Assignment 1 — Basic Skills in R

BUAD6701 – Spring 2019

---

1. Objectives

The purpose of this assignment is to practice certain skills in the R language that are commonly used in the development of models of the type we will be studying this semester.

2. What You Will Need

- Access to a Windows computer with R, and to the following files, which can be downloaded from the Class Schedule page of the course web site:
  - vehiclename.csv
  - rating.csv
  - year.csv
  - cylinders.csv

3. What You Will Hand In

Submit a script file called Assignment1*firstnamelastname*.R via Blackboard - Assignment 1. See the Preliminaries section below for instructions on how to create this file.

4. Due Date

Not before Friday April 12th, 2019 at 23:59 – consult the Class Schedule page of the course web site for the official due date and time.

Absent extenuating circumstances, assignments submitted after the due date will be penalized 20% of the assignment's value and assignments that are more than 3 days late will receive a grade of 0%. No extensions to the due date will be granted – no exceptions.

5. Note on Collaboration

This is a Category C assignment. While you may receive verbal assistance from, and provide verbal assistance to, another student in the current class, the work that you submit must have been created by you individually without any physical or digital assistance from another person. This prohibition includes (but is not limited to) copying someone else's code, whether visually or electronically. **To do otherwise is an Honor Code violation**. Obtaining help from the TA's or from the instructor is permitted.

6. Preliminaries:

To get set up for the assignment, follow these steps:

1. Create a directory (using Windows Explorer) somewhere on your hard drive (or network drive) called Assignment1.
2. Download the assignment files from the web site into this directory.
3. Start Rstudio and set the working directory to this directory.
   a. On the main menu, choose Session/Set Working Directory/Choose Directory…
4. Create a new script and save it as Assignment1*firstnamelastname*.R, where *firstnamelastname* is your first name and your last name (no spaces in the whole file name, capitalize first letters of all words, as in Assignment1DavidMurray.R).
5. Enter the R statements in the script window as you work through the assignment, testing them as you go with Ctrl+Enter as we have been doing in class.
6. **Important Notes**:
   - Place an rm(list=ls()) statement at the very beginning of your script
   - Each time you begin a new question, place the following comments in your script (where *n* is the question number):

     ##
     ##### QUESTION *n* ######################
     ##

   - Your script should contain only those statements that are necessary to perform the tasks required … there should be no extra statements therein. In particular, do not include statements that display the contents of vectors or data frames unless requested to do so and do not include statements that install packages.
   - Be sure that there is no "path" information in your script. All file references must be to your working directory so that path information is not required in your script.
   - If you wish to stop and resume later, save your script and close RStudio. When you wish to resume, just double-click your script file to start Rstudio, use the Session menu item to set your working directory to "To Source File Location" and run your entire script.

7. Assignment Questions:

Question 1 (20%):

i. Evaluate the following formula without using R (i.e.: on paper or in Excel), where a, b and c are the values 2, 5 and 7 respectively:

   a. $\frac{a}{b^3} - 1 + \frac{2c^2}{a}$

ii. Create a vector named **vec1** containing the values 2, 5 and 7

iii. To confirm that you understand the order in which R performs arithmetic operations, do the following (in a single R statement):

   Create a variable named **res1** which evaluates the mathematical expression above using the three elements of **vec1**, whatever they may be. That is, in the expression above,

   - Instead of the **a,** refer directly to the first element in the vector
   - Instead of the **b**, refer directly to the second element of the vector
   - Instead of the **c**, refer directly to the third element of the vector.

   Your code should not define intermediate variables named a, b and c, for example.

iv. Display the value of **res1**.

v. Change **vec1** so that its values are 60, 23 and 17

vi. Compute **res1** again by executing the same expression as in step iii again.

vii. Display **res1**.

Question 2 (30%):

i. Create four vectors named vehiclename, rating, year, and cylinders from the four comma-separated value (.csv) files provided. Note that the vehiclename and rating are character variables while the values in cylinders and year are numeric.

   a. Recall the scan() function. Note that when creating a numeric vector with the scan() function, you must use the following syntax:

      what=numeric( )

ii. Create a data frame named vehicleinfo using these vectors (in the order vehiclename, rating, year, cylinders). Do not suppress the conversion of strings to factors. That is, do not include a stringsAsFactors=F parameter.

iii. Display the structure of the data frame and confirm that the vehiclename and rating columns are Factors and the cylinders and year variables are numeric (if they're not, recreate the vectors with the correct modes, then recreate the data frame)

   a. Recall the str(…) function

Your data frame should look like this:

```
'data.frame':   396 obs. of  4 variables:
 $ vehiclename: Factor w/ 304 levels "amc ambassador brougham",..: 49 36 231 14 161 141 54 223 241 2 ...
 $ rating     : Factor w/ 6 levels "Acceptable","Excellent",..: 1 2 6 6 2 3 6 3 1 6 ...
 $ year       : num  70 70 70 70 70 70 70 70 70 70 ...
 $ cylinders  : num  8 8 8 8 8 8 8 8 8 8 ...
```

Question 3 (30%):

i. Display the second row of the data frame.

ii. Display the first 10 rows of the "cylinders" column of the data frame using the head() function. **Do not use the "dollar-sign" notation to do so**.

iii. Display the contents of the first and fourth columns, but just for rows 3 through 7. Do this in a single statement to produce the following:

```
                vehiclename cylinders
3 plymouth satellite         8
4      amc rebel sst         8
5        ford torino         8
6   ford galaxie 500         8
7   chevrolet impala         8
```

iv. Display the first 10 values in the rating column **using the "dollar-sign" notation**. The results should look like this:

```
 [1] Acceptable Excellent  Very Poor  Very Poor  Excellent  Good
 [7] Very Poor  Good       Acceptable Very Poor
Levels: Acceptable Excellent Good Poor Very Good Very Poor
```

Question 4 (20%):

i. Change the rating variable to an ordinal factor. The order should be as follows:

   a. Very Poor
   b. Poor
   c. Acceptable
   d. Good
   e. Very Good
   f. Excellent

   Note that you are being asked to change the variable in the data frame, not the vector with the same name. To do this, you must refer to the variable in the data frame (perhaps using the "dollar-sign" notation) on both the left-hand and right-hand sides of the <- assignment operator.

ii. Display the structure of vehicleinfo. It should look like this:

```
'data.frame':   396 obs. of  4 variables:
 $ vehiclename: Factor w/ 304 levels "amc ambassador brougham",..: 49 36 231 14 161 141 54 223 241 2 ...
 $ rating     : Ord.factor w/ 6 levels "Very Poor"<"Poor"<..: 3 6 1 1 6 4 1 4 3 1 ...
 $ year       : num  70 70 70 70 70 70 70 70 70 70 ...
 $ cylinders  : num  8 8 8 8 8 8 8 8 8 8 ...
```