Assignment 1b: Replicating *Sexual Orientation and Labor Economics* by Cara Brown (1998)

Daniel Sánchez

1/12/23

This document walks through the replication of Cara Brown's 1998 paper *Sexual Orientation and Labor Economics* published at *Feminist Economics*.

Preliminaries

I need the following R libraries to make this code work:

```
library(tidyverse) # For data wrangling
library(haven) # Reading stata files
```

I load the data from the given file below:

```
df<-
   read_dta('data/Census_2016_Hierarchial.dta')
```

Exploratory Analysis & Data Wrangling

I run some exploratory analysis which motivates my wrangling/cleaning of the data set to make it fit for the analysis below.

```
# First, rename the income variable (empin) to something more usable

df<-
   df %>%
   rename(income = 'empin')
```

```
# Now, replace the invalid values (88,888,888, 99,999,999) with actual NAs in R.

df <-
  df %>%
  mutate(income_clean =
    case_when(
      income %in% c(88888888, 99999999) ~ as.numeric(NA),
      TRUE ~ income
    )
  )
```

I now group by household and count the number of individuals in them.

```
households <-
  df %>%
  group_by(HH_ID) %>%
  summarise(people = n())
```