Homework 4

Download the homework as a quarto document [here](homework/_homework_4.qmd) or as a Word document [here](homework/_homework_4.docx).

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.

library(tidyverse)

1. Read in the .csv file you used for yesterday’s homework from <https://github.com/louisahsmith/id543/raw/main/data/nlsy.csv>. This time, read the column names and missing values in directly. Recall that 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.
2. Create a variable that is 1 if an observation has an income between 20,000 and 30,000, and 0 otherwise. Calculate the proportion of people in the dataset who fit that criterion.
3. 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.
4. Join the full nlsy dataset with the kids dataset you read in in class from <https://github.com/louisahsmith/id543/raw/main/data/nlsy-child.csv>. Use a join function that results in all the moms and dads without kids in the dataset included. How many rows does it have?

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/id543/raw/main/data/nlsy-child.csv",   
 skip = 1, col\_names = kids\_cols,   
 na = c("-1", "-2", "-3", "-4", "-5", "-7", "-6"))

1. How many of each group are there: a) moms with kids, b) moms without kids, c) dads with kids, d) dads without kids?
2. Here’s a code snippet:

regions <- tibble(region = 1:4,  
 region\_cat = c("Northeast", "North Central", "South", "West"))

With this regions dataset, use a join function to add the region\_cat variable to the joined parent/kid dataset.

1. Using your region\_cat variable, calculate the average birthweight in grams within each region. Remember the variable “bwt\_kid” is in ounces.