Week 4 Assignment

The answers to the following six questions should be placed in a single R script. Place your R script in a public repository on github and submitting a link to the script here. Label your answers using comments so that they can be clearly and quickly found within the script.

Week 4 assignment is due end of day on Tuesday September 23th. *Solutions to all assignment exercises will be posted on Wednesday September 24th.*

1. Read in the week-4-price-data.csv and week-4-make-model-data.csv files as data frames and then merge them by the ModelNumber key. Leave the "all" parameters as their defaults. How many observations end up in the result? Is this what you would have expected?

2. Use the data sets from the previous question, but this time merge them so that the rows from the price-data table all appear, even if there is no match in the make-model table.

3. Take your result from question 2 and subset it so that only the 2010 vehicles are included.

4. Take your result from question 2 and subset it so that only the red cars that cost more than $10,000 are included.

5. Take your result from question 4 and subset it so that the ModelNumber and Color columns are removed.

6. Write a function that takes as input a character vector and returns a numeric vector with the numbers of characters in each of the elements in the original vector.

Week 4 Challenge Exercises. ***Not Required****. You'll build up your skills faster by just trying to tackle these:*

7. Write a function that takes two character vectors of equal length and concatenates them element by element with a space as the separator. Have the function die gracefully if the vectors are not the same length.

8. Write a function that takes a character or factor vector and determines the number of distinct elements in the vector, the most commonly occurring element, the number of times the most commonly occurring element occurs, and the number of missing values. (Be sure to handle ties gracefully.) Have the function return a named list with the desired information in a logical order.

9. Write a function that takes a logical vector and determines the number of true values, the number of false values, the proportion of true values, and the number of missing values. Have the function return a named list with the desired information in a logical order.

10. Write a function that takes a character vector and returns the substring of three characters that begins with the first vowel in the string. Have the function handle gracefully substrings where this isn't possible.