**Spreadsheet Best Practices Workshop Fall 2019**

**Companion worksheet**

The inspiration for this workshop is the data carpentry workshop on spreadsheets.<http://www.datacarpentry.org/spreadsheet-ecology-lesson/>

Resources: Headfirst Excel O'Reilly Books – <http://uclid.uc.edu/record=b6626353~S39>

Hadley Wickham – Tidy Data <https://www.jstatsoft.org/article/view/v059i10>

Karl W. Broman & Kara H. Woo, *Data Organization in Spreadsheets*, Vol. 72, Issue 1, 2018, The American Statistician. <https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989>

Workshop Materials – <https://github.com/RebeccaOlson/UCL_Workshops>

Our goal is to create a machine readable spreadsheet.

**Structuring data in spreadsheets**

The cardinal rules of using spreadsheet programs for data:

1. Spreadsheets are meant to be machine readable
2. Put all your (**variables in columns)** - the thing you're measuring, like 'weight' or 'temperature'.
3. Put each (**observation) in its own row**.
4. One piece of (information) per cell. **Don't combine multiple pieces of information in one cell**. Sometimes it just seems like one thing, but think if that's the only way you'll want to be able to use or sort that data.
5. **Leave the raw data (raw)** - don't mess with it! Make a copy of the spreadsheet and work in the copy.
6. Export the cleaned data to a (**text based) format** like CSV. This ensures that anyone can use the data, and is the format required by most data repositories.
7. Also remember that a spreadsheet must also be machine readable.

**Blank Cells vs. Zero value – are they the same?**

Open the **SAFI\_messy** file.

How many tabs do you see?

How many mini spreadsheets do you see?

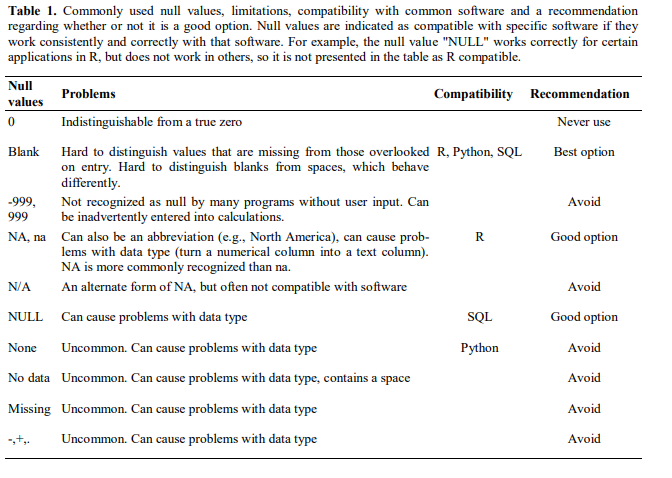
What formatting problems do you see?

For Excel we can use the term NULL to indicate that a value for the variable was not taken. Excel will also process a blank as a Null value. Other programs treat this expression and blank cells differently and you should be cautious when working between programs. The table below indicates what expression is best for each program.

Null Values Table From White et al, 2013, [Nine simple ways to make it easier to (re)use your data.](https://ojs.library.queensu.ca/index.php/IEE/article/view/4608)

Ideas in Ecology and Evolution - doi: 10.4033/iee.2013.6b.6f

Null Values.



**Metadata**

The information about the data – context, full information.

|  |
| --- |
| Open SAFI\_clean.  What metadata do you need to understand the variables? |

**Dates:**

**Note:** Dates are especially problematic for Excel. Here is a good article about the issues with date standards and how Excel causes issues.

Briney, Kristin A.. 2018. "The Problem with Dates: Applying ISO 8601 to Research Data Management." Journal of eScience Librarianship 7(2): e1147. <https://doi.org/10.7191/jeslib.2018.1147>

Download/open the **SAFI\_dates** file.

Note the dates in both tabs.

What do you notice?

Best practice tip = separate out the date into month, day, years.

|  |
| --- |
| Step 1. Highlight the Interview\_Date column. Check the format (should be DATE – select the correct formatting).  Step 2. Insert 3 empty columns next to the “interview\_Date” column. On the top, enter Month, Day, Year (or Day, Month, Year, depending upon tap)  Step 3. Make sure all three columns are formatted as numbers, with 0 decimal places.  Enter in the Month column in the first data cell, =**MONTH(A2).** Look for the little box in the lower right corner of that cell. Drag it down to apply the formula to the whole column. Repeat with **=DAY(A2), =YEAR(A2)**  Try entering just 11/1.  What do you notice? |

**Data Validation:**

For more information, see Microsoft’s tips for applying validation to cells. <https://support.office.com/en-us/article/Apply-data-validation-to-cells-29FECBCC-D1B9-42C1-9D76-EFF3CE5F7249>

**QA for Spreadsheets**.

You can require a range or list for entries to a cell through using the Data Tools settings on Excel.

Best practice tip: set up range/list parameters BEFORE data entry.

|  |
| --- |
| **SAFI\_Clean**  **For numbers**:  Select the column **no\_membrs**.  From the **DATA** tab, select **DATA TOOLS**, **DATA VALIDATION**/**VALIDATION TOOLS**.  Under **Settings**, select “Whole number” from the **ALLOW** menu.  Select Minimum (1) and Maximum (30).  Under **Input Message** enter “Invalid Number” as the title, and an error message.  Click OK.  Try entering a value outside of 1-30 in the column.  **For entries on a list**:  Select the column **responden\_wall\_type**  From the **DATA** tab, select **DATA TOOLS**, **DATA VALIDATION**/**VALIDATION TOOLS**.  Under **Settings**, select “List” from the **ALLOW** menu.  In the **Source** field, enter the values you want to be accepted, separated by commas.  Enter **grass, muddaub, burntbricks, sunbricks, cement**  Create an **INPUT** message then click ok. |

Contact us at [ASKDATA@UC.EDU](mailto:ASKDATA@UC.EDU)

Research & Data Services,

University of Cincinnati Libraries

<https://libraries.uc.edu/rds>