

Assignment 2: Coding Basics

Siyang Chen

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics (ENV872L) on coding basics in R.

Directions

1. Change “Student Name” on line 3 (above) with your name.
2. Use the lesson as a guide. It contains code that can be modified to complete the assignment.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document. Space for your answers is provided in this document and is indicated by the “>” character. If you need a second paragraph be sure to start the first line with “>”. You should notice that the answer is highlighted in green by RStudio.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file. You will need to have the correct software installed to do this (see Software Installation Guide) Press the **Knit** button in the RStudio scripting panel. This will save the PDF output in your Assignments folder.
6. After Knitting, please submit the completed exercise (PDF file) to the dropbox in Sakai. Please add your last name into the file name (e.g., “Salk_A02_CodingBasics.pdf”) prior to submission.

The completed exercise is due on Thursday, 24 January, 2019 before class begins.

Basics Day 1

1. Generate a sequence of numbers from one to 100, increasing by fours. Assign this sequence a name.
2. Compute the mean and median of this sequence.
3. Ask R to determine whether the mean is greater than the median.
4. Insert comments in your code to describe what you are doing.

```
# Generate a sequence from 1 to 100, increasing by 4
seq(1, 100, 4)
```

```
## [1] 1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89
## [24] 93 97
```

```
# Assign the sequence to seq_by_4
seq_by_4 <- seq(1, 100, 4)
```

```
# Compute the mean of the sequence
mean(seq_by_4)
```

```
## [1] 49
```

```
# Compute the median of the sequence
median(seq_by_4)
```

```
## [1] 49
```

```
# Determine whether the mean is greater than the median
mean(seq_by_4) > median(seq_by_4)
```

```
## [1] FALSE
```

Basics Day 2

5. Create a series of vectors, each with four components, consisting of (a) names of students, (b) test scores out of a total 100 points, and (c) whether or not they have passed the test (TRUE or FALSE) with a passing grade of 50.
6. Label each vector with a comment on what type of vector it is.
7. Combine each of the vectors into a data frame. Assign the data frame an informative name.
8. Label the columns of your data frame with informative titles.

```
# Create vectors
name <- c('John', 'Sara', 'Janet', 'Paul', 'Alice') # Character vector
score <- c(80, 75, 45, 60, 100) # Numeric vector
passing <- c(TRUE, TRUE, FALSE, TRUE, TRUE) # Logical vector

# Combine vectors into dataframe
student_grades <- data.frame(name, score, passing)

# Label dataframe columns
names(student_grades) <- c('Name', 'Score', 'Passed'); View(student_grades)
```

```
## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/
## Resources/modules/R_de.so'' had status 1
```

9. QUESTION: How is this data frame different from a matrix?

ANSWER: Data frame is a two-dimensional structure that is more general than a matrix. Data frame can have columns of different types of data, but matrix can only have columns with same type of data.

10. Create a function with an if/else statement. Your function should determine whether a test score is a passing grade of 50 or above (TRUE or FALSE). You will need to choose either the `if` and `else` statements or the `ifelse` statement. The name of your function should be informative.
11. Apply your function to the vector with test scores that you created in number 5.

```
# Create a function with if/else statement
pass_or_not <- function(x){
  if(x > 50) {TRUE}
  else {FALSE}
}

# Or, create a function with ifelse statement
pass_or_not_ifelse <- function(x){
  ifelse(x>50, TRUE, FALSE)
}

# Apply function to the score vector
pass_or_not_ifelse(score)
```

```
## [1] TRUE TRUE FALSE TRUE TRUE
```

12. QUESTION: Which option of `if` and `else` vs. `ifelse` worked? Why?

ANSWER: Only the ifelse statement work in this situation. When running the if and else statement, there's a warning message saying that the condition/vector has more than one elements and thus only the first element will be used. In other words, the if and else statement only run for the first number in the score vector.