Problem Set 1

The purpose of this problem set is to get comfortable with R and its facilities. We shall spend most of the time doing some basic computations. If you are a good programmer you will finish these computations quickly. First start by opening R, create a new script, and save it to your hard drive. At each computation you may use `help` or `??` for assistance.

Problem 1: Importing data

1. Set the working directory
2. Import the dataset WestRoxbury.csv (available on Brightspace) into RStudio. Pay attention to the type of datafile and the associated command (*read.csv*). Check that the data set has been loaded correctly. Are variables names part of the data (as a first row)? If so, it is incorrect.

   Alternatively, type and adapt the following command:

   ```
   housing.df <- read.csv("...filepath ...  /WestRoxbury.csv", header = TRUE)
   ```

Problem 2: Basic data exploration

1. Explore the following functions or commands in your dataset
   (a) `dim()` to find the dimensions of the dataset
   (b) `head()` to show the first x rows
   (c) `View()` to show all the data in a new tab
2. Show the first 10 rows of the first column only
3. Show the first 10 rows of each of the columns
4. Show the fifth row of the first 10 columns
5. Show the whole first column
6. Find the mean of the first column
7. Find summary statistics for each column
8. Check the type of variables using the command `str()`. Which of the variables are categorical variables?
9. Convert the categorical variables to factors and identify the levels of the factor variables using the commands

   `dataname$variablename= as.factor(dataname$variablename)` and

   `levels(dataname$variablename)`
10. Export your new data using the command

    `write.csv(dataname, "new_dataname.csv")`

Problem 3: Programming

1. Define `v1 <- c(1, 2, 2, 1)` and `v2 <- c(2, 3, 3, 2)`. Do element-wise addition, subtraction, and multiplication. Perform concatenation between `v1` and `v2` into a new vector `v3`.

2. Define a 2×3 matrix, `mA`, of your choice.

   (a) Print the minimum and the maximum of each row.

   (b) Compute the sum of each column and print it as a vector.

   (c) Sort in ascending order the elements on the first column of `mA`

3. Consider the function $y = f(x)$ defined by:

$$f(x) = \begin{cases} -x^3, & \text{if } x \leq 0 \\ x^2, & \text{if } x \in (0, 1] \\ \sqrt{x}, & \text{if } x > 1. \end{cases}$$

   Supposing that you are given x, write an R expression for y using `if` statements.

4. Let $h(x, n) = 1 + x + x^2 + x^3 + \cdots + x^n = \sum_{i=0}^{n} x^i$. For a given x and n write an R function that calculates $h(x, n)$ using a `for` loop.

5. Write a function that achieves the same result as in 4 but using a `while` loop.

6. Give R expressions that return the following matrices `mB` and `mC` without explicitly writing it element by element:

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 2 & 3 & 0 \\ 0 & 6 & 0 & 0 \\ 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix}$$

7. A room contains 100 toggle switches, originally all turned off. 100 people enter the room in turn. The first one toggles every switch, the second one toggles every second switch, the one third every third switch, and so on until the last person, who toggles the last switch only. At the end of this process, which switches are turned on? *Note: This requires a little thinking. Don't give up!*

Optional: Get more R practice

"Lab: Introduction to R" in the ISL textbook (p. 42, 2nd edition) includes some extra (beginner-level) exercises with solutions. Also consider the Applied Exercises 8, 9, and 10 (p. 54, 2nd edition), which are enjoyable to perform, more exam-relevant, and won't need more than 1h of labor in total.