# Using R to Replicate NCHS Data Brief No. 368

### Prevalence of Tooth Loss Among Older Adults: United States, 2015–2018.

Data Brief No. 368 (https://www.cdc.gov/nchs/products/databriefs/db368.htm)

Load packages.

```
library(gt)
library(haven)
library(srvyr)
library(survey)
library(tidyverse)
```

Read demographic files keeping variables of interest.

```
DEMO_I <- read_xpt("https://wwwn.cdc.gov/nchs/nhanes/2015-2016/demo_i.xpt") %>%
  select(SEQN, SDDSRVYR, RIAGENDR, RIDAGEYR, RIDRETH3, DMDEDUC2, SDMVSTRA, SDMVPSU, WTMEC2YR)
DEMO_J <- read_xpt("https://wwwn.cdc.gov/nchs/nhanes/2017-2018/demo_j.xpt") %>%
  select(SEQN, SDDSRVYR, RIAGENDR, RIDAGEYR, RIDRETH3, DMDEDUC2, SDMVSTRA, SDMVPSU, WTMEC2YR)
```

Append demographic files and create new variables.

```
DEMO <- bind_rows(DEMO_I, DEMO_J) %>%
  mutate(
    gender = case_when(
      RIAGENDR == 1 ~ "Men", RIAGENDR == 2 ~ "Women"),
    age = case_when(
      RIDAGEYR %in% 65:69 ~ "65-69", RIDAGEYR %in% 70:74 ~ "70-74",
      RIDAGEYR >= 75 ~ "75 and over"),
    age = factor(age,
      levels = c("65-69", "70-74", "75 and over")),
    race_ethnicity = case_when(
      RIDRETH3 %in% 1:2 ~ "Hispanic", RIDRETH3 == 3 ~ "Non-Hispanic white",
      RIDRETH3 == 4 ~ "Non-Hispanic black"),
    race_ethnicity = factor(race_ethnicity,
      levels = c("Non-Hispanic white", "Non-Hispanic black", "Hispanic")),
    education = case_when(
      DMDEDUC2 %in% 1:2 ~ "Less than high school education",
      DMDEDUC2 %in% 3:5 ~ "High school education or greater"),
    education = factor(education,
      levels = c("Less than high school education", "High school education or greater"))) %>%
  select(SEQN, gender, age, race_ethnicity, education, SDMVSTRA, SDMVPSU, WTMEC2YR)
```

Read oral health files keeping variables of interest.

```
OHXDEN_I <- read_xpt("https://wwwn.cdc.gov/nchs/nhanes/2015-2016/ohxden_i.xpt") %>%
  select(SEQN, OHX02TC:OHX31TC)
OHXDEN_J <- read_xpt("https://wwwn.cdc.gov/nchs/nhanes/2017-2018/ohxden_j.xpt") %>%
  select(SEQN, OHX02TC:OHX31TC)
```

Append oral health files and create variable for edentulism or complete tooth loss.

```
OHXDEN <- bind_rows(OHXDEN_I, OHXDEN_J) %>%
  mutate(
    tc = str_c(OHX02TC, OHX03TC, OHX04TC, OHX05TC, OHX06TC, OHX07TC, OHX08TC,
               OHX09TC, OHX10TC, OHX11TC, OHX12TC, OHX13TC, OHX14TC, OHX15TC,
               OHX18TC, OHX19TC, OHX20TC, OHX21TC, OHX22TC, OHX23TC, OHX24TC,
               OHX25TC, OHX26TC, OHX27TC, OHX28TC, OHX29TC, OHX30TC, OHX31TC),
    edentulism = case_when(str_detect(tc, "[4]{28}") ~ 1,
                           str_detect(tc, "[1-5]{28}") ~ 0)) %>%
  select(SEQN, edentulism)
```

Join demographic and oral health data.

```
One <- left_join(DEMO, OHXDEN, by = "SEQN")
```

Define survey design.

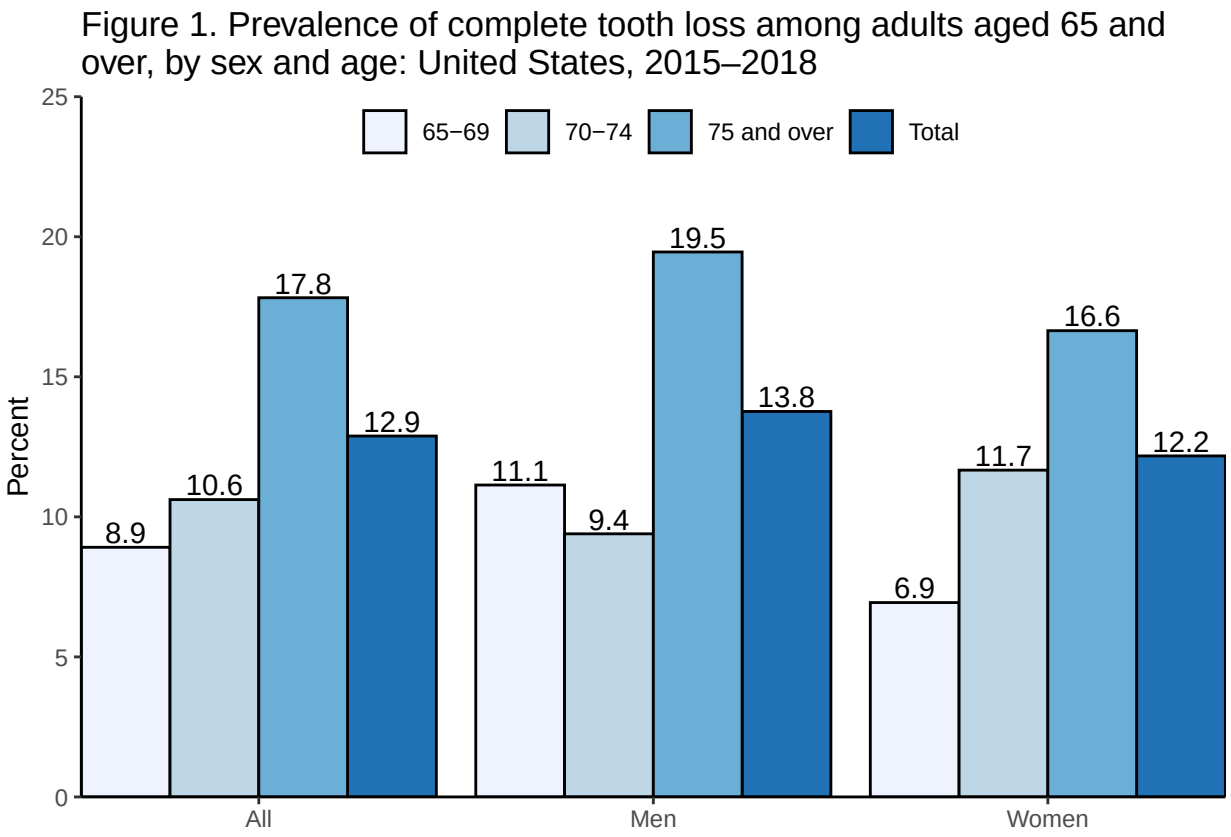
```
NHANES <- One %>%
  as_survey_design(id = SDMVPSU, strata = SDMVSTRA, nest = TRUE, weight = WTMEC2YR)

NHANES <- NHANES %>%
  filter(!is.na(age) & !is.na(edentulism))
```

Create data for Figure 1.

```
t1 <- NHANES %>%
  summarize(gender = "All", age = "Total", n = unweighted(n()), percent = survey_mean(edentulism) * 100)

t2 <- NHANES %>%
  group_by(age) %>%
  summarize(gender = "All", n = unweighted(n()), percent = survey_mean(edentulism) * 100)

t3 <- NHANES %>%
  group_by(gender) %>%
  summarize(age = "Total", n = unweighted(n()), percent = survey_mean(edentulism) * 100)

t4 <- NHANES %>%
  group_by(gender, age) %>%
  summarize(n = unweighted(n()), percent = survey_mean(edentulism) * 100)

table1 <- bind_rows(t1, t2, t3, t4)
```

Create Figure 1.

```
ggplot(table1, aes(gender, percent, fill = age)) +
  geom_bar(stat = "identity", position = position_dodge(), color = "black") +
  geom_text(aes(label = sprintf("%.1f", percent)),
    position = position_dodge(0.9), vjust = -0.2) +
  labs(title = paste("Figure 1. Prevalence of complete tooth loss among adults aged 65 and",
    "over, by sex and age: United States, 2015-2018", sep = "\n"),
    x = element_blank(),
    y = "Percent",
    fill = element_blank()) +
  coord_cartesian(ylim = c(0, 25), expand = FALSE) +
  scale_fill_brewer(palette = "Blues") +
  theme_classic() +
  theme(legend.position=c(.5, .95), legend.direction = "horizontal")
```




Chi-square tests of association.

```
svychisq(~edentulism+age, NHANES)$p.value %>% as.numeric()
```

```
## [1] 6.204857e-06
```

```
svychisq(~edentulism+age, subset(NHANES, gender == "Men"))$p.value %>% as.numeric()
```

```
## [1] 0.0005152575
```

```
svychisq(~edentulism+age, subset(NHANES, gender == "Women"))$p.value %>% as.numeric()
```

```
## [1] 0.001021262
```

All - Test of linear trend.

```
summary(svyglm(edentulism~as.numeric(age), NHANES, family = quasibinomial()))$coefficients
```

```
##                   Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)     -2.8122051 0.27554629 -10.205926 4.147809e-11
## as.numeric(age)  0.4166127 0.09060133   4.598307 7.735880e-05
```

Men - Pairwise comparisons of proportions.

```
svyttest(edentulism~age, subset(NHANES,
  gender == "Men" & age %in% c("65-69", "70-74")))$p.value %>% as.numeric()
```

```
## [1] 0.4562162
```

```
svyttest(edentulism~age, subset(NHANES,
  gender == "Men" & age %in% c("65-69", "75 and over")))$p.value %>% as.numeric()
```

```
## [1] 0.001993754
```

```
svyttest(edentulism~age, subset(NHANES,
  gender == "Men" & age %in% c("70-74", "75 and over")))$p.value %>% as.numeric()
```

```
## [1] 5.530099e-05
```

Women - Test of linear trend.

```
summary(svyglm(edentulism~as.numeric(age), subset(NHANES,
  gender == "Women"), family = quasibinomial()))$coefficients
```

```
##                   Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     -3.0455927  0.3602727 -8.453577 2.577050e-09
## as.numeric(age)  0.4823287  0.1342962  3.591528 1.197418e-03
```

Create table showing data for Figure 1.

```
table1 %>%
  pivot_wider(names_from = gender, values_from = c(n, percent, percent_se)) %>%
  gt() %>%
    tab_spanner(label = "All", columns = c(n_All, percent_All, percent_se_All)) %>%
    tab_spanner(label = "Men", columns = c(n_Men, percent_Men, percent_se_Men)) %>%
    tab_spanner(label = "Women", columns = c(n_Women, percent_Women, percent_se_Women)) %>%
    fmt_number(columns = c(n_All, n_Men, n_Women), decimals = 0) %>%
    fmt_number(columns = c(percent_All, percent_se_All, percent_Men, percent_se_Men,
      percent_Women, percent_se_Women), decimals = 1) %>%
    cols_label(age = "Age",
      n_All = "n", percent_All = "Percent", percent_se_All = "SE",
      n_Men = "n", percent_Men = "Percent", percent_se_Men = "SE",
      n_Women = "n", percent_Women = "Percent", percent_se_Women = "SE")
```

| | All | | | Men | | | Women | | |
|---|---|---|---|---|---|---|---|---|---|
| Age | n | Percent | SE | n | Percent | SE | n | Percent | SE |
| Total | 2,583 | 12.9 | 1.3 | 1,301 | 13.8 | 1.6 | 1,282 | 12.2 | 1.4 |
| 65-69 | 808 | 8.9 | 1.6 | 399 | 11.1 | 2.4 | 409 | 6.9 | 1.7 |
| 70-74 | 637 | 10.6 | 1.5 | 339 | 9.4 | 1.7 | 298 | 11.7 | 2.1 |
| 75 and over | 1,138 | 17.8 | 1.5 | 563 | 19.5 | 1.8 | 575 | 16.6 | 2.0 |

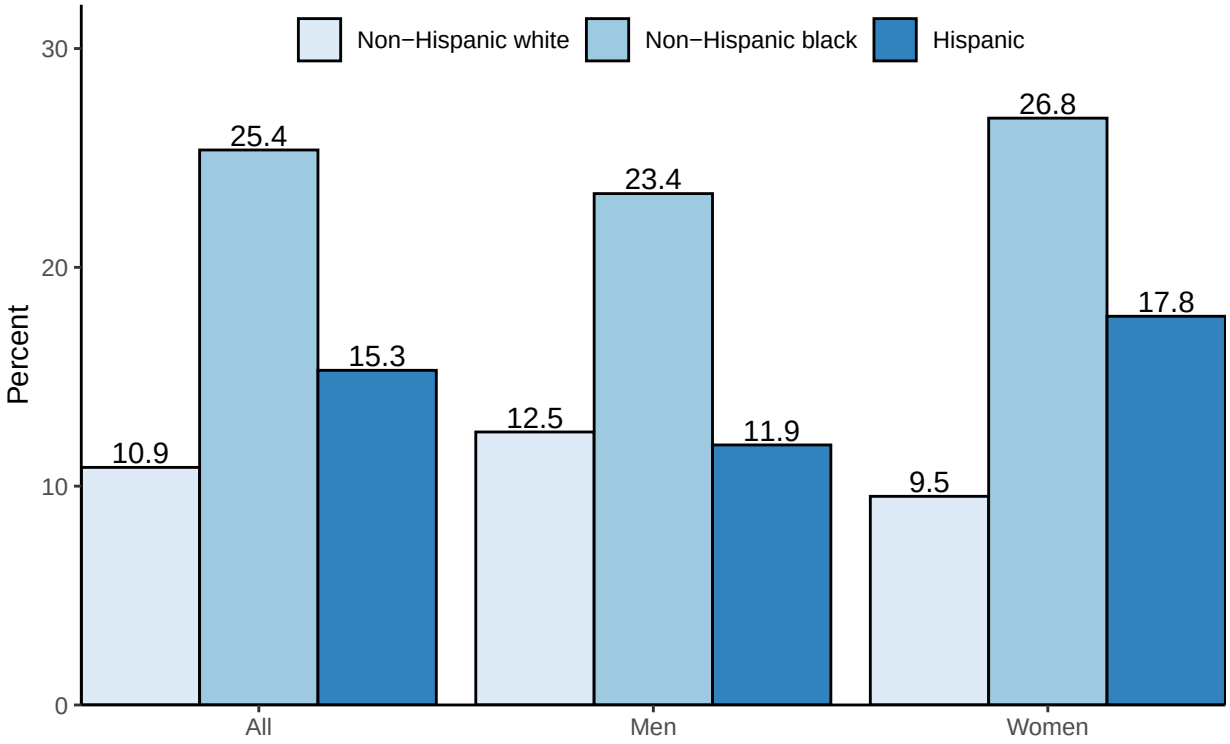
Create data for Figure 2.

```
t1 <- NHANES %>%
  filter(!is.na(race_ethnicity)) %>%
  group_by(race_ethnicity) %>%
  summarize(gender = "All", n = unweighted(n()), percent = survey_mean(edentulism) * 100)

t2 <- NHANES %>%
  filter(!is.na(race_ethnicity)) %>%
  group_by(gender, race_ethnicity) %>%
  summarize(n = unweighted(n()), percent = survey_mean(edentulism) * 100)

table2 <- bind_rows(t1, t2)
```

Create Figure 2.

```
ggplot(table2, aes(gender, percent, fill = race_ethnicity)) +
  geom_bar(stat = "identity",
    position = position_dodge(), color = "black") +
  geom_text(aes(label = sprintf("%.1f", percent)), position = position_dodge(0.9), vjust = -0.2) +
  labs(title = paste("Figure 2. Prevalence of complete tooth loss among adults aged 65 and",
    "over, by sex and race and Hispanic origin: United States, 2015-2018", sep = "\n"),
    x = element_blank(),
    y = "Percent",
    fill = element_blank()) +
  coord_cartesian(ylim = c(0, 32), expand = FALSE) +
  scale_fill_brewer(palette = "Blues") +
  theme_classic() +
  theme(legend.position=c(.5, .95), legend.direction = "horizontal")
```

Figure 2. Prevalence of complete tooth loss among adults aged 65 and over, by sex and race and Hispanic origin: United States, 2015–2018




Chi-square tests of association.

```
svychisq(~edentulism+race_ethnicity, NHANES)$p.value %>% as.numeric()
```

```
## [1] 1.558057e-05
```

```
svychisq(~edentulism+race_ethnicity, subset(NHANES, gender == "Men"))$p.value %>% as.numeric()
```

```
## [1] 0.008930217
```

```
svychisq(~edentulism+race_ethnicity, subset(NHANES, gender == "Women"))$p.value %>% as.numeric()
```

```
## [1] 1.419613e-06
```

All - Pairwise comparisons of proportions.

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  race_ethnicity %in% c("Non-Hispanic white", "Non-Hispanic black")))$p.value %>% as.numeric()
```

```
## [1] 5.300531e-05
```

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  race_ethnicity %in% c("Non-Hispanic white", "Hispanic")))$p.value %>% as.numeric()
```

```
## [1] 0.09104472
```

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  race_ethnicity %in% c("Non-Hispanic black", "Hispanic")))$p.value %>% as.numeric()
```

```
## [1] 0.00870443
```

Men - Pairwise comparisons of proportions.

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  gender == "Men" & race_ethnicity %in% c("Non-Hispanic white", "Non-Hispanic black")))$p.value %>% as.n
```

```
## [1] 0.007369109
```

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  gender == "Men" & race_ethnicity %in% c("Non-Hispanic white", "Hispanic")))$p.value %>% as.numeric()
```

```
## [1] 0.847378
```

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  gender == "Men" & race_ethnicity %in% c("Non-Hispanic black", "Hispanic")))$p.value %>% as.numeric()
```

```
## [1] 0.01081931
```

Women - Pairwise comparisons of proportions.

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  gender == "Women" & race_ethnicity %in% c("Non-Hispanic white", "Non-Hispanic black")))$p.value %>% as
```

```
## [1] 2.115126e-05
```

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  gender == "Women" & race_ethnicity %in% c("Non-Hispanic white", "Hispanic")))$p.value %>% as.numeric()
```

```
## [1] 0.01129873
```

```
svyttest(edentulism~race_ethnicity, subset(NHANES,
  gender == "Women" & race_ethnicity %in% c("Non-Hispanic black", "Hispanic")))$p.value %>% as.numeric()
```

```
## [1] 0.04890223
```

Non-Hispanic white - Pairwise comparison of proportions.

```
svyttest(edentulism~gender, subset(NHANES,
  race_ethnicity == "Non-Hispanic white"))$p.value %>% as.numeric()
```

```
## [1] 0.09061324
```

Non-Hispanic black - Pairwise comparison of proportions.

```
svyttest(edentulism~gender, subset(NHANES,
  race_ethnicity == "Non-Hispanic black"))$p.value %>% as.numeric()
```

```
## [1] 0.3587525
```

Hispanic - Pairwise comparison of proportions.

```
svyttest(edentulism~gender, subset(NHANES,
  race_ethnicity == "Hispanic"))$p.value %>% as.numeric()
```

```
## [1] 0.05075984
```

Create table showing data for Figure 2.

```
table2 %>%
  pivot_wider(names_from = gender, values_from = c(n, percent, percent_se)) %>%
  gt() %>%
    tab_spanner(label = "All", columns = c(n_All, percent_All, percent_se_All)) %>%
    tab_spanner(label = "Men", columns = c(n_Men, percent_Men, percent_se_Men)) %>%
    tab_spanner(label = "Women", columns = c(n_Women, percent_Women, percent_se_Women)) %>%
    fmt_number(columns = c(n_All, n_Men, n_Women), decimals = 0) %>%
    fmt_number(columns = c(percent_All, percent_se_All, percent_Men, percent_se_Men,
      percent_Women, percent_se_Women), decimals = 1) %>%
    cols_align(columns = race_ethnicity, align = "left") %>%
    cols_label(race_ethnicity = "Race & Hispanic Origin",
      n_All = "n", percent_All = "Percent", percent_se_All = "SE",
      n_Men = "n", percent_Men = "Percent", percent_se_Men = "SE",
      n_Women = "n", percent_Women = "Percent", percent_se_Women = "SE")
```

| | All | | | Men | | | Women | | |
|---|---|---|---|---|---|---|---|---|---|
| Race & Hispanic Origin | n | Percent | SE | n | Percent | SE | n | Percent | SE |
| Non-Hispanic white | 1,217 | 10.9 | 1.6 | 641 | 12.5 | 1.9 | 576 | 9.5 | 1.7 |
| Non-Hispanic black | 510 | 25.4 | 2.8 | 263 | 23.4 | 2.8 | 247 | 26.8 | 3.6 |
| Hispanic | 554 | 15.3 | 1.8 | 249 | 11.9 | 2.5 | 305 | 17.8 | 2.2 |

Create data for Figure 3.

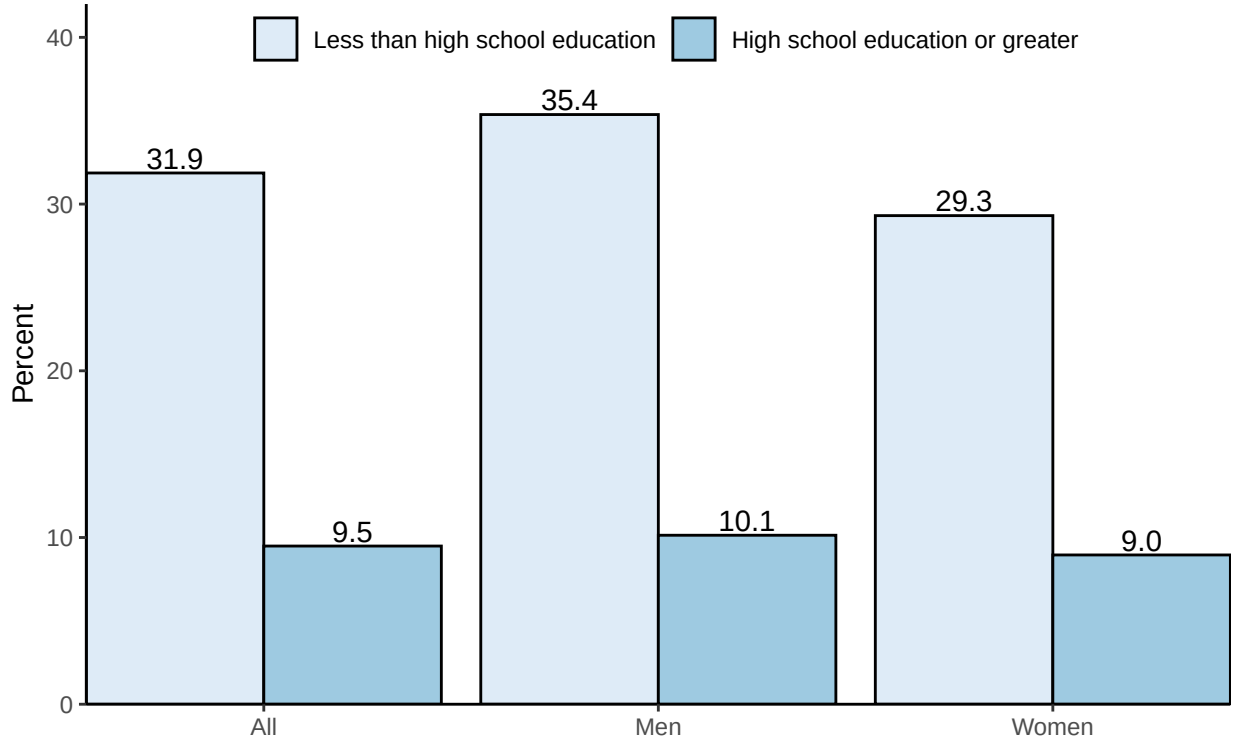
```
t1 <- NHANES %>%
  filter(!is.na(education)) %>%
  group_by(education) %>%
  summarize(gender = "All", n = unweighted(n()), percent = survey_mean(edentulism) * 100)
```

```
t2 <- NHANES %>%
  filter(!is.na(education)) %>%
  group_by(gender, education) %>%
  summarize(n = unweighted(n()), percent = survey_mean(edentulism) * 100)

table3 <- bind_rows(t1, t2)
```

Create Figure 3.

```
ggplot(table3, aes(gender, percent, fill = education)) +
  geom_bar(stat = "identity", position = position_dodge(), color = "black") +
  geom_text(aes(label = sprintf("%.1f", percent)),
    position = position_dodge(0.9), vjust = -0.2) +
  labs(title = paste("Figure 3. Prevalence of complete tooth loss among adults aged 65 and",
    "over, by sex and education level: United States, 2015-2018", sep = "\n"),
    x = element_blank(),
    y = "Percent",
    fill = element_blank()) +
  coord_cartesian(ylim = c(0, 42), expand = FALSE) +
  scale_fill_brewer(palette = "Blues") +
  theme_classic() +
  theme(legend.position=c(.5, .95), legend.direction = "horizontal")
```

Figure 3. Prevalence of complete tooth loss among adults aged 65 and over, by sex and education level: United States, 2015–2018




Chi-square tests of association.

```
svychisq(~edentulism+education, NHANES)$p.value %>% as.numeric()
```

```
## [1] 5.776786e-09
```

```
svychisq(~edentulism+education, subset(NHANES, gender == "Men"))$p.value %>% as.numeric()
```

```
## [1] 1.074444e-07
```

```
svychisq(~edentulism+education, subset(NHANES, gender == "Women"))$p.value %>% as.numeric()
```

```
## [1] 3.678882e-08
```

Create table showing data for Figure 3.
*(Note: The standard error for women with less than a high school education is 3.7 in the Data Brief. Assume difference due to rounding or typo.)*

```
table3 %>%
  pivot_wider(names_from = gender, values_from = c(n, percent, percent_se)) %>%
  gt() %>%
    tab_spanner(label = "All", columns = c(n_All, percent_All, percent_se_All)) %>%
    tab_spanner(label = "Men", columns = c(n_Men, percent_Men, percent_se_Men)) %>%
    tab_spanner(label = "Women", columns = c(n_Women, percent_Women, percent_se_Women)) %>%
    fmt_number(columns = c(n_All, n_Men, n_Women), decimals = 0) %>%
    fmt_number(columns = c(percent_All, percent_se_All, percent_Men, percent_se_Men,
      percent_Women, percent_se_Women), decimals = 1) %>%
    cols_align(columns = education, align = "left") %>%
    cols_label(education = "Education Level",
      n_All = "n", percent_All = "Percent", percent_se_All = "SE",
      n_Men = "n", percent_Men = "Percent", percent_se_Men = "SE",
      n_Women = "n", percent_Women = "Percent", percent_se_Women = "SE")
```

| | All | | | Men | | | Women | | |
|---|---|---|---|---|---|---|---|---|---|
| Education Level | n | Percent | SE | n | Percent | SE | n | Percent | SE |
| Less than high school education | 696 | 31.9 | 3.1 | 335 | 35.4 | 4.0 | 361 | 29.3 | 3.6 |
| High school education or greater | 1,876 | 9.5 | 1.2 | 960 | 10.1 | 1.5 | 916 | 9.0 | 1.2 |