Module 1, Assignment 2: Getting to Know the Team

Ellen Bledsoe

2023-09-05

Assignment Description

Purpose

The goal of this assignment is to get comfortable using R to look at 1- and 2-dimensional data sets.

Task

Write R code to successfully answer each question below.

Criteria for Success

- Code is within the provided code chunks
- Code is commented with brief descriptions of what the code does
- Code chunks run without errors
- Code produces the correct result

Due Date

Sept 15 at midnight MST

Assignment Questions

Remember to comment your code and run each chunk of code as you go!

Each question is worth 2 points.

Vectors

Let's start working with vectors, or 1-dimensional data, first.

Run this chunk of code to create a vector of data to use.

```
# vector with counts of penguins
counts <- c(2, 9, 4, 3, 6, 7, 1, 0, 3)
```

1. What data class is `counts`? Write a line of code that tells you.

```r
# data class for counts
class(counts)
```

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

2. Write a line of code that pulls out the 2nd value in the `counts` vector.

```r
# second value of counts vector
counts[2]
```

```
## [1] 9
```

3. Calculate the average number of penguins that were counted.

```r
# mean number of penguins
mean(counts)
```

```
## [1] 3.888889
```

Data Frames

Now that we've practiced with vectors, let's move on to 2-dimensional data.

Remember that quiz we took in class with fun/silly questions about our trip to Antarctica? It's time to play around with that data!

The following code chunk will read in the data. Be sure to run it before try to answer the questions! A bunch of stuff will pop up, but it should work just fine.

```r
team_data <- read.csv("../data/team_antarctica_responses.csv")
```

Running the code chunk above produces a message that gives us some useful information, even before we look at the data set (alternatively, you can check it out in the environment tab to the right).

- What are the dimensions of the data?
- What are the names of the columns in the data?
- What data *class* do you expect to find in each column (i.e., numbers, character strings, T/F, etc.)

4. Take a look at the data frame using the `head()` function. Typically, `head()` provides the first 6 rows of data. Modify one of the arguments in `head()` so that the line of code prints the first 10 rows. (If you aren't sure how to do that, remember how you can look for help about functions!)

```r
head(team_data, 10)
```

```
##   uniqueID fishing_skill               swim cold_tolerance good_with_animals
## 1        1             1                Yes              1              TRUE
## 2        2             1                Yes              3              TRUE
## 3        3             3                Yes              4              TRUE
## 4        4             1 I can dog paddle              4              TRUE
## 5        5             1                Yes              3              TRUE
```

```
## 6          6                 1                Yes               2               TRUE
## 7          7                 2                Yes               5               TRUE
## 8          8                 1                Yes               3               TRUE
## 9          9                 1                Yes               2               TRUE
## 10        10                 1                Yes               2               TRUE
##     remote_location parka_color      flag_mascot distance_mi
## 1                 3        Blue     Leopard Seal       8,791
## 2                 4        Blue     Leopard Seal       8,790
## 3                 1      Orange       Sea Spider        9184
## 4                 3       White             Orca        8625
## 5                 4        Blue             Orca    8,346.83
## 6                 2        Blue  Emperor Penguin       9,637
## 7                 5       White  Emperor Penguin       8,277
## 8                 3       Black  Emperor Penguin       8,347
## 9                 4        Blue  Emperor Penguin        8290
## 10                3        Blue             Orca    3,985.31
```

5. Using what you know about sub-setting data frames, write a line of code the pulls out the parka color for UniqueID 9 (row 9). (Hint: count the columns!)

```
team_data[9, 7]
```

```
## [1] "Blue"
```

When we have a large data set like this, it is often helpful to summarize the data in some way. The next few questions will help use get a better understanding of the content of the data set.

6. On average, how did people rate their ability to be in a remote location?

```
mean(team_data$remote_location)
```

```
## [1] 3.148148
```

7. What are the minimum and maximum distances that would be traveled by a team member to get to Antarctica? Use the `min()` and `max()` functions.

```
# minimum distance
min(team_data$distance_mi)
```

```
## [1] "1200"
```

```
# maximum distance
max(team_data$distance_mi)
```

```
## [1] "9514"
```

8. Create a data frame that only includes rows of data for people who rated their fishing skills as a 5. (Hint: numbers do not need quotation marks around them).

```
team_data[team_data$fishing_skill == 1, ]
```

```
##    uniqueID fishing_skill          swim cold_tolerance good_with_animals
## 1         1             1            Yes              1              TRUE
## 2         2             1            Yes              3              TRUE
## 4         4             1 I can dog paddle            4              TRUE
## 5         5             1            Yes              3              TRUE
## 6         6             1            Yes              2              TRUE
## 8         8             1            Yes              3              TRUE
## 9         9             1            Yes              2              TRUE
## 10       10             1            Yes              2              TRUE
## 12       12             1            Yes              5              TRUE
## 14       14             1            Yes              3              TRUE
## 15       15             1            Yes              4              TRUE
## 16       16             1            Yes              5              TRUE
## 17       17             1            Yes              3             FALSE
## 18       18             1            Yes              4              TRUE
## 20       20             1            Yes              2              TRUE
## 21       21             1            Yes              1              TRUE
## 23       23             1            Yes              3              TRUE
## 24       24             1            Yes              3              TRUE
##    remote_location parka_color      flag_mascot distance_mi
## 1                3        Blue     Leopard Seal       8,791
## 2                4        Blue     Leopard Seal       8,790
## 4                3       White             Orca        8625
## 5                4        Blue             Orca    8,346.83
## 6                2        Blue  Emperor Penguin       9,637
## 8                3       Black  Emperor Penguin       8,347
## 9                4        Blue  Emperor Penguin        8290
## 10               3        Blue             Orca    3,985.31
## 12               4       Black             Orca        9514
## 14               2        Blue       Sea Spider        8290
## 15               4        Blue  Emperor Penguin       14148
## 16               5       Black             Orca        7800
## 17               1       Black  Emperor Penguin        8290
## 18               3      Orange     Leopard Seal        8290
## 20               5       White       Sea Spider        7000
## 21               2       White       Sea Spider       8,625
## 23               2       Black       Sea Spider        1200
## 24               3        Blue             Orca       13448
```

9. Write a line of code that tells us what data class the `good_with_animals` column is.

```
class(team_data$good_with_animals)
```

```
## [1] "logical"
```

10. Calculate the average cold tolerance of people who want blue parkas. Use these steps to think through how to answer this question.

   a. First, think about how to create a data frame with only people who want blue parkas
   b. Next, think about how you select the column with the cold tolerance data

c. Finally, think about how to calculate the average

```
mean(team_data[team_data$parka_color == "Blue", 4])
```

```
## [1] 2.846154
```

Bonus (up to 2 points)!

What animal should be on our team flag?

First, create a vector called `mascot` that has only the values from the **flag_mascot** column. Next, use the **table()** function on the `mascot`. This will give you the number of times each option was chosen. According to the results, which animal should be on our team flag?

Answer:

```
# data frame
mascot <- team_data$flag_mascot
# mascot count
table(mascot)
```

```
## mascot
## Emperor Penguin    Leopard Seal            Orca      Sea Spider
##               6               9               6               6
```