BRFSS Data

Ethan Walker

July 29, 2019

Initial look, clean oregon_brfss_2014

```
head(oregon_brfss_2014, 5)
```

```
## # A tibble: 5 x 15
##    SEQNO `@_AGE_G`   SEX EDUCA EMPLOY1 MARITAL VETERAN3 `@_CASTHM1`
##    <int>     <int> <int> <chr> <chr>   <chr>   <chr>          <int>
## 1 2.01e9         5     2 4     5       1       2                  1
## 2 2.01e9         4     1 3     8       2       2                  2
## 3 2.01e9         2     1 6     1       1       2                  1
## 4 2.01e9         6     2 4     2       1       2                  1
## 5 2.01e9         3     2 6     1       1       2                  1
## # ... with 7 more variables: PNEUVAC3 <chr>, SMOKE100 <chr>,
## #   SMOKDAY2 <chr>, `@_RFSMOK3` <int>, USENOW3 <chr>, `@_BMI5` <chr>,
## #   `@_BMI5CAT` <chr>
```

```
clean_data_oregon <- oregon_brfss_2014 %>%
  mutate(id = SEQNO,
         age_cat = `@_AGE_G`,
         sex = SEX,
         education_cat = EDUCA,
         employment_cat = EMPLOY1,
         marital_cat = MARITAL,
         veteran = VETERAN3,
         asthma = `@_CASTHM1`,
         pneum_vax = PNEUVAC3,
         smoke_100 = SMOKE100,
         smoke_now = SMOKDAY2,
         tobacco_now = USENOW3,
         current_smoker = `@_RFSMOK3`,
         bmi = `@_BMI5`,
         bmi_cat = `@_BMI5CAT`) %>%
  select(id, age_cat, sex, education_cat, employment_cat, marital_cat, veteran,
         asthma, pneum_vax, smoke_100, smoke_now, tobacco_now, current_smoker,
         bmi, bmi_cat) %>%
  mutate(age_cat = factor(age_cat, levels = c(1,2,3,4,5,6),
                          labels = c("18-24", "25-34", "35-44", "45-54", "55-64", "65+")),
         sex = factor(sex, levels = c(1,2), labels = c("male", "female")),
         education_cat = factor(education_cat, levels = c(1,2,3,4,5,6),
                          labels = c("none", "elementary", "some high school",
                                     "high school grad", "some college", "college grad")),
         employment_cat = factor(employment_cat, levels = c(1,2,3,4,5,6,7,8),
                          labels = c("employed for wages", "self-employed", "out of work 1yr+",
                                     "out of work <1yr", "homemaker", "student",
                                     "retired", "unable to work")),
         marital_cat = factor(marital_cat, levels = c(1,2,3,4,5,6),
```

```
                          labels = c("married", "divorced", "widowed",
                                     "separated", "never married", "unmarried couple")),
         veteran = factor(veteran, levels = c(1,2), labels = c("yes", "no")),
         asthma = factor(asthma, levels = c(1,2), labels = c("no", "yes")),
         pneum_vax = factor(pneum_vax, levels = c(1,2), labels = c("yes", "no")),
         smoke_100 = factor(smoke_100, levels = c(1,2), labels = c("yes", "no")),
         smoke_now = factor(smoke_now, levels = c(1,2,3),
                            labels = c("every day", "some days", "not at all")),
         tobacco_now = factor(tobacco_now, levels = c(1,2,3),
                            labels = c("every day", "some days", "not at all")),
         current_smoker = factor(current_smoker, levels = c(1,2), labels = c("no", "yes")),
         bmi = as.numeric(bmi),
         bmi_cat = factor(bmi_cat, levels = c(1,2,3,4),
                          labels = c("underweight", "normal", "overweight", "obese")))

head(clean_data_oregon, 5)
```

```
## # A tibble: 5 x 15
##       id age_cat sex   education_cat employment_cat marital_cat veteran
##    <int> <fct>   <fct> <fct>         <fct>          <fct>       <fct>
## 1 2.01e9 55-64   fema~ high school ~ homemaker      married     no
## 2 2.01e9 45-54   male  some high sc~ unable to work divorced    no
## 3 2.01e9 25-34   male  college grad  employed for ~ married     no
## 4 2.01e9 65+     fema~ high school ~ self-employed  married     no
## 5 2.01e9 35-44   fema~ college grad  employed for ~ married     no
## # ... with 8 more variables: asthma <fct>, pneum_vax <fct>,
## #   smoke_100 <fct>, smoke_now <fct>, tobacco_now <fct>,
## #   current_smoker <fct>, bmi <dbl>, bmi_cat <fct>
```

Check data

Compare numbers in each factor to make sure levels/NAs were coded correctly

```
table(clean_data_oregon$age_cat)
```

```
##
## 18-24 25-34 35-44 45-54 55-64   65+
##    56    98   120   144   218   364
```

```
table(oregon_brfss_2014$`@_AGE_G`)
```

```
##
##   1   2   3   4   5   6
##  56  98 120 144 218 364
```

```
table(clean_data_oregon$sex)
```

```
##
##   male female
##    407    593
```

```
table(oregon_brfss_2014$SEX)
```

```
##
```

```
##   1   2
## 407 593
```

```
table(clean_data_oregon$education_cat)
```

```
##
##             none       elementary some high school high school grad
##                3               10               44              220
##     some college     college grad
##              328              390
```

```
table(oregon_brfss_2014$EDUCA)
```

```
##
##   .   1   2   3   4   5   6   9
##   2   3  10  44 220 328 390   3
```

```
table(clean_data_oregon$employment_cat)
```

```
##
## employed for wages      self-employed   out of work 1yr+
##                367                 93                 31
##   out of work <1yr          homemaker            student
##                 17                 51                 31
##            retired     unable to work
##                346                 51
```

```
table(oregon_brfss_2014$EMPLOY1)
```

```
##
##   .   1   2   3   4   5   6   7   8   9
##   4 367  93  31  17  51  31 346  51   9
```

```
table(clean_data_oregon$marital_cat)
```

```
##
##          married         divorced          widowed        separated
##              513              152              125               16
##    never married unmarried couple
##              145               41
```

```
table(oregon_brfss_2014$MARITAL)
```

```
##
##   .   1   2   3   4   5   6   9
##   1 513 152 125  16 145  41   7
```

```
table(clean_data_oregon$veteran)
```

```
##
## yes  no
## 124 874
```

```
table(oregon_brfss_2014$VETERAN3)
```

```
##
##   .   1   2   9
##   1 124 874   1
```

```r
table(clean_data_oregon$asthma)
```

```
##
##  no yes
## 898  95
```

```r
table(oregon_brfss_2014$`@_CASTHM1`)
```

```
##
##   1   2   9
## 898  95   7
```

```r
table(clean_data_oregon$pneum_vax)
```

```
##
## yes  no
## 390 451
```

```r
table(oregon_brfss_2014$PNEUVAC3)
```

```
##
##   .   1   2   7   9
##  47 390 451 110   2
```

```r
table(clean_data_oregon$smoke_100)
```

```
##
## yes  no
## 442 512
```

```r
table(oregon_brfss_2014$SMOKE100)
```

```
##
##   .   1   2   7   9
##  37 442 512   7   2
```

```r
table(clean_data_oregon$smoke_now)
```

```
##
##  every day  some days not at all
##         90         30        322
```

```r
table(oregon_brfss_2014$SMOKDAY2)
```

```
##
##   .   1   2   3
## 558  90  30 322
```

```r
table(clean_data_oregon$tobacco_now)
```

```
##
##  every day  some days not at all
##         15          8        937
```

```r
table(oregon_brfss_2014$USENOW3)
```

```
##
##   .   1   2   3   9
##  38  15   8 937   2
```

```r
table(clean_data_oregon$current_smoker)
```

```
##
##  no yes
## 834 120
```

```r
table(oregon_brfss_2014$`@_RFSMOK3`)
```

```
##
##   1   2   9
## 834 120  46
```

```r
table(clean_data_oregon$bmi_cat)
```

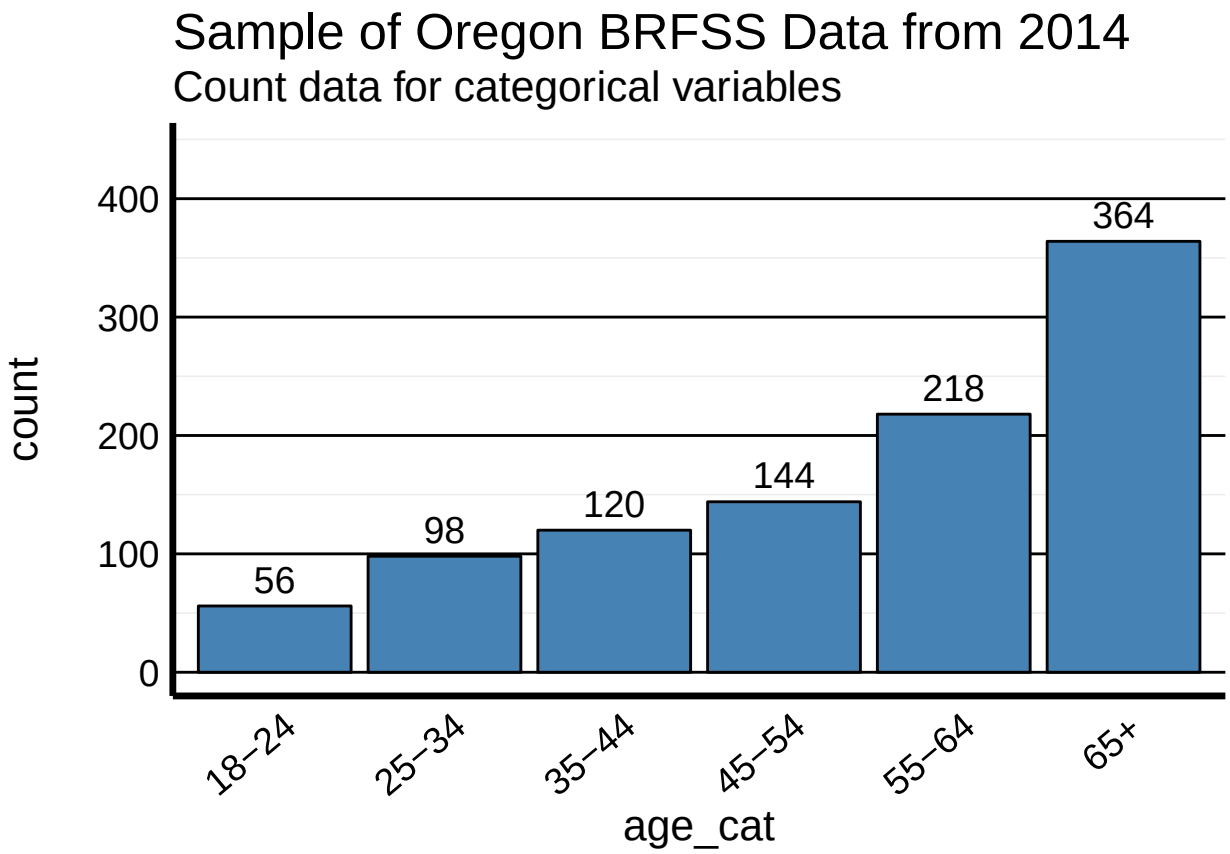
```
##
## underweight      normal  overweight       obese
##          18         322         336         258
```

```r
table(oregon_brfss_2014$`@_BMI5CAT`)
```

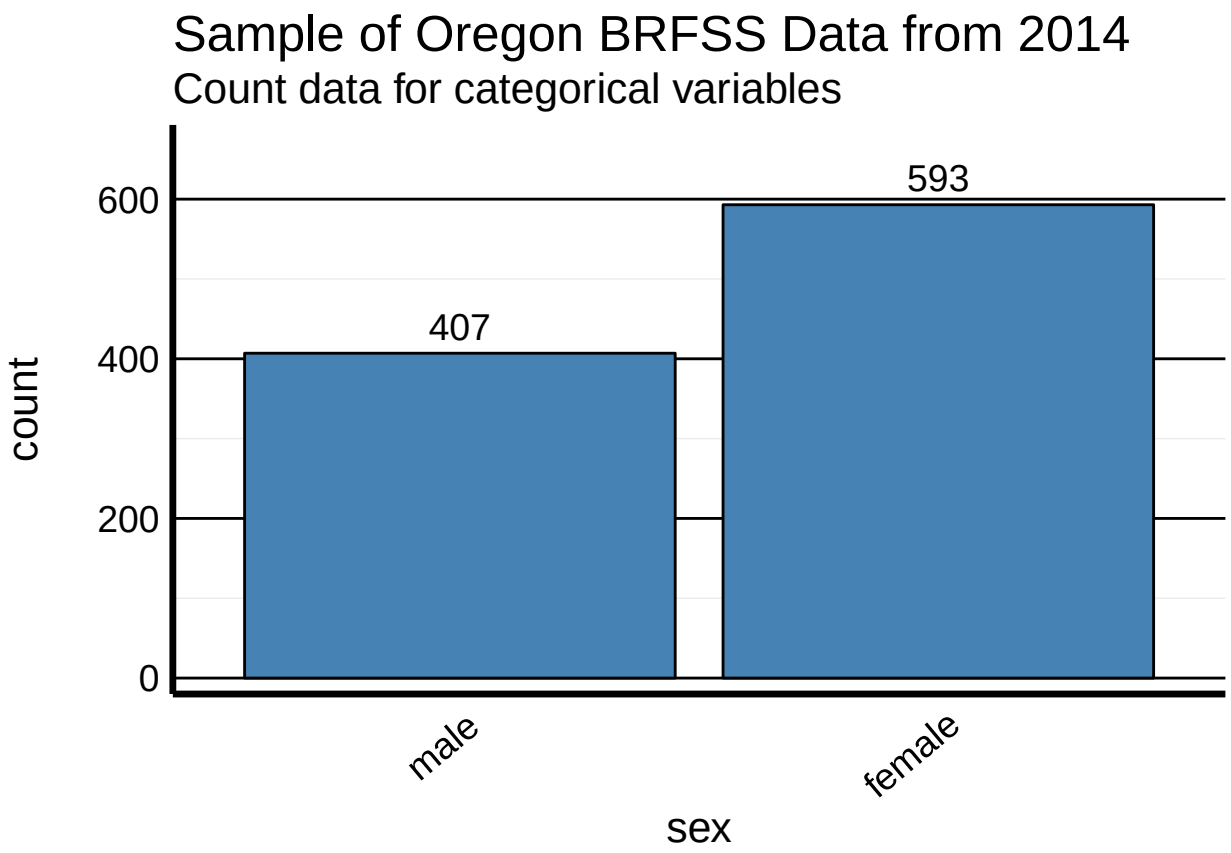
```
##
##   .   1   2   3   4
##  66  18 322 336 258
```

Function for plotting each categorical variable

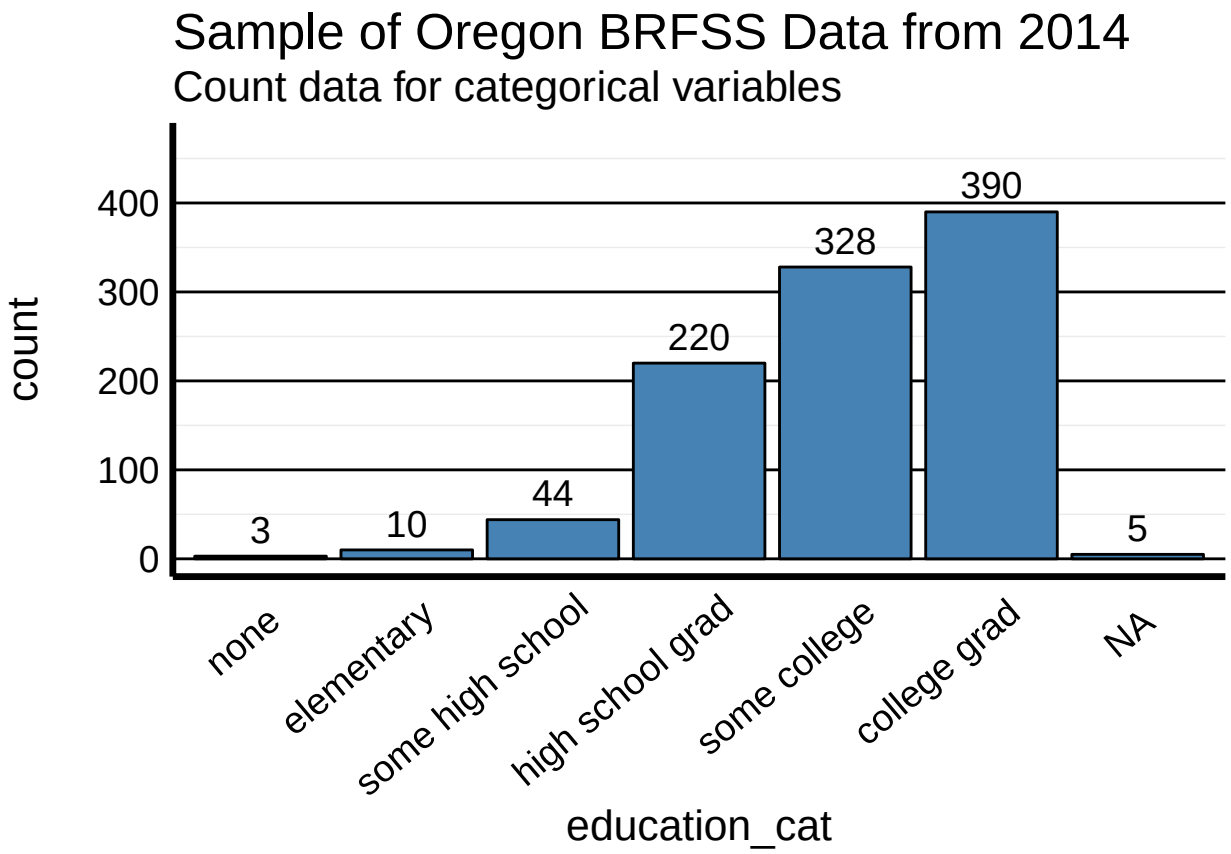
```r
hist_function <- function(df = clean_data_oregon, var) {
hist_plot <- df %>%
  rename(result = var) %>%
  ggplot(aes(result)) +
    geom_bar(stat = "count", fill = "steelblue", color = "black") +
    geom_text(stat = "count", aes(label = ..count..), vjust = -0.5, size = 5) +
    ggtitle("Sample of Oregon BRFSS Data from 2014",
            subtitle = "Count data for categorical variables") +
    xlab(var) +
    theme_minimal() +
    scale_y_continuous(expand = expand_scale(add = c(20, 100))) +
    theme(axis.line = element_line(size = 1.2),
          panel.grid.major.y = element_line(size = 0.5, color = "black"),
          panel.grid.major.x = element_blank(),
          axis.text.x = element_text(size = 14, color = "black", angle = 40,
                                     hjust = .9, vjust = .9),
          axis.text.y = element_text(size = 14, color = "black"),
          axis.title.y = element_text(size = 16,
                                      margin = margin(t = 0, r = 20, b = 0, l = 0)),
          title = element_text(size = 16, color = "black"))
hist_plot
}

hist_function(clean_data_oregon, var = "age_cat")
```



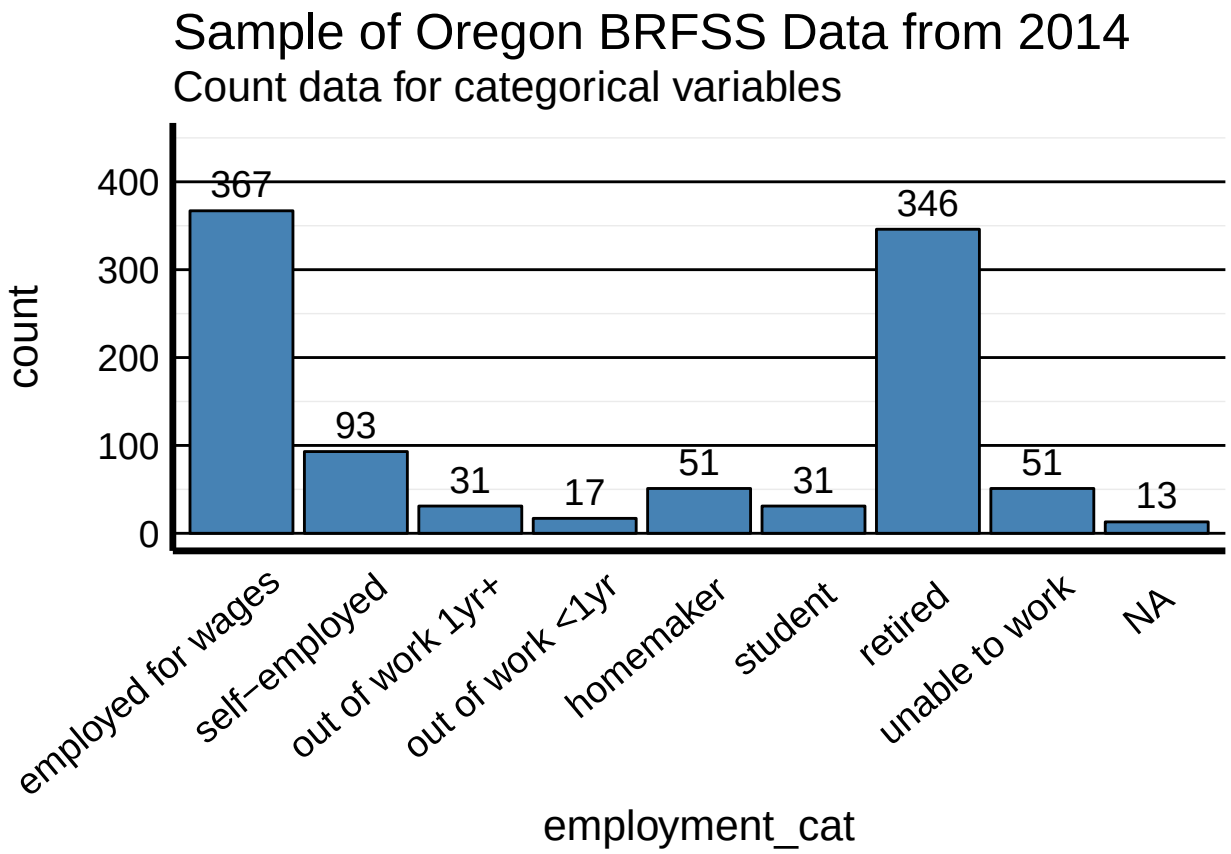

```
hist_function(clean_data_oregon, var = "sex")
```



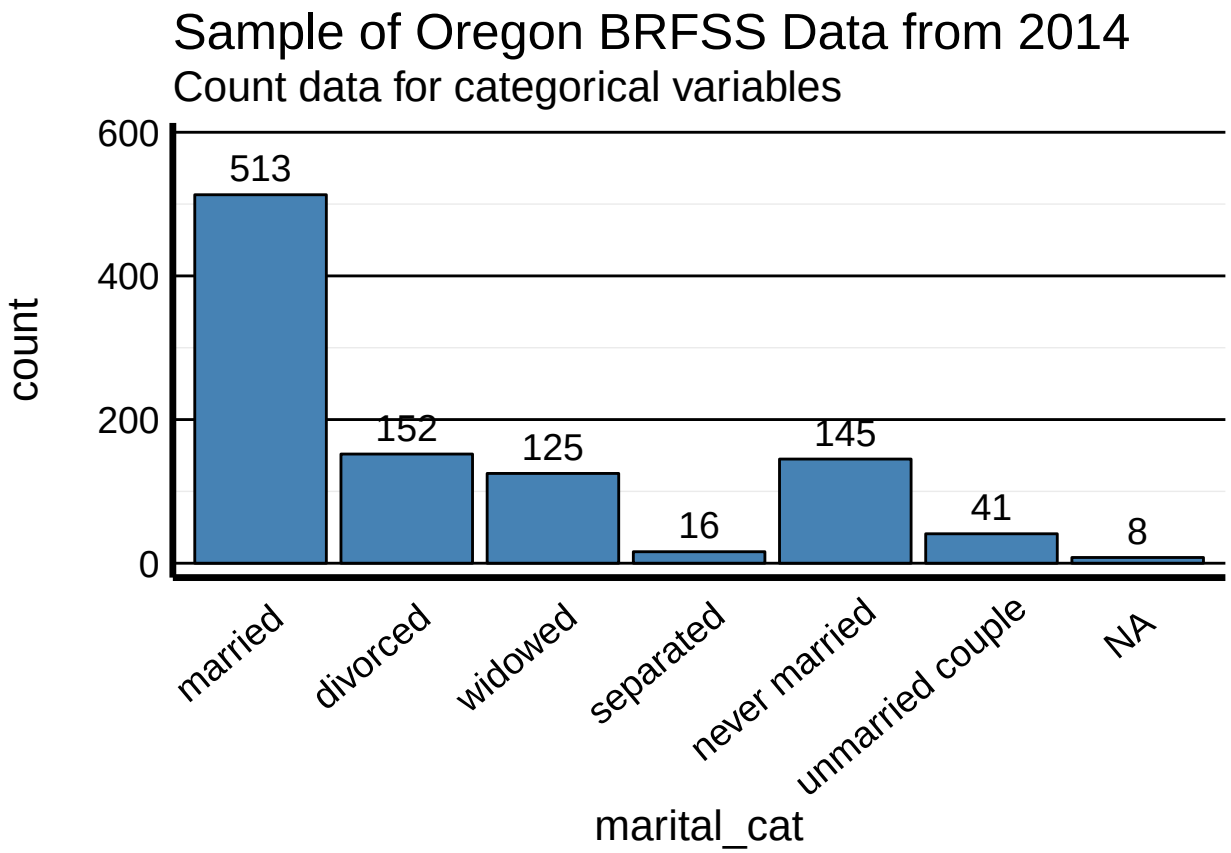

```
hist_function(clean_data_oregon, var = "education_cat")
```



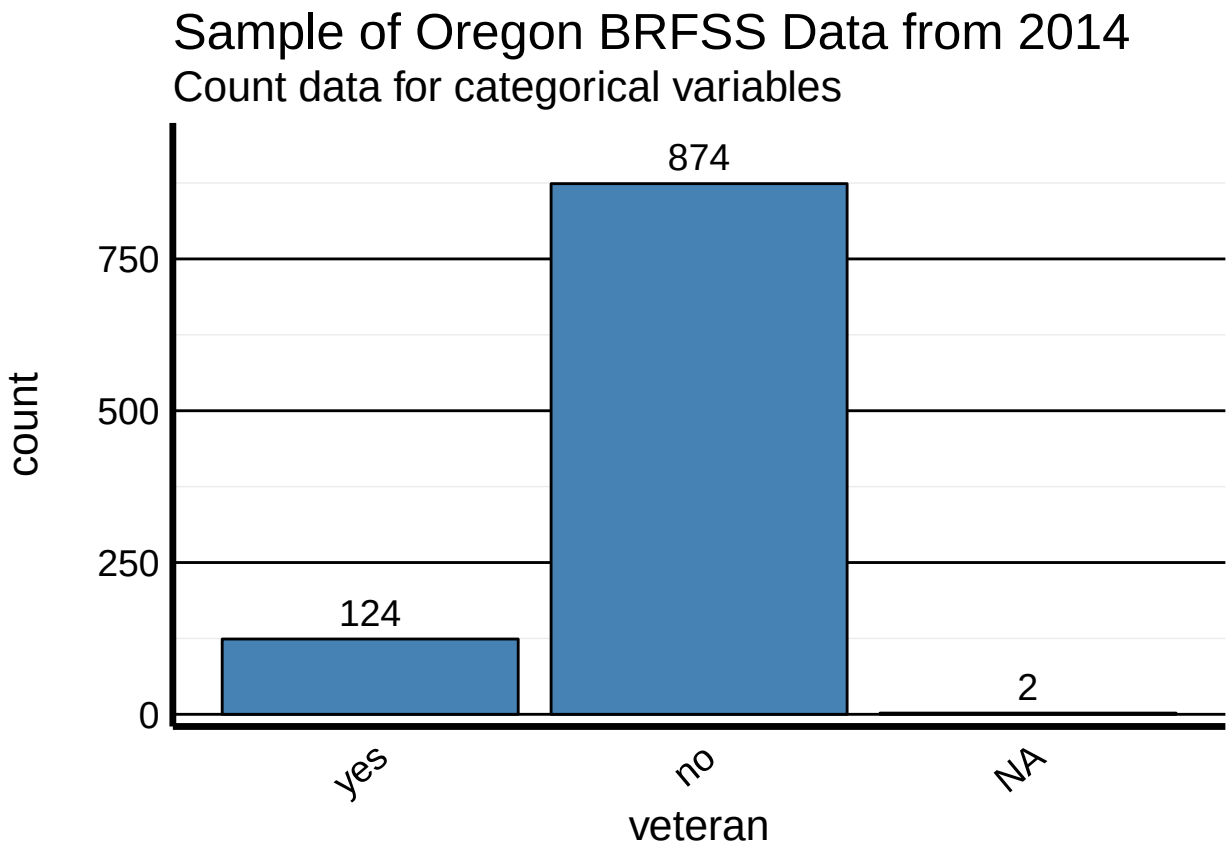

```
hist_function(clean_data_oregon, var = "employment_cat")
```



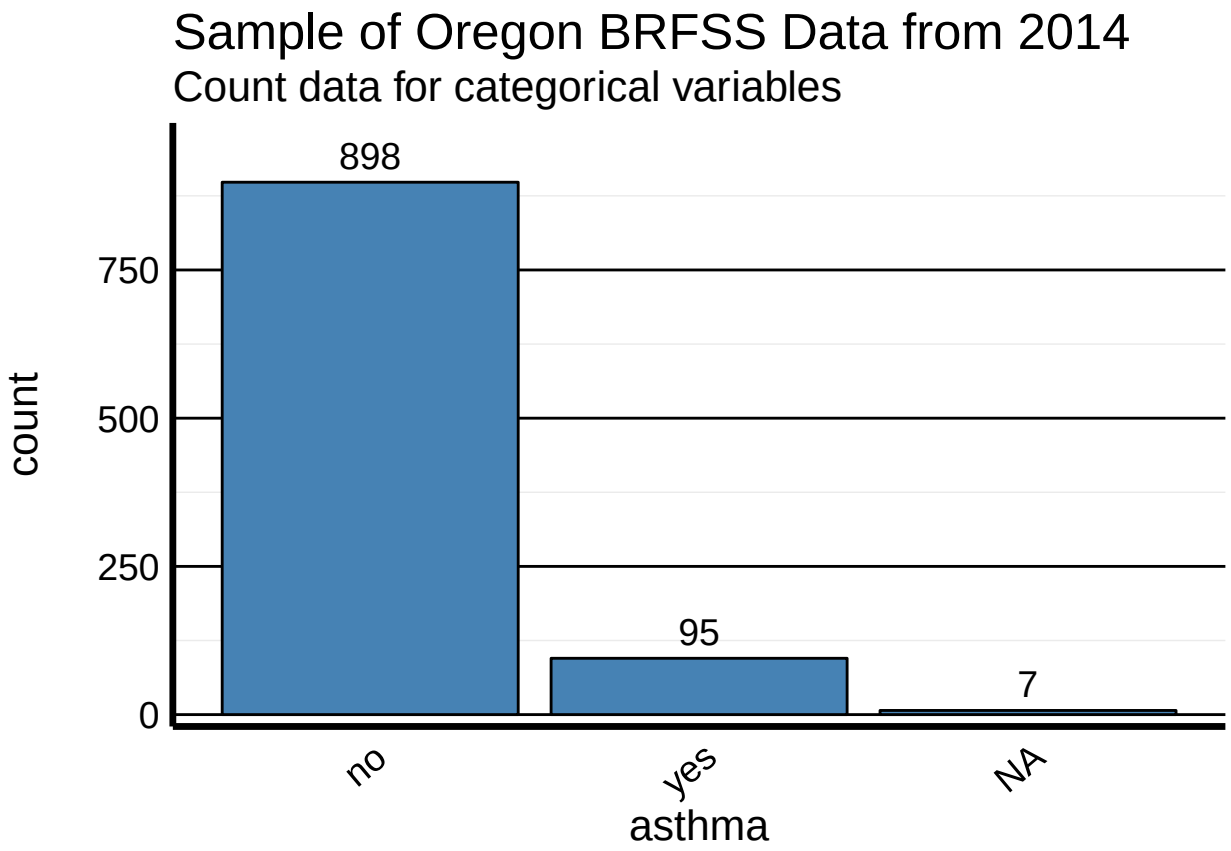

```
hist_function(clean_data_oregon, var = "marital_cat")
```



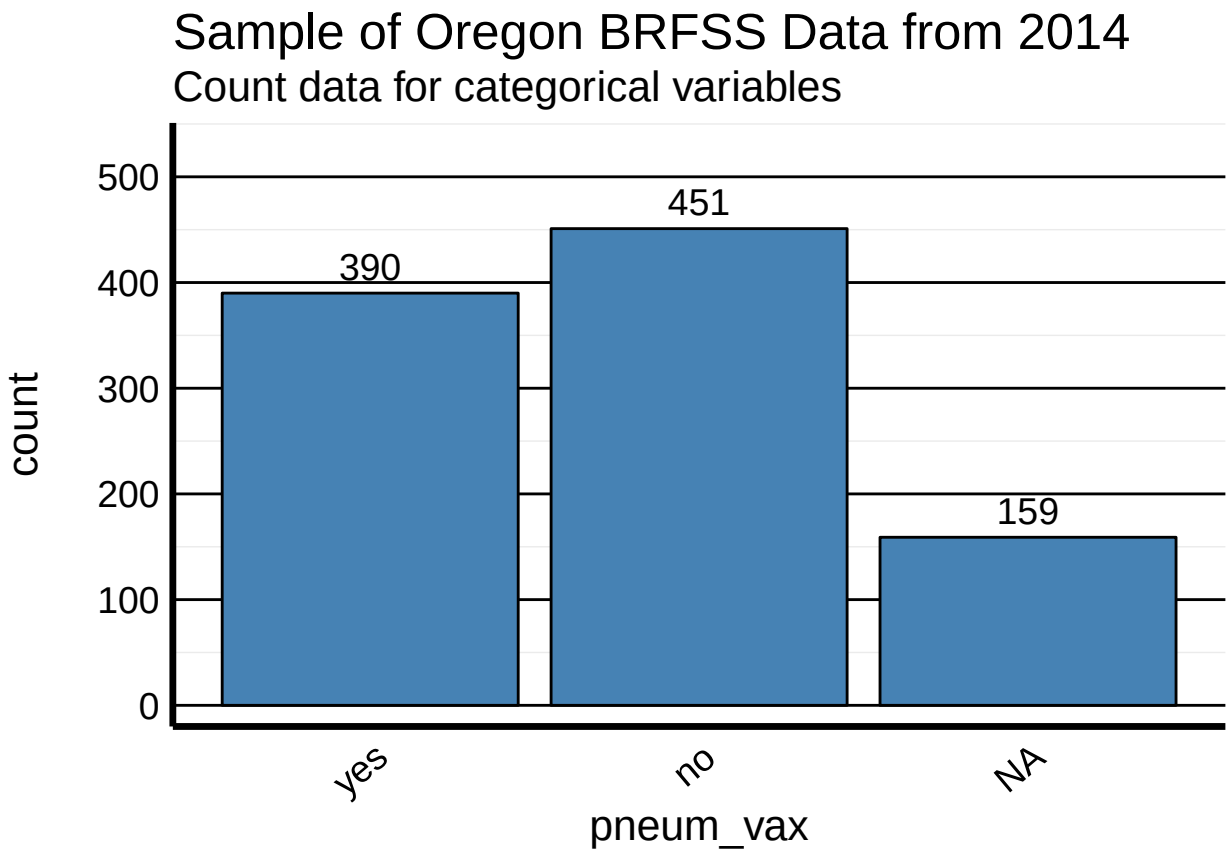

```
hist_function(clean_data_oregon, var = "veteran")
```



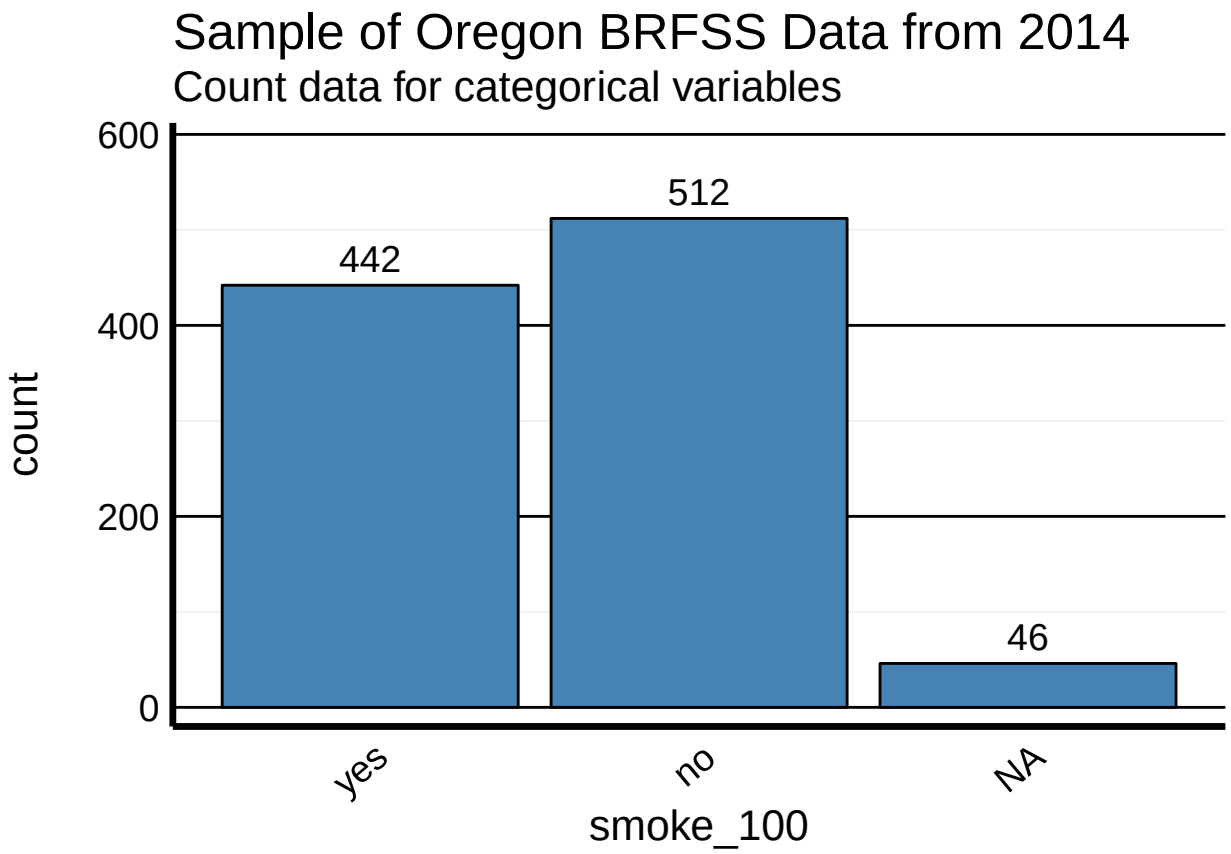

```
hist_function(clean_data_oregon, var = "asthma")
```



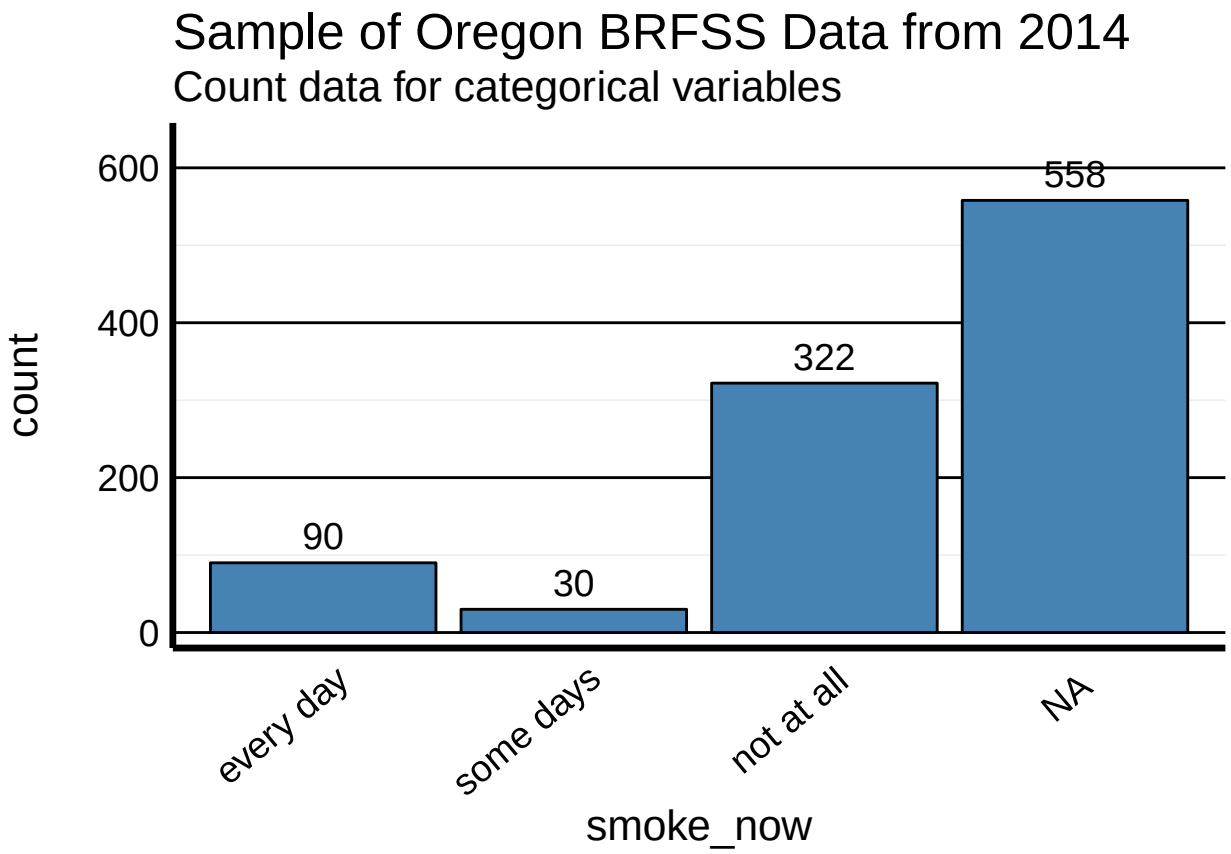

```
hist_function(clean_data_oregon, var = "pneum_vax")
```



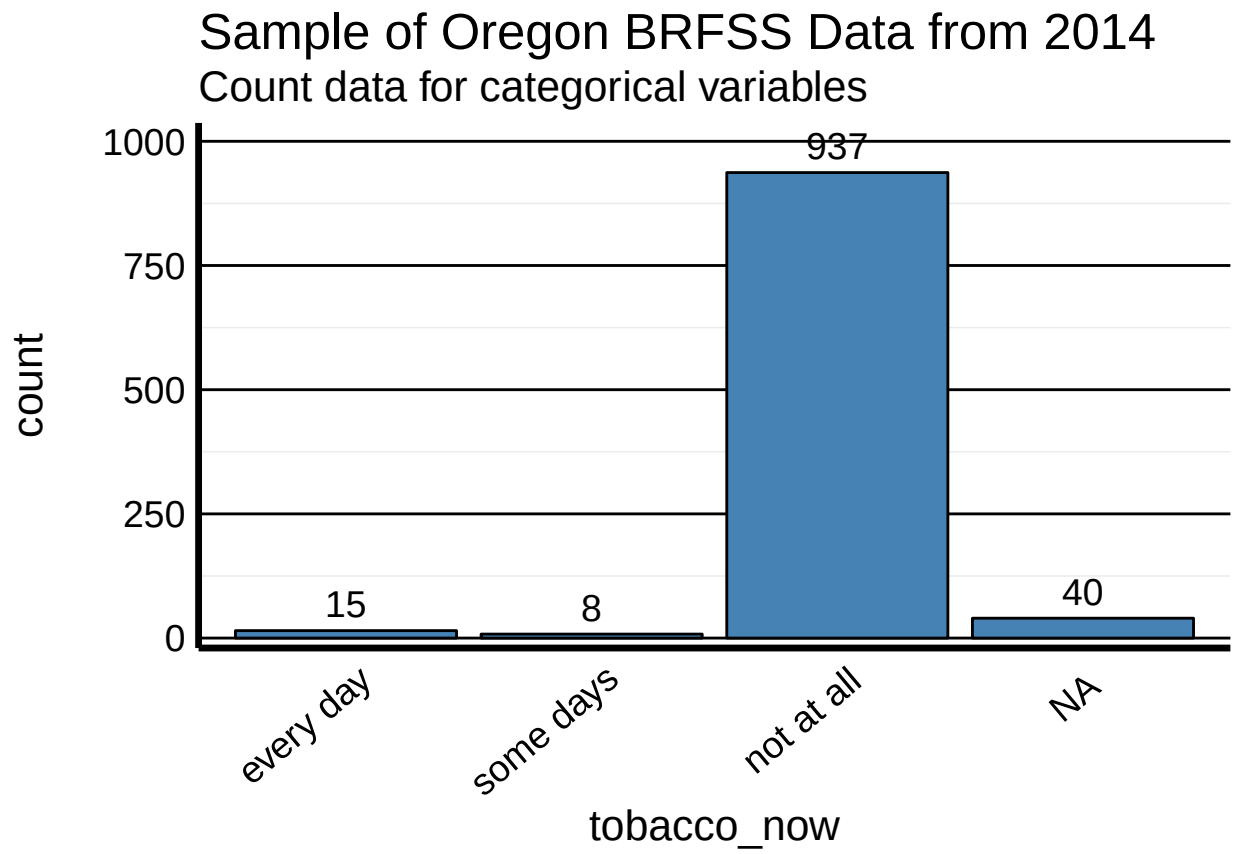

```
hist_function(clean_data_oregon, var = "smoke_100")
```



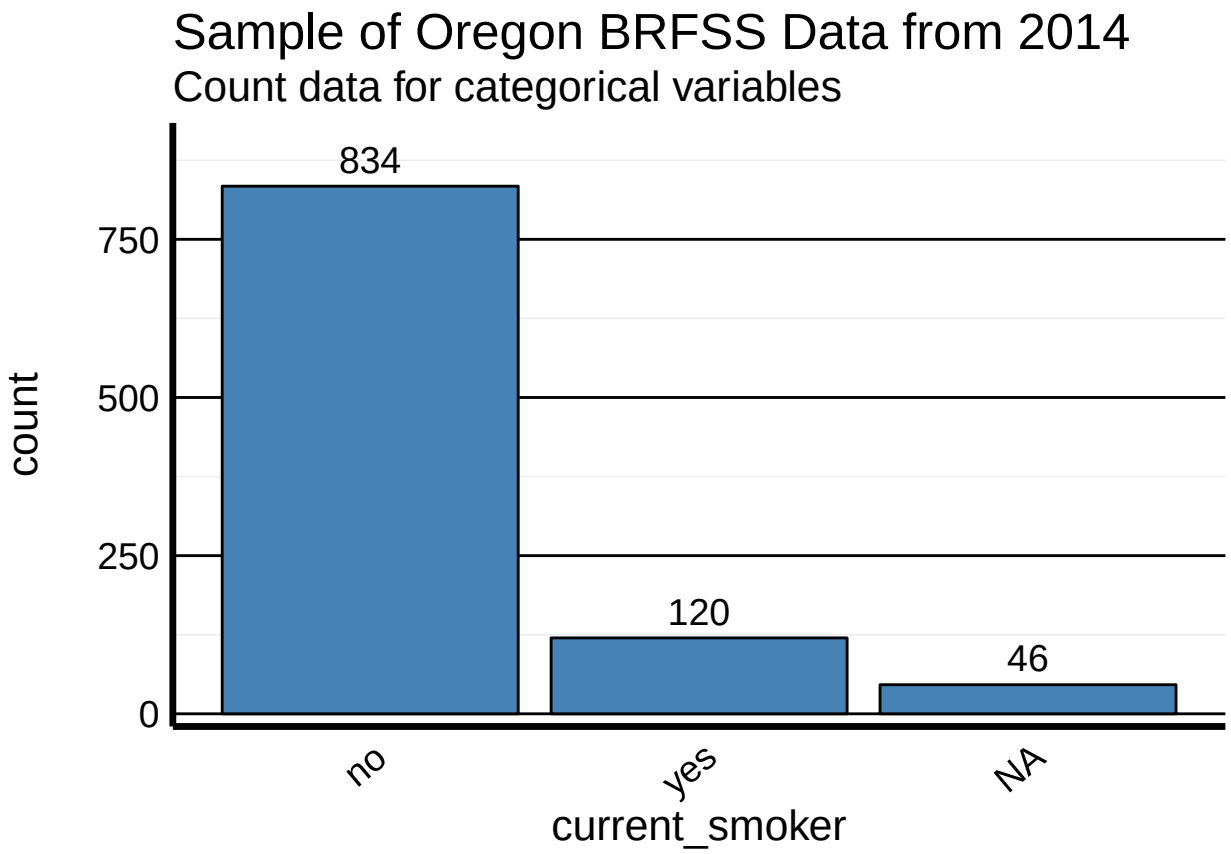

```
hist_function(clean_data_oregon, var = "smoke_now")
```



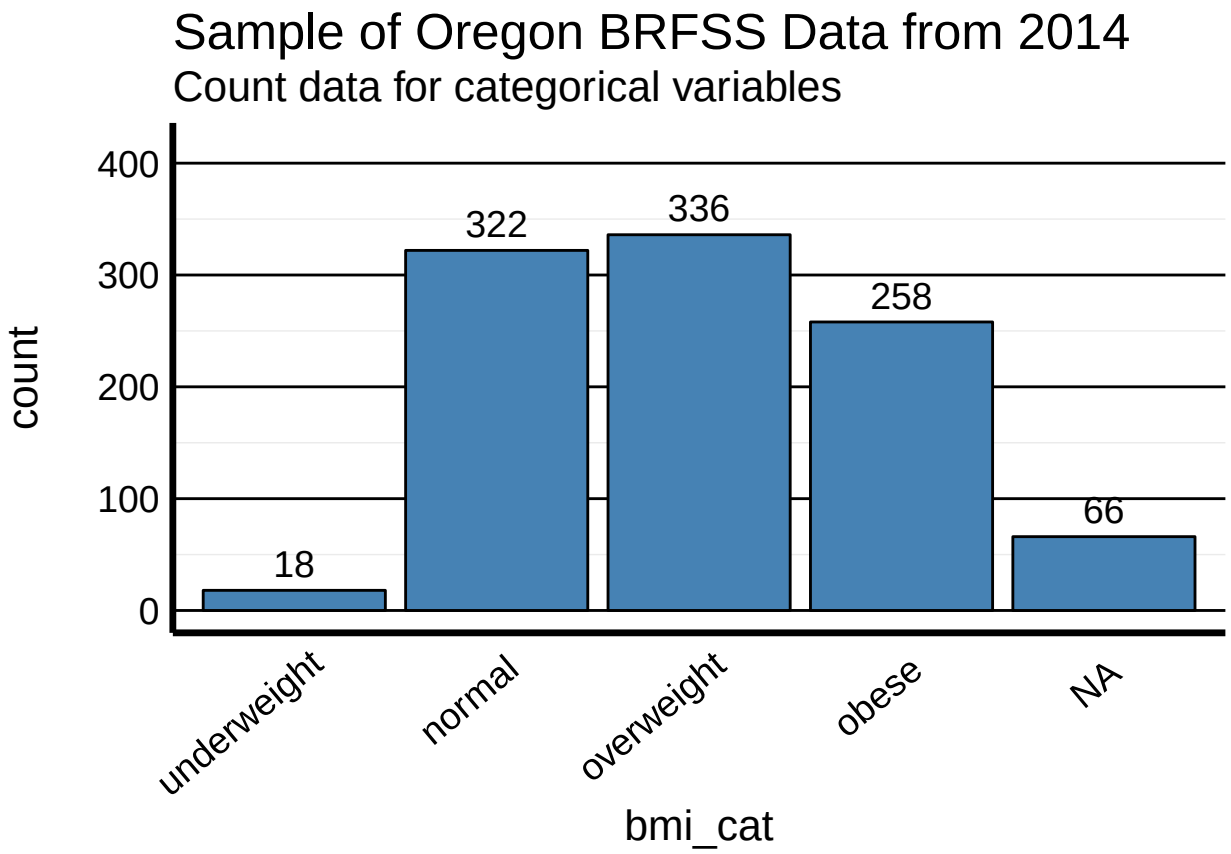

```
hist_function(clean_data_oregon, var = "tobacco_now")
```




```
hist_function(clean_data_oregon, var = "current_smoker")
```




```
hist_function(clean_data_oregon, var = "bmi_cat")
```



Save original data as RDS