

Examining the burden of overdose deaths across US States

GH 811 In-class Exercise 1

January 29, 2019

Background

The Centers for Disease Control tracks drug overdose deaths in the United States using multiple cause-of-death records from the National Vital Statistics System. Overdose deaths are primarily related to pain relievers and heroin.

Data

The data for this exercise are available on the website of the Centers for Disease Control: <http://www.cdc.gov/drugoverdose/data/statedeaths.html> Scroll down to below the table to find the link to the CSV file.

Objectives

You have 45 minutes to complete the following steps. In cases where answers are required include these in your script using the # sign to comment out the text. At the end of the exercise, you will save your R script as “last name_ex1.R” and submit the script and dataset by email to acstokes@bu.edu

1. Pre-processing step
 - a. Open CSV file in MS Excel. Remove the commas from the columns reporting the number of overdose deaths in 2013 and 2014.
 - b. Save over the original CSV file with this modified file.
2. Read in the CSV file to R Studio
 - a. Establish a folder on your computer for GH 811 Exercise 1
 - b. Download the CSV file to that directory
 - c. Establish a new script file in R Studio
 - d. Set directory, point R to the directory and read in the data using read.csv
3. Browse data in R Studio
 - a. Briefly describe the contents of the database (remember to use # sign to comment out this text in your script file).
 - b. How many observations are there? How many variables?
4. Clean data
 - a. In this exercise, we only need the state identifiers, the number of deaths from overdose in 2013 and 2014 and the overdose death rate in 2014. Create a new dataset containing just these variables.

- b. What class do each of the four variables belong to? If the state variable is not a character variable, convert it to this format.
5. Explore your data
- a. How many overdose deaths were there in total in 2013?
 - b. How many in 2014?
 - c. Which state had the highest rate of overdose deaths in 2014?
 - d. Which state had the lowest rate of overdose deaths in 2014?
 - e. How many overdose deaths were there in MA in 2013 and 2014 combined?
 - f. What was the mean number of overdose deaths in MA in these two years?
 - g. How many more overdose deaths were there in MA in 2014 compared to 2013?
6. Save your dataset to your computer
- a. Consult help file for write.csv to learn about the proper syntax for this command
 - b. Save your dataset to your working directory
 - c. Open saved CSV file. Does it contain the same number of columns as your dataset in R? If not, use help file to determine appropriate syntax for eliminating the extra column.
 - d. Re-save to directory and open the new CSV file to confirm that it worked.
 - e. Send script and data file to instructor.