Exercise #3

Crime Analytics (CJUS 6106)

2023-02-13

Instructions

For this assignment, you will be using the functions reviewed in the Data Cleaning and Management lecture to create the NLSY97 data set you used for Exercise #2. In doing so, you will need to read the original .csv file into R, rename each column, replace negative values with their correct values (-4 = 0, all else is NA), create an age variable using the interview year and year of birth variables, create labels for each column, and recode several character variables into factor variables. You will have both data files available to you so you can see what the final data set should look like - you are merely recreating the syntax I used to make the data set for Exercise #2. You should submit both a .rmd file and knitted PDF by 11:59pm on the due date indicated on Canvas.

Question 1

Read the nls97_raw.csv file into R as a data frame.

Question 2

Rename each column so that the variable names reflect those in the Exercise 2 data set. You can use the colnames() function to do this or you can create copies of each variable and combine them into a new data frame.

Question 3

Replace the -4 values in each column with zeroes and the remaining negative values (-5, -3, -2, and -1) with NA values.

Question 4

Create a new variable called age98 using the interview year and birth year variables. Make sure that this new variable is created within your new data frame.

Question 5

Label each column with the labels provided in the Exercise 2 data frame. You can see these labels when you double-click on the data frame object and inspect the text underneath each variable name.

Question 6

Recode the household structure, sex, and race variables to be consistent with their values in the Exercise 2 data frame. You will need to consult the codebook file I provide on Canvas to know what each numeric value represents. This file can be opened in a text editor like Notepad.