

Introduction to R

- Turn in the knitted Word document from R Markdown to the dropbox on D2L.
- Make sure to include your name and assignment number in the header.
- You MUST include all R code in order to receive full credit.

Exercises:

1. In this problem you will read in a data set, get the structure of the data set, and then perform basic arithmetic on the elements of the data set. You will also have to deal with missing values.
 - (a) Read in the data set called `FinalData.csv`. How many observations do we have? How many variables? What are the names of the variables? Are they categorical or quantitative?
 - (b) Add up the values of the fourth row and first column and the twelfth row and third column.
 - (c) Print the first ten values in the second column.
 - (d) Find the mean of all the non-missing values in the third column.
 - (e) Find the proportion of observations that fall in each category for variable `y`.
2. In this problem we will write a function that contains an if/else statement and a for loop. Then we will use the `apply()` function to run this created function on all the numeric columns in the above data set.
 - (a) Create a function that reads in a vector of numeric values. In this function, you will loop through all the elements of the vector. You must use a for loop! For each element:
 - If that element is missing, then set it equal to zero;
 - If that element is negative, then multiply the value by -1;
 - If that element is non-negative, then add two to the value.Then, sum up all of the new values and return this sum.
 - (b) Use the `apply()` to run this new function on all numeric variables in the above data set.
 - (c) Use only the `ifelse()` and `sum()` functions to get the same results as the previous step. In other words, remove the for loop and function part of this problem.

3. This problem will use functions from the tidyverse package to manipulate data sets and create scatterplots.
- (a) Load the `tidyverse` package into R.
 - (b) Use the pipe command or create intermediary objects to do the following:
 - i. Use `mutate()` function to create a new variable that computes the difference between values in column three and column one;
 - ii. Use the `filter()` function to only include observations that got values equal to B or C for variable `y` (remove any observations that got value A for variable `y`);
 - iii. Use the `group_by()` and `summarize()` functions to get the mean difference from the first step for each of the remaining values in `y`;
 - iv. Arrange them from largest to smallest.
 - (c) Use the `ggplot()` function to create a scatterplot with `x` as the explanatory variable (horizontal axis) and `var.1` as the response variable (vertical axis). Make observations with different values of `y` different colors in the scatterplot. Add the best fitting curve for each of the different values of `y`.