R Appendix

Courtney Johnson

December 15, 2019

# PAP SMEAR CODE  
  
# read data   
#getting pap data from cancer df  
pap\_dat <- read\_csv("./data/cancerxx.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, fpx, #identifiers  
 wtfa\_sa, #weights  
 strat\_p, psu\_p, #for design  
 region,   
 paphad1, # Ever had Pap smear/Pap test  
 papfrst1, #Age when had first Pap test  
 pap6yr1, # Number of Pap tests, last 6 years   
 rpap1\_m1, # Month of most recent Pap test   
 rpap1y1, # Year of most recent Pap test   
 rpap1n1, # Time ago date of most recent Pap test: # of units   
 rpap1t1, # Time ago date of most recent Pap test: Time unit   
 rpap21, # Most recent Pap test, time categories   
 rpap3a1, # Most recent Pap test, time categories (using 2005 method)   
 rpap3b1, # Most recent Pap test, time categories (using 2000 method)   
 hpvhrd, # Ever heard of HPV   
 hpvpap, # Had HPV test with Pap test   
 paprea2, # Main reason had Pap/Pap or HPV   
 papabn3, # Pap test results in last 3 years   
 papnot2, # Most important reason never had Pap or HPV test   
 mdrecp1, # Doctor recommended Pap test   
 paphpvpy) #Paid for Pap or HPV test out of pocket   
  
## getting covariates  
fam\_dat <- read\_csv("./data/familyxx/familyxx.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, #identifiers  
 rat\_cat4, rat\_cat5) # Ratio of family income to the poverty threshol  
  
pers\_dat <- read\_csv("./data/personsx/personsx.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, fpx, #identifiers  
 age\_p, #age  
 educ1, #education  
 sex, #gender  
 notcov, cover, cover65, cover65o, #coverage > 65, 65+, alternate 65+  
 la1ar, #limitation  
 lcondrt, #limitation is chronic  
 lachronr, #chronic limitation  
 hiscodi3, #ethnicity recode,  
 racreci3, yrsinus, plborn)#race recode  
  
adult\_dat <- read\_csv("./data/samadult/samadult.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, fpx, #identifiers  
 ausualpl, ahcplrou, ahcplknd, #Usual source of care - different options  
 fla1ar) #functional limitation  
  
# data manipulation  
  
# join into one pap smear data set  
pap\_dat <- pap\_dat %>%   
 left\_join(adult\_dat, by = c("hhx", "fmx", "fpx")) %>%   
 left\_join(pers\_dat, by = c("hhx", "fmx", "fpx")) %>%   
 left\_join(fam\_dat, by = c("hhx", "fmx"))  
  
##outcome##  
#disregarding clustered structure: looking for 3 or less as recommended  
pap\_dat %>%   
 count(rpap21)  
  
pap\_dat %>%   
 count(rpap3b1) #choose 2000 method of q for best comparison (also less NA)  
  
pap\_dat <- pap\_dat %>%   
 mutate(paprec\_3bcat = if\_else(rpap3b1 <= 3, 1, 0))  
  
##covariates##  
pap\_dat %>%   
 ggplot() +  
 geom\_histogram(aes(x = age\_p, weight = wtfa\_sa), bins = 10)  
  
pap\_dat <- pap\_dat %>%   
 mutate(age\_cat = case\_when(age\_p >= 25 & age\_p < 40 ~ "25–39",  
 age\_p >= 40 & age\_p < 50 ~ "40–49",  
 age\_p >= 50 & age\_p < 65 ~ "50–64",  
 age\_p >= 65 ~ "65+"))  
  
pap\_dat %>%   
 count(educ1)  
  
# Less than high school: < 13  
# High school graduate: 13 or 14  
# Some college or AA degree 15, 16, 17  
# College graduate (BA/BS) 18, 19, 20, 21  
  
pap\_dat <- pap\_dat %>%   
 mutate(educ\_cat = case\_when(educ1 < 13 ~ "Less than high school",  
 educ1 >= 13 & educ1 < 15 ~ "High school",  
 educ1 >= 15 & educ1 < 18 ~ "Some college",  
 educ1 >= 18 & educ1 <= 21 ~ "College graduate"))  
  
# Family income/poverty ratio   
# <200%   
# 200–299%  
# 300–399%   
# 400–499%   
# ≥500%  
  
# 01 Under 0.50 1718 4.06  
# 02 0.50 - 0.74 1566 3.70  
# 03 0.75 - 0.99 1953 4.62  
# 04 1.00 - 1.24 1852 4.38  
# 05 1.25 - 1.49 1657 3.92  
# 06 1.50 - 1.74 1658 3.92  
# 07 1.75 - 1.99 1500 3.55  
# 08 2.00 - 2.49 3145 7.44  
# 09 2.50 - 2.99 2504 5.92  
# 10 3.00 - 3.49 2339 5.53  
# 11 3.50 - 3.99 1746 4.13  
# 12 4.00 - 4.49 1834 4.34  
# 13 4.50 - 4.99 1316 3.11  
# 14 5.00 and over 8318 19.67  
# 15 Less than 1.00 (no further detail) 845 2.00  
# 16 1.00 - 1.99 (no further detail) 1381 3.27  
# 17 2.00 and over (no further detail) 3615 8.55  
# 96 Undefinable 548 1.30  
# 99 Unknown 2793 6.60   
  
pap\_dat %>% count(rat\_cat5)  
pap\_dat <- pap\_dat %>%   
 mutate(finc\_cat = case\_when(rat\_cat5 <= 7 | rat\_cat5 %in% c(15, 16) ~ "<200%",  
 rat\_cat5 %in% c(8, 9) ~ "200–299%",   
 rat\_cat5 %in% c(10, 11) ~ "300–399%",  
 rat\_cat5 >= 18 & educ1 <= 21 ~ "400–499%",  
 rat\_cat5 == 14 ~">=500%",  
 rat\_cat5 == 17 ~">=200%, no further detail",  
 rat\_cat5 %in% c(96, 99) ~ "Unknown")) %>%   
 mutate(finc\_cat = factor(finc\_cat, levels = c("<200%", "200–299%", "300–399%", "400–499%", ">=500%",   
 ">=200%, no further detail", "Unknown")))  
  
#usually go when sick  
# 1 Yes   
# 2 There is NO place   
# 3 There is MORE THAN ONE place   
# 7 Refused   
# 8 Not ascertained   
# 9 Don't know   
pap\_dat <- pap\_dat %>%   
 mutate(ausualpl\_cat = case\_when(ausualpl == 2 ~ "No",  
 ausualpl %in% c(1, 3) ~ "Yes",  
 ausualpl %in% c(7, 8, 9) ~ "Other"))  
  
# Health insurance  
# None  
# Public  
# Private/military  
# NOTCOV Frequency Percent  
# ------------------------------------------  
# 1 Not covered   
# 2 Covered   
# 7 Refused   
# 8 Not ascertained   
# 9 Don't know 1102   
# cover  
# 1 Private   
# 2 Medicaid and other public   
# 3 Other coverage   
# 4 Uninsured 10381   
# 5 Don't know 1030   
# COVER65 Frequency Percent  
# ---------------------------------------------------------------------  
# 1 Private   
# 2 Dual eligible   
# 3 Medicare Advantage   
# 4 Medicare only excluding Medicare Advantage   
# 5 Other coverage   
# 6 Uninsured   
# 7 Don't know   
  
pap\_dat %>% count(cover)  
pap\_dat %>% count(!is.na(cover65))  
pap\_dat %>% count(cover65o)  
  
pap\_dat <- pap\_dat %>%   
 mutate(cover\_cat = case\_when(notcov == 1 | cover == 4 | cover65 == 6 ~ "None",  
 cover == 2 | cover65 %in% 2:4 ~ "Public",  
 cover %in% c(1, 3) | cover65 %in% c(1, 5) ~ "Private/Military"))  
  
# Chronic disability  
# Yes  
# No  
# LCONDRT Frequency  
# 1 At least one condition causing limitation of activity is chronic   
# 2 No condition causing limitation of activity is chronic   
# 9 Unknown if any condition causing limitation of activity is chronic   
  
pap\_dat %>% count(lcondrt)  
pap\_dat <- pap\_dat %>%   
 mutate(lcond\_chronic\_cat = if\_else(lcondrt == 1, "Yes", "No"))  
  
  
# Race/ethnicity  
# hispanic, nonhispanic white, nonhispanic black, nonhispanic asian, nonhispanic alaska native/american indian  
pap\_dat %>% count(racreci3)  
pap\_dat %>% count(hiscodi3)  
# race (white, black, asian, alaska native/american indian)  
# RACRECI3 Frequency Percent  
# ------------------------------------------------------------------  
# 1 White   
# 2 Black   
# 3 Asian   
# 4 All other race groups (See file layout)   
# HHC.200\_01.000: Race/ethnicity recode  
# HISCODI3 Frequency Percent  
# -------------------------------------------------------------  
# 1 Hispanic   
# 2 Non-Hispanic White   
# 3 Non-Hispanic Black   
# 4 Non-Hispanic Asian   
# 5 Non-Hispanic All other race groups   
  
pap\_dat <- pap\_dat %>%   
 mutate(race\_cat = case\_when(racreci3 == 1 ~ "White",  
 racreci3 == 2 ~ "Black",  
 racreci3 == 3 ~ "Asian",  
 racreci3 == 4 ~ "AN/AI"),  
 eth\_cat = case\_when(hiscodi3 == 1 ~ "Hispanic",  
 hiscodi3 == 2 ~ "Non-Hispanic White",  
 hiscodi3 == 3 ~ "Non-Hispanic Black",  
 hiscodi3 == 4 ~ "Non-Hispanic Asian",  
 hiscodi3 == 5 ~ "Non-Hispanic AN/AI"))  
  
  
pap\_dat <- pap\_dat %>%  
 mutate(imm\_stat = case\_when(yrsinus < 4 ~ "In U.S. < 10 yrs",  
 yrsinus == 4 | yrsinus == 5 ~ "In U.S. >= 10 yrs",  
 plborn == 1 ~ "Born in U.S."))  
  
  
#create indicator inclusion criteria variable  
##filter age less than 25 per cancer paper  
##filter to only women  
pap\_dat %>% count(sex)  
pap\_dat <- pap\_dat %>%   
 mutate(inc = if\_else(age\_p >= 25 & sex == 2, 1, 0))  
#create indicator variable and include it in domain analysis   
  
# fit survey design  
  
des <- svydesign(ids = ~psu\_p, strata = ~strat\_p,   
 weights = ~wtfa\_sa, nest = TRUE, data = pap\_dat)  
pap\_dat %>% select(ends\_with("cat")) %>% names()  
pap\_dat %>% count(paprec\_3bcat) #unweighted sample size  
  
# get overll pap smear screening rate   
svyby(~paprec\_3bcat, by = ~inc, svymean, na.rm = TRUE, design = des, vartype = "ci") %>% filter(inc == 1)  
pap\_dat %>% count(paprec\_3bcat, inc) %>% drop\_na() %>% filter(inc == 1) %>% summarize(n = sum(n))  
  
# descripive statistics  
  
# function to get screening rate  
pct\_func <- function(outcome = "paprec\_3bcat", inclusion = "inc", var1 = "age\_cat", var2 = NULL) {  
 .outcome = reformulate(outcome)  
 .by = reformulate(c(inclusion, var1, var2))  
 svyby(.outcome , by = .by, svymean, na.rm = TRUE, design = des, vartype = "ci")   
   
}  
  
##replace this function with an unweighted total   
tot\_func <- function(outcome = "paprec\_3bcat", inclusion = "inc", var1 = "age\_cat", var2 = NULL) {  
outcome\_sym = rlang::sym(outcome)  
inc\_sym = rlang::sym(inclusion)  
var1\_sym = rlang::sym(var1)  
var2\_sym = ifelse(is.null(var2), rlang::sym(" "), rlang::sym(var2))  
  
if (is.null(var2)) {  
 pap\_dat %>%   
 count(!!outcome\_sym, !!inc\_sym, !!var1\_sym) %>% drop\_na() %>%   
 group\_by(!!var1\_sym, !!inc\_sym) %>% filter(!!inc\_sym == 1) %>%   
 summarize(n = sum(n))  
} else {  
 pap\_dat %>%   
 count(!!outcome\_sym, !!inc\_sym, !!var2\_sym, !!var1\_sym) %>% drop\_na() %>%   
 group\_by(!!var2\_sym, !!var1\_sym, !!inc\_sym) %>% filter(!!inc\_sym == 1) %>%   
 summarize(n = sum(n))  
  
}  
}  
  
tot\_func(var2 = NULL, var1 = "race\_cat")  
  
# create table  
pap\_by <- pap\_dat %>%   
 select(ends\_with("cat"), imm\_stat, -paprec\_3bcat) %>%   
 mutate(ausualpl\_cat = fct\_explicit\_na(ausualpl\_cat)) %>%   
 names() %>%   
 tibble(var = .) %>%   
 mutate(pct = map(var, ~pct\_func(var1 = .x))) %>%   
 mutate(tot = map(var, ~tot\_func(var1 = .x))) %>%   
 mutate(pct\_byage = map(var, ~pct\_func(var2 = "age\_cat", var1 = .x))) %>%   
 mutate(tot\_byage = map(var, ~tot\_func(var2 = "age\_cat", var1 = .x)))  
  
get\_comp\_tables <- function(tablepct, tabletot, var) {  
 tabletot <- tabletot %>% filter(inc == 1) %>% rename\_all(~paste0("t\_", .x))   
 tablepct %>%   
 filter(inc == 1) %>%   
 bind\_cols(tabletot) %>%   
 mutate\_at(vars(paprec\_3bcat, ci\_l, ci\_u), ~round(.x\*100, 1)) %>%   
 mutate(pct = paprec\_3bcat) %>%   
 mutate(pct\_ci = paste0("(", ci\_l, ", ", ci\_u, ")")) %>%   
 mutate(tot = t\_n) %>%   
 select(var, tot, pct, pct\_ci) %>%   
 rename(levels = var) %>%   
 as\_tibble()  
}  
  
pap\_by <- pap\_by %>%   
 mutate(comp\_tbl = pmap(list(x = pct, y = tot, z = var), function(x, y, z)   
 {get\_comp\_tables(tablepct = x, tabletot = y, var = z)} ))  
  
pap\_sel <- pap\_by %>%   
 select(var, comp\_tbl) %>%   
 unnest\_wider(comp\_tbl) %>%   
 unnest(-var)  
   
pap\_sel %>%   
 dplyr::select(var, levels, tot, pct, pct\_ci) %>%   
 filter(!levels %in% c("Unknown", "Other")) %>%   
 knitr::kable(names = c("Variable", "Levels", "Total", "Percent"))  
  
get\_by\_tables <- function(tablepct, tabletot, var) {  
 tabletot <- tabletot %>% filter(inc == 1) %>% rename\_all(~paste0("t\_", .x))   
 tablepct %>%   
 filter(inc == 1) %>%   
 bind\_cols(tabletot) %>%   
 mutate\_at(vars(paprec\_3bcat, ci\_l, ci\_u), ~round(.x\*100, 1)) %>%   
 mutate(pct = paste0(paprec\_3bcat, " (", ci\_l, ", ", ci\_u, ")")) %>%   
 mutate(tot = t\_n) %>%   
 select(age\_cat, var, pct, tot) %>%   
 rename(levels = var) %>%   
 as\_tibble()  
}  
  
new\_ausualpl\_tot <- (pap\_by %>%   
 filter(var == "ausualpl\_cat") %>%   
 pull(tot\_byage))[[1]] %>% ungroup() %>%   
 add\_case(age\_cat = "65+", ausualpl\_cat = "Other", inc = 1, n = 0)  
  
pap\_strat <- pap\_by %>%   
 dplyr::select(var, pct\_byage, tot\_byage) %>%   
 mutate(tot\_byage = if\_else(var == "ausualpl\_cat", list(new\_ausualpl\_tot), tot\_byage)) %>%   
 mutate(comp\_tbl = pmap(list(x = pct\_byage, y = tot\_byage, z = var), function(x, y, z)   
 {get\_by\_tables(tablepct = x, tabletot = y, var = z)} )) %>%   
 dplyr::select(var, comp\_tbl) %>% unnest()   
  
pap\_strat %>%   
 filter(var != "age\_cat") %>%   
 rename(n = tot) %>%   
 pivot\_wider(names\_from = age\_cat, values\_from = c(n, pct), names\_prefix = "Age\_", names\_sep = "\_") %>%   
 select(var, levels, ends\_with("39"), ends\_with("49"), ends\_with("64"), ends\_with("+")) %>%  
 knitr::kable()  
  
  
# barplots for presentation  
p1 <- pap\_by %>%   
 filter(var == "ausualpl\_cat") %>%   
 select(var, pct\_byage) %>%   
 unnest(pct\_byage) %>%   
 filter(inc == 1) %>%   
 filter(!ausualpl\_cat %in% c("Unknown", "Other")) %>%   
 ggplot(aes(x = ausualpl\_cat, y = 100\*paprec\_3bcat, fill = ausualpl\_cat)) +  
 geom\_col() +  
 geom\_errorbar(aes(ymin = 100\*ci\_l, ymax = 100\*ci\_u)) + ylim(0, 100) + coord\_flip() +  
 facet\_grid(~age\_cat) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") +   
 labs(y = "", x = "Usual Source of Care")  
  
p2 <- pap\_by %>%   
 filter(var == "cover\_cat") %>%   
 select(var, pct\_byage) %>%   
 unnest(pct\_byage) %>%   
 filter(inc == 1) %>%   
 filter(!cover\_cat %in% c("Unknown", "Other")) %>%   
 ggplot(aes(x = cover\_cat, y = 100\*paprec\_3bcat, fill = cover\_cat)) +  
 geom\_col() +  
 geom\_errorbar(aes(ymin = 100\*ci\_l, ymax = 100\*ci\_u)) +  
 facet\_grid(~age\_cat) +   
 ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") + ylim(0, 100) + coord\_flip() +  
 labs(y = "", x = "Insurance Coverage")  
  
p3 <- pap\_by %>%   
 filter(var == "educ\_cat") %>%   
 select(var, pct\_byage) %>%   
 unnest(pct\_byage) %>%   
 filter(inc == 1) %>%   
 ggplot(aes(x = educ\_cat, y = 100\*paprec\_3bcat, fill = educ\_cat)) +  
 geom\_col() +  
 geom\_errorbar(aes(ymin = 100\*ci\_l, ymax = 100\*ci\_u)) +  
 facet\_grid(~age\_cat) +   
 ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") + ylim(0, 100) +  
 coord\_flip() +   
 labs(y = "Percent Had Pap Smear, Last 3 years", x = "Education")  
  
p4 <- pap\_by %>%   
 filter(var == "imm\_stat") %>%   
 select(var, pct\_byage) %>%   
 unnest(pct\_byage) %>%   
 filter(inc == 1) %>%   
 ggplot(aes(x = imm\_stat, y = 100\*paprec\_3bcat, fill = imm\_stat)) +  
 geom\_col() +  
 geom\_errorbar(aes(ymin = 100\*ci\_l, ymax = 100\*ci\_u)) +  
 facet\_grid(~age\_cat) +   
 ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") + ylim(0, 100) +  
 coord\_flip() +   
 labs(y = "Percent Had Pap Smear, Last 3 years", x = "Immigration Status")  
  
library(patchwork)  
p1 / p2 / p3   
  
pap\_formod <- pap\_dat %>%   
 select(psu\_p, strat\_p, wtfa\_sa, ends\_with("cat"), imm\_stat, hpvhrd, paphad1, mdrecp1, inc) %>%   
 mutate(imm\_stat = if\_else(imm\_stat == "Born in U.S.", "Born in U.S.", "Immigrated")) %>%   
 mutate(lcond\_chronic\_cat = forcats::fct\_explicit\_na(lcond\_chronic\_cat, "None Reported")) %>%   
 mutate(finc\_cat = case\_when(finc\_cat == "<200%" ~ "<200%",  
 finc\_cat == "Unknown" ~ "Unkwown",   
 TRUE ~ ">=200%"))  
  
# fit logistic model  
pap\_formod <- pap\_formod %>%   
 mutate\_at(vars(hpvhrd, paphad1, mdrecp1), ~factor(.x)) %>%   
 mutate(ausualpl\_cat = factor(ausualpl\_cat, levels = c("Yes", "No"))) %>%   
 mutate(eth\_cat = if\_else(eth\_cat == "Hispanic", "Hispanic", "Not Hispanic"))  
  
pap\_formod %>% ggplot(aes(x = paprec\_3bcat)) + geom\_histogram() + facet\_grid(~inc)  
  
vars = pap\_dat %>% select(ends\_with("cat"), imm\_stat, inc) %>% select(-paprec\_3bcat) %>% names()  
vars = vars[1:9]  
varssym = map(vars, ~sym(.x))  
  
.form = reformulate(response = "paprec\_3bcat", termlabels = c(vars) )  
  
map(.x = varssym, ~count(pap\_formod %>% filter(inc == 1) %>% filter(!is.na(paprec\_3bcat)), !!.x))  
  
  
des2 <- svydesign(ids = ~psu\_p, strata = ~strat\_p,   
 weights = ~wtfa\_sa, nest = TRUE, data = pap\_formod)  
  
mod <- svyglm(paprec\_3bcat ~ age\_cat + educ\_cat +   
 ausualpl\_cat + finc\_cat + cover\_cat + lcond\_chronic\_cat +   
 eth\_cat + race\_cat + imm\_stat, design = des2,   
 family = binomial,  
 subset = inc == 1)   
  
jtools::summ(mod, exp = TRUE, confint = TRUE, pvals = FALSE, digits = 2)  
  
# test significance of individual terms/term groups  
sig1 = tibble(vars) %>%   
 mutate(test = map(vars, ~regTermTest(mod, .x, method = "LRT"))) %>%   
 mutate(pval = map\_dbl(test, ~.x$p)) %>%   
 arrange(desc(pval)) %>%   
 mutate(sig = if\_else(pval > 0.05, 0, 1))  
  
sig1  
  
  
new\_coef <- stringr::str\_remove(names(coef(mod)), "^[^\_]\*\_cat") #make pretty names  
new\_coef <- stringr::str\_remove(new\_coef, "^[^\_]\*\_stat") #make pretty names  
new\_coef <- stringr::str\_remove(new\_coef, "^[^\_]\*\_chronic\_cat") #make pretty names  
coef <- names(coef(mod)) #assign pretty names  
names(coef) <- new\_coef #names original coef vector with pretty names  
coef <- coef[-1] #remove intercept  
  
jtools::plot\_summs(mod, ci\_level = 0.95,   
 coefs = coef,  
 exp = TRUE, robust = TRUE) + labs(title = "Pap Smear Significant Predictors")  
  
# MAMMOGRAM CODE  
  
# read in data  
# read in cancer module data  
cancer = read\_csv("./data/cancerxx.csv") %>%  
 select(HHX, FMX, FPX, WTFA\_SA, STRAT\_P, PSU\_P, REGION, MAMHAD, MAM6YR,   
 RMAM1\_MT, RMAM1YR, RMAM1N, RMAM1T, RMAM2, RMAM3A, RMAM3B, MAMPAY,   
 MAMREAS, MDRECMAM, MAMDNBR, MAMABN1, MFOLLOW0, MFOLLO01, MFOLLO02,   
 MFOLLO03, MFOLLO04, MFOLLO05, MNOTFOL1, MAMMODE, MAMCAN1)  
  
# read in adult data  
adult = read\_csv("./data/samadult.csv") %>%  
 select(HHX, FMX, FPX, AUSUALPL, AHCPLROU, AHCPLKND, FLA1AR)  
  
# read in family data  
family = read\_csv("./data/familyxx.csv") %>%  
 select(HHX, FMX, RAT\_CAT4, RAT\_CAT5)  
  
# read in person data  
person = read\_csv("./data/personsx.csv") %>%  
 select(HHX, FMX, FPX, AGE\_P, EDUC1, SEX, NOTCOV, COVER65, COVER65O, LA1AR,  
 LCONDRT, LACHRONR, HISCODI3, RACRECI3, COVER, YRSINUS, PLBORN)  
  
# join these data files into one for analysis  
mam\_dat = cancer %>%  
 left\_join(adult, by = c("HHX", "FMX", "FPX")) %>%  
 left\_join(person, by = c("HHX", "FMX", "FPX")) %>%  
 left\_join(family, by = c("HHX", "FMX"))   
  
# CREATE VARIABLES FOR ANALYSIS: OUTCOME AND DEMOGRAPHIC  
  
# outcome is having a mammogram in the last 2 years:RMAM3A = 1,2  
# create immigration status variable based on PLBORN and YRSINUS  
mam\_dat = mam\_dat %>%  
 mutate(mam\_2 = if\_else(RMAM3A <= 2, 1, 0),  
 imm\_stat = case\_when(YRSINUS < 4 ~ "In U.S. < 10 yrs",  
 YRSINUS == 4 | YRSINUS == 5 ~ "In U.S. >= 10 yrs",  
 PLBORN == 1 ~ "Born in U.S."))  
  
# create the age category  
mam\_dat = mam\_dat %>%   
 mutate(age\_cat = case\_when(AGE\_P >= 25 & AGE\_P < 40 ~ "25–39",  
 AGE\_P >= 40 & AGE\_P < 50 ~ "40–49",  
 AGE\_P >= 50 & AGE\_P < 65 ~ "50–64",  
 AGE\_P >= 65 ~ "65+"))  
# create educ category  
mam\_dat = mam\_dat %>%   
 mutate(educ\_cat = case\_when(EDUC1 < 13 ~ "Less than high school",  
 EDUC1 >= 13 & EDUC1 < 15 ~ "High school",  
 EDUC1 >= 15 & EDUC1 < 18 ~ "Some college",  
 EDUC1 >= 18 & EDUC1 <= 21 ~ "College graduate"))  
  
# create financial category  
mam\_dat = mam\_dat %>%   
 mutate(finc\_cat = case\_when(RAT\_CAT5 <= 7 | RAT\_CAT5 %in% c(15, 16) ~ "<200%",  
 RAT\_CAT5 %in% c(8, 9) ~ "200–299%",   
 RAT\_CAT5 %in% c(10, 11) ~ "300–399%",  
 RAT\_CAT5 >= 18 & EDUC1 <= 21 ~ "400–499%",  
 RAT\_CAT5 == 14 ~">=500%",  
 RAT\_CAT5 == 17 ~">=200%, no further detail",  
 RAT\_CAT5 %in% c(96, 99) ~ "Unknown"))  
  
# create usual care category  
mam\_dat = mam\_dat %>%   
 mutate(ausualpl\_cat = case\_when(AUSUALPL == 2 ~ "No",  
 AUSUALPL %in% c(1, 3) ~ "Yes",  
 AUSUALPL %in% c(7, 8, 9) ~ "Other"))  
# create coverage status  
mam\_dat = mam\_dat %>%   
 mutate(cover\_cat = case\_when(NOTCOV == 1 | COVER == 4 | COVER65 == 6 ~ "None",  
 COVER == 2 | COVER65 %in% 2:4 ~ "Public",  
 COVER %in% c(1, 3) | COVER65 %in% c(1, 5) ~ "Private/Military"))  
  
# create disability variable  
mam\_dat = mam\_dat %>%   
 mutate(lcond\_chronic\_cat = if\_else(LCONDRT == 1, "Yes", "No"))  
  
# create race & ethnic categories  
mam\_dat = mam\_dat %>%   
 mutate(race\_cat = case\_when(RACRECI3 == 1 ~ "White",  
 RACRECI3 == 2 ~ "Black",  
 RACRECI3 == 3 ~ "Asian",  
 RACRECI3 == 4 ~ "AN/AI"),  
 eth\_cat = case\_when(HISCODI3 == 1 ~ "Hispanic",  
 HISCODI3 == 2 ~ "Non-Hispanic White",  
 HISCODI3 == 3 ~ "Non-Hispanic Black",  
 HISCODI3 == 4 ~ "Non-Hispanic Asian",  
 HISCODI3 == 5 ~ "Non-Hispanic AN/AI"))  
  
# fit survey design  
# create domain variable for inclusion criteria  
mam\_dat = mam\_dat %>%  
 mutate(domain = if\_else(SEX == 2 & AGE\_P >= 40, 1, 0))  
  
# create survey design object  
des = svydesign(ids = ~PSU\_P, strata = ~STRAT\_P, weights = ~WTFA\_SA, nest = TRUE, data = mam\_dat)  
  
# create tables   
# percent of women who have had mammogram in the last two years by age  
  
age\_pct = svyby(~mam\_2, by = ~domain+age\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("ci", "se"))  
age\_pct %>% filter(domain == 1) %>% select(-domain, -se) %>% knitr::kable()  
  
  
# percent of women who have had mammogram by education  
edu\_pct = svyby(~mam\_2, by = ~domain+educ\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
edu\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(educ\_cat) %>%  
 summarise(count = n())  
  
edu\_pct = left\_join(edu\_pct, edu\_counts, by = "educ\_cat")  
  
edu\_pct %>%   
 knitr::kable()  
  
# percent of women who have had mammogram by financial category  
finc\_pct = svyby(~mam\_2, by = ~domain+finc\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
finc\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(finc\_cat) %>%  
 summarise(count = n())  
  
finc\_pct = left\_join(finc\_pct, finc\_counts, by = "finc\_cat")  
  
finc\_pct %>% knitr::kable()  
  
  
# percent of women who have had mammogram by usual care category  
ausualp\_pct = svyby(~mam\_2, by = ~domain+ausualpl\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
usual\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(ausualpl\_cat) %>%  
 summarise(count = n())  
  
ausualp\_pct = left\_join(ausualp\_pct, usual\_counts, by = "ausualpl\_cat")  
  
ausualp\_pct %>% knitr::kable()  
  
  
# percent of women who have had mammogram by health coverage  
cover\_pct = svyby(~mam\_2, by = ~domain+cover\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
cover\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(cover\_cat) %>%  
 summarise(count = n())  
  
cover\_pct = left\_join(cover\_pct, cover\_counts, by = "cover\_cat")  
  
cover\_pct %>% knitr::kable()  
  
  
# percent of women who have had mammogram by chronic conditions  
lcond\_chronic\_pct = svyby(~mam\_2, by = ~domain+lcond\_chronic\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
chronic\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(lcond\_chronic\_cat) %>%  
 summarise(count = n())  
  
lcond\_chronic\_pct = left\_join(lcond\_chronic\_pct, chronic\_counts, by = "lcond\_chronic\_cat")  
  
lcond\_chronic\_pct %>% knitr::kable()  
  
  
# percent of women who have had mammogram by race  
race\_pct = svyby(~mam\_2, by = ~domain+race\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
race\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(race\_cat) %>%  
 summarise(count = n())  
  
race\_pct = left\_join(race\_pct, race\_counts, by = "race\_cat")  
  
race\_pct %>% knitr::kable()  
  
  
# percent of women who have had mammogram by ethnicity  
eth\_pct = svyby(~mam\_2, by = ~domain+eth\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
eth\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(eth\_cat) %>%  
 summarise(count = n())  
  
eth\_pct = left\_join(eth\_pct, eth\_counts, by = "eth\_cat")  
  
eth\_pct %>% knitr::kable()  
  
# percent of women who have had mammogram by immigration status  
imm\_pct = svyby(~mam\_2, by = ~domain+imm\_stat, svymean, na.rm = TRUE,  
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
imm\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(imm\_stat) %>%  
 summarise(count = n())  
  
imm\_pct = left\_join(imm\_pct, imm\_counts, by = "imm\_stat")  
  
imm\_pct %>% knitr::kable()  
  
# create tables by age group  
  
# percent of women who have had mammogram by education and age   
edu\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + educ\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
edu\_tab = edu\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
edu\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, educ\_cat) %>%  
 summarise(count = n())  
  
edu\_tab = left\_join(edu\_tab, edu\_counts2, by = c("age\_cat", "educ\_cat"))  
  
edu\_tab %>% knitr::kable()  
  
  
# percent of women who have had mammogram by financial category and age  
finc\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + finc\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
  
finc\_tab = finc\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
finc\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, finc\_cat) %>%  
 summarise(count = n())  
  
finc\_tab = left\_join(finc\_tab, finc\_counts2, by = c("age\_cat", "finc\_cat"))  
  
finc\_tab %>% knitr::kable()  
  
  
# percent of women who have had mammogram by usual care and age  
ausualp\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + ausualpl\_cat,  
 svymean, na.rm = TRUE, design = des,   
 vartype = c("se", "ci"))  
  
usual\_tab = ausualp\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
usual\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, ausualpl\_cat) %>%  
 summarise(count = n())  
  
usual\_tab = left\_join(usual\_tab, usual\_counts2, by = c("age\_cat", "ausualpl\_cat"))  
  
usual\_tab %>% knitr::kable()  
  
  
# percent of women who have had mammogram by health coverage and age  
cover\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + cover\_cat, svymean,  
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
  
ins\_tab = cover\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)  
  
ins\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, cover\_cat) %>%  
 summarise(count = n())  
  
ins\_tab = left\_join(ins\_tab, ins\_counts2, by = c("age\_cat", "cover\_cat"))  
  
ins\_tab %>% knitr::kable()  
  
  
  
# percent of women who have had mammogram by chronic conditions and age  
lcond\_chronic\_pct\_strat = svyby(~mam\_2,   
 by = ~domain + age\_cat + lcond\_chronic\_cat,  
 svymean, na.rm = TRUE, design = des,  
 vartype = c("se", "ci"))  
  
dis\_tab = lcond\_chronic\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)  
  
dis\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, lcond\_chronic\_cat) %>%  
 summarise(count = n())  
  
dis\_tab = left\_join(dis\_tab, dis\_counts2, by = c("age\_cat", "lcond\_chronic\_cat"))  
  
dis\_tab %>% knitr::kable()  
  
  
# percent of women who have had mammogram by race and age  
race\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + race\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
  
race\_tab = race\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
race\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, race\_cat) %>%  
 summarise(count = n())  
  
race\_tab = left\_join(race\_tab, race\_counts2, by = c("age\_cat", "race\_cat"))  
  
race\_tab %>% knitr::kable()  
  
  
# percent of women who have had mammogram by ethnicity and age  
eth\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + eth\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
  
eth\_tab = eth\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
eth\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, eth\_cat) %>%  
 summarise(count = n())  
  
eth\_tab = left\_join(eth\_tab, eth\_counts2, by = c("age\_cat", "eth\_cat"))  
  
eth\_tab %>% knitr::kable()  
  
# percent of women who have had mammogram by immigration and age  
imm\_pct\_strat = svyby(~mam\_2, by = ~domain + age\_cat + imm\_stat, svymean,  
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
  
imm\_tab = imm\_pct\_strat %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
imm\_counts2 = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat, imm\_stat) %>%  
 summarise(count = n())  
  
imm\_tab = left\_join(imm\_tab, imm\_counts2, by = c("age\_cat", "imm\_stat"))  
  
imm\_tab %>% knitr::kable()  
  
# overall percent of women who have had mammogram  
total = svyby(~mam\_2, by = ~domain + age\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci"))  
  
tot\_tab = total %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
tot\_counts = filter(mam\_dat, domain == 1) %>%  
 group\_by(age\_cat) %>%  
 summarise(count = n())  
  
tot\_tab = left\_join(tot\_tab, tot\_counts, by = "age\_cat")  
  
tot\_tab %>% knitr::kable()  
  
all\_counts = filter(mam\_dat, domain == 1)  
  
tot\_pct = svyby(~mam\_2, by = ~domain, svymean, na.rm = TRUE, design = des,  
 vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se) %>%  
 mutate(age\_cat = "40+",  
 count = 12483)  
  
# COMBINE TABLES BY AGE WITH TABLES OVERALL  
  
# overall  
  
tot\_tab = rbind(tot\_pct, tot\_tab)  
  
tot\_tab2 = tot\_tab %>%  
 mutate(type = "Total",  
 level = "-")  
  
# education  
edu\_pct2 = edu\_pct %>%  
 mutate(age\_cat = "40+")  
  
edu\_tab = rbind(edu\_pct2, edu\_tab)  
  
# factor levels for presentation  
edu\_tab2 = edu\_tab %>%  
 mutate(type = "Education") %>%  
 rename(level = educ\_cat) %>%  
 mutate(level = factor(level, levels = c("Less than high school", "High school", "Some college", "College graduate"))) %>%  
 arrange(level)  
  
# financial category  
finc\_pct2 = finc\_pct %>%  
 mutate(age\_cat = "40+")  
  
finc\_tab = rbind(finc\_pct2, finc\_tab)  
  
# factor levels for presentation  
finc\_tab2 = finc\_tab %>%  
 mutate(type = "Family Income Poverty Ratio") %>%  
 rename(level = finc\_cat) %>%  
 mutate(level = factor(level, levels = c("<200%", ">=200%, no further detail", "200–299%", "300–399%", "400–499%", ">=500%", "Unknown"))) %>%  
 arrange(level)  
  
# usual care category  
ausualp\_pct2 = ausualp\_pct %>%  
 mutate(age\_cat = "40+")  
  
usual\_tab = rbind(ausualp\_pct2, usual\_tab)  
  
  
# factor levels for presentation  
usual\_tab2 = usual\_tab %>%  
 mutate(type = "Usual Source of Care") %>%  
 rename(level = ausualpl\_cat) %>%  
 mutate(level = factor(level, levels = c("No", "Yes", "Other"))) %>%  
 arrange(level)  
  
# insurance coverage category  
cover\_pct2 = cover\_pct %>%  
 mutate(age\_cat = "40+")  
  
ins\_tab = rbind(cover\_pct2, ins\_tab)  
  
# factor levels for presentation  
ins\_tab2 = ins\_tab %>%  
 mutate(type = "Insurance Type") %>%  
 rename(level = cover\_cat) %>%  
 mutate(level = factor(level, levels = c("None", "Public", "Private/Military"))) %>%  
 arrange(level)  
  
# chronic condition category  
lcond\_chronic\_pct2 = lcond\_chronic\_pct %>%  
 mutate(age\_cat = "40+")  
  
dis\_tab = rbind(lcond\_chronic\_pct2, dis\_tab)  
  
# factor levels for presentation  
dis\_tab2 = dis\_tab %>%  
 mutate(type = "Chronic Disability") %>%  
 rename(level = lcond\_chronic\_cat) %>%  
 mutate(level = factor(level, levels = c("Yes", "No"))) %>%  
 arrange(level)  
  
# ethnic category  
eth\_pct2 = eth\_pct %>%  
 mutate(age\_cat = "40+")  
  
eth\_tab = rbind(eth\_pct2, eth\_tab)  
  
# factor levels for presentation  
eth\_tab2 = eth\_tab %>%  
 mutate(type = "Ethnicity") %>%  
 rename(level = eth\_cat) %>%  
 mutate(level = factor(level, levels = c("Hispanic", "Non-Hispanic White", "Non-Hispanic Black", "Non-Hispanic AN/AI", "Non-Hispanic Asian"))) %>%  
 arrange(level)  
  
# race category  
race\_pct2 = race\_pct %>%  
 mutate(age\_cat = "40+")  
  
race\_tab = rbind(race\_pct2, race\_tab)  
  
# factor levels for presentation  
race\_tab2 = race\_tab %>%  
 mutate(type = "Race") %>%  
 rename(level = race\_cat) %>%  
 mutate(level = factor(level, levels = c("White", "Black", "AN/AI", "Asian"))) %>%  
 arrange(level)  
  
# immigration category  
imm\_pct2 = imm\_pct %>%  
 mutate(age\_cat = "40+")  
  
imm\_tab = rbind(imm\_pct2, imm\_tab)  
  
# factor levels for presentation  
imm\_tab2 = imm\_tab %>%  
 mutate(type = "Immigration") %>%  
 rename(level = imm\_stat) %>%  
 mutate(level = factor(level, levels = c("In U.S. < 10 yrs", "In U.S. >= 10 yrs", "Born in U.S."))) %>%  
 arrange(level)  
  
# create table of percentages of women who have gotten mammograms within the last two years (still need to add CIs)  
tab\_one = rbind(tot\_tab2, edu\_tab2, finc\_tab2, usual\_tab2, ins\_tab2, dis\_tab2, eth\_tab2, race\_tab2, imm\_tab2) %>%  
 mutate(mam\_2 = round(mam\_2\*100, 1),  
 ci\_l = round(ci\_l\*100, 1),  
 ci\_u = round(ci\_u\*100, 1),  
 CI = str\_c(ci\_l, ", ", ci\_u)) %>%  
 rename(Percent = mam\_2,  
 N = count) %>%  
 select(-ci\_l, -ci\_u) %>%  
 pivot\_wider(names\_from = age\_cat, values\_from = c(N, Percent, CI)) %>%  
 janitor::clean\_names() %>%  
 select(type, level, n\_40, percent\_40, ci\_40, n\_40\_49, percent\_40\_49, ci\_40\_49, n\_50\_64, percent\_50\_64, ci\_50\_64, everything())  
  
# barplots for presentation in class  
  
# sample size  
usual\_size = mam\_dat %>%  
 filter(domain == 1) %>%  
 group\_by(ausualpl\_cat) %>%  
 summarise(count = n()) %>%  
 filter(ausualpl\_cat != "NA")  
sum(pull(usual\_size, count))  
  
# usual source of care barchart  
usual\_tab %>%  
 filter(age\_cat != "40+" & ausualpl\_cat != "Other") %>%  
 ggplot(aes(x = ausualpl\_cat, y = mam\_2, fill = ausualpl\_cat)) +  
 geom\_col() +  
 scale\_y\_continuous(limits = c(0,1), breaks = c(0, 0.25, 0.5, 0.75, 1)) +  
 geom\_errorbar(aes(ymin = ci\_l, ymax = ci\_u)) +  
 facet\_grid(~age\_cat) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") +  
 labs(y = "Percent Had Mammogram, Last 2 Years", x = "Usual Source of Care (Have/Have Not)", title = "N = 12483")  
  
# sample size  
ins\_size = mam\_dat %>%  
 filter(domain == 1) %>%  
 group\_by(cover\_cat) %>%  
 summarise(count = n()) %>%  
 filter(cover\_cat != "NA")  
sum(pull(ins\_size, count))  
  
# insurance type barchart  
ins\_tab %>%  
 filter(age\_cat != "40+") %>%  
 ggplot(aes(x = cover\_cat, y = mam\_2, fill = cover\_cat)) +  
 geom\_col() +   
 scale\_y\_continuous(limits = c(0,1), breaks = c(0, 0.25, 0.5, 0.75, 1)) +  
 geom\_errorbar(aes(ymin = ci\_l, ymax = ci\_u)) +  
 facet\_grid(~age\_cat) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") +  
 labs(y = "Percent Had Mammogram, Last 2 Years", x = "Type of Insurance Coverage", title = "N = 12436")  
  
# fit models  
# make financial status and ethnicity only two levels   
mam\_dat2 = mam\_dat %>%  
 mutate(finc\_cat2 = if\_else(finc\_cat == "<200%", finc\_cat,  
 if\_else(finc\_cat == "Unknown", finc\_cat, ">=200%")),  
 eth\_cat2 = if\_else(eth\_cat == "Hispanic", eth\_cat, "Non-Hispanic"),  
 imm\_stat2 = if\_else(imm\_stat == "Born in U.S.", imm\_stat, "Immigrated"),  
 ausualpl\_cat2 = replace(ausualpl\_cat, ausualpl\_cat == "Other", NA),  
 lcond\_chronic\_cat2 = if\_else(lcond\_chronic\_cat == "Yes", "Yes (Chronic)", lcond\_chronic\_cat))  
  
# refit the design object  
des2 = svydesign(ids = ~PSU\_P, strata = ~STRAT\_P, weights = ~WTFA\_SA, nest = TRUE, data = mam\_dat2)  
  
  
# fit the full model on all categorical variables for the included women  
mam2\_fit = svyglm(mam\_2 ~ as.factor(age\_cat) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2) + as.factor(race\_cat) + as.factor(eth\_cat2) + as.factor(imm\_stat2),   
 design = des2, subset = domain == 1, family = binomial(link = "logit"))  
  
summary(mam2\_fit)  
summ(mam2\_fit)  
  
# Rao-scott LRT ANOVA  
  
# test significance of full model  
regTermTest(mam2\_fit, ~ as.factor(age\_cat) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2) + as.factor(race\_cat) + as.factor(eth\_cat2) + as.factor(imm\_stat2),  
 method = "LRT")  
  
# test significance of individual terms/term groups  
regTermTest(mam2\_fit, "as.factor(age\_cat)",   
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(educ\_cat)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(finc\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(ausualpl\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(cover\_cat)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(lcond\_chronic\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(race\_cat)",  
 method = "LRT") # not sig  
  
regTermTest(mam2\_fit, "as.factor(eth\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit, "as.factor(imm\_stat2)",  
 method = "LRT") # not sig  
  
# fit reduced model with significant predictors  
mam2\_fit2 = svyglm(mam\_2 ~ as.factor(age\_cat) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2) + as.factor(eth\_cat2),  
 design = des2, subset = domain == 1,   
 family = binomial(link = "logit"))  
  
summary(mam2\_fit2)  
summ(mam2\_fit2) # 0.05-0.20   
  
# Rao-Scott of full model  
regTermTest(mam2\_fit2, ~ as.factor(age\_cat) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2) + as.factor(eth\_cat2))  
  
# single term/term group significance  
regTermTest(mam2\_fit2, "as.factor(age\_cat)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit2, "as.factor(educ\_cat)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit2, "as.factor(finc\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit2, "as.factor(ausualpl\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit2, "as.factor(cover\_cat)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit2, "as.factor(lcond\_chronic\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(mam2\_fit2, "as.factor(eth\_cat2)",  
 method = "LRT") # sig  
  
  
# give labels to the coefficients to look nicer on the OR graph  
  
coef <- names(coef(mam2\_fit2))  
coef\_new = stringr::str\_remove(coef, "^[^\_]\*\_cat[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_cat2[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_chronic")  
coef\_new = stringr::str\_remove(coef\_new, "^\*\_cat2[)]")  
names(coef) <- coef\_new  
coef1 = coef[-1] # remove intercept   
  
model\_coef = broom::tidy(mam2\_fit2, conf.int = TRUE, conf.level = 0.95, exponentiate = TRUE) %>%  
 mutate(term = case\_when(term == "(Intercept)" ~ "Intercept",  
 term == "as.factor(age\_cat)50–64" ~ "50-64 vs <50",  
 term == "as.factor(age\_cat)65+" ~ "65+ vs <50",  
 term == "as.factor(educ\_cat)High school" ~ "High School Degree vs College Degree",  
 term == "as.factor(educ\_cat)Less than high school" ~ "Less than High School vs College Degree",  
 term == "as.factor(educ\_cat)Some college" ~ "Some college vs College Degree",  
 term == "as.factor(finc\_cat2)>=200%" ~ ">=200% vs <200% Poverty Level",  
 term == "as.factor(ausualpl\_cat2)Yes" ~ "Usual Source of Care vs No Usual Source of Care",  
 term == "as.factor(cover\_cat)Private/Military" ~ "Private/Military Insurance vs No Insurance",  
 term == "as.factor(cover\_cat)Public" ~ "Public Insurance vs No Insurance",  
 term == "as.factor(lcond\_chronic\_cat2)Yes (Chronic)" ~ "Chronic Condition vs No Chronic Condition",  
 term == "as.factor(eth\_cat2)Non-Hispanic" ~ "Non-Hispanic vs Hispanic"),  
 estimate = round(estimate, 2),  
 conf.low = round(conf.low, 2),  
 conf.high = round(conf.high, 2)) %>%  
 select(term, estimate, conf.low, conf.high)  
  
model\_coef %>% knitr::kable()  
  
# plot ORs with CIs  
jtools::plot\_summs(mam2\_fit2, coefs = coef1, exp = TRUE) +  
 labs(title = "Mammogram Significant Predictors")   
  
# print percentages  
tab\_one %>% knitr::kable()  
  
  
# COLORECTAL CODE  
  
# data import  
cancer = read\_csv("./data/cancerxx.csv") %>%   
 janitor::clean\_names() %>%   
 select(hhx, fmx, fpx, wtfa\_sa, strat\_p, psu\_p, region, hfobhad1,   
 rhfo2\_mt, rhfo2yr, rhfo2n, rhfo2t, rhfo2, rhfob3a,  
 rhfob3b, hfobrea2, fobhad1, rfob2\_mt, rfob2yr,  
 rfob2n, rfob2t, rofob3a, rofob3b, rfobres1)  
  
adult = read\_csv("./data/samadult.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, fpx, ausualpl, ahcplrou, ahcplknd, fla1ar)  
  
family = read\_csv("./data/familyxx.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, rat\_cat4, rat\_cat5)  
  
person = read\_csv("./data/personsx.csv") %>%  
 janitor::clean\_names() %>%   
 select(hhx, fmx, fpx, age\_p, educ1, sex, notcov, cover65, cover65o,   
 la1ar, lcondrt, lachronr, hiscodi3, racreci3, cover, yrsinus, plborn)  
  
col\_dat = cancer %>%  
 left\_join(adult, by = c("hhx", "fmx", "fpx")) %>%  
 left\_join(person, by = c("hhx", "fmx", "fpx")) %>%  
 left\_join(family, by = c("hhx", "fmx"))  
  
# data manipulation  
#home stool test within last year  
col\_dat = col\_dat %>%  
 mutate(col\_2 = if\_else(rhfob3a <= 1, 1, 0),  
 imm\_stat = case\_when(yrsinus < 4 ~ "In U.S. < 10 yrs",  
 yrsinus == 4 | yrsinus == 5 ~ "In U.S. >= 10 yrs",  
 plborn == 1 ~ "Born in U.S."))  
  
# create the age category  
col\_dat = col\_dat %>%   
 mutate(age\_cat = case\_when(age\_p >= 25 & age\_p < 40 ~ "25–39",  
 age\_p >= 40 & age\_p < 50 ~ "40–49",  
 age\_p >= 50 & age\_p < 65 ~ "50–64",  
 age\_p >= 65 ~ "65+"))  
# create educ category  
col\_dat = col\_dat %>%   
 mutate(educ\_cat = case\_when(educ1 < 13 ~ "Less than high school",  
 educ1 >= 13 & educ1 < 15 ~ "High school",  
 educ1 >= 15 & educ1 < 18 ~ "Some college",  
 educ1 >= 18 & educ1 <= 21 ~ "College graduate"))  
  
# create financial category  
col\_dat = col\_dat %>%   
 mutate(finc\_cat = case\_when(rat\_cat5 <= 7 | rat\_cat5 %in% c(15, 16) ~ "<200%",  
 rat\_cat5 %in% c(8, 9) ~ "200–299%",   
 rat\_cat5 %in% c(10, 11) ~ "300–399%",  
 rat\_cat5 >= 18 & educ1 <= 21 ~ "400–499%",  
 rat\_cat5 == 14 ~">=500%",  
 rat\_cat5 == 17 ~">=200%, no further detail",  
 rat\_cat5 %in% c(96, 99) ~ "Unknown"))  
  
# create as usual category  
col\_dat = col\_dat %>%   
 mutate(ausualpl\_cat = case\_when(ausualpl == 2 ~ "No",  
 ausualpl %in% c(1, 3) ~ "Yes",  
 ausualpl %in% c(7, 8, 9) ~ "Other"))  
# coverage status  
col\_dat = col\_dat %>%   
 mutate(cover\_cat = case\_when(notcov == 1 | cover == 4 | cover65 == 6 ~ "None",  
 cover == 2 | cover65 %in% 2:4 ~ "Public",  
 cover %in% c(1, 3) | cover65 %in% c(1, 5) ~  
 "Private/Military"))  
  
# disability  
col\_dat = col\_dat %>%   
 mutate(lcond\_chronic\_cat = if\_else(lcondrt == 1, "Yes", "No"))  
  
# race  
col\_dat = col\_dat %>%   
 mutate(race\_cat = case\_when(racreci3 == 1 ~ "White",  
 racreci3 == 2 ~ "Black",  
 racreci3 == 3 ~ "Asian",  
 racreci3 == 4 ~ "AN/AI"),  
 eth\_cat = case\_when(hiscodi3 == 1 ~ "Hispanic",  
 hiscodi3 == 2 ~ "Non-Hispanic White",  
 hiscodi3 == 3 ~ "Non-Hispanic Black",  
 hiscodi3 == 4 ~ "Non-Hispanic Asian",  
 hiscodi3 == 5 ~ "Non-Hispanic AN/AI"))  
  
# fit survey design  
  
col\_dat = col\_dat %>%  
 mutate(domain = if\_else(age\_p >= 50, 1, 0))  
  
des = svydesign(ids = ~ psu\_p, strata = ~ strat\_p,   
 weights = ~ wtfa\_sa, nest = TRUE, data = col\_dat)  
  
# tables  
  
#age percentage  
age\_pct = svyby(~col\_2, by = ~domain + age\_cat + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("ci", "se"))  
age\_pct %>%   
 filter(domain == 1) %>% select(-domain) %>% knitr::kable()  
  
  
#education  
edu\_pct = svyby(~col\_2, by = ~domain + educ\_cat + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
edu\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(educ\_cat) %>%  
 summarise(count = n())  
  
edu\_pct = left\_join(edu\_pct, edu\_counts, by = "educ\_cat")  
  
edu\_pct %>%   
 knitr::kable()  
  
  
#finc  
finc\_pct = svyby(~col\_2, by = ~domain + finc\_cat + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
finc\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(finc\_cat) %>%  
 summarise(count = n())  
  
finc\_pct = left\_join(finc\_pct, finc\_counts, by = "finc\_cat")  
  
finc\_pct %>% knitr::kable()  
  
  
#usual care  
ausualp\_pct = svyby(~col\_2, by = ~domain + ausualpl\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
usual\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(ausualpl\_cat) %>%  
 summarise(count = n())  
ausualp\_pct = left\_join(ausualp\_pct, usual\_counts, by = "ausualpl\_cat")  
  
ausualp\_pct %>% knitr::kable()  
  
  
#health coverage  
cover\_pct = svyby(~col\_2, by = ~domain + cover\_cat + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
cover\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(cover\_cat) %>%  
 summarise(count = n())  
cover\_pct = left\_join(cover\_pct, cover\_counts, by = "cover\_cat")  
  
cover\_pct %>% knitr::kable()  
  
  
#chronic conditions  
lcond\_chronic\_pct = svyby(~col\_2, by = ~domain + lcond\_chronic\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
chronic\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(lcond\_chronic\_cat) %>%  
 summarise(count = n())  
  
lcond\_chronic\_pct = left\_join(lcond\_chronic\_pct, chronic\_counts, by = "lcond\_chronic\_cat")  
  
lcond\_chronic\_pct %>% knitr::kable()  
  
  
#race  
race\_pct = svyby(~col\_2, by = ~domain + race\_cat + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
  
race\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(race\_cat) %>%  
 summarise(count = n())  
race\_pct = left\_join(race\_pct, race\_counts, by = "race\_cat")  
  
race\_pct %>% knitr::kable()  
  
  
#ethnicity  
eth\_pct = svyby(~col\_2, by = ~domain + eth\_cat + sex, svymean, na.rm = TRUE,  
 design = des, vartype = c("se", "ci")) %>%   
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
  
eth\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(eth\_cat) %>%  
 summarise(count = n())  
eth\_pct = left\_join(eth\_pct, eth\_counts, by = "eth\_cat")  
  
eth\_pct %>% knitr::kable()  
  
  
#immigration  
imm\_pct = svyby(~col\_2, by = ~domain + imm\_stat + sex, svymean, na.rm = TRUE,  
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
imm\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(imm\_stat) %>%  
 summarise(count = n())  
imm\_pct = left\_join(imm\_pct, imm\_counts, by = "imm\_stat")  
  
imm\_pct %>% knitr::kable()  
  
# tables by age group  
  
#education  
edu\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + educ\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
edu\_tab = edu\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)  
  
edu\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, educ\_cat) %>%  
 summarise(count = n())  
edu\_tab = left\_join(edu\_tab, edu\_counts2, by = c("age\_cat", "educ\_cat"))  
  
edu\_tab %>% knitr::kable()  
  
  
#finc  
finc\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + finc\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
  
finc\_tab = finc\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
finc\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, finc\_cat) %>%  
 summarise(count = n())  
finc\_tab = left\_join(finc\_tab, finc\_counts2, by = c("age\_cat", "finc\_cat"))  
  
finc\_tab %>% knitr::kable()  
  
  
#usual care  
ausualp\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + ausualpl\_cat + sex,   
 svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci"))  
usual\_tab = ausualp\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
usual\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, ausualpl\_cat) %>%  
 summarise(count = n())  
usual\_tab = left\_join(usual\_tab, usual\_counts2, by = c("age\_cat", "ausualpl\_cat"))  
  
usual\_tab %>% knitr::kable()  
  
  
#health coverage  
cover\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + cover\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
ins\_tab = cover\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)  
  
ins\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, cover\_cat) %>%  
 summarise(count = n())  
ins\_tab = left\_join(ins\_tab, ins\_counts2, by = c("age\_cat", "cover\_cat"))  
  
ins\_tab %>% knitr::kable()  
  
  
#chronic conditions  
lcond\_chronic\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat +   
 lcond\_chronic\_cat + sex,  
 svymean, na.rm = TRUE, design = des,  
 vartype = c("se", "ci"))  
dis\_tab = lcond\_chronic\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
dis\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, lcond\_chronic\_cat) %>%  
 summarise(count = n())  
dis\_tab = left\_join(dis\_tab, dis\_counts2, by = c("age\_cat", "lcond\_chronic\_cat"))  
  
dis\_tab %>% knitr::kable()  
  
  
#race  
race\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + race\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
race\_tab = race\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
race\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, race\_cat) %>%  
 summarise(count = n())  
race\_tab = left\_join(race\_tab, race\_counts2, by = c("age\_cat", "race\_cat"))  
  
race\_tab %>% knitr::kable()  
  
  
#ethnicity  
eth\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + eth\_cat + sex, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
eth\_tab = eth\_pct\_strat %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
eth\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, eth\_cat) %>%  
 summarise(count = n())  
eth\_tab = left\_join(eth\_tab, eth\_counts2, by = c("age\_cat", "eth\_cat"))  
  
eth\_tab %>% knitr::kable()  
  
  
#immigration  
imm\_pct\_strat = svyby(~col\_2, by = ~domain + age\_cat + imm\_stat + sex, svymean,  
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
imm\_tab = imm\_pct\_strat %>%  
 filter(domain == 1) %>%  
 select(-domain, -se)  
  
imm\_counts2 = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat, imm\_stat) %>%  
 summarise(count = n())  
imm\_tab = left\_join(imm\_tab, imm\_counts2, by = c("age\_cat", "imm\_stat"))  
  
imm\_tab %>% knitr::kable()  
  
  
#total gotten colorectal  
total = svyby(~col\_2, by = ~domain + age\_cat + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci"))  
  
tot\_tab = total %>%   
 filter(domain == 1) %>%   
 select(-domain, -se)   
  
tot\_counts = filter(col\_dat, domain == 1) %>%  
 group\_by(age\_cat) %>%  
 summarise(count = n())  
tot\_tab = left\_join(tot\_tab, tot\_counts, by = "age\_cat")  
  
tot\_tab %>% knitr::kable()  
  
all\_counts = filter(col\_dat, domain == 1)  
  
tot\_pct = svyby(~col\_2, by = ~domain + sex, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 select(-domain, -se) %>%  
 mutate(age\_cat = "50+",  
 count = 17056)  
  
# combine tables  
tot\_tab = rbind(tot\_pct, tot\_tab)  
  
tot\_tab2 = tot\_tab %>%  
 mutate(type = "Total",  
 level = "-")  
  
edu\_pct2 = edu\_pct %>%  
 mutate(age\_cat = "50+")  
  
edu\_tab = rbind(edu\_pct2, edu\_tab)  
  
edu\_tab2 = edu\_tab %>%  
 mutate(type = "Education") %>%  
 rename(level = educ\_cat) %>%  
 mutate(level = factor(level,   
 levels = c("Less than high school", "High school",   
 "Some college", "College graduate"))) %>%  
 arrange(level)  
  
finc\_pct2 = finc\_pct %>%  
 mutate(age\_cat = "50+")  
  
finc\_tab = rbind(finc\_pct2, finc\_tab)  
  
finc\_tab2 = finc\_tab %>%  
 mutate(type = "Family Income Poverty Ratio") %>%  
 rename(level = finc\_cat) %>%  
 mutate(level = factor(level, levels = c("<200%", ">=200%, no further detail", "200–299%", "300–399%", "400–499%", ">=500%", "Unknown"))) %>%  
 arrange(level)  
  
ausualp\_pct2 = ausualp\_pct %>%  
 mutate(age\_cat = "50+")  
  
usual\_tab = rbind(ausualp\_pct2, usual\_tab)  
  
usual\_tab2 = usual\_tab %>%  
 mutate(type = "Usual Source of Care") %>%  
 rename(level = ausualpl\_cat) %>%  
 mutate(level = factor(level, levels = c("No", "Yes", "Other"))) %>%  
 arrange(level)  
  
cover\_pct2 = cover\_pct %>%  
 mutate(age\_cat = "50+")  
  
ins\_tab = rbind(cover\_pct2, ins\_tab)  
  
ins\_tab2 = ins\_tab %>%  
 mutate(type = "Insurance Type") %>%  
 rename(level = cover\_cat) %>%  
 mutate(level = factor(level, levels = c("None", "Public", "Private/Military"))) %>%  
 arrange(level)  
  
lcond\_chronic\_pct2 = lcond\_chronic\_pct %>%  
 mutate(age\_cat = "50+")  
  
dis\_tab = rbind(lcond\_chronic\_pct2, dis\_tab)  
  
dis\_tab2 = dis\_tab %>%  
 mutate(type = "Chronic Disability") %>%  
 rename(level = lcond\_chronic\_cat) %>%  
 mutate(level = factor(level, levels = c("Yes", "No"))) %>%  
 arrange(level)  
  
eth\_pct2 = eth\_pct %>%  
 mutate(age\_cat = "50+")  
  
eth\_tab = rbind(eth\_pct2, eth\_tab)  
  
eth\_tab2 = eth\_tab %>%  
 mutate(type = "Ethnicity") %>%  
 rename(level = eth\_cat) %>%  
 mutate(level = factor(level, levels = c("Hispanic", "Non-Hispanic White", "Non-Hispanic Black", "Non-Hispanic AN/AI", "Non-Hispanic Asian"))) %>%  
 arrange(level)  
  
race\_pct2 = race\_pct %>%  
 mutate(age\_cat = "50+")  
  
race\_tab = rbind(race\_pct2, race\_tab)  
  
race\_tab2 = race\_tab %>%  
 mutate(type = "Race") %>%  
 rename(level = race\_cat) %>%  
 mutate(level = factor(level, levels = c("White", "Black", "AN/AI", "Asian"))) %>%  
 arrange(level)  
  
imm\_pct2 = imm\_pct %>%  
 mutate(age\_cat = "50+")  
  
imm\_tab = rbind(imm\_pct2, imm\_tab)  
  
imm\_tab2 = imm\_tab %>%  
 mutate(type = "Immigration") %>%  
 rename(level = imm\_stat) %>%  
 mutate(level = factor(level, levels = c("In U.S. < 10 yrs", "In U.S. >= 10 yrs", "Born in U.S."))) %>%  
 arrange(level)  
  
# create table of percentages of people who have gotten fobts within the last year  
tab\_one = rbind(tot\_tab2, edu\_tab2, finc\_tab2, usual\_tab2, ins\_tab2, dis\_tab2, eth\_tab2, race\_tab2, imm\_tab2) %>%  
 mutate(col\_2 = round(col\_2\*100, 1),  
 ci\_l = round(ci\_l\*100, 1),  
 ci\_u = round(ci\_u\*100, 1),  
 CI = str\_c(ci\_l, ", ", ci\_u)) %>%  
 rename(Percent = col\_2,  
 N = count) %>%  
 select(-ci\_l, -ci\_u) %>%  
 pivot\_wider(names\_from = c(sex, age\_cat), values\_from = c(N, Percent, CI)) %>%  
 janitor::clean\_names() %>%  
 select(Type = type, Level = level,   
 "Number Men Age 50-64" = n\_1\_50\_64,  
 "Percent Men Age 50-64" = percent\_1\_50\_64,   
 "CI Men Age 50-64" = ci\_1\_50\_64,   
 "Number Men Age 65+" = n\_1\_65,  
 "Percent Men Age 65+" = percent\_1\_65,   
 "CI Men Age 65+" = ci\_1\_65,  
 "Number Women Age 50-64" = n\_2\_50\_64,  
 "Percent Women Age 50-64" = percent\_2\_50\_64,   
 "CI Women Age 50-64" = ci\_2\_50\_64,   
 "Number Women Age 65+" = n\_2\_65,  
 "Percent Women Age 65+" = percent\_2\_65,   
 "CI Women Age 65+" = ci\_2\_65)  
  
# print percentages  
tab\_one %>% knitr::kable()  
  
# barplots for in class presentation  
# usual source of care barchart  
usual\_tab %>%  
 mutate(sex = if\_else(sex == 1, "M", "F")) %>%   
 filter(age\_cat != "50+" & ausualpl\_cat != "Other") %>%  
 ggplot(aes(x = ausualpl\_cat, y = col\_2, fill = ausualpl\_cat)) +  
 geom\_col() +  
 scale\_y\_continuous(limits = c(0,1), breaks = c(0, 0.25, 0.5, 0.75, 1)) +  
 geom\_errorbar(aes(ymin = ci\_l, ymax = ci\_u)) +  
 facet\_grid(~age\_cat + sex) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") +  
 labs(y = "Percent Had Colorectal Screen, Last 5 Years", x = "Usual Source of Care (Have/Have Not)")  
  
# insurance type barchart  
ins\_tab %>%  
 mutate(sex = if\_else(sex == 1, "M", "F")) %>%  
 filter(age\_cat != "50+") %>%  
 ggplot(aes(x = cover\_cat, y = col\_2, fill = cover\_cat)) +  
 geom\_col() +   
 scale\_y\_continuous(limits = c(0,1), breaks = c(0, 0.25, 0.5, 0.75, 1)) +  
 geom\_errorbar(aes(ymin = ci\_l, ymax = ci\_u)) +  
 facet\_grid(~age\_cat + sex) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none", axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 labs(y = "Percent Had Colorectal Screen, Last 5 Years", x = "Type of Insurance Coverage")  
  
# fit model  
  
# get sample sizes  
# 17056  
col\_temp = col\_dat %>%  
 filter(domain == 1)  
  
#7436  
col\_tempm = col\_temp %>%   
 filter(sex == 1)  
  
#9620  
col\_tempf = col\_temp %>%   
 filter(sex == 2)  
  
# 8678  
col\_temp50 = col\_temp %>%  
 filter(age\_cat == "50–64")  
  
# 8378  
col\_temp65 = col\_temp %>%  
 filter(age\_cat == "65+")  
  
col\_dat2 = col\_dat %>%  
 mutate(finc\_cat2 = if\_else(finc\_cat == "<200%", finc\_cat,  
 if\_else(finc\_cat == "Unknown", finc\_cat, ">=200%")),  
 eth\_cat2 = if\_else(eth\_cat == "Hispanic", eth\_cat, "Non-Hispanic"),  
 sex = if\_else(sex == 1, "M", "F"))  
  
des2 = svydesign(ids = ~psu\_p, strata = ~strat\_p, weights = ~wtfa\_sa, nest = TRUE,  
 data = col\_dat2)  
  
col2\_fit = svyglm(col\_2 ~ as.factor(age\_cat) + as.factor(sex) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat) + as.factor(race\_cat) + as.factor(eth\_cat2) + as.factor(imm\_stat),  
 design = des2, subset = domain == 1,   
 family = binomial(link = "logit"))  
  
summary(col2\_fit)  
summ(col2\_fit)  
  
  
# Rao-scott LRT ANOVA  
# test significance of full model  
  
regTermTest(col2\_fit, ~ as.factor(age\_cat) + as.factor(sex) + as.factor(educ\_cat) +  
 as.factor(finc\_cat2) + as.factor(ausualpl\_cat) + as.factor(cover\_cat) +  
 as.factor(lcond\_chronic\_cat) + as.factor(race\_cat) +  
 as.factor(eth\_cat2) + as.factor(imm\_stat), method = "LRT")  
  
# test significance of individual terms/term groups  
regTermTest(col2\_fit, "as.factor(age\_cat)",   
 method = "LRT") # not sig  
  
regTermTest(col2\_fit, "as.factor(sex)",   
 method = "LRT") # not sig  
  
regTermTest(col2\_fit, "as.factor(educ\_cat)",  
 method = "LRT") # not sig  
  
regTermTest(col2\_fit, "as.factor(finc\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(col2\_fit, "as.factor(ausualpl\_cat)",  
 method = "LRT") # not sig  
  
regTermTest(col2\_fit, "as.factor(cover\_cat)",  
 method = "LRT") # not sig  
  
regTermTest(col2\_fit, "as.factor(lcond\_chronic\_cat)",  
 method = "LRT") # sig  
  
regTermTest(col2\_fit, "as.factor(race\_cat)",  
 method = "LRT") # sig  
  
regTermTest(col2\_fit, "as.factor(eth\_cat2)",  
 method = "LRT") # not sig  
  
regTermTest(col2\_fit, "as.factor(imm\_stat)",  
 method = "LRT") # not sig  
  
  
# fit reduced model with significant predictors  
col2\_fit2 = svyglm(col\_2 ~ as.factor(finc\_cat2) + as.factor(lcond\_chronic\_cat) +  
 as.factor(race\_cat),  
 design = des2, subset = domain == 1,   
 family = binomial(link = "logit"))  
summary(col2\_fit2)  
summ(col2\_fit2)  
  
# Rao-Scott of full model  
regTermTest(col2\_fit2, ~ as.factor(finc\_cat2) + as.factor(lcond\_chronic\_cat) +  
 as.factor(race\_cat))  
  
# single term/term group significance  
regTermTest(col2\_fit2, "as.factor(finc\_cat2)",  
 method = "LRT") # sig  
  
regTermTest(col2\_fit2, "as.factor(lcond\_chronic\_cat)",  
 method = "LRT") # sig  
  
regTermTest(col2\_fit2, "as.factor(race\_cat)",  
 method = "LRT") # sig  
  
  
# give labels to the coefficients to look nicer on the OR graph  
coef <- names(coef(col2\_fit2))  
coef\_new = stringr::str\_remove(coef, "^[^\_]\*\_cat[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_cat2[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_stat[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_cat[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*lcond\_chronic\_cat[)]")  
names(coef) <- coef\_new  
coef1 = coef[-1] # remove intercept   
coef2 = coef1[-7] # remove coefficient for Usual Care = Other  
coef3c = coef2[-2] # remove income = unknown  
  
model\_coef = broom::tidy(col2\_fit2, conf.int = TRUE,   
 conf.level = 0.95, exponentiate = TRUE) %>%  
 mutate(term = case\_when(term == "(Intercept)" ~ "Intercept",  
 term == "as.factor(finc\_cat2)>=200%" ~ ">=200% vs <200% Poverty Level",  
 term == "as.factor(race\_cat)Asian" ~ "Asian vs American Indian/Alaskan Native",  
 term == "as.factor(race\_cat)Black" ~ "Black vs American Indian/Alaskan Native",  
 term == "as.factor(race\_cat)White" ~ "White vs American Indian/Alaskan Native"))   
  
model\_coef[is.na(model\_coef)] = "Chronic Condition vs No Chronic Condition"  
model\_coef = model\_coef[-3,]  
  
model\_coef %>% knitr::kable()  
  
  
# plot ORs with CIs  
jtools::plot\_summs(col2\_fit2, coefs = coef3c, exp = TRUE) +  
 labs(title = "Colorectal Screen Significant Predictors") +  
 scale\_x\_continuous(limits = c(0, 6), breaks = c(0, 2, 4, 6))  
  
# PSA CODE  
  
# data preparation  
  
cancer\_df = read\_csv("cancerxx.csv") %>%   
 janitor::clean\_names() %>%   
 dplyr::select(hhx, fmx, fpx, #identifiers  
 wtfa\_sa, #weights  
 strat\_p, psu\_p, #for design  
 region,   
 psahad, #ever has a psa test  
 rpsa1\_mt, #month of most recent psa test  
 rpsa1n:psaexp)  
  
fam\_dat = read\_csv("familyxx.csv") %>%  
 janitor::clean\_names() %>%   
 dplyr::select(hhx, fmx, #identifiers  
 rat\_cat4, rat\_cat5) # Ratio of family income to the poverty threshold   
  
pers\_dat = read\_csv("personsx.csv") %>%  
 janitor::clean\_names() %>%   
 dplyr::select(hhx, fmx, fpx, #identifiers  
 age\_p, #age  
 educ1, #education  
 sex, #gender  
 notcov, cover, cover65, cover65o, #coverage > 65, 65+, alternate 65+  
 la1ar, #limitation  
 lcondrt, #limitation is chronic  
 lachronr, #chronic limitation  
 hiscodi3, #ethnicity recode,  
 racreci3, #race recode  
 yrsinus, #immigration status (for people not born in the US)  
 plborn # place born  
 )  
  
adult\_dat = read\_csv("samadult.csv") %>%  
 janitor::clean\_names() %>%   
 dplyr::select(hhx, fmx, fpx, #identifiers  
 ausualpl, ahcplrou, ahcplknd, #Usual source of care - different options  
 fla1ar) #functional limitation  
  
psa\_df = cancer\_df %>%   
 left\_join(adult\_dat, by = c("hhx", "fmx", "fpx")) %>%   
 left\_join(pers\_dat, by = c("hhx", "fmx", "fpx")) %>%   
 left\_join(fam\_dat, by = c("hhx", "fmx"))  
head(psa\_df)  
  
# data manipulation  
psa\_df=psa\_df %>%   
 mutate(educ\_cat = case\_when(educ1 < 13 ~ "Less than high school",  
 educ1 >= 13 & educ1 < 15 ~ "High school",  
 educ1 >= 15 & educ1 < 18 ~ "Some college",  
 educ1 >= 18 & educ1 <= 21 ~ "College graduate"))  
  
psa\_df=psa\_df %>%   
 mutate(finc\_cat = case\_when(rat\_cat5 <= 7 | rat\_cat5 %in% c(15, 16) ~ "<200%",  
 rat\_cat5 %in% c(8, 9) ~ "200–299%",   
 rat\_cat5 %in% c(10, 11) ~ "300–399%",  
 rat\_cat5 >= 18 & educ1 <= 21 ~ "400–499%",  
 rat\_cat5 == 14 ~">=500%",  
 rat\_cat5 == 17 ~">=200%, no further detail",  
 rat\_cat5 %in% c(96, 99) ~ "Unknown"))  
  
psa\_df = psa\_df %>%   
 mutate(ausualpl\_cat = case\_when(ausualpl == 2 ~ "No",  
 ausualpl %in% c(1, 3) ~ "Yes",  
 ausualpl %in% c(7, 8, 9) ~ "Other"))  
  
psa\_df <- psa\_df %>%   
 mutate(cover\_cat = case\_when(notcov == 1 | cover == 4 | cover65 == 6 ~ "None",  
 cover == 2 | cover65 %in% 2:4 ~ "Public",  
 cover %in% c(1, 3) | cover65 %in% c(1, 5) ~ "Private/Military"))  
  
psa\_df <- psa\_df %>%   
 mutate(lcond\_chronic\_cat = if\_else(lcondrt == 1, "Yes", "No"))  
  
psa\_df <- psa\_df %>%   
 mutate(race\_cat = case\_when(racreci3 == 1 ~ "White",  
 racreci3 == 2 ~ "Black",  
 racreci3 == 3 ~ "Asian",  
 racreci3 == 4 ~ "AN/AI"),  
 eth\_cat = case\_when(hiscodi3 == 1 ~ "Hispanic",  
 hiscodi3 == 2 ~ "Non-Hispanic White",  
 hiscodi3 == 3 ~ "Non-Hispanic Black",  
 hiscodi3 == 4 ~ "Non-Hispanic Asian",  
 hiscodi3 == 5 ~ "Non-Hispanic AN/AI"),  
 yrus\_cat = case\_when(yrsinus < 4 ~ "In U.S. < 10 yrs",  
 yrsinus == 4 | yrsinus == 5 ~ "In U.S. >= 10 yrs",  
 plborn == 1 ~ "Born in U.S.")) %>%   
 mutate\_at(vars(psaadv, psadisav, psaexp, psasugg), ~factor(.x))  
  
psa\_df =  
 psa\_df %>%  
 # create domain variable for inclusion criteria  
 mutate(domain = if\_else(sex == 1 & age\_p >= 50, 1,0),  
 age\_cat = case\_when(age\_p >= 50 & age\_p < 65 ~ "50–64",  
 age\_p >= 65 ~ "65+"),  
 imm\_stat = case\_when(yrsinus < 4 ~ "In U.S. < 10 yrs",  
 yrsinus == 4 | yrsinus == 5 ~ "In U.S. >= 10 yrs",  
 plborn == 1 ~ "Born in U.S."))   
  
# fit survey design  
  
psa\_df %>% count(rpsa1\_mt)  
  
psa\_df =   
 psa\_df %>%   
 mutate(psa\_1yr = case\_when(rpsa1\_mt %in% c(1:12) ~ 1,  
 rpsa1\_mt %in% c(96,97,99) ~ 0)) %>%   
 filter(strat\_p != 70)  
  
psa\_df %>%   
 filter(domain==1) %>%   
 summarise(n())  
  
des = svydesign(ids = ~psu\_p, strata = ~strat\_p, weights = ~wtfa\_sa, nest = TRUE, data = psa\_df)  
  
# tables  
  
# percent of men who have had PSA in the last two years by age  
age\_pct = svyby(~psa\_1yr, by = ~domain+age\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("ci", "se"))  
age\_pct %>% filter(domain == 1) %>% dplyr::select(-domain, -se) %>% knitr::kable()  
  
# percent of men who have had PSA by education  
edu\_pct = svyby(~psa\_1yr, by = ~domain+educ\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
edu\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(educ\_cat) %>%  
 summarise(count = n())  
edu\_pct = left\_join(edu\_pct, edu\_counts, by = "educ\_cat")  
edu\_pct %>%   
 knitr::kable()  
  
# percent of men who have had PSA by financial category  
finc\_pct = svyby(~psa\_1yr, by = ~domain+finc\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
finc\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(finc\_cat) %>%  
 summarise(count = n())  
finc\_pct = left\_join(finc\_pct, finc\_counts, by = "finc\_cat")  
finc\_pct %>% knitr::kable()  
  
# percent of men who have had PSA by usual care category  
ausualp\_pct = svyby(~psa\_1yr, by = ~domain+ausualpl\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
usual\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(ausualpl\_cat) %>%  
 summarise(count = n())  
ausualp\_pct = left\_join(ausualp\_pct, usual\_counts, by = "ausualpl\_cat")  
ausualp\_pct %>% knitr::kable()  
  
# percent of men who have had PSA by health coverage  
cover\_pct = svyby(~psa\_1yr, by = ~domain+cover\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
cover\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(cover\_cat) %>%  
 summarise(count = n())  
cover\_pct = left\_join(cover\_pct, cover\_counts, by = "cover\_cat")  
cover\_pct %>% knitr::kable()  
  
# percent of men who have had PSA by chronic conditions  
lcond\_chronic\_pct = svyby(~psa\_1yr, by = ~domain+lcond\_chronic\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
chronic\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(lcond\_chronic\_cat) %>%  
 summarise(count = n())  
lcond\_chronic\_pct = left\_join(lcond\_chronic\_pct, chronic\_counts, by = "lcond\_chronic\_cat")  
lcond\_chronic\_pct %>% knitr::kable()  
  
# percent of men who have had PSA by race  
race\_pct = svyby(~psa\_1yr, by = ~domain+race\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
race\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(race\_cat) %>%  
 summarise(count = n())  
race\_pct = left\_join(race\_pct, race\_counts, by = "race\_cat")  
race\_pct %>% knitr::kable()  
  
# percent of men who have had PSA by ethnicity  
eth\_pct = svyby(~psa\_1yr, by = ~domain+eth\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
eth\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(eth\_cat) %>%  
 summarise(count = n())  
eth\_pct = left\_join(eth\_pct, eth\_counts, by = "eth\_cat")  
eth\_pct %>% knitr::kable()  
  
# percent of men who have had PSA by immigration status  
imm\_pct = svyby(~psa\_1yr, by = ~domain+imm\_stat, svymean, na.rm = TRUE,  
 design = des, vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
imm\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(imm\_stat) %>%  
 summarise(count = n())  
imm\_pct = left\_join(imm\_pct, imm\_counts, by = "imm\_stat")  
imm\_pct %>% knitr::kable()  
  
# tables by age group  
  
# percent of men who have had PSA by education and age   
edu\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + educ\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
edu\_tab = edu\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)   
edu\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, educ\_cat) %>%  
 summarise(count = n())  
edu\_tab = left\_join(edu\_tab, edu\_counts2, by = c("age\_cat", "educ\_cat"))  
edu\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by financial category and age  
finc\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + finc\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
finc\_tab = finc\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)   
finc\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, finc\_cat) %>%  
 summarise(count = n())  
finc\_tab = left\_join(finc\_tab, finc\_counts2, by = c("age\_cat", "finc\_cat"))  
finc\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by usual care and age  
ausualp\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + ausualpl\_cat,  
 svymean, na.rm = TRUE, design = des,   
 vartype = c("se", "ci"))  
usual\_tab = ausualp\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)   
usual\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, ausualpl\_cat) %>%  
 summarise(count = n())  
usual\_tab = left\_join(usual\_tab, usual\_counts2, by = c("age\_cat", "ausualpl\_cat"))  
usual\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by health coverage and age  
cover\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + cover\_cat, svymean,  
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
ins\_tab = cover\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)  
ins\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, cover\_cat) %>%  
 summarise(count = n())  
ins\_tab = left\_join(ins\_tab, ins\_counts2, by = c("age\_cat", "cover\_cat"))  
ins\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by chronic conditions and age  
lcond\_chronic\_pct\_strat = svyby(~psa\_1yr,   
 by = ~domain + age\_cat + lcond\_chronic\_cat,  
 svymean, na.rm = TRUE, design = des,  
 vartype = c("se", "ci"))  
dis\_tab = lcond\_chronic\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)  
dis\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, lcond\_chronic\_cat) %>%  
 summarise(count = n())  
dis\_tab = left\_join(dis\_tab, dis\_counts2, by = c("age\_cat", "lcond\_chronic\_cat"))  
dis\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by race and age  
race\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + race\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
race\_tab = race\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)   
race\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, race\_cat) %>%  
 summarise(count = n())  
race\_tab = left\_join(race\_tab, race\_counts2, by = c("age\_cat", "race\_cat"))  
race\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by ethnicity and age  
eth\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + eth\_cat, svymean,   
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
eth\_tab = eth\_pct\_strat %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)   
eth\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, eth\_cat) %>%  
 summarise(count = n())  
eth\_tab = left\_join(eth\_tab, eth\_counts2, by = c("age\_cat", "eth\_cat"))  
eth\_tab %>% knitr::kable()  
  
# percent of men who have had PSA by immigration and age  
imm\_pct\_strat = svyby(~psa\_1yr, by = ~domain + age\_cat + imm\_stat, svymean,  
 na.rm = TRUE, design = des, vartype = c("se", "ci"))  
imm\_tab = imm\_pct\_strat %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se)  
imm\_counts2 = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat, imm\_stat) %>%  
 summarise(count = n())  
imm\_tab = left\_join(imm\_tab, imm\_counts2, by = c("age\_cat", "imm\_stat"))  
imm\_tab %>% knitr::kable()  
  
# overall percent of men who have had PSA  
total = svyby(~psa\_1yr, by = ~domain + age\_cat, svymean, na.rm = TRUE,   
 design = des, vartype = c("se", "ci"))  
tot\_tab = total %>%   
 filter(domain == 1) %>%   
 dplyr::select(-domain, -se)   
tot\_counts = filter(psa\_df, domain == 1) %>%  
 group\_by(age\_cat) %>%  
 summarise(count = n())  
tot\_tab = left\_join(tot\_tab, tot\_counts, by = "age\_cat")  
tot\_tab %>% knitr::kable()  
all\_counts = filter(psa\_df, domain == 1)  
tot\_pct = svyby(~psa\_1yr, by = ~domain, svymean, na.rm = TRUE, design = des,  
 vartype = c("se", "ci")) %>%  
 filter(domain == 1) %>%  
 dplyr::select(-domain, -se) %>%  
 mutate(age\_cat = "50+",  
 count = 7426)  
  
# COMBINE TABLES BY AGE WITH TABLES OVERALL  
# overall  
tot\_tab = rbind(tot\_pct, tot\_tab)  
tot\_tab2 = tot\_tab %>%  
 mutate(type = "Total",  
 level = "-")  
# education  
edu\_pct2 = edu\_pct %>%  
 mutate(age\_cat = "50+")  
edu\_tab = rbind(edu\_pct2, edu\_tab)  
# factor levels for presentation  
edu\_tab2 = edu\_tab %>%  
 mutate(type = "Education") %>%  
 rename(level = educ\_cat) %>%  
 mutate(level = factor(level, levels = c("Less than high school", "High school", "Some college", "College graduate"))) %>%  
 arrange(level)  
# financial category  
finc\_pct2 = finc\_pct %>%  
 mutate(age\_cat = "50+")  
finc\_tab = rbind(finc\_pct2, finc\_tab)  
# factor levels for presentation  
finc\_tab2 = finc\_tab %>%  
 mutate(type = "Family Income Poverty Ratio") %>%  
 rename(level = finc\_cat) %>%  
 mutate(level = factor(level, levels = c("<200%", ">=200%, no further detail", "200–299%", "300–399%", "400–499%", ">=500%", "Unknown"))) %>%  
 arrange(level)  
# usual care category  
ausualp\_pct2 = ausualp\_pct %>%  
 mutate(age\_cat = "50+")  
usual\_tab = rbind(ausualp\_pct2, usual\_tab)  
# factor levels for presentation  
usual\_tab2 = usual\_tab %>%  
 mutate(type = "Usual Source of Care") %>%  
 rename(level = ausualpl\_cat) %>%  
 mutate(level = factor(level, levels = c("No", "Yes", "Other"))) %>%  
 arrange(level)  
# insurance coverage category  
cover\_pct2 = cover\_pct %>%  
 mutate(age\_cat = "50+")  
ins\_tab = rbind(cover\_pct2, ins\_tab)  
# factor levels for presentation  
ins\_tab2 = ins\_tab %>%  
 mutate(type = "Insurance Type") %>%  
 rename(level = cover\_cat) %>%  
 mutate(level = factor(level, levels = c("None", "Public", "Private/Military"))) %>%  
 arrange(level)  
# chronic condition category  
lcond\_chronic\_pct2 = lcond\_chronic\_pct %>%  
 mutate(age\_cat = "50+")  
dis\_tab = rbind(lcond\_chronic\_pct2, dis\_tab)  
# factor levels for presentation  
dis\_tab2 = dis\_tab %>%  
 mutate(type = "Chronic Disability") %>%  
 rename(level = lcond\_chronic\_cat) %>%  
 mutate(level = factor(level, levels = c("Yes", "No"))) %>%  
 arrange(level)  
# ethnic category  
eth\_pct2 = eth\_pct %>%  
 mutate(age\_cat = "50+")  
eth\_tab = rbind(eth\_pct2, eth\_tab)  
# factor levels for presentation  
eth\_tab2 = eth\_tab %>%  
 mutate(type = "Ethnicity") %>%  
 rename(level = eth\_cat) %>%  
 mutate(level = factor(level, levels = c("Hispanic", "Non-Hispanic White", "Non-Hispanic Black", "Non-Hispanic AN/AI", "Non-Hispanic Asian"))) %>%  
 arrange(level)  
# race category  
race\_pct2 = race\_pct %>%  
 mutate(age\_cat = "50+")  
race\_tab = rbind(race\_pct2, race\_tab)  
# factor levels for presentation  
race\_tab2 = race\_tab %>%  
 mutate(type = "Race") %>%  
 rename(level = race\_cat) %>%  
 mutate(level = factor(level, levels = c("White", "Black", "AN/AI", "Asian"))) %>%  
 arrange(level)  
# immigration category  
imm\_pct2 = imm\_pct %>%  
 mutate(age\_cat = "50+")  
imm\_tab = rbind(imm\_pct2, imm\_tab)  
# factor levels for presentation  
imm\_tab2 = imm\_tab %>%  
 mutate(type = "Immigration") %>%  
 rename(level = imm\_stat) %>%  
 mutate(level = factor(level, levels = c("In U.S. < 10 yrs", "In U.S. >= 10 yrs", "Born in U.S."))) %>%  
 arrange(level)  
# create table of percentages of women who have gotten mammograms within the last two years (still need to add CIs)  
tab\_one = rbind(tot\_tab2, edu\_tab2, finc\_tab2, usual\_tab2, ins\_tab2, dis\_tab2, eth\_tab2, race\_tab2, imm\_tab2) %>%  
 mutate(psa\_1yr = round(psa\_1yr\*100, 1),  
 ci\_l = round(ci\_l\*100, 1),  
 ci\_u = round(ci\_u\*100, 1),  
 CI = str\_c(ci\_l, ", ", ci\_u)) %>%  
 rename(Percent = psa\_1yr,  
 N = count) %>%  
 dplyr::select(-ci\_l, -ci\_u) %>%  
 pivot\_wider(names\_from = age\_cat, values\_from = c(N, Percent, CI)) %>%  
 janitor::clean\_names() %>%   
 dplyr::select(type, level, n\_50, percent\_50, ci\_50, n\_50\_64, percent\_50\_64, ci\_50\_64, n\_65, percent\_65, ci\_65, everything())  
# print percentages  
tab\_one %>% knitr::kable()  
  
# create barplots for in class presentation  
  
# sample size  
usual\_size = psa\_df %>%  
 filter(domain == 1) %>%  
 group\_by(ausualpl\_cat) %>%  
 summarise(count = n()) %>%  
 filter(ausualpl\_cat != "NA")  
sum(pull(usual\_size, count))  
  
# usual source of care barchart  
usual\_tab %>%  
 filter(age\_cat != "50+" & ausualpl\_cat != "Other") %>%  
 ggplot(aes(x = ausualpl\_cat, y = psa\_1yr, fill = ausualpl\_cat)) +  
 geom\_col() +  
 scale\_y\_continuous(limits = c(0,1), breaks = c(0, 0.25, 0.5, 0.75, 1)) +  
 geom\_errorbar(aes(ymin = ci\_l, ymax = ci\_u)) +  
 facet\_grid(~age\_cat) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") +  
 labs(y = "Percent Had Mammogram, Last 2 Years", x = "Usual Source of Care (Have/Have Not)", title = "N = 7426")  
  
# sample size  
ins\_size = psa\_df %>%  
 filter(domain == 1) %>%  
 group\_by(cover\_cat) %>%  
 summarise(count = n()) %>%  
 filter(cover\_cat != "NA")  
sum(pull(ins\_size, count))  
# insurance type barchart  
ins\_tab %>%  
 filter(age\_cat != "50+") %>%  
 ggplot(aes(x = cover\_cat, y = psa\_1yr, fill = cover\_cat)) +  
 geom\_col() +   
 scale\_y\_continuous(limits = c(0,1), breaks = c(0, 0.25, 0.5, 0.75, 1)) +  
 geom\_errorbar(aes(ymin = ci\_l, ymax = ci\_u)) +  
 facet\_grid(~age\_cat) + ggthemes::theme\_few() + ggthemes::scale\_fill\_few() + theme(legend.position = "none") +  
 labs(y = "Percent Had Mammogram, Last 2 Years", x = "Type of Insurance Coverage", title = "N = 7426")  
  
# fit model  
  
# make financial status and ethnicity only two levels   
psa\_df2 = psa\_df %>%  
 mutate(finc\_cat2 = if\_else(finc\_cat == "<200%", finc\_cat,  
 if\_else(finc\_cat == "Unknown", finc\_cat, ">=200%")),  
 eth\_cat2 = if\_else(eth\_cat == "Hispanic", eth\_cat, "Non-Hispanic"),  
 imm\_stat2 = if\_else(imm\_stat == "Born in U.S.", imm\_stat, "Immigrated"),  
 ausualpl\_cat2 = replace(ausualpl\_cat, ausualpl\_cat == "Other", NA),  
 lcond\_chronic\_cat2 = if\_else(lcond\_chronic\_cat == "Yes", "Yes (Chronic)", lcond\_chronic\_cat))  
  
des2 = svydesign(ids = ~psu\_p, strata = ~strat\_p, weights = ~wtfa\_sa, nest = TRUE, data = psa\_df2)  
  
model1 = svyglm(psa\_1yr ~ as.factor(age\_cat) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2) + as.factor(race\_cat) + as.factor(eth\_cat2) + as.factor(imm\_stat2),   
 design = des2, subset = domain == 1, family = binomial(link = "logit"))   
summary(model1)  
summ(model1)   
   
regTermTest(model1, ~ as.factor(age\_cat) + as.factor(educ\_cat) + as.factor(finc\_cat2) + as.factor(ausualpl\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2) + as.factor(race\_cat) + as.factor(eth\_cat2) + as.factor(imm\_stat2),  
 method = "LRT")  
  
# test significance of individual terms/term groups  
regTermTest(model1, "as.factor(age\_cat)",   
 method = "LRT") # not sig  
regTermTest(model1, "as.factor(educ\_cat)",  
 method = "LRT") # not sig  
regTermTest(model1, "as.factor(finc\_cat2)",  
 method = "LRT") # sig  
regTermTest(model1, "as.factor(ausualpl\_cat2)",  
 method = "LRT") # not sig  
regTermTest(model1, "as.factor(cover\_cat)",  
 method = "LRT") # sig  
regTermTest(model1, "as.factor(lcond\_chronic\_cat2)",  
 method = "LRT") # sig  
regTermTest(model1, "as.factor(race\_cat)",  
 method = "LRT") # not sig  
regTermTest(model1, "as.factor(eth\_cat2)",  
 method = "LRT") # not sig  
regTermTest(model1, "as.factor(imm\_stat2)",  
 method = "LRT") # not sig  
  
# fit reduced model with significant predictors  
# remove finc\_cat2, cover\_cat, lcond\_chronic\_cat2  
model2 = svyglm(psa\_1yr ~ as.factor(finc\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2),   
 design = des2, subset = domain == 1, family = binomial(link = "logit"))   
summary(model2)  
summ(model2)  
  
  
# Rao-Scott of full model  
regTermTest(model2, ~as.factor(finc\_cat2) + as.factor(cover\_cat) + as.factor(lcond\_chronic\_cat2))  
  
# single term/term group significance  
regTermTest(model2, "as.factor(finc\_cat2)",  
 method = "LRT") # sig  
regTermTest(model2, "as.factor(cover\_cat)",  
 method = "LRT") # sig  
regTermTest(model2, "as.factor(lcond\_chronic\_cat2)",  
 method = "LRT") # sig  
  
# give labels to the coefficients to look nicer on the OR graph  
coef = names(coef(model2))  
coef\_new = stringr::str\_remove(coef, "^[^\_]\*\_cat[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_cat2[)]")  
coef\_new = stringr::str\_remove(coef\_new, "^[^\_]\*\_chronic")  
coef\_new = stringr::str\_remove(coef\_new, "^\*\_cat2[)]")  
names(coef) <- coef\_new  
coef1 = coef[-1] # remove intercept   
model\_coef = broom::tidy(model2, conf.int = TRUE, conf.level = 0.95, exponentiate = TRUE) %>%  
 mutate(term = case\_when(term == "(Intercept)" ~ "Intercept",  
 term == "as.factor(age\_cat)50–64" ~ "50-64 vs <50",  
 term == "as.factor(age\_cat)65+" ~ "65+ vs <50",  
 term == "as.factor(educ\_cat)High school" ~ "High School Degree vs College Degree",  
 term == "as.factor(educ\_cat)Less than high school" ~ "Less than High School vs College Degree",  
 term == "as.factor(educ\_cat)Some college" ~ "Some college vs College Degree",  
 term == "as.factor(finc\_cat2)>=200%" ~ ">=200% vs <200% Poverty Level",  
 term == "as.factor(ausualpl\_cat2)Yes" ~ "Usual Source of Care vs No Usual Source of Care",  
 term == "as.factor(cover\_cat)Private/Military" ~ "Private/Military Insurance vs No Insurance",  
 term == "as.factor(cover\_cat)Public" ~ "Public Insurance vs No Insurance",  
 term == "as.factor(lcond\_chronic\_cat2)Yes (Chronic)" ~ "Chronic Condition vs No Chronic Condition",  
 term == "as.factor(eth\_cat2)Non-Hispanic" ~ "Non-Hispanic vs Hispanic"))  
model\_coef %>% knitr::kable(digits = 2)  
# plot ORs with CIs  
jtools::plot\_summs(model2, coefs = coef1, exp = TRUE) +  
 labs(title = "PSA Significant Predictors")