



## School of Computing & Information Technology

# ISCG8026 Introduction to Data Science Semester 1 – 2020

## Lab Exercises 1 & 2

Due Date: 15 March 2020, 13.00

Weight: 10%

For this set of lab exercises, you will be working in groups of 2-3 people. It is worth **10%** of your final grade and it must be submitted on Moodle by **13.00 on Sunday 15 March 2020**.

1. Install R & Rstudio on your laptop, if you are working on your laptop and haven't got the packages installed (refer to slides for URLs/instruction). The R packages should already be installed on the lab PCs.
2. Set up a Github account, if you don't have one already - you may use your own name or a pseudonym. If you are working in a group, *each individual member has to set up their own Github account*.
3. Create a repo called `datascienceunitec`. Include the link to your GitHub accounts
4. Create a text file called *HelloWorld.md*. Add the line "## This is a markdown file" (without the quotation marks) to the document. Push the document to the `datascienceunitec` repo you created on Github. Submit the links to the *HelloWorld.md* file on your Github repo. [0.5]
5. Fork the data sharing repository here: <https://github.com/nnuntalid/DataScienceS12020>  
Submit the link to the forked repository on your Github account. [0.25]
6. In R the following are all atomic data types EXCEPT: (Select all that apply) [0.5]
  - a) table
  - b) numeric
  - c) character
  - d) matrix
  - e) list
  - f) complex
  - g) integer
  - h) data frame
  - i) array
  - j) logical
7. What is the class of the object defined by the expression `x <- c(4, "a", TRUE)`? [0.25]
  - a) Logical
  - b) integer
  - c) character
  - d) numeric

e) mixed

8. If I have two vectors `x <- c(1,3, 5)` and `y <- c(3, 2, 10)`, what is produced by the expression `cbind(x, y)`? [0.5]
- a) a 2 by 3 matrix
  - b) a 2 by 2 matrix
  - c) a vector of length 3
  - d) a matrix with 2 columns and 3 rows
  - e) a 3 by 3 matrix
  - f) a vector of length 2
9. Suppose I have a list defined as `x <- list(2, "a", "b", TRUE)`. What does `x[[2]]` give me? Select all that apply. [0.25]
- a) a list containing character vector with the letter "a".
  - b) a list containing the number 2 and the letter "a".
  - c) a character vector with the elements "a" and "b".
  - d) a character vector of length 1.
  - e) a character vector containing the letter "a".
10. Suppose I have a vector `x <- 1:4` and `y <- 2:3`. What is produced by the expression `x + y`? [0.25]
- a) a numeric vector with the values 1, 2, 5, 7.
  - b) a numeric vector with the values 3, 5, 3, 4.
  - c) an numeric vector with the values 3, 5, 5, 7.
  - d) an error
  - e) a warning
  - f) an integer vector with the values 3, 5, 5, 7.
  - g) an integer vector with the values 3, 5, 3, 4.
11. Suppose I have a vector `x <- c(3, 5, 1, 10, 12, 6)` and I want to set all elements of this vector that are less than 6 to be equal to zero. What R code achieves this? Select all that apply. [0.5]
- a) `x[x == 0] < 6`
  - b) `x[x %in% 1:5] <- 0`
  - c) `x[x > 0] <- 6`
  - d) `x[x > 6] <- 0`
  - e) `x[x < 6] == 0`
  - f) `x[x != 6] <- 0`
  - g) `x[x == 6] <- 0`
  - h) `x[x >= 6] <- 0`
  - i) `x[x < 6] <- 0`
  - j) `x[x <= 5] <- 0`
  - k) `x[x == 0] <- 6`
12. Use the Session 1&2 Data Set (available on Moodle) to answer Questions 12-17. Extract the first 2 rows of the data frame and print them to the console. What does the output look like? [0.25]

13. How many observations (i.e. rows) are in this data frame? [0.25]

- a) 45
- b) 129
- c) 153
- d) 160

14. What is the value of Ozone in the 47th row? [0.25]

- a) 63
- b) 18
- c) 34
- d) 21

15. How many missing values are in the Ozone column of this data frame? [0.25]

- a) 37
- b) 9
- c) 43
- d) 78

16. What is the mean of the Ozone column in this dataset? Exclude missing values (coded as NA) from this calculation. [0.25]

- a) 42.1
- b) 31.5
- c) 18.0
- d) 53.2

17. What was the maximum Ozone value in the month of May (i.e. Month is equal to 5)? [0.25]

- a) 97
- b) 115
- c) 18
- d) 100

18. Suppose I define the following function in R: [0.25]

```
cube <- function(x, n) {  
  x^3  
}
```

What is the result of running: cube(3)

- a) An error is returned because 'n' is not specified in the call to 'cube'
- b) The number 27 is returned

- c) A warning is given with no value returned.
- d) The users is prompted to specify the value of 'n'.

19. The following code will produce a warning in R. [0.5]

```
x <- 1:10
if(x > 5) {
  x <- 0
}
```

Why?

- a) There are no elements in 'x' that are greater than 5
- b) You cannot set 'x' to be 0 because 'x' is a vector and 0 is a scalar.
- c) The syntax of this R expression is incorrect.
- d) 'x' is a vector of length 10 and 'if' can only test a single logical statement.
- e) The expression uses curly braces.

20. Consider the following function [0.5]

```
f <- function(x) {
  g <- function(y) {
    y + z
  }
  z <- 4
  x + g(x)
}
```

If I then run in R

```
z <- 10
f(3)
```

What value is returned?

- a) 10
- b) 7
- c) 16
- d) 4

21. Consider the following expression: [0.25]

```
x <- 5
y <- if(x < 3) {
  NA
} else {
  10
}
```

What is the value of 'y' after evaluating this expression?

- a) 3
- b) 10
- c) 5
- d) NA

22. Consider the following R function

[0.5]

```
h <- function(x, y = NULL, d = 3L) {

  z <- cbind(x, d)
  if(!is.null(y))
    z <- z + y
  else
    z <- z + f
  g <- x + y / z

  if(d == 3L)
    return(g)
  g <- g + 10

  g
}
```

Which symbol in the above function is a free variable?

- a) f
- b) z
- c) d
- d) l
- e) g

23. What is an environment in R?

[0.25]

- a) an R package that only contains data
- b) a collection of symbol/value pairs

- c) a list whose elements are all functions
- d) a special type of function

24. The R language uses what type of scoping rule for resolving free variables? [0.25]

- a) global scoping
- b) compilation scoping
- c) dynamic scoping
- d) lexical scoping

25. How are free variables in R functions resolved? [0.5]

- a) The values of free variables are searched for in the working directory
- b) The values of free variables are searched for in the environment in which the function was defined
- c) The values of free variables are searched for in the global environment
- d) The values of free variables are searched for in the environment in which the function was called

26. What is one of the consequences of the scoping rules used in R? [0.5]

- a) R objects cannot be larger than 100 MB
- b) All objects must be stored in memory
- c) Functions cannot be nested
- d) All objects can be stored on the disk

27. In R, what is the parent frame? [0.5]

- a) It is the environment in which a function was called
- b) It is the environment in which a function was defined
- c) It is always the global environment
- d) It is the package search list

28. Take a look at the 'iris' dataset that comes with R. The data can be loaded with the

```
code: library(datasets)
```

```
data(iris)
```

A description of the dataset can be found by

```
running ?iris
```

There will be an object called 'iris' in your workspace. In this dataset, what is the mean of 'Sepal.Length' for the species *virginica*? **Please round your answer to the nearest whole number.** [0.5]

29. Continuing with the 'iris' dataset from the previous Question, what R code returns a vector of the means of the variables 'Sepal.Length', 'Sepal.Width', 'Petal.Length', and 'Petal.Width'? [0.5]

- a) colMeans(iris)
- b) apply(iris, 1, mean)
- c) apply(iris, 2, mean)
- d) apply(iris[, 1:4], 1, mean)
- e) apply(iris[, 1:4], 2, mean)
- f) rowMeans(iris[, 1:4])

30. What's the value of f(3)?

[0.5]

```
y<-10  
f<-function (x) {  
  y <- 2  
  y^2 + g(x)  
}  
g<-function(x) {  
  x*y  
}
```

### **Late Submission of Assignments**

Assignments submitted after the due date and time without having received an extension through Special Assessment Circumstances (SAC) will be penalised according to the following:

- 10% of marks deducted if submitted within 24hrs of the deadline
- 20% of marks deducted if submitted after 24hrs and up to 48hrs of the deadline
- No grade will be awarded for an assignment that is submitted later than 48hrs after the deadline

Assignments submitted in more than 48 hours late will not be marked unless Special Assessment Circumstances apply. So, it is better to submit an incomplete assignment on time.

### **Special Assessment Circumstances**

A student, who due to circumstances beyond his or her control, misses a test, final exam or an assignment deadline or considers his or her performance in a test, final exam or an assignment to have been adversely affected, should complete the Special Assessment Circumstances (SAC) form available from Student Central. Within any semester, a student may have only one SAC per course. When requesting an SAC for an assignment, the SAC application form must be submitted (along with the work completed to date) within the time frame of the extension requested; i.e. if the Doctor's certificate is for one (1) day, then the SAC application form and work completed must be submitted within one (1) day.

### **Assistance to other Students**

Students themselves can be an excellent resource to assist the learning of fellow students, but there are issues that arise in assessments that relate to the type and amount of assistance given by students to other students. It is important to recognise what types of assistance are beneficial to another's learning and also what types of assistance are unacceptable in an assessment.

#### **Beneficial Assistance**

- Study Groups.
- Discussion.
- Sharing reading material.
- Testing another student's programming work using the executable code and giving them the results of that testing.

#### **Unacceptable Assistance**

- Working together on one copy of the assessment and submitting it as own work.
- Giving another student your work.
- Copying someone else's work. This includes work done by someone not on the course.
- Changing or correcting another student's work.
- Copying from books, Internet etc. and submitting it as own work. Anything taken directly from another source must be acknowledged correctly: show the source alongside the quotation.

Do you want to do the best that you can do on this assignment and improve your grades? You could:

- Talk it over with your lecturer
- Visit Student Success and Achievement for learning advice and support (in Te Puna)
- Visit the Centre for Pacific Development and Support