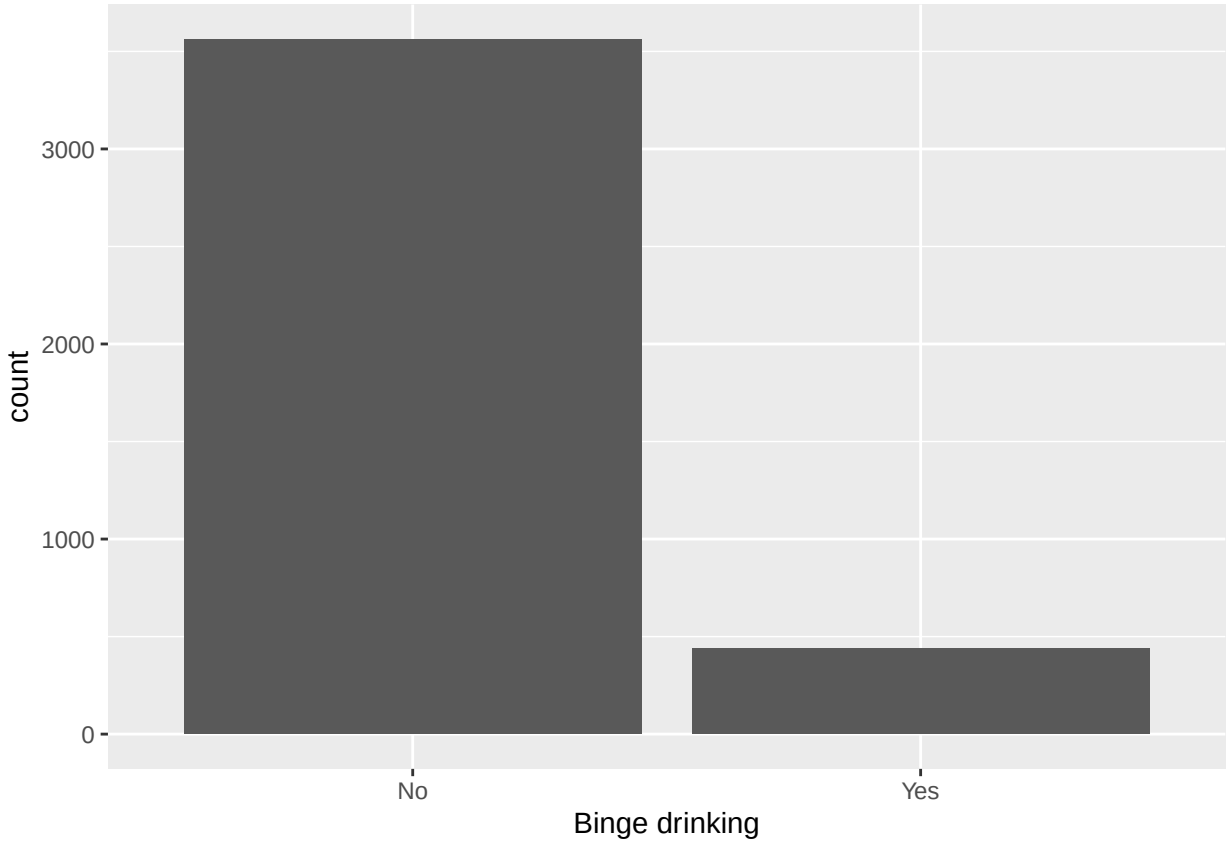


Binge drinking

Table

```
## # A tibble: 2 x 3
##   Category Count   Percentage
##   <chr>    <table> <table>
## 1 No       3562    89
## 2 Yes       438    11
```

Bar plot



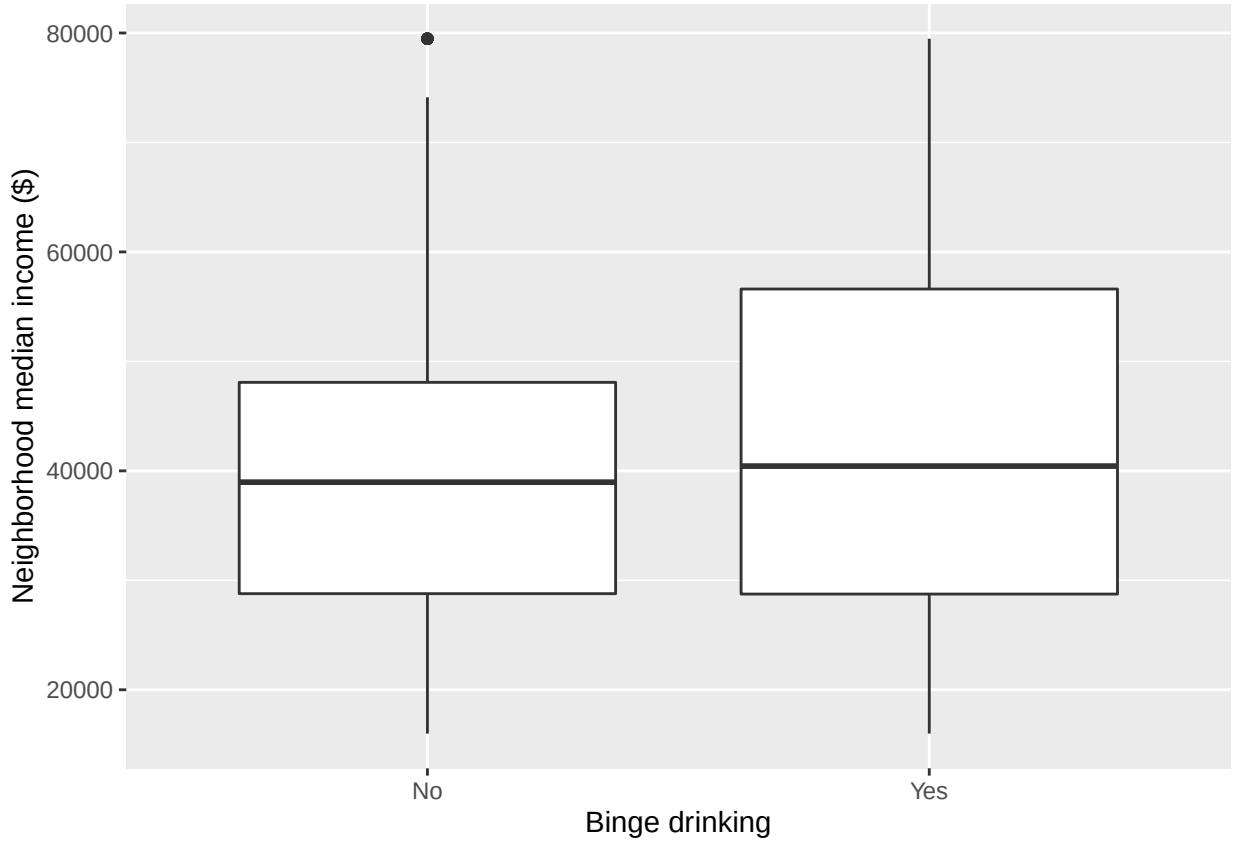
Bivariate analyses by outcome

Neighborhood median income

Table

```
## # A tibble: 2 x 6
##   `Binge drinking`   Mean Median `Standard deviation` Minimum Maximum
##   <fct>             <dbl>  <dbl>                <dbl>   <dbl>   <dbl>
## 1 No               40139.  38965               14740.   16000   79475
## 2 Yes              42333.  40435               16866.   16000   79475
```

Box plots

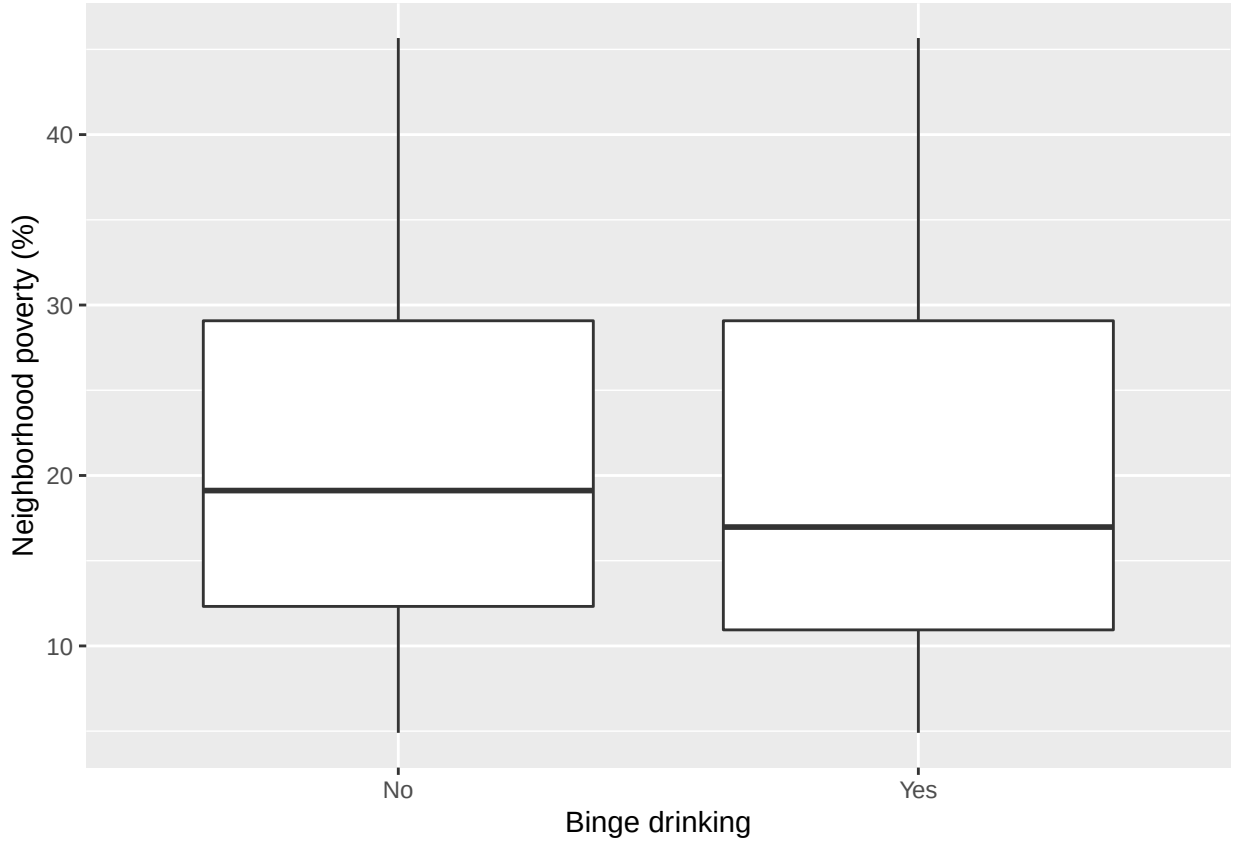


Neighborhood percent in poverty

Table

```
## # A tibble: 2 x 6
##   `Binge drinking`  Mean Median `Standard deviation` Minimum Maximum
##   <fct>            <dbl>  <dbl>                <dbl>   <dbl>   <dbl>
## 1 No                20.8   19.1                 10.6    4.90    45.7
## 2 Yes               20.6   17.0                 11.3    4.90    45.7
```

Box plots



Correlation between neighborhood poverty and neighborhood income

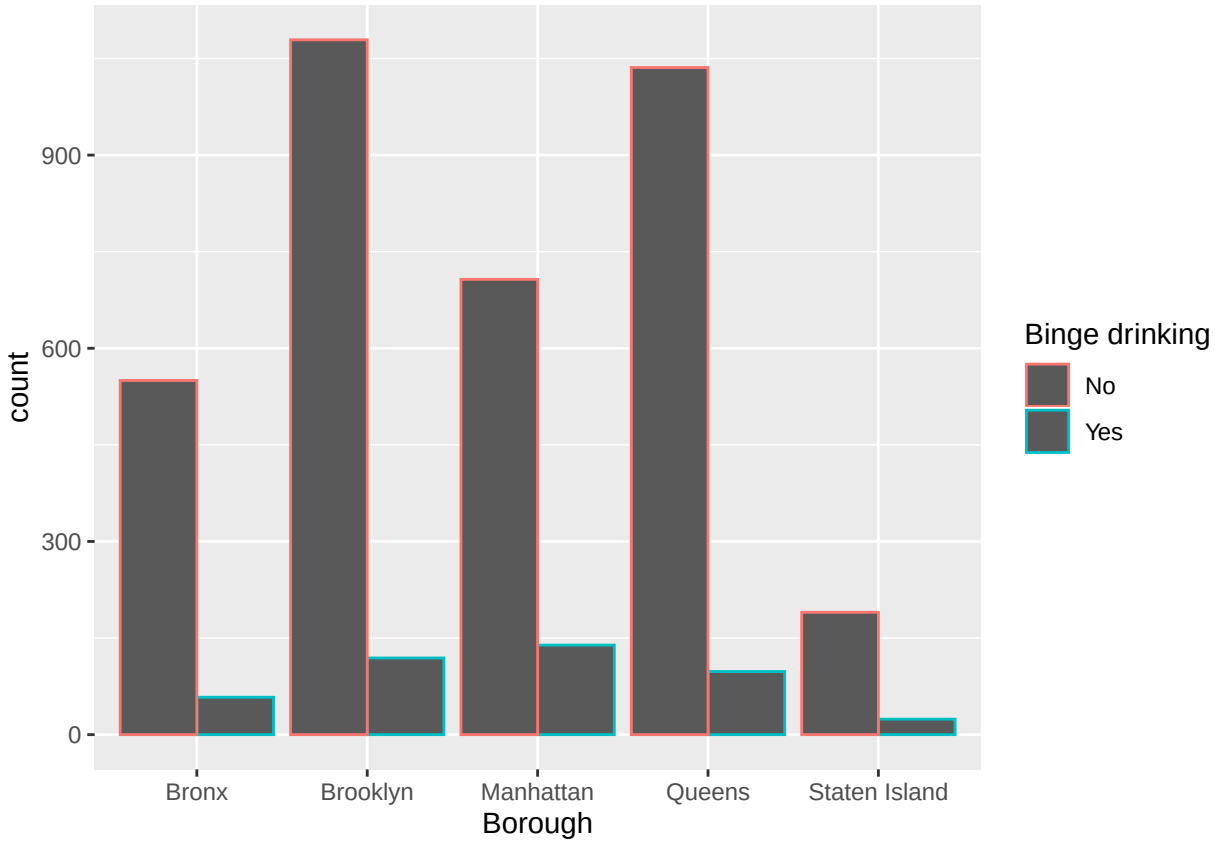
```
##
##  Pearson's product-moment correlation
##
## data:  nyses_analyze$`Neighborhood median income ($)` and nyses_analyze$`Neighborhood poverty (%)`
## t = -127.38, df = 3998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9016850 -0.8894228
## sample estimates:
##        cor
## -0.8957241
```

Borough

Table

```
## # A tibble: 5 x 7
##   Borough    No `% No (row)` `% No (column)`   Yes `% Yes (row)`
##   <chr>   <int>        <dbl>            <dbl> <int>         <dbl>
## 1 Bronx     550         90.5             15.4    58           9.5
## 2 Brookl~  1079         90.1             30.3   119           9.9
## 3 Manhat~   707         83.6             19.8   139          16.4
## 4 Queens   1036         91.4             29.1    98           8.6
## 5 Staten~   190         88.8              5.3    24          11.2
## # ... with 1 more variable: `% Yes (column)` <dbl>
```

Bar plot

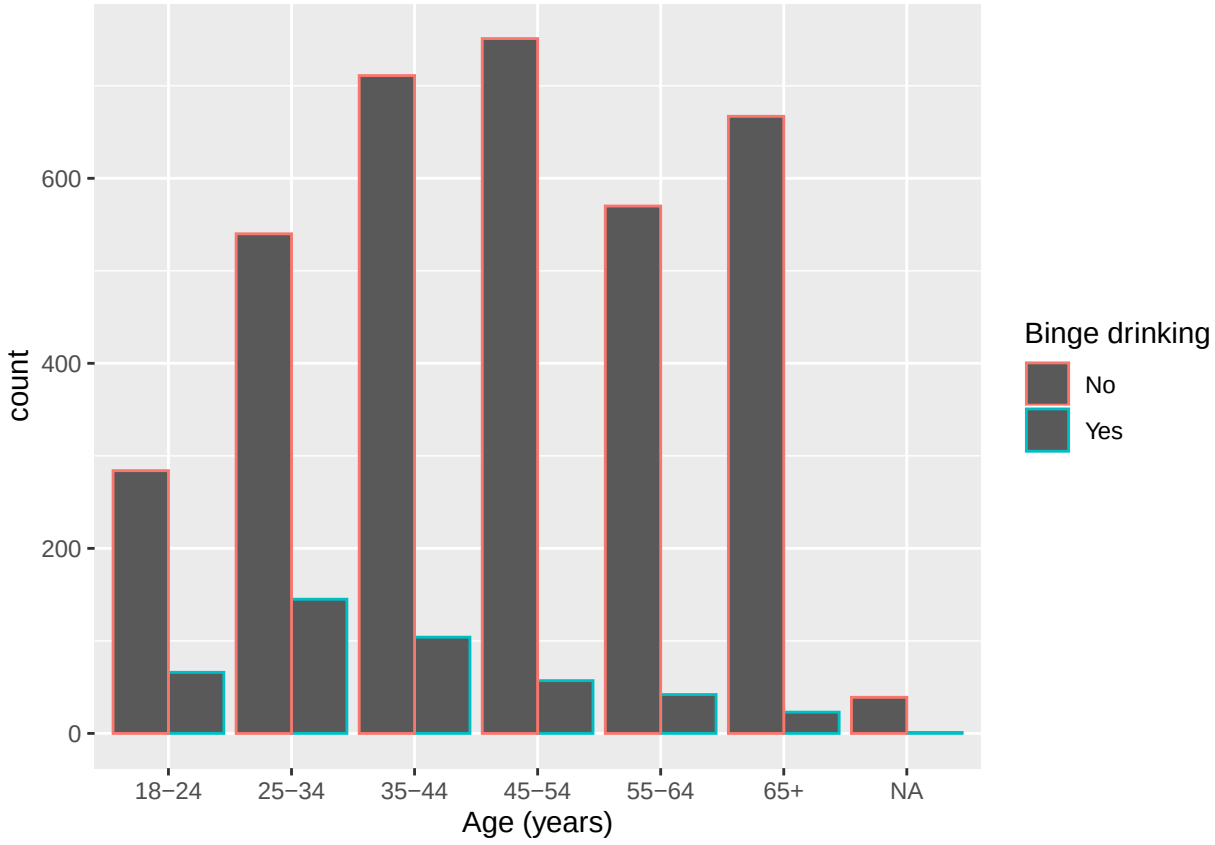


Age

Table

```
## # A tibble: 7 x 7
##   `Age (years)`    No `% No (row)` `% No (column)`   Yes `% Yes (row)`
##   <chr>         <int>        <dbl>            <dbl> <int>         <dbl>
## 1 18-24           284         81.1             8       66          18.9
## 2 25-34           540         78.8            15.2    145          21.2
## 3 35-44           711         87.2            20      104          12.8
## 4 45-54           751         92.9            21.1     57           7.1
## 5 55-64           570         93.1            16       42           6.9
## 6 65+             667         96.7            18.7     23           3.3
## 7 <NA>             39         97.5             1.10     1           2.5
## # ... with 1 more variable: `% Yes (column)` <dbl>
```

Bar plot

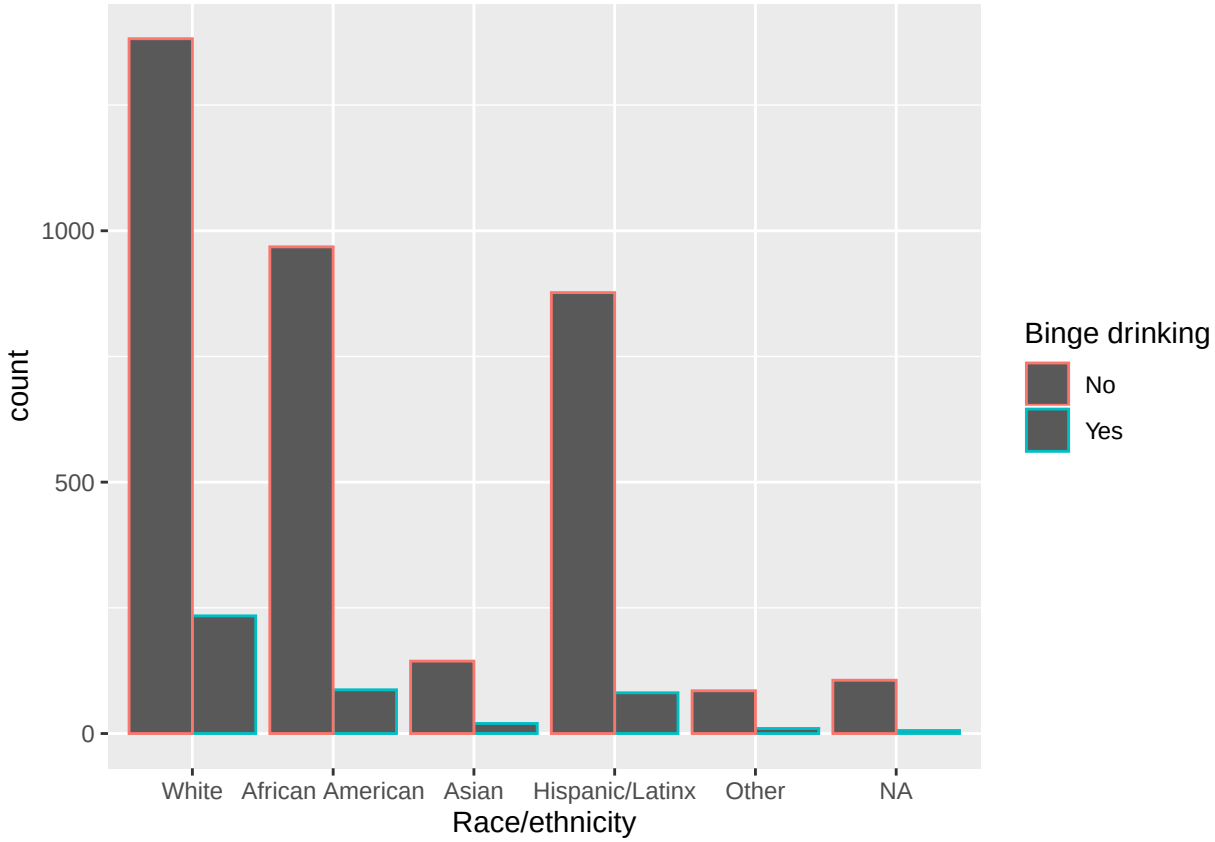


Race/ethnicity

Table

```
## # A tibble: 6 x 7
##   `Race/ethnicity`     No `% No (row)` `% No (column)`   Yes `% Yes (row)`
##   <chr>             <int>        <dbl>            <dbl> <int>         <dbl>
## 1 White              1382         85.5             38.8   234          14.5
## 2 African American    968         91.8             27.2    87           8.2
## 3 Asian               144         87.8              4      20          12.2
## 4 Hispanic/Latinx     877         91.5             24.6    81           8.5
## 5 Other                85         89.5              2.4    10          10.5
## 6 <NA>                106         94.6              3       6           5.4
## # ... with 1 more variable: `% Yes (column)` <dbl>
```

Bar plot

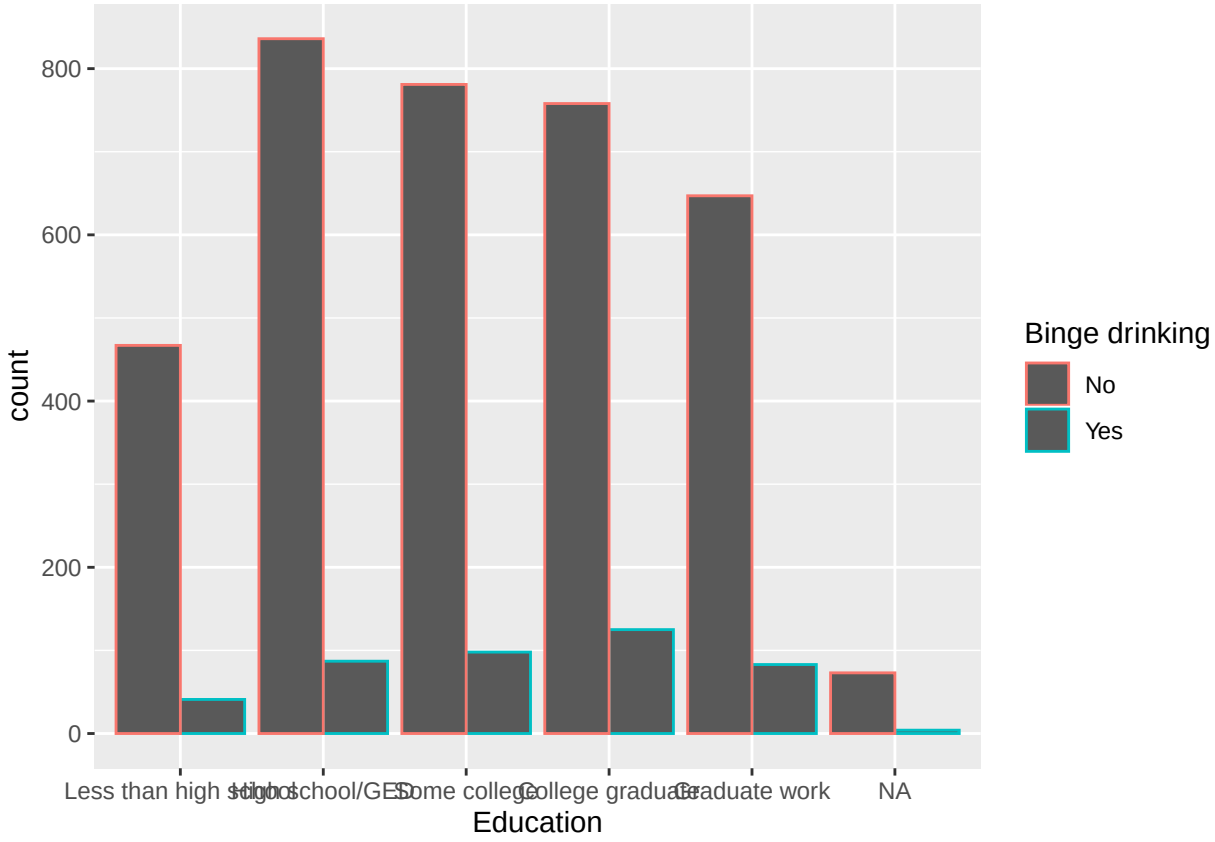


Education

Table

```
## # A tibble: 6 x 7
##   Education    No `% No (row)` `% No (column)`   Yes `% Yes (row)`
##   <chr>     <int>        <dbl>            <dbl> <int>         <dbl>
## 1 Less tha~   467         91.9             13.1    41           8.1
## 2 High sch~   836         90.6             23.5    87           9.4
## 3 Some col~   781         88.9             21.9    98          11.1
## 4 College ~   758         85.8             21.3   125          14.2
## 5 Graduate~   647         88.6             18.2    83          11.4
## 6 <NA>         73         94.8              2       4           5.2
## # ... with 1 more variable: `% Yes (column)` <dbl>
```

Bar plot

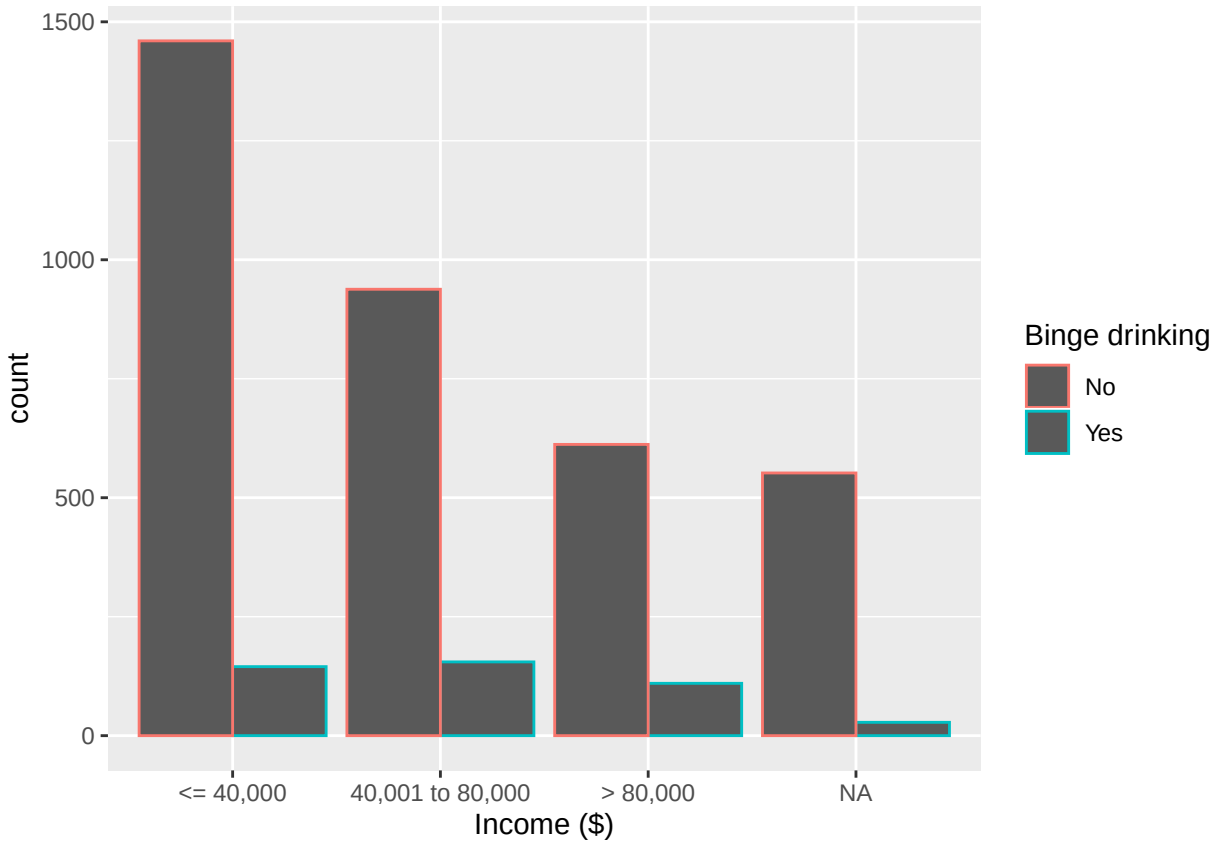


Income

Table

```
## # A tibble: 4 x 7
##   Income    No `% No (row)` `% No (column)`   Yes `% Yes (row)` `% Yes (column)`
##   <chr>  <int>        <dbl>           <dbl> <int>         <dbl>             <dbl>
## 1 <= 40~  1460         91              41     145          9                33.1
## 2 40,00~   938         85.8            26.3   155         14.2              35.4
## 3 > 80,~   612         84.8            17.2   110         15.2              25.1
## 4 <NA>     552         95.2            15.5    28          4.8               6.4
```

Bar plot




```
knitr::opts_chunk$set(echo = FALSE,
                      warning = FALSE,
                      message = FALSE)


# need to call the libraries every time you begin a new R session
library(dplyr)
library(ggplot2)
library(nnet)
library(tidyverse)
library(tableone)
```

```r
library(xtable)
library(knitr)
library(tableone)
library(kableExtra)

# read in data - suppose the file dataset.csv contains continuous variables var1 and var2, and a binary
nyses_raw_data <- read_csv("NYSES data for class.csv")

# make working data file
nyses_to_edit <- nyses_raw_data

# id
nyses_to_edit$id <- nyses_to_edit$QKEY2

# community district
nyses_to_edit$`Community district` <- nyses_to_edit$cd

# neighborhood median income
nyses_to_edit$`Neighborhood median income ($)` <- nyses_to_edit$medinc

# neighborhood percent below poverty
nyses_to_edit$`Neighborhood poverty (%)` <- nyses_to_edit$pbelowpv

# borough
borough_labels <-
  c("Bronx",
    "Brooklyn",
    "Manhattan",
    "Queens",
    "Staten Island") # (value = order)

nyses_to_edit$Borough <-
  factor(nyses_to_edit$boro,
         labels = borough_labels)

# age
age_labels <-
  c("18-24",
    "25-34",
    "35-44",
    "45-54",
    "55-64",
    "65+") # (value = order)

nyses_to_edit$`Age (years)` <-
  factor(nyses_to_edit$agecat,
         labels = age_labels)

# race/ethnicity
race_labels <-
  c("White",
    "African American",
    "Asian",
```

```
    "Hispanic/Latinx",
    "Other") # (value = order)

nyses_to_edit$`Race/ethnicity` <-
  factor(nyses_to_edit$racecat, labels = race_labels)

# education
ed_labels <-
  c("Less than high school",
    "High school/GED",
    "Some college",
    "College graduate",
    "Graduate work") # (value = order)

nyses_to_edit$`Education` <-
  factor(nyses_to_edit$edcat, labels = ed_labels)

# income
income_labels <-
  c("<= 40,000",
    "40,001 to 80,000",
    "> 80,000") # (value = order)

nyses_to_edit$`Income ($)` <-
  factor(nyses_to_edit$inc3cat, labels = income_labels)

# binge drinking
binge_labels <-
  c("No",
    "Yes") # (value = order)

nyses_to_edit$`Binge drinking` <-
  factor(nyses_to_edit$binge,
         labels = binge_labels)

nyses_analyze <-
  nyses_to_edit %>%
  select(id,
         `Community district`,
         `Neighborhood median income ($)`,
         `Neighborhood poverty (%)`,
         Borough,
         `Age (years)`,
         `Race/ethnicity`,
         Education,
         `Income ($)`,
         `Binge drinking`)


# create a list of variables for the table
# (not including the stratification variable)
table_one_variables <- c("Neighborhood median income ($)",
                         "Neighborhood poverty (%)",
```

```
                          "Borough",
                          "Age (years)",
                          "Race/ethnicity",
                          "Education",
                          "Income ($)")

# create a list of which ones are categorical (factor)
factor_variables <- c("Borough",
                      "Age (years)",
                      "Race/ethnicity",
                      "Education",
                      "Income ($)")

table_1 <- CreateTableOne(vars = table_one_variables,
                          factorVars = factor_variables,
                          strata = "Binge drinking",
                          data = nyses_analyze,
                          test = FALSE,
                          includeNA = TRUE)

save(table_1,
     file = "table_1.rdata")

# print(table.1) # Standard output

# Creates a formatted table, using kable from the knitr package
# Would want to clean this up for publication purposes:
hi <- kable(print(table_1,
                  showAllLevels = TRUE,
                  printToggle = FALSE,
                  noSpaces = TRUE,
                  catDigits = 1,
                  contDigits = 1),
            col.names = c("Level", "No binge drinking", "Binge drinking"),
            caption=paste("Descriptive statistics for participants in",
                          "NYSES cohort, stratified by binge drinking."))


# hi <- kable(print(table_1,
#                   showAllLevels = TRUE,
#                   printToggle = FALSE,
#                   noSpaces = TRUE,
#                   catDigits=1,
#                   contDigits=1),
#       caption=paste("Descriptive statistics for participants in",
#                     "NYSES cohort, stratified by binge drinking."))

hi


nyses_analyze %>%
  summary()
```

```r
# neighborhood median income
nyses_analyze %>%
  summarise(Mean = mean(`Neighborhood median income ($)`),
            Median = median(`Neighborhood median income ($)`),
            `Standard deviation` = sd(`Neighborhood median income ($)`),
            Minimum = range(`Neighborhood median income ($)`)[1],
            Maximum = range(`Neighborhood median income ($)`)[2])

# neighborhood median income
ggplot(data = nyses_analyze,
       aes(y = `Neighborhood median income ($)`)) +
  geom_boxplot()


# neighborhood percent in poverty
nyses_analyze %>%
  ungroup() %>%
  summarise(Mean = mean(`Neighborhood poverty (%)`),
            Median = median(`Neighborhood poverty (%)`),
            `Standard deviation` = sd(`Neighborhood poverty (%)`),
            Minimum = range(`Neighborhood poverty (%)`)[1],
            Maximum = range(`Neighborhood poverty (%)`)[2])

# neighborhood median income
ggplot(data = nyses_analyze,
       aes(y = `Neighborhood median income ($)`)) +
  geom_boxplot()


borough_table <-
  tibble(Category = c(borough_labels),
         Count = table(nyses_analyze$Borough, useNA = "ifany"),
         Percentage = round(prop.table(table(nyses_analyze$Borough,
                                             useNA = "ifany")),3)*100)

borough_table


nyses_analyze %>%
  ggplot(aes(x = Borough)) +
  geom_bar()


age_table <-
  tibble(Category = c(age_labels, "NA"),
         Count = table(nyses_analyze$`Age (years)`, useNA = "ifany"),
         Percentage = round(prop.table(table(nyses_analyze$`Age (years)`, useNA = "ifany")),3)*100)

age_table


nyses_analyze %>%
  ggplot(aes(x = `Age (years)`)) +
```

```
  geom_bar()


re_table <-
  tibble(Category = c(race_labels, "NA"),
         Count = table(nyses_analyze$`Race/ethnicity`, useNA = "ifany"),
         Percentage = round(prop.table(table(nyses_analyze$`Race/ethnicity`,
                                             useNA = "ifany")),3)*100)

re_table


nyses_analyze %>%
  ggplot(aes(x = `Race/ethnicity`)) +
  geom_bar()


ed_table <-
  tibble(Education = c(ed_labels, "NA"),
         Count = table(nyses_analyze$Education, useNA = "ifany"),
         Percentage = round(prop.table(table(nyses_analyze$Education,
                                             useNA = "ifany")),3)*100)

ed_table


nyses_analyze %>%
  ggplot(aes(x = Education)) +
  geom_bar()


income_table <-
  tibble(Category = c(income_labels, "NA"),
         Count = table(nyses_analyze$`Income ($)`, useNA = "ifany"),
         Percentage = round(prop.table(table(nyses_analyze$`Income ($)`,
                                             useNA = "ifany")),3)*100)

income_table


nyses_analyze %>%
  ggplot(aes(x = `Income ($)`)) +
  geom_bar()


binge_table <-
  tibble(Category = c(binge_labels),
         Count = table(nyses_analyze$`Binge drinking`, useNA = "ifany"),
         Percentage = round(prop.table(table(nyses_analyze$`Binge drinking`,
                                             useNA = "ifany")),3)*100)

binge_table
```

```r
nyses_analyze %>%
  ggplot(aes(x = `Binge drinking`)) +
  geom_bar()


# neighborhood median income
nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  summarise(Mean = mean(`Neighborhood median income ($)`),
            Median = median(`Neighborhood median income ($)`),
            `Standard deviation` = sd(`Neighborhood median income ($)`),
            Minimum = range(`Neighborhood median income ($)`)[1],
            Maximum = range(`Neighborhood median income ($)`)[2])


# neighborhood median income
ggplot(data = nyses_to_edit,
       aes(x = `Binge drinking`,
           y = `Neighborhood median income ($)`)) +
  geom_boxplot()


# neighborhood percent in poverty
nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  summarise(Mean = mean(`Neighborhood poverty (%)`),
            Median = median(`Neighborhood poverty (%)`),
            `Standard deviation` = sd(`Neighborhood poverty (%)`),
            Minimum = range(`Neighborhood poverty (%)`)[1],
            Maximum = range(`Neighborhood poverty (%)`)[2])


# neighborhood median income
ggplot(data = nyses_to_edit,
       aes(x = `Binge drinking`,
           y = `Neighborhood poverty (%)`)) +
  geom_boxplot()


cor.test(nyses_analyze$`Neighborhood median income ($)`,
         nyses_analyze$`Neighborhood poverty (%)`)


basic_two_borough <- table(nyses_analyze$Borough,
                           nyses_analyze$`Binge drinking`,
                           useNA = "ifany")

prop_col_borough <- round(prop.table(basic_two_borough, 2), 3) * 100

prop_row_borough <- round(prop.table(basic_two_borough, 1), 3) * 100

full_two_table_borough <-
  tibble(Borough = c(borough_labels),
```

```
         No = basic_two_borough[,1],
         `% No (row)` = prop_row_borough[,1],
         `% No (column)` = prop_col_borough[,1],
         Yes = basic_two_borough[,2],
         `% Yes (row)` = prop_row_borough[,2],
         `% Yes (column)` = prop_col_borough[,2])

full_two_table_borough


nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  ggplot(aes(x = Borough, col = `Binge drinking`)) +
  geom_bar(position = position_dodge())


basic_two_age <- table(nyses_analyze$`Age (years)`,
                           nyses_analyze$`Binge drinking`,
                           useNA = "always")

prop_col_age <- round(prop.table(basic_two_age, 2), 3) * 100

prop_row_age <- round(prop.table(basic_two_age, 1), 3) * 100

full_two_table_age <-
  tibble(`Age (years)` = c(age_labels, NA),
         No = basic_two_age[,1],
         `% No (row)` = prop_row_age[,1],
         `% No (column)` = prop_col_age[,1],
         Yes = basic_two_age[,2],
         `% Yes (row)` = prop_row_age[,2],
         `% Yes (column)` = prop_col_age[,2])

full_two_table_age


nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  ggplot(aes(x = `Age (years)`, col = `Binge drinking`)) +
  geom_bar(position = position_dodge())


basic_two_race <- table(nyses_analyze$`Race/ethnicity`,
                           nyses_analyze$`Binge drinking`,
                           useNA = "ifany")

prop_col_race <- round(prop.table(basic_two_race, 2), 3) * 100

prop_row_race <- round(prop.table(basic_two_race, 1), 3) * 100

full_two_table_race <-
  tibble(`Race/ethnicity` = c(race_labels, NA),
         No = basic_two_race[,1],
```

```
         `% No (row)` = prop_row_race[,1],
         `% No (column)` = prop_col_race[,1],
         Yes = basic_two_race[,2],
         `% Yes (row)` = prop_row_race[,2],
         `% Yes (column)` = prop_col_race[,2])

full_two_table_race


nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  ggplot(aes(x = `Race/ethnicity`, col = `Binge drinking`)) +
  geom_bar(position = position_dodge())


basic_two_education <- table(nyses_analyze$Education,
                             nyses_analyze$`Binge drinking`,
                             useNA = "ifany")

prop_col_education <- round(prop.table(basic_two_education, 2), 3) * 100

prop_row_education <- round(prop.table(basic_two_education, 1), 3) * 100

full_two_table_education <-
  tibble(Education = c(ed_labels, NA),
         No = basic_two_education[,1],
         `% No (row)` = prop_row_education[,1],
         `% No (column)` = prop_col_education[,1],
         Yes = basic_two_education[,2],
         `% Yes (row)` = prop_row_education[,2],
         `% Yes (column)` = prop_col_education[,2])

full_two_table_education


nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  ggplot(aes(x = Education, col = `Binge drinking`)) +
  geom_bar(position = position_dodge())


basic_two_income <- table(nyses_analyze$`Income ($)`,
                             nyses_analyze$`Binge drinking`,
                             useNA = "ifany")

prop_col_income <- round(prop.table(basic_two_income, 2), 3) * 100

prop_row_income <- round(prop.table(basic_two_income, 1), 3) * 100

full_two_table_income <-
  tibble(Income = c(income_labels, NA),
         No = basic_two_income[,1],
         `% No (row)` = prop_row_income[,1],
```

```
         `% No (column)` = prop_col_income[,1],
         Yes = basic_two_income[,2],
         `% Yes (row)` = prop_row_income[,2],
         `% Yes (column)` = prop_col_income[,2])

full_two_table_income


nyses_analyze %>%
  group_by(`Binge drinking`) %>%
  ggplot(aes(x = `Income ($)`, col = `Binge drinking`)) +
  geom_bar(position = position_dodge())
```