Gov 51 - Pset 1

Tyler Dang - https://github.com/tylerdang25/Gov-51—Pset-1.git

2. Get to Know Your Data

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.6
v forcats   1.0.1     v stringr   1.6.0
v ggplot2   4.0.1     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.2
v purrr     1.2.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```
library(dplyr)
library(knitr)
```

```
acs2024 <- read.csv('/Users/tyler/Documents/Gov 51/gov51-ps1/data/raw/acs2024.csv')
```

FROM THE DOCUMENTATION:

TRANTIME reports the total amount of time, in minutes, that it usually took the respondent to get from home to work last week.

A 000 for TRANTIME represents cases in which there is no data available for that observation.

PERWT indicates how many persons in the U.S. population are represented by a given person in an IPUMS sample. Basically, for each observation in this data set, the PERWT value indicates how many real people in the U.S. population are represented by that observation.

```
acs2024 <- acs2024 %>%
  mutate(female = ifelse(SEX == 2, 1, 0))

acs2024 <- acs2024 %>%
  mutate(educ_lhs = ifelse(EDUC %in% 1:5, 1, 0),
    educ_hs = ifelse(EDUC == 6, 1, 0),
    educ_somecol = ifelse(EDUC %in% 07:09, 1, 0),
    educ_college = ifelse(EDUC == 10, 1, 0),
    educ_adv = ifelse(EDUC == 11, 1, 0))

acs2024 <- acs2024 %>%
  mutate(employed = ifelse(EMPSTAT == 1, 1, 0),
    unemployed = ifelse(EMPSTAT == 2, 1, 0),
    nlf = ifelse(EMPSTAT == 3, 1, 0))
```

If we included special codes that represent "N/A" codes, we may accidentally misreport in
summary statistics the extent that a certain population appears. For example, if we included
0 as a part of the range for education less than high school, we would drastically increase the
extent that those with an education level less than high school make up the data set instead
of accurately showing that the information may just be missing.

```
acs2024table <- data.frame(
  Variable = c("Age", "Female", "Less than High School", "High School Only",
              "Some College", "College Only", "Advanced Degree", "Employed",
              "Unemployed", "Not in Labor Force", "Commute Time (mins)",
              "Total Income ($)"),

  N = c(length(acs2024$AGE), length(acs2024$female),length(acs2024$educ_lhs),
      length(acs2024$educ_hs), length(acs2024$educ_somecol),
      length(acs2024$educ_college), length(acs2024$educ_adv),
      length(acs2024$employed), length(acs2024$unemployed),
      length(acs2024$nlf), length(acs2024$TRANTIME), length(acs2024$INCTOT)),

  Mean = c(mean(acs2024$AGE), mean(acs2024$female), mean(acs2024$educ_lhs),
        mean(acs2024$educ_hs), mean(acs2024$educ_somecol),
        mean(acs2024$educ_college), mean(acs2024$educ_adv),
        mean(acs2024$employed), mean(acs2024$unemployed),mean(acs2024$nlf),
        mean(acs2024$TRANTIME), mean(acs2024$INCTOT) / 1e6),

  SD = c(sd(acs2024$AGE), sd(acs2024$female),sd(acs2024$educ_lhs),
      sd(acs2024$educ_hs), sd(acs2024$educ_somecol),
      sd(acs2024$educ_college), sd(acs2024$educ_adv), sd(acs2024$employed),
```

```
        sd(acs2024$unemployed),sd(acs2024$nlf), sd(acs2024$TRANTIME),
        sd(acs2024$INCTOT) / 1e6),

  Min = c(min(acs2024$AGE), min(acs2024$female), min(acs2024$educ_lhs),
          min(acs2024$educ_hs), min(acs2024$educ_somecol),
          min(acs2024$educ_college), min(acs2024$educ_adv),
          min(acs2024$employed), min(acs2024$unemployed), min(acs2024$nlf),
          min(acs2024$TRANTIME), min(acs2024$INCTOT) / 1e6),

  Max = c(max(acs2024$AGE), max(acs2024$female),max(acs2024$educ_lhs),
          max(acs2024$educ_hs), max(acs2024$educ_somecol),
          max(acs2024$educ_college), max(acs2024$educ_adv),
          max(acs2024$employed), max(acs2024$unemployed), max(acs2024$nlf),
          max(acs2024$TRANTIME), max(acs2024$INCTOT) / 1e6)
)

#acs2024table

acs2024rounded <- acs2024table
acs2024rounded[, -1] <- round(acs2024table[, -1], 2)
#acs2024rounded

knitr::kable(acs2024rounded,
            caption = "Table 1: Summary Statistics for 2024 ACS Sample",
            align = "lccccc", digits = 2)
```

Table 1: Table 1: Summary Statistics for 2024 ACS Sample

Variable	N	Mean	SD	Min	Max
Age	3422888	43.39	24.03	0.00	96
Female	3422888	0.51	0.50	0.00	1
Less than High School	3422888	0.19	0.39	0.00	1
High School Only	3422888	0.30	0.46	0.00	1
Some College	3422888	0.18	0.38	0.00	1
College Only	3422888	0.17	0.38	0.00	1
Advanced Degree	3422888	0.11	0.32	0.00	1
Employed	3422888	0.47	0.50	0.00	1
Unemployed	3422888	0.02	0.14	0.00	1
Not in Labor Force	3422888	0.35	0.48	0.00	1
Commute Time (mins)	3422888	10.81	19.83	0.00	195
Total Income ($)	3422888	1.54	3.54	-0.01	10