Data Management

Data management is the "creation, storage, analysis, dissemination, and preservation of your research data" (https://www.lib.umn.edu/datamanagement).

Why is this important? Why should it be part of your role?

What are pros and cons of each of the following ways of storing a data set?

Method 1

Johnson class is Bold, Olson class is Italic

Alice	Sex	F		
	Age	7		
	IncomGroup	1		
	Scores	67	79	71
Sophia	Sex	F		
	Age	8		
	IncomeGroup	2		
	Scores	89	90	93

Method 2

Class	Student	Time	Score	Sex	Age	IncomeGroup
Johnson	Alice	1	67	F	7	1
Johnson	Alice	2	79	F	7	1
Johnson	Alice	3	71	F	7	1
Johnson	Bob	1	58	M	7	2
Johnson	Bob	2	59	M	7	2
Johnson	Bob	3	55	M	7	2
Johnson	Carol	1	80	F	8	2
Johnson	Carol	2	82	F	8	2
Johnson	Carol	3	86	F	8	2
Johnson	Daniel	1	55	M	7	3
Johnson	Daniel	2	60	M	7	3
Johnson	Daniel	3	42	M	7	3

Method 3

Sheet1

Class	Student	Score1	Score2	Score3	Sex	Age	IncomeGroup
Johnson	Alice	67	79	71	F	7	1
Johnson	Bob	58	59	55	M	7	2
Johnson	Carol	80	82	86	F	8	2
Johnson	Daniel	55	60	42	M	7	3
Olson	Sophia	89	90	93	F	8	2
Olson	Emma	70	76	81	F	9	1
Olson	Jacob	50	60	65	M	9	1
Olson	Mason	80	83	70	M	8	3

What kinds of metadata do you wish you had on this data set? How might you store it?

_____

_____

_____

What do you do when you find an error in the data set? How about an outlier that you're not sure what to do with?

_____

_____

_____

Discuss the balance between sharing your data and keeping it confidential. When are cases when one or the other might be necessary or preferred?

_____

_____

_____

Name: _____

What was new or interesting to you today?

_____

_____

_____

_____