EDA

The objective of the Exploratory Data Analysis (EDA) is to investigate the data so as to discover patterns, to spot anomalies, to handle missing values, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

Imports

```
library(rstudioapi)
library(tidyverse)
library(magrittr)
library(DescTools)
library(gridExtra)
```

Loading data

```
current_path = rstudioapi::getActiveDocumentContext()$path
setwd(dirname(current_path))

X_train = read_csv("../data/training_set_features.csv")
Y_train = read_csv("../data/training_set_labels.csv")

df_train = X_train %>%
  inner_join(Y_train, by='respondent_id')

dimensions = dim(df_train)
sprintf("The dataset contains %d rows and %d columns", dimensions[1], dimensions[2]) %>% cat()
```

The dataset contains 26707 rows and 38 columns

```
features = names(X_train)[-1]
targets = names(Y_train)[-1]
```

```
df_train %>% glimpse()
```

```
## Rows: 26,707
## Columns: 38
## $ respondent_id               <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 1...
## $ h1n1_concern                <dbl> 1, 3, 1, 1, 2, 3, 0, 1, 0, 2, 2, 1, 1, ...
## $ h1n1_knowledge              <dbl> 0, 2, 1, 1, 1, 1, 0, 0, 2, 1, 1, 2, 1, ...
## $ behavioral_antiviral_meds   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ behavioral_avoidance        <dbl> 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, ...
## $ behavioral_face_mask        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ behavioral_wash_hands       <dbl> 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, ...
## $ behavioral_large_gatherings <dbl> 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, ...
## $ behavioral_outside_home     <dbl> 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, ...
## $ behavioral_touch_face       <dbl> 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, ...
```

```
## $ doctor_recc_h1n1            <dbl> 0, 0, NA, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,...
## $ doctor_recc_seasonal        <dbl> 0, 0, NA, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1,...
## $ chronic_med_condition       <dbl> 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, ...
## $ child_under_6_months        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ...
## $ health_worker               <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ health_insurance            <dbl> 1, 1, NA, NA, NA, NA, NA, 1, NA, 1, 0, ...
## $ opinion_h1n1_vacc_effective <dbl> 3, 5, 3, 3, 3, 5, 4, 5, 4, 4, 4, 3, 3, ...
## $ opinion_h1n1_risk           <dbl> 1, 4, 1, 3, 3, 2, 1, 2, 1, 2, 1, 2, 2, ...
## $ opinion_h1n1_sick_from_vacc <dbl> 2, 4, 1, 5, 2, 1, 1, 1, 1, 2, 2, 2, 1, ...
## $ opinion_seas_vacc_effective <dbl> 2, 4, 4, 5, 3, 5, 4, 4, 4, 4, 5, 4, 5, ...
## $ opinion_seas_risk           <dbl> 1, 2, 1, 4, 1, 4, 2, 2, 2, 2, 4, 2, 4, ...
## $ opinion_seas_sick_from_vacc <dbl> 2, 4, 2, 1, 4, 4, 1, 1, 1, 2, 4, 1, 1, ...
## $ age_group                   <chr> "55 - 64 Years", "35 - 44 Years", "18 -...
## $ education                   <chr> "< 12 Years", "12 Years", "College Grad...
## $ race                        <chr> "White", "White", "White", "White", "Wh...
## $ sex                         <chr> "Female", "Male", "Male", "Female", "Fe...
## $ income_poverty              <chr> "Below Poverty", "Below Poverty", "<= $...
## $ marital_status              <chr> "Not Married", "Not Married", "Not Marr...
## $ rent_or_own                 <chr> "Own", "Rent", "Own", "Rent", "Own", "O...
## $ employment_status           <chr> "Not in Labor Force", "Employed", "Empl...
## $ hhs_geo_region              <chr> "oxchjgsf", "bhuqouqj", "qufhixun", "lr...
## $ census_msa                  <chr> "Non-MSA", "MSA, Not Principle  City", ...
## $ household_adults            <dbl> 0, 0, 2, 0, 1, 2, 0, 2, 1, 0, 2, 1, 1, ...
## $ household_children          <dbl> 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 2, 0, ...
## $ employment_industry         <chr> NA, "pxcmvdjn", "rucpziij", NA, "wxleye...
## $ employment_occupation       <chr> NA, "xgwztkwe", "xtkaffoo", NA, "emcorr...
## $ h1n1_vaccine                <dbl> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, ...
## $ seasonal_vaccine            <dbl> 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, ...
```

- All of the columns available are categorical
- Some of them are labeled as numeric values and others are still characters
- The features `hhs_geo_region`, `employment_industry`, `employment_occupation` are random character strings

Targets

Checking class balance.

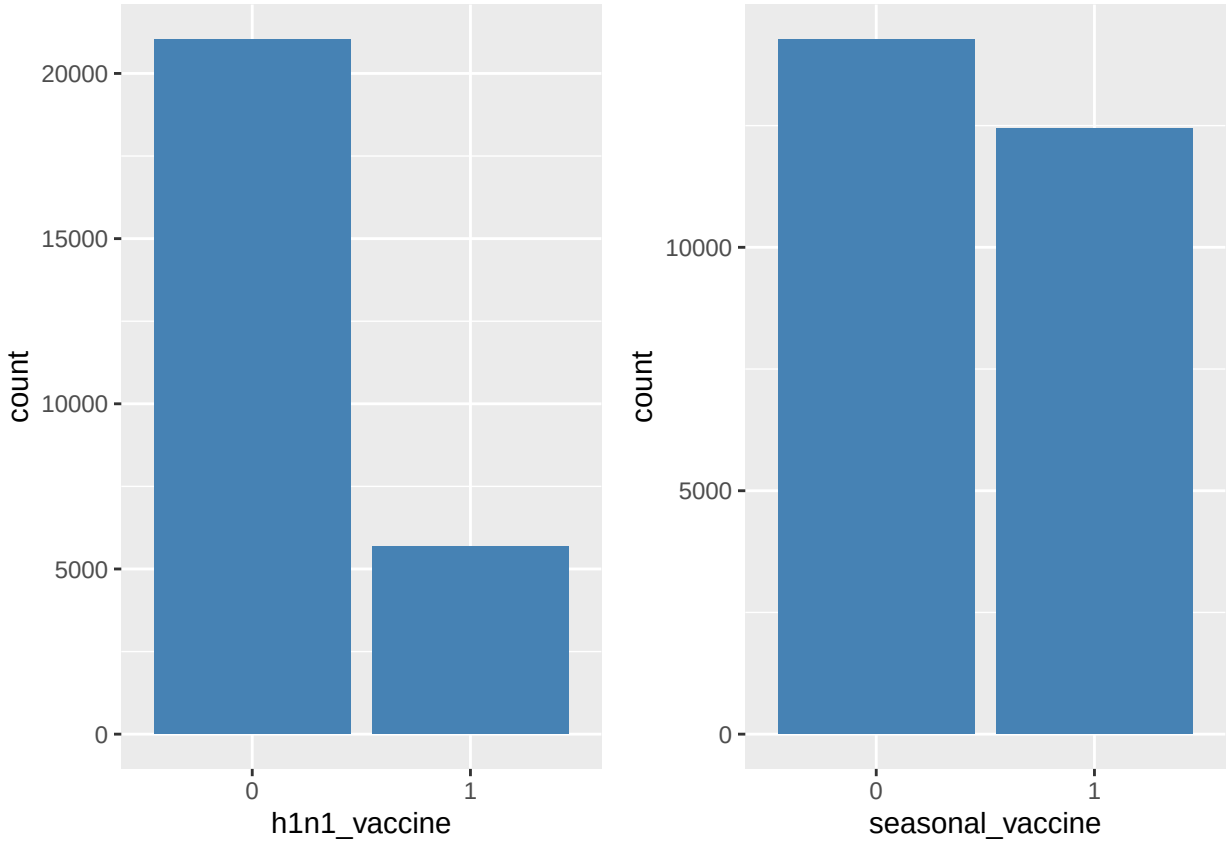
```
for(target in targets){
  percent = df_train[[target]] %>% mean()
  sprintf("- %.1f%% of the observations of the column %s are 1's \n", percent * 100, target) %>% cat()
}
```

- 21.2% of the observations of the column h1n1_vaccine are 1's
- 46.6% of the observations of the column seasonal_vaccine are 1's

The classes of the variable corresponding to Whether the respondent received H1N1 flu vaccine is imbalanced, much more people didn't received the H1N1 vaccine.

```
target = "h1n1_vaccine"

df_train = df_train %>%
  mutate_at(targets, factor)

plots = list()
i = 1
```

```
for(target in targets){
  p = df_train %>%
    ggplot(aes_string(x=target)) +
    geom_bar(fill="steelblue")
  plots[[i]] = p
  i = i + 1
}

do.call("grid.arrange", c(plots, ncol=2))
```



Features

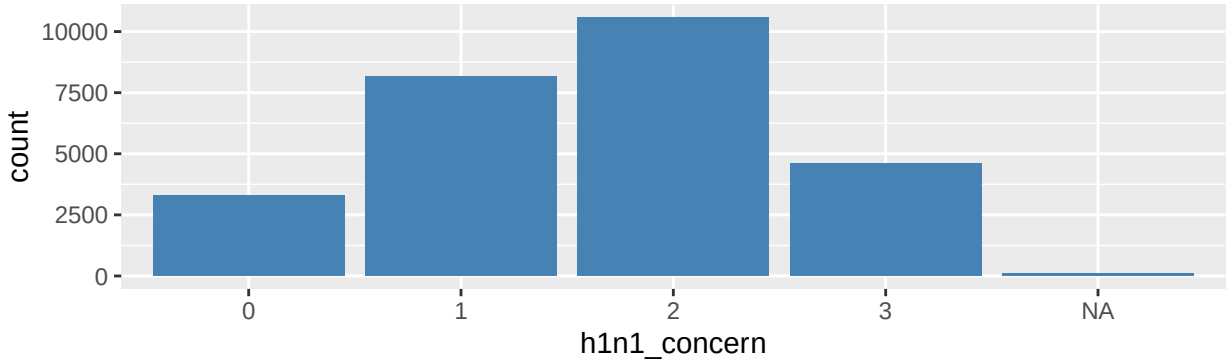
```
df_train %>%
  select_if(is.character)
```

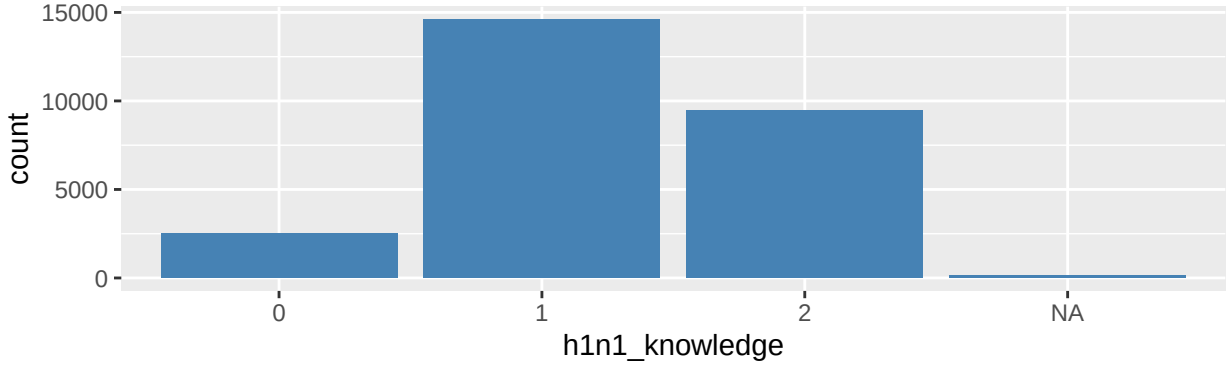
```
## # A tibble: 26,707 x 12
##    age_group education race  sex   income_poverty marital_status rent_or_own
##    <chr>     <chr>     <chr> <chr> <chr>          <chr>          <chr>
##  1 55 - 64 ~ < 12 Yea~ White Fema~ Below Poverty  Not Married    Own
##  2 35 - 44 ~ 12 Years  White Male  Below Poverty  Not Married    Rent
##  3 18 - 34 ~ College ~ White Male  <= $75,000, A~ Not Married    Own
##  4 65+ Years 12 Years  White Fema~ Below Poverty  Not Married    Rent
##  5 45 - 54 ~ Some Col~ White Fema~ <= $75,000, A~ Married        Own
##  6 65+ Years 12 Years  White Male  <= $75,000, A~ Married        Own
##  7 55 - 64 ~ < 12 Yea~ White Male  <= $75,000, A~ Not Married    Own
```

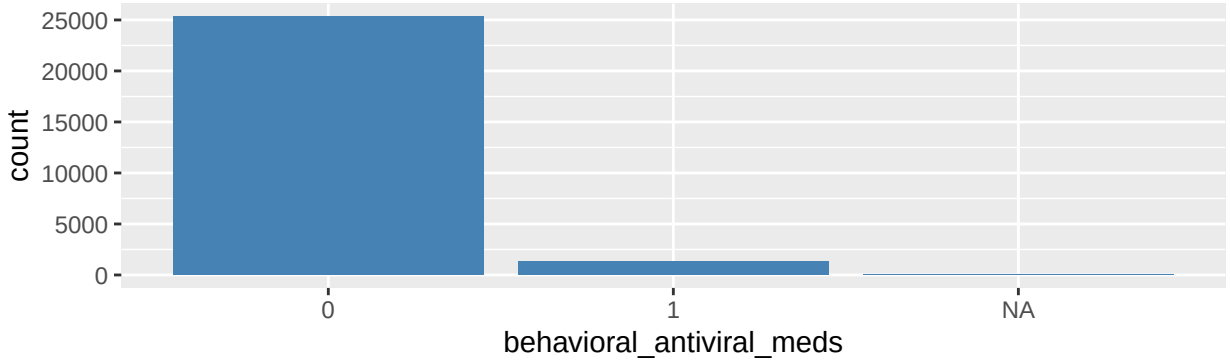
```
##  8 45 - 54 ~ Some Col~ White Fema~ <= $75,000, A~ Married        Own
##  9 45 - 54 ~ College ~ White Male  > $75,000      Married        Own
## 10 55 - 64 ~ 12 Years  White Male  <= $75,000, A~ Not Married    Own
## # ... with 26,697 more rows, and 5 more variables: employment_status <chr>,
## #   hhs_geo_region <chr>, census_msa <chr>, employment_industry <chr>,
## #   employment_occupation <chr>
```

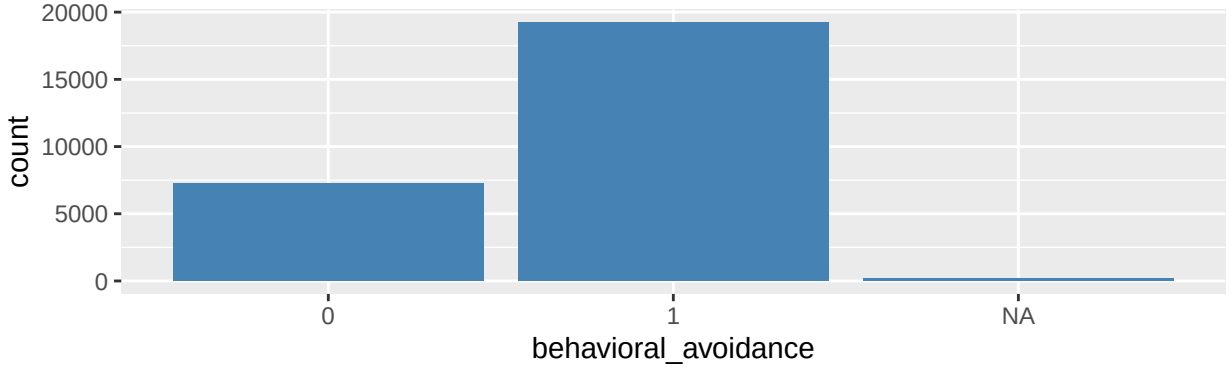
Distribution of each feature:

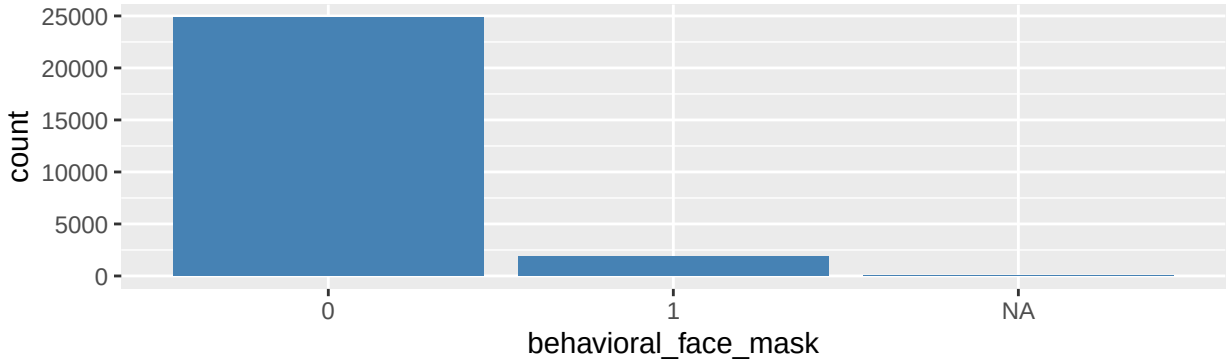
```
plots = list()
i = 1

for(feature in features){
  p = df_train %>%
    ggplot(aes_string(x=feature)) +
    geom_bar(fill="steelblue")

  plots[[i]] = p
  if(i %% 4 == 0){
    do.call("grid.arrange", c(plots[(i-3): i], nrow=4))
  }
  i = i + 1
}
```


count
10000
7500
5000
2500
0
0
1
2
3
NA
h1n1_concern

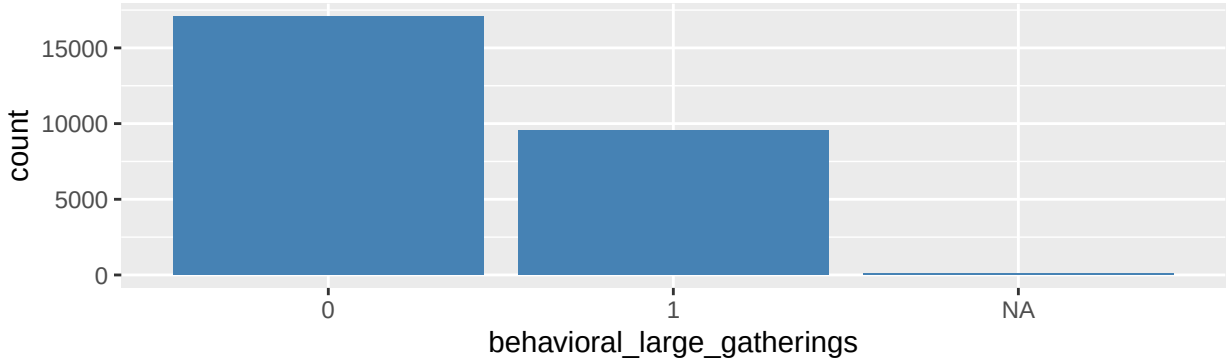

count
15000
10000
5000
0
0
1
2
NA
h1n1_knowledge

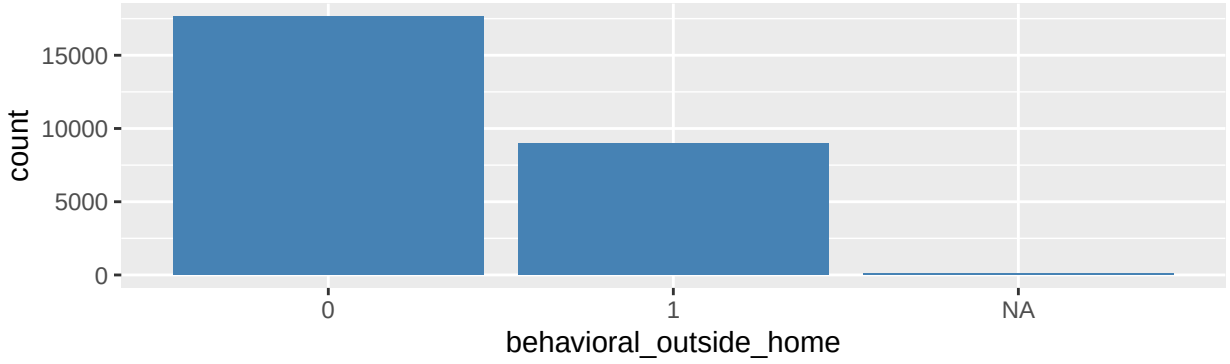

count
25000
20000
15000
10000
5000
0
0
1
NA
behavioral_antiviral_meds

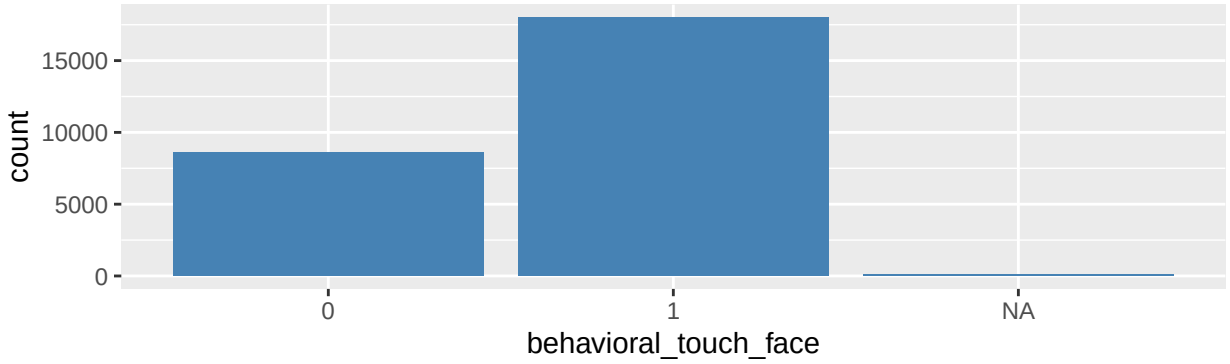

count
20000
15000
10000
5000
0
0
1
NA
behavioral_avoidance

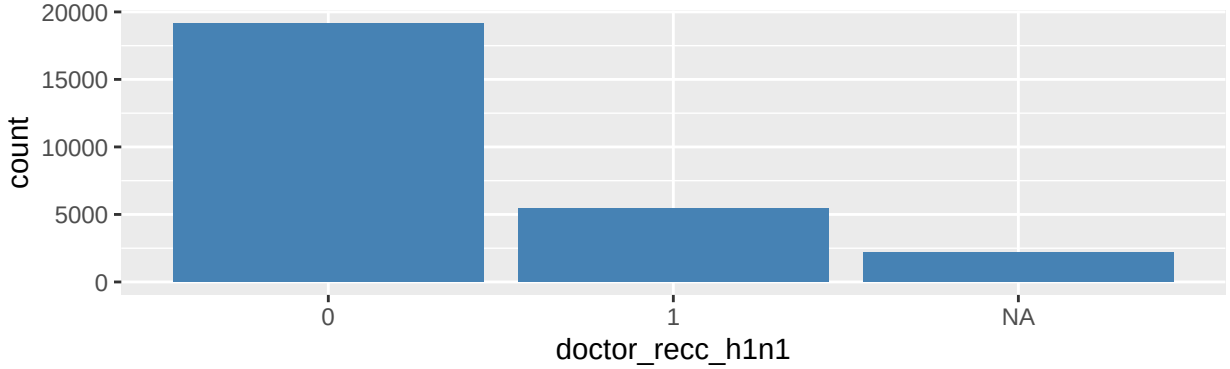

25000
20000
15000
10000
5000
0
count
0
1
NA
behavioral_face_mask

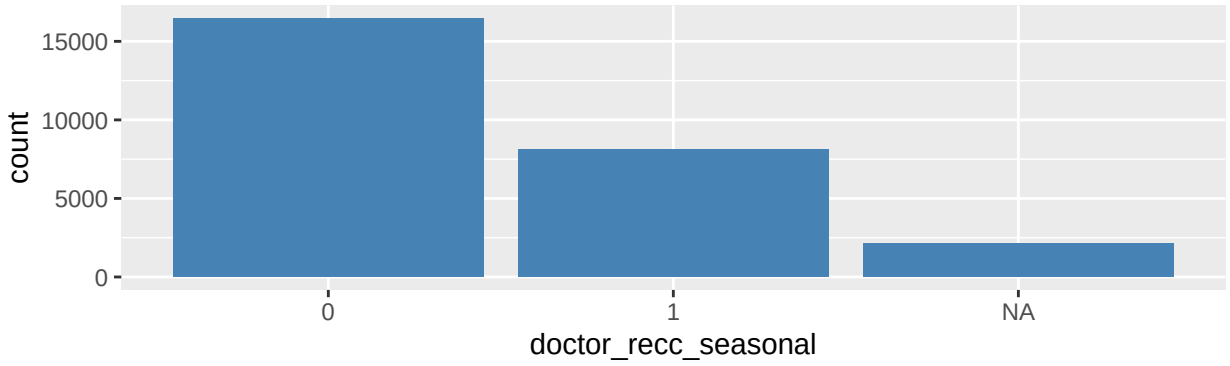

20000
15000
10000
5000
0
count
0
1
NA
behavioral_wash_hands

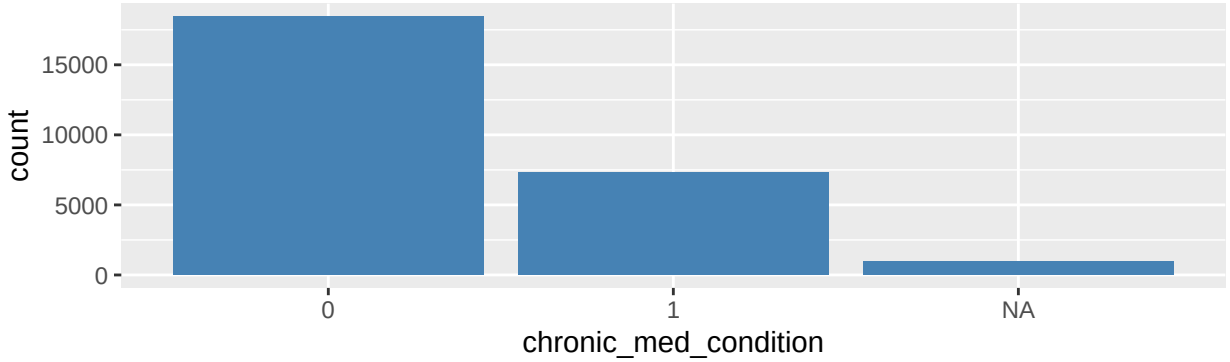

15000
10000
5000
0
count
0
1
NA
behavioral_large_gatherings

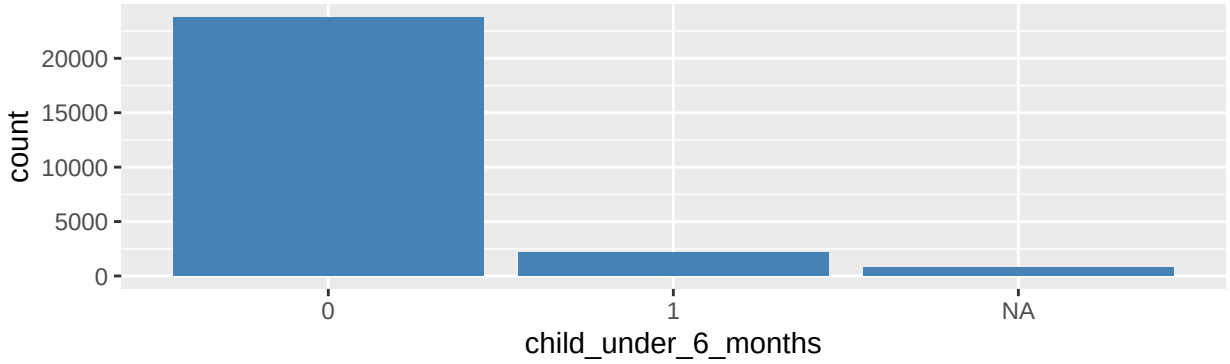

15000
10000
5000
0
count
0
1
NA
behavioral_outside_home

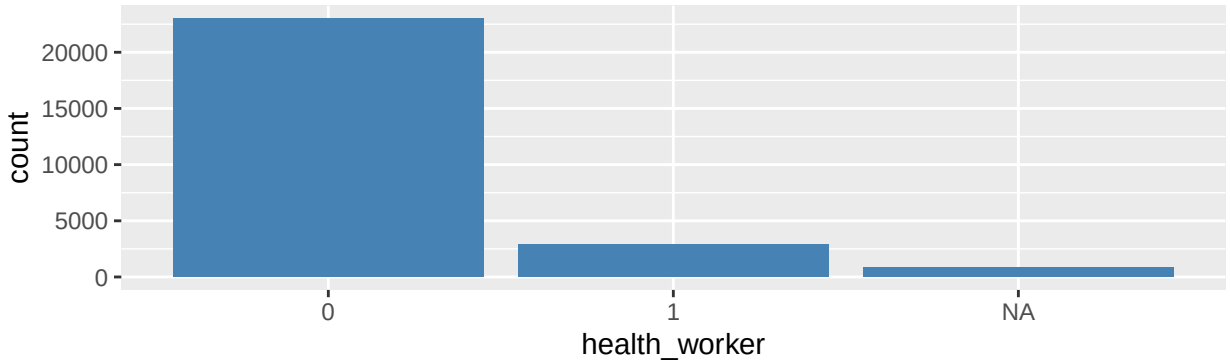

count
15000
10000
5000
0
0
1
NA
behavioral_touch_face

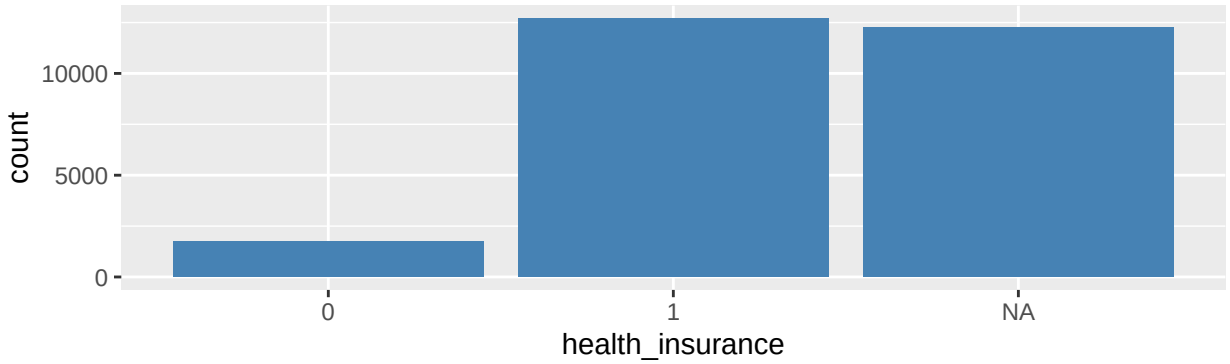

count
20000
15000
10000
5000
0
0
1
NA
doctor_recc_h1n1

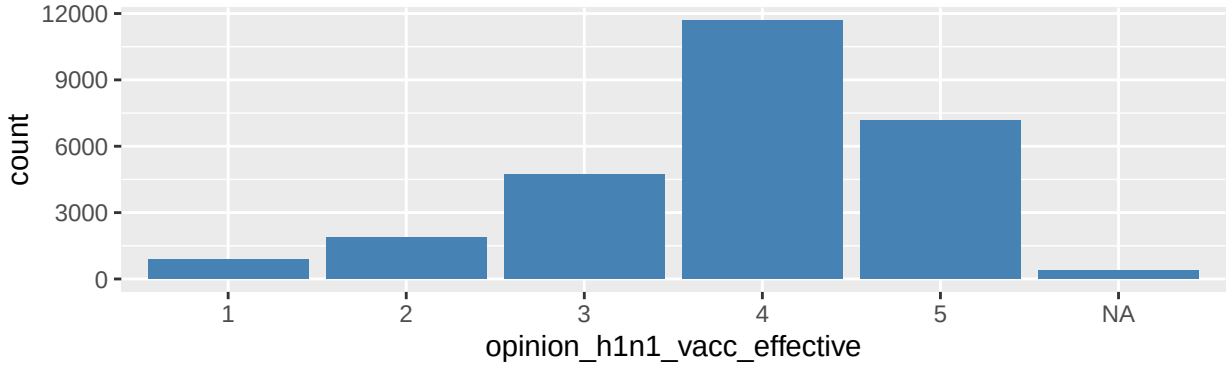

count
15000
10000
5000
0
0
1
NA
doctor_recc_seasonal

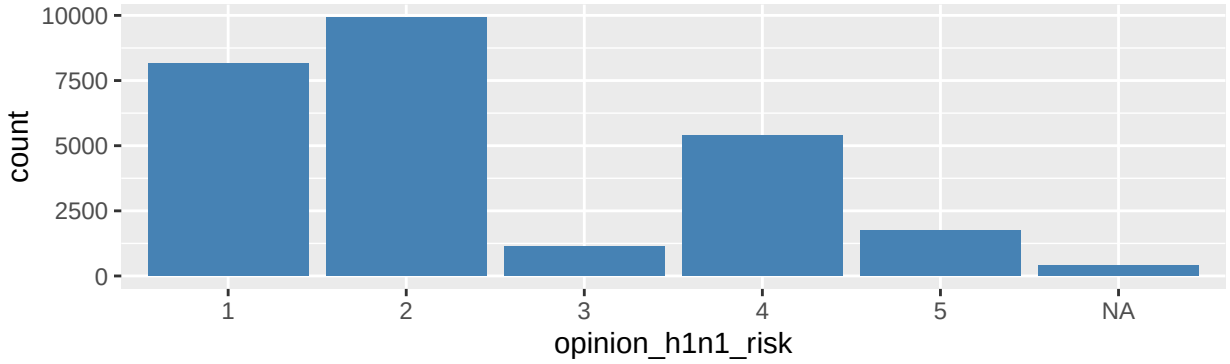

count
15000
10000
5000
0
0
1
NA
chronic_med_condition

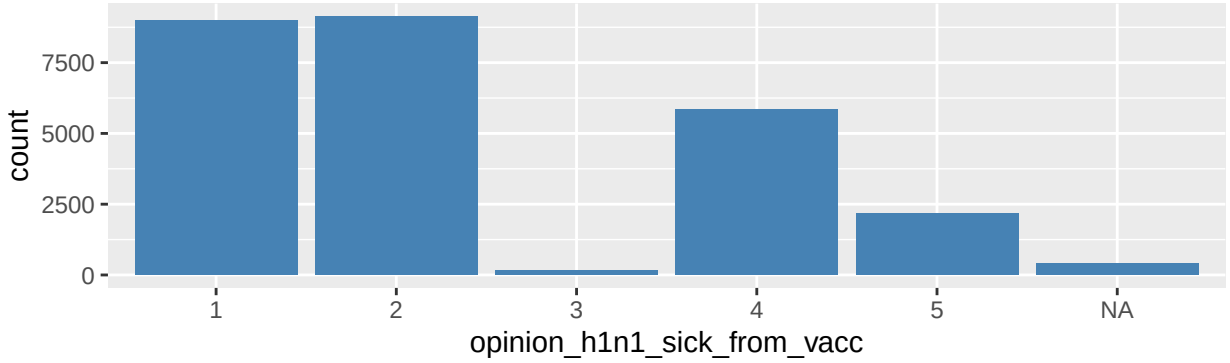

count
20000
15000
10000
5000
0
0
1
NA
child_under_6_months

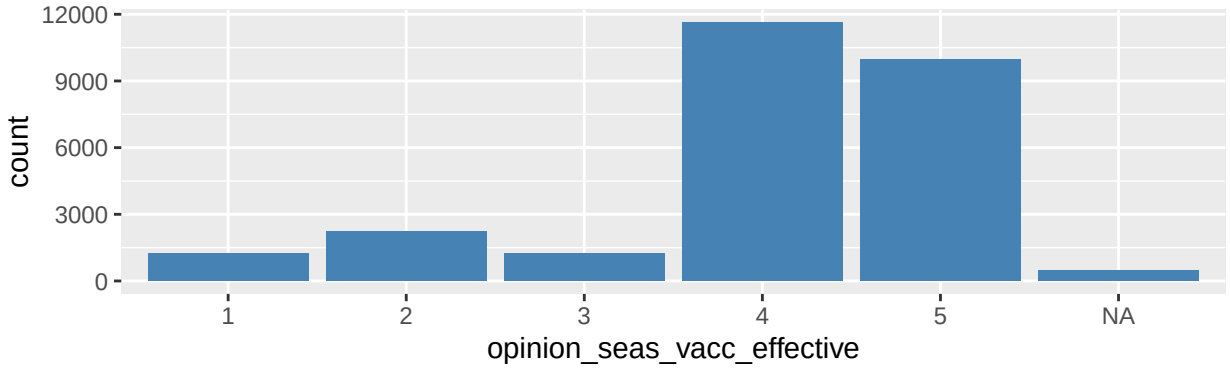

count
20000
15000
10000
5000
0
0
1
NA
health_worker

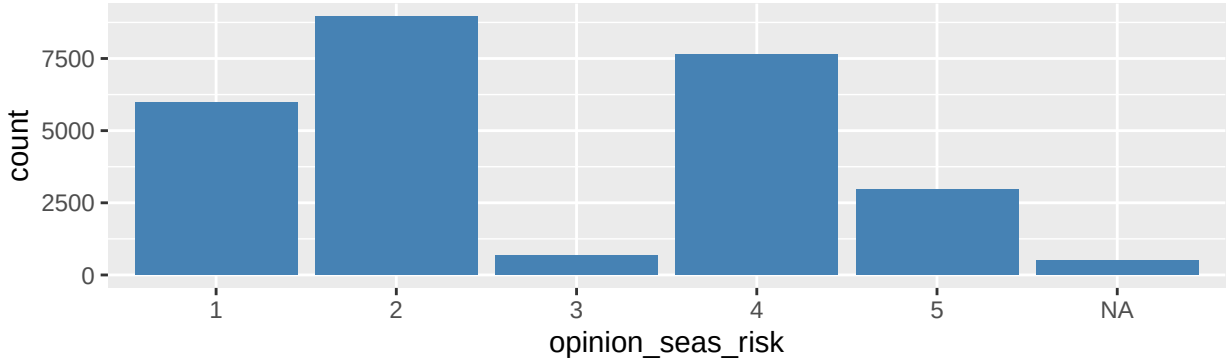

count
10000
5000
0
0
1
NA
health_insurance

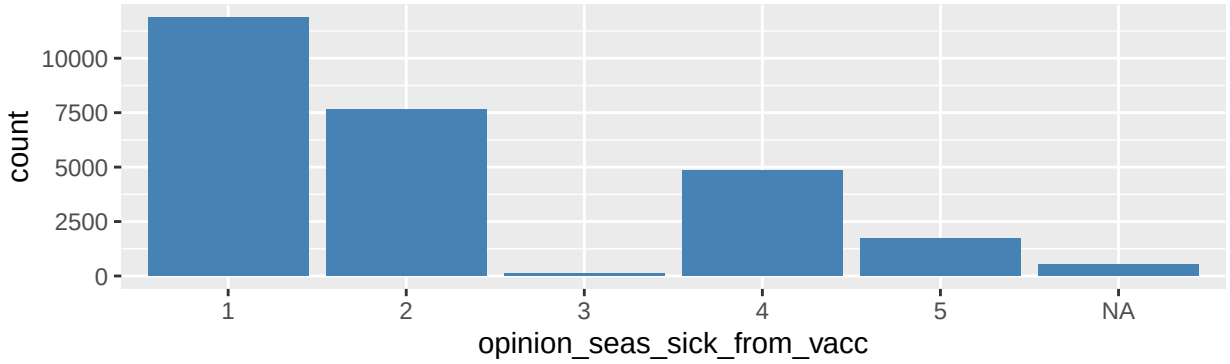

count
12000
9000
6000
3000
0
1
2
3
4
5
NA
opinion_h1n1_vacc_effective

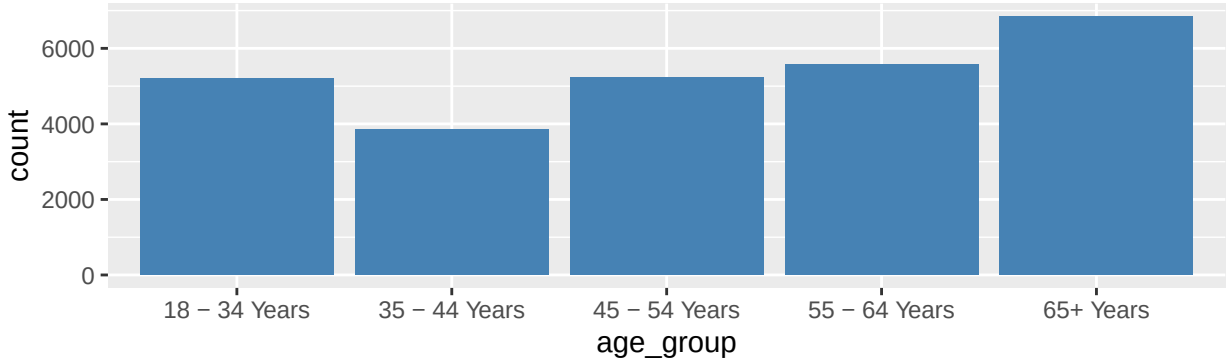

10000
7500
5000
2500
0
count
1
2
3
4
5
NA
opinion_h1n1_risk

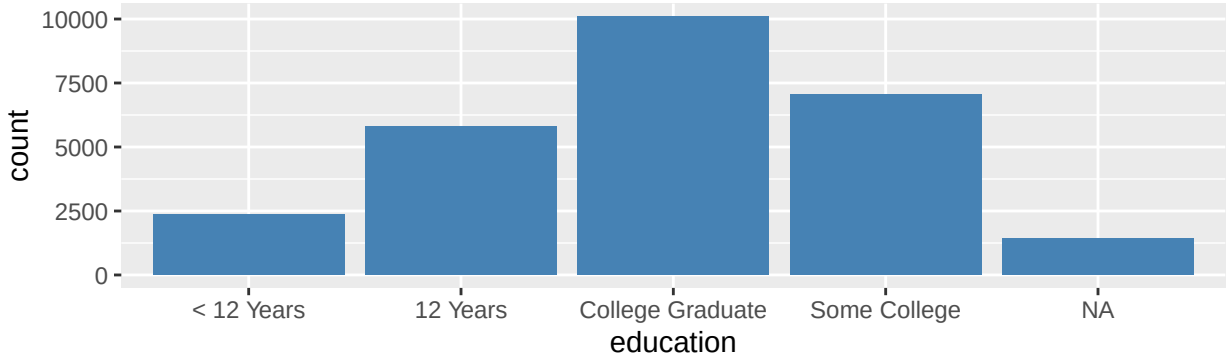

7500
5000
2500
0
count
1
2
3
4
5
NA
opinion_h1n1_sick_from_vacc

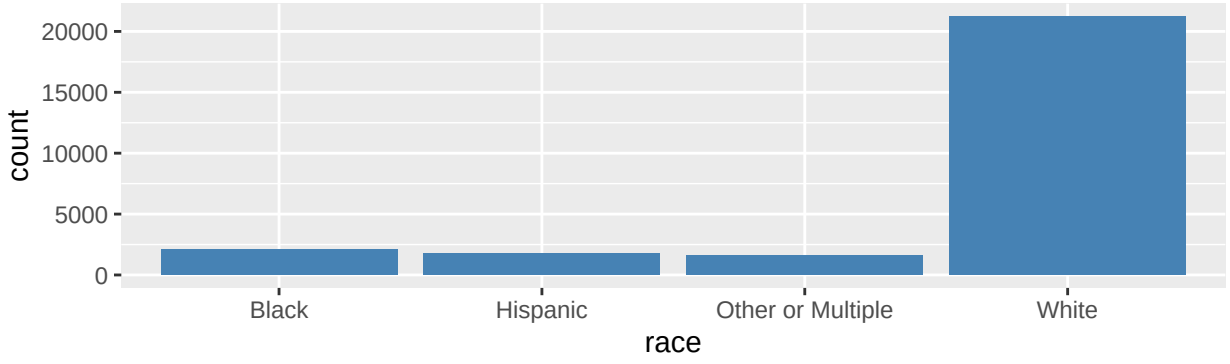

12000
9000
6000
3000
0
count
1
2
3
4
5
NA
opinion_seas_vacc_effective

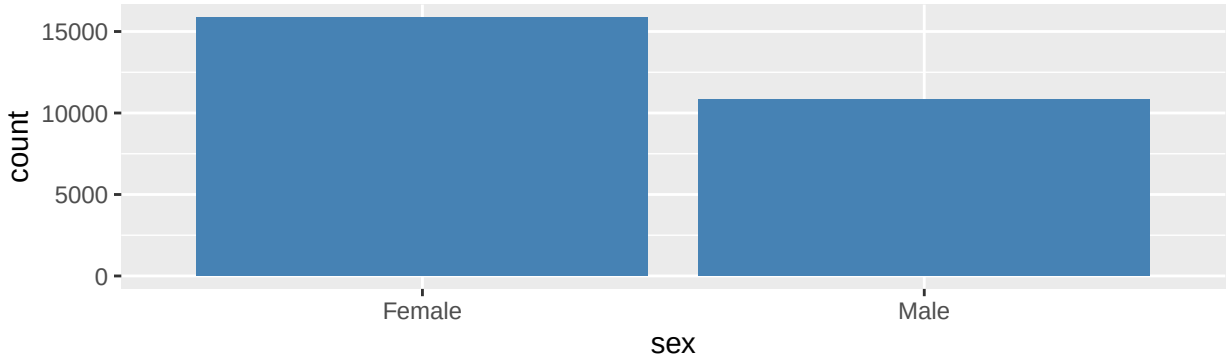

7500
5000
2500
0
count
1
2
3
4
5
NA
opinion_seas_risk

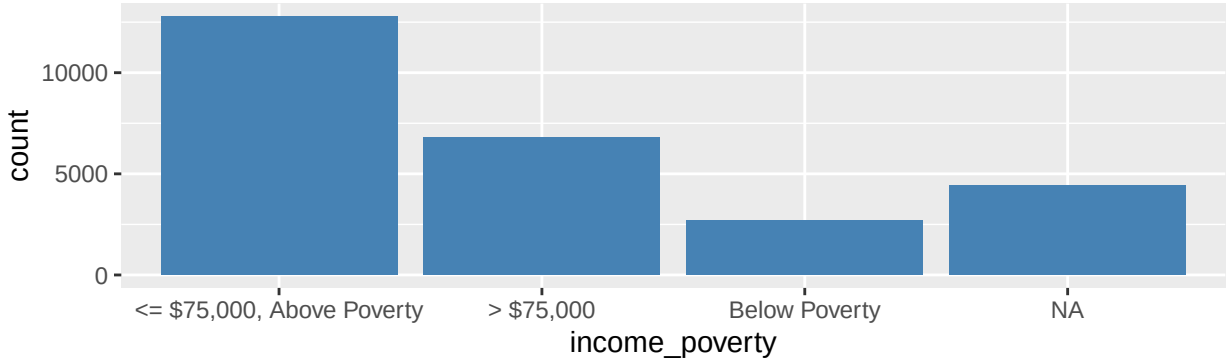

count
10000
7500
5000
2500
0
1
2
3
4
5
NA
opinion_seas_sick_from_vacc

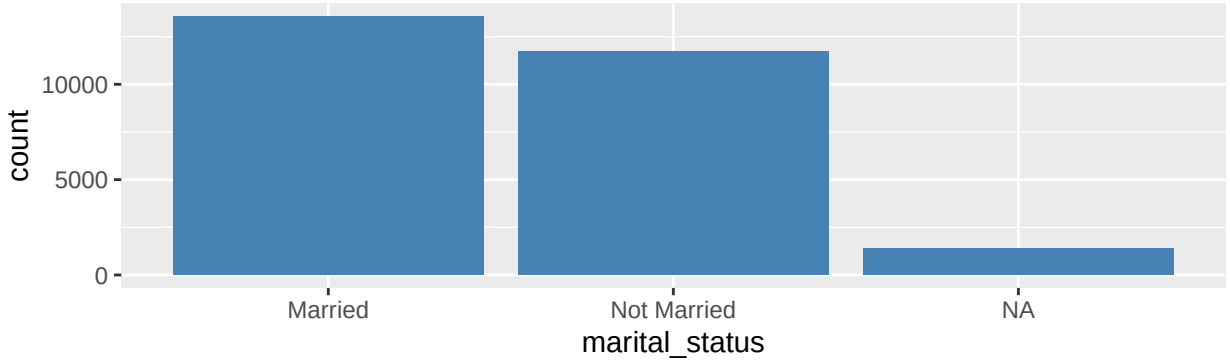

count
6000
4000
2000
0
18 − 34 Years
35 − 44 Years
45 − 54 Years
55 − 64 Years
65+ Years
age_group

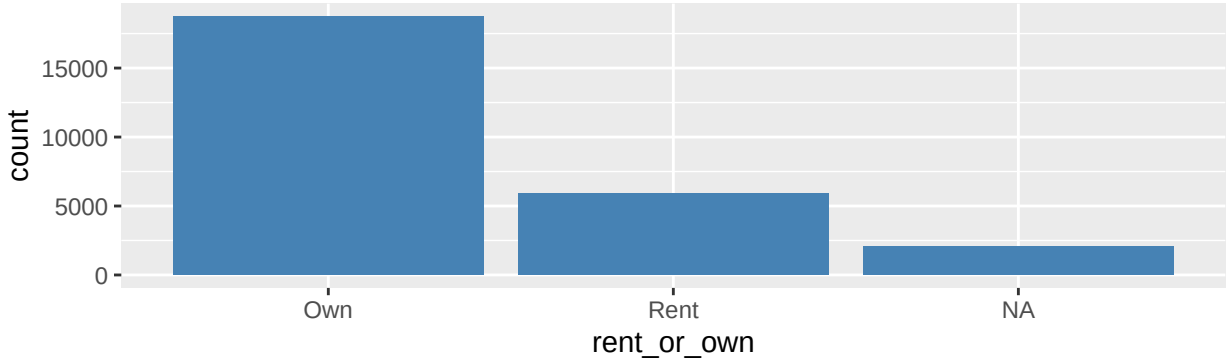

count
10000
7500
5000
2500
0
< 12 Years
12 Years
College Graduate
Some College
NA
education

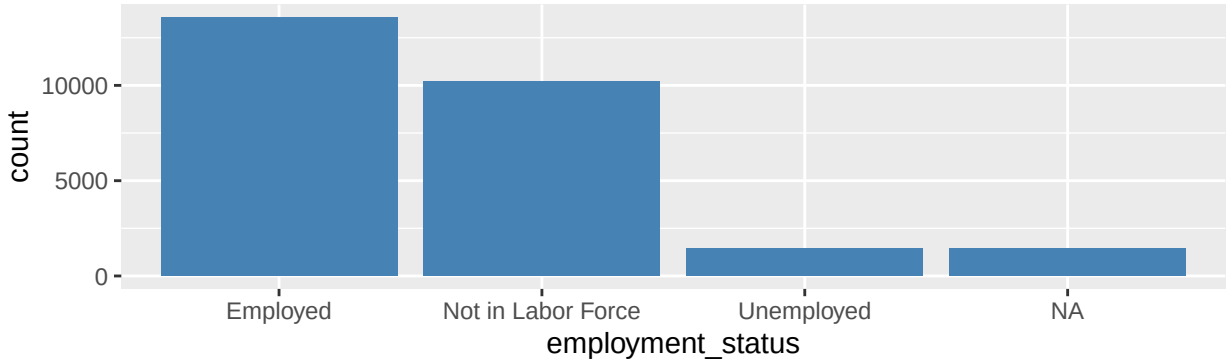

count
20000
15000
10000
5000
0
Black
Hispanic
Other or Multiple
White
race

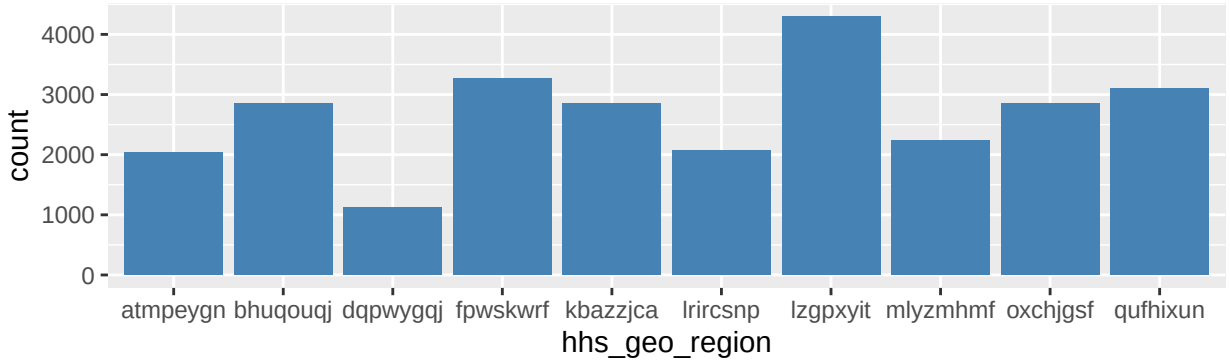

15000
10000
5000
0
count
Female
Male
sex

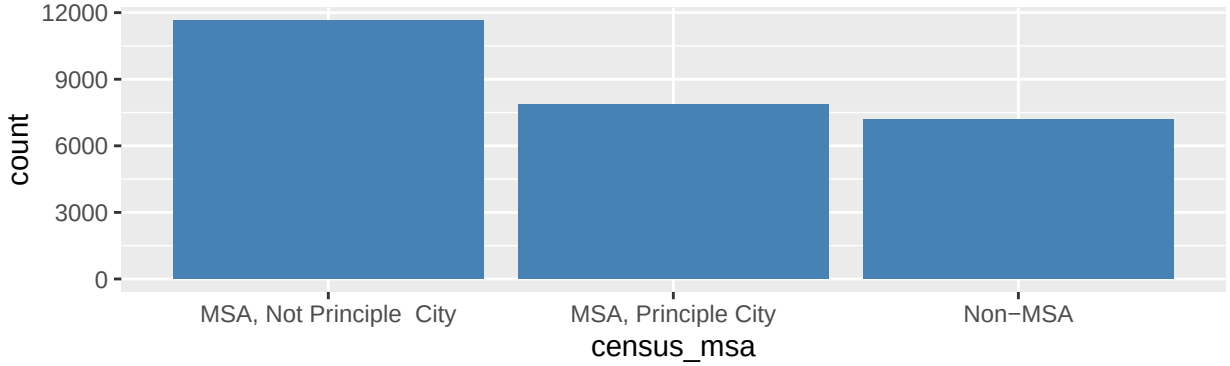

10000
5000
0
count
<= $75,000, Above Poverty
> $75,000
Below Poverty
NA
income_poverty

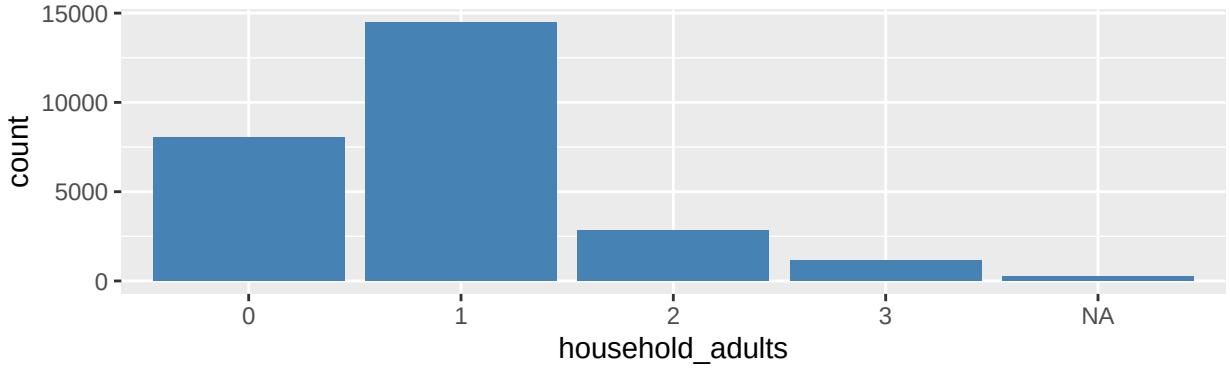

10000
5000
0
count
Married
Not Married
NA
marital_status


15000
10000
5000
0
count
Own
Rent
NA
rent_or_own





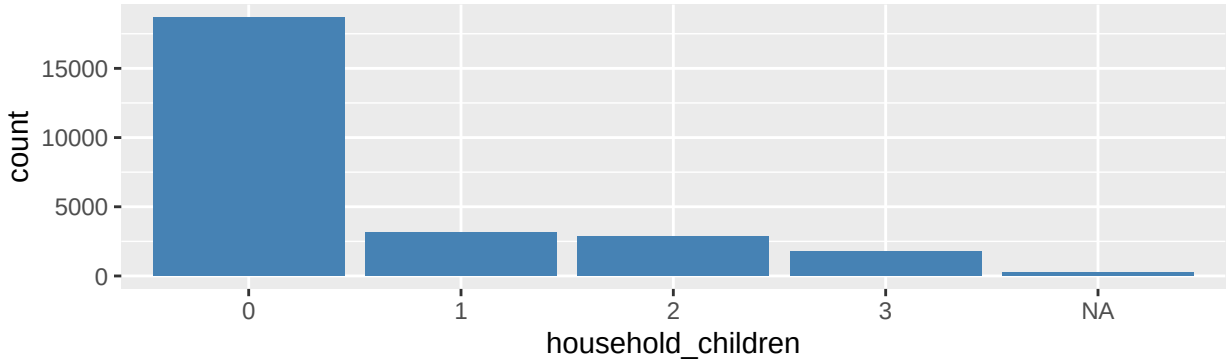


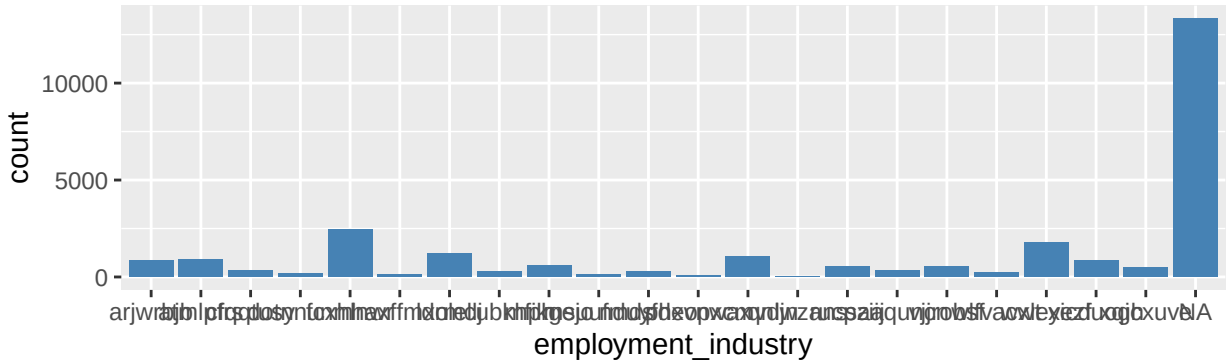


Distribution of each feature by target

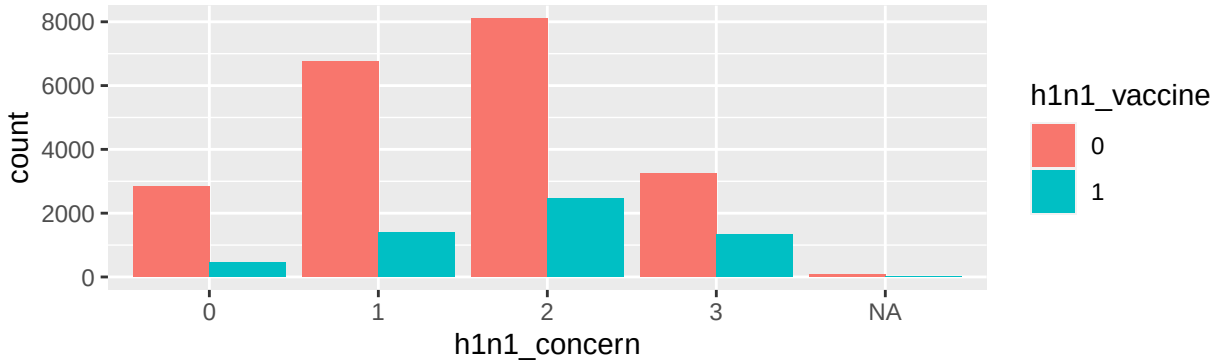
```
for(target in targets){
  for(feature in features){
    p = df_train %>%
```

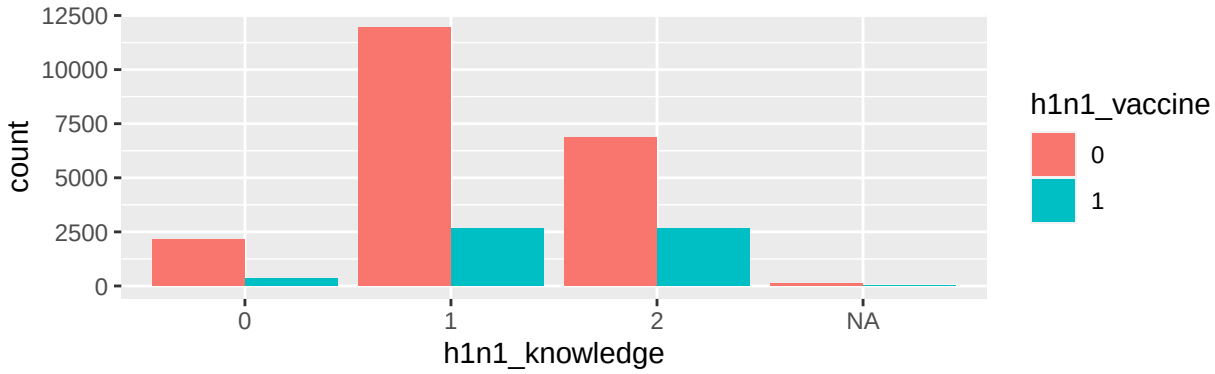
```
      ggplot(aes_string(x=feature, fill=target)) +
      geom_bar(position="dodge")

    plots[[i]] = p
    if(i %% 4 == 0){
      do.call("grid.arrange", c(plots[(i-3): i], nrow=4))
    }
    i = i + 1
  }
}
```

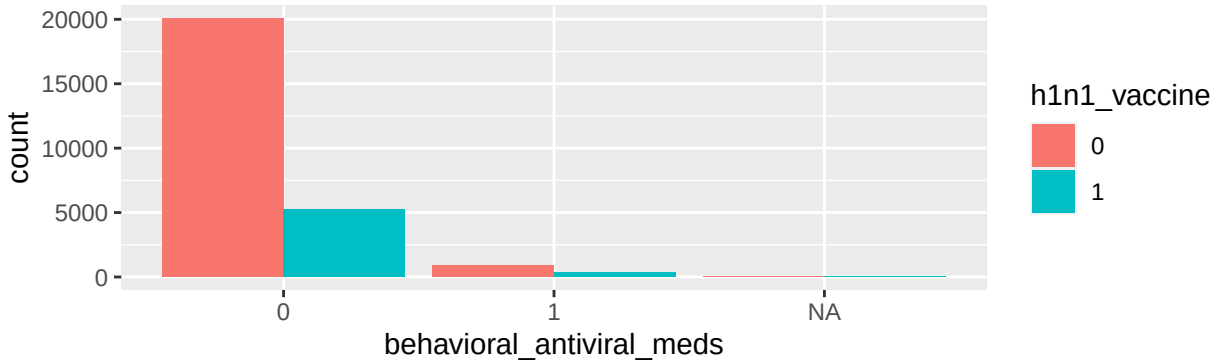

count
15000
10000
5000
0
0
1
2
3
NA
household_children

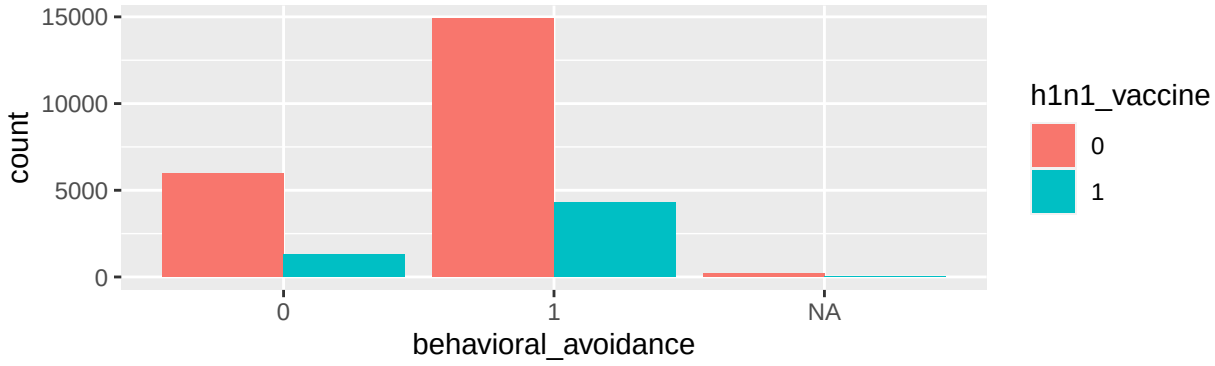

count
10000
5000
0
employment_industry

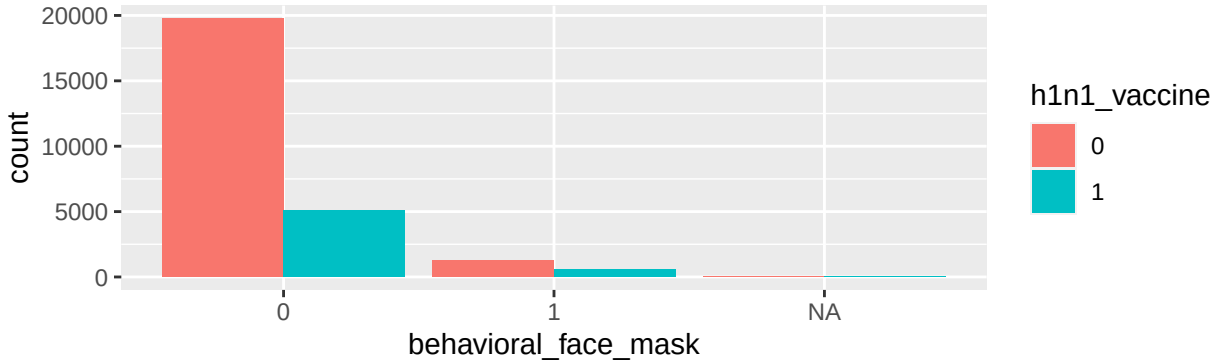

count
10000
5000
0
employment_occupation

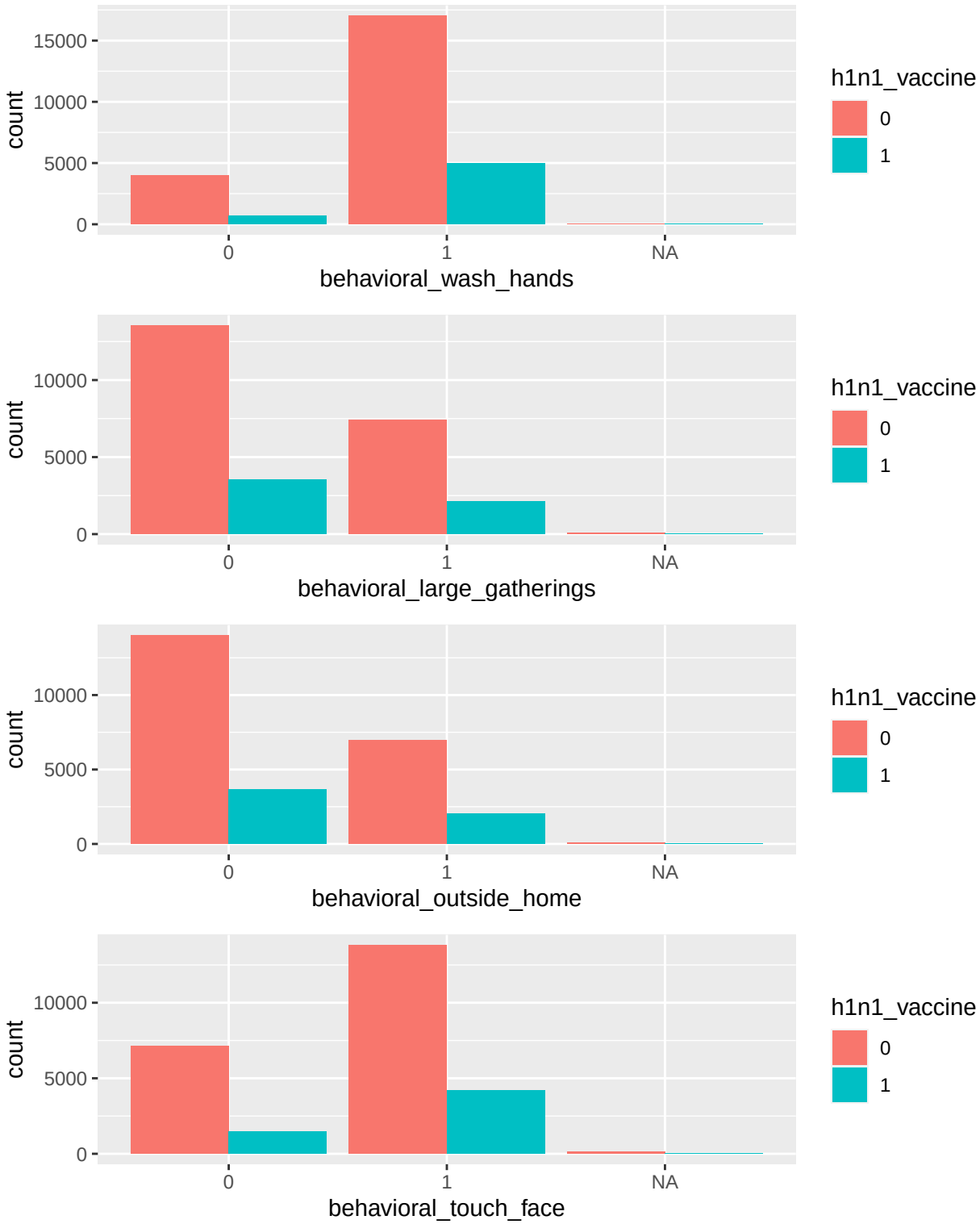

count
8000
6000
4000
2000
0
0
1
2
3
NA
h1n1_concern
h1n1_vaccine
0
1

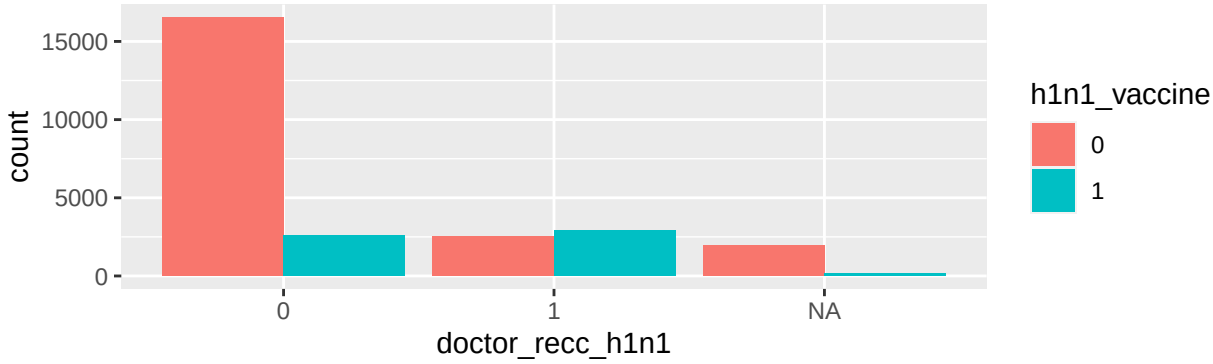

12500
10000
7500
5000
2500
0
count
0
1
2
NA
h1n1_knowledge
h1n1_vaccine
0
1

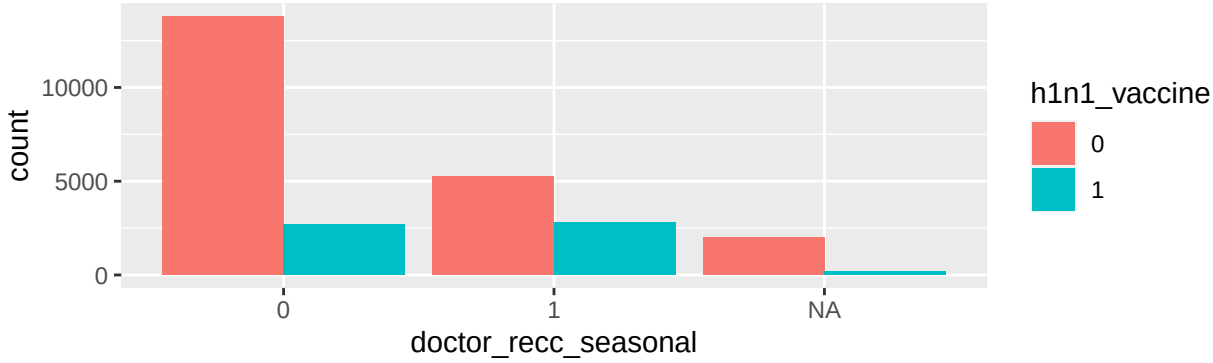

20000
15000
10000
5000
0
count
0
1
NA
behavioral_antiviral_meds
h1n1_vaccine
0
1

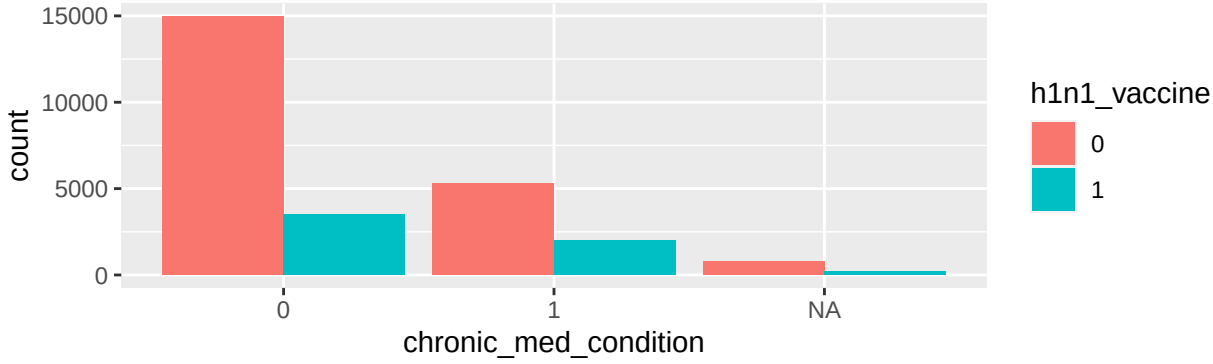

15000
10000
5000
0
count
0
1
NA
behavioral_avoidance
h1n1_vaccine
0
1

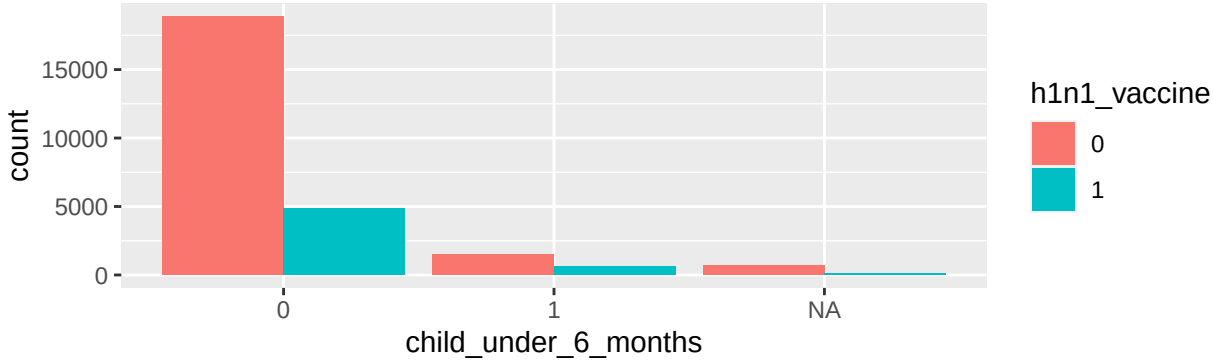

20000
15000
10000
5000
0
count
0
1
NA
behavioral_face_mask
h1n1_vaccine
0
1

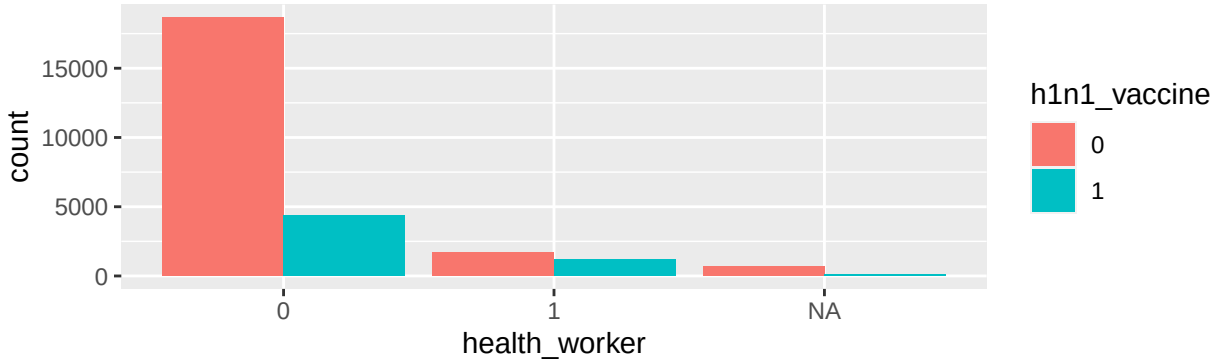

count
15000
10000
5000
0
0
1
NA
behavioral_wash_hands
h1n1_vaccine
0
1
count
10000
5000
0
0
1
NA
behavioral_large_gatherings
h1n1_vaccine
0
1
count
10000
5000
0
0
1
NA
behavioral_outside_home
h1n1_vaccine
0
1
count
10000
5000
0
0
1
NA
behavioral_touch_face
h1n1_vaccine
0
1

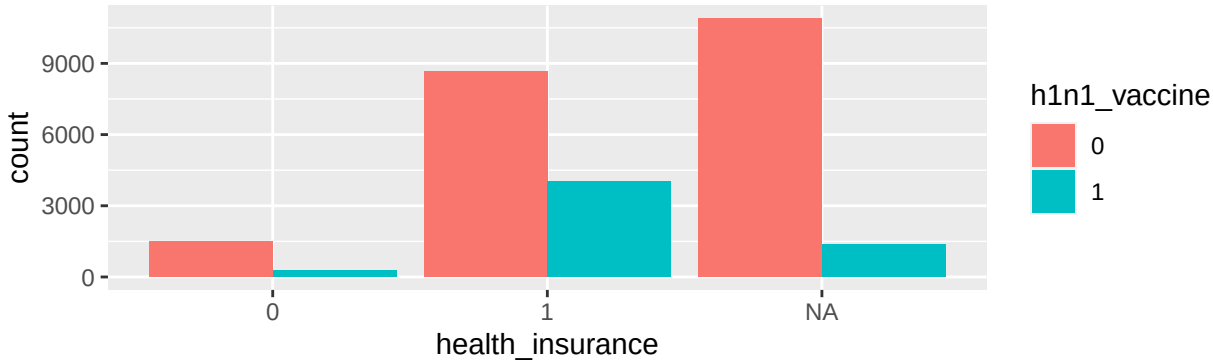

15000
10000
5000
0
count
0
1
NA
doctor_recc_h1n1
h1n1_vaccine
0
1

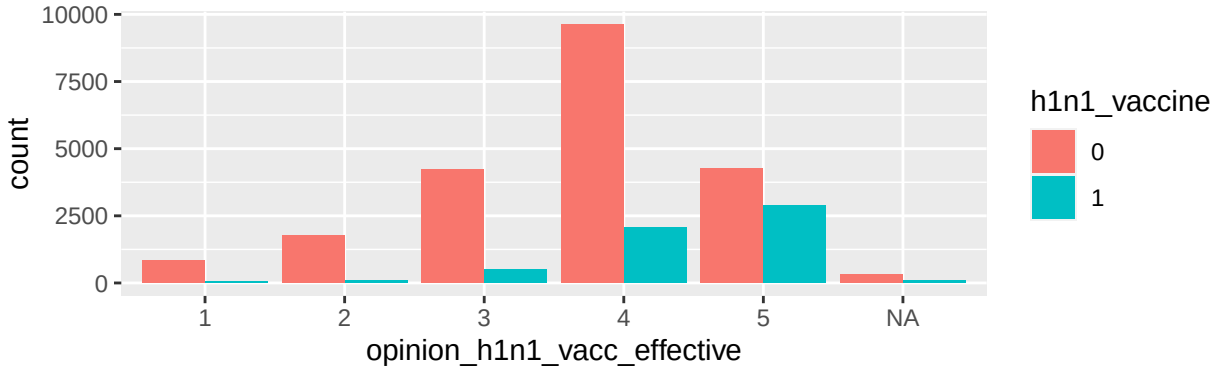

10000
5000
0
count
0
1
NA
doctor_recc_seasonal
h1n1_vaccine
0
1

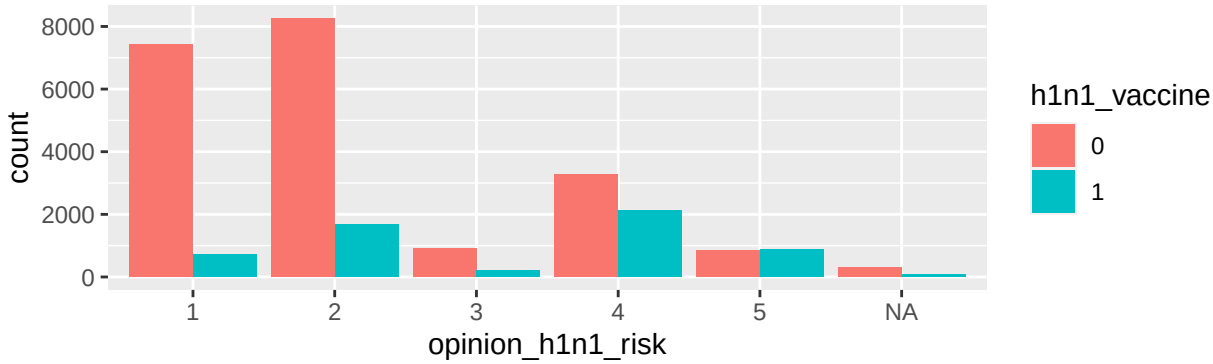

15000
10000
5000
0
count
0
1
NA
chronic_med_condition
h1n1_vaccine
0
1

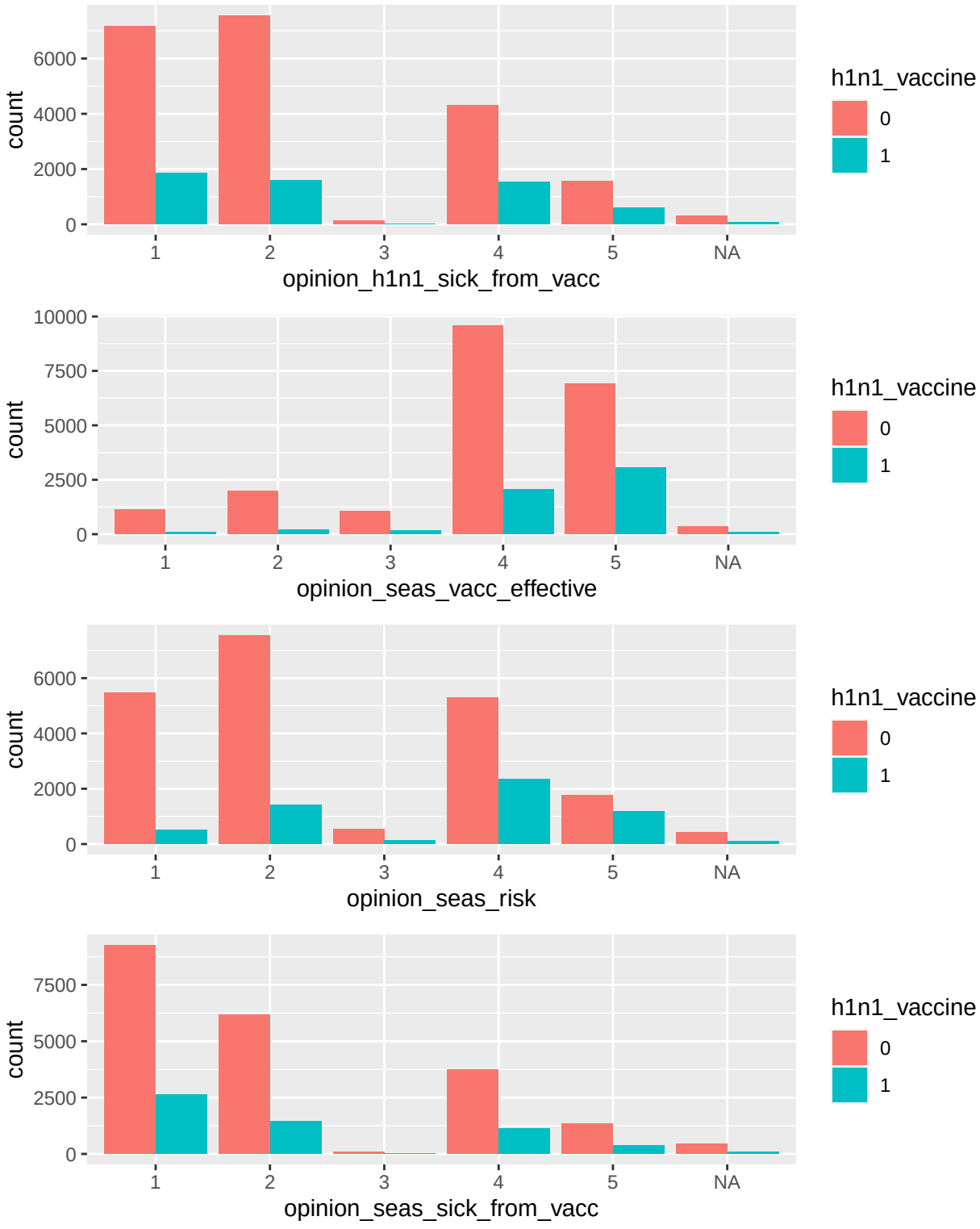

15000
10000
5000
0
count
0
1
NA
child_under_6_months
h1n1_vaccine
0
1

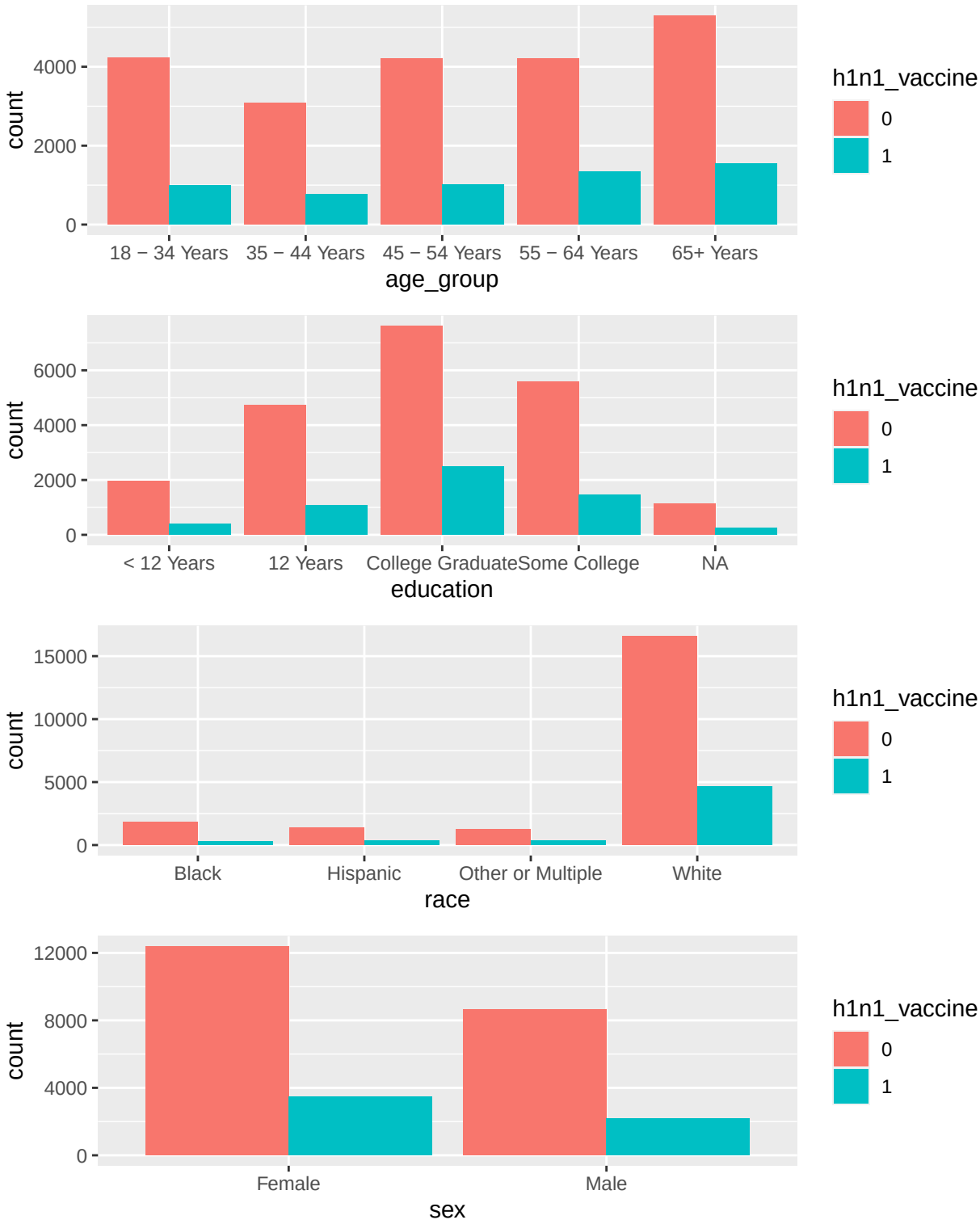

count
15000
10000
5000
0
0
1
NA
health_worker
h1n1_vaccine
0
1

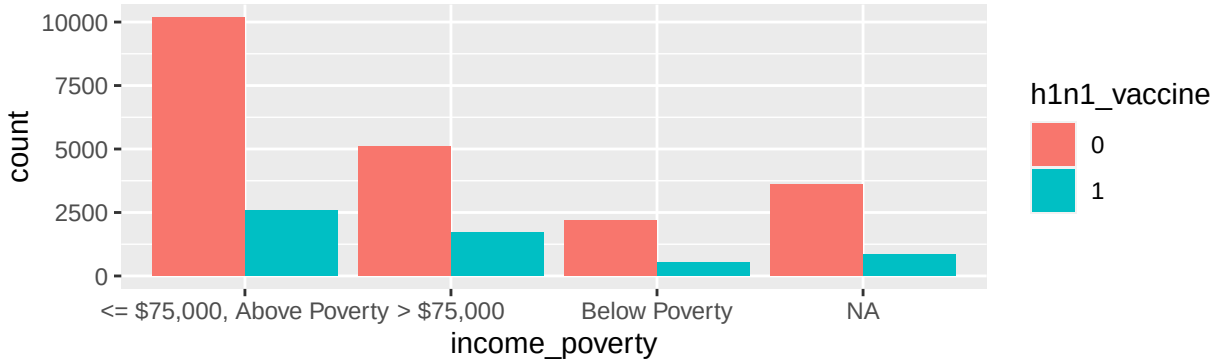

count
9000
6000
3000
0
0
1
NA
health_insurance
h1n1_vaccine
0
1

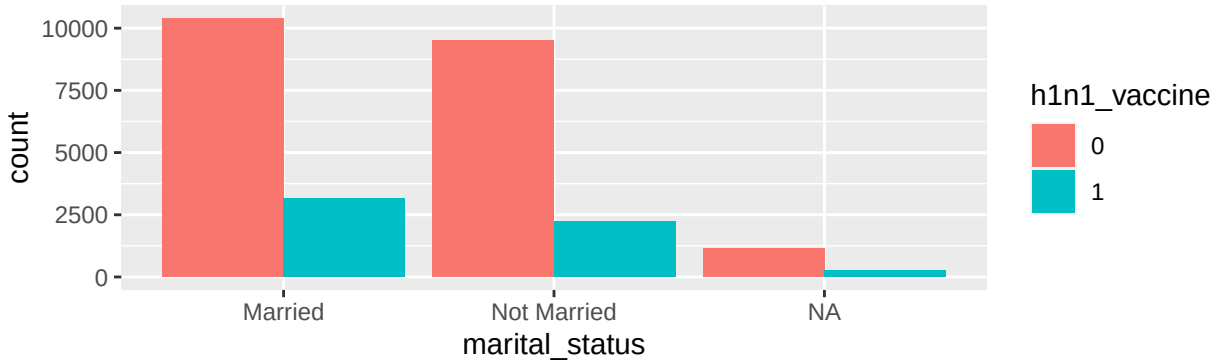

count
10000
7500
5000
2500
0
1
2
3
4
5
NA
opinion_h1n1_vacc_effective
h1n1_vaccine
0
1

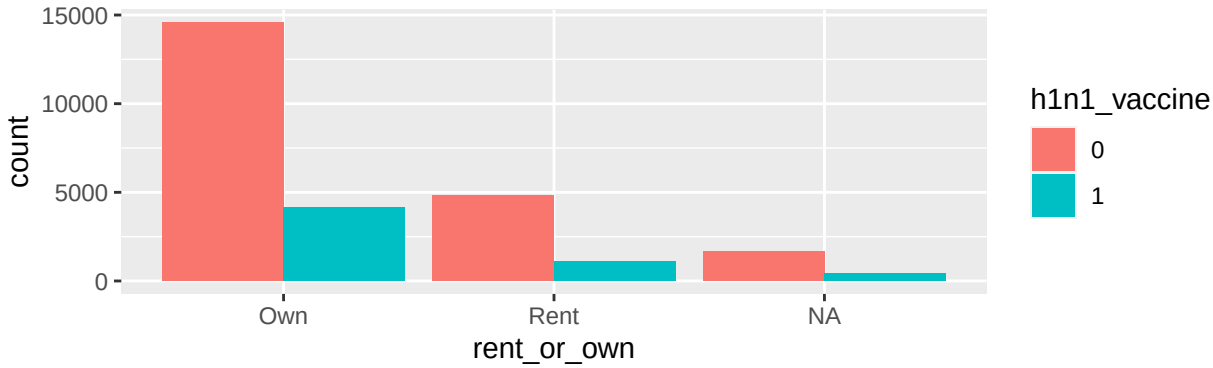

count
8000
6000
4000
2000
0
1
2
3
4
5
NA
opinion_h1n1_risk
h1n1_vaccine
0
1

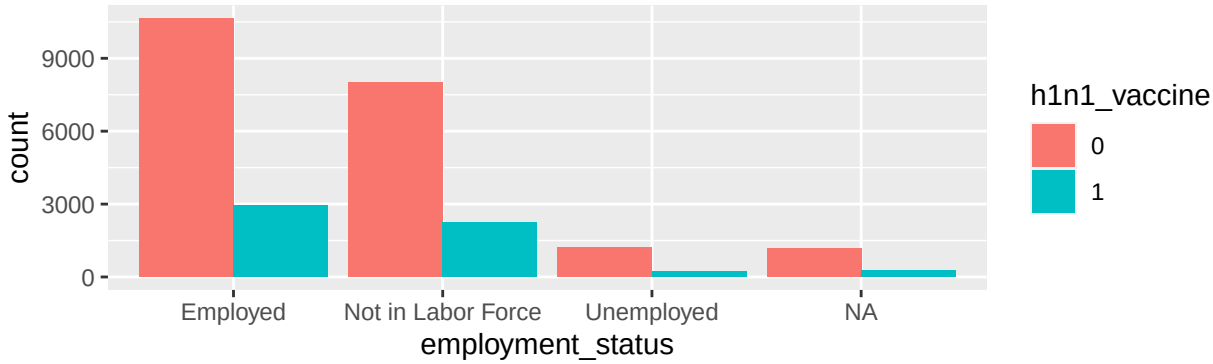

count
6000
4000
2000
0
1
2
3
4
5
NA
opinion_h1n1_sick_from_vacc
h1n1_vaccine
0
1
count
10000
7500
5000
2500
0
1
2
3
4
5
NA
opinion_seas_vacc_effective
h1n1_vaccine
0
1
count
6000
4000
2000
0
1
2
3
4
5
NA
opinion_seas_risk
h1n1_vaccine
0
1
count
7500
5000
2500
0
1
2
3
4
5
NA
opinion_seas_sick_from_vacc
h1n1_vaccine
0
1

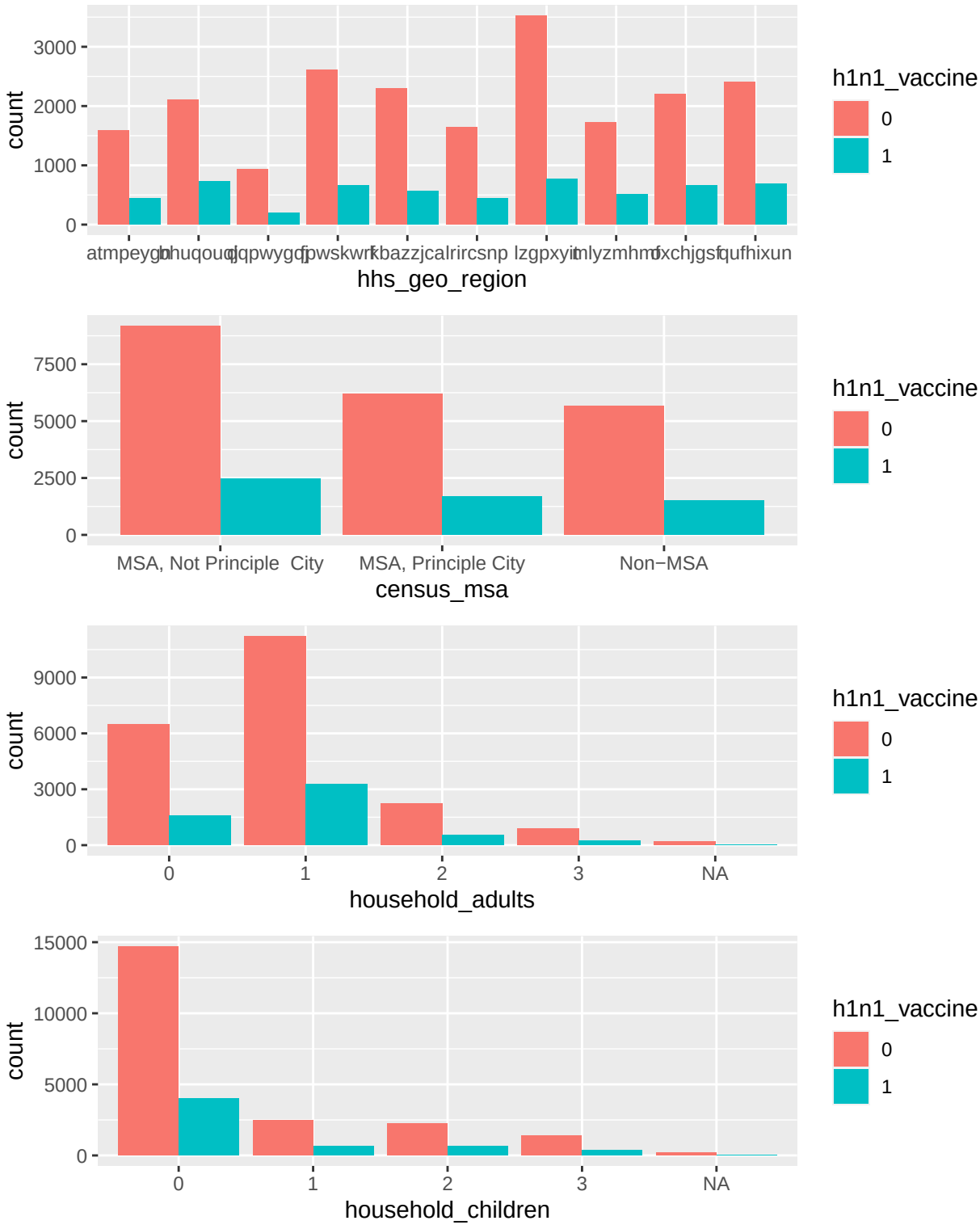

count
4000
2000
0
18 − 34 Years
35 − 44 Years
45 − 54 Years
55 − 64 Years
65+ Years
age_group
h1n1_vaccine
0
1
count
6000
4000
2000
0
< 12 Years
12 Years
College Graduate
Some College
NA
education
h1n1_vaccine
0
1
count
15000
10000
5000
0
Black
Hispanic
Other or Multiple
White
race
h1n1_vaccine
0
1
count
12000
8000
4000
0
Female
Male
sex
h1n1_vaccine
0
1

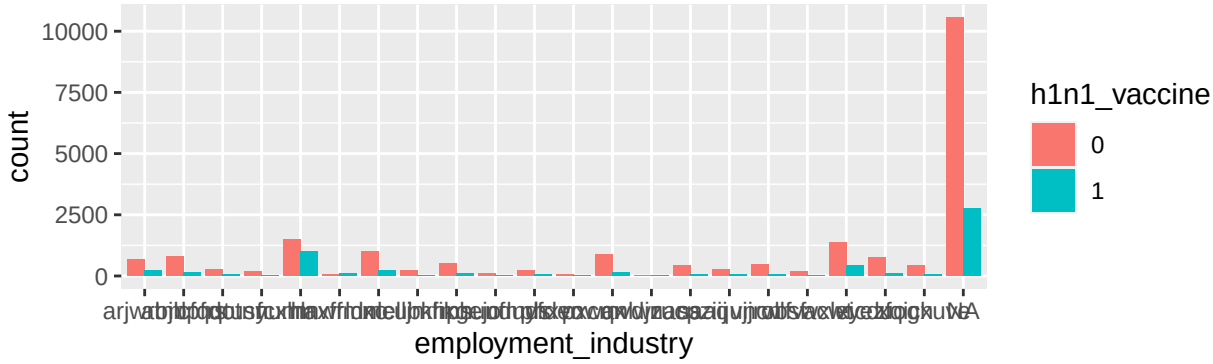

10000
7500
5000
2500
0
count
<= $75,000, Above Poverty
> $75,000
Below Poverty
NA
income_poverty
h1n1_vaccine
0
1


10000
7500
5000
2500
0
count
Married
Not Married
NA
marital_status
h1n1_vaccine
0
1

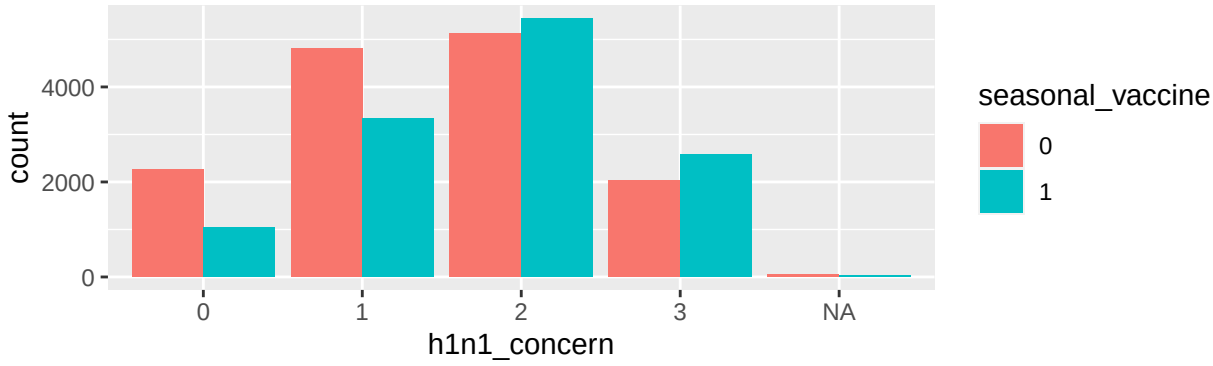

15000
10000
5000
0
count
Own
Rent
NA
rent_or_own
h1n1_vaccine
0
1

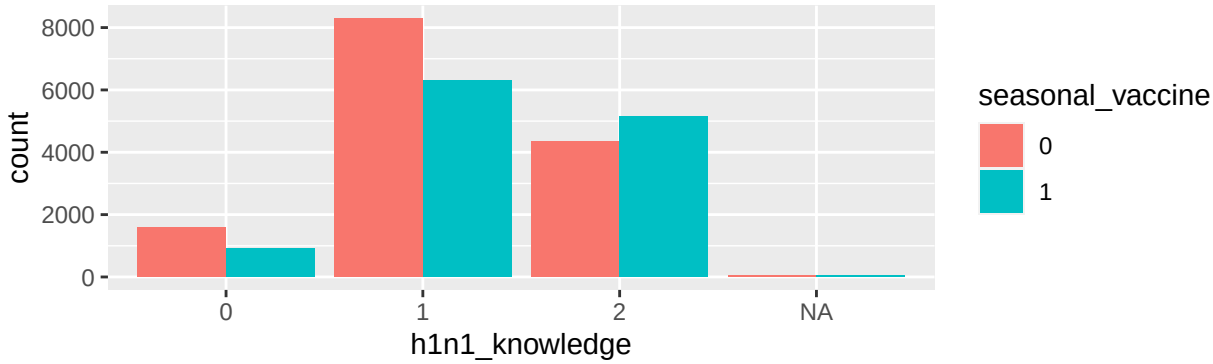

9000
6000
3000
0
count
Employed
Not in Labor Force
Unemployed
NA
employment_status
h1n1_vaccine
0
1

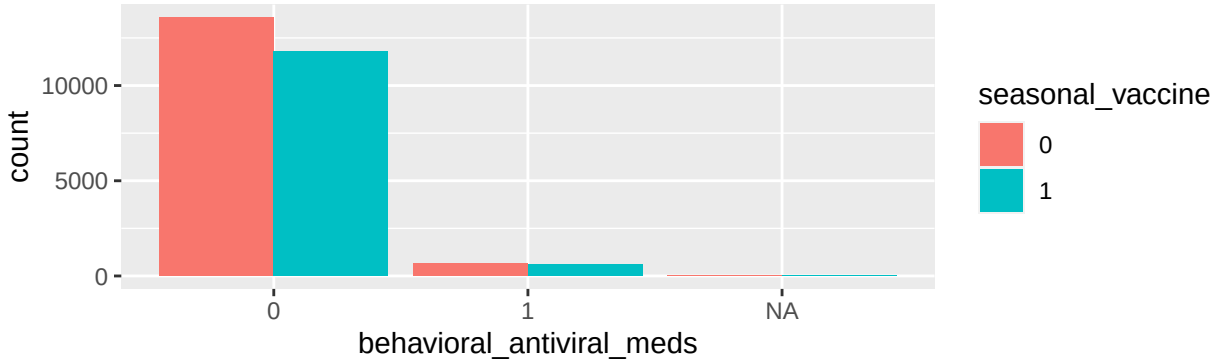

count
3000
2000
1000
0
atmpeygb
bhuqouqd
qjqpwygq
fpwskwrk
kbazzjca
lrircsnp
lzgpxyit
mlyzmhmd
oxchjgsf
qufhixun
hhs_geo_region
h1n1_vaccine
0
1
count
7500
5000
2500
0
MSA, Not Principle City
MSA, Principle City
Non−MSA
census_msa
h1n1_vaccine
0
1
count
9000
6000
3000
0
0
1
2
3
NA
household_adults
h1n1_vaccine
0
1
count
15000
10000
5000
0
0
1
2
3
NA
household_children
h1n1_vaccine
0
1

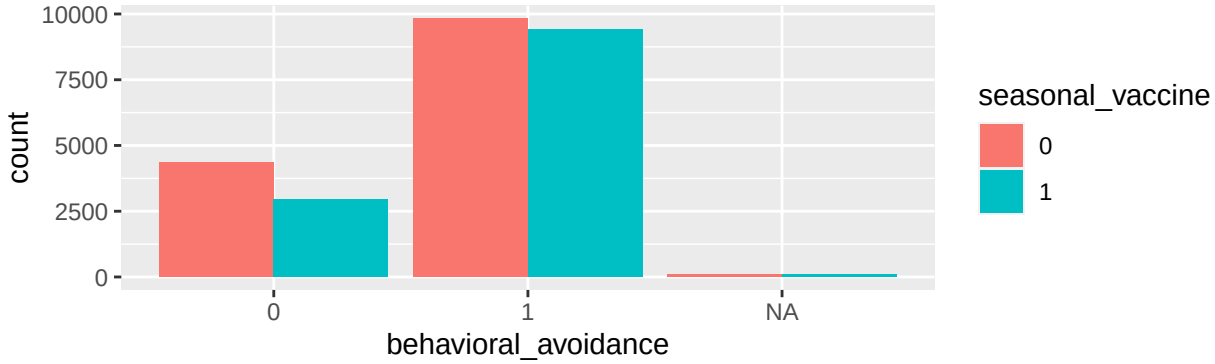

count
10000
7500
5000
2500
0
employment_industry
h1n1_vaccine
0
1

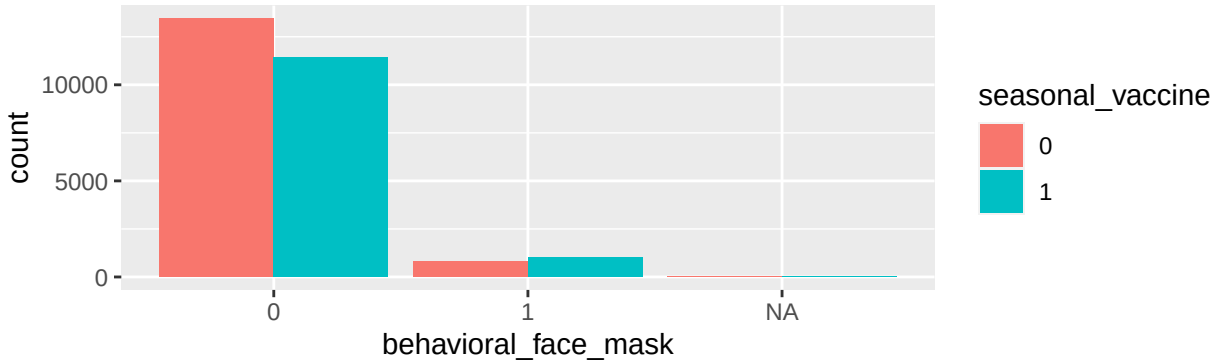

count
9000
6000
3000
0
employment_occupation
h1n1_vaccine
0
1

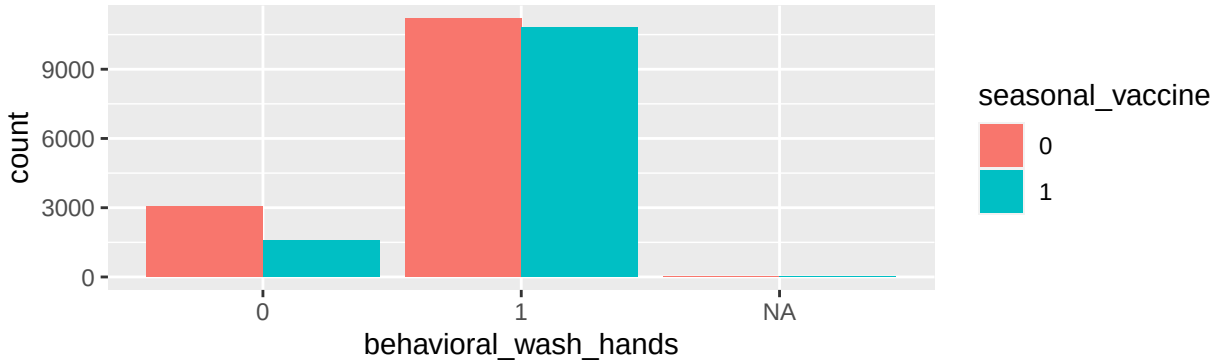

count
4000
2000
0
0
1
2
3
NA
h1n1_concern
seasonal_vaccine
0
1

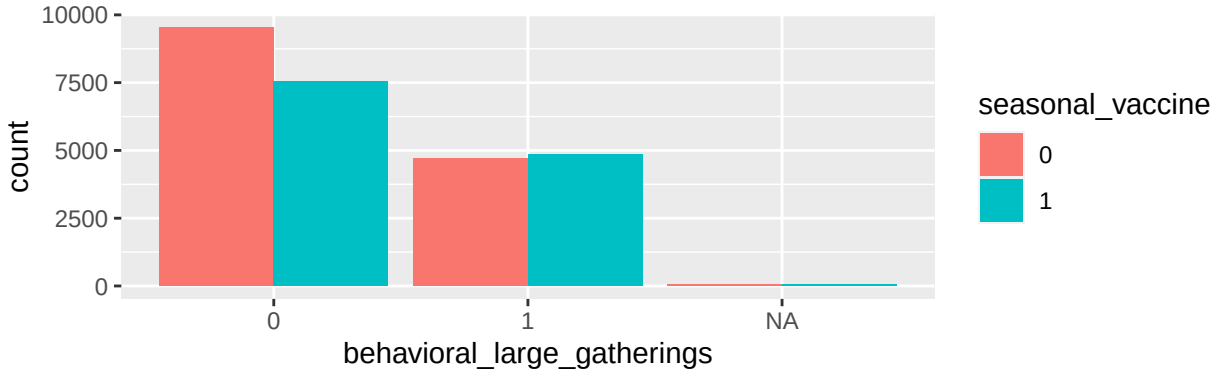

count
8000
6000
4000
2000
0
0
1
2
NA
h1n1_knowledge
seasonal_vaccine
0
1

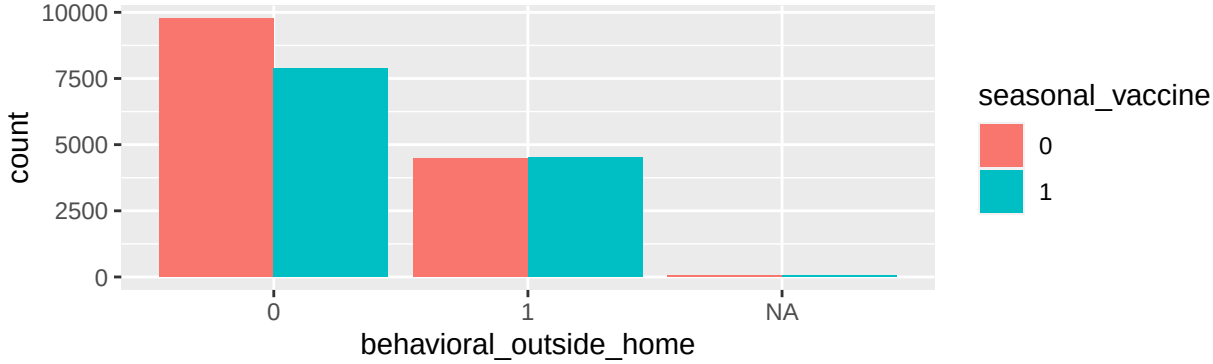

count
10000
5000
0
0
1
NA
behavioral_antiviral_meds
seasonal_vaccine
0
1

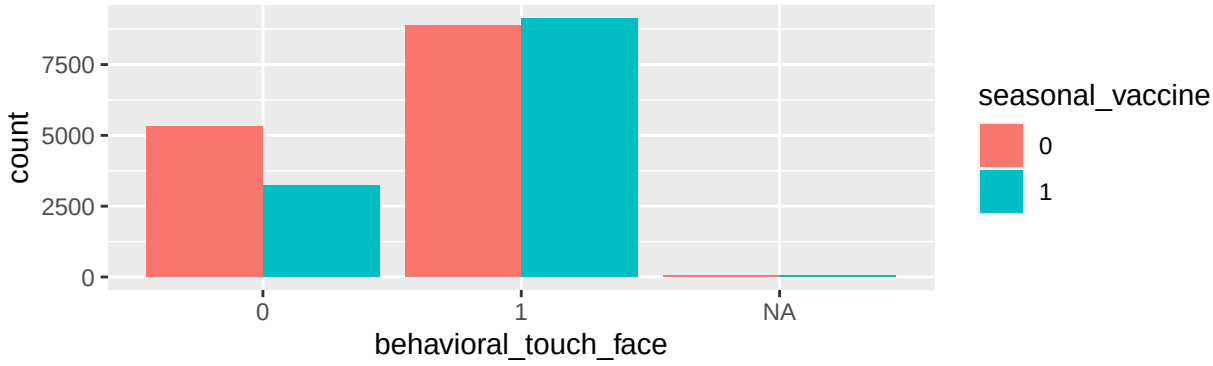

count
10000
7500
5000
2500
0
0
1
NA
behavioral_avoidance
seasonal_vaccine
0
1

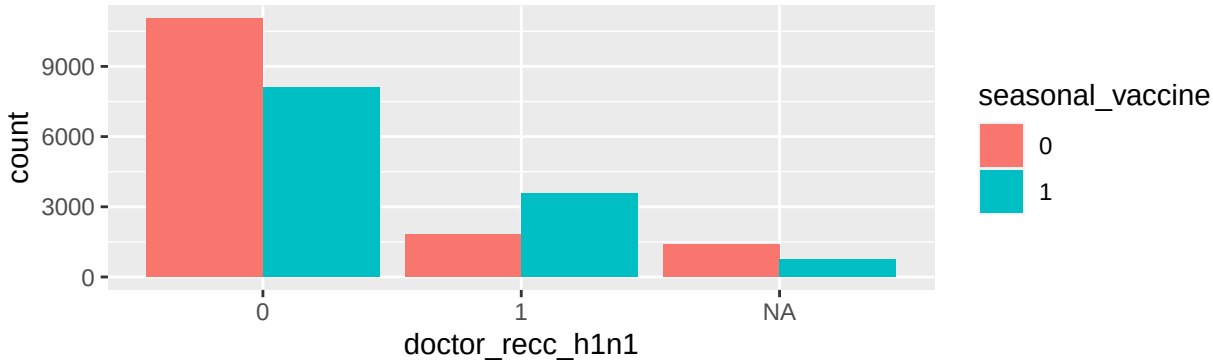

count
10000
5000
0
0
1
NA
behavioral_face_mask
seasonal_vaccine
0
1

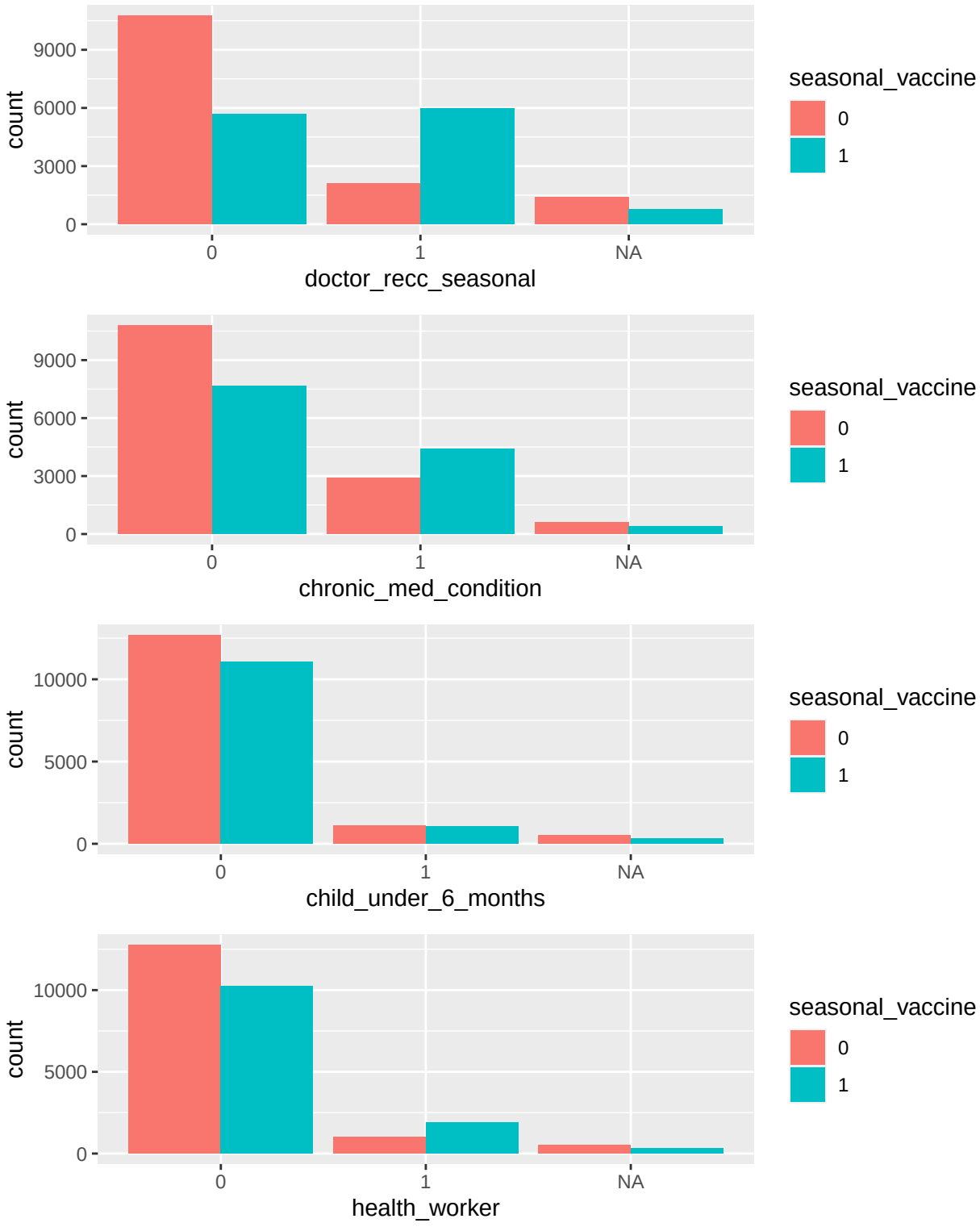

count
9000
6000
3000
0
0
1
NA
behavioral_wash_hands
seasonal_vaccine
0
1

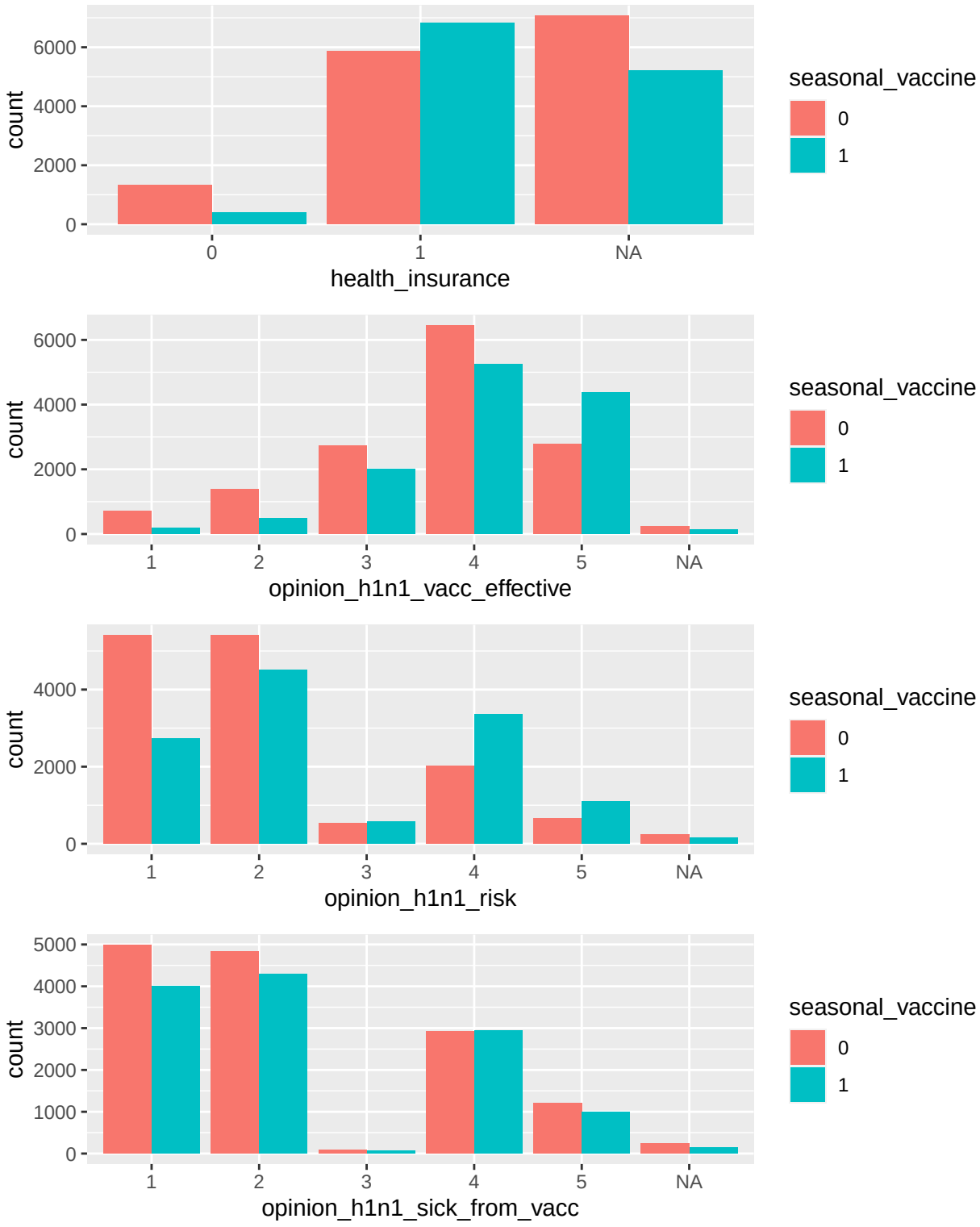

10000
7500
5000
2500
0
count
0
1
NA
behavioral_large_gatherings
seasonal_vaccine
0
1

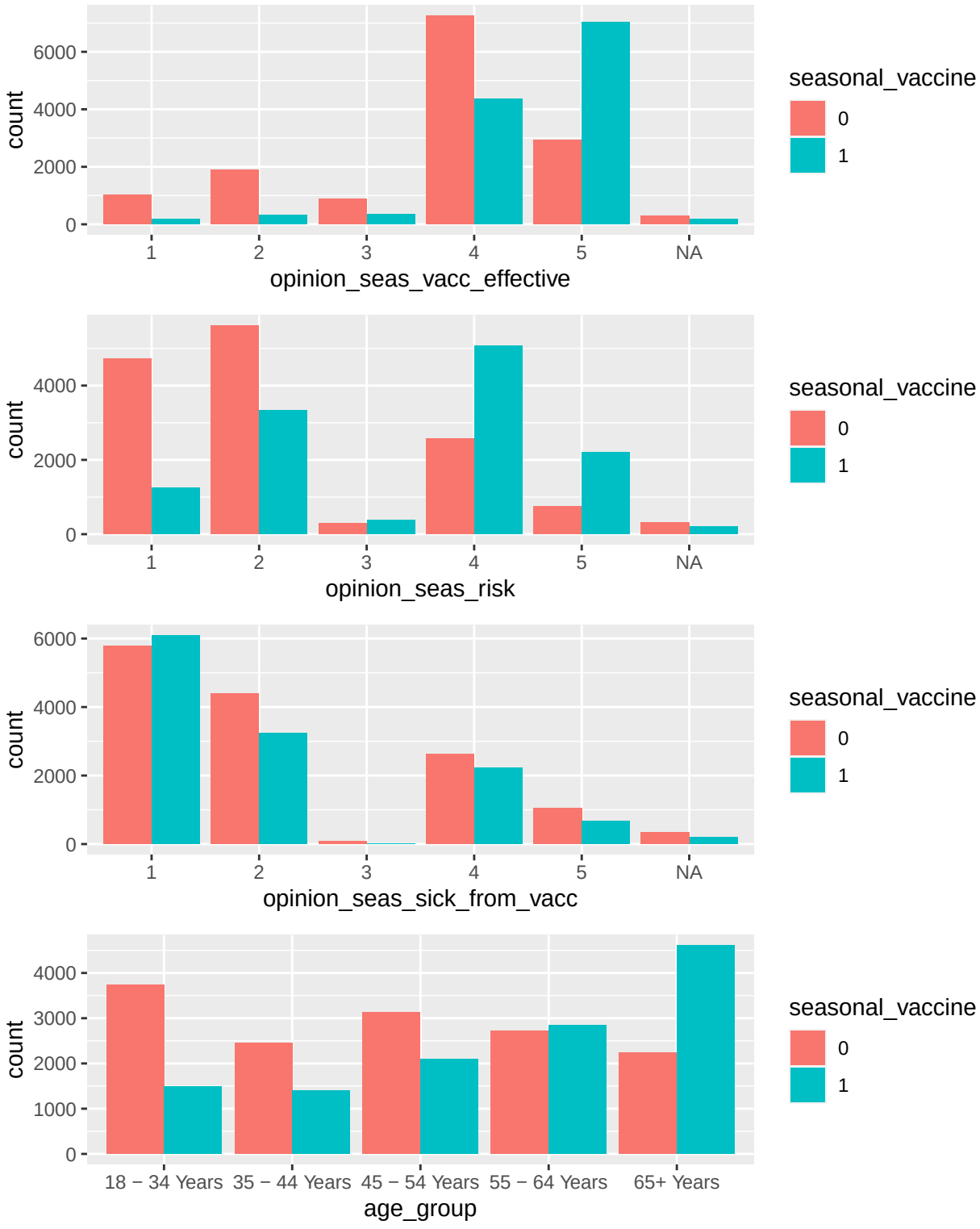

10000
7500
5000
2500
0
count
0
1
NA
behavioral_outside_home
seasonal_vaccine
0
1

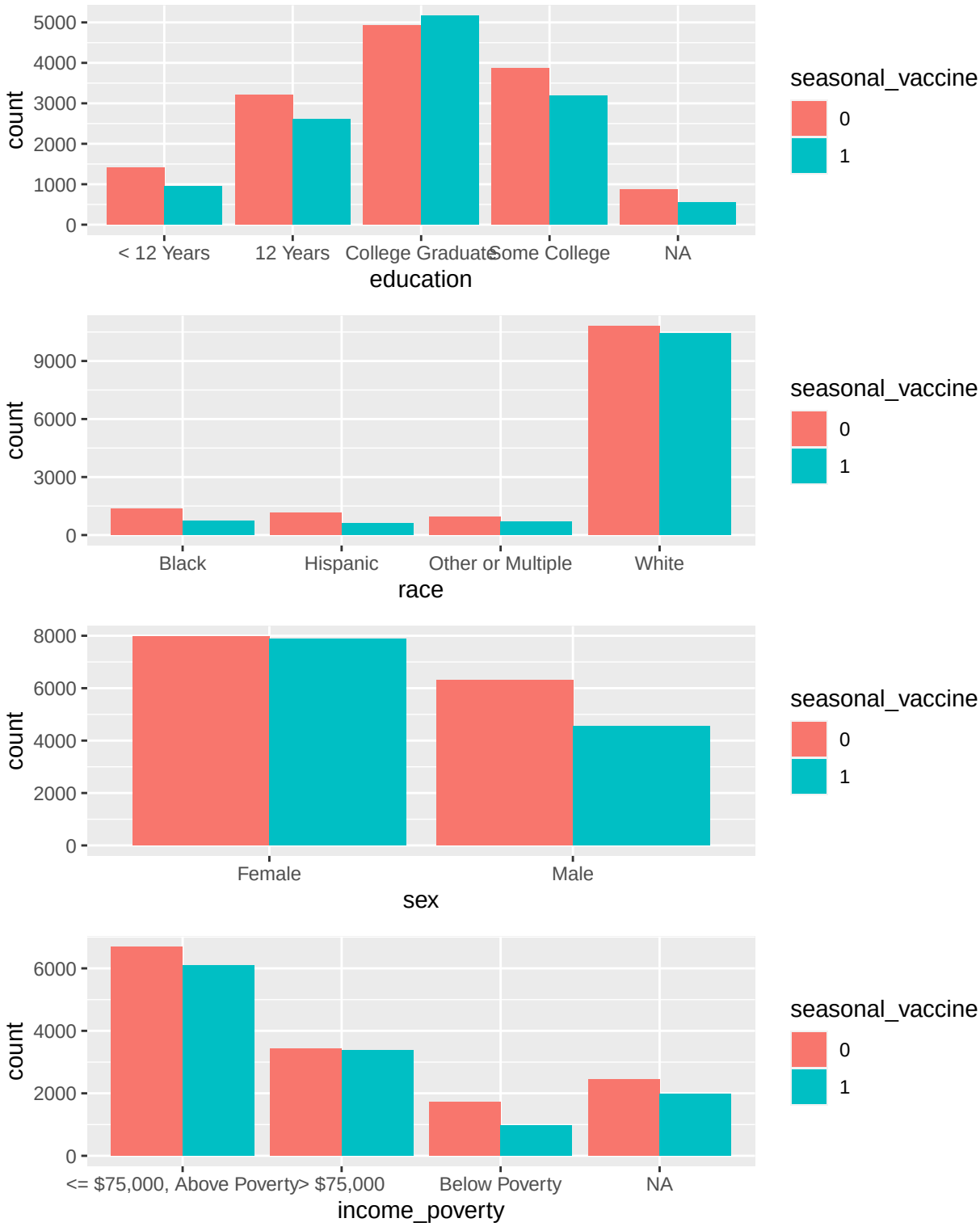

7500
5000
2500
0
count
0
1
NA
behavioral_touch_face
seasonal_vaccine
0
1

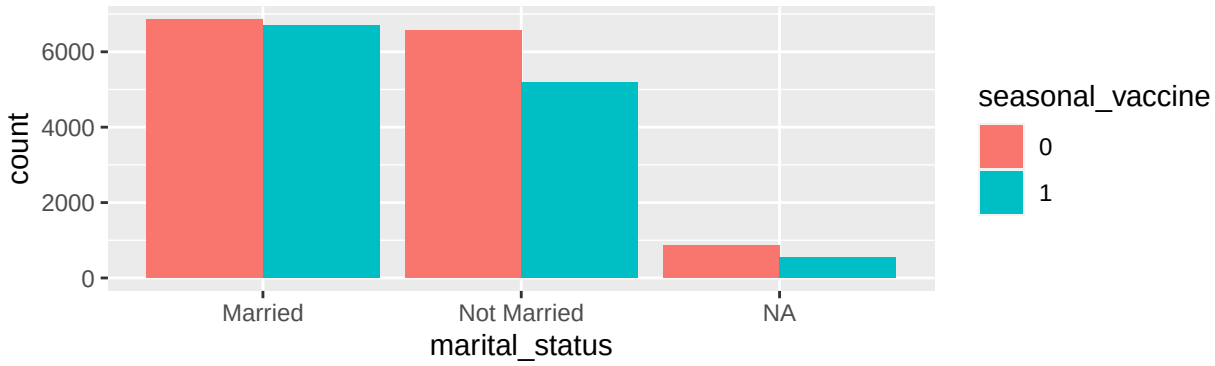

9000
6000
3000
0
count
0
1
NA
doctor_recc_h1n1
seasonal_vaccine
0
1

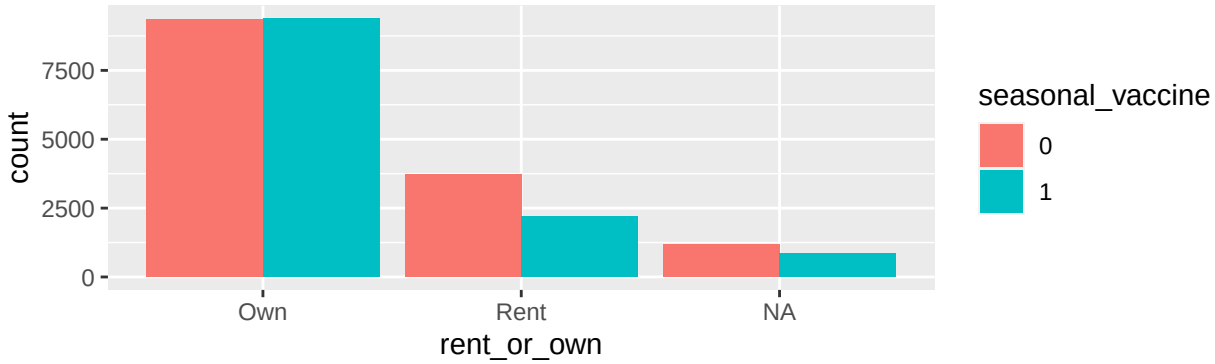

count
9000
6000
3000
0
0
1
NA
doctor_recc_seasonal
seasonal_vaccine
0
1
count
9000
6000
3000
0
0
1
NA
chronic_med_condition
seasonal_vaccine
0
1
count
10000
5000
0
0
1
NA
child_under_6_months
seasonal_vaccine
0
1
count
10000
5000
0
0
1
NA
health_worker
seasonal_vaccine
0
1

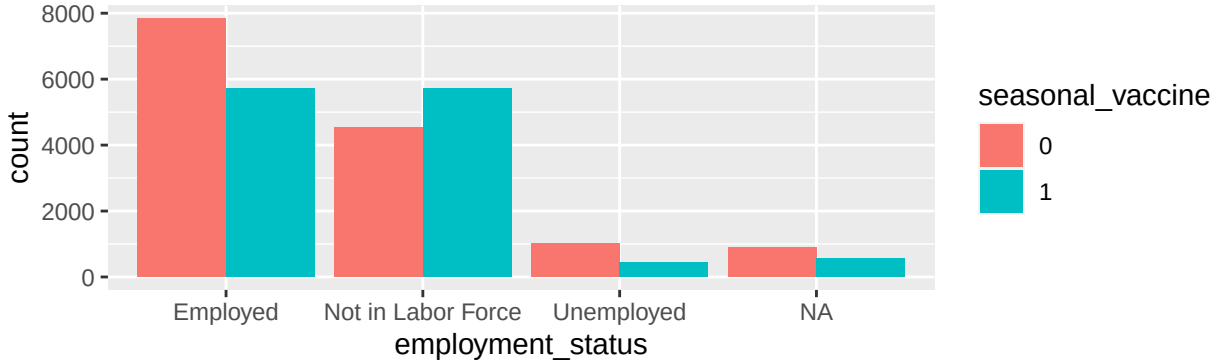

count
6000
4000
2000
0
0
1
NA
health_insurance
seasonal_vaccine
0
1
count
6000
4000
2000
0
1
2
3
4
5
NA
opinion_h1n1_vacc_effective
seasonal_vaccine
0
1
count
4000
2000
0
1
2
3
4
5
NA
opinion_h1n1_risk
seasonal_vaccine
0
1
count
5000
4000
3000
2000
1000
0
1
2
3
4
5
NA
opinion_h1n1_sick_from_vacc
seasonal_vaccine
0
1

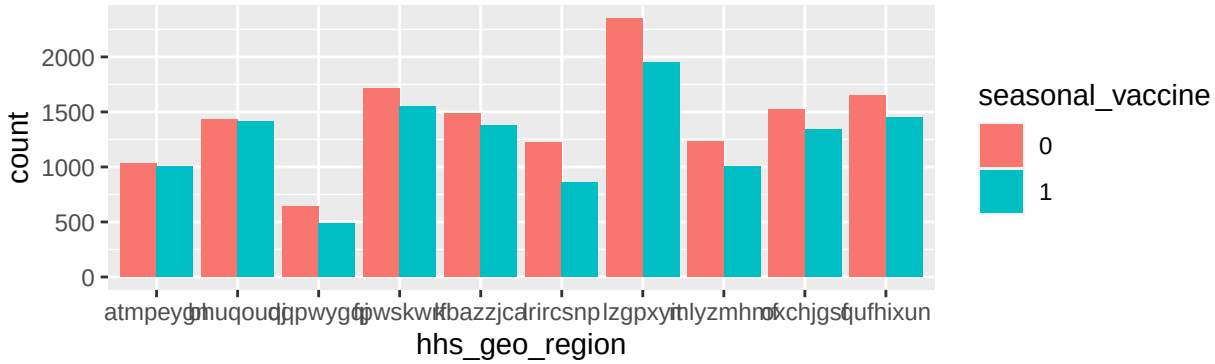

count
6000
4000
2000
0
1
2
3
4
5
NA
opinion_seas_vacc_effective
seasonal_vaccine
0
1
count
4000
2000
0
1
2
3
4
5
NA
opinion_seas_risk
seasonal_vaccine
0
1
count
6000
4000
2000
0
1
2
3
4
5
NA
opinion_seas_sick_from_vacc
seasonal_vaccine
0
1
count
4000
3000
2000
1000
0
18 − 34 Years
35 − 44 Years
45 − 54 Years
55 − 64 Years
65+ Years
age_group
seasonal_vaccine
0
1

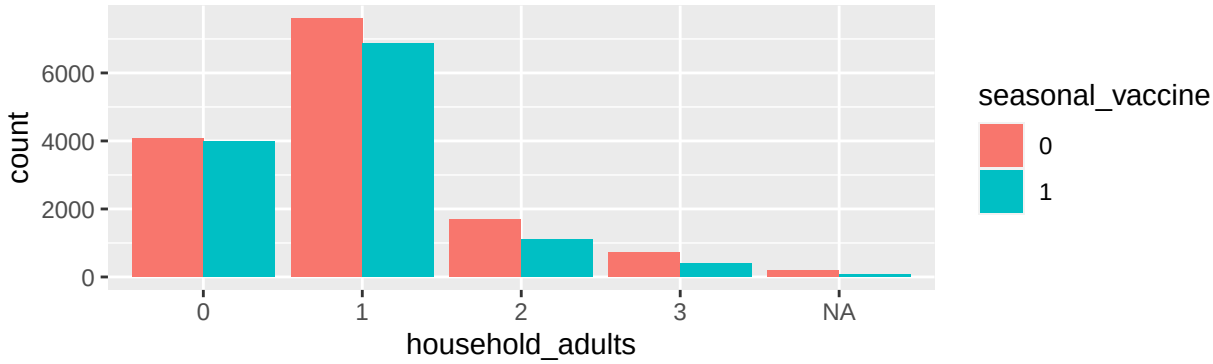

count
5000
4000
3000
2000
1000
0
< 12 Years
12 Years
College Graduate
Some College
NA
education
seasonal_vaccine
0
1
count
9000
6000
3000
0
Black
Hispanic
Other or Multiple
White
race
seasonal_vaccine
0
1
count
8000
6000
4000
2000
0
Female
Male
sex
seasonal_vaccine
0
1
count
6000
4000
2000
0
<= $75,000, Above Poverty
> $75,000
Below Poverty
NA
income_poverty
seasonal_vaccine
0
1

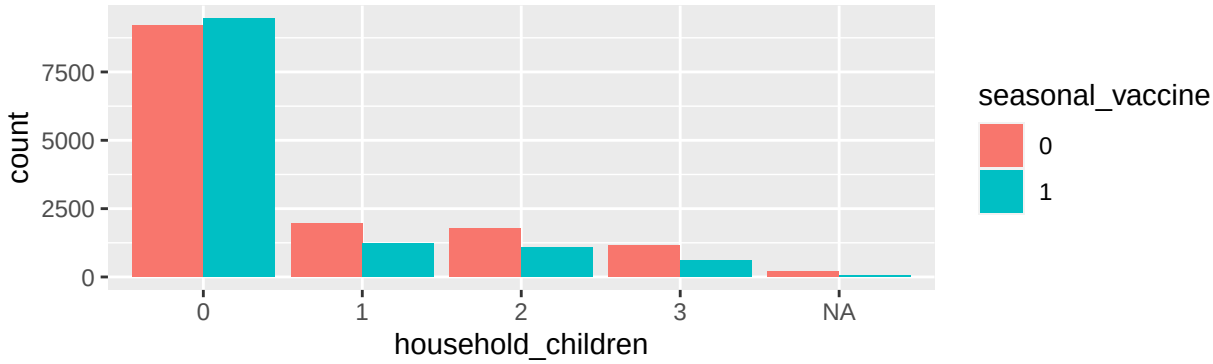

count
6000
4000
2000
0
Married
Not Married
NA
marital_status
seasonal_vaccine
0
1

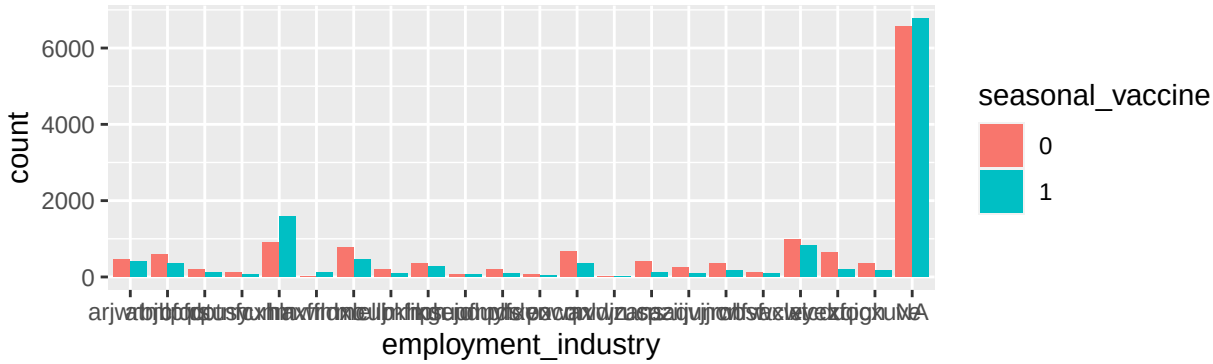

count
7500
5000
2500
0
Own
Rent
NA
rent_or_own
seasonal_vaccine
0
1


count
8000
6000
4000
2000
0
Employed
Not in Labor Force
Unemployed
NA
employment_status
seasonal_vaccine
0
1


count
2000
1500
1000
500
0
hhs_geo_region
seasonal_vaccine
0
1

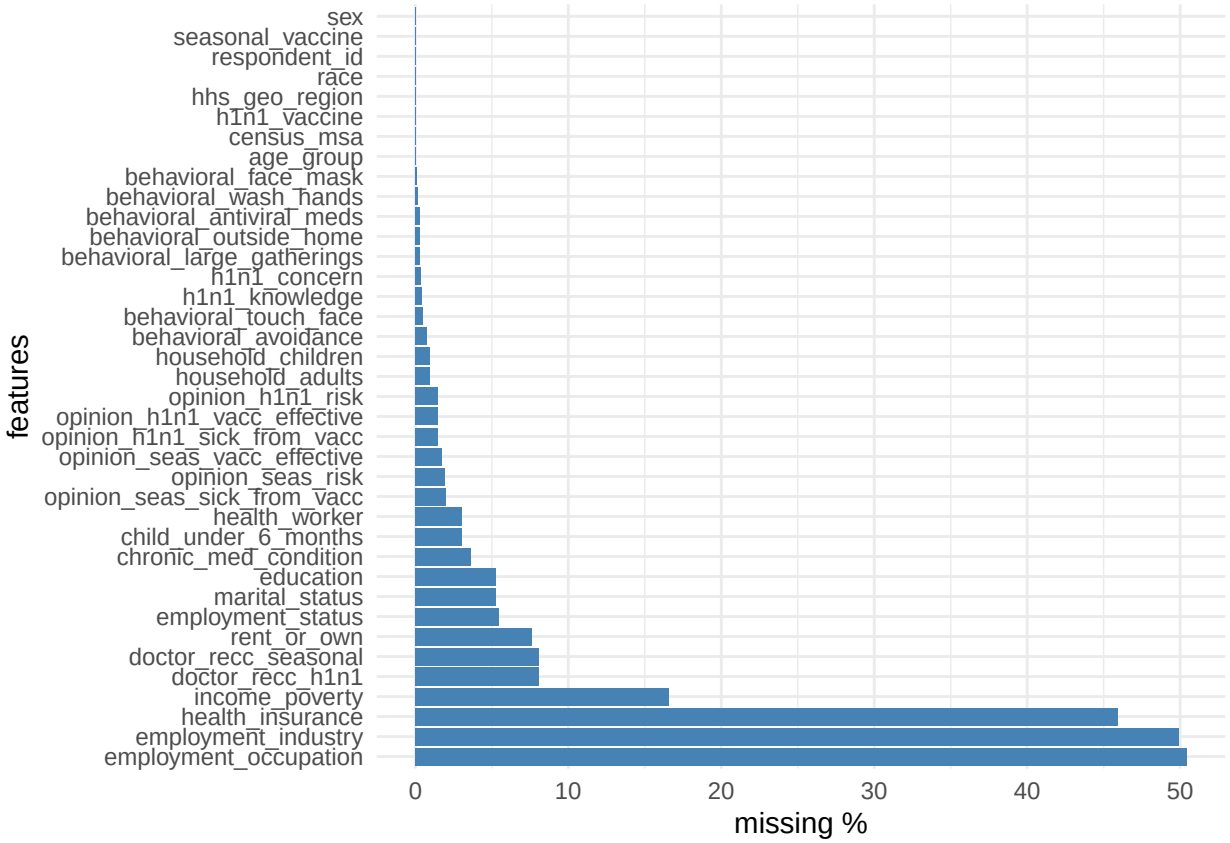






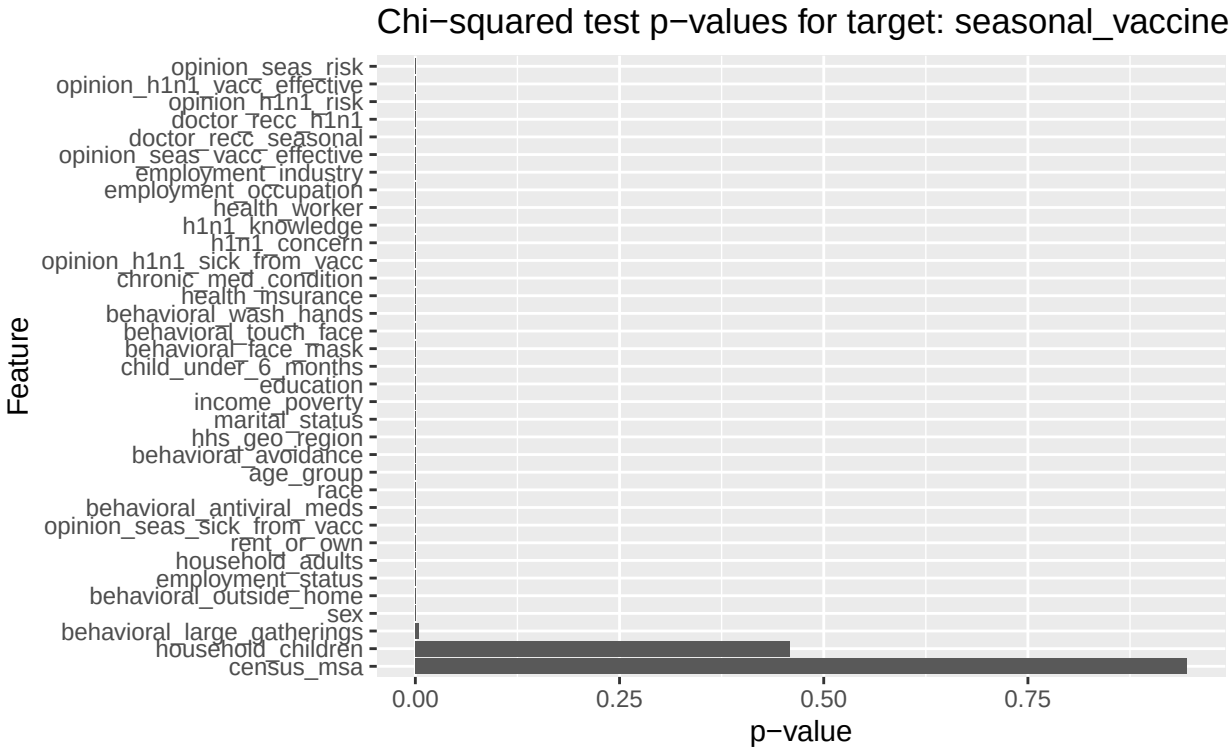

```
df_train %>% summarise_all(funs(sum(is.na(.))/n()*100)) %>%
  gather(key="feature", value="missing_pct") %>%
  ggplot(aes(x=reorder(feature,-missing_pct),y=missing_pct)) +
  geom_bar(stat="identity", fill="steelblue")+
  labs(y = "missing %", x = "features") +
```

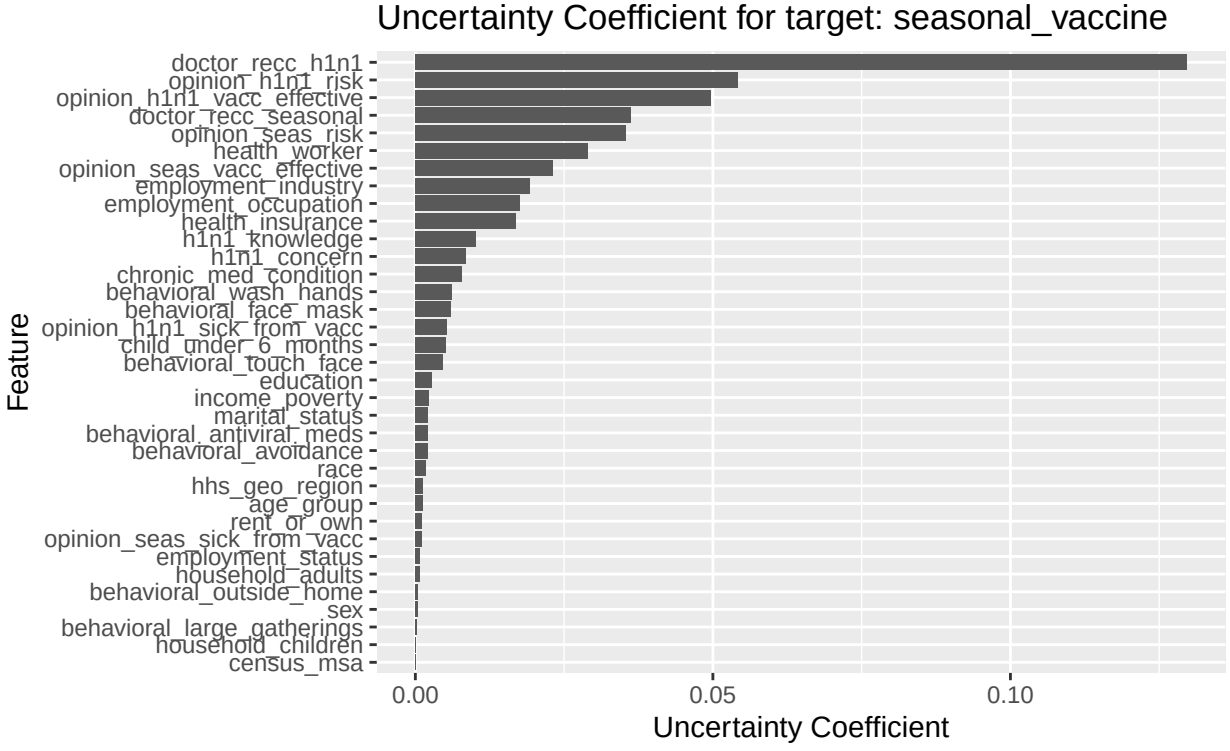
```
  coord_flip() +
  theme_minimal()
```

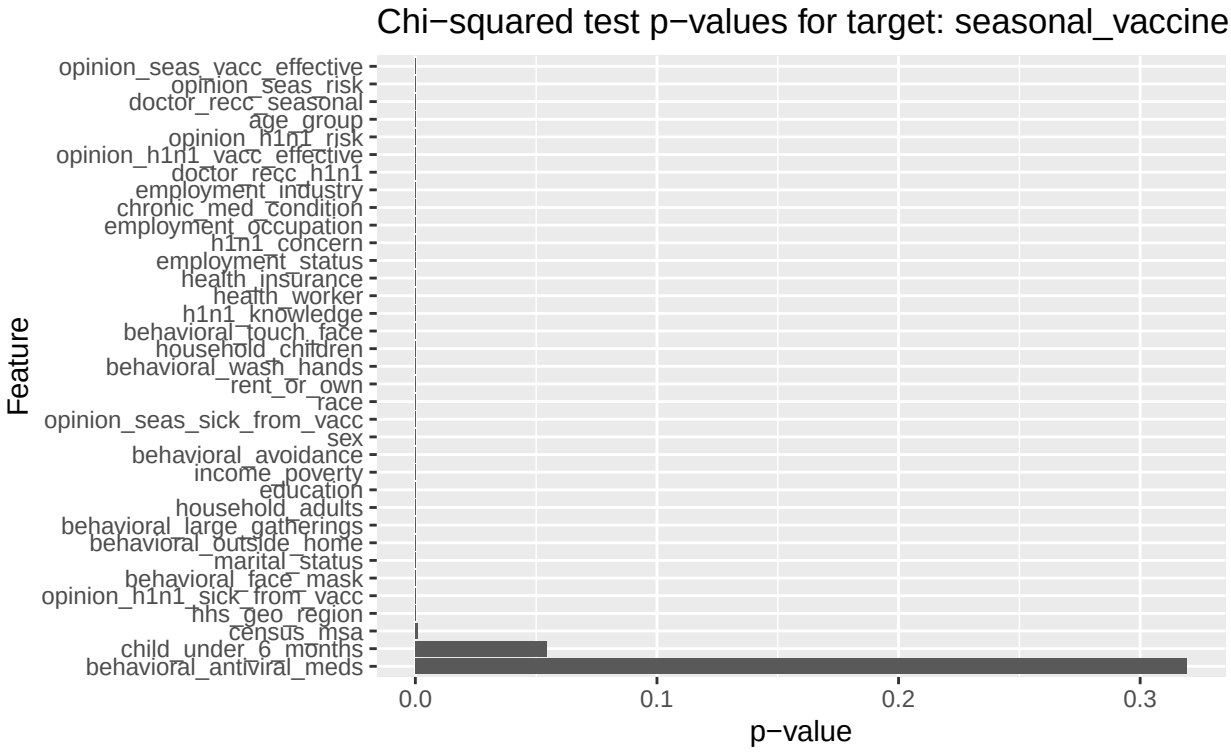


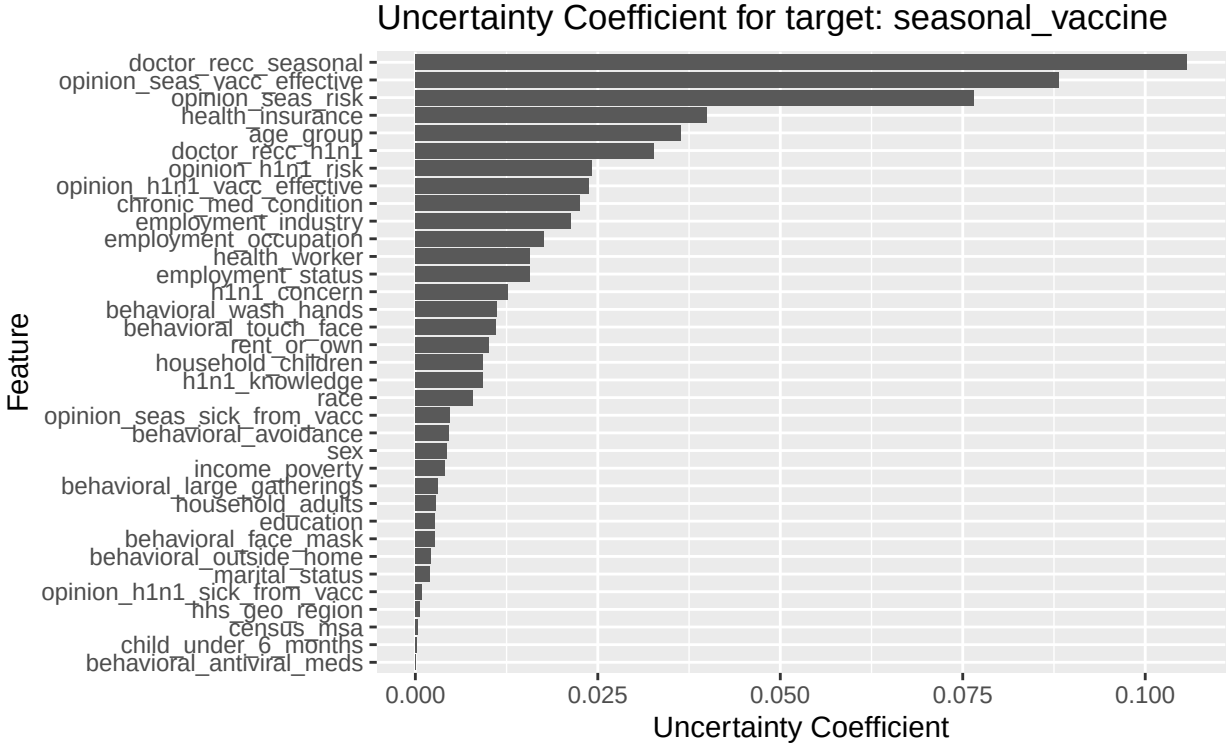

```
chisq_values = c()
uncert_values = c()
target_values = c()

for(target in targets){
  for(feature in features){
    contingency = df_train %>%
      select(target, feature) %>%
      table()

    p.value = chisq.test(contingency)$p.value

    target_values = c(target_values, target)
    chisq_values = c(chisq_values, p.value)
    uncert_values = c(uncert_values, UncertCoef(contingency))
  }
}

results = tibble(feature = rep(features, 2),
                 target = target_values,
                 p.value = chisq_values,
                 uncertainty = uncert_values)
```

```
for(target_value in targets){
  p1 = results %>%
    filter(target == target_value) %>%
    filter(p.value < 1) %>%
    ggplot(aes(x=reorder(feature, -p.value), y=p.value)) +
    geom_bar(stat='identity') +
    xlab("Feature") +
    ylab("p-value") +
    ggtitle(sprintf("Chi-squared test p-values for target: %s", target)) +
    coord_flip()

  p2 = results %>%
    filter(target == target_value) %>%
    filter(p.value < 1) %>%
    ggplot(aes(x=reorder(feature, uncertainty), y=uncertainty)) +
    geom_bar(stat='identity') +
    xlab("Feature") +
    ylab("Uncertainty Coefficient") +
    ggtitle(sprintf("Uncertainty Coefficient for target: %s", target)) +
    coord_flip()

  # print(p1)
  # print(p2)
  do.call("grid.arrange", c(list(p1, p2), nrow=2))
}
```


Chi−squared test p−values for target: seasonal_vaccine
opinion_seas_risk
opinion_h1n1_vacc_effective
opinion_h1n1_risk
doctor_recc_h1n1
doctor_recc_seasonal
opinion_seas_vacc_effective
employment_industry
employment_occupation
health_worker
h1n1_knowledge
h1n1_concern
opinion_h1n1_sick_from_vacc
chronic_med_condition
health_insurance
behavioral_wash_hands
behavioral_touch_face
behavioral_face_mask
child_under_6_months
education
income_poverty
marital_status
hhs_geo_region
behavioral_avoidance
age_group
race
behavioral_antiviral_meds
opinion_seas_sick_from_vacc
rent_or_own
household_adults
employment_status
behavioral_outside_home
sex
behavioral_large_gatherings
household_children
census_msa
Feature
0.00
0.25
0.50
0.75
p−value


Uncertainty Coefficient for target: seasonal_vaccine
doctor_recc_h1n1
opinion_h1n1_risk
opinion_h1n1_vacc_effective
doctor_recc_seasonal
opinion_seas_risk
health_worker
opinion_seas_vacc_effective
employment_industry
employment_occupation
health_insurance
h1n1_knowledge
h1n1_concern
chronic_med_condition
behavioral_wash_hands
behavioral_face_mask
opinion_h1n1_sick_from_vacc
child_under_6_months
behavioral_touch_face
education
income_poverty
marital_status
behavioral_antiviral_meds
behavioral_avoidance
race
hhs_geo_region
age_group
rent_or_own
opinion_seas_sick_from_vacc
employment_status
household_adults
behavioral_outside_home
sex
behavioral_large_gatherings
household_children
census_msa
Feature
0.00
0.05
0.10
Uncertainty Coefficient


Chi−squared test p−values for target: seasonal_vaccine
Feature
opinion_seas_vacc_effective
opinion_seas_risk
doctor_recc_seasonal
age_group
opinion_h1n1_risk
opinion_h1n1_vacc_effective
doctor_recc_h1n1
employment_industry
chronic_med_condition
employment_occupation
h1n1_concern
employment_status
health_insurance
health_worker
h1n1_knowledge
behavioral_touch_face
household_children
behavioral_wash_hands
rent_or_own
race
opinion_seas_sick_from_vacc
sex
behavioral_avoidance
income_poverty
education
household_adults
behavioral_large_gatherings
behavioral_outside_home
marital_status
behavioral_face_mask
opinion_h1n1_sick_from_vacc
hhs_geo_region
census_msa
child_under_6_months
behavioral_antiviral_meds
0.0
0.1
0.2
0.3
p−value


Uncertainty Coefficient for target: seasonal_vaccine
Feature
doctor_recc_seasonal
opinion_seas_vacc_effective
opinion_seas_risk
health_insurance
age_group
doctor_recc_h1n1
opinion_h1n1_risk
opinion_h1n1_vacc_effective
chronic_med_condition
employment_industry
employment_occupation
health_worker
employment_status
h1n1_concern
behavioral_wash_hands
behavioral_touch_face
rent_or_own
household_children
h1n1_knowledge
race
opinion_seas_sick_from_vacc
behavioral_avoidance
sex
income_poverty
behavioral_large_gatherings
household_adults
education
behavioral_face_mask
behavioral_outside_home
marital_status
opinion_h1n1_sick_from_vacc
hhs_geo_region
census_msa
child_under_6_months
behavioral_antiviral_meds
0.000
0.025
0.050
0.075
0.100
Uncertainty Coefficient