Tips for Encouraging Accuracy in Data Entry by Students
Permanent Forest Plot Project
Ecological Research as Education Network (EREN)
Primary Author: Kathleen LoGiudice, Union College

There are several points in data collection and processing that present opportunities for errors. Quality in actual data collection is addressed elsewhere. Here I suggest a way to maintain high quality through the data entry process.
Suggestions:
- Lines of data should be continuously numbered on the paper data sheets (i.e., a set of 6 pages numbered 1 – 300).
- Paper data sheets should precisely match Excel data entry forms (see example Excel file). If there are additional columns on the Excel data entry form, consider hiding column during the data entry process and filling them in as a block after the entry process is complete (for example collection date, which may be included only on the top of the paper data sheet but will be a separate column on the data entry form). If students must skip columns in the data entry forms, it makes errors much more likely.

Process:
- Collect all data sheets upon return from the field.
- Either photocopy sheets (2 copies of each sheet) or scan to high quality pdf file and post on Blackboard or a webpage accessible to the class.
- Post the blank Excel data entry form on GoogleDocs with lines numbered to match the paper data entry sheets.
- Assign blocks of data by line number to students. Each student should have 2 blocks of numbers, one that they are responsible to enter and one that they proofread.
- The data entry and quality checking is best done in 2 steps:
  - Instruct students to enter data by a given due date. There should be a column at the far right of the data for the initials of the person who entered the data.
    - Download the first version of the dataset before the proofreading period begins for quality checking later. Leave the original version posted on Googledocs.
  - Instruct students to proofread the data by a given due date. There should be a column at the far right of the data for the initials of the person who proofread the data.
    - Download the second version and compare to the first version using the technique described below to compare the two versions. Students' grades could be impacted by the accuracy of their data entry. For example, if they make a mistake (as documented by the proofreader), they lose points. If they found a mistake in their proofreading process, they earn points.
- To compare the two data sheet versions (the original as input and the proofread version):
  - Copy both the original and the proofread data sheets into the same workbook (named "Original" and "Proofread")
  - In a 3rd worksheet, use the "If" function to compare the cells in the Original data sheet to the Proofread version. This can be done as follows:
    - The original worksheet is titled "Original" and the proofread version is named "Proofread". Copy the title rows from the datasheets into the third "Compare" worksheet. In the In the first data cell (in this example, this is cell A4) enter: =IF(Original!A4=Proofread!A4,"ok","ERROR"). This will compare the contents of cell A4 in the Original sheet to the contents of cell A4 in the proofread sheet. If

they are the same, it will return "ok", if they are different, it will return "ERROR" and you will know that the proofreader found an error in the originally entered data. This can be copied to the entire "Compare" worksheet to highlight any differences between the original and the proofread sheets. You can then double check against the paper data sheets to be sure that the proofreader actually got it right and assign or remove credit accordingly.

- o There is an Excel workbook with example data included to illustrate this.
- Of course, accuracy depends upon your students completing these assignments. It may help to assure that the proofreading is done to announce that you will spot check a certain number of lines of data. Students should also initial the lines of data to show that they have been proofread.
- Students should be warned about using autocopy, which creates a series (adds one) to most numbers when it is used.