# Week 5 assignment ECOL 4130L

## Forensic Data Management

This week we’ve been focusing on where our data comes from and how we can understand what it’s saying. I realize that the wealth of files associated with the NEON data can be overwhelming, and hope that this week has helped to demystify them. In this assignment you’re going to get some experience with the far more common problem of poorly documented data. Attached to this assignment you will find an excel workbook called AZ-small-mammals.xlsx, that has 2 sheets. This workbook records data from 2 years of rodent trapping at a field site in Arizona. The data were collected by multiple scientists, and they did not collect the data in a standardized way. They also wrote no metadata for this dataset. Your job is to help make sense of these messy data. This job has 2 parts:

1. Combine all of the data into a single excel sheet, saved as a .csv file, with a single row for each rodent capture (see example), and a column for each piece of data about the rodent captures. Be sure to follow data best-practices like separating pieces of data into individual columns, not combining numbers and words in a single cell, and not using color or shading to encode information (see Wednesday’s lecture for more details).
2. Write metadata for this sheet in the style of NEON “variables” spreadsheets. This metadata spreadsheet should include the variable name, description, variable type (e.g “string” for words, “real” for numbers), and units for each column in your main spreadsheet.

See rubric for additional details and guidelines.

|  |  |  |  |
| --- | --- | --- | --- |
| **Grading Category** | Full Credit | Partial Credit | No Credit |
| **Data File** | | | |
| Data file structure (1 pt) | Data is saved as a single csv file with all data in continuous columns. | Data file is saved as another file type but still in continuous columns | Data file is made up of multiple sheets or multiple sets of columns on a single sheet |
| Data completeness (1 pt) | All capture records are included as separate rows in the data sheet | Some capture records are missing but greater than 75% of capture data is still present | Many capture records are missing or combined into single rows of data |
| Variable Columns (2 pt) | There is a single column for each variable included in the original data and the column is named according to established conventions (clear meaning, no spaces) | A small number of variables are missing or have been combined inappropriately into single columns, some variables are named inappropriately | Many variables missing, inappropriately combined into fewer columns, or named inappropriately |
| Data formatting (1 pt) | Words and numbers are not placed in the same cell. Colors are not used to contain information. | Some minor data formatting issues. | Many mixed data type cells are included and/or vital information is contained using color/shading |
| **Metadata File** | | | |
| Variables included (1 pt) | Metadata file contains a row for each variable column in the Data file | Metadata file contains a row for most variable columns in the Data file | Metadata file is missing rows for many variable columns in the Data file |
| Variable description (2 pts) | Metadata descriptions are brief but sufficient to understand the variable and exist for all variables | Metadata descriptions for some variables are missing or are difficult to understand | Metadata descriptions for many variables are missing or are difficult to understand |
| Data type (1 pt) | Data type column entry is included for all variables and is accurate | Data type missing or inaccurate for some variables | Data type missing or inaccurate for many variables |
| Data units (1 pt) | Data units column entry is included for all variables and is accurate | Data units column entry is missing or inaccurate for some variables | Data units column entry is missing or inaccurate for many variables |