

# Introduction to Spreadsheets

## Working with Data

1. Each row is one observation.
2. Each cell should only have one value.
3. Use separate columns for year, month, and day, or be vigilant on how you deal with your date column. Data Carpentries can provide more information on storing [dates as data](https://datacarpentry.org/spreadsheets-socialsci/03-dates-as-data/index.html).
4. Avoid problematic null values. The best option is to leave a cell with no or missing data blank. NA or NULL are you next best options. See the table further down for more details.
5. Define the accepted values for each column. If possible, use data validation to help ensure consistency.
6. Keep a separate copy of your original raw data.
7. Keep raw data in a CSV (comma separated values) file. You can save each sheet in a workbook as its own CSV.
8. Use a system to name your files and spreadsheet columns. If you don't have one, consider this [naming convention system](https://www2.staffingindustry.com/Editorial/Archived-Blog-Posts/Adam-Pode-s-Blog/Probably-the-best-file-naming-convention-ever).
9. Don't rely on formatting to convey any information, and avoid destructive formatting, like merging cells.
10. Freeze the top row of your spreadsheet.
11. Use keyboard shortcuts to make common tasks simpler.

## Overview of Sample Datasets

### BookSample

A very basic spreadsheet. Try opening BookSample.csv with a text editor to see how spreadsheets are represented in CSVs.

### SkyrimMuseumSample

**Original:** A simple spreadsheet that could be imported into a site like Omeka.

**GenericTagColumns:** An attempt to avoid multiple tag values in one cell. This is not a solution for organized data. Avoid creating generic columns for items with variable numbers, like tags.

**SpecificTagColumns:** An attempt to avoid multiple tag values in one cell. This is not ideal for analysis, and it becomes unwieldy with more tag options. Avoid.

**Objects:** Part of the solution - a table with each of the object data, minus the tags.

**Tags:** The other part of the solution - a table associating each object with its tags. Use this and the Objects sheet as an example for similar datasets.

### SurveySample

**Original:** A very, *very*, simplified version of some real data. Each household surveyed is given space for information about up to twelve members. This is an extremely problematic way to deal with this type of data.

**Households and Members:** The equivalent sheets to Objects and Tags from the SkyrimMuseumSample. Household data receives its own spreadsheet, and each member is associated with the household they are from.

**BetterHouseholds:** Households, but with the Livestock column expanded into four columns to prevent multiple values in a single cell.

## Useful Spreadsheet Features

### Data Validation

Microsoft support has information on using [data validation in Excel](https://support.microsoft.com/en-us/office/apply-data-validation-to-cells-29fecbcc-d1b9-42c1-9d76-eff3ce5f7249#ID0EAADAAA=Windows) for Windows, Mac, or Web.

Data validation in Google Sheets will be reasonably similar. The linked instructions are for [creating an in-cell dropdown with data validation](https://www.groovypost.com/howto/add-google-docs-in-cell-dropdown-validation-spreadsheets-review/). For other types of data validation, follow steps 1 and 2, but choose a different option instead of "List from a range."

### Freezing Rows (or Columns)

Freeze your header row to easily refer to it when scrolling through your data.

* [Freeze rows in Excel](https://support.microsoft.com/en-us/office/freeze-panes-to-lock-rows-and-columns-dab2ffc9-020d-4026-8121-67dd25f2508f)
* [Freeze rows in Google Sheets](https://spreadsheetpoint.com/how-to-freeze-rows-in-google-sheets/)

### Keyboard Shortcuts

Use keyboard shortcuts to simplify common tasks.

* [Keyboard shortcuts in Excel](https://support.microsoft.com/en-us/office/keyboard-shortcuts-in-excel-1798d9d5-842a-42b8-9c99-9b7213f0040f)
* [Keyboard shortcuts in Google Sheets](https://support.google.com/docs/answer/181110?co=GENIE.Platform%3DDesktop)

## Additional Resources and Figures

This workshop relied heavily on content from the Data Carpentries lesson [Data Organization in Spreadsheets for Social Scientists](https://datacarpentry.org/spreadsheets-socialsci/). Data Carpentries also has good lessons for learning to use OpenRefine, a very useful data cleaning tool. You can try [Data Cleaning with OpenRefine for Social Scientists](https://datacarpentry.org/openrefine-socialsci/) or [Data Cleaning with OpenRefine for Ecologists](https://datacarpentry.org/OpenRefine-ecology-lesson/). OU Libraries periodically offers Data Carpentries and Software Carpentries workshops.

### Null Values Table

This table was simplified from the one included in the Data Carpentries Data Organization in Spreadsheets for Social Scientists lesson on the [Formatting Problems](https://datacarpentry.org/spreadsheets-socialsci/02-common-mistakes/index.html) page.

| Null Value | Problems | Recommendation |
| --- | --- | --- |
| 0 | Indistinguishable from a true zero | Never use |
| Blank | Hard to distinguish values that are missing from those overlooked on entry. Hard to distinguish blanks from spaces, which behave differently. | Best option |
| -999 or 999 | Not recognized as null by many programs without user input. Can be inadvertently entered into calculations. | Avoid |
| NA or na | Can also be an abbreviation (e.g. North America), can cause problems with data type (turn a numerical column into a text column). NA is more commonly recognized than na. | Good option |
| N/A | An alternate form of NA, but often not compatible with software | Avoid |
| NULL | Can cause problems with data type | Good option |
| None | Uncommon. Can cause problems with data type | Avoid |
| No data | Uncommon. Can cause problems with data type, contains a space | Avoid |
| Missing | Uncommon. Can cause problems with data type | Avoid |
| - or + or . | Uncommon. Can cause problems with data type | Avoid |