DNSC - Quiz 2

Others...

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

Python is open source / SAS proprietary

SAS is not free for commercial use / python is

SAS has support available / python does not

SAS holds data on-disk / python is in-mem

Python is general purpose / SAS is domain specific

2. **(1 pt.)** Give an example of, or describe, a nominal variable?

A nominal variable is an unordered categorical variable

examples include: car make/models, colors, events, states, etc...

3.) **(1 pts.)** Given Data Set 1 below, write out the results of selecting the rows where X1 ≥ 3.

ID	X1	X2
123A	0	W
1234	5	X
1235	8	Y
1235B	3	Z

Data Set 1

ID	X1	X2
1234	5	X
1235	8	Y
1235B	3	Z

→ row order does not matter

4.) Given Data Set 2 below, write out the results of transposing Data Set 2. You may disregard column and/or row names for the transposed set. (2 pts.)

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

Data Set 2

123C	1234	1235	1235	1234	1235	1235B
5	6	7	8	9	10	11

2 pts.

4.) ~~**(3 pts.)**~~ Given Data Set 2 above, write out the results of grouping Data Set 2 by ID and summing X3.

ID	SUM_X3
123C	5
1234	15
1235	25
1235B	11

row order does not matter

aggregated X3 col name does not matter

5. **(3 pts.)** 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.) Fill any resulting numeric missing values with the sentinel value -9999 and any missing character values with the sentinel value ZZZZ.

ID	X3	X1	X2
123C	5	-9999	ZZZZ
1234	6	5	W
1235	7	8	Y
1235	8	8	Y
1234	9	5	W
1235	10	8	Y
1235B	11	3	Z

row order does not matter

column order does not matter

ID can be repeated