Homework 3

The homework is in this format to facilitate you showing your code and output. Render the document when you are finished and submit it on Canvas. If you have trouble rendering to Word, try format: html instead. If you are still having trouble, you can always copy and paste your code and output (or screenshots of the output) into a Word document. What is important is that we can see both your code and its output.

1. Start a new R session. Before you can read in the data, load the tidyverse set of packages. Then read in the data.
2. Create a variable that is 1 if an observation has an income between 20,000 and 30,000, and 0 otherwise. Using summarize(), calculate the proportion of people in the dataset who fit that criterion.
3. Repeat the previous question, but separately for males and females.
4. The summarize() function recently gained another argument: .by =. Look up the documentation for summarize() and use this argument to repeat the previous question.
5. Recreate the summary() function using summarize() (i.e., produce all the same statistics) for the age\_bir variable. That is, you should get all the same values as summary(nlsy$age\_bir) but using summarize() instead.
6. Join the full nlsy dataset with the kids dataset you read in in class from <https://github.com/louisahsmith/data/raw/main/nlsy/nlsy-child.csv>. Use a join function that results in *all* the moms and dads in the nlsy dataset included, even if they have no kids. How many rows does it have? Why is it more than the number of rows in the nlsy dataset?

For your convenience, the code to read in the nlsy kids dataset is already provided below.

kids\_cols <- c("id\_kid", "id\_mom", "sex\_kid", "dob\_kid", "agebir\_mom", "bwt\_kid")

nlsy\_kids <- read\_csv("https://github.com/louisahsmith/data/raw/main/nlsy/nlsy-child.csv",
 skip = 1, col\_names = kids\_cols,
 na = c("-1", "-2", "-3", "-4", "-5", "-7", "-6"))

Error in read\_csv("https://github.com/louisahsmith/data/raw/main/nlsy/nlsy-child.csv", : could not find function "read\_csv"

1. In the merged dataset from the last question, how many males and females are there *without* kids?
2. Calculate the average birthweight in grams by race/ethnicity. Remember the variable “bwt\_kid” is in ounces.