Module 1, Assignment 2: Getting to Know the Team

Ellen Bledsoe

2026-01-29

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

Thursday, Sept 11 at 9:30 am MST

Assignment Questions

Each question is worth 2 points unless otherwise specified.

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. It represents the number of penguins counted at different sites.

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

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

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

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

2. Use conditional sub-setting to return values from the vector tha are greater than 5.

```
counts[counts > 5]
```

```
## [1] 9 6 7
```

3. Calculate the average number of penguins that were counted. (Hint: the `mean()` function will calculate the average.)

```
# 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.

```
# using read.csv because we haven't started using the tidyverse yet
# file path on Posit Cloud is read.csv("team_antarctica.csv")
team_data <- read.csv("../data/team_antarctica.csv")
```

4. Using either a function or looking in the environment, answer the following questions (1 point 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?

Answer:

```
# using code is optional
dim(team_data) # 22 rows, 10 columns
```

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

```
colnames(team_data)
```

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

```
str(team_data)
```

```
## 'data.frame':    34 obs. of  10 variables:
##  $ team                : chr  "University of Arizona" "University of Arizona" "University of Arizona
##  $ cooking_skill       : int  5 1 4 3 1 3 5 1 5 3 ...
##  $ fishing_skill       : int  1 2 3 1 1 1 3 1 3 1 ...
##  $ swimming            : chr  "Yes" "No" "Yes" "Yes" ...
##  $ cold_tolerance      : int  3 4 4 4 1 4 3 3 4 4 ...
##  $ working_with_animals: logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ remote_location     : int  1 3 4 4 3 2 2 3 3 3 ...
##  $ parka_color         : chr  "Black" "Black" "Black" "Blue" ...
##  $ team_flag           : chr  "Leopard Seal" "Orca" "Leopard Seal" "Leopard Seal" ...
##  $ distance_mi         : chr  "8800" "8275" "8791" "8218.69" ...
```

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!)

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

```
##                    team cooking_skill fishing_skill swimming cold_tolerance
## 1  University of Arizona             5             1      Yes              3
## 2  University of Arizona             1             2       No              4
## 3  University of Arizona             4             3      Yes              4
## 4  University of Arizona             3             1      Yes              4
## 5  University of Arizona             1             1      Yes              1
## 6  University of Arizona             3             1      Yes              4
## 7  University of Arizona             5             3      Yes              3
## 8  University of Arizona             1             1      Yes              3
## 9  University of Arizona             5             3      Yes              4
## 10 University of Arizona             3             1      Yes              4
##    working_with_animals remote_location parka_color       team_flag distance_mi
## 1                  TRUE               1       Black    Leopard Seal        8800
## 2                  TRUE               3       Black            Orca        8275
## 3                  TRUE               4       Black    Leopard Seal        8791
## 4                  TRUE               4        Blue    Leopard Seal     8218.69
## 5                  TRUE               3       Black    Leopard Seal        8000
## 6                  TRUE               2       White Emperor Penguin       9,000
## 7                  TRUE               2        Blue            Orca       9,000
## 8                  TRUE               3      Orange Emperor Penguin        8223
## 9                  TRUE               3       Black            Orca        8243
## 10                 TRUE               3      Orange Emperor Penguin      10,000
```

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] "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.

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

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

```
## [1] 3.352941
```

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] "10,000"
```

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

```
## [1] "9143"
```

9. Use conditional sub-setting to produce a new 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 == 5, ]
```

```
##                    team cooking_skill fishing_skill swimming cold_tolerance
## 16 University of Arizona             3             5      Yes              5
## 18 University of Arizona             4             5      Yes              5
```

```
##    working_with_animals remote_location parka_color    team_flag distance_mi
## 16                 TRUE               4         Red Leopard Seal       10000
## 18                 TRUE               4       White Leopard Seal       10000
```

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", 5])
```

```
## [1] 3.25
```

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: Leopard seal

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

```
## mascot
## Emperor Penguin    Leopard Seal            Orca      Sea Spider
##              12              15               5               2
```