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)
library(gt)
library(marginaleffects)
library(ggeffects)
library(labelled)
library(kableExtra)
```

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)

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

Recodes

```
#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")
```

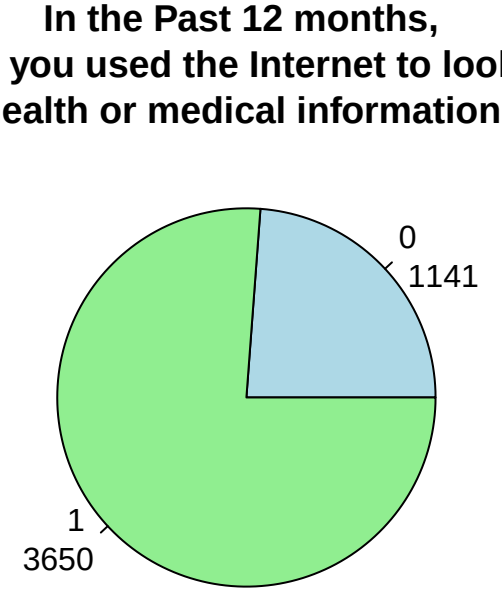
```r
# Trust in Health Info on SM
data$MisleadingHealthInfo <- factor(data$MisleadingHealthInfo,
                                    levels = c(5, 4, 3, 2, 1))
levels(data$MisleadingHealthInfo) <- c("1", "2", "3", "4", "5")


#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
```

```r
# 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

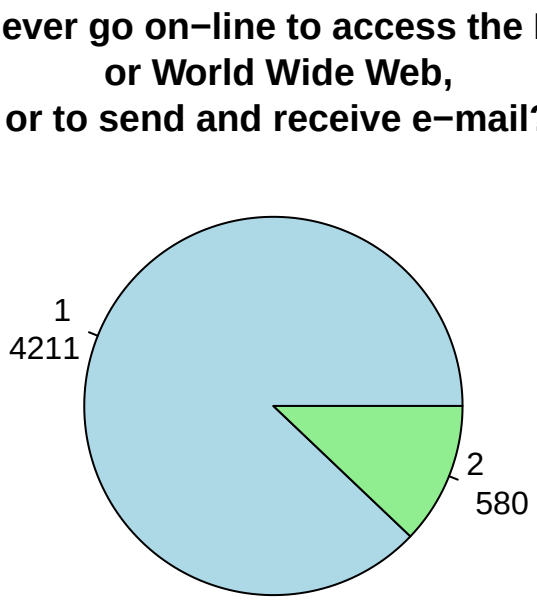
Dependent VAR

```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))
```



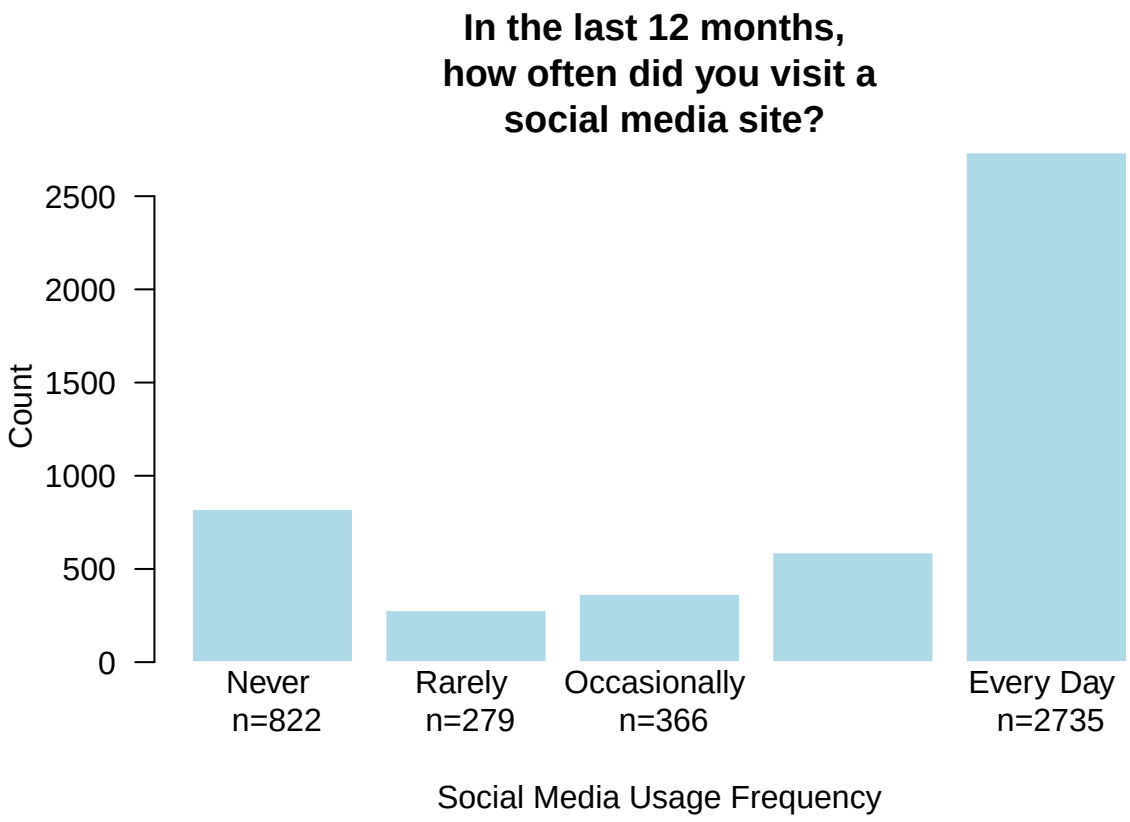
Compare the results against how many people use the internet in general

```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

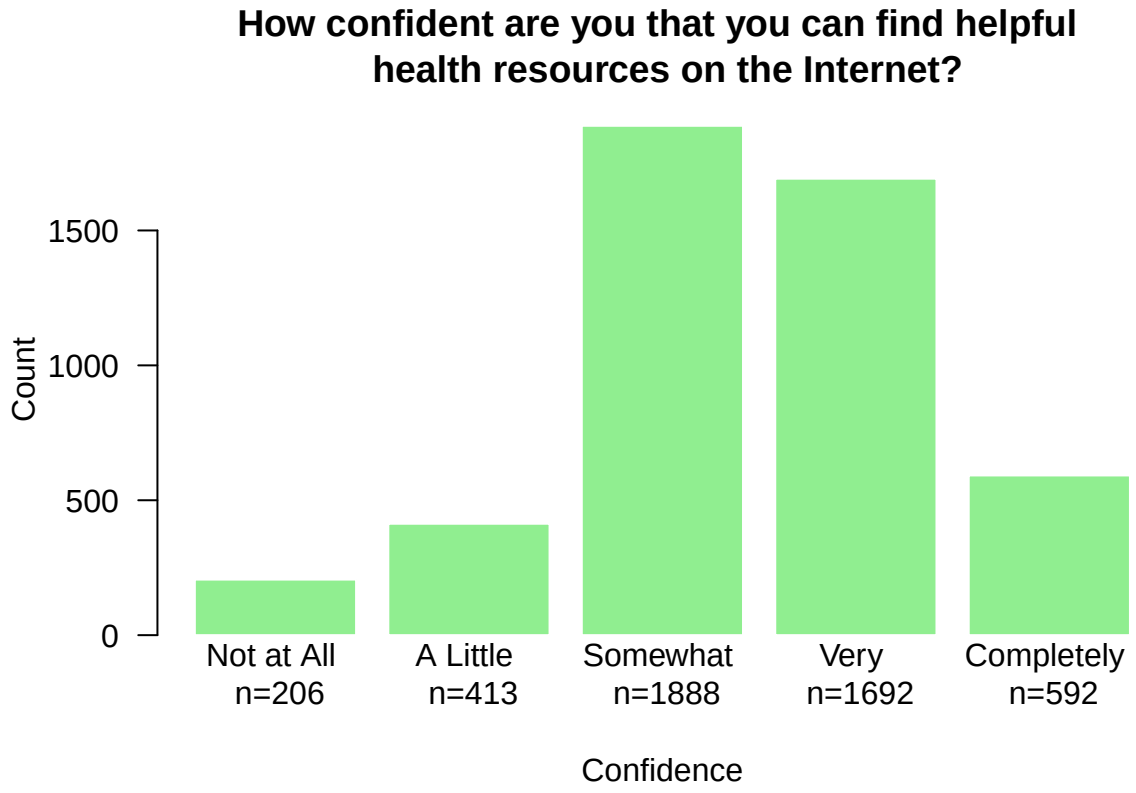
Category	Count
Dial-Up	24
High-Speed	706
Cellular	260
Multiple	3215

```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")
```

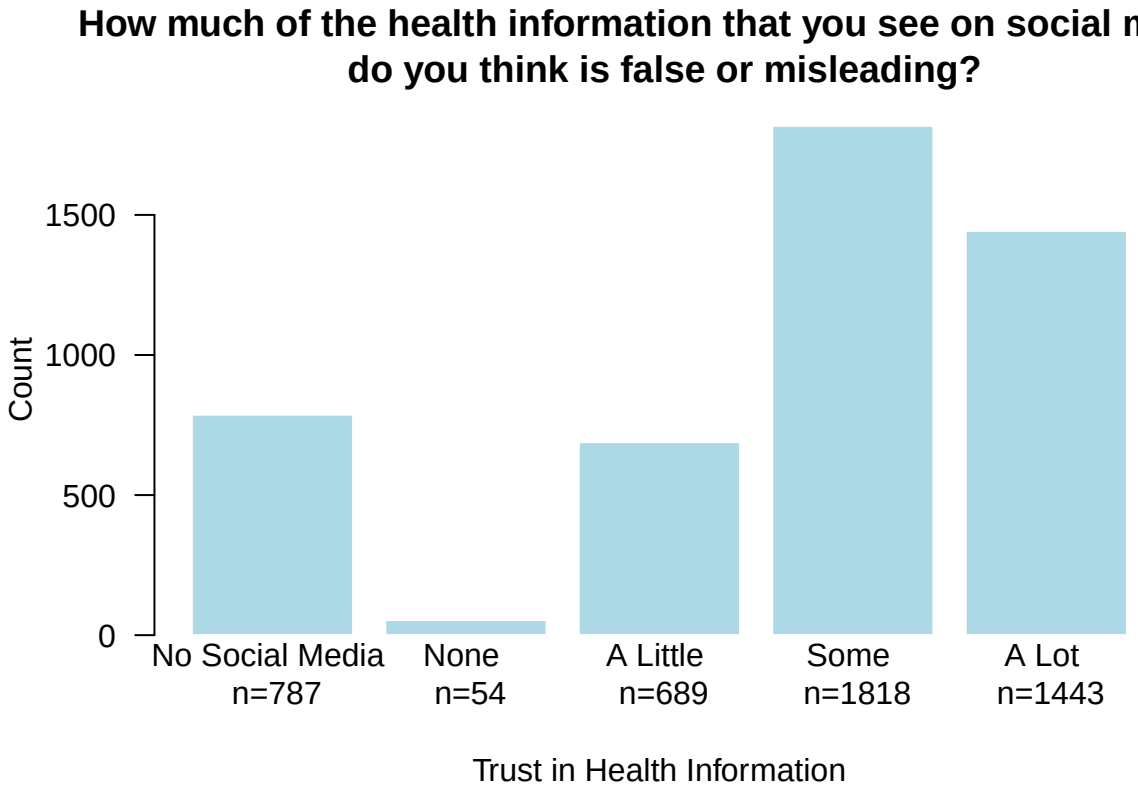
Table 2: Device Type

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)
```



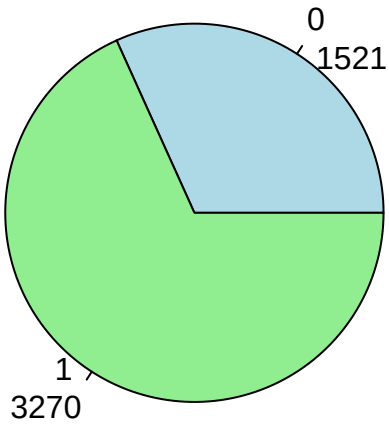

```r
#trust in health information on social media

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))
```

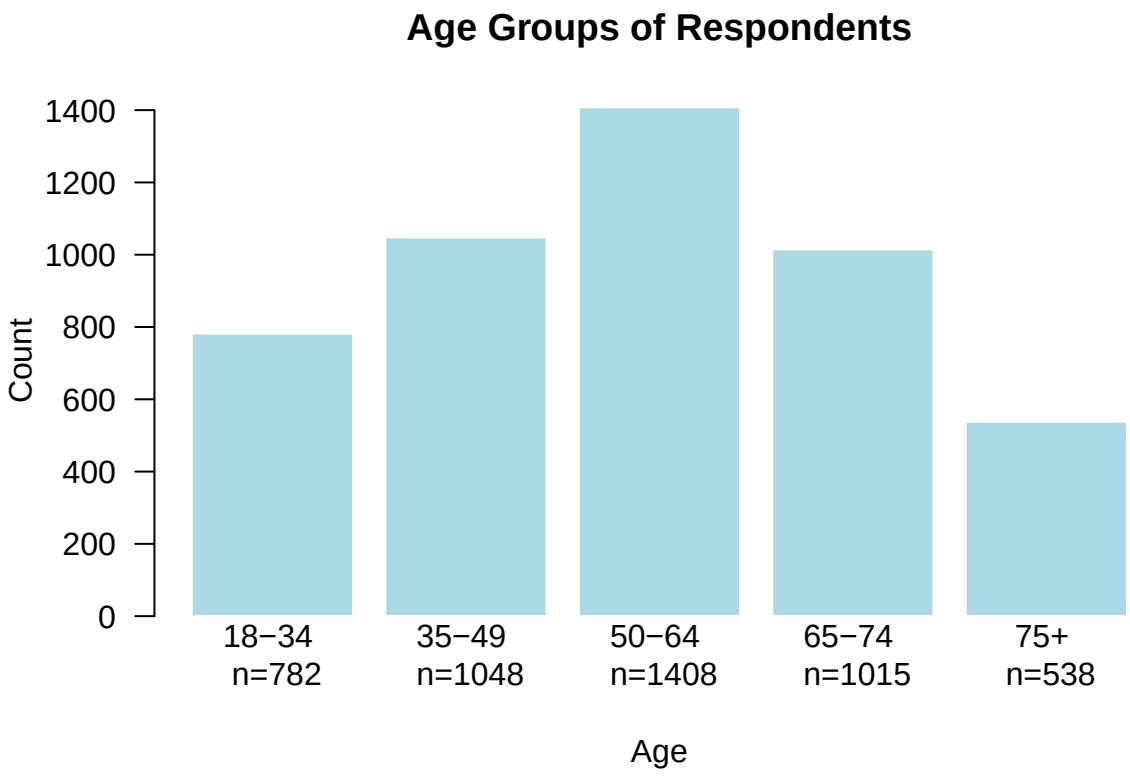
Have you ever had Cancer or another Chronic Condition?
(i.e., Diabetes, High BP, Heart Condition, Lung Disease, Depressior



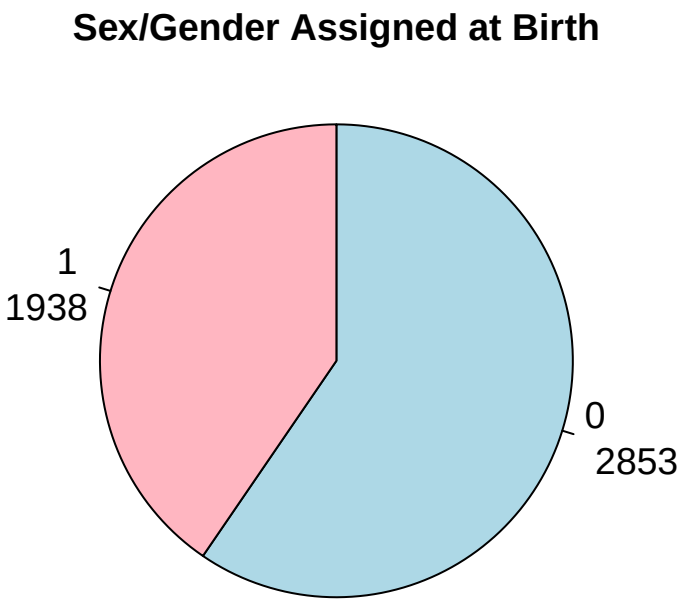

```r
# Age
median(data$Age)
```

```
## [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

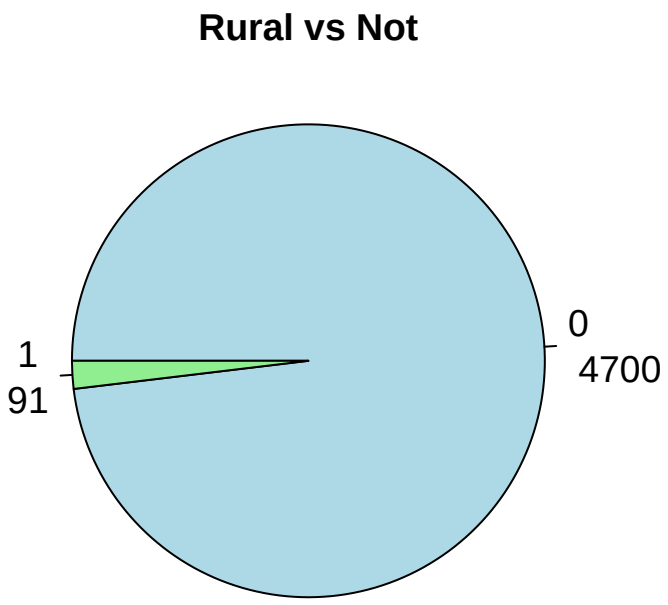
```
# Education
freq_table <- table(data$EducA)
```

```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")
```

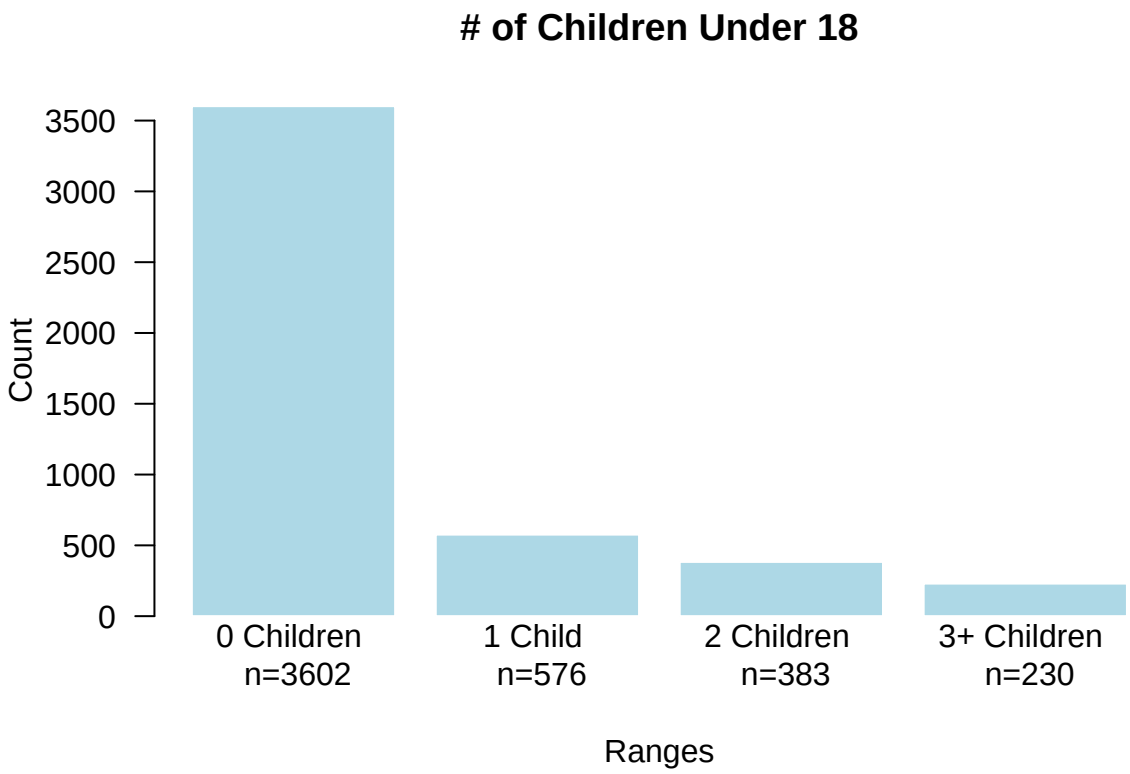
Table 4: Education

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
)
```

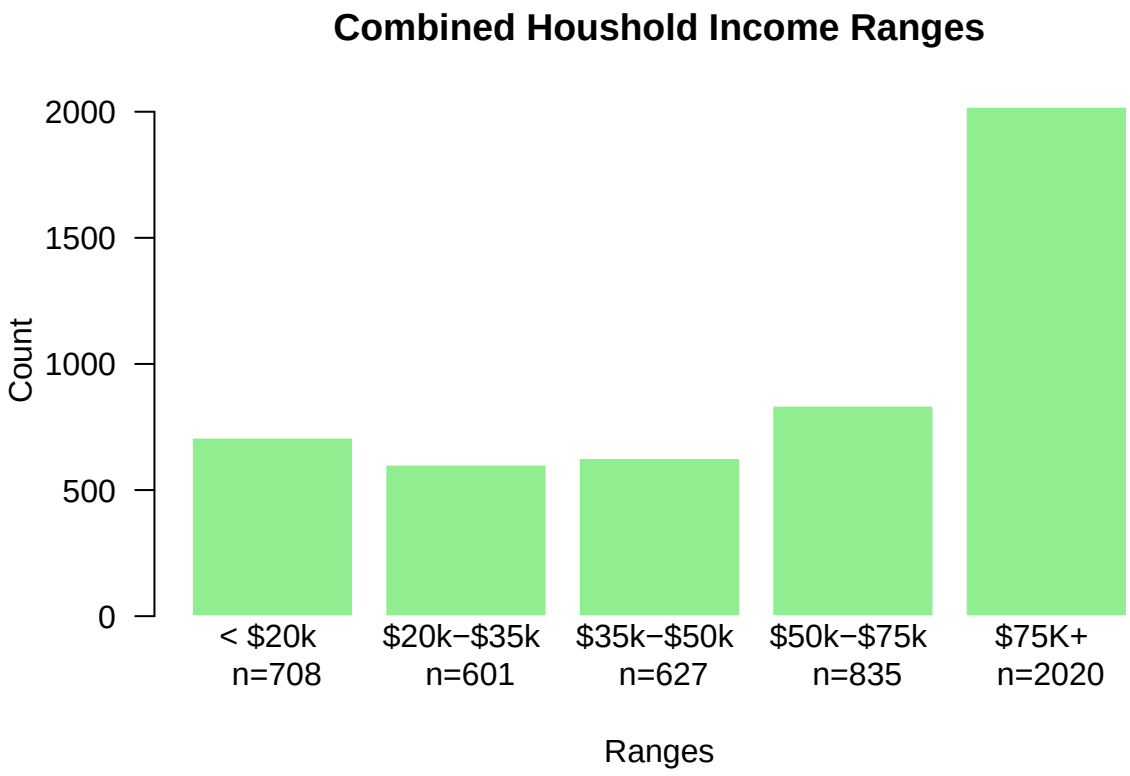



```r
# How many Children Under 18

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)
```

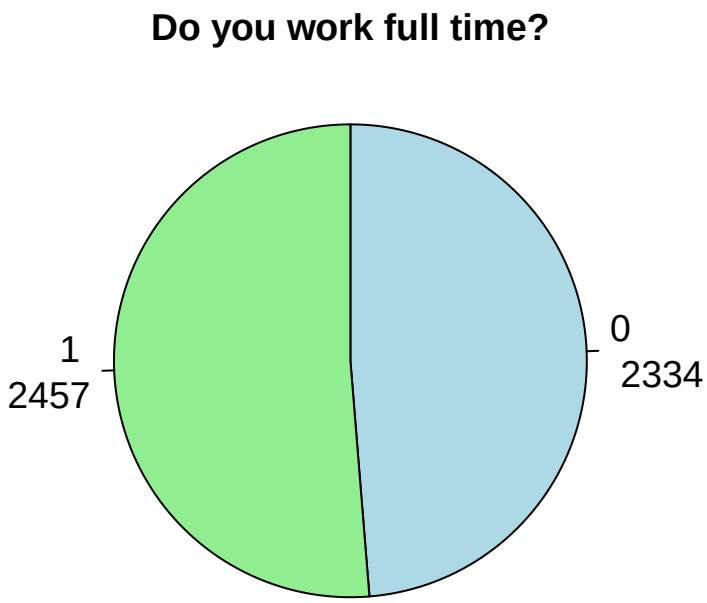


Economic Controls

```
#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)
```

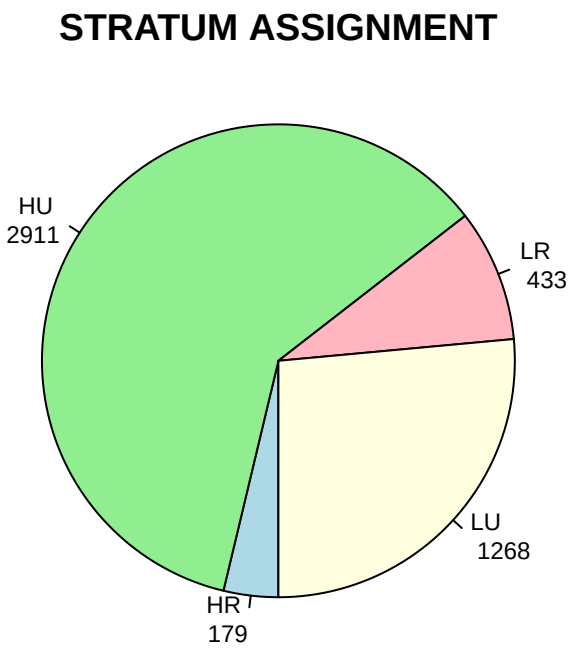



```r
#Full Time Work
freq_table <- table(data$WorkFullTime)

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
)
```



More Recoding for Data Analysis

```
#ACCESS TYPE
table(data$InternetAccessType)
```

```
##
##    1    2    3    4
##   24  706  260 3215
```

```
data$InternetAccessType_dummy <- ifelse(data$InternetAccessType == 4, 1, 0)
table(data$InternetAccessType_dummy)
```

```
##
##    0    1
##  990 3215
```

```
# DEVICE
table(data$HAVEDEVICE_CAT)
```

```
##
##    0    1    2    3    4
##   85  157 1452  256 2841
```

```
data$HAVEDEVICE_CAT_recode <- ifelse(data$HAVEDEVICE_CAT == 3, 0,
                                     ifelse(data$HAVEDEVICE_CAT == 4, 2, 1))
table(data$HAVEDEVICE_CAT_recode)
```

```
##
##    0    1    2
##  256 1694 2841
```

```
#VISTIED
table(data$SocMed_Visited)
```

```
##
##    1    2    3    4    5
##  822  279  366  589 2735
```

```
data$SocMed_Visited <- as.numeric(data$SocMed_Visited)

#CONFIDENCE
table(data$ConfidentInternetHealth)
```

```
##
##    1    2    3    4    5
##  206  413 1888 1692  592
```

```
data$ConfidentInternetHealth <- as.numeric(data$ConfidentInternetHealth)

#RACE
table(data$RaceEthn5)
```

```
##
##    1    2    3    4    5
## 2800  755  814  255  167
```

```
data$RaceEthn5_recoded <- factor(data$RaceEthn5,
                         levels = c(1, 2, 3, 4, 5),
                         label = c("NHWhite", "NHBlack", "Hispanic", "NHAsian", "NHOther"))

# Check the table after factoring
table(data$RaceEthn5_recoded)
```

```
##
##  NHWhite  NHBlack Hispanic  NHAsian  NHOther
##     2800      755      814      255      167
```

```
#EDUCATION
table(data$EducA)
```

```
##
##    1    2    3    4
##  240  798 1362 2391
```

```
data$EducA_recoded <- factor(data$EducA,
                         levels = c(1, 2, 3, 4),
                         label = c("< High School", "HS Grad", "Some College", "College Grad +"))

# Check the table after factoring
table(data$EducA_recoded)
```

```
##
##  < High School        HS Grad   Some College College Grad +
##            240            798           1362           2391
```

```r
#CHILDREN
data$ChildrenInHH_recode <- as.numeric(data$ChildrenInHH_recode)
```

LOGIT REGRESSIONS

```r
# Logistic Regression Models

## Model 1: Frequency of internet usage as a predictor
model1 <- glm(Electronic2_HealthInfo ~ SocMed_Visited , family = binomial, data = data)
summary(model1)
```

```
##
## Call:
## glm(formula = Electronic2_HealthInfo ~ SocMed_Visited, family = binomial,
##     data = data)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.39276    0.08003  -4.908 9.21e-07 ***
## SocMed_Visited  0.43065    0.02115  20.366  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5260.0  on 4790  degrees of freedom
## Residual deviance: 4836.5  on 4789  degrees of freedom
## AIC: 4840.5
##
## Number of Fisher Scoring iterations: 3
```

```r
## Model 2: Alongside Controls
model2 <- glm(Electronic2_HealthInfo ~ SocMed_Visited +
                InternetAccessType_dummy + HAVEDEVICE_CAT_recode + ConfidentInternetHealth +
                MisleadingHealthInfo, family = binomial, data = data)
summary(model2)
```

```
##
## Call:
## glm(formula = Electronic2_HealthInfo ~ SocMed_Visited + InternetAccessType_dummy +
##     HAVEDEVICE_CAT_recode + ConfidentInternetHealth + MisleadingHealthInfo,
##     family = binomial, data = data)
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -1.96106    0.23262  -8.430  < 2e-16 ***
## SocMed_Visited             0.02758    0.04002   0.689 0.490706
## InternetAccessType_dummy   0.39718    0.10594   3.749 0.000177 ***
## HAVEDEVICE_CAT_recode      0.40673    0.08283   4.910 9.09e-07 ***
## ConfidentInternetHealth    0.67053    0.05670  11.827  < 2e-16 ***
## MisleadingHealthInfo2      0.10626    0.45781   0.232 0.816453
## MisleadingHealthInfo3      0.36051    0.19076   1.890 0.058775 .
## MisleadingHealthInfo4      0.71408    0.16914   4.222 2.42e-05 ***
```

```
## MisleadingHealthInfo5     0.80294    0.17057   4.707 2.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3288.6  on 4204  degrees of freedom
## Residual deviance: 2984.6  on 4196  degrees of freedom
##   (586 observations deleted due to missingness)
## AIC: 3002.6
##
## Number of Fisher Scoring iterations: 5
```

```
## Model 3: Adding demographic factors
model3 <- glm(Electronic2_HealthInfo ~ SocMed_Visited +
                InternetAccessType_dummy + HAVEDEVICE_CAT_recode + ConfidentInternetHealth +
                MisleadingHealthInfo +
                HealthStatus+ Age+ BirthGender+ RaceEthn5_recoded+ EducA_recoded + PR_RUCA_2010 +
                ChildrenInHH_recode, family = binomial, data = data)
summary(model3)
```

```
##
## Call:
## glm(formula = Electronic2_HealthInfo ~ SocMed_Visited + InternetAccessType_dummy +
##     HAVEDEVICE_CAT_recode + ConfidentInternetHealth + MisleadingHealthInfo +
##     HealthStatus + Age + BirthGender + RaceEthn5_recoded + EducA_recoded +
##     PR_RUCA_2010 + ChildrenInHH_recode, family = binomial, data = data)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -2.634513   0.434991  -6.056 1.39e-09 ***
## SocMed_Visited               0.022322   0.041420   0.539 0.589945
## InternetAccessType_dummy     0.313614   0.113244   2.769 0.005617 **
## HAVEDEVICE_CAT_recode        0.339854   0.086112   3.947 7.93e-05 ***
## ConfidentInternetHealth      0.640111   0.058870  10.873  < 2e-16 ***
## MisleadingHealthInfo2        0.252874   0.464404   0.545 0.586089
## MisleadingHealthInfo3        0.472667   0.196729   2.403 0.016277 *
## MisleadingHealthInfo4        0.711372   0.173823   4.093 4.27e-05 ***
## MisleadingHealthInfo5        0.736622   0.176922   4.164 3.13e-05 ***
## HealthStatus                 0.396203   0.109135   3.630 0.000283 ***
## Age                         -0.002899   0.003620  -0.801 0.423255
## BirthGender                 -0.478655   0.099724  -4.800 1.59e-06 ***
## RaceEthn5_recodedNHBlack    -0.432701   0.136572  -3.168 0.001533 **
## RaceEthn5_recodedHispanic   -0.091097   0.142040  -0.641 0.521298
## RaceEthn5_recodedNHAsian     0.186621   0.254123   0.734 0.462721
## RaceEthn5_recodedNHOther    -0.230423   0.260700  -0.884 0.376769
## EducA_recodedHS Grad         0.444005   0.220761   2.011 0.044300 *
## EducA_recodedSome College    0.919110   0.212801   4.319 1.57e-05 ***
## EducA_recodedCollege Grad +  1.609023   0.213549   7.535 4.89e-14 ***
## PR_RUCA_2010                 1.206912   0.529472   2.279 0.022639 *
## ChildrenInHH_recode          0.122188   0.069457   1.759 0.078544 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 3288.6  on 4204  degrees of freedom
## Residual deviance: 2820.9  on 4184  degrees of freedom
##   (586 observations deleted due to missingness)
## AIC: 2862.9
##
## Number of Fisher Scoring iterations: 5
```

```
## Model 4: Adding demographic and economic factors
model4 <- glm(Electronic2_HealthInfo ~ SocMed_Visited +
                InternetAccessType_dummy + HAVEDEVICE_CAT_recode + ConfidentInternetHealth +
                MisleadingHealthInfo +
                HealthStatus+ Age+ BirthGender+ RaceEthn5_recoded+ EducA_recoded + PR_RUCA_2010 +
                ChildrenInHH_recode + HHInc + WorkFullTime, family = binomial, data = data)
summary(model4)
```

```
##
## Call:
## glm(formula = Electronic2_HealthInfo ~ SocMed_Visited + InternetAccessType_dummy +
##     HAVEDEVICE_CAT_recode + ConfidentInternetHealth + MisleadingHealthInfo +
##     HealthStatus + Age + BirthGender + RaceEthn5_recoded + EducA_recoded +
##     PR_RUCA_2010 + ChildrenInHH_recode + HHInc + WorkFullTime,
##     family = binomial, data = data)
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -2.723792   0.445644  -6.112 9.84e-10 ***
## SocMed_Visited              0.017525   0.041546   0.422 0.673151
## InternetAccessType_dummy    0.263120   0.114537   2.297 0.021605 *
## HAVEDEVICE_CAT_recode       0.276613   0.087442   3.163 0.001559 **
## ConfidentInternetHealth     0.623110   0.058922  10.575  < 2e-16 ***
## MisleadingHealthInfo2       0.398562   0.463518   0.860 0.389865
## MisleadingHealthInfo3       0.554347   0.198505   2.793 0.005229 **
## MisleadingHealthInfo4       0.743364   0.174494   4.260 2.04e-05 ***
## MisleadingHealthInfo5       0.748459   0.177652   4.213 2.52e-05 ***
## HealthStatus                0.421652   0.110337   3.821 0.000133 ***
## Age                        -0.005470   0.003839  -1.425 0.154172
## BirthGender                -0.523544   0.101422  -5.162 2.44e-07 ***
## RaceEthn5_recodedNHBlack   -0.322966   0.139184  -2.320 0.020317 *
## RaceEthn5_recodedHispanic  -0.024896   0.143118  -0.174 0.861900
## RaceEthn5_recodedNHAsian    0.211336   0.254979   0.829 0.407198
## RaceEthn5_recodedNHOther   -0.183339   0.261189  -0.702 0.482716
## EducA_recodedHS Grad        0.353611   0.221845   1.594 0.110946
## EducA_recodedSome College   0.769278   0.215193   3.575 0.000350 ***
## EducA_recodedCollege Grad +  1.339382   0.221210   6.055 1.41e-09 ***
## PR_RUCA_2010                1.204253   0.529176   2.276 0.022863 *
## ChildrenInHH_recode         0.106672   0.069659   1.531 0.125684
## HHInc                       0.192698   0.038704   4.979 6.40e-07 ***
## WorkFullTime               -0.198635   0.115988  -1.713 0.086796 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3288.6  on 4204  degrees of freedom
```

```
## Residual deviance: 2796.1  on 4182  degrees of freedom
##   (586 observations deleted due to missingness)
## AIC: 2842.1
##
## Number of Fisher Scoring iterations: 5
```

SUMMARY OF MODEL $

```
model_sum <- list("Model 4" = model4)
modelsummary(model_sum, output = "huxtable")
```

SUMMARY OF ALL MODELS

```
model_list<-list("Baseline" = model1, "Health Related Controls" = model2,
                 "Add Demographics" = model3,
                 "Add Economics" = model4)
modelsummary(model_list, output = "huxtable")
```

GLOBAL F TEST

```
anova(model3, model4)
```

MARGINAL EFFECTS

AVERAGE PREDICTIONS

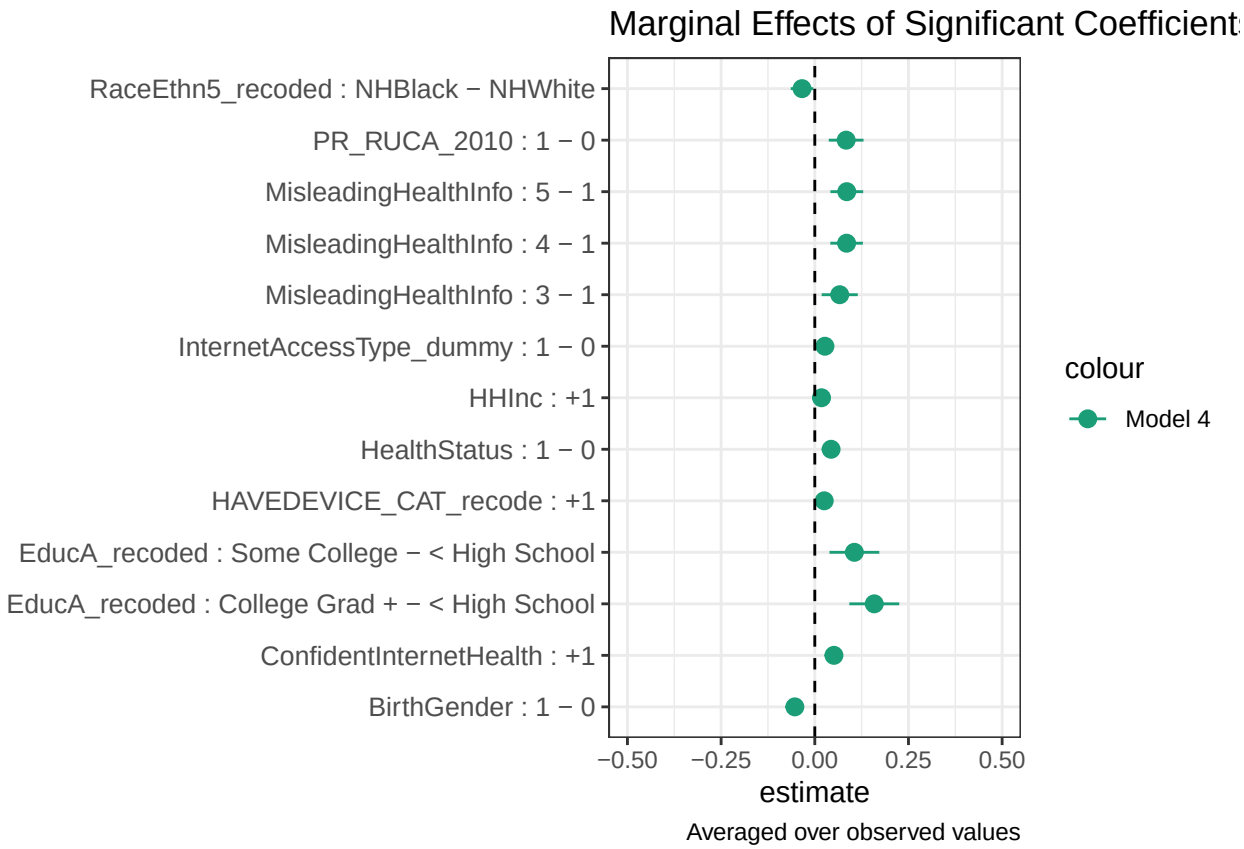
```
predicitions <- avg_predictions(model4, variables='SocMed_Visited')|>
  tibble()

predicitions
```

```
#put in presentable table

predictions <- avg_predictions(model4, variables = 'SocMed_Visited') |>
  tibble()

predictions$SocMed_Visited <- c("Never", "Rarely", "Occasionally", "Frequently", "Every Day")

# Create and display the table with custom row names
predictions |>
  gt() |>
  tab_header(
    title = "Average Predicted Probabilities for SocMed_Visited",
    subtitle = "Based on Logistic Regression Model"
  ) |>
  fmt_number(
    columns = everything(),
    decimals = 3
  ) |>
  cols_label(
```

```
    SocMed_Visited = "Social Media Usage",
    estimate = "Estimate",
    std.error = "Std. Error",
    statistic = "Statistic",
    p.value = "P-value",
    s.value = "S-value",
    conf.low = "Conf. Low",
    conf.high = "Conf. High"
  )
```

COUNTERFACTUAL APPROACH

```
mfx <- avg_comparisons(model4) |>
  tidy() |>
  mutate(contrast = paste(term, ":", contrast))

mfx_significant <- mfx |>
  filter(p.value < 0.05)

ggplot(mfx_significant, aes(x = estimate, y = contrast, color= "Model 4",
                xmin = conf.low,
                xmax = conf.high)) +
  geom_pointrange(position = position_dodge(width = 0.5)) +
  theme_bw() +
  xlim(c(-0.5, 0.5)) +
  geom_vline(xintercept = 0, lty = 2) +
  scale_color_brewer(palette = "Dark2") +
  labs(
    title = "Marginal Effects of Significant Coefficients",
    caption = "Averaged over observed values"
  ) +
  theme(
    axis.title.y = element_blank(),
    axis.text.y = element_text(size = 10)
  )
```



CHECK THE MARGINAL EFFECTS

```
#HIGH SM USAGE
logit_effect1 <-predict(model4,
                        newdata=data.frame(
                          SocMed_Visited = 5,
                          InternetAccessType_dummy = 1,
                          HAVEDEVICE_CAT_recode = 2,
                          ConfidentInternetHealth =5,
                          MisleadingHealthInfo = "5",
                          HealthStatus = 1,
                          BirthGender= 0,
                          RaceEthn5_recoded = "NHWhite",
                          EducA_recoded = "College Grad +",
                          PR_RUCA_2010 = 1,
                          HHInc = 5,
                          Age = 56,
                          WorkFullTime= 1,
                          ChildrenInHH_recode = 0),
                          type='response')


logit_effect2 <-predict(model4,
                        newdata=data.frame(
                          SocMed_Visited = 5,
```

```
                            InternetAccessType_dummy = 0,
                            HAVEDEVICE_CAT_recode = 0,
                            ConfidentInternetHealth =1,
                            MisleadingHealthInfo = "1",
                            HealthStatus = 0,
                            BirthGender= 1,
                            RaceEthn5_recoded = "NHBlack",
                            EducA_recoded = "< High School",
                            PR_RUCA_2010 = 0,
                            HHInc = 1,
                            Age = 56,
                            WorkFullTime= 1,
                            ChildrenInHH_recode = 0),
                            type='response')

logit_effect1- logit_effect2
```

```
##         1
## 0.955558
```

```
# LOW SM USAGE
logit_effect3 <-predict(model4,
                        newdata=data.frame(
                          SocMed_Visited = 1,
                          InternetAccessType_dummy = 1,
                          HAVEDEVICE_CAT_recode = 2,
                          ConfidentInternetHealth =5,
                          MisleadingHealthInfo = "5",
                          HealthStatus = 1,
                          BirthGender= 0,
                          RaceEthn5_recoded = "NHWhite",
                          EducA_recoded = "College Grad +",
                          PR_RUCA_2010 = 1,
                          HHInc = 5,
                          Age = 56,
                          WorkFullTime= 1,
                          ChildrenInHH_recode = 0),
                          type='response')


logit_effect4 <-predict(model4,
                        newdata=data.frame(
                          SocMed_Visited = 1,
                          InternetAccessType_dummy = 0,
                          HAVEDEVICE_CAT_recode = 0,
                          ConfidentInternetHealth =1,
                          MisleadingHealthInfo = "1",
                          HealthStatus = 0,
                          BirthGender= 1,
                          RaceEthn5_recoded = "NHBlack",
                          EducA_recoded = "< High School",
                          PR_RUCA_2010 = 0,
                          HHInc = 1,
                          Age = 56,
```

```
                          WorkFullTime= 1,
                          ChildrenInHH_recode = 0),
                          type='response')

logit_effect3- logit_effect4
```

```
##         1
## 0.9578761
```

DIAGNOSTICS AND ROBUSTNESS CHECKS

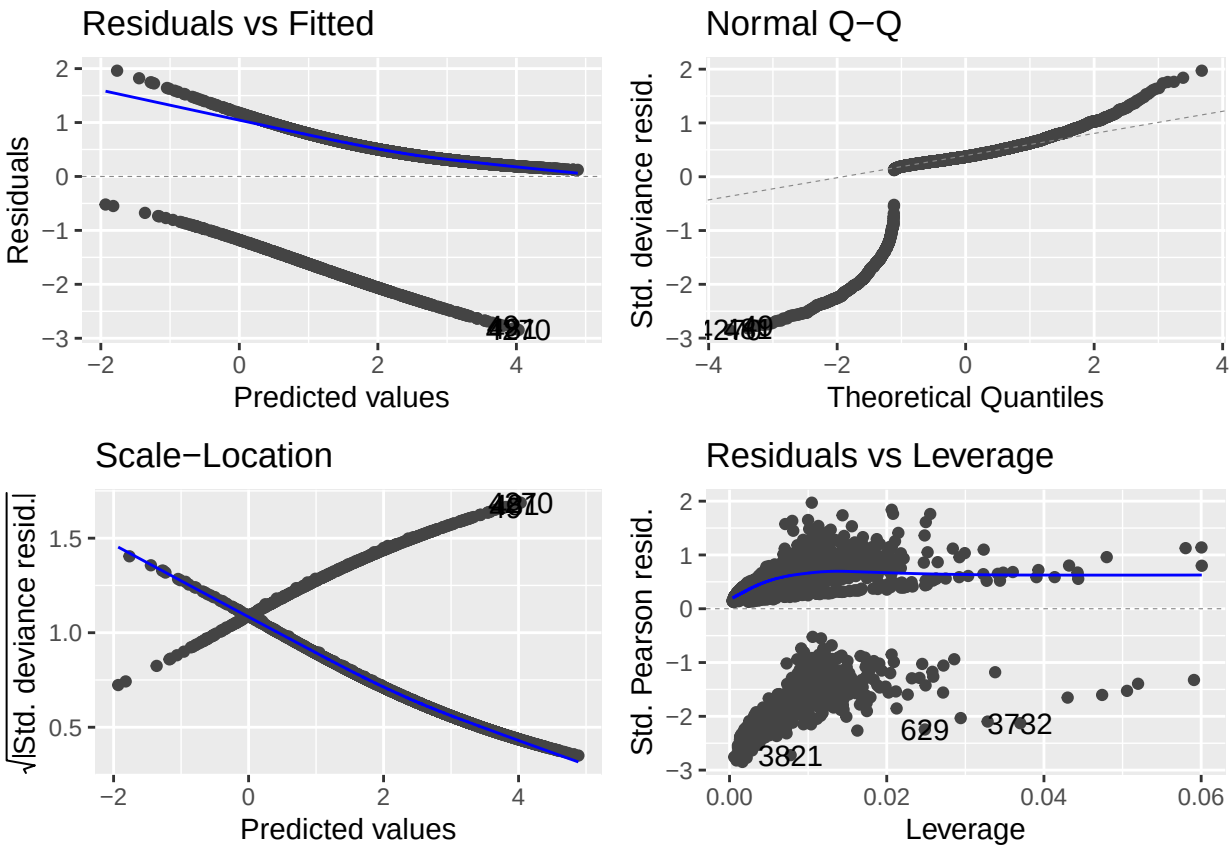
Muliti-colinearity

```
vif(model4)
```

```
##                             GVIF Df GVIF^(1/(2*Df))
## SocMed_Visited          1.771390  1        1.330936
## InternetAccessType_dummy 1.211956  1        1.100889
## HAVEDEVICE_CAT_recode   1.128336  1        1.062231
## ConfidentInternetHealth 1.092967  1        1.045451
## MisleadingHealthInfo    1.919090  4        1.084893
## HealthStatus            1.140703  1        1.068037
## Age                     1.743546  1        1.320434
## BirthGender             1.066099  1        1.032521
## RaceEthn5_recoded       1.223364  4        1.025521
## EducA_recoded           1.235169  3        1.035828
## PR_RUCA_2010            1.010404  1        1.005188
## ChildrenInHH_recode     1.226045  1        1.107269
## HHInc                   1.444332  1        1.201804
## WorkFullTime            1.397146  1        1.182009
```

Autoplot

```
autoplot(model4)
```



Robust Standard Errors

```
robust_se <- sqrt(diag(vcovHC(model4, type = "HC0")))
coefficients <- coef(model4)  # Get model coefficients
z_stats <- coefficients / robust_se
conf_int_low <- coefficients - 1.96 * robust_se
conf_int_high <- coefficients + 1.96 * robust_se

results_table <- data.frame(
  Variable = names(coefficients),
  Coefficient = coefficients,
  Robust_SE = robust_se,
  Z_Statistic = z_stats,
  Conf_Low = conf_int_low,
  Conf_High = conf_int_high
)


# Create the table and format it
results_table %>%
  gt() %>%
  tab_header(
    title = "Robust Standard Errors and Model Statistics"
  ) %>%
  cols_label(
    Variable = "Coefficient",
    Coefficient = "Estimate",
```

```
    Robust_SE = "Robust SE",
    Z_Statistic = "Z-Statistic",
    Conf_Low = "95% Conf. Low",
    Conf_High = "95% Conf. High"
  ) %>%
  fmt_number(
    columns = vars(Coefficient, Robust_SE, Z_Statistic, Conf_Low, Conf_High),
    decimals = 3
  )
```

```
## Warning: Since gt v0.3.0, `columns = vars(...)` has been deprecated.
## * Please use `columns = c(...)` instead.
```

	Model 4
(Intercept)	-2.724
	(0.446)
SocMed_Visited	0.018
	(0.042)
InternetAccessType_dummy	0.263
	(0.115)
HAVEDEVICE_CAT_recode	0.277
	(0.087)
ConfidentInternetHealth	0.623
	(0.059)
MisleadingHealthInfo2	0.399
	(0.464)
MisleadingHealthInfo3	0.554
	(0.199)
MisleadingHealthInfo4	0.743
	(0.174)
MisleadingHealthInfo5	0.748
	(0.178)
HealthStatus	0.422
	(0.110)
Age	-0.005
	(0.004)
BirthGender	-0.524
	(0.101)
RaceEthn5_recodedNHBlack	-0.323
	(0.139)
RaceEthn5_recodedHispanic	-0.025
	(0.143)
RaceEthn5_recodedNHAsian	0.211
	(0.255)
RaceEthn5_recodedNHOther	-0.183
	(0.261)

	Baseline	Health Related Controls	Add Demographics	Add Economics
(Intercept)	-0.393	-1.961	-2.635	-2.724
	(0.080)	(0.233)	(0.435)	(0.446)
SocMed_Visited	0.431	0.028	0.022	0.018
	(0.021)	(0.040)	(0.041)	(0.042)
InternetAccessType_dummy		0.397	0.314	0.263
		(0.106)	(0.113)	(0.115)
HAVEDEVICE_CAT_recode		0.407	0.340	0.277
		(0.083)	(0.086)	(0.087)
ConfidentInternetHealth		0.671	0.640	0.623
		(0.057)	(0.059)	(0.059)
MisleadingHealthInfo2		0.106	0.253	0.399
		(0.458)	(0.464)	(0.464)
MisleadingHealthInfo3		0.361	0.473	0.554
		(0.191)	(0.197)	(0.199)
MisleadingHealthInfo4		0.714	0.711	0.743
		(0.169)	(0.174)	(0.174)
MisleadingHealthInfo5		0.803	0.737	0.748
		(0.171)	(0.177)	(0.178)
HealthStatus			0.396	0.422
			(0.109)	(0.110)
Age			-0.003	-0.005
			(0.004)	(0.004)
BirthGender			-0.479	-0.524
			(0.100)	(0.101)
RaceEthn5_recodedNHBlack			-0.433	-0.323
			(0.137)	(0.139)
RaceEthn5_recodedHispanic			-0.091	-0.025
			(0.142)	(0.143)
RaceEthn5_recodedNHAsian			0.187	0.211
			(0.254)	(0.255)
RaceEthn5_recodedNHOther			-0.230	-0.183
			(0.261)	(0.261)

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
4.18e+03	2.82e+03			
4.18e+03	2.8e+03	2	24.8	4.17e-06

SocMed_Visited	estimate	std.error	statistic	p.value	s.value	conf.low	conf.high
1	0.863	0.0128	67.2	0	Inf	0.837	0.888
2	0.864	0.00899	96.1	0	Inf	0.847	0.882
3	0.866	0.00589	147	0	Inf	0.855	0.878
4	0.868	0.00493	176	0	Inf	0.858	0.878
5	0.87	0.00689	126	0	Inf	0.856	0.883

Average Predicted Probabilities for SocMed_Visited

Based on Logistic Regression Model

Social Media Usage	Estimate	Std. Error	Statistic	P-value	S-value	Conf. Low	Conf. High
Never	0.863	0.013	67.175	0.000	Inf	0.837	0.888
Rarely	0.864	0.009	96.131	0.000	Inf	0.847	0.882
Occasionally	0.866	0.006	147.024	0.000	Inf	0.855	0.878
Frequently	0.868	0.005	176.070	0.000	Inf	0.858	0.878
Every Day	0.870	0.007	126.162	0.000	Inf	0.856	0.883

Robust Standard Errors and Model Statistics

Coefficient	Estimate	Robust SE	Z-Statistic	95% Conf. Low	95% Conf. High
(Intercept)	-2.724	0.462	-5.890	-3.630	-1.817
SocMed_Visited	0.018	0.044	0.402	-0.068	0.103
InternetAccessType_dummy	0.263	0.116	2.268	0.036	0.491
HAVEDEVICE_CAT_recode	0.277	0.087	3.166	0.105	0.448
ConfidentInternetHealth	0.623	0.063	9.963	0.501	0.746
MisleadingHealthInfo2	0.399	0.450	0.886	-0.483	1.280
MisleadingHealthInfo3	0.554	0.204	2.713	0.154	0.955
MisleadingHealthInfo4	0.743	0.182	4.086	0.387	1.100
MisleadingHealthInfo5	0.748	0.184	4.075	0.388	1.108
HealthStatus	0.422	0.109	3.881	0.209	0.635
Age	-0.005	0.004	-1.408	-0.013	0.002
BirthGender	-0.524	0.100	-5.247	-0.719	-0.328
RaceEthn5_recodedNHBlack	-0.323	0.139	-2.320	-0.596	-0.050
RaceEthn5_recodedHispanic	-0.025	0.140	-0.178	-0.299	0.250
RaceEthn5_recodedNHAsian	0.211	0.257	0.822	-0.293	0.715
RaceEthn5_recodedNHOther	-0.183	0.285	-0.644	-0.741	0.375
EducA_recodedHS Grad	0.354	0.239	1.482	-0.114	0.821
EducA_recodedSome College	0.769	0.233	3.295	0.312	1.227
EducA_recodedCollege Grad +	1.339	0.239	5.608	0.871	1.808
PR_RUCA_2010	1.204	0.531	2.269	0.164	2.245
ChildrenInHH_recode	0.107	0.073	1.469	-0.036	0.249
HHInc	0.193	0.038	5.035	0.118	0.268
WorkFullTime	-0.199	0.117	-1.701	-0.427	0.030