

Name:

NetId/Email:

KEY

1. (1 pt.) State an important difference between R and SAS.

Open source vs. proprietary
in-memory vs. on disk
object oriented vs. procedural

many packages vs. fewer products

make your own functions vs. macro language
etc.

2. (1 pt.) Give an example of, or describe, an ordinal variable?

an ordered, categorical variable
hardness of minerals tax brackets
e.g. good, better, best grades
street numbers etc.

3.) (2 pts.) Given Data Set 1 below, write out the results of selecting the rows where $X1 \geq 3$.

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Data Set 1

ID	X1	X2
1235	3	C
1235B	4	D

variable header is not mandatory

row order does not matter

variable order does not matter

4.) (3 pts.) Given Data Set 2 below, write out the results of grouping Data Set 2 by ID and averaging X3.

ID	X3
123A	5
1234	6
1235	7
1235B	8
1234	9
1235	10
1235B	11

Data Set 2

variable header is not mandatory

ID	X3
1234	7.5
1235	8.5
1235B	9.5
123A	5

variable name does not matter

row order does not matter

variable order does not matter

2. (3 pt.) Given Data Sets 1 and 2 above, write out the results of **left** joining Data Set 1 onto Data Set 2 by the key variable shared between the two data sets. (Let Data Set 2 be the left table.)

ID	X3	X1	X2
123A	5	1	A
1234	6	2	B
1235	7	3	C
1235B	8	4	D
1234	9	2	B
1235	10	3	C
1235B	11	4	D

variable header not mandatory
variable order does not matter
row order does not matter
ID variable can be repeated