

## Homework # 1

### Fundamentals, R Data, Data Wrangling, Visualization I

This assignment covers the contents of the first four weeks of the lab where we (i) installed R and R Studio, (ii) introduced both programs together with (iii) the fundamentals of working with them, and learnt the basics of (iv) R Data, (v) data cleansing/wrangling and (vi) visualization.

Submit an R script with code that produces the tasks requested in this assignment. I will evaluate your submission by running your code. For written/explanatory answers you can either add comments to your code (I think this will make both of our lives easier), or submit a pdf/word/etc document with anything you want/need to explain. Please be as concise as possible in your written answers, if any. As long as you answer what you are supposed to, I will not grade based on the length of your explanation.

You can work in groups, consult class materials and code found online; but please write your own code and submit one assignment per student.

**Due:** End of day on February 16/17 (POLI 001 / 002), 2023

**How:** Canvas assignment submission link

**File:** hw01-Lastname.R

**Max grade:** 10 points

## 1 Fundamentals

- (a) [0.25 pts] Install and load the `legislatoR` R package.
- (b) [0.5 pts] Using appropriate commands learn and describe something about the package.
- (c) [0.75 pts] Create a vector of numbers ranging from 5 to 130 increasing by 3 at a time. Show your vector. What's the type and class of this vector? Change it to integer.
- (d) [1 pt] Create a list of 5 elements: the vector you created, an empty element, a Boolean produced by asking some logic operation to R, the names of three things you like most in life (activities, food/drinks, favourite movies/books, whatever), and the absolute value of the negative of your vector squared. Show your list. What's the type and class of this list? What about the elements in it?
- (e) [2 pts] Create two vectors, `a` and `b`, that contain the same number of numbers. Using a loop and an `if...else` statement create a vector `c` that equals 'b is greater than a' if  $b > a$ , 'b is smaller than a' if  $a > b$ , and 'they are equal' otherwise. Show your `c` vector.  
*HINT: You'll need to create an empty `c` object in advance, and use `[ ]` to index accordingly when making use of your loop iterations.*

## 2 Bringing In and Creating Data

- (a) [0.25 pts] Install and load the `poliscidata` R package. Assign the `nes` data set from this package to an object in your environment.
- (b) [0.25 pts] Import to your environment the `hw01-data.csv` file.
- (c) [1 pt] Create a data frame with three variables setting seed by `set.seed(0311)`: `var1` is a binary variable having 100 observations, which are randomly drawn from binomial distribution,

the probability of drawing 1 is 0.5, and the same for 0 (you're modeling individual Bernoulli trials); `var2` is a continuous variable of 100 observations randomly drawn from a normal distribution with mean 0 and standard deviation of 1; `var3` is a continuous variable of 100 observations randomly drawn from a uniform distribution ranging from 0 to 1. Save these data as a CSV (comma-separated values) file in your working directory. Is there a way you can save this without row names?

### 3 Data Cleansing and Visualization

- (a) [1 pt] Plot() the `married` variable from the `nes` data set. Using the function arguments properly, specify a graph title, labels for X and Y axes, and two different colors for the variable's categories. What type of plot/visualization does `plot()` produce? Describe what the plot shows.
- (b) [3 pts] Using the `hw01-data` data:
  - (i) [0.5 pts] Create a new `expenditure_log` variable that is the logarithmic transformation of the `expenditure` variable.
  - (ii) [0.5 pts] Create a new `winning` variable indicating 1 if `voteshare` is greater than 0.5, 0 otherwise.
  - (iii) [0.5 pts] Change the variable name of `gender` into `gender_text`.
  - (iv) [0.5 pts] Create a new `incumbent_gender` variable with four categories that indicate whether the observation corresponds to an 'Incumbent male', 'Incumbent female', 'Non-incumbent male' or a 'Nonincumbent female'.
  - (v) [1 pts] Produce a cross tab and a plot of your choice to show the frequency distribution of the `incumbent_gender` considering the `winning` variable. Describe what you see.