Code

ALA Mode (Group 2): Anja Shahu, Anna Wuest, Ligia Flores

11/4/2020

```r
library(tidyverse)
library(RColorBrewer)
```

```r
url <- "https://meps.ahrq.gov/mepsweb/data_files/pufs/h209dat.zip"
download.file(url, temp <- tempfile())
meps_path <- unzip(temp, exdir = tempdir())
source("https://meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h209/h209ru.txt")
unlink(temp)
```

```r
# creating a reduced data frame including only the variables that we'll be considering
h209red <- data.frame("pap" = h209$ADPAP42,
                      "region" = h209$REGION18,
                      "race" = h209$RACETHX,
                      "age" = h209$AGE18X,
                      "marital_stat" = h209$MARRY18X, # newly added
                      "educ" = h209$EDUCYR, # newly added
                      "smoke_freq" = h209$OFTSMK53, # newly added
                      "income_indiv" = h209$TTLP18X,
                      "income_fam" = h209$FAMINC18,
                      "income_percpov" = h209$POVLEV18,
                      "hrsworked_rd1" = h209$HOUR31H,
                      "hrsworked_rd2" = h209$HOUR42H,
                      "hrsworked_rd3" = h209$HOUR53H,
                      "limitation" = h209$ACTLIM31, # newly added
                      "menhlth_rd1" = h209$MNHLTH31, # already included
                      "menhlth_rd2" = h209$MNHLTH42, # already included
                      "menhlth_rd3" = h209$MNHLTH53, # already included
                      "genhlth_rd1" = h209$RTHLTH31,
                      "genhlth_rd2" = h209$RTHLTH42,
                      "genhlth_rd3" = h209$RTHLTH53,
                      "totexp" = h209$TOTEXP18,
                      "outofpocket_exp" = h209$TOTSLF18,
                      "afford_care" = h209$AFRDCA42,
                      "have_usc" = h209$HAVEUS42,
                      "dist_from_usc" = h209$TMTKUS42,
                      "rch_usc_byphn" = h209$PHNREG42,
                      "usc_offhrs_nw" = h209$OFFHOU42,
                      "usc_asks_abt_trts" = h209$TREATM42,
                      "usc_asks_hlp_dec" = h209$DECIDE42,
                      "usc_expln_options" = h209$EXPLOP42,
                      "usc_spk_lang" = h209$PRVSPK42,
                      "usc_gender" = h209$GENDRP42,
                      "inscov_gen_2018" = h209$INSCOV18)
```

```r
rm(h209) # remove original data set from environment

h209red <- h209red %>%
  as_tibble() %>%
  filter(pap != -1) %>% # filtering out the people who were not asked pap smear question
  filter(age >= 21 & age <= 65) # filtering to women ages 21-65 (note there was 1 inapplicable person t

## inputting NAs into hours worked variables
h209red$hrsworked_rd1[h209red$hrsworked_rd1 == -1] <- NA
h209red$hrsworked_rd2[h209red$hrsworked_rd2 == -1] <- NA
h209red$hrsworked_rd3[h209red$hrsworked_rd3 == -1] <- NA

# hours worked (rounded average)
h209red <- h209red %>% rowwise() %>%
  mutate(hrs_worked_avg = round(mean(c(hrsworked_rd1, hrsworked_rd2, hrsworked_rd3), na.rm = TRUE)))

h209red$hrs_worked_avg[is.nan(as.numeric(h209red$hrs_worked_avg))] <- NA

summary(h209red$hrs_worked_avg) # too many missing variables to use
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.00   30.00   40.00   36.36   40.00  168.00    1654
```

```r
# re-calculating the mental and general health variables

# perceived mental heath NA
h209red$menhlth_rd1[h209red$menhlth_rd1 == -8 ]<- NA
h209red$menhlth_rd1[h209red$menhlth_rd1 == -1 ]<- NA

h209red$menhlth_rd2[h209red$menhlth_rd2 == -7 ]<- NA
h209red$menhlth_rd2[h209red$menhlth_rd2 == -8 ]<- NA

h209red$menhlth_rd3[h209red$menhlth_rd3 == -7 ]<- NA
h209red$menhlth_rd3[h209red$menhlth_rd3 == -8 ]<- NA
h209red$menhlth_rd3[h209red$menhlth_rd3 == -1 ]<- NA

# perceived mental health (rounded average)
h209red <- h209red %>% rowwise() %>%
  mutate(menhlth_avg = round(mean(c(menhlth_rd1, menhlth_rd2, menhlth_rd3), na.rm=TRUE)))

summary(h209red$menhlth_avg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   2.115   3.000   5.000
```

```r
# perceived mental health as factor (rounded average)
h209red <- h209red %>%
  mutate(menhlth_avg_f = factor(menhlth_avg,
                                levels = c("5", "4", "3", "2", "1"))) %>%
  mutate(menhlth_avg_f = fct_recode(menhlth_avg_f,
                                    "poor" = "5",
                                    "fair" = "4",
                                    "good" = "3",
                                    "very good" = "2",
                                    "excellent" = "1"))
```

```r
# re-calculating the general health variables

# perceived general heath NA
h209red$genhlth_rd1[h209red$genhlth_rd1 == -8 ] <- NA
h209red$genhlth_rd1[h209red$genhlth_rd1 == -1 ] <- NA

h209red$genhlth_rd2[h209red$genhlth_rd2 == -8 ] <- NA

h209red$genhlth_rd3[h209red$genhlth_rd3 == -7 ] <- NA
h209red$genhlth_rd3[h209red$genhlth_rd3 == -8 ] <- NA
h209red$genhlth_rd3[h209red$genhlth_rd3 == -1 ] <- NA

# perceived mental health (rounded average)
h209red <- h209red %>% rowwise() %>%
  mutate(genhlth_avg = round(mean(c(genhlth_rd1, genhlth_rd2, genhlth_rd3), na.rm=TRUE)))

summary(h209red$genhlth_avg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   2.351   3.000   5.000
```

```r
# perceived general health as factor (rounded average)
h209red <- h209red %>%
  mutate(genhlth_avg_f = factor(genhlth_avg,
                                levels = c("5", "4", "3", "2", "1"))) %>%
  mutate(genhlth_avg_f = fct_recode(genhlth_avg_f,
                               "poor" = "5",
                               "fair" = "4",
                               "good" = "3",
                               "very good" = "2",
                               "excellent" = "1"))

# creating factor versions of other categorical variables

# pap status
h209red <- h209red %>%
  mutate(pap_f = factor(pap,
                        levels = c("1", "2", "-15"))) %>%
  mutate(pap_f = fct_recode(pap_f,
                            "yes" = "1",
                            "no" = "2",
                            NULL = "-15"))

# region
h209red <- h209red %>%
  mutate(region_f = factor(region,
                           levels = c("1", "2", "3", "4"))) %>%
  mutate(region_f = fct_recode(region_f,
                               "northeast" = "1",
                               "midwest" = "2",
                               "south" = "3",
                               "west" = "4"))
```

```r
# race
h209red <- h209red %>%
  mutate(race_f = factor(race,
                         levels = c("2", "1", "3", "4", "5"))) %>%
  mutate(race_f = fct_recode(race_f,
                             "white" = "2",
                             "hispanic" = "1",
                             "black" = "3",
                             "asian" = "4",
                             "other or multiple races" = "5"))

# marital status
h209red <- h209red %>%
  mutate(marital_stat_f = factor(marital_stat,
                                 levels = c("5", "1", "2", "3", "4"))) %>%
  mutate(marital_stat_f = fct_recode(marital_stat_f,
                                     "never married" = "5",
                                     "married" = "1",
                                     "widowed" = "2",
                                     "divorced" = "3",
                                     "seperated" = "4"))
# education
h209red <- h209red %>%
  mutate(educ_f = factor(educ)) %>%
  mutate(educ_f = fct_collapse(educ_f,
                               "none or any elementary" = c("0", "1", "2", "3", "4", "5", "6", "7", "8")
                               "any high school" = c("9", "10", "11", "12"),
                               "any college" = c("13", "14", "15", "16", "17"),
                               NULL = "-15",
                               NULL = c("-8", "-7")))

# smoking frequency
h209red <- h209red %>%
  mutate(smoke_freq_f = factor(smoke_freq,
                               levels = c("3", "2", "1", "-8", "-7", "-1"))) %>%
  mutate(smoke_freq_f = fct_recode(smoke_freq_f,
                                   "never" = "3",
                                   "some days" = "2",
                                   "every day" = "1",
                                   NULL = "-8",
                                   NULL = "-7",
                                   NULL = "-1"))

# limitation
h209red <- h209red %>%
  mutate(limitation_f = factor(limitation,
                               levels = c("2", "1", "-8", "-7", "-1"))) %>%
  mutate(limitation_f = fct_recode(limitation_f,
                                   "no" = "2",
                                   "yes" = "1",
                                   NULL = "-8",
                                   NULL = "-7",
                                   NULL = "-1"))
```

```r
# ability to afford care
h209red <- h209red %>%
  mutate(afford_care_f = factor(afford_care,
                                levels = c("2", "1", "-8", "-7"))) %>%
  mutate(afford_care_f = fct_recode(afford_care_f,
                                    "no" = "2",
                                    "yes" = "1",
                                    NULL = "-8",
                                    NULL = "-7"))


# usual source of care status
h209red <- h209red %>%
  mutate(have_usc_f = factor(have_usc,
                             levels = c("2", "1", "-8", "-7"))) %>%
  mutate(have_usc_f = fct_recode(have_usc_f,
                                 "no" = "2",
                                 "yes" = "1",
                                 NULL = "-8",
                                 NULL = "-7"))


# distance from provider
h209red <- h209red %>%
  mutate(dist_from_usc = ifelse(have_usc_f == "no",
                                -100,
                                dist_from_usc)) %>% # creating level for not having a provider
  mutate(dist_from_usc_f = factor(dist_from_usc,
                                  levels = c("1", "2", "3", "4", "5", "6", "-100", "-8", "-7", "-1"))) %
  mutate(dist_from_usc_f = fct_recode(dist_from_usc_f,
                                      "<15" = "1",
                                      "15 to 30" = "2",
                                      "31 to 60" = "3",
                                      "61 to 90" = "4",
                                      "91 to 120" = "5",
                                      ">120" = "6",
                                      "no usc" = "-100",
                                      NULL = "-8",
                                      NULL = "-7",
                                      NULL = "-1"))


# ability to reach provider by phone
h209red <- h209red %>%
  mutate(rch_usc_byphn = ifelse(have_usc_f == "no",
                                -100,
                                rch_usc_byphn)) %>% # creating level for not having a provider
  mutate(rch_usc_byphn_f = factor(rch_usc_byphn,
                                  levels = c("4", "3", "2", "1", "-100", "-8", "-7", "-1"))) %>%
  mutate(rch_usc_byphn_f = fct_recode(rch_usc_byphn_f,
                                      "not at all difficult" = "4",
                                      "not too difficult" = "3",
                                      "somewhat difficult" = "2",
                                      "very difficult" = "1",
                                      "no usc" = "-100",
```

```r
                                          NULL = "-8",
                                          NULL = "-7",
                                          NULL = "-1"))


# provider offers office hours during nights/weekends
h209red <- h209red %>%
  mutate(usc_offhrs_nw = ifelse(have_usc_f == "no",
                                -100,
                                usc_offhrs_nw)) %>% # creating level for not having a provider
  mutate(usc_offhrs_nw_f = factor(usc_offhrs_nw,
                                  levels = c("-100", "2", "1", "-8", "-7", "-1"))) %>%
  mutate(usc_offhrs_nw_f = fct_recode(usc_offhrs_nw_f,
                                      "no usc" = "-100",
                                      "no" = "2",
                                      "yes" = "1",
                                      NULL = "-8",
                                      NULL = "-7",
                                      NULL = "-1"))
# provider asks about treatments
h209red <- h209red %>%
  mutate(usc_asks_abt_trts = ifelse(have_usc_f == "no",
                                    -100,
                                    usc_asks_abt_trts)) %>% # creating level for not having a provider
  mutate(usc_asks_abt_trts_f = factor(usc_asks_abt_trts,
                                      levels = c("-100", "2", "1", "-8", "-7", "-1"))) %>%
  mutate(usc_asks_abt_trts_f = fct_recode(usc_asks_abt_trts_f,
                                          "no usc" = "-100",
                                          "no" = "2",
                                          "yes" = "1",
                                          NULL = "-8",
                                          NULL = "-7",
                                          NULL = "-1"))


# provider asks person to help make decisions
h209red <- h209red %>%
  mutate(usc_asks_hlp_dec = ifelse(have_usc_f == "no",
                                   -100,
                                   usc_asks_hlp_dec)) %>% # creating level for not having a provider
  mutate(usc_asks_hlp_dec_f = factor(usc_asks_hlp_dec,
                                     levels = c("-100", "1", "2", "3", "4", "-8", "-7", "-1"))) %>%
  mutate(usc_asks_hlp_dec_f = fct_recode(usc_asks_hlp_dec_f,
                                         "no usc" = "-100",
                                         "never" = "1",
                                         "sometimes" = "2",
                                         "usually" = "3",
                                         "always" = "4",
                                         NULL = "-8",
                                         NULL = "-7",
                                         NULL = "-1"))


# provider presents and explains all options
h209red <- h209red %>%
  mutate(usc_expln_options = ifelse(have_usc_f == "no",
```

```r
                                               -100,
                                      usc_expln_options)) %>% # creating level for not having a provider
    mutate(usc_expln_options_f = factor(usc_expln_options,
                                        levels = c("-100", "2", "1", "-8", "-7", "-1"))) %>%
    mutate(usc_expln_options_f = fct_recode(usc_expln_options_f,
                                            "no usc" = "-100",
                                            "no" = "2",
                                            "yes" = "1",
                                            NULL = "-8",
                                            NULL = "-7",
                                            NULL = "-1"))


# insurance indicator in 2018
h209red <- h209red %>%
    mutate(inscov_gen_2018_f = factor(inscov_gen_2018,
                                      levels = c("1", "2", "3"))) %>%
    mutate(inscov_gen_2018_f = fct_recode(inscov_gen_2018_f,
                                          "any private" = "1",
                                          "public only" = "2",
                                          "uninsured" = "3"))

# creating combined provider availability variable using 1) distance 2) ability to reach by phone 3) of

h209red <- h209red %>%
    mutate(dist_from_usc = ifelse(have_usc_f == "no", 0, dist_from_usc))
h209red$dist_from_usc[h209red$dist_from_usc == -8] <- NA
h209red$dist_from_usc[h209red$dist_from_usc == -7] <- NA

h209red <- h209red %>%
    mutate(rch_usc_byphn = ifelse(have_usc_f == "no", 0, rch_usc_byphn))
h209red$rch_usc_byphn[h209red$rch_usc_byphn == -8] <- NA
h209red$rch_usc_byphn[h209red$rch_usc_byphn == -7] <- NA

h209red <- h209red %>%
    mutate(usc_offhrs_nw = ifelse(have_usc_f == "no", 0, usc_offhrs_nw))
h209red$usc_offhrs_nw[h209red$usc_offhrs_nw == -8] <- NA
h209red$usc_offhrs_nw[h209red$usc_offhrs_nw == -7] <- NA
h209red$usc_offhrs_nw[h209red$usc_offhrs_nw == -1] <- NA

# creating binary access variables to use for making combined score

# give 0 to those w/o provider
# give 1 to people who have to travel 30+ minutes
# and give 2 to people who are within 30 min
h209red <- h209red %>% mutate(dist_from_usc_bin = ifelse(dist_from_usc %in% c(1, 2), 2,
                                                  ifelse(dist_from_usc %in% c(3, 4, 5, 6), 1,
                                                         dist_from_usc)))


# give 0 to those w/o provider
# and give 1 to people who answer somewhat difficult or very difficult
# give 2 to people who answer not at all difficult or not too difficult
h209red <- h209red %>% mutate(rch_usc_byphn_bin = ifelse(rch_usc_byphn %in% c(1, 2), 1,
                                                  ifelse(rch_usc_byphn %in% c(3, 4, 5, 6), 2,
                                                         rch_usc_byphn)))
```

```r
# give 0 to those w/o provider
# and give 1 to people whose provider does not offer office hours during night/weekend
# give 2 to people whose provider does offer
h209red <- h209red %>% mutate(usc_offhrs_nw_bin = ifelse(usc_offhrs_nw == 1, 2,
                                              ifelse(usc_offhrs_nw == 2, 1,
                                                     usc_offhrs_nw)))


# finally creating combined availability score from binary variables
h209red <- h209red %>%
  mutate(usc_access_score = dist_from_usc_bin + rch_usc_byphn_bin + usc_offhrs_nw_bin)
```

```r
# creating combined provider satisfaction variable using 1) asking about treatments 2) asks person to h

h209red <- h209red %>%
  mutate(usc_asks_abt_trts = ifelse(have_usc_f == "no", 0, usc_asks_abt_trts))
h209red$usc_asks_abt_trts[h209red$usc_asks_abt_trts == -8] <- NA
h209red$usc_asks_abt_trts[h209red$usc_asks_abt_trts == -7] <- NA

h209red <- h209red %>%
   mutate(usc_asks_hlp_dec = ifelse(have_usc_f == "no", 0, usc_asks_hlp_dec))
h209red$usc_asks_hlp_dec[h209red$usc_asks_hlp_dec == -8] <- NA
h209red$usc_asks_hlp_dec[h209red$usc_asks_hlp_dec == -7] <- NA

h209red <- h209red %>%
  mutate(usc_expln_options = ifelse(have_usc_f == "no", 0, usc_expln_options))
h209red$usc_expln_options[h209red$usc_expln_options == -8] <- NA
h209red$usc_expln_options[h209red$usc_expln_options == -7] <- NA


# creating binary access variables to use for making combined score

# give 0 to those w/o provider
# give 1 to those who answered no
# and give 2 to those who answered yes
h209red <- h209red %>% mutate(usc_asks_abt_trts_bin = ifelse(usc_asks_abt_trts == 1, 2,
                                                ifelse(usc_asks_abt_trts == 2, 1,
                                                       usc_asks_abt_trts)))

# give 0 to those w/o provider
# and give 1 to those who answered never or sometimes
# give 2 to those who answered usually or always
h209red <- h209red %>% mutate(usc_asks_hlp_dec_bin = ifelse(usc_asks_hlp_dec %in% c(1, 2), 1,
                                               ifelse(usc_asks_hlp_dec %in% c(3, 4), 2,
                                                      usc_asks_hlp_dec)))

# give 0 to those w/o provider
# and give 1 to people who answer no
# give 2 to those who answered yes
h209red <- h209red %>% mutate(usc_expln_options_bin = ifelse(usc_expln_options == 1, 2,
                                                ifelse(usc_expln_options == 2, 1,
                                                       usc_expln_options)))

# finally creating combined satisfaction score from binary variables
```

```
h209red <- h209red %>%
  mutate(usc_satisf_score = usc_asks_abt_trts_bin + usc_asks_hlp_dec_bin + usc_expln_options_bin)
```

```
# creating our data set of usable variables
df <- h209red %>% select(pap_f, age, income_indiv, income_fam, totexp,
                         outofpocket_exp, menhlth_avg_f, genhlth_avg_f,
                         region_f, race_f, marital_stat_f, educ_f, smoke_freq_f,
                         limitation_f, afford_care_f, have_usc_f, usc_access_score,
                         usc_satisf_score, inscov_gen_2018_f)
# note: didn't include the seperate access and satisfaction variables since we made the combined score
```

Bar plots demographics

Models

```
# cost related
m1 <- glm(formula = pap_f ~ afford_care_f + inscov_gen_2018_f + income_indiv + income_fam, family = bin
summary(m1)
```

```
##
## Call:
## glm(formula = pap_f ~ afford_care_f + inscov_gen_2018_f + income_indiv +
##     income_fam, family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2302  -0.8017  -0.6608   1.1405   2.7540
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -8.230e-01  6.764e-02 -12.167  < 2e-16 ***
## afford_care_fyes         -2.428e-01  1.103e-01  -2.201 0.027735 *
## inscov_gen_2018_fpublic only  2.850e-01  7.874e-02   3.620 0.000295 ***
## inscov_gen_2018_funinsured   9.464e-01  9.590e-02   9.868  < 2e-16 ***
## income_indiv             -1.089e-05  1.339e-06  -8.131 4.27e-16 ***
## income_fam               -1.006e-06  6.556e-07  -1.534 0.124936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6880.2  on 6026  degrees of freedom
## Residual deviance: 6540.7  on 6021  degrees of freedom
##   (609 observations deleted due to missingness)
## AIC: 6552.7
##
## Number of Fisher Scoring iterations: 4
```

```
# accessibility related
m2 <- glm(formula = pap_f ~ limitation_f + have_usc_f + usc_access_score, family = binomial(), data = d
summary(m2)
```

```
##
## Call:
## glm(formula = pap_f ~ limitation_f + have_usc_f + usc_access_score,
```

9

```
##       family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1767  -0.6962  -0.6755   1.4189   1.8432
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.55175    0.05145 -10.725  < 2e-16 ***
## limitation_fyes    0.55018    0.11186   4.918 8.73e-07 ***
## have_usc_fyes     -1.14806    0.30285  -3.791  0.00015 ***
## usc_access_score   0.06769    0.05797   1.168  0.24291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6085.1  on 5279  degrees of freedom
## Residual deviance: 5927.6  on 5276  degrees of freedom
##   (1356 observations deleted due to missingness)
## AIC: 5935.6
##
## Number of Fisher Scoring iterations: 4
```

```
# provider characteristics
m3 <- glm(formula = pap_f ~ usc_satisf_score, family = binomial(), data = df)
summary(m3)
```

```
##
## Call:
## glm(formula = pap_f ~ usc_satisf_score, family = binomial(),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9687  -0.7092  -0.6640   1.4015   1.8001
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.51289    0.05027  -10.20   <2e-16 ***
## usc_satisf_score  -0.14780    0.01148  -12.87   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6326.1  on 5533  degrees of freedom
## Residual deviance: 6161.8  on 5532  degrees of freedom
##   (1102 observations deleted due to missingness)
## AIC: 6165.8
##
## Number of Fisher Scoring iterations: 4
```

```
#demographics
m4 <- glm(formula = pap_f ~ region_f, family = binomial(), data = df)
```

```r
summary(m4)
```

```
##
## Call:
## glm(formula = pap_f ~ region_f, family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8344  -0.8344  -0.7365   1.5648   1.7714
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.16629    0.07739 -15.071   <2e-16 ***
## region_fmidwest  -0.16913    0.10442  -1.620   0.1053
## region_fsouth     0.29010    0.08956   3.239   0.0012 **
## region_fwest      0.10086    0.09723   1.037   0.2996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6903.4  on 6040  degrees of freedom
## Residual deviance: 6869.0  on 6037  degrees of freedom
##   (595 observations deleted due to missingness)
## AIC: 6877
##
## Number of Fisher Scoring iterations: 4
```

```r
m4 <- glm(formula = pap_f ~ race_f, family = binomial(), data = df)
summary(m4)
```

```
##
## Call:
## glm(formula = pap_f ~ race_f, family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0625  -0.8590  -0.6561   1.2969   1.8120
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.42643    0.04600 -31.007  < 2e-16 ***
## race_fhispanic               0.64601    0.07193   8.982  < 2e-16 ***
## race_fblack                  0.61940    0.08469   7.313  2.6e-13 ***
## race_fasian                  1.14996    0.11540   9.965  < 2e-16 ***
## race_fother or multiple races 0.41649   0.16408   2.538   0.0111 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6903.4  on 6040  degrees of freedom
## Residual deviance: 6746.6  on 6036  degrees of freedom
##   (595 observations deleted due to missingness)
```

```
## AIC: 6756.6
##
## Number of Fisher Scoring iterations: 4
```

```r
m4 <- glm(formula = pap_f ~  age, family = binomial(), data = df)
summary(m4)
```

```
##
## Call:
## glm(formula = pap_f ~ age, family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7891  -0.7786  -0.7681   1.6238   1.6681
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.960151   0.102100  -9.404   <2e-16 ***
## age         -0.002233   0.002324  -0.961    0.337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6903.4  on 6040  degrees of freedom
## Residual deviance: 6902.5  on 6039  degrees of freedom
##   (595 observations deleted due to missingness)
## AIC: 6906.5
##
## Number of Fisher Scoring iterations: 4
```

```r
m4 <- glm(formula = pap_f ~ marital_stat_f, family = binomial(), data = df)
summary(m4)
```

```
##
## Call:
## glm(formula = pap_f ~ marital_stat_f, family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9320  -0.7298  -0.6739   1.4445   1.7854
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -0.60901    0.04943 -12.322  < 2e-16 ***
## marital_stat_fmarried  -0.75769    0.06627 -11.434  < 2e-16 ***
## marital_stat_fwidowed  -0.04229    0.16548  -0.256    0.798
## marital_stat_fdivorced -0.57816    0.10246  -5.643 1.68e-08 ***
## marital_stat_fseperated -0.23367   0.15974  -1.463    0.144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6903.4  on 6040  degrees of freedom
```

```
## Residual deviance: 6762.7  on 6036  degrees of freedom
##    (595 observations deleted due to missingness)
## AIC: 6772.7
##
## Number of Fisher Scoring iterations: 4
```

```r
m4 <- glm(formula = pap_f ~ educ_f, family = binomial(), data = df)
summary(m4)
```

```
##
## Call:
## glm(formula = pap_f ~ educ_f, family = binomial(), data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9650  -0.9209  -0.6594   1.4058   1.8070
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.41516    0.04214 -33.584  < 2e-16 ***
## educ_fany high school        0.77672    0.06193  12.542  < 2e-16 ***
## educ_fnone or any elementary 0.89264    0.13188   6.768  1.3e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6844.8  on 6002  degrees of freedom
## Residual deviance: 6669.0  on 6000  degrees of freedom
##    (633 observations deleted due to missingness)
## AIC: 6675
##
## Number of Fisher Scoring iterations: 4
```

Forward and backward selection

```r
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(leaps)
library(LogisticDx)
library(ResourceSelection)
```

```
## ResourceSelection 0.3-5   2019-07-22
```

```r
cc_df <- df %>% drop_na()

mod_forw <- step(glm(pap_f ~ 1, data = cc_df, family = binomial()), ~ age + income_indiv + income_fam +
```

```
## Start:  AIC=5731.12
## pap_f ~ 1
##
##                      Df Deviance    AIC
## + income_indiv        1   5525.2 5529.2
## + inscov_gen_2018_f   2   5556.4 5562.4
## + usc_satisf_score    1   5577.1 5581.1
## + educ_f              2   5591.3 5597.3
## + have_usc_f          1   5594.9 5598.9
## + usc_access_score    1   5605.8 5609.8
## + income_fam          1   5611.2 5615.2
## + race_f              4   5610.1 5620.1
## + marital_stat_f      4   5610.5 5620.5
## + outofpocket_exp     1   5682.3 5686.3
## + smoke_freq_f        2   5692.3 5698.3
## + totexp              1   5699.6 5703.6
## + region_f            3   5705.6 5713.6
## + limitation_f        1   5718.4 5722.4
## + genhlth_avg_f       4   5712.6 5722.6
## + menhlth_avg_f       4   5716.9 5726.9
## <none>                    5729.1 5731.1
## + age                 1   5728.0 5732.0
## + afford_care_f       1   5729.0 5733.0
##
## Step:  AIC=5529.19
## pap_f ~ income_indiv
##
##                      Df Deviance    AIC
## + usc_satisf_score    1   5407.3 5413.3
## + have_usc_f          1   5418.9 5424.9
## + usc_access_score    1   5429.6 5435.6
## + race_f              4   5436.4 5448.4
## + marital_stat_f      4   5442.0 5454.0
## + inscov_gen_2018_f   2   5448.3 5456.3
## + educ_f              2   5470.5 5478.5
## + totexp              1   5500.5 5506.5
## + outofpocket_exp     1   5504.9 5510.9
## + smoke_freq_f        2   5506.7 5514.7
## + region_f            3   5507.3 5517.3
## + income_fam          1   5517.9 5523.9
## <none>                    5525.2 5529.2
## + age                 1   5524.3 5530.3
## + limitation_f        1   5524.3 5530.3
## + afford_care_f       1   5524.6 5530.6
## + genhlth_avg_f       4   5521.1 5533.1
## + menhlth_avg_f       4   5523.3 5535.3
##
## Step:  AIC=5413.34
## pap_f ~ income_indiv + usc_satisf_score
##
##                      Df Deviance    AIC
## + race_f              4   5338.4 5352.4
## + marital_stat_f      4   5339.5 5353.5
## + educ_f              2   5355.8 5365.8
```

```
## + inscov_gen_2018_f  2    5359.8 5369.8
## + smoke_freq_f        2    5387.5 5397.5
## + totexp              1    5397.4 5405.4
## + age                 1    5398.3 5406.3
## + outofpocket_exp     1    5399.1 5407.1
## + limitation_f        1    5399.6 5407.6
## + usc_access_score    1    5401.7 5409.7
## + region_f            3    5398.3 5410.3
## + income_fam          1    5403.9 5411.9
## + have_usc_f          1    5405.1 5413.1
## + afford_care_f       1    5405.2 5413.2
## <none>                     5407.3 5413.3
## + genhlth_avg_f       4    5400.0 5414.0
## + menhlth_avg_f       4    5404.1 5418.1
##
## Step:  AIC=5352.4
## pap_f ~ income_indiv + usc_satisf_score + race_f
##
##                      Df Deviance    AIC
## + marital_stat_f      4    5272.1 5294.1
## + educ_f             2    5289.2 5307.2
## + inscov_gen_2018_f  2    5295.5 5313.5
## + smoke_freq_f       2    5305.4 5323.4
## + age                1    5327.2 5343.2
## + limitation_f       1    5328.2 5344.2
## + totexp             1    5332.6 5348.6
## + income_fam         1    5333.1 5349.1
## + outofpocket_exp    1    5334.4 5350.4
## + usc_access_score   1    5334.6 5350.6
## + region_f           3    5331.7 5351.7
## <none>                    5338.4 5352.4
## + have_usc_f         1    5337.2 5353.2
## + afford_care_f      1    5337.2 5353.2
## + genhlth_avg_f      4    5331.9 5353.9
## + menhlth_avg_f      4    5333.5 5355.5
##
## Step:  AIC=5294.1
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f
##
##                      Df Deviance    AIC
## + educ_f             2    5224.2 5250.2
## + age                1    5238.7 5262.7
## + inscov_gen_2018_f  2    5238.2 5264.2
## + smoke_freq_f       2    5248.4 5274.4
## + limitation_f       1    5264.6 5288.6
## + totexp             1    5265.9 5289.9
## + region_f           3    5263.5 5291.5
## + usc_access_score   1    5268.4 5292.4
## + outofpocket_exp    1    5269.8 5293.8
## + afford_care_f      1    5270.0 5294.0
## <none>                    5272.1 5294.1
## + have_usc_f         1    5271.2 5295.2
## + income_fam         1    5272.0 5296.0
## + genhlth_avg_f      4    5267.4 5297.4
```

```
## + menhlth_avg_f       4    5269.5 5299.5
##
## Step:  AIC=5250.18
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f
##
##                     Df Deviance    AIC
## + age                1    5197.2 5225.2
## + inscov_gen_2018_f  2    5197.0 5227.0
## + smoke_freq_f       2    5209.0 5239.0
## + limitation_f       1    5218.0 5246.0
## + totexp             1    5218.2 5246.2
## + region_f           3    5216.1 5248.1
## + usc_access_score   1    5221.6 5249.6
## + afford_care_f      1    5222.0 5250.0
## <none>                    5224.2 5250.2
## + income_fam         1    5222.5 5250.5
## + outofpocket_exp    1    5222.6 5250.6
## + have_usc_f         1    5223.9 5251.9
## + genhlth_avg_f      4    5220.9 5254.9
## + menhlth_avg_f      4    5222.5 5256.5
##
## Step:  AIC=5225.19
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age
##
##                     Df Deviance    AIC
## + inscov_gen_2018_f  2    5169.1 5201.1
## + smoke_freq_f       2    5183.7 5215.7
## + totexp             1    5187.7 5217.7
## + region_f           3    5188.9 5222.9
## + outofpocket_exp    1    5194.1 5224.1
## + usc_access_score   1    5194.2 5224.2
## + afford_care_f      1    5194.3 5224.3
## + limitation_f       1    5194.8 5224.8
## + income_fam         1    5195.1 5225.1
## <none>                    5197.2 5225.2
## + have_usc_f         1    5197.1 5227.1
## + genhlth_avg_f      4    5193.7 5229.7
## + menhlth_avg_f      4    5195.3 5231.3
##
## Step:  AIC=5201.12
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f
##
##                     Df Deviance    AIC
## + smoke_freq_f       2    5155.4 5191.4
## + totexp             1    5162.4 5196.4
## + afford_care_f      1    5163.0 5197.0
## + usc_access_score   1    5165.2 5199.2
## + limitation_f       1    5165.4 5199.4
## + outofpocket_exp    1    5165.5 5199.5
## + income_fam         1    5166.3 5200.3
## <none>                    5169.1 5201.1
```

```
## + region_f          3   5164.4 5202.4
## + have_usc_f         1   5168.9 5202.9
## + genhlth_avg_f      4   5165.8 5205.8
## + menhlth_avg_f      4   5166.8 5206.8
##
## Step:  AIC=5191.39
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f + smoke_freq_f
##
##                    Df Deviance    AIC
## + afford_care_f     1   5148.2 5186.2
## + totexp            1   5148.2 5186.2
## + usc_access_score  1   5151.3 5189.3
## + income_fam        1   5151.8 5189.8
## + outofpocket_exp   1   5151.8 5189.8
## + limitation_f      1   5152.4 5190.4
## <none>                 5155.4 5191.4
## + region_f          3   5150.7 5192.7
## + have_usc_f        1   5155.2 5193.2
## + genhlth_avg_f     4   5151.6 5195.6
## + menhlth_avg_f     4   5153.1 5197.1
##
## Step:  AIC=5186.2
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f + smoke_freq_f + afford_care_f
##
##                    Df Deviance    AIC
## + totexp            1   5141.3 5181.3
## + usc_access_score  1   5144.2 5184.2
## + limitation_f      1   5144.6 5184.6
## + income_fam        1   5145.2 5185.2
## + outofpocket_exp   1   5145.3 5185.3
## <none>                 5148.2 5186.2
## + region_f          3   5143.3 5187.3
## + have_usc_f        1   5148.0 5188.0
## + genhlth_avg_f     4   5144.7 5190.7
## + menhlth_avg_f     4   5145.9 5191.9
##
## Step:  AIC=5181.3
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f + smoke_freq_f + afford_care_f +
##     totexp
##
##                    Df Deviance    AIC
## + limitation_f      1   5134.5 5176.5
## + usc_access_score  1   5137.7 5179.7
## + income_fam        1   5138.2 5180.2
## <none>                 5141.3 5181.3
## + outofpocket_exp   1   5140.5 5182.5
## + region_f          3   5136.5 5182.5
## + have_usc_f        1   5141.1 5183.1
## + genhlth_avg_f     4   5137.1 5185.1
## + menhlth_avg_f     4   5138.8 5186.8
##
```

```
## Step:  AIC=5176.45
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f + smoke_freq_f + afford_care_f +
##     totexp + limitation_f
##
##                     Df Deviance    AIC
## + usc_access_score   1   5130.6 5174.6
## + income_fam         1   5131.5 5175.5
## <none>                   5134.5 5176.5
## + outofpocket_exp    1   5133.8 5177.8
## + region_f           3   5129.9 5177.9
## + have_usc_f         1   5134.2 5178.2
## + genhlth_avg_f      4   5131.1 5181.1
## + menhlth_avg_f      4   5132.7 5182.7
##
## Step:  AIC=5174.57
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f + smoke_freq_f + afford_care_f +
##     totexp + limitation_f + usc_access_score
##
##                     Df Deviance    AIC
## + income_fam         1   5127.5 5173.5
## <none>                   5130.6 5174.6
## + have_usc_f         1   5129.5 5175.5
## + region_f           3   5125.8 5175.8
## + outofpocket_exp    1   5129.9 5175.9
## + genhlth_avg_f      4   5127.2 5179.2
## + menhlth_avg_f      4   5128.9 5180.9
##
## Step:  AIC=5173.54
## pap_f ~ income_indiv + usc_satisf_score + race_f + marital_stat_f +
##     educ_f + age + inscov_gen_2018_f + smoke_freq_f + afford_care_f +
##     totexp + limitation_f + usc_access_score + income_fam
##
##                     Df Deviance    AIC
## <none>                   5127.5 5173.5
## + have_usc_f         1   5126.6 5174.6
## + region_f           3   5122.6 5174.6
## + outofpocket_exp    1   5126.8 5174.8
## + genhlth_avg_f      4   5124.6 5178.6
## + menhlth_avg_f      4   5125.9 5179.9
```

```r
summary(mod_forw)
```

```
##
## Call:
## glm(formula = pap_f ~ income_indiv + usc_satisf_score + race_f +
##     marital_stat_f + educ_f + age + inscov_gen_2018_f + smoke_freq_f +
##     afford_care_f + totexp + limitation_f + usc_access_score +
##     income_fam, family = binomial(), data = cc_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7407  -0.7832  -0.5617   0.9481   2.8781
##
```

```
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -9.898e-01  1.494e-01  -6.624 3.50e-11 ***
## income_indiv                  -1.217e-05  1.593e-06  -7.635 2.25e-14 ***
## usc_satisf_score              -2.057e-01  4.580e-02  -4.491 7.08e-06 ***
## race_fhispanic                 1.912e-01  9.191e-02   2.080  0.03748 *
## race_fblack                    8.373e-02  1.031e-01   0.812  0.41675
## race_fasian                    1.174e+00  1.421e-01   8.266  < 2e-16 ***
## race_fother or multiple races  4.397e-02  1.942e-01   0.226  0.82092
## marital_stat_fmarried         -7.718e-01  9.170e-02  -8.417  < 2e-16 ***
## marital_stat_fwidowed         -1.507e-01  2.051e-01  -0.735  0.46238
## marital_stat_fdivorced        -5.037e-01  1.278e-01  -3.941 8.11e-05 ***
## marital_stat_fseperated       -6.055e-01  1.912e-01  -3.167  0.00154 **
## educ_fany high school          4.210e-01  7.748e-02   5.433 5.53e-08 ***
## educ_fnone or any elementary   2.328e-01  1.634e-01   1.425  0.15423
## age                            1.646e-02  3.236e-03   5.085 3.67e-07 ***
## inscov_gen_2018_fpublic only  -4.238e-02  9.472e-02  -0.447  0.65456
## inscov_gen_2018_funinsured     5.881e-01  1.137e-01   5.173 2.30e-07 ***
## smoke_freq_fsome days          3.238e-01  1.549e-01   2.090  0.03665 *
## smoke_freq_fevery day          3.861e-01  1.095e-01   3.527  0.00042 ***
## afford_care_fyes              -3.262e-01  1.247e-01  -2.616  0.00889 **
## totexp                        -9.526e-06  3.284e-06  -2.901  0.00372 **
## limitation_fyes                3.535e-01  1.328e-01   2.661  0.00779 **
## usc_access_score               9.712e-02  4.882e-02   1.989  0.04666 *
## income_fam                     1.368e-06  7.793e-07   1.755  0.07929 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5729.1  on 4976  degrees of freedom
## Residual deviance: 5127.5  on 4954  degrees of freedom
## AIC: 5173.5
##
## Number of Fisher Scoring iterations: 5
```

```r
full_mod <- glm(pap_f ~ age + income_indiv + income_fam + totexp + outofpocket_exp + menhlth_avg_f + ge
summary(full_mod)
```

```
##
## Call:
## glm(formula = pap_f ~ age + income_indiv + income_fam + totexp +
##     outofpocket_exp + menhlth_avg_f + genhlth_avg_f + region_f +
##     race_f + marital_stat_f + educ_f + smoke_freq_f + limitation_f +
##     afford_care_f + have_usc_f + usc_access_score + usc_satisf_score +
##     inscov_gen_2018_f, family = binomial(), data = cc_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7150  -0.7807  -0.5576   0.9373   2.8742
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -1.178e+00  4.552e-01  -2.589 0.009621 **
## age                            1.717e-02  3.287e-03   5.224 1.75e-07 ***
```

```
## income_indiv                     -1.207e-05  1.597e-06  -7.562 3.98e-14 ***
## income_fam                        1.339e-06  7.877e-07   1.699 0.089251 .
## totexp                           -8.289e-06  3.502e-06  -2.367 0.017953 *
## outofpocket_exp                  -2.142e-05  2.561e-05  -0.836 0.403068
## menhlth_avg_ffair                 3.066e-01  4.000e-01   0.766 0.443416
## menhlth_avg_fgood                 1.334e-01  3.937e-01   0.339 0.734654
## menhlth_avg_fvery good            2.130e-01  3.989e-01   0.534 0.593306
## menhlth_avg_fexcellent            1.060e-01  4.071e-01   0.260 0.794671
## genhlth_avg_ffair                 4.761e-02  3.003e-01   0.159 0.874019
## genhlth_avg_fgood                 3.553e-02  3.018e-01   0.118 0.906292
## genhlth_avg_fvery good           -5.268e-02  3.117e-01  -0.169 0.865794
## genhlth_avg_fexcellent            1.610e-01  3.267e-01   0.493 0.622205
## region_fmidwest                  -1.331e-01  1.235e-01  -1.078 0.281134
## region_fsouth                     7.997e-02  1.088e-01   0.735 0.462206
## region_fwest                     -6.199e-02  1.177e-01  -0.527 0.598354
## race_fhispanic                    1.793e-01  9.531e-02   1.881 0.059932 .
## race_fblack                       2.036e-02  1.078e-01   0.189 0.850177
## race_fasian                       1.181e+00  1.438e-01   8.215  < 2e-16 ***
## race_fother or multiple races     5.212e-02  1.949e-01   0.267 0.789123
## marital_stat_fmarried            -7.789e-01  9.220e-02  -8.448  < 2e-16 ***
## marital_stat_fwidowed            -1.584e-01  2.063e-01  -0.768 0.442634
## marital_stat_fdivorced           -5.182e-01  1.284e-01  -4.035 5.45e-05 ***
## marital_stat_fseperated          -6.223e-01  1.926e-01  -3.231 0.001236 **
## educ_fany high school             4.208e-01  7.810e-02   5.388 7.14e-08 ***
## educ_fnone or any elementary      2.179e-01  1.646e-01   1.324 0.185358
## smoke_freq_fsome days             3.258e-01  1.553e-01   2.098 0.035927 *
## smoke_freq_fevery day             3.860e-01  1.106e-01   3.490 0.000483 ***
## limitation_fyes                   3.306e-01  1.422e-01   2.324 0.020106 *
## afford_care_fyes                 -3.274e-01  1.276e-01  -2.566 0.010293 *
## have_usc_fyes                    -4.864e-01  4.411e-01  -1.103 0.270187
## usc_access_score                  1.457e-01  6.452e-02   2.258 0.023929 *
## usc_satisf_score                 -1.615e-01  6.107e-02  -2.645 0.008175 **
## inscov_gen_2018_fpublic only     -3.898e-02  9.621e-02  -0.405 0.685320
## inscov_gen_2018_funinsured        5.656e-01  1.148e-01   4.925 8.42e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5729.1  on 4976  degrees of freedom
## Residual deviance: 5115.0  on 4941  degrees of freedom
## AIC: 5187
##
## Number of Fisher Scoring iterations: 5
mod_back <- step(full_mod, direction = "backward")

## Start:  AIC=5187
## pap_f ~ age + income_indiv + income_fam + totexp + outofpocket_exp +
##     menhlth_avg_f + genhlth_avg_f + region_f + race_f + marital_stat_f +
##     educ_f + smoke_freq_f + limitation_f + afford_care_f + have_usc_f +
##     usc_access_score + usc_satisf_score + inscov_gen_2018_f
##
##                   Df Deviance    AIC
## - menhlth_avg_f    4   5117.7 5181.7
```

```
## - genhlth_avg_f       4    5119.0 5183.0
## - outofpocket_exp     1    5115.7 5185.7
## - have_usc_f          1    5116.2 5186.2
## - region_f            3    5120.3 5186.3
## <none>                     5115.0 5187.0
## - income_fam          1    5117.8 5187.8
## - usc_access_score    1    5120.1 5190.1
## - limitation_f        1    5120.3 5190.3
## - totexp              1    5121.4 5191.4
## - afford_care_f       1    5121.8 5191.8
## - usc_satisf_score    1    5121.9 5191.9
## - smoke_freq_f        2    5129.7 5197.7
## - inscov_gen_2018_f   2    5143.4 5211.4
## - educ_f              2    5144.1 5212.1
## - age                 1    5142.4 5212.4
## - race_f              4    5181.4 5245.4
## - income_indiv        1    5176.2 5246.2
## - marital_stat_f      4    5192.8 5256.8
##
## Step:  AIC=5181.69
## pap_f ~ age + income_indiv + income_fam + totexp + outofpocket_exp +
##     genhlth_avg_f + region_f + race_f + marital_stat_f + educ_f +
##     smoke_freq_f + limitation_f + afford_care_f + have_usc_f +
##     usc_access_score + usc_satisf_score + inscov_gen_2018_f
##
##                      Df Deviance    AIC
## - genhlth_avg_f       4    5120.7 5176.7
## - outofpocket_exp     1    5118.5 5180.5
## - region_f            3    5122.7 5180.7
## - have_usc_f          1    5118.9 5180.9
## <none>                     5117.7 5181.7
## - income_fam          1    5120.4 5182.4
## - usc_access_score    1    5122.9 5184.9
## - limitation_f        1    5123.4 5185.4
## - totexp              1    5124.3 5186.3
## - afford_care_f       1    5124.4 5186.4
## - usc_satisf_score    1    5124.7 5186.7
## - smoke_freq_f        2    5132.7 5192.7
## - inscov_gen_2018_f   2    5145.7 5205.7
## - age                 1    5144.8 5206.8
## - educ_f              2    5147.1 5207.1
## - race_f              4    5183.9 5239.9
## - income_indiv        1    5178.8 5240.8
## - marital_stat_f      4    5195.5 5251.5
##
## Step:  AIC=5176.7
## pap_f ~ age + income_indiv + income_fam + totexp + outofpocket_exp +
##     region_f + race_f + marital_stat_f + educ_f + smoke_freq_f +
##     limitation_f + afford_care_f + have_usc_f + usc_access_score +
##     usc_satisf_score + inscov_gen_2018_f
##
##                      Df Deviance    AIC
## - outofpocket_exp     1    5121.5 5175.5
## - region_f            3    5125.8 5175.8
```

```
## - have_usc_f          1   5121.8 5175.8
## <none>                    5120.7 5176.7
## - income_fam           1   5123.9 5177.9
## - usc_access_score     1   5125.8 5179.8
## - limitation_f         1   5127.4 5181.4
## - afford_care_f        1   5127.5 5181.5
## - totexp               1   5127.6 5181.6
## - usc_satisf_score     1   5127.8 5181.8
## - smoke_freq_f         2   5135.7 5187.7
## - inscov_gen_2018_f    2   5148.8 5200.8
## - age                  1   5147.6 5201.6
## - educ_f               2   5150.1 5202.1
## - race_f               4   5186.8 5234.8
## - income_indiv         1   5182.5 5236.5
## - marital_stat_f       4   5199.1 5247.1
##
## Step:  AIC=5175.5
## pap_f ~ age + income_indiv + income_fam + totexp + region_f +
##     race_f + marital_stat_f + educ_f + smoke_freq_f + limitation_f +
##     afford_care_f + have_usc_f + usc_access_score + usc_satisf_score +
##     inscov_gen_2018_f
##
##                       Df Deviance    AIC
## - region_f             3   5126.6 5174.6
## - have_usc_f           1   5122.6 5174.6
## <none>                    5121.5 5175.5
## - income_fam           1   5124.6 5176.6
## - usc_access_score     1   5126.5 5178.5
## - limitation_f         1   5128.3 5180.3
## - afford_care_f        1   5128.6 5180.6
## - usc_satisf_score     1   5128.8 5180.8
## - totexp               1   5131.0 5183.0
## - smoke_freq_f         2   5136.5 5186.5
## - inscov_gen_2018_f    2   5149.2 5199.2
## - age                  1   5148.0 5200.0
## - educ_f               2   5151.1 5201.1
## - race_f               4   5188.0 5234.0
## - income_indiv         1   5183.9 5235.9
## - marital_stat_f       4   5200.2 5246.2
##
## Step:  AIC=5174.56
## pap_f ~ age + income_indiv + income_fam + totexp + race_f + marital_stat_f +
##     educ_f + smoke_freq_f + limitation_f + afford_care_f + have_usc_f +
##     usc_access_score + usc_satisf_score + inscov_gen_2018_f
##
##                       Df Deviance    AIC
## - have_usc_f           1   5127.5 5173.5
## <none>                    5126.6 5174.6
## - income_fam           1   5129.5 5175.5
## - usc_access_score     1   5131.2 5177.2
## - afford_care_f        1   5133.5 5179.5
## - limitation_f         1   5133.6 5179.6
## - usc_satisf_score     1   5133.9 5179.9
## - totexp               1   5136.3 5182.3
```

```
## - smoke_freq_f      2   5141.7 5185.7
## - age                1   5153.0 5199.0
## - educ_f             2   5156.4 5200.4
## - inscov_gen_2018_f  2   5158.3 5202.3
## - race_f             4   5193.0 5233.0
## - income_indiv       1   5189.0 5235.0
## - marital_stat_f     4   5203.7 5243.7
##
## Step:  AIC=5173.54
## pap_f ~ age + income_indiv + income_fam + totexp + race_f + marital_stat_f +
##     educ_f + smoke_freq_f + limitation_f + afford_care_f + usc_access_score +
##     usc_satisf_score + inscov_gen_2018_f
##
##                       Df Deviance    AIC
## <none>                    5127.5 5173.5
## - income_fam           1  5130.6 5174.6
## - usc_access_score     1  5131.5 5175.5
## - limitation_f         1  5134.5 5178.5
## - afford_care_f        1  5134.6 5178.6
## - totexp               1  5137.3 5181.3
## - smoke_freq_f         2  5142.5 5184.5
## - usc_satisf_score     1  5147.8 5191.8
## - age                  1  5153.5 5197.5
## - educ_f               2  5157.0 5199.0
## - inscov_gen_2018_f    2  5159.5 5201.5
## - race_f               4  5193.8 5231.8
## - income_indiv         1  5190.0 5234.0
## - marital_stat_f       4  5204.6 5242.6
```

```
summary(mod_back)
```

```
##
## Call:
## glm(formula = pap_f ~ age + income_indiv + income_fam + totexp +
##     race_f + marital_stat_f + educ_f + smoke_freq_f + limitation_f +
##     afford_care_f + usc_access_score + usc_satisf_score + inscov_gen_2018_f,
##     family = binomial(), data = cc_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7407  -0.7832  -0.5617   0.9481   2.8781
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -9.898e-01  1.494e-01  -6.624 3.50e-11 ***
## age                         1.646e-02  3.236e-03   5.085 3.67e-07 ***
## income_indiv               -1.217e-05  1.593e-06  -7.635 2.25e-14 ***
## income_fam                  1.368e-06  7.793e-07   1.755  0.07929 .
## totexp                     -9.526e-06  3.284e-06  -2.901  0.00372 **
## race_fhispanic              1.912e-01  9.191e-02   2.080  0.03748 *
## race_fblack                 8.373e-02  1.031e-01   0.812  0.41675
## race_fasian                 1.174e+00  1.421e-01   8.266  < 2e-16 ***
## race_fother or multiple races  4.397e-02  1.942e-01   0.226  0.82092
## marital_stat_fmarried      -7.718e-01  9.170e-02  -8.417  < 2e-16 ***
## marital_stat_fwidowed      -1.507e-01  2.051e-01  -0.735  0.46238
```

```
## marital_stat_fdivorced       -5.037e-01  1.278e-01  -3.941 8.11e-05 ***
## marital_stat_fseperated      -6.055e-01  1.912e-01  -3.167  0.00154 **
## educ_fany high school         4.210e-01  7.748e-02   5.433 5.53e-08 ***
## educ_fnone or any elementary  2.328e-01  1.634e-01   1.425  0.15423
## smoke_freq_fsome days         3.238e-01  1.549e-01   2.090  0.03665 *
## smoke_freq_fevery day         3.861e-01  1.095e-01   3.527  0.00042 ***
## limitation_fyes               3.535e-01  1.328e-01   2.661  0.00779 **
## afford_care_fyes             -3.262e-01  1.247e-01  -2.616  0.00889 **
## usc_access_score              9.712e-02  4.882e-02   1.989  0.04666 *
## usc_satisf_score             -2.057e-01  4.580e-02  -4.491 7.08e-06 ***
## inscov_gen_2018_fpublic only -4.238e-02  9.472e-02  -0.447  0.65456
## inscov_gen_2018_funinsured    5.881e-01  1.137e-01   5.173 2.30e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5729.1  on 4976  degrees of freedom
## Residual deviance: 5127.5  on 4954  degrees of freedom
## AIC: 5173.5
##
## Number of Fisher Scoring iterations: 5
```

```
hoslem.test(as.numeric(cc_df$pap_f) - 1, fitted(mod_forw), g = 10)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  as.numeric(cc_df$pap_f) - 1, fitted(mod_forw)
## X-squared = 9.085, df = 8, p-value = 0.3352
```

```
hoslem.test(as.numeric(cc_df$pap_f) - 1, fitted(mod_back), g = 10)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  as.numeric(cc_df$pap_f) - 1, fitted(mod_back)
## X-squared = 9.085, df = 8, p-value = 0.3352
```

```
gof(mod_forw, g = 9)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
##       chiSq df pVal
## PrI       1  2    2
## drI       2  2    4
## PrG       1  1    1
## drG       2  1    3
## PrCT      1  1    1
## drCT      2  1    3
##             val df pVal
## HL chiSq      9  3    5
## mHL F         8  4    1
## OsRo Z        3  5    4
## SstPgeq0.5 Z  2  5    8
```
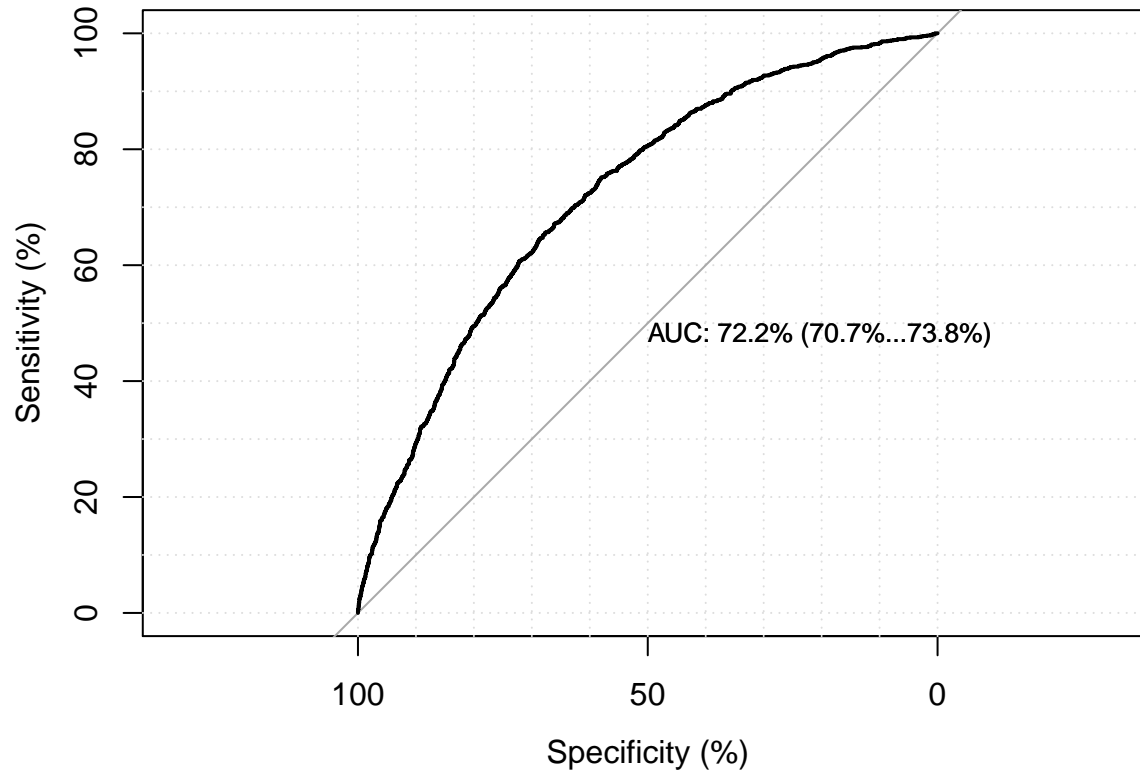
```
## SstPl0.5 Z          4   5    3
## SstBoth chiSq        6   2    7
## SllPgeq0.5 chiSq     1   1    9
## SllPl0.5 chiSq       5   1    2
## SllBoth chiSq        7   2    6
```

```r
gof(mod_back, g = 9)
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

Receiver Operating Curve


```
##        chiSq df pVal
## PrI       1  2    2
## drI       2  2    4
## PrG       1  1    1
## drG       2  1    3
## PrCT      1  1    1
## drCT      2  1    3
##                 val df pVal
## HL chiSq          9  3    5
## mHL F             8  4    1
## OsRo Z            3  5    4
## SstPgeq0.5 Z      2  5    8
## SstPl0.5 Z        4  5    3
## SstBoth chiSq     6  2    7
## SllPgeq0.5 chiSq  1  1    9
## SllPl0.5 chiSq    5  1    2
## SllBoth chiSq     7  2    6
```