

1. I imported the data from the website given in CANVAS and then used library(haven) since it was a SAS file. I also opened the file in SAS Studio to observe the variables. Using knitr::opts_chunk\$set() to change the default values of chunk options in the SAS file.
2. The variables that were needed were outputted, renamed and re-coded. This was done using the dplyr package.
3. We defined the survey and apply weights to the data set for the study. In the data set the questions that were asked (pulled out) were that, Have you had an HIV test in the last year?, The last time you had sexual intercourse, did you use a condom?, The age group?, Race?, and, How many sexual partners have you had in the last year, to date?.
4. The hypothesis we are observing in this dataset for the variables stated above is where the use of condoms has any association with HIV testing from individuals in the study. We have found that there is no specific association for condom use and HIV testing within this data set in this study for those individuals. The confounders such as agegrp, newrace, and sexpartner that have some type of association with HIV testing. Associated with HIV testing are the newrace variable (ethnicities: black & hispanic) which is shown in the logistic regression model below. This model also shows that the individuals who were 45 years of age and older have an confounder association with HIV testing.
5. The data set was cleaned; this outputted Table 1 as well as appropriate graphs which depicted said information. This was done using the tidyr and kableExtra package. Table 1 shows the data set of the population within the study. In total, there are 5026 samples (individuals). White, North African, Middle Eastern, Non-Hispanic - 33.6% Hispanic - 30.8% Black, Non-Hispanic - 20.1% Asian/PI, non-Hispanic - 12.3% Other, non-Hispanic - 3.3% Above is a breakdown of what is displayed in the data in terms of Race and Ethnicity (newrace variable).

18-24 - 7.4% 25-44 - 47.5% 45-64 - 34.4% 65+ - 10.7% Above shows the age groups (agegrp variable) within the study. Majority of the study contains ages ranging from 25-64.

Out of 5026 samples, 35.5% of them have done HIV testing within the last 12 months and 29.9% of them have had condom use in the last 12 months.

81.3% of the 5026 samples have had 1 sexual partner during the last 12 months.

6. Table and visuals. Regression analysis proved my hypothesis wrong which was if there is association between condom use and HIV testing.

```
library(haven)
chs21 <- read_sas("https://www1.nyc.gov/assets/doh/downloads/sas/episrv/chs2020_public.sas7bdat")
```

```
library(dplyr)
chs21_recode <-
  transform(chs21,
    newrace = recode_factor(newrace, `1`="White/N Afri/MidEastern, non-Hispanic",
                                `2`="Black, non-Hispanic",
                                `3`="Hispanic",
                                `4`="Asian/PI, non-Hispanic",
                                `5`="Other, non-Hispanic"),
    agegroup = recode_factor(agegroup,
                            `1` = "18-24",
                            `2` = "25-44",
                            `3` = "45-64",
                            `4` = "65+"),
    hiv12months20 = recode_factor(hiv12months20,
```

```

                `2` = "No", # put first to make "No" the reference group
                `1` = "Yes"),
condom20 = recode_factor(condom20,
                `2` = "No",
                `1` = "Yes"),
## may or may not want sexpartner as an ordered factor: if ordered,
## R will add a test for trend to regression analyses
## survey strata:
strata = as.character(strata))

chs21_recode$Race <- chs21_recode$newrace
chs21_recode$AgeGroup <- chs21_recode$agegroup
chs21_recode$HIVTesting <- chs21_recode$hiv12months20
chs21_recode$CondomUse <- chs21_recode$condom20
chs21_recode$NumberofSexPartners <- factor(chs21$sexpartner)

```

```

library(survey)
chs.dsgn <-
  svydesign(
    ids = ~ 1,
    strata = ~ strata,
    weights = ~ wt21_dual, # match to current year dataset
    data = chs21_recode, # match to current year dataset
    nest = TRUE,
    na.rm = TRUE
  )

```

```

library(svydiags)
fit <- svyglm(HIVTesting ~ CondomUse + Race + AgeGroup + NumberofSexPartners,
  design = chs.dsgn,
  family="quasibinomial")

```

```

library(tidyr)
library(kableExtra)

```

```

## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'

```

```

chs21_clean <- chs21_recode %>% dplyr::select(Race, AgeGroup, HIVTesting, CondomUse, NumberofSexPartners)
tableone::CreateTableOne(data = chs21_clean)

```

```

##
##
##      n      Overall
## Race (%)
##   White/N Afri/MidEastern, non-Hispanic 1691 (33.6)
##   Black, non-Hispanic                    1009 (20.1)
##   Hispanic                              1546 (30.8)
##   Asian/PI, non-Hispanic                  616 (12.3)
##   Other, non-Hispanic                     164 ( 3.3)
## AgeGroup (%)

```

```
##      18-24                      372 ( 7.4)
##      25-44                      2388 (47.5)
##      45-64                      1727 (34.4)
##      65+                        539 (10.7)
## HIVTesting = Yes (%)            1786 (35.5)
## CondomUse = Yes (%)            1504 (29.9)
## NumberofSexPartners (%)
##      1                          0 ( 0.0)
##      2                        4086 (81.3)
##      3                         366 ( 7.3)
##      4                         574 (11.4)
```

```
summary(fit)
```

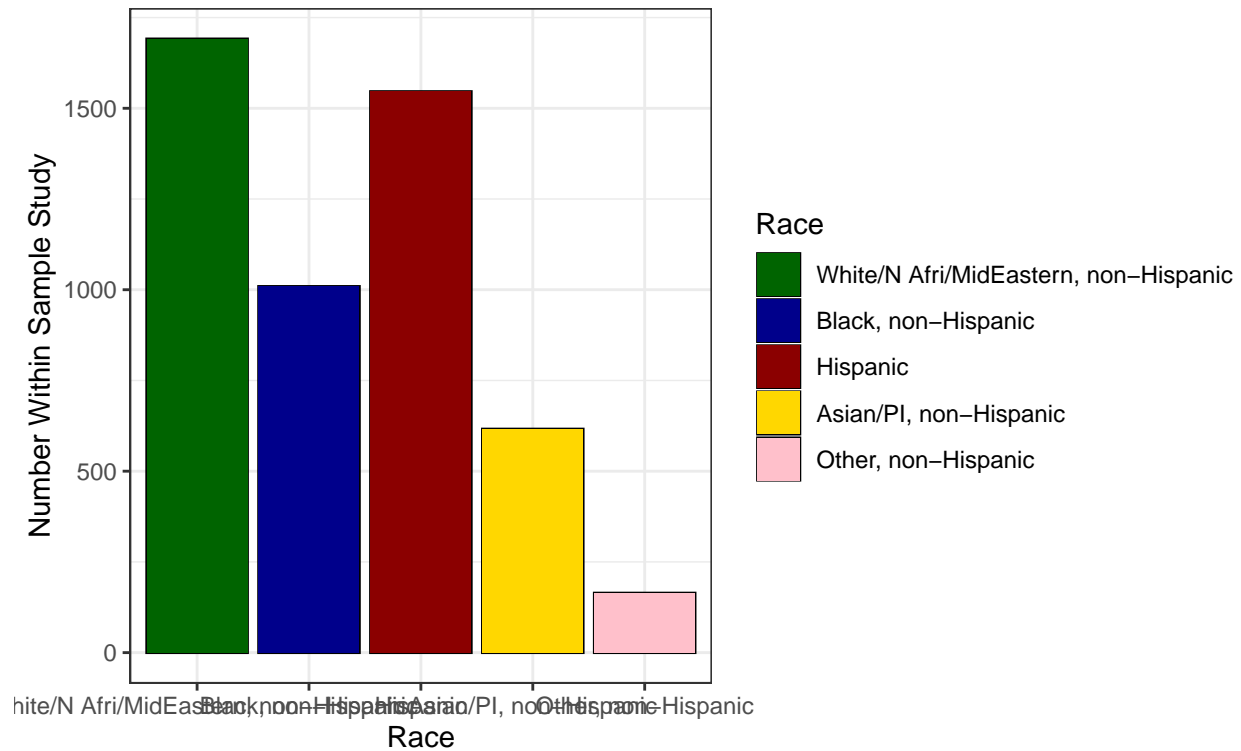
```
##
## Call:
## svyglm(formula = HIVTesting ~ CondomUse + Race + AgeGroup + NumberofSexPartners,
##      design = chs.dsgn, family = "quasibinomial")
##
## Survey design:
## svydesign(ids = ~1, strata = ~strata, weights = ~wt21_dual, data = chs21_recode,
##      nest = TRUE, na.rm = TRUE)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.55083    0.18499  -8.383  < 2e-16 ***
## CondomUseYes      0.09951    0.10500   0.948  0.343328
## RaceBlack, non-Hispanic  1.43754    0.13040  11.024  < 2e-16 ***
## RaceHispanic      1.32700    0.11856  11.193  < 2e-16 ***
## RaceAsian/PI, non-Hispanic 0.22918    0.17573   1.304  0.192250
## RaceOther, non-Hispanic  0.72968    0.28898   2.525  0.011602 *
## AgeGroup25-44      0.20023    0.16630   1.204  0.228633
## AgeGroup45-64     -0.39751    0.17796  -2.234  0.025546 *
## AgeGroup65+      -1.18743    0.25780  -4.606  4.21e-06 ***
## NumberofSexPartners3  0.62052    0.17748   3.496  0.000476 ***
## NumberofSexPartners4  1.22001    0.15883   7.681  1.89e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.051299)
##
## Number of Fisher Scoring iterations: 4
```

```
library(tidyverse)
library(broom)
coef_table <- tidy(fit, exponentiate = TRUE, conf.int = TRUE, conf.level = 0.95)
```

```
library(ggplot2)
ggplot(chs21_clean, aes(x = Race, fill = Race)) +
  geom_bar(position = "stack", stat = "count", color = "black") +
  geom_bar() + labs(title = "Race Population Within the Study", subtitle = "Data obtained from the 2020
xlab("Race") +
  ylab("Number Within Sample Study") + scale_fill_manual(values=c("dark green", "dark blue", "dark red
theme_bw()
```

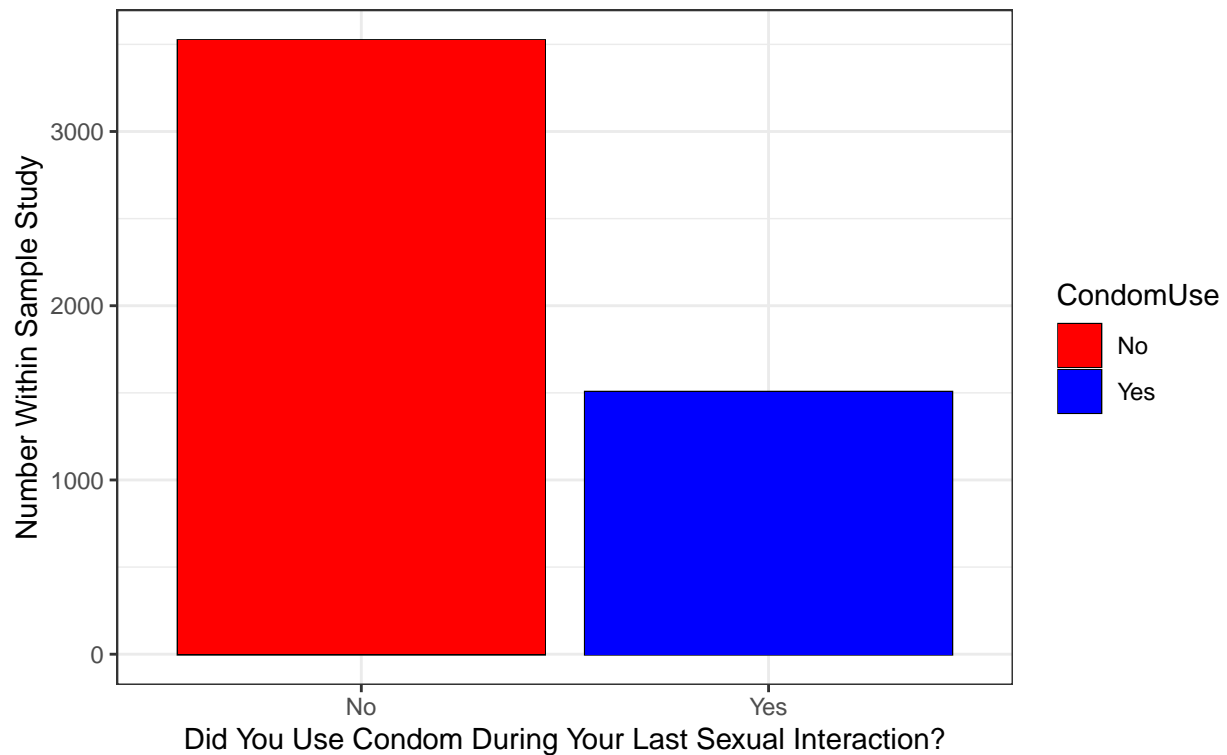
Race Population Within the Study

Data obtained from the 2020 NYC DOHMH CHS



```
ggplot(chs21_clean, aes(x = CondomUse, fill = CondomUse)) +
  geom_bar(position = "stack", stat = "count", color = "black") +
  geom_bar() + labs(title = "The Use of Condoms Among Population Within the Study", subtitle = "Data ob
  theme_bw()
```

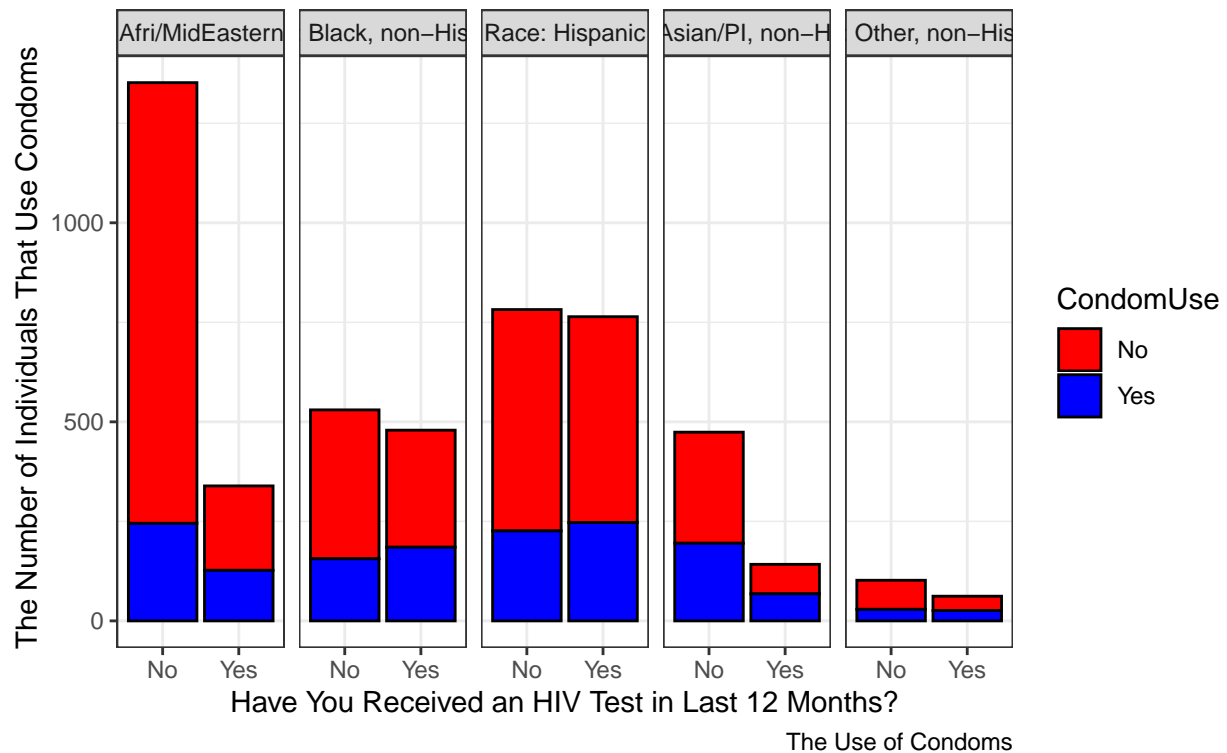
The Use of Condoms Among Population Within the Study
Data obtained from the 2020 NYC DOHMH CHS



```
ggplot(chs21_clean, aes(x = HIVTesting, fill = CondomUse)) +
  geom_bar(position = "stack", stat = "count", color = "black") +
  facet_grid(~Race, labeller = label_both) +
  labs(title = "The Use of Condoms and HIV Testing Status Distributed by Race",
        subtitle = "Data obtained from the 2020 NYC DOHMH CHS",
        caption = "The Use of Condoms") +
  xlab("Have You Received an HIV Test in Last 12 Months?") +
  ylab("The Number of Individuals That Use Condoms") +
  scale_fill_manual(values=c("red", "blue")) +
  theme_bw()
```

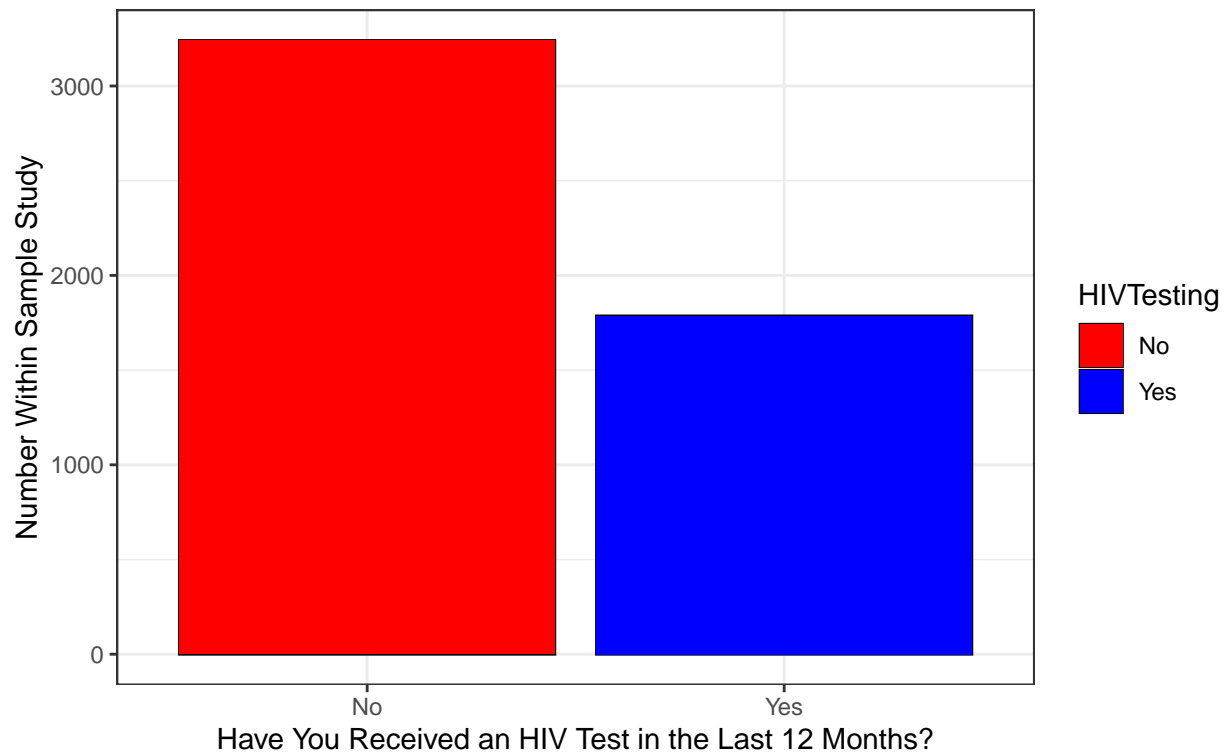
The Use of Condoms and HIV Testing Status Distributed by Race

Data obtained from the 2020 NYC DOHMH CHS



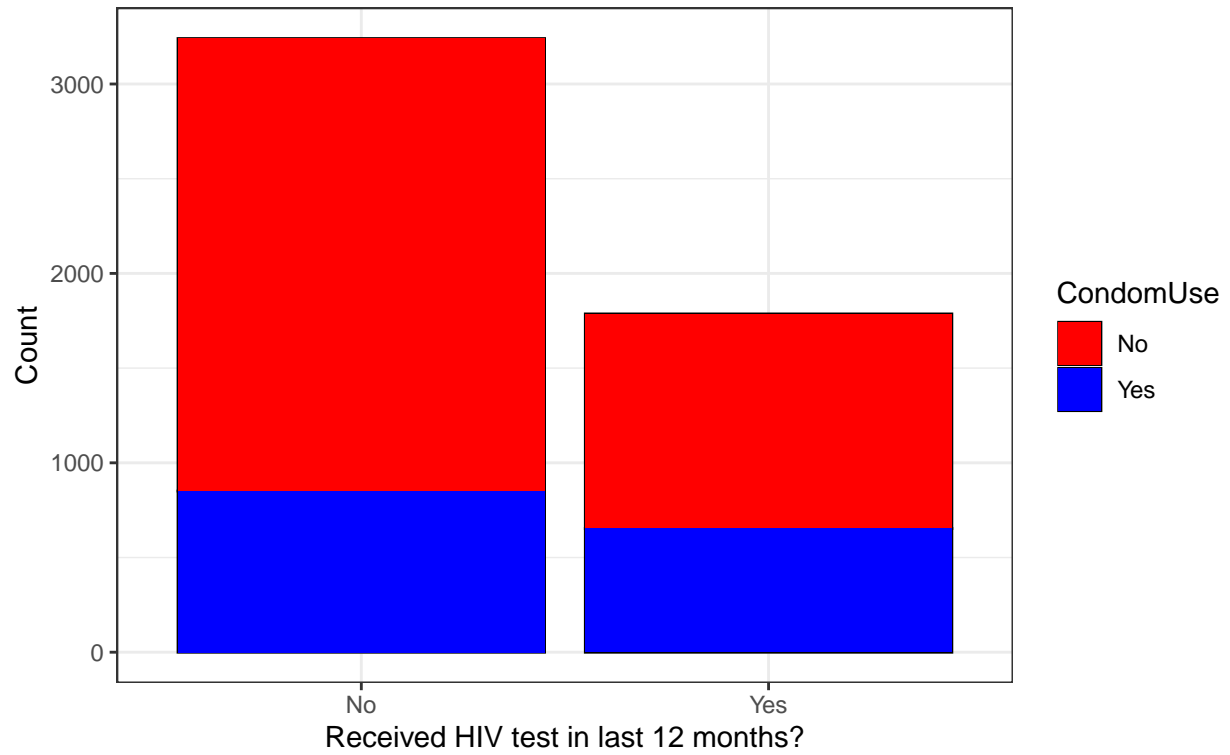
```
ggplot(chs21_clean, aes(x = HIVTesting, fill = HIVTesting)) +
  geom_bar(position = "stack", stat = "count", color = "black") +
  geom_bar() + labs(title = "HIV Testing Status in Study Population", subtitle = "Data obtained from the")
theme_bw()
```

HIV Testing Status in Study Population
Data obtained from the 2020 NYC DOHMH CHS



```
ggplot(chs21_clean, aes(x = HIVTesting, fill = CondomUse)) +  
  geom_bar(position = "stack", stat = "count", color = "black") +  
  geom_bar() + labs(title = "HIV Testing Status and Condom Use in Study Population", subtitle = "Data obtained from the 2020 NYC DOHMH CHS") +  
  scale_fill_manual(values=c("red", "blue")) +  
  theme_bw()
```

HIV Testing Status and Condom Use in Study Population
Data obtained from the 2020 NYC DOHMH CHS



Attribution: Adil Shakil