Homework 3

Include all of your code in a quarto document (echo: true). Render your document to html, docx, or pdf and submit. As always, there are many different ways to solve these problems, but focus on the functions and skills we have been learning in class.

Use the following code to read in a raw version of the NLSY dataset we’ve been using:

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.2 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
✔ purrr 1.0.1   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

nlsy <- read\_csv("https://github.com/louisahsmith/id543/raw/main/data/nlsy.csv")

Rows: 12686 Columns: 14  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
dbl (14): H0012400, H0012500, H0022300, H0022500, R0000100, R0009100, R01736...  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

The variables in this dataset correspond to c("glasses", "eyesight", "sleep\_wkdy", "sleep\_wknd", "id", "nsibs", "samp", "race\_eth", "sex", "region", "income", "res\_1980", "res\_2002", "age\_bir") in this order. All negative numbers indicate missing values.

1. Create a version of the dataset with the variable names as specified above.
2. Replace missing values in income and region with NA.
3. Create a factor variable that has the value “sister”, “brother”, or “only child” depending on a person’s sex and number of siblings. Recall that men have sex = 1 and women have sex = 2. E.g. if the individual has siblings and is male, they are a “brother” to their siblings.
4. Create a dataset with *only* the sisters who are missing income and region. How many are there?
5. Move id to the beginning (first column) of the dataset and remove the samp, res\_1980, and res\_2002 variables.
6. Combine your answers to questions 1:5, including reading in the data, using pipes. Your answer should be one continuous sequence of code that reads in the data and stores the changed dataset in an object.