Module 1 Assignment 2

Ellen Bledsoe

2024-09-17

Purpose

The goal of this assignment is to explore computational reproducibility and apply the base R coding skills we've learned and practiced in class and lab.

Task

Write R code to successfully answer each question below and/or write text to successfully answer the question.

Criteria for Success

- Code is within the provided code chunks
- Code chunks run without errors
- Code produces the correct result
  - Code attempts will clear logical direction will get half points
  - Code that produces the correct answer will receive full points
- Text answers correctly address the question asked

Due Date

Sept 24 at 1pm MST

Assignment Questions

This assignment is worth 20 points total. Each question is worth 1 point.

Definitions (1 point each)

In your own words, define/describe the following terms. These don't need to be technical descriptions but rather how you are thinking about them.

Grading note: I've copied these definitions from my slides. If their definitions are the same, they have copied directly from the slides and not used their own words, so take 0.5 points off.

1. *Reproducibility*: ability to repeat the original study using the same data, materials, and methods

2. *Open science*: Scientific research conducted and communicated in an honest, accessible, and transparent way, such that–at a minimum–a study can be reproduced and/or replicated.
3. *R*: R refers to both the programming language and the software that interprets scripts written in the language.
4. *RStudio*: RStudio is an integrated development environment (IDE) that helps us interact with R more easily.

Vectors (1 point each)

5. Assign the number 120 to an object called `birds`.

```
birds <- 120
birds
```

```
## [1] 120
```

6. Divide the object `birds` by 5. Save the result as an object called `fewer_birds`.

```
fewer_birds <- birds / 5
fewer_birds
```

```
## [1] 24
```

7. Create a vector called `snails` with the following values: 7, 2, 19, 4, 0, 2

```
snails <- c(7, 2, 19, 4, 0, 2)
snails
```

```
## [1]  7  2 19  4  0  2
```

8. Write a line of code that determines what *class* of data is in the `snails` vector.

```
class(snails)
```

```
## [1] "numeric"
```

9. Run the following line of code. Based on the output, explain what the line of code does (~ 1 sentence).

```
snails * fewer_birds
```

```
## [1] 168  48 456  96   0  48
```

Answer:

10. Create a character vector called `rhymes` that contains the following values: cat, rat, bat, hat

```
rhymes <- c("cat", "rat", "bat", "hat")
rhymes
```

```
## [1] "cat" "rat" "bat" "hat"
```

Let's Bring in the Cactus Data

For the following questions, we will be using the cactus pad data that we collected a few weeks ago. Run the following code chunk to bring ("read") the .csv file into R for us to use.

```
cactus <- read.csv("../data_raw/CactusPads_Fall2024.csv")
```

This code chunk uses the `$` operator to pull out individuals columns from 2D data and treat them as vectors. Here, I'm saving the cactus pad length column as a vector called `pad_length`. Once you run this code chunk, you should see `pad_length` in your environment.

```
pad_length <- cactus$length_in
pad_length
```

```
##   [1]  7.250  7.500  9.250  9.500  8.000  4.500  9.250  7.500  6.000  7.000
##  [11] 10.000 10.500 13.500 10.000 16.000 10.000 11.000 12.000 11.000 11.000
##  [21]  8.500  7.500 12.000 10.000  7.500 12.000  9.500  8.000  8.500  7.000
##  [31] 10.750 13.500 12.500  8.500  7.750 11.500 13.000 15.500 12.000 11.500
##  [41]  3.125  5.875  5.625  5.625  3.750  4.625  4.500  7.000  6.125  6.000
##  [51]  8.500  9.250  7.400  3.200  7.500  9.700  8.500  9.300  9.000 10.500
##  [61]  5.000  7.000 10.500 13.250 10.000  7.500  9.000  9.000  7.500  9.000
##  [71] 15.000 12.500 15.000 12.000 15.200 10.000  8.500 16.000 10.800 13.000
##  [81]  6.500  6.000  6.300  7.500  9.000  9.100  7.500  7.800  8.600  7.000
##  [91]  9.000  5.000  4.500  6.000  5.500  6.000  4.500  3.500  5.500  6.000
## [101] 13.500 10.000 12.000 12.000 11.000 10.000  9.500 14.000 13.000  9.500
## [111]  7.500 10.500  8.500  9.500 10.000  9.000  9.800 11.000 10.200 11.000
```

Below, I'm doing the same thing with a couple more columns.

```
pad_width <- cactus$width_in
damage <- cactus$damage

pad_width
```

```
##   [1]  2.500  2.250  4.000  3.500  5.500  1.750  4.500  3.500  2.500  3.000
##  [11] 10.500  9.000  6.500 10.500  5.500  7.500  7.500  8.000  7.500  9.000
##  [21]  3.500  3.500  5.500  6.000  3.000  6.000  7.000  4.500  4.000  4.000
##  [31]  7.500 11.500  9.300  6.500  4.500  8.000  7.000 11.500  8.250 10.500
##  [41]  3.250  4.500  5.375  4.250  4.125  4.500  4.000  5.750  4.625  4.750
##  [51]  4.700  5.500  2.750  1.600  4.500  4.500  4.500  5.200  6.500  4.900
##  [61]  2.500  4.000  6.000  7.500  6.000  3.000  4.000  6.000  4.000  4.500
##  [71] 12.000  9.000  9.500  9.000  5.800  3.000 13.800  5.500  8.000  9.000
##  [81]  2.500  1.500  2.800  3.500  4.000  2.600  3.500  3.400  3.100  3.200
##  [91]  6.500  5.000  4.000  4.000  5.000  5.000  4.000  1.500  4.500  4.750
## [101]  9.000  8.500  8.500  8.000  8.000  7.000  7.000 10.500 10.000  6.000
## [111]  5.500  7.000  5.000  5.800  6.900  4.300  5.000  6.200  7.000  6.800
```

```
damage
```

```
##   [1] "None" "None" "None" "None" "Some" "None" "Some" "Some" "None" "None"
##  [11] "Most" "Most" "None" "Most" "None" "Some" "All"  "Some" "Some" "Some"
##  [21] "None" "None" "None" "None" "None" "None" "None" "None" "None" "None"
##  [31] "None" "Some" "Some" "None" "Some" "None" "None" "Some" "Some" "Most"
##  [41] "Some" "Some" "Some" "Most" "Some" "Some" "Some" "Some" "Some" "Some"
##  [51] "None" "Some" "None" "None" "Some" "Some" "None" "None" "Some" "Most"
##  [61] "None" "Some" "Most" "None" "None" "None" "None" "Some" "None" "None"
##  [71] "None" "None" "None" "Some" "Most" "None" "All"  "None" "Some" "Some"
##  [81] "None" "None" "None" "Some" "Some" "None" "Some" "Some" "None" "None"
##  [91] "Some" "Some" "Some" "Some" "Most" "Most" "Most" "Most" "Some" "Some"
## [101] "None" "Most" "None" "Some" "None" "Some" "None" "Most" "Some" "None"
## [111] "None" "Some" "None" "None" "None" "Some" "None" "Some" "Most" "Most"
```

---

11. Write a line of code that pulls out the 4th value in the `pad_width` vector.

```
pad_width[4]
```

```
## [1] 3.5
```

12. Write a line of code that saves the 42nd value in the `pad_width` column as an object called `answer`.

```
answer <- pad_width[42]
answer
```

```
## [1] 4.5
```

13. Write a line of code that pulls out the 6th through the 10th value in the `pad_width` vector.

```
pad_width[6:10]
```

```
## [1] 1.75 4.50 3.50 2.50 3.00
```

14. Using the `length()` function, find the number of elements in the `damage` vector.

```
length(damage)
```

```
## [1] 120
```

15. Write a line of code that returns all of the elements of the `damage` vector which have no damage on the pad.

```
damage[damage == "None"]
```

```
##  [1] "None" "None" "None" "None" "None" "None" "None" "None" "None" "None"
## [11] "None" "None" "None" "None" "None" "None" "None" "None" "None" "None"
## [21] "None" "None" "None" "None" "None" "None" "None" "None" "None" "None"
## [31] "None" "None" "None" "None" "None" "None" "None" "None" "None" "None"
## [41] "None" "None" "None" "None" "None" "None" "None" "None" "None" "None"
## [51] "None" "None" "None" "None" "None" "None"
```

16. Combining your answers for Questions 13 and 14, write a line of code that determines how many recorded cactus pads had no damage to them.

```
length(damage[damage == "None"])
```

```
## [1] 56
```

17. Write a line of code that returns all of the elements in the `pad_length` vector which are greater than 10.

```
pad_length[pad_length > 10]
```

```
##  [1] 10.50 13.50 16.00 11.00 12.00 11.00 11.00 12.00 12.00 10.75 13.50 12.50
## [13] 11.50 13.00 15.50 12.00 11.50 10.50 10.50 13.25 15.00 12.50 15.00 12.00
## [25] 15.20 16.00 10.80 13.00 13.50 12.00 12.00 11.00 14.00 13.00 10.50 11.00
## [37] 10.20 11.00
```

18. Calculate both the average (using the `mean()`) and the standard deviation (`sd()`) for all of the cactus pad lengths.

```
mean(pad_length)
```

```
## [1] 9.090833
```

```
sd(pad_length)
```

```
## [1] 2.898737
```

19. Calculate both the average (`mean()`) and the standard deviation (`sd()`) for the cactus pad widths.

```
mean(pad_width)
```

```
## [1] 5.702292
```

```
sd(pad_width)
```

```
## [1] 2.505664
```

20. Compare the results of Questions 18 and 19. Are cactus lengths or widths larger, on average? Which of the two sets of data has greater variability, and what does that mean? (~2 sentences)

Answer: Cactus pad lengths are longer, on average, than cactus pad widths. The lengths also have more variability in their values, meaning that the lengths are further away from the mean than the widths are.

Turning in Your Assignment

Follow these steps to successfully turn in your assignment on D2L.

1. Click the `Knit` button up near the top of this document. This should produce a PDF file that shows up in the `Files` panel on the bottom-right of your screen.
2. Click the empty box to the left of the PDF file.
3. Click on the blue gear near the top of the `Files` panel and choose Export.
4. Put your last name at the front of the file name when prompted, then click the Download button. The PDF file of your assignment is now in your "Downloads" folder on your device.
5. Head over to D2L and navigate to Module 1 Assignment 2. Submit the PDF file that you just downloaded.