Module 1 Assignment 3

Ellen Bledsoe

2024-09-24

Purpose

The goal of this assignment is to explore both 1D (vectors) and 2D (data frames) data in R.

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 get half points
    - Code that produces the correct answer will receive full points

- Text answers correctly address the question asked

Due Date

Oct 1 at 1pm MST

Assignment Questions

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

Vectors

First, let's remind ourselves a little bit about working with vectors.

Run the following line of code to create a vector called `counts`.

```
counts <- c(63, 69, 60, 65, NA, 68, 61, 70, 61, 59, 64, 69, 63, 63, 72, 65, 64, 70, 63, 65)
```

1. Use the `mean()` function to calculate the mean value of `counts`. Do not add any arguments beyond the name of the vector.

Hint: remember to check the answer key!

```
mean(counts)
```

```
## [1] NA
```

2. The answer to Q1 should yield an odd result caused by the `NA` value in `counts`. To resolve this, use the help function to remind yourself about the argument `na.rm = TRUE` in the `mean()` function.

Issue a revised command to calculate the mean value in `counts`.

```
mean(counts, na.rm = TRUE)
```

```
## [1] 64.94737
```

3. Write a line of code that selects values less than 66 from the `counts` vector. The `NA` values with be included in this vector, which is fine.

```
counts[counts < 66]
```

```
##  [1] 63 60 65 NA 61 61 59 64 63 63 65 64 63 65
```

4. Create a character vector called `pet_names` which includes the following names: Mystery, Hazel, Spud, Pancho

```
pet_names <- c("Mystery", "Hazel", "Spud", "Pancho")
pet_names
```

```
## [1] "Mystery" "Hazel"   "Spud"    "Pancho"
```

5. Write a line of code using *sub-setting via indexing* to pull out the element "Spud" from the vector `pet_names`.

```
pet_names[3]
```

```
## [1] "Spud"
```

6. Now, write a line of code using *conditional sub-setting* that pulls out the element "Spud".

```
pet_names[pet_names == "Spud"]
```

```
## [1] "Spud"
```

Data Frames

Now that we have reminded ourselves how to work with vectors, we are now going to move on to working with 2-dimensional data in the form of data frames. Each column in a data frame is a vector.

For the remainder of the assignment, we will be using the cactus data we collected during our first lab.

Once you run the code chunk below, the data frame will be saved in your environment as the object `cactus`.

7. Take a look at the `cactus` data frame. How many rows does this data frame have? How many columns? You can either use code to figure this out or look at the object in the environment. Explain how you figured it out (e.g., if you looked in the environment, how did you know which number meant what). (2 points)

Rows: 118

Columns: 11

Explanation:

```
# optional space for code to answer the question above
# str(pads)
# dim(pads)
```

8. Look at the first 6 rows of data. You can either do this by using a function or by using index sub-setting.

```
head(cactus)
```

```
##   group_id paddle_id length_in width_in depth_in spines insects damage location
## 1        1         1      7.25     2.50     0.25      N       N   None    Eigth
## 2        1         2      7.50     2.25     0.25      N       N   None  Seventh
## 3        1         3      9.25     4.00     0.50      N       Y   None    Sixth
## 4        1         4      9.50     3.50     0.25      N       N   None    Fifth
## 5        1         5      8.00     5.50     0.50      N       N   Some   Fourth
## 6        1         6      4.50     1.75     0.25      N       N   None    Fifth
##                 species   size
## 1 Opuntia ficus-indica Medium
## 2 Opuntia ficus-indica Medium
## 3 Opuntia ficus-indica Medium
## 4 Opuntia ficus-indica Medium
## 5 Opuntia ficus-indica Medium
## 6 Opuntia ficus-indica Medium
```

9. Use a function to calculate the mean *length* of all the cactus pads that were measured.

```
mean(cactus$length_in, na.rm = TRUE)
```

```
## [1] 9.123729
```

10. Use the `round()` function to around the mean of the cactus pad *lengths* (Q9) to only 1 place after the decimal.
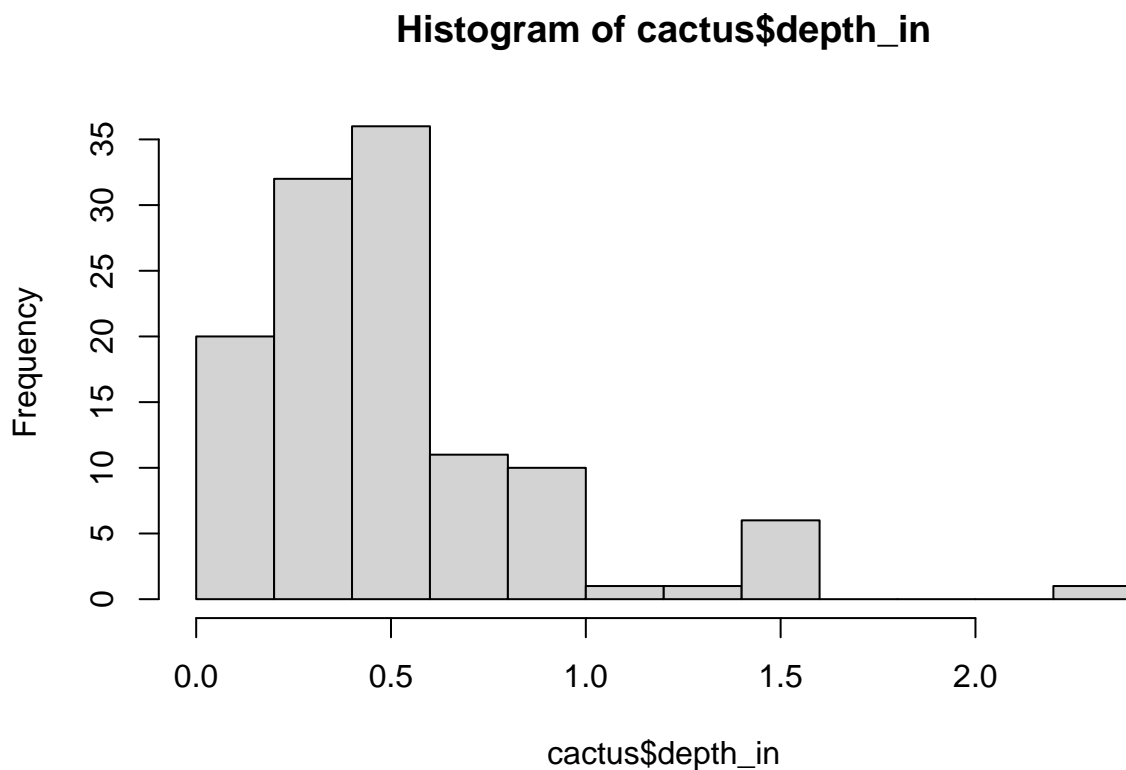
```
mean_length <- mean(cactus$length_in, na.rm = T)
round(mean_length, 1)
```

```
## [1] 9.1
```

```
# Any variation of the above or below works
# round(mean(cactus$length_in, na.rm = TRUE), digits = 1)
```

11. Create a histogram with the *depth* of all the cactus pads that were measured. No need to use `na.rm = TRUE` here; the histogram function automatically drops `NA` values.

```
hist(cactus$depth_in)
```

Histogram of cactus$depth_in

12. In 2-3 sentences, interpret the histogram. Describe what the axes represent, the approximate range of the data, where the majority of values fall, the distribution/skew of the data, etc. (2 points)

Answer: (Example) The x-axis indicates groupings/bins of depth measurements for each cactus pad and the y-axis indicates the frequency with which measurements are in each group. While the range of the data is from about 0-2.5 the majority of values seem to fall between 0.2-0.6. The distribution is right-skewed.

13. Run the following line of code. By looking at the code and the output, describe what the following line of code is doing. Make sure to include each part of the code. (As a point of reference, this took me 5 sentences in the answer key). (2 points)

```
cactus[cactus$spines == "N", 9]
```

```
##  [1] "Eigth"    "Seventh"  "Sixth"    "Fifth"    "Fourth"   "Fifth"
##  [7] "Fourth"   "Fourth"   "Fourth"   "Seventh"  "Fifth"    "Fifth"
## [13] "Second"   "Fifth"    "Sixth"    "Third"    "Fourth"   "Third"
## [19] "Second"   "Second"   "Ninth"    "Seventh"  "Ninth"    "Sixth"
## [25] "Fifth"    "Sixth"    "Sixth"    "Eigth"    "Seventh"  "Sixth"
## [31] "Fifth"    "Fourth"   "Fourth"   "Fifth"    "Seventh"  "Sixth"
## [37] "Eigth"    "Ninth"    "Seventh"  "Fourth"   "Sixth"    "Ninth"
## [43] "Eleventh" "Tenth"    "Ninth"    "Eigth"    "Seventh"  "Sixth"
## [49] "Second"   "Third"    "First"    "Eleventh" "Tenth"    "Tenth"
## [55] "Tenth"    "Tenth"    "Ninth"    "Ninth"    "Eigth"    "Seventh"
## [61] "Sixth"
```

Answer: This line of code pulls out depth values for the cactus pads that do not have spines. We first tell R that we want to select something from the `cactus` data frame. The [] indicate that we are subsetting. We set a condition in the rows section that indicates we only want rows in which the value in the spines column is "N", indicating no spines. After the comma, we specify that we want values from the 9th column, which is `depth_in`.

14. Create 2 different data frames: one called `ficus` with only rows that have *Opuntia ficus-indica* in the species column and one called `engel` with only rows which have *Opuntia engelmannii* in the species column. The data frames should have *all* columns.

Hint: neither dataframe should have 0 rows; if that has happened, you have a typo somewhere.

```
ficus <- cactus[cactus$species == "Opuntia ficus-indica", ]
engel <- cactus[cactus$species == "Opuntia engelmannii", ]
```

15. Calculate the mean pad width for each species.

```
mean_ficus <- mean(ficus$width_in, na.rm = TRUE)
mean_ficus
```

```
## [1] 4.411207
```

```
mean_engel <- mean(engel$width_in, na.rm = TRUE)
mean_engel
```

```
## [1] 8.32875
```

16. In a similar fashion, calculate the standard deviation (`sd()`) for the width of both species.

```
sd_ficus <- sd(ficus$width_in, na.rm = TRUE)
sd_ficus
```

```
## [1] 1.528388
```

```
sd_engel <- sd(engel$width_in, na.rm = TRUE)
sd_engel
```

[1] 2.12681

17. Interpret the results from Q15 and Q16. What do they tell us about the width of cactus pads and the distribution of the data?

    Answer: (Exmaple) Opuntia engelmannii seems to have wider pads on average than *O. ficus-indica.* The *O. engelmannii* data also has a higher standard deviation, meaning that there is more variation in the data points (they are less close to the mean than the *ficus* data points).

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 3. Submit the PDF file that you just downloaded.