
Homework for Lecture 7: Data Frames

The following homework assignment should be completed as an RMarkdown document. Only show your working code in the code chunks, and place comments/explanations in the “white space” to help me evaluate your work. Please submit both the knitted RMarkdown document (.doc) and the .Rmd file. You may discuss the assignment with your classmates, but your write-up and work should be done independently of your classmates. The use of AI is considered a violation of academic policy for this assignment. When in doubt, ask me what is and is not allowed.

Answer the following questions. When you run code, make sure you explain what you are doing and what the output is telling us. This should be done in the written “white” section of the RMarkdown document. The “grey” section is where your code should go, and when you knit the document, the output will appear below the code.

1. Create a header *# Exercise 1*. Then answer the following questions. Make sure you also explain the code you are using, what the output is showing us, and whether the result makes sense.

(a) Create a data frame that replicates the following table:

Name	Age	Working
Kermit	18	FALSE
Gonzo	39	TRUE
Fozzie	22	TRUE
Statler	83	FALSE
Beaker	24	TRUE

- (b) Using a logical vector, display all of the observations where the person is over the age of 30.
- (c) Using a logical vector, display all of the observations where the person is either not working or under the age of 30.
2. Create a header *# Exercise 2*. Then answer the following questions. Make sure you also explain the code you are using, what the output is showing us, and whether the result makes sense.
- (a) Carry out a random sample of size 30 from the following list of names: Huey, Dewey, Louie, Donald, and Scrooge. Save this vector as **duck_character**.
- (b) Carry out a random sample of size 30 from the sequence of numbers 1 through 100 without replacement. Save this vector as **appearances**.
- (c) Carry out a random sample of size 30 from TRUE and FALSE. Save this vector as **director**.

**** MORE PROBLEMS ON PAGE 2 ****

-
- (d) Create a data frame called `duck.df` that combines the three variables above into a single data frame.
 - (e) Display just the first 4 observations.
 - (f) Display all of the observations dealing with Dewey.
 - (g) Display all of the observations where `appearances` is greater than 40.
 - (h) Display all of the observations where `appearances` is greater than 40 or the character is a director.
 - (i) Display all of the observations where `appearances` is greater than 40 and the character is a director.
 - (j) Display all of the observations where the character is named Huey or Scrooge, and the character has less than 60 appearances.
 - (k) Display all of the observations where the character is named Huey, or the character is Scrooge and has less than 60 appearances.
 - (l) In the written space below your code, explain the difference between the previous two questions. Why do they not necessarily give the same result?