Module 1, Assignment 1: Getting to Know the Team

Ellen Bledsoe

2024-01-18

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

Jan 25 at 11am MST

Assignment Questions

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`? Use a function to find out.

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

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

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

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

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

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

```
# 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 those fun/silly questions about our trip to Antarctica? It's time to play around with that data!

The following code chunk will read the data into RStudio. Be sure to run it before try to answer the questions! A bunch of stuff will pop up, but it should work just fine. Once you run the code, you will see a dataframe called `team_data` show up in the environment.

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

4. Using either a function or looking in the environment, answer the following questions (2 points each).

- What are the dimensions of the data (how many rows and how many columns)?
- What are the names of the columns in the data?
- What is the data *class* for each column? Also, determine which of these classes are qualitative vs. quantitative.

Answer:

```
# using code is optional
dim(team_data) # 34 rows, 10 columns
```

```
## [1] 34 10
```

```
colnames(team_data)
```

```
##  [1] "uniqueID"          "cooking_skill"     "fishing_skill"
##  [4] "swimming"          "cold_tolerance"    "working_with_animals"
##  [7] "remote_location"   "parka_color"       "team_flag"
## [10] "distance_mi"
```

```r
str(team_data)
```

```
## 'data.frame':    34 obs. of  10 variables:
##  $ uniqueID           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ cooking_skill      : int  2 4 5 4 4 1 3 4 3 2 ...
##  $ fishing_skill      : int  2 4 5 2 4 1 4 1 3 2 ...
##  $ swimming           : chr  "I can dog paddle" "Yes" "Yes" "Yes" ...
##  $ cold_tolerance     : int  2 5 5 3 3 1 3 4 1 4 ...
##  $ working_with_animals: logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ remote_location    : int  1 5 5 5 4 4 3 3 3 3 ...
##  $ parka_color        : chr  "Blue" "Blue" "Blue" "Red" ...
##  $ team_flag          : chr  "Emperor Penguin" "Orca" "Emperor Penguin" "Orca" ...
##  $ distance_mi        : num  8497 9056 8821 8932 3957 ...
```

```r
# uniqueID (integer/numeric)
# cooking_skill (integer/numeric)
# fishing_skill (integer/numeric)
# swimming (character/text)
# cold_tolerance (integer/numeric)
# working_with_animals (logical)
# remote_location (integer/numeric)
# parka_color (character/text)
# team_flag (character/text)
# distance (numeric)

# integer/numeric = quantitative
# character and logical = qualitative
```

5. 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 cooking_skill fishing_skill          swimming cold_tolerance
## 1         1             2             2 I can dog paddle              2
## 2         2             4             4              Yes              5
## 3         3             5             5              Yes              5
## 4         4             4             2              Yes              3
## 5         5             4             4              Yes              3
## 6         6             1             1              Yes              1
## 7         7             3             4              Yes              3
## 8         8             4             1              Yes              4
## 9         9             3             3              Yes              1
## 10       10             2             2              Yes              4
##    working_with_animals remote_location parka_color       team_flag distance_mi
## 1                  TRUE               1        Blue Emperor Penguin     8496.70
## 2                  TRUE               5        Blue            Orca     9056.17
## 3                  TRUE               5        Blue Emperor Penguin     8821.00
## 4                  TRUE               5         Red            Orca     8932.00
## 5                  TRUE               4       Black    Leopard Seal     3957.00
## 6                  TRUE               4        Blue            Orca     8024.00
```

```
## 7                    TRUE             3      Black Emperor Penguin      8276.00
## 8                    TRUE             3        Red            Orca      8400.00
## 9                    TRUE             3       Blue            Orca      8276.00
## 10                   TRUE             3      Black    Leopard Seal      8290.00
```

6. Using what you know about sub-setting data frames, write a line of code the pulls out the parka color for Unique ID 4. (Hint: count the columns!)

```
team_data[4, 8]
```

```
## [1] "Red"
```

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.

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

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

```
## [1] 3.676471
```

8. 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] 1732
```

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

```
## [1] 9269
```

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 `team_flag` 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$team_flag
# mascot count
table(mascot)
```

```
## mascot
## Emperor Penguin     Leopard Seal             Orca      Sea Spider
##              11                8               12               3
```