

```
# Tutorial 2: Basics of R"
# Submission Deadline: 5 Sept 2022 9am
```

## ## Learning Objectives

```
# In this tutorial, we will review and practice applying the concepts
related to the Basics of R.
# We will focus mainly on doing simple data manipulations with R objects
such as vectors and dataframes.
# R objects such as matrix and lists which are less used in this module
will be covered again later in the module.
```

```
#- Part 1 will be done during the lab session in Week 3. You may save
your answers as "T2-1[matric no].R". You will
# need to show your TA your answers to part 1 to earn your lab credits.
#- Type your answers for questions in Part 2 using R script and save your
file as "T2-[matric no].R" (eg if your
# matric number is A12345J then save your file as T2-A12345.R) and
upload to CANVAS.
#- You will discuss the answers to questions in Part 2 during the
Tutorial session in week 5.
```

```
# Note that we use the back ticks (` `) to denote an R object eg `Orange`
means Orange is an R object. When you are
# asked to assign an output to `df2`, the R object is simply df2 and not
`df2`.
```

```
# We also use the hex sign (#) to denote a comment. Anything that appears
after a # on a line will not be executed by R.
# You can use it to provide comments of your answers. You should also use
it to label the question numbers for your
# answers and to provide any textual answers that are required by the
questions.
```

```
# Part 1: To be completed in Week 3 Lab
```

```
#### 1) We will start by exploring the built-in dataset called
`ToothGrowth`. To find out more about this dataset, type ?ToothGrowth in
the R command line.
```

```
# - What do each of the following functions do? (Hint: You may use the
Help menu or ?<function> where <function> is the function name e.g. ?
summary, to find out)
```

```
# i) summary()
# ii) head()
# iii) tail()
# iv) str()
```

```
# Type your answers below.
```

```
#### 2) Selecting data
```

# - There are several variables in `ToothGrowth`. Using Base R and dplyr functions, can you perform (i), (ii) and (iii)?

# i) Extract the column `supp`  
# ii) Extract rows where `supp` is equal to "VC" and `dose` is less than 1 and assign the output to df2  
# iii) Extract the values of `len` where `supp` is equal to "VC"  
# iv) Try to perform the above operations (i, ii, iii) again but this time, assign the output to df2.1, df2.2  
#       and df2.3 respectively.  
# v) Use the class function to check the class attribute for each of the outputs. Use is.data.frame function to check whether the output is a dataframe or a vector.

# Type your answers below.

#i

#ii

#iii

#### 3) Adding/Removing/Changing data columns for Toothgrowth data.

# - i) Change the variable name from `len` to `length` and assign the output to df3.1  
# - ii) Increase the value of len by 0.5 if supp is equal to OJ and assign the output to df3.2  
# - iii) Remove the column `dose` from the data and assign the output to df3.3  
# - iv) Increase the value of `dose` by 0.1 for all records and rename `dose` to `dose.new` and assign output to df3.4  
# - v) Create a new variable `high.dose` and assign it a value of "TRUE" if `dose` is more than 1 and "FALSE" if  
#       `dose` is less than or equal to 1. Assign the dataframe with the new variable `high.dose` to df3.5.  
#   Export df3.5 to a csv file. Discuss what is the r code to export as an excel file (.xlsx).

# Type your answers below.

# i

# ii

# etc.

#### 4) Sorting

# - i) There are two functions in Base R "sort" and "order" to perform sorting. How do these two functions differ?  
#       Try to do a sort with each function on ToothGrowth\$len.  
# - ii) Using a base R function (e.g. order), how can you sort the dataframe `ToothGrowth` in decreasing order of `len`?  
# - iii) What dplyr functions can you use to sort `ToothGrowth` in increasing order of `len`?  
#       Can you also sort the dataframe in decreasing order of `len`?

# Type your answers below.

# i

# ii

# iii

#### 5) Factors

# - i) Check if `supp` is a factor vector. First type `ToothGrowth$supp`. What do you observe with the output?

# - ii) Next use `is.factor()` and `is.ordered()` to check if `supp` is a factor and if so whether it is an ordered factor.

# - iii) Now suppose we find that vitamin C (VC) is a superior supplement compared to orange juice (OJ), and we

# want to order `supp` such that VC is a higher level than OJ, how could we do this?

# Type your answers below.

# i

# ii

# iii

### PART 2 (15 marks)

# For this part of the tutorial, you will be using the built-in dataset `trees`.

# This dataset provides measurements of the diameter, height and volume of timber in 31 felled black cherry trees.

# Note that the diameter (in inches) is erroneously labelled Girth in the data. It is measured at 4 ft 6 in above the ground.

# The 3 variables are defined as follows:

# - Girth: Tree diameter (rather than girth, actually) in inches

# - Height: Height in ft

# - Volume: Volume of timber in cubic ft

#### 1) Inspect the dataset (2 marks)

# - Use the functions you have learnt in Part 1 of this tutorial to inspect the dataset.

# Describe this dataset in terms of the number of observations, number of variables, and type of variables.

#### 2) Data Extraction (6 marks)

# - i) Assign the dataset `trees` to `dft` (Note: 0 is the capital letter of o and not the number zero)

# - ii) Extract the columns `Height` and `Volume` from `dft` and assign it to `dft2ii`.

# Export `dft2ii` as a csv file. (2 marks)

# - iii) Using Base R functions, extract the rows from `dt` where `Volume` is greater than 22.

```
#           How many rows are extracted?
# - iv) Using dplyr functions, remove the `Volume` column and retain only
the rows where `Girth` is greater than 12
#           and Height is less than 78 and assign this output to `dft2iv`.
How many observations are there in `dft2iv`?[2 marks)
```

#### 3) Variables (4 marks)

```
# - i) Rename the variable in `dft` from `Girth` to `Diameter`
# - ii) Convert the values in `Diameter` from inches to centimeters
[ hint: 1 inch = 2.54cm]
# - iii) Create a new *factor* variable in `dft` called `Size`. `Size` is
an ordered factor with two values "Small"
#           and "Large". Trees are considered "Large" if their volume is
larger than 30 or height is greater than 80,
#           otherwise they are considered "Small". Assign the values to the
variable `Size` based on this definition. (2 marks)
```

```
# You may use dplyr or base R functions for this question part.
```

#### 4) Sorting (3 marks)

```
# - i) Using base R, sort `dft` in increasing order of `Size`. How many
large and small trees are there?
# - ii) Using dplyr, sort `dft` in decreasing order of `Size` followed by
decreasing order of `Volume`. The output
#           should have the observations arranged in decreasing order of Height
first and within the same level of Size,
#           the observations should be arranged in decreasing order of Volume. (2
marks)
```