

Data Carpentry

Cardiff University, 29-30 October 2019

Spreadsheets: Introduction

- **Good data practices**
- Spreadsheets can be used for:
 - Data entry
 - Organizing data
 - Subsetting and sorting data
 - Statistics
 - Plotting

Spreadsheets: Formatting data

- Put all your variables in columns.
- Put each observation in its own row.
- Don't combine multiple pieces of information in one cell.
- Leave the raw data raw - don't change it!
- Export the cleaned data to a text-based format like CSV (comma-separated values) format.

Spreadsheets: Common mistakes

- Avoid using multiple tables within one spreadsheet.
- Avoid spreading data across multiple tabs.
- Record zeros as zeros.
- Use an appropriate null value to record missing data.
- Don't use formatting to convey information or to make your spreadsheet look pretty.
- Place comments in a separate column.
- Record units in column headers.
- Include only one piece of information in a cell.
- Avoid spaces, numbers and special characters in column headers.
- Avoid special characters in your data.
- Record metadata in a separate plain text file.

Spreadsheets: Quality control

- Use data validation to prevent accidentally entering invalid data.
- Use sorting to check for invalid data.
- Use conditional formatting (cautiously) to check for invalid data.