Implementation Science and Evaluation #16:

DATA MANAGEMENT (II)

PRINCIPLES of good data management

MACHINE READABILITY

Structure

- Relevant fields can be extracted and used without human input

File Type

- Generic Formats (e.g., CSV which can be presented in excel)

Column Headers:

- Only use **alphanumeric** and these 2 special characters `. _`
- Use "**and**" instead of "**&**"
- **Each header must be unique**

TIDY DATA

Each **column** represents a variable (e.g., Name)

In a Dataset:

Each **row** is an **observation** (e.g., A Student)

Textual Variables

Should not have line breaks within cells

Name	Teacher	CCA _ and _ Leadership_Points	Score _ 1	Score _ 2	Attendance
Alexis	Ms Teo	780	88	72	84
Beatrice	Mr Lim	1000	96	84	90
Charles	Mr Tan Mr Lim	1,000	87 marks	62	0.9
David	Mr Tan	1200	55	65	92

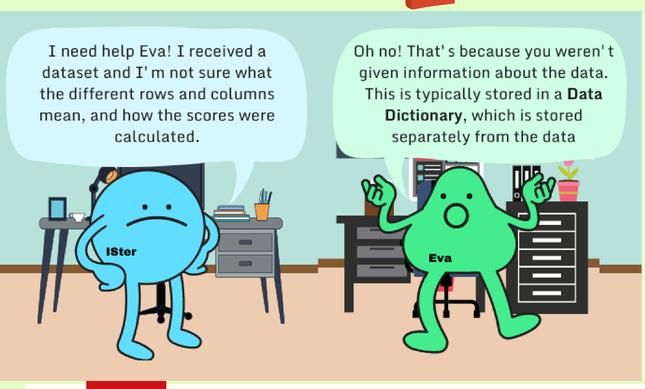
Numeric Variables

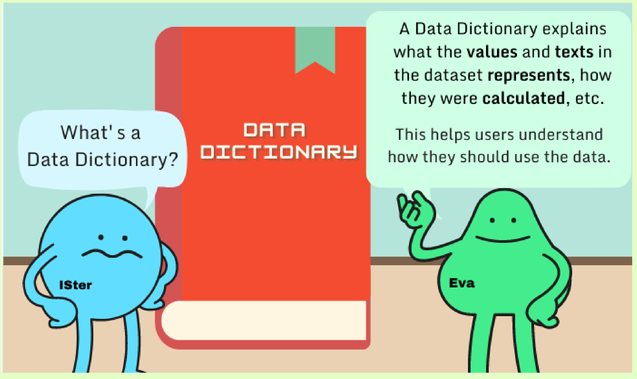
- Expressed as **full numbers**
- **Should not** have **commas**
- **Should not** have **units**
- Express **percentages** either in **decimal** (0.9) or **in full** (90%), but not both

Be **consistent** with the **format** throughout the column!

Refer to Infoposter #14 on Data Management (I) to find out more!

HAVE A DATA DICTIONARY





Tips for a Good Data Dictionary

- Define **what each column means** as clearly as possible
- For calculated or derived values, provide the **formula**

DATA DICTIONARY EXAMPLE

Data Field	Explanation	Value / Code	Special Markers / Null
Name	The Full Name of the student		
Gender	The Gender of the student	M = Male F = Female	"9999" = Missing Data
Score	Total Score = Section A score + Section B score + Section C score	0 (lowest) 100 (highest)	A null value means that student is absent from ALL tests

- Define the **range of possible values** and what each value means (if applicable)
- Indicate what **special markers** (e.g., 9999) or **null value** (i.e. no value input) means

PITFALLS of data entry

Results in **unwanted categories** during analysis

1 WHITESPACES

Eg: Male & Male

Ensure that your labels do not contain any unnecessary spaces

2 INCONSISTENT CAPITALISATION

Eg: Male & male

Ensure that your labels have consistent capitalisation

3 INCONSISTENT CATEGORICAL LABELS

Eg: Moderator & Mod

Ensure that your categorical labels are consistent

Want to know more?

Here is a toolkit that provides a step-by-step guide on how to make a Data Dictionary:
https://www.secoda.co/blog/how-to-create-a-data-dictionary-a-step-by-step-guide