GVPT728 Project Final Code

Yael Beshaw

2025-01-23

Download Potentially Necessary Packages

```
library(tidyverse)
library(tidycensus)
library(dplyr)
library(readr)
library(readxl)
library(haven)
library(ggfortify)
library(car)
library(huxtable)
library(lmtest)
library(ggdist)
library(fixest)
library(sandwich)
library(lmtest)
library(lme4)
library(modelsummary)
library(tableone)
library(knitr)
```

Import Original Dataset

```
hints6 <- read_sav("/Users/yaelbeshaw/R Scripts and Projects/NYU-APSTA-GE-2011/HINTS6_SPSS/hints6_publi
```

Variables of Interest for this Study

Dependent Variable

Health Information-Seeking Behaviors:

Electronic2_HealthInfo: "In the past 12 months, have you used the Internet to look for health or medical information?" (Binary)

Independent Variable

Frequency of Internet Usage:

SocMed_Visited: Frequency of visiting social media sites in the past 12 months (ordinal or categorical).

Controls

Internet Access

Type of internet access

1. Internet_DialUp
2. Internet_HighSpeed
3. Internet_Cell

Device ownership or access

HAVEDEVICE_CAT

Access Satisfaction

InternetConnection: Satisfaction with internet connection for health-related needs (ordinal scale).

Health Literacy:

ConfidentInternetHealth: "How confident are you that you can find helpful health resources on the Internet?" (Ordinal scale)

Trust in Information:

MisleadingHealthInfo: "How much of the health information that you see on social media do you think is false or misleading?" (Ordinal scale)

Health Status:

EverHadCancer

MedConditions_Diabetes, MedConditions_HighBP, MedConditions_HeartCondition, MedConditions_LungDisease, MedConditions_Depression

Demographics:

Age; Age and AgeGrpB

Gender; BirthGender

Race; RaceEthn5

Educational attainment; EducA

Geographic: CENSDIV

Urban vs Rural; PR_RUCA_2010: USDA 2010 Primary Rural-Urban Community Area Code

Economic Factors:

Income; HHInc

Employment: WorkFullTime(R4)

How many children: ChildrenInHH

Stratum Classification

STRATUM

HR High minority rural area HU High minority urban area LR Low minority rural area LU Low minority urban area

Household ID

HHID

Select Variables

```
select_data <- hints6 |>
  select(Electronic2_HealthInfo, #binary 1 or 2 *done
         UseInternet, #binary 1 or 2 *done
         SocMed_Visited, # 1 to 5 (every day to never) --flip *done
         Internet_DialUp, # binary 1 or 2 *done
         Internet_HighSpeed, # binary 1 or 2 *done
         Internet_Cell, # binary 1 or 2 *done
         HAVEDEVICE_CAT, #recode 4 as 0 *done
         InternetConnection, # 1 to 5 (extremely satisfied to not at all) -flip *done
         ConfidentInternetHealth, #1 to 5 (completely confident to not at all) -flip *done
         MisleadingHealthInfo, #1 to 5 (a lot to none (4), I dont use social (5)) -flip with 1 =0 *done
         EverHadCancer, #dummy- healthstatus *done
         MedConditions_Diabetes, #dummy - healthstatus *done
         MedConditions_HighBP, #dummy- healthstatus *done
         MedConditions_HeartCondition, #dummy - healthstatus *done
         MedConditions_LungDisease, #dummy- healthstatus *done
         MedConditions_Depression, #dummy- healthstatus *done
         Age, #continuous (18 to 99) use AgeGrpB for visualization *done
         BirthGender, # 0 == Male, 1 == Female *done
         RaceEthn5, # NHWhite, NHBlack, Hispanic, NHAsian, NHOther *done
         EducA, # Less than HS, HS, Some College, Collge and Beyond *done
         CENSDIV, # New England, Middle Atlantic, E/W North Central, South Atlantic, E/W South Central,
         PR_RUCA_2010, # Metropolitan, Micropolitian, Small Town, Rural (code 0 or 1) *done
         HHInc, # 1 to 5, increasing *done
         WorkFullTime, # 1 or 2 0,1 *done
         ChildrenInHH, # 0 to 9 *done
         AgeGrpB,
         HHID)
```

Data Preprocessing

Remove any missing data

```
clean_data <- select_data %>%
  mutate(across(everything(), ~ na_if(., -9))) %>%
  mutate(across(everything(), ~ na_if(., -7))) %>%
  mutate(across(everything(), ~ na_if(., -6))) %>%
  mutate(across(everything(), ~ na_if(., -5))) %>%
  mutate(across(everything(), ~ na_if(., -4))) %>%
  mutate(across(everything(), ~ na_if(., -2)))


data <- na.omit(clean_data)
```

Join the new dataset with the stratum assignment using HHID, use later

```
stratum <- hints6 |>
  select(HHID,
         STRATUM)
```

```r
data <- inner_join(data, stratum, by = "HHID")
```

Recodes

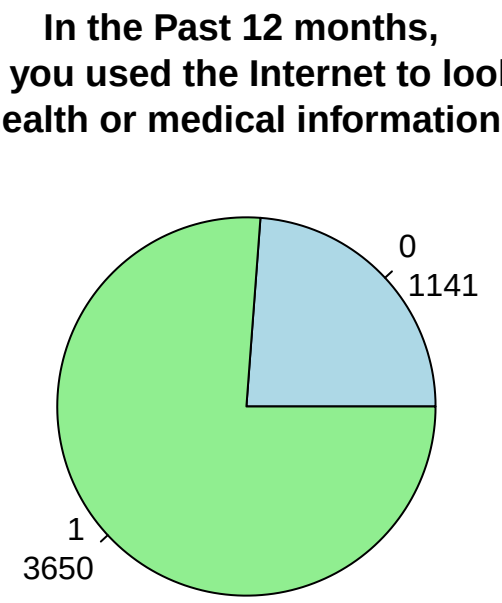
```r
#For Binary Variables- Code 1's and 2's to 0 and 1

#DV
data$Electronic2_HealthInfo <- ifelse(data$Electronic2_HealthInfo == 1,
                                      1, 0)
# Type of Internet
data$Internet_DialUp <- ifelse(data$Internet_DialUp == 1,
                                      1, 0)
data$Internet_HighSpeed <- ifelse(data$Internet_HighSpeed == 1,
                                      1, 0)
data$Internet_Cell <- ifelse(data$Internet_Cell == 1,
                                      1, 0)

#Gender
data$BirthGender <- ifelse(data$BirthGender == 1,
                                      1, 0)

#Employment
data$WorkFullTime<- ifelse(data$WorkFullTime == 1,
                                      1, 0)


# Flip the ordinal variables from least to greatest

#Fr(SM)
data$SocMed_Visited <- factor(data$SocMed_Visited,
                                    levels = c(5, 4, 3, 2, 1))
levels(data$SocMed_Visited) <- c("1", "2", "3", "4", "5")


# Internet Connection Satisfaction
data$InternetConnection <- factor(data$InternetConnection,
                                    levels = c(5, 4, 3, 2, 1))
levels(data$InternetConnection) <- c("1", "2", "3", "4", "5")


# Confidence in Getting Health Info (Literacy)
data$ConfidentInternetHealth <- factor(data$ConfidentInternetHealth,
                                    levels = c(5, 4, 3, 2, 1))
levels(data$ConfidentInternetHealth) <- c("1", "2", "3", "4", "5")


# Trust in Health Info on SM
data$MisleadingHealthInfo <- factor(data$MisleadingHealthInfo,
                                    levels = c(5, 4, 3, 2, 1))
levels(data$MisleadingHealthInfo) <- c("1", "2", "3", "4", "5")
```

```r
#Recode Variables

# Rural vs Not
data$PR_RUCA_2010 <- ifelse(data$PR_RUCA_2010== 1 |
                              data$PR_RUCA_2010== 4|
                              data$PR_RUCA_2010== 7,
                            0,
                            1)

# Any chronic conditions + cancer as a 1 or 0
data$HealthStatus <- ifelse(data$EverHadCancer == 1 |
                                data$MedConditions_Diabetes == 1 |
                                data$MedConditions_HighBP == 1 |
                                data$MedConditions_HeartCondition == 1 |
                                data$MedConditions_LungDisease == 1 |
                                data$MedConditions_Depression == 1,
                                1,
                                0)

# Devices, none is 0 and multiple is 4 instead
data$HAVEDEVICE_CAT <- ifelse(data$HAVEDEVICE_CAT == 4, 0,
                                ifelse(data$HAVEDEVICE_CAT == 5, 4,
                                       data$HAVEDEVICE_CAT))


# Turn internet type into one variable
data$InternetAccessType <- NA

# Assign any combinations as 4 = Multiple
data$InternetAccessType[data$Internet_DialUp == 1 &
                          data$Internet_HighSpeed == 1] <- 4

data$InternetAccessType[data$Internet_DialUp == 1 &
                          data$Internet_Cell == 1] <- 4

data$InternetAccessType[data$Internet_HighSpeed == 1 &
                          data$Internet_Cell == 1] <- 4

# Assign individual categories as their own
data$InternetAccessType[data$Internet_DialUp == 1 &
                          is.na(data$InternetAccessType)] <- 1

data$InternetAccessType[data$Internet_HighSpeed == 1 &
                          is.na(data$InternetAccessType)] <- 2

data$InternetAccessType[data$Internet_Cell == 1 &
                          is.na(data$InternetAccessType)] <- 3


# Recode the number of children, 0 = None up to 3+

data$ChildrenInHH_recode <- ifelse(data$ChildrenInHH >= 3, 3, as.character(data$ChildrenInHH))
```

Descriptive Statisitics Visualizations

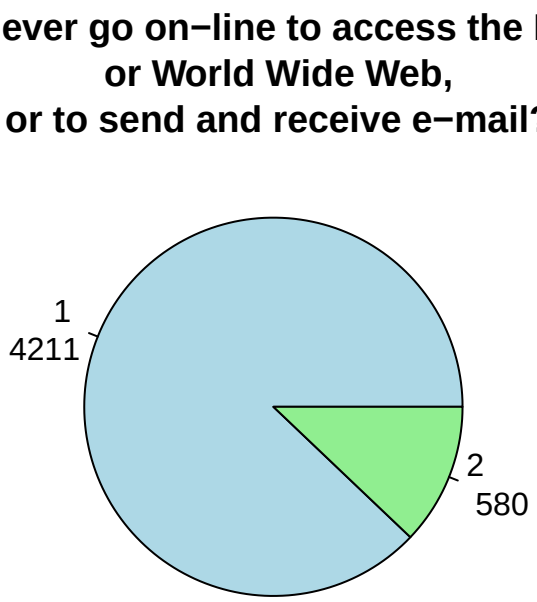
Dependent VAR

```r
#DEPENDENT VARIABLE

freq_table <- table(data$Electronic2_HealthInfo)

pie(freq_table,
    main = "In the Past 12 months, \n have you used the Internet to look for \n health or medical infor
    col = c("lightblue", "lightgreen"),
    labels = paste(names(freq_table), "\n", freq_table))
```



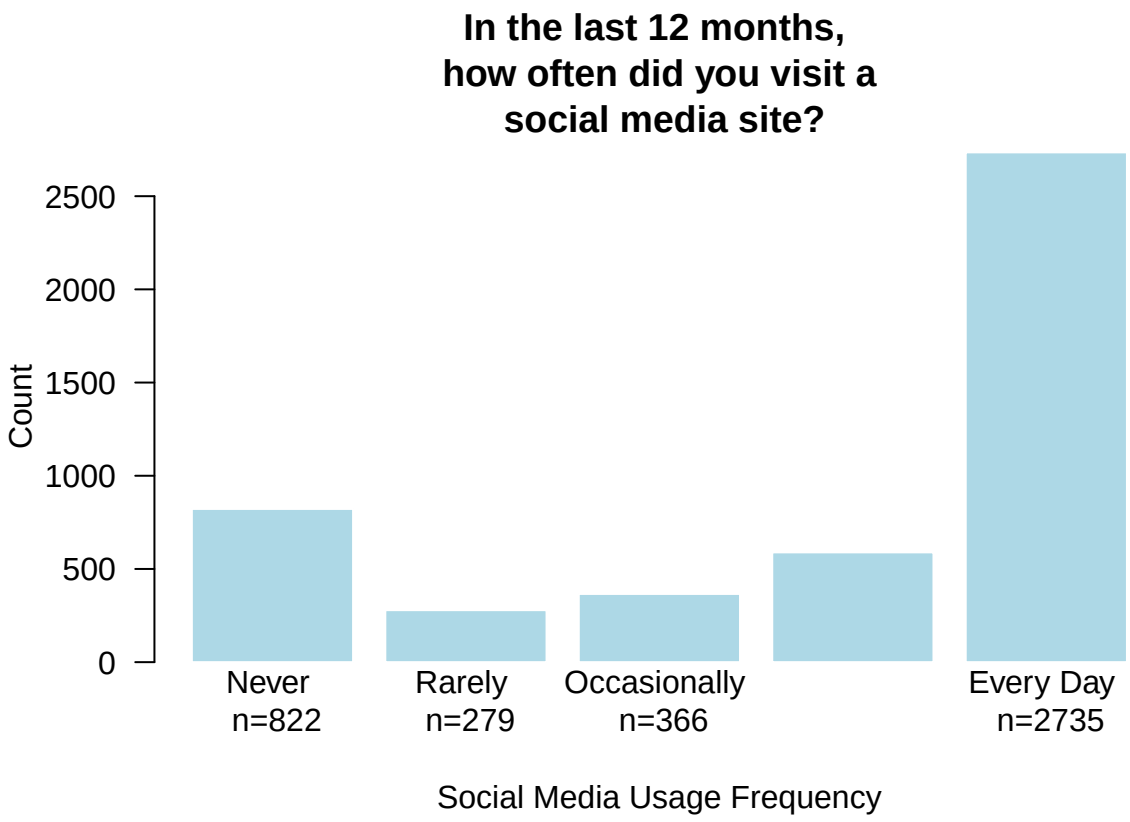
Compare the results against how many people use the internet in general

```r
freq_table <- table(data$ UseInternet)

pie(freq_table,
    main = "Do you ever go on-line to access the Internet \n or World Wide Web,\n or to send and receiv
    col = c("lightblue", "lightgreen"),
    labels = paste(names(freq_table), "\n", freq_table))
```



Independent VAR

```
freq_table <- table(data$SocMed_Visited)

x_labels <- c("Never \n n=822", "Rarely \n n=279", "Occasionally \n n=366", "Frequently \n n= 589", "Ev

# Create the barplot
barplot(freq_table,
        main = "In the last 12 months, \n how often did you visit a \n social media site?",
        xlab = "Social Media Usage Frequency",
        ylab = "Count",
        col = "lightblue",
        border = "white",
        las = 1,
        ylim = c(0, max(freq_table) + 5),
        names.arg = x_labels)
```



Health Information Related Controls

```
#internet type
freq_table <- table(data$InternetAccessType)

names(freq_table) <- c("Dial-Up", "High-Speed", "Cellular", "Multiple")

category_count_table <- data.frame(
  Category = names(freq_table),
  Count = as.vector(freq_table)
)

kable(category_count_table,
      caption = "Access Type")
```

Table 1: Access Type

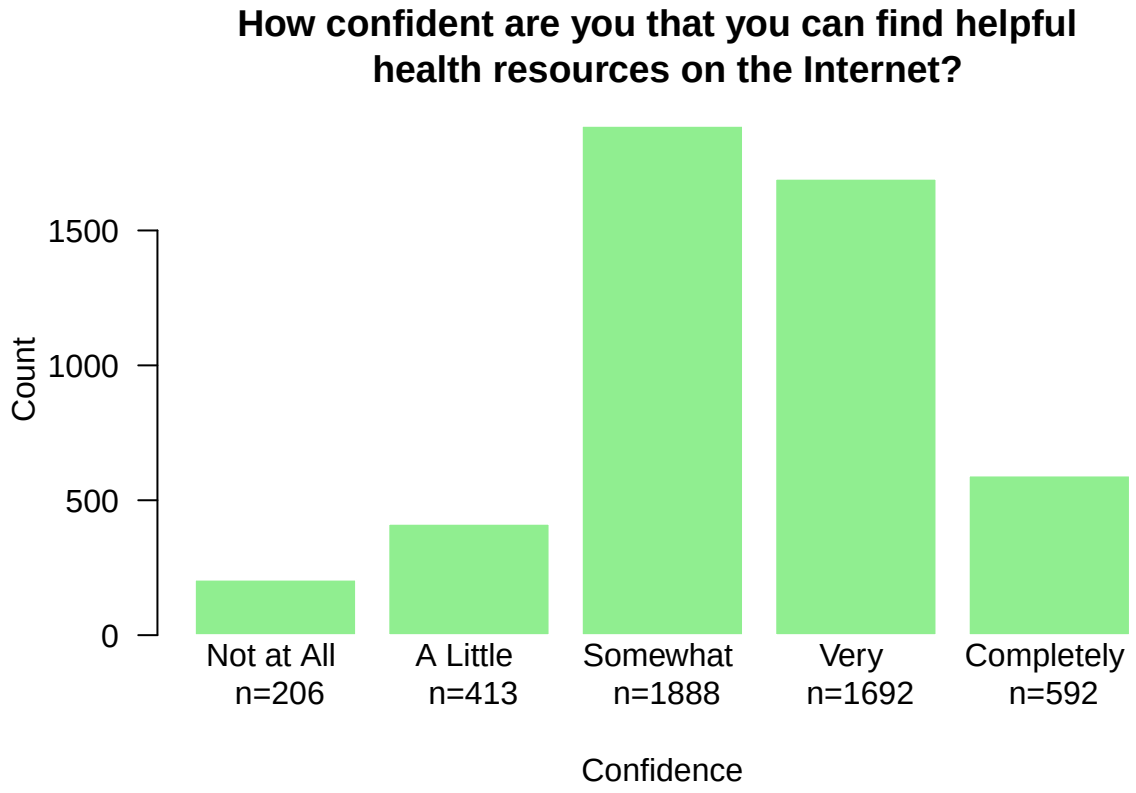
Category	Count
Dial-Up	24
High-Speed	706
Cellular	260
Multiple	3215

```r
#device type
freq_table <- table(data$HAVEDEVICE_CAT)

names(freq_table) <- c("Tablet Computer Only", "Smartphone Only", "Basic Cell Only",
                       "None", "Multiple Devices")

category_count_table <- data.frame(
  Category = names(freq_table),
  Count = as.vector(freq_table)
)

kable(category_count_table,
      caption = "Device Type")
```

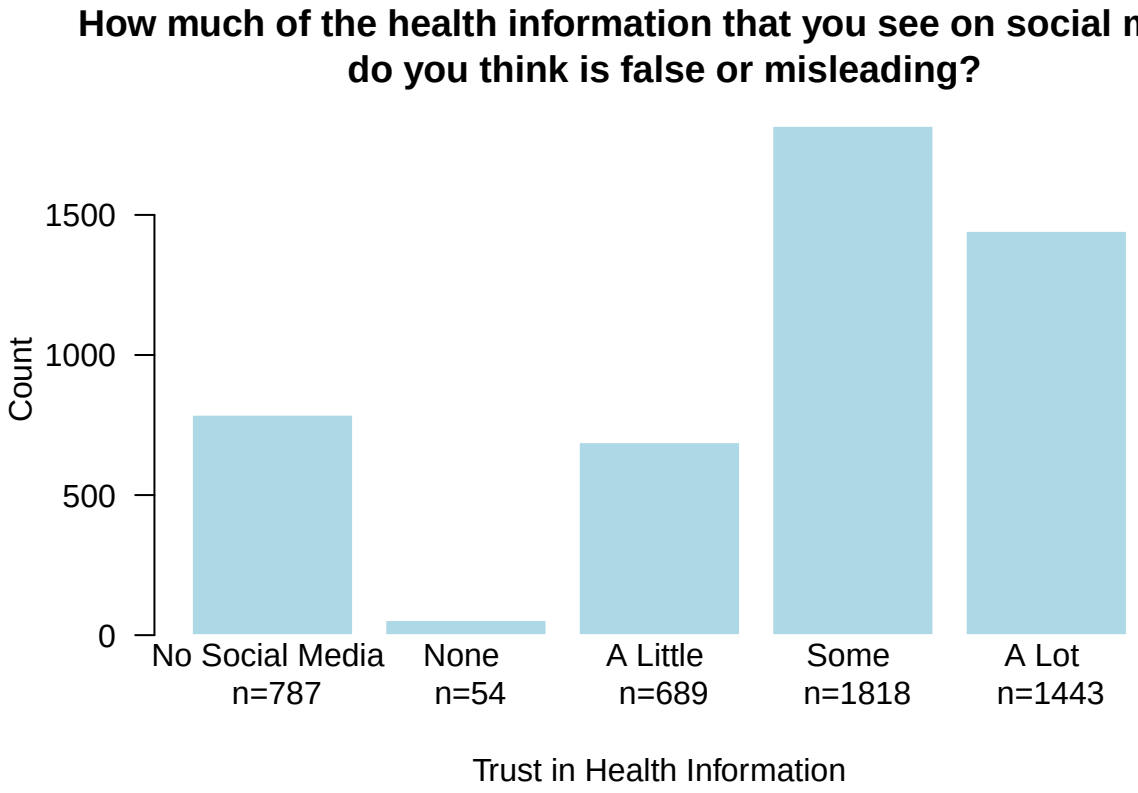
Table 2: Device Type

Category	Count
Tablet Computer Only	85
Smartphone Only	157
Basic Cell Only	1452
None	256
Multiple Devices	2841

```r
#health literacy

freq_table <- table(data$ConfidentInternetHealth)

x_labels <- c("Not at All \n n=206", "A Little \n n=413", "Somewhat \n n=1888", "Very \n n=1692", "Compl

# Create the barplot
barplot(freq_table,
        main = "How confident are you that you can find helpful \n health resources on the Internet?",
        xlab = "Confidence",
        ylab = "Count",
        col = "lightgreen",
        border = "white",
        las = 1,
        ylim = c(0, max(freq_table) + 5),
        names.arg = x_labels)
```



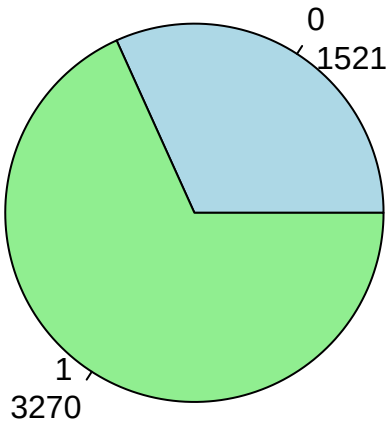

```r
#trust in health information on social media

freq_table <- table(data$MisleadingHealthInfo)

x_labels <- c("No Social Media \n n=787", "None \n n=54", "A Little \n n=689", "Some \n n=1818", "A Lot

barplot(freq_table,
        main = "How much of the health information that you see on social media \n do you think is fals
        xlab = "Trust in Health Information",
        ylab = "Count",
        col = "lightblue",
        border = "white",
        las = 1,
        ylim = c(0, max(freq_table) + 5),
        names.arg = x_labels)
```



Demographic Controls

```
#DEMOGRAPHIC CONTROLS

# Health Status
freq_table <- table(data$HealthStatus)

pie(freq_table,
    main = "Have you ever had Cancer or another Chronic Condition? \n (i.e., Diabetes, High BP, Heart C
    col = c("lightblue", "lightgreen"),
    labels = paste(names(freq_table), "\n", freq_table))
```

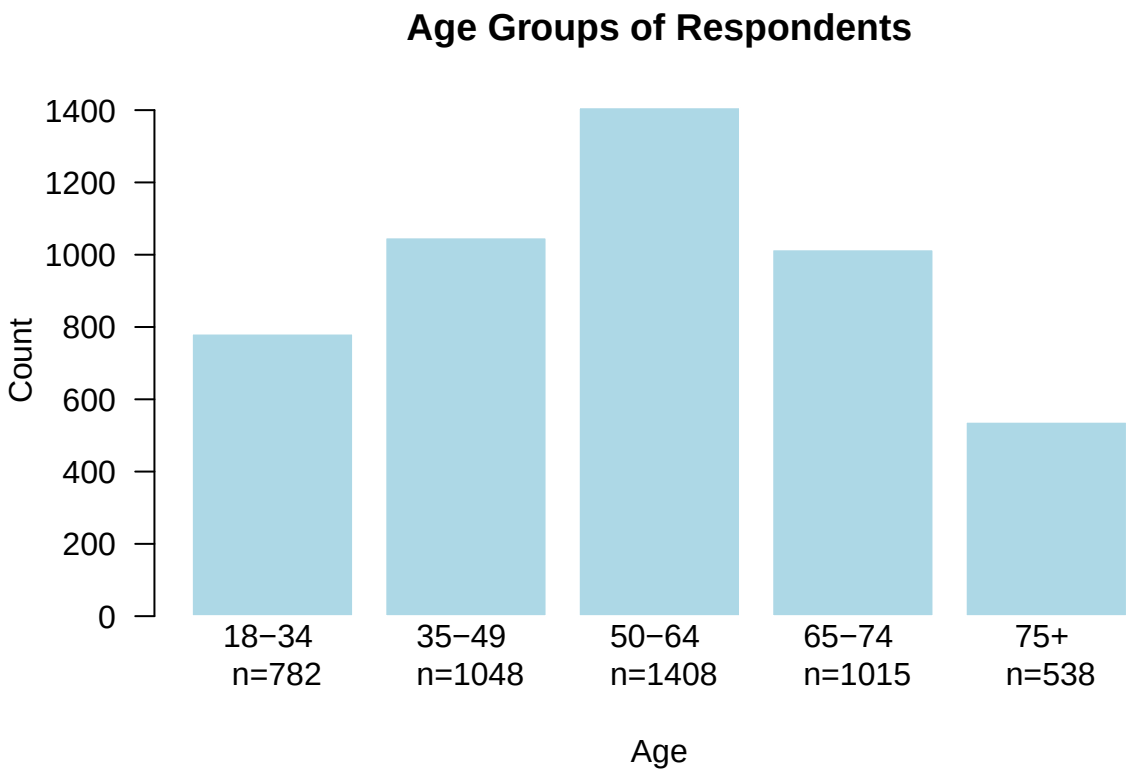
Have you ever had Cancer or another Chronic Condition?
(i.e., Diabetes, High BP, Heart Condition, Lung Disease, Depressior



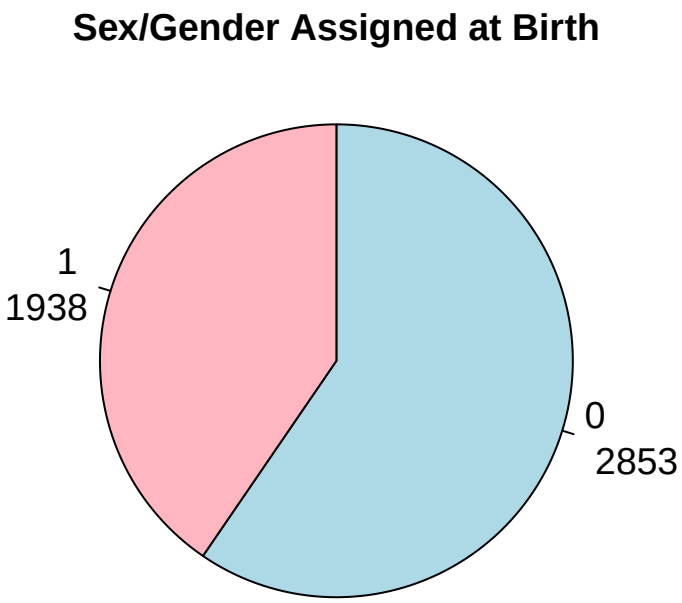

```r
# Age
median(data$Age)
```

```
## [1] 56
```

```r
freq_table <- table(data$AgeGrpB)

x_labels <- c("18-34 \n n=782", "35-49 \n n=1048", "50-64 \n n=1408", "65-74 \n n=1015", "75+ \n n=538")

barplot(freq_table,
        main = "Age Groups of Respondents",
        xlab = "Age",
        ylab = "Count",
        col = "lightblue",
        border = "white",
        las = 1,
        ylim = c(0, max(freq_table) + 5),
        names.arg = x_labels)
```




```r
# SEX at BIRTH
freq_table <- table(data$BirthGender)

pie(freq_table,
    main = "Sex/Gender Assigned at Birth",
    col = c("lightblue", "lightpink"),
    labels = paste(names(freq_table), "\n", freq_table),
    radius = 1,
    cex = 1.2,
    clockwise = TRUE,
    border = "black",
    init.angle = 90
)
```




```
# Race
freq_table <- table(data$RaceEthn5)

names(freq_table) <- c("NHWHite", "NHBlack", "Hispanic",
                       "NHAsian", "NHOthers")

category_count_table <- data.frame(
  Category = names(freq_table),
  Count = as.vector(freq_table)
)

kable(category_count_table,
      caption = "Race")
```

Table 3: Race

Category	Count
NHWHite	2800
NHBlack	755
Hispanic	814
NHAsian	255
NHOthers	167

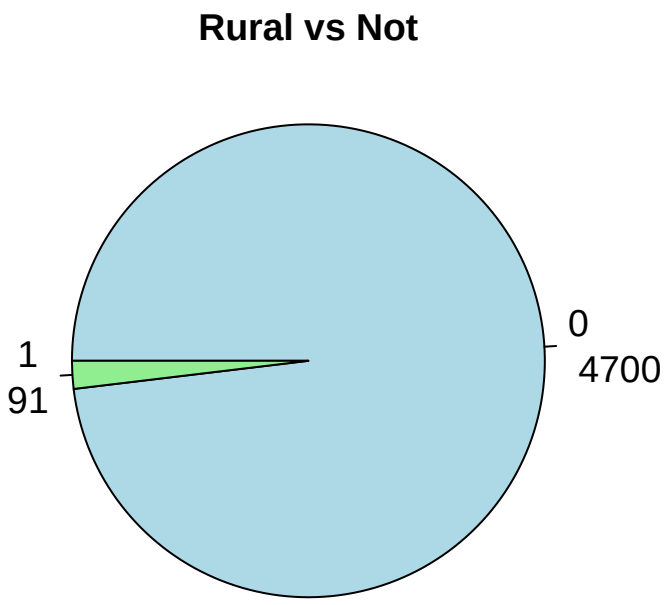
```
# Education
freq_table <- table(data$EducA)
```

```r
names(freq_table) <- c("< High School", "High School Grad", "Some College",
                       "College Grad +")

category_count_table <- data.frame(
  Category = names(freq_table),
  Count = as.vector(freq_table)
)

kable(category_count_table,
      caption = "Education")
```

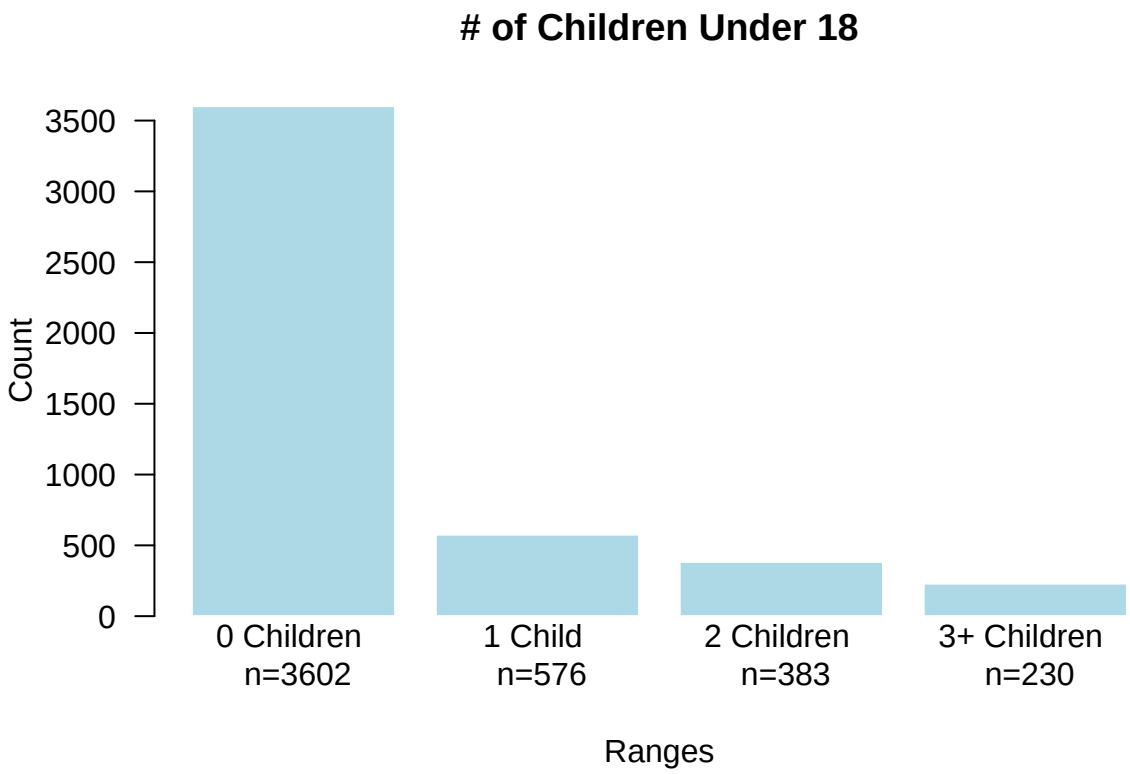
Table 4: Education

Category	Count
< High School	240
High School Grad	798
Some College	1362
College Grad +	2391

```r
# Rural vs Not
freq_table <- table(data$PR_RUCA_2010)

pie(freq_table,
    main = "Rural vs Not",
    col = c("lightblue", "lightgreen"),
    labels = paste(names(freq_table), "\n", freq_table),
    radius = 1,
    cex = 1.2,
    clockwise = TRUE,
    border = "black",
    init.angle = 180
)
```

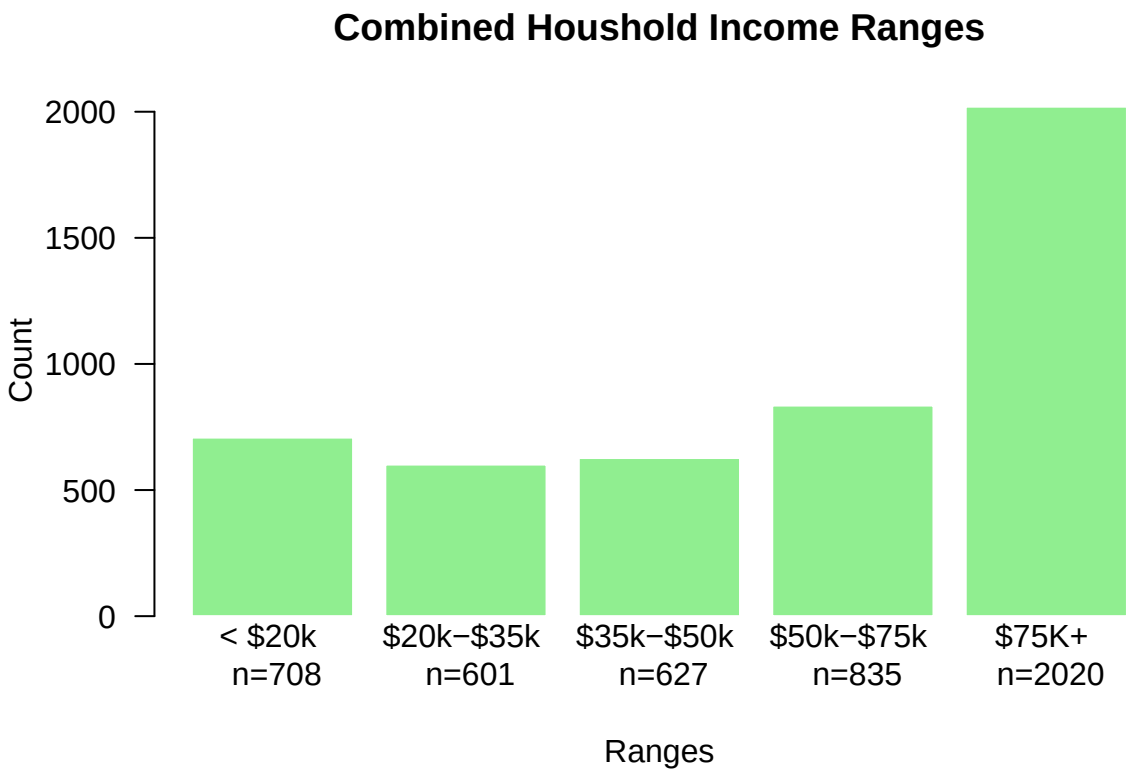



```r
# How many Children Under 18

freq_table <- table(data$ChildrenInHH_recode)

x_labels <- c("0 Children \n n=3602", "1 Child \n n=576", "2 Children \n n=383", "3+ Children \n n=230")

# Create the barplot
barplot(freq_table,
        main = "# of Children Under 18",
        xlab = "Ranges",
        ylab = "Count",
        col = "lightblue",
        border = "white",
        las = 1,
        ylim = c(0, max(freq_table) + 5),
        names.arg = x_labels)
```

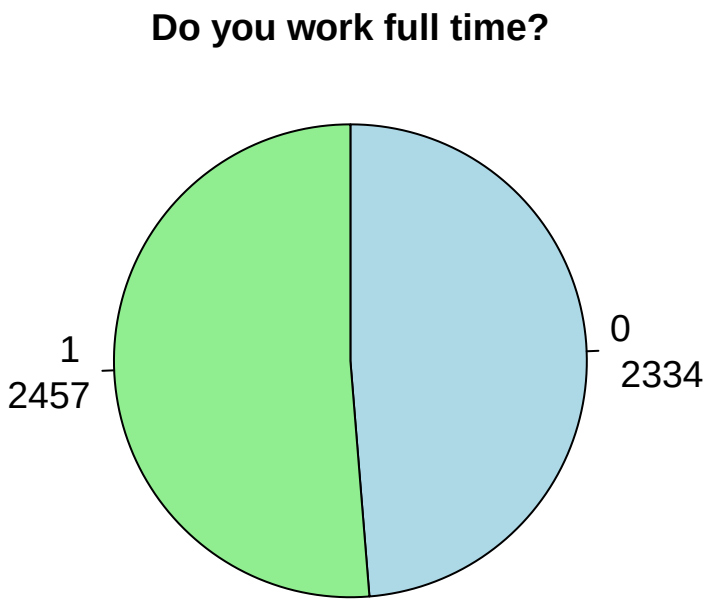


Economic Controls

```
#HHInc

freq_table <- table(data$HHInc)

x_labels <- c("< $20k \n n=708", "$20k-$35k \n n=601", "$35k-$50k \n n=627", "$50k-$75k \n n=835", "$75

# Create the barplot
barplot(freq_table,
        main = "Combined Houshold Income Ranges",
        xlab = "Ranges",
        ylab = "Count",
        col = "lightgreen",
        border = "white",
        las = 1,
        ylim = c(0, max(freq_table) + 5),
        names.arg = x_labels)
```




```r
#Full Time Work
freq_table <- table(data$WorkFullTime)

pie(freq_table,
    main = "Do you work full time?",
    col = c("lightblue", "lightgreen"),
    labels = paste(names(freq_table), "\n", freq_table),
    radius = 1,
    cex = 1.2,
    clockwise = TRUE,
    border = "black",
    init.angle = 90
)
```



Stratification Assignment

```
#Stratum Classification

freq_table <- table(data$STRATUM)

pie(freq_table,
    main = "STRATUM ASSIGNMENT",
    col = c("lightblue", "lightgreen", "lightpink", "lightyellow"),
    labels = paste(names(freq_table), "\n", freq_table),
    radius = 1,
    cex = 0.75,
    clockwise = TRUE,
    border = "black",
    init.angle = 270
)
```

STRATUM ASSIGNMENT