Problem Set 3

MA615

2022-09-29

Instructions

1. The due date and time for this assignment is Friday, September 30, 1430 Eastern time.
2. Your file name most be **problemSet_3.R**.
3. Make sure to include a vector named **myName** at the top of problemSet_3.R which contains your name as a string.
4. For each problem, there is an **Submit** instruction.
5. Some of the questions include an example of the output that the problem describes. These examples do not necessarily describe the answer you are being asked to submit. They are there to guide your work. The examples have been formatted for visual clarity. Formatting is not required for your answers.

Problems

Use the Iris built-in data.

1. Put all rows of Species 'versicolor' in a new data frame. Call this data frame: iris.vers.
   Submit: ans_1 where ans_1 = iris.ver.

   Expected result (first 5 rows)

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
54	5.5	2.3	4.0	1.3	versicolor
55	6.5	2.8	4.6	1.5	versicolor

2. Make a vector called 'sepal.dif' with the difference between 'Sepal.Length' and 'Sepal.Width' of 'versicolor' plants.
   Submit: ans_2 where ans_2 = sepal.dif.

3. Update (add) 'iris.vers' with the new column 'sepal.dif' added as the right-most column.
   Submit: ans_3 where ans_3 = iris.vers.

Use the mtcars built-in data.

4. Using the apply family of functionals, check the class of each variable in 'mtcars'. Store the result from the apply function in a vector x.
   Submit: ans_4 where ans_4 = x.

5. Change 'am', 'cyl' and 'vs' to integer and store the new dataset as 'newmtc'. As you did in problem 4, check the class of each variable in 'mtcars' and store the result from the apply function in a vector x .
   Submit: ans_5 where ans_5 = x.

	x
mpg	numeric
cyl	integer
disp	numeric
hp	numeric
drat	numeric
wt	numeric
qsec	numeric
vs	integer
am	integer
gear	numeric
carb	numeric

6. Round the 'newmtc' data frame to one digit.
   Submit: ans_6 where ans_6 = newmtc (updated).

Expected result (first 5 rows)

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
21.0	6	160	110	3.9	2.6	16.5	0	1	4	4
21.0	6	160	110	3.9	2.9	17.0	0	1	4	4
22.8	4	108	93	3.9	2.3	18.6	1	1	4	1
21.4	6	258	110	3.1	3.2	19.4	1	0	3	1
18.7	8	360	175	3.1	3.4	17.0	0	0	3	2

Use the Iris built-in data.

7. Use dplyr to filter the Iris data frame for all data of Species 'virginica' with a 'Sepal.Width' of greater than 3.5. Store the result of this operation in a data frame named iris_7.
   Submit: ans_7 where ans_7 = iris_7.

8. How would you use R Base to get a data frame of all data of Species 'virginica' with a 'Sepal.Width' of greater than 3.5, but without the last column Species in the data frame? Call this data frame iris_8.
   Submit: ans_8 where ans_8 = iris_8.

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
110	7.2	3.6	6.1	2.5
118	7.7	3.8	6.7	2.2
132	7.9	3.8	6.4	2.0

9. Get the row IDs of the rows matching the two filtering criteria provided above. store the row IDs in a variable r_id.
   Submit: ans_9 where ans_9 = r_id.

Use the Diamonds built-in data.

10. How many observations of diamonds have a cut of 'ideal' and have less than 0.21 carat? Store your answer in a variable diam_10.
    Submit: ans_10 where ans_10 = diam_10.
11. How many observations of diamonds have a combined 'x' + 'y' + 'z' dimension greater than 40? Store your answer in a variable diam_11. **Submit:** ans_11 where ans_11 = diam_11.
12. How many observations of diamonds have either a price above 10.000 USD or a depth of at least 70? Store your answer in a variable diam_12. **Submit:** ans_12 where ans_12 = diam_12.
13. Make a data frame with observations '67' and '982' of variables color and y. Store your answer in a variable diam_13. **Submit:** ans_13 where ans_13 = diam_13.
14. Make a data frame with the full info on observations '453', '792' and '10489'. Store your answer in a variable diam_14. **Submit:** ans_14 where ans_14 = diam_14.
15. Get the first 10 rows of the dataset 'diamonds' with the variables 'x', 'y', 'z'. Store your answer in a tibble diam_15. **Submit:** ans_15 where ans_15 = diam_15.
16. Create the object 'newdiam' which is a subset of the first 1000 rows of 'diamonds'. Be sure that newdiam is a tibble. **Submit:** ans_16 where ans_16 = newdiam.
17. Order 'newdiam' according to price, starting with the lowest. Store your answer in a tibble newdiam_17. **Submit:** ans_17 where ans_17 = newdiam_17.
18. Use 'dplyr', 'sample_n' to get the object 'diam750' which contains 750 randomly sampled observations of 'diamonds'. Use set.seed(56). Be sure that diam750 is a tibble. **Submit:** ans_18 where ans_18 = diam750.
19. Get a summary of diam750. Assign the result to variable sum_diam750.
    Submit: ans_19 where ans_19 = sum_diam750.