Richard Gresham
CPSC 375-01
September 09, 2022

Homework #1

Prepare your answers as a **single PDF file**.
Group work: You may work in groups of 1-3. Include all group member names in the PDF file. You may work with students in both sections (375-01). Only one person in the group should submit to Canvas.
Due: check on Canvas.

1. Use the in-built dataset, `iris`, for this problem. Write code to:

a. Get number of rows (Hint: nrow)
   Input: nrow(iris)
   Output: 150
b. Get number of columns (Hint: ncol)
   Input: ncol(iris)
   Output: 5
c. Show first 10 rows
   Input: iris[1:10,]
   Output:

	Sepal.Length	Sepal.Width
1	5.1	3.5
2	4.9	3.0
3	4.7	3.2
4	4.6	3.1
5	5.0	3.6
6	5.4	3.9
7	4.6	3.4
8	5.0	3.4
9	4.4	2.9
10	4.9	3.1

	Petal.Length	Petal.Width
1	1.4	0.2
2	1.4	0.2
3	1.3	0.2
4	1.5	0.2
5	1.4	0.2
6	1.7	0.4
7	1.4	0.3
8	1.5	0.2
9	1.4	0.2
10	1.5	0.1

Species
1 setosa
2 setosa
3 setosa
4 setosa
5 setosa
6 setosa
7 setosa
8 setosa
9 setosa
10 setosa

d. Show column Sepal.Length
Input: iris[,"Sepal.Length"]
Output:
[1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8
[14] 4.3 5.8 5.7 5.4 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0
[27] 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5 4.9 5.0 5.5 4.9 4.4
[40] 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0 6.4
[53] 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6
[66] 6.7 5.6 5.8 6.2 5.6 5.9 6.1 6.3 6.1 6.4 6.6 6.8 6.7
[79] 6.0 5.7 5.5 5.5 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5 5.5
[92] 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8 7.1 6.3
[105] 6.5 7.6 4.9 7.3 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5
[118] 7.7 7.7 6.0 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1 6.4 7.2
[131] 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.9 5.8
[144] 6.8 6.7 6.7 6.3 6.5 6.2 5.9

e. Calculate the mean Sepal.Length?
Input: mean(iris$Sepal.Length)
Output: [1] 5.843333

f. Show all rows where Sepal.Length > 7.6
Input: iris[iris$Sepal.Length > 7.6,]
Output:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
118	7.7	3.8	6.7	2.2	virginica
119	7.7	2.6	6.9	2.3	virginica
123	7.7	2.8	6.7	2.0	virginica
132	7.9	3.8	6.4	2.0	virginica
136	7.7	3.0	6.1	2.3	virginica

i. What are the row indexes where Sepal.Length > 7.6? (Hint: which)
Input: which(iris$Sepal.Length > 7.6)
Output: [1] 118 119 123 132 136

g. Show all rows where Species is “setosa”
Input: iris[iris$Species == "setosa",]
Output:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3.0	1.4	0.1
14	4.3	3.0	1.1	0.1
15	5.8	4.0	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3
19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3
21	5.4	3.4	1.7	0.2
22	5.1	3.7	1.5	0.4
23	4.6	3.6	1.0	0.2
24	5.1	3.3	1.7	0.5
25	4.8	3.4	1.9	0.2
26	5.0	3.0	1.6	0.2
27	5.0	3.4	1.6	0.4
28	5.2	3.5	1.5	0.2
29	5.2	3.4	1.4	0.2
30	4.7	3.2	1.6	0.2
31	4.8	3.1	1.6	0.2
32	5.4	3.4	1.5	0.4
33	5.2	4.1	1.5	0.1
34	5.5	4.2	1.4	0.2
35	4.9	3.1	1.5	0.2

36	5.0	3.2	1.2	0.2
37	5.5	3.5	1.3	0.2
38	4.9	3.6	1.4	0.1
39	4.4	3.0	1.3	0.2
40	5.1	3.4	1.5	0.2
41	5.0	3.5	1.3	0.3
42	4.5	2.3	1.3	0.3
43	4.4	3.2	1.3	0.2
44	5.0	3.5	1.6	0.6
45	5.1	3.8	1.9	0.4
46	4.8	3.0	1.4	0.3
47	5.1	3.8	1.6	0.2
48	4.6	3.2	1.4	0.2
49	5.3	3.7	1.5	0.2
50	5.0	3.3	1.4	0.2

	Species
1	setosa
2	setosa
3	setosa
4	setosa
5	setosa
6	setosa
7	setosa
8	setosa
9	setosa
10	setosa
11	setosa
12	setosa
13	setosa
14	setosa
15	setosa
16	setosa
17	setosa
18	setosa
19	setosa
20	setosa
21	setosa
22	setosa
23	setosa
24	setosa
25	setosa
26	setosa
27	setosa
28	setosa

29 setosa
30 setosa
31 setosa
32 setosa
33 setosa
34 setosa
35 setosa
36 setosa
37 setosa
38 setosa
39 setosa
40 setosa
41 setosa
42 setosa
43 setosa
44 setosa
45 setosa
46 setosa
47 setosa
48 setosa
49 setosa
50 setosa

h. Show all rows where Sepal.Length > 3.0 and Species is “setosa”
Input: iris[iris$Species == "setosa" & iris$Sepal.Length > 3.0,]
Output:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3.0	1.4	0.1
14	4.3	3.0	1.1	0.1
15	5.8	4.0	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3

19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3
21	5.4	3.4	1.7	0.2
22	5.1	3.7	1.5	0.4
23	4.6	3.6	1.0	0.2
24	5.1	3.3	1.7	0.5
25	4.8	3.4	1.9	0.2
26	5.0	3.0	1.6	0.2
27	5.0	3.4	1.6	0.4
28	5.2	3.5	1.5	0.2
29	5.2	3.4	1.4	0.2
30	4.7	3.2	1.6	0.2
31	4.8	3.1	1.6	0.2
32	5.4	3.4	1.5	0.4
33	5.2	4.1	1.5	0.1
34	5.5	4.2	1.4	0.2
35	4.9	3.1	1.5	0.2
36	5.0	3.2	1.2	0.2
37	5.5	3.5	1.3	0.2
38	4.9	3.6	1.4	0.1
39	4.4	3.0	1.3	0.2
40	5.1	3.4	1.5	0.2
41	5.0	3.5	1.3	0.3
42	4.5	2.3	1.3	0.3
43	4.4	3.2	1.3	0.2
44	5.0	3.5	1.6	0.6
45	5.1	3.8	1.9	0.4
46	4.8	3.0	1.4	0.3
47	5.1	3.8	1.6	0.2
48	4.6	3.2	1.4	0.2
49	5.3	3.7	1.5	0.2
50	5.0	3.3	1.4	0.2

	Species
1	setosa
2	setosa
3	setosa
4	setosa
5	setosa
6	setosa
7	setosa
8	setosa
9	setosa
10	setosa
11	setosa

12	setosa
13	setosa
14	setosa
15	setosa
16	setosa
17	setosa
18	setosa
19	setosa
20	setosa
21	setosa
22	setosa
23	setosa
24	setosa
25	setosa
26	setosa
27	setosa
28	setosa
29	setosa
30	setosa
31	setosa
32	setosa
33	setosa
34	setosa
35	setosa
36	setosa
37	setosa
38	setosa
39	setosa
40	setosa
41	setosa
42	setosa
43	setosa
44	setosa
45	setosa
46	setosa
47	setosa
48	setosa
49	setosa
50	setosa

i. Get the largest value of Sepal.Length
Input: iris[which(iris$Sepal.Length == max(iris$Sepal.Length)),]
Output:
Sepal.Length Sepal.Width Petal.Length Petal.Width
132 7.9 3.8 6.4 2
Species
132 virginica

i. Get the row index that contains this value
Input: which(iris$Sepal.Length == max(iris$Sepal.Length))
Output: [1] 132

j. What Species corresponds to this largest Sepal.Length?
Input: iris[which(iris$Sepal.Length == max(iris$Sepal.Length)),"Species"]
Output:
[1] virginica
Levels: setosa versicolor virginica

For each question, give (1) the code and (2) the output.

2. Consider the answer posted to Quora.com to "Why is R great for Data Science?. Answer one of the following questions.

The author lists 5 parts of the R ecosystem, the 5th being "community". Write 4-5 sentences about any one online community where members discuss R. (Include the URL, how active is the community, what types of people post here, how "friendly" it is to newcomers, etc.)

https://community.rstudio.com/#

This is a great community area where people discuss R and most especially R using R-studio. The community is very active with posts and replies occurring every day from students to software developers and data analysts who use R on the daily basis. This community is also supported by RStudio which is one of the largest R programming software. This site has tutorials on installing certain software, help areas where people answer others questions and overall create a friendly environment to learn R.

OR (if you know Python)

The author says "Note that in python, data frame manipulation will require numpy and pandas external packages (and the syntax is more cumbersome)". Do you agree with this statement? Justify your answer in 4-5 sentences.