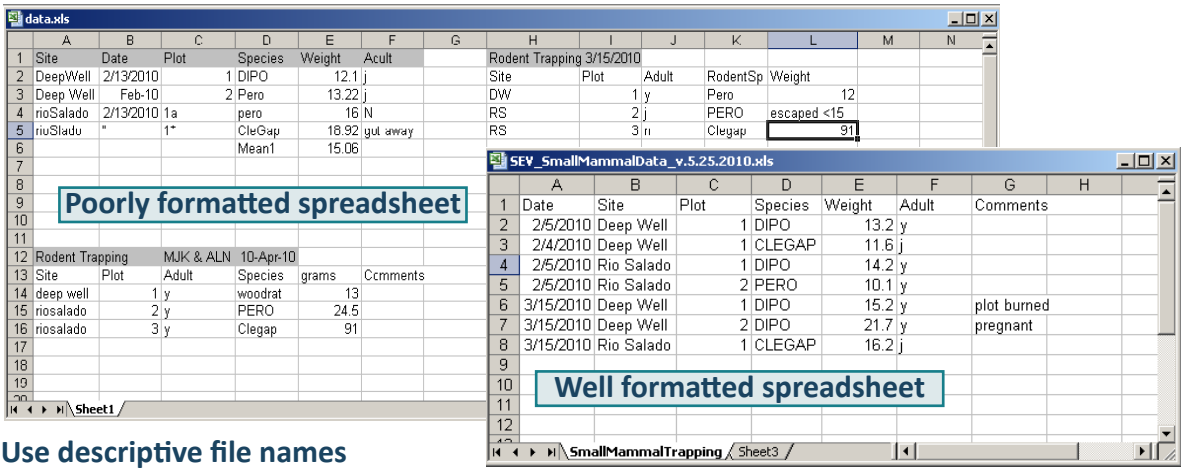
Lesson 4: Data Entry & Manipulation

Best Practices for spreadsheets & data files

- Create only one table per spreadsheet and enter data consistently. Do not leave empty rows or columns.
- Enter data using consistent formats. For example, always use the same formats for dates.
- Name columns consistently and without spaces or special characters.
- Order similar columns in multiple tables in a similar way.
- Use consistent formats, codes, spellings, etc. and do not mix data types in a single column.
- Store data in a format that will be accessibly by any (or many) applications now or in the future.
- Keep your raw data raw. Make a copy in another file for doing calculations and manipulations.



GOAL of DATA ENTRY:

To create **valid data sets** that are **organized** to support **ease of use.**

Use descriptive file names

Example:

SEV_SmallMammaData_v.5.25.2010.csv

Project = SEV
Type of data = SmallMammalData
Date version was created = v5.25.2010
File type = csv

Data entry tools & data validations

Spreadsheets and Google Docs are two options for data entry. Many data entry tools offer data validation that provides users with lists of choices or restricts entries to a specific data type or size. These features can improve data quality by preventing data entry errors.

Definition: Relational Database

A relational database consists of a set of tables, defined relationships between those tables (which are related to which and how), and a powerful command language called Structured Query Language (SQL) that facilitates data manipulation. Forms can be created to make date entry into a database much easier.

Spreadsheets	VS.	Relational Databases
Great for charts, graphs, calculations		Easy to query for portions of data
Flexible about cell content (#s & text in one column)		Data fields are types (only #s OR text in one column
Easy to use		Steeper learning curve than spreadsheet
No record integrity (columns sort independently)		Record integrity (columns can not sort independently
Harder to maintain as complexity & size grow		Better with large amounts of data

Software options for data manipulation

>> Analyzing, subsetting, or trans forming data <<

- SAS: outstanding support
- SPSS: user-friendly interface
- Matlab: toolboxes for different discipline
- R: free, quality graphics, lots of helpful forums

Local contact information