Coding Challenge 2

Data Cleaning and Management

Due: 5pm, Friday, October 26th, 2018

In this coding challenge, you will be tasked with cleaning and combining data sets. In the real world, data is (generally) not provided to you in a format that is ready for analysis. Usually some form of data cleaning or rearranging is required before you can begin your data analysis.

Download and read into R the four `.csv` files from Piazza, titled: `cdcdeath_g`, `cdcdeath_r`, `Crime_G`, `Crime_R`. The two "death" datasets have been adapted from the CDC Wonder data that you will be using for your projects. The "crime" datasets were obtained from one of the outside sources you could use for your projects, The Uniform Crime Reporting Program Data.

Submission: Please submit two `.csv` files: one for gender and one for race, and an `.Rmd` file with your code into a folder titled `lastname_firstname_CC2` within your Google Drive folder by **5pm, Friday, October 26th, 2018**.

Question 1 (2 points)

Use `View()` to look at the four datasets. We want to combine the gender datasets and the race datasets together to make two datasets (one for gender, one for race). What potential problems do you think there are with this data that would cause issues with combining them right now?

Question 2 (10 points)

You are not allowed to use any loops for these questions.

Part A (3 points)

Using the `Crime_R` data, change the levels of the `RACE` variable to match those in the `death_r` data, so that both datasets have the same labels for each level. (*Note: Assume that Indian falls under American Indian or Alaska Native.*)

Part B (2 points)

Using the `Crime_G` data, make the `GENDER` variable binary, such that "Male" = 0 and "Female" = 1.

Part C (2 points)

Using the `death_g` data and using a *different* function than **Part B**, make the `GENDER` variable binary, such that "M" = 0 and "F" = 1.

Part D (3 points)

Sort all the datasets by the `STATE` variable. (*Hint: Use the* `order()` *function.*)

Question 3 (8 points)

Part A (5 points)

Combine the “gender” datasets and the “race” datasets two make two datasets, one for gender and one for race. (*Hint: Look up a function that will allow you to combine datasets by certain columns.*)

Bonus (Up to 5 points)

Create two columns that display the death rates and arrest rates in the “gender” and “race” data. Make sure to label the columns appropriately.

Part B (3 points)

Export the two datasets as `.csv` files with the following filenames: `lastname_firstname_gender` and `lastname_firstname_race`