````
---
title: "APEX Colorectal Cancer Screening - 2022 BRFSS data"
format: html
editor: Swathi.Pagadala
Date: 09/30/2024
---

## Quarto

**Required packages**

```{r}
pacman::p_load("tidyverse", "janitor", "tidylog", "survey", "foreign", "magrittr")
library(tidyverse)
install.packages("reshape2")
library(reshape2)
library(knitr)
library(tidyr)
install.packages("survey")
library(survey)
install.packages("gtsummary")
install.packages("flextable")
library(gtsummary)
library(flextable)
install.packages("writexl")
library(writexl)
```

**Here we import the raw 2022 BRFSS file**

```{r}
BRFSS.Raw <- read.xport("C:/Users/Intern/Desktop/LLCP2022.XPT")
```

**Triming down the data set to things that we think we want**

```{r}
BRFSS.Select <- BRFSS.Raw %>% select(X_STATE, X_LLCPWT, X_LLCPWT2, X_STSTR, X_PSU,
X_RACE1, X_SEX, X_AGEG5YR, X_EDUCAG, INCOME3,CHECKUP1,EMPLOY1, MSCODE, X_METSTAT,
X_URBSTAT,MEDCOST1,PERSDOC3,X_HLTHPLN,HADSIGM4, COLNSIGM,COLNTES1, SIGMTES1,
LASTSIG4, COLNCNCR, VIRCOLO1, VCLNTES2, SMALSTOL, STOLTEST, STOOLDN2, BLDSTFIT,
SDNATES1,X_HADCOLN, X_CLNSCP1, X_HADSIGM, X_SGMSCP1, X_SGMS101, X_RFBLDS5,
X_STOLDN1, X_VIRCOL1, X_SBONTI1, X_CRCREC2, X_INCOMG1) # DEDICATED HEALTH CARE
PROVIDER
FIPS <- c("1", "2", "4", "5", "6", "8", "9", "10", "11", "12", "13", "15", "16",
"17", "18", "19", "20", "21", "22", "23", "24", "25", "26", "27", "28", "29", "30",
"31", "32", "33", "34", "35", "36", "37", "38", "39", "40", "41", "42", "44", "45",
"46", "47", "48", "49", "50", "51", "53", "54", "55", "56")
State_Code <- c("AL", "AK", "AZ", "AR", "CA", "CO", "CT", "DE", "DC", "FL", "GA",
"HI", "ID", "IL", "IN", "IA", "KS", "KY", "LA", "ME", "MD", "MA", "MI", "MN", "MS",
````
````

```
"MO", "MT", "NE", "NV", "NH", "NJ", "NM", "NY", "NC", "ND", "OH", "OK", "OR", "PA",
"RI", "SC", "SD", "TN", "TX", "UT", "VT", "VA", "WA", "WV", "WI", "WY")
State <- c("Alabama", "Alaska", "Arizona", "Arkansas", "California", "Colorado",
"Connecticut", "Delaware", "District of Columbia", "Florida", "Georgia", "Hawaii",
"Idaho", "Illinois", "Indiana", "Iowa", "Kansas", "Kentucky", "Louisiana", "Maine",
"Maryland", "Massachusetts", "Michigan", "Minnesota", "Mississippi", "Missouri",
"Montana", "Nebraska", "Nevada", "New Hampshire", "New Jersey", "New Mexico", "New
York", "North Carolina", "North Dakota", "Ohio", "Oklahoma", "Oregon",
"Pennsylvania", "Rhode Island", "South Carolina", "South Dakota", "Tennessee",
"Texas", "Utah", "Vermont", "Virginia", "Washington", "West Virginia", "Wisconsin",
"Wyoming")
State.Info <- data.frame(FIPS, State_Code, State)
# Convert the X_STATE column from numeric to character
BRFSS.Select$X_STATE <- as.character(BRFSS.Select$X_STATE)


```

Create indicators for what we are interested

```{r}
# We want to create indicators for what we are interested

BRFSS.Select %<>% mutate(ALLCRCtests = case_when(X_CRCREC2 == 1 ~ TRUE,
                                                 X_CRCREC2 == 2 ~ FALSE,
                                                 X_CRCREC2 == 3 ~ FALSE,
                                                 TRUE ~ NA))
```

Create sub population groups

```{r}
BRFSS.Select %<>% mutate (AGE = case_when(X_AGEG5YR == 1~ "18-24",
                                          X_AGEG5YR == 2~ "25-29",
                                          X_AGEG5YR == 3~ "30-34",
                                          X_AGEG5YR == 4~ "35-39",
                                          X_AGEG5YR == 5~ "40-44",
                                          X_AGEG5YR == 6~ "45-49",
                                          X_AGEG5YR == 7~ "50-54",
                                          X_AGEG5YR == 8~ "55-59",
                                          X_AGEG5YR == 9~ "60-64",
                                          X_AGEG5YR == 10~ "65-69",
                                          X_AGEG5YR == 11~ "70-74",
                                          X_AGEG5YR == 12~ "75-79",
                                          X_AGEG5YR == 13~ "80 AND MORE",
                                          TRUE ~ NA))


BRFSS.Select %<>% mutate(EDUCATION = case_when(X_EDUCAG == 1 ~ "LTHS",
                                               X_EDUCAG == 2 ~ "HS",
```

```
                                                X_EDUCAG == 3 ~ "SC",
                                                X_EDUCAG == 4 ~ "CG",
                                                TRUE ~ NA))

BRFSS.Select$EDUCATION <- factor(BRFSS.Select$EDUCATION,
                                 levels = c("LTHS", "HS", "SC", "CG"),
                                 labels = c("Less Than High School", "High School
Graduate", "Attended College", "College Graduate"))

BRFSS.Select %<>% mutate(INCOME = case_when(X_INCOMG1 == 1 ~ "LTH15K",
                                            X_INCOMG1 == 2 ~ "15-25K",
                                            X_INCOMG1  == 3 ~ "25-35K",
                                            X_INCOMG1 == 4 ~ "35-50K",
                                            X_INCOMG1== 5~ "50-100K",
                                            X_INCOMG1== 6~ "100-200K",
                                            X_INCOMG1== 7~ "200K AND MORE",
                                            TRUE ~ NA))

BRFSS.Select$INCOME <- factor(BRFSS.Select$INCOME,
                              levels = c("LTH15K", "15-25K", "25-35K",
"35-50K","50-100K","100-200K","200K AND MORE"),
                              labels = c("<15,000$", "$15,000-$25,000",
"$25,000-$35,000",
"$35,000-$50,000","$50,000-$100,000","$100,000-$200,000","$200,000 AND MORE"))

BRFSS.Select %<>% mutate(RACE = case_when(X_RACE1 == 1 ~ "WHITE",
                                          X_RACE1 == 2 ~ "BLACK",
                                          X_RACE1  == 3 ~ "AI/AN",
                                          X_RACE1 == 4 ~ "ASIAN",
                                          X_RACE1== 5~ "NH/OP",
                                          X_RACE1== 7~ "MULTIRACIAL",
                                          X_RACE1== 8~ "HISPANIC",
                                          TRUE ~ NA))
BRFSS.Select$RACE <- factor(BRFSS.Select$RACE,
                            levels = c("WHITE", "BLACK", "AI/AN",
"ASIAN","NH/OP","MULTIRACIAL","HISPANIC"),
                            labels = c("WHITE", "BLACK", "AI/AN",
"ASIAN","NH/OP","MULTIRACIAL","HISPANIC"))

BRFSS.Select %<>% mutate(AVOIDEDCARE = case_when(MEDCOST1 == 1 ~"YES",
                                                 MEDCOST1 == 2 ~"NO",
                                                 TRUE ~ NA))
BRFSS.Select %<>% mutate(HCPROVIDER = case_when(PERSDOC3 == 1 ~ "YES",
                                                PERSDOC3 == 2 ~ "YES",
                                                PERSDOC3==3~ "NO",
                                                TRUE ~ NA))
BRFSS.Select %<>% mutate(INSURANCE = case_when(X_HLTHPLN == 1 ~ "YES",
                                               X_HLTHPLN == 2 ~ "No",
                                               TRUE ~ NA))
```

```
BRFSS.Select$SEX <- factor(BRFSS.Select$X_SEX,
                           levels = c("1", "2"),
                           labels = c("MALE", "FEMALE"))

```

Checking all the tables

```{r}
table(BRFSS.Select$AVOIDEDCARE, useNA = "always")
table(BRFSS.Select$HCPROVIDER, useNA = "always")
table(BRFSS.Select$INSURANCE, useNA = "always")
table(BRFSS.Select$X_METSTAT, useNA = "always")
table(BRFSS.Select$SEX, useNA = "always")
table(BRFSS.Select$INCOME, useNA = "always")
table(BRFSS.Select$EDUCATION, useNA = "always")
table(BRFSS.Select$RACE, useNA = "always")
table(BRFSS.Select$ALLCRCtests, useNA="always")
table (BRFSS.Select$AGE, useNA="always")
```

Mutating Age variable to groups 45-75

```{data <-BRFSS.Select %>%}
  filter(AGE %in% c ("45-49","50-54","55-59", "60-64","65-69","70-74","75-79"))

unique(data$AGE)

data_filtered<-data%>%
  mutate(Age =case_when(
    AGE %in% c("45-49","50-54") ~ "45-54",AGE %in% c("55-59", "60-64") ~ "50-64
Yrs",
    AGE %in% c("65-69","70-74","75-79")~ "65-75 Yrs"
  ))
```

Creating Table 1

```{r}
mytable1 <- tbl_summary(
data=data_filtered,
include
=c(Age,SEX,INCOME,EDUCATION,RACE,INSURANCE,AVOIDEDCARE,HCPROVIDER,X_METSTAT),
by= ALLCRCtests,
type=all_dichotomous()~"categorical",
label=list(SEX ~"Sex",
           Age ~"Age",
           INCOME ~"Income",
           EDUCATION ~"Education",
           RACE ~"Race",
```

```
           INSURANCE ~"Insurance",
           AVOIDEDCARE ~"Avoided care due to cost",
           HCPROVIDER~"Have 1 or more health care provider",
          X_METSTAT ~ "Metropolitan/Non Metropolitan"),

  missing="no"
  )%>%
  add_n()%>%
  add_overall ()%>%
  modify_header (label="")%>%
  modify_spanning_header(c("stat_1","stat_2") ~ "**Met USPSTF recommendations for
testing**")%>%
  add_stat_label()%>%
  bold_labels()

  print(mytable1)
```

Creating the Survey Object and setting the primary sample unit adjustment

```{r}
BRFSS.Design <- svydesign(
    id = ~X_PSU,
    strata = ~X_STSTR,
    nest = T,
    weights = ~X_LLCPWT,
    data = data_filtered)

  options(survey.lonely.psu = "adjust")
```

Prevalence

```{r}
#National level

Results.Object <- svyciprop(~ ALLCRCtests, BRFSS.Design, method = "xlogit", level =
0.95, na.rm = TRUE)

Estimate <- as.numeric(Results.Object)
StandardError <- SE(Results.Object)
Results <- data.frame(Estimate, StandardError)


#State level

Results.AGE <- svyby(~ALLCRCtests, ~X_STATE + Age, BRFSS.Design, svyciprop, method
= "xlogit", level = 0.95, na.rm = TRUE)
Results.INC <- svyby(~ALLCRCtests, ~X_STATE + INCOME, BRFSS.Design, svyciprop,
method = "xlogit", level = 0.95, na.rm = TRUE)
```

```
Results.EDU <- svyby(~ALLCRCtests,~X_STATE+ EDUCATION,BRFSS.Design,svyciprop,
method = "xlogit", level = 0.95, na.rm = TRUE)
Results.sex <- svyby(~ALLCRCtests,~X_STATE+SEX,BRFSS.Design,svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.rrace <-svyby(~ALLCRCtests,~X_STATE+ RACE, BRFSS.Design,svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.state<- svyby(~ALLCRCtests, ~X_STATE, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)

# Join the State.Info database with the results of the state-level analysis
Results.state %<>% left_join(State.Info, by = c("X_STATE" = "FIPS"))


# This is National level survey with sub populations and ALL crc tests

Results.RACE1 <- svyby(~ALLCRCtests, ~ RACE, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.INCOME1 <- svyby(~ALLCRCtests, ~INCOME, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.INSURANCE1 <- svyby(~ALLCRCtests, ~INSURANCE, BRFSS.Design, svyciprop,
method = "xlogit", level = 0.95, na.rm = TRUE)
Results.EDUCATION1 <- svyby(~ALLCRCtests, ~EDUCATION, BRFSS.Design, svyciprop,
method = "xlogit", level = 0.95, na.rm = TRUE)
Results.SEX1 <- svyby(~ALLCRCtests, ~SEX, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.Age1 <- svyby(~ALLCRCtests, ~Age, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.hcP <-svyby(~ALLCRCtests, ~HCPROVIDER, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.MET <-svyby(~ALLCRCtests, ~X_METSTAT, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)
Results.States <-svyby(~ALLCRCtests, ~X_STATE, BRFSS.Design, svyciprop, method =
"xlogit", level = 0.95, na.rm = TRUE)

```

PLOTS OF DEMOGRAPHICS AND CRC

```{r}
ggplot(Results.EDUCATION1, aes(x= EDUCATION, y=ALLCRCtests, fill=EDUCATION))+
  geom_col(position="dodge")+
  geom_text(aes(label =scales::percent(ALLCRCtests, accuracy =0.01)),
                position = position_stack(vjust=0.9),
                colour ="white", size =3)+
  scale_y_continuous(labels= scales::percent)+
  labs(title ="education and crc tests",
       x="EDUCATION",
       y="Percentage of people who had CRC screening")+
  coord_flip()+
  theme_minimal()+
```

```
  theme (axis.text.x=element_text(angle=0, hjust=0.5, margin =margin(t=10)))

ggplot(Results.INCOME1, aes(x= INCOME, y=ALLCRCtests, fill=INCOME))+
  geom_col(position="dodge")+
  geom_text(aes(label =scales::percent(ALLCRCtests, accuracy =0.01)),
            position = position_stack(vjust=0.9),
            colour ="white", size =3)+
  scale_y_continuous(labels= scales::percent)+
  labs(title ="INCOME and crc tests",
       x="INCOME",
       y="Percentage of people who had CRC screening")+
  coord_flip()+
  theme_minimal()+
  theme (axis.text.x=element_text(angle=0, hjust=0.5, margin =margin(t=10)))

ggplot(Results.RACE1, aes(x= RACE, y=ALLCRCtests, fill=RACE))+
  geom_col(position="dodge")+
  geom_text(aes(label =scales::percent(ALLCRCtests, accuracy =0.01)),
            position = position_stack(vjust=0.9),
            colour ="white", size =3)+
  scale_y_continuous(labels= scales::percent)+
  labs(title ="RACE and crc tests",
       x="RACE",
       y="Percentage of people who had CRC screening")+
  coord_flip()+
  theme_minimal()+
  theme (axis.text.x=element_text(angle=0, hjust=0.5, margin =margin(t=10)))

ggplot(Results.SEX1, aes(x=SEX, y=ALLCRCtests, fill=SEX))+
  geom_col(position="dodge")+
  geom_text(aes(label =scales::percent(ALLCRCtests, accuracy =0.01)),
            position = position_stack(vjust=0.9),
            colour ="white", size =3)+
  scale_y_continuous(labels= scales::percent)+
  labs(title ="SEX and crc tests",
       x="SEX",
       y="People who had CRC screening")+
  theme_minimal()+
  theme (axis.text.x=element_text(angle=0, hjust=0.5, margin =margin(t=10)))

ggplot(Results.Age1, aes(x= Age, y=ALLCRCtests, fill=Age))+
  geom_col(position="dodge")+
  geom_text(aes(label =scales::percent(ALLCRCtests, accuracy =0.01)),
            position = position_stack(vjust=0.9),
            colour ="white", size =3)+
  scale_y_continuous(labels= scales::percent)+
  labs(title ="AGE and crc tests",
       x="AGE",
       y="People who had CRC screening")+
  theme_minimal()+
```

```
  theme (axis.text.x=element_text(angle=0, hjust=0.5, margin =margin(t=10)))
```