BASIC DATA SUMMARY AND ANALYSIS MODULE FOR AN ECOLOGY LABORATORY COURSE

***Robert F. Smith** , Lycoming College Department of Biology, 700 College Pl, Williamsport, PA 17701*

smithr@Lycoming.edu / www.rfsmithecology.com

INTRODUCTION

- Contemporary topics in ecology and environmental science that examine ecological patterns and processes require:
  - Quantitative approaches
  - Analysis of large spatial or temporal scale data sets
- Experiential learning requires **tools that are:**
  - **Reasonably simple**
  - **Effective and meaningful**
- Tools must be useful for data **1) organization and management; 2) summary; and 3) analysis**
- Microsoft Excel can be a useful tool for undergraduate students
  - Adequate capabilities
  - Familiar to students
  - Broadly applicable to students in numerous sub-disciplines in biology (Table 1)
- This module is used at the beginning of a 2nd year undergraduate ecology course.

*Table 1. Partial list of fields from O*net OnLine (https://www.onetonline.org/) that listed Microsoft Excel a "hot technology" (meaning that it is "a technology requirement frequently included in employer job postings").*

Field	Lists MS Excel as a "hot technology"
Aquaculture manager	Y
Soil and water conservationists	Y
Park naturalist	Y
Fish and Game Wardens	Y
Forest and Conservation Technicians	Y
Natural Sciences Managers	Y
Park Naturalists	Y
Zoologists and Wildlife Biologists	Y
Gas Plant Operators	Y
Methane/Landfill Gas Generation System Technicians	Y
Biofuels Processing Technicians	Y
Physician Assistants	Y
Physical Therapists	Y
Clinical Nurse Specialists	Y

GOALS AND INTENDED LEARNING OUTCOMES

General purpose of this module:

- Increase student proficiency using MS Excel.
- Teach students the importance of data organization and summary for accurate analysis.
- Teach students about data structures (e.g., tabular, matrix, etc.)

Specific learning outcomes for students:

- **Learn specific approaches and protocols** for organizing, summarizing, and analyzing simple and complex datasets using MS Excel
- Develop a **semester-long resource** for students
- Develop **a conceptual understanding** of the basic approaches for data organization and summary (and the importance of each)
- Understand approaches to working with different data structures (e.g., tabular, array/matrix, etc.)

CHALLENGES AND PITFALLS FOR STUDENT LEARNING

- Resistance to using the lab as a semester-long resource
- Demonstrating relevance of these skills:
  - For course assignments
  - As a skill needed beyond graduation
- Can discourage critical thinking
  - Much of the lab is a walkthrough of specific procedures (i.e., a 'canned lab')
  - Students must apply the skills to an assignment, but often have difficulty applying information to novel situations

FUTURE WORK

- Teach additional analysis, summary, and display tools:
  - Box and whisker plots
  - Pivot graphs (directly from pivot tables)
- Create online tutorials as supplemental resources (and encourage continued use as a resource)
- Parallel introduction of the R statistical program (possibly)
- Direct references to specific components of other assignments

ACKNOWLEDGEMENTS I thank the students in my BIO224 (Ecology) classes for their feedback. I want to thank Rob Bertin (College of the Holy Cross) for valuable feedback about how to teach quantitative concepts in general ecology courses. This work was supported, in part, by the NSF, Engineering, and Education for Sustainability (SEES) Fellowship Grant GEO-1215896NSF.

IMPLEMENTATION: LAB SCALE

- Students work through a demonstration first that introduces specific skills (below)
- Data is too big to cheat
- Design uses examples that are common barriers to completing other labs throughout the semester

Basic operations

- Rows, columns, cell coordinates, and worksheets (i.e., tabs)
- Shortcuts for navigating worksheets, copy/paste, and formatting
- Sorting data

Formulas

- Mathematical operators (+, -, >, <, =, etc.)
- Using formulas
- Basic (e.g., sum) and advanced (e.g., count, average, min, max, stdev, etc.) formulas
- 'vlookup' function (relating tabular data)
- Logical arguments (i.e., if/then)

Pivot table

- Checking data integrity, making lists of responses, etc.
- Creating a matrix from tabular data

Histogram

- Analysis toolpack
- Using the tool and creating the plot
- Examining distributions (informal)

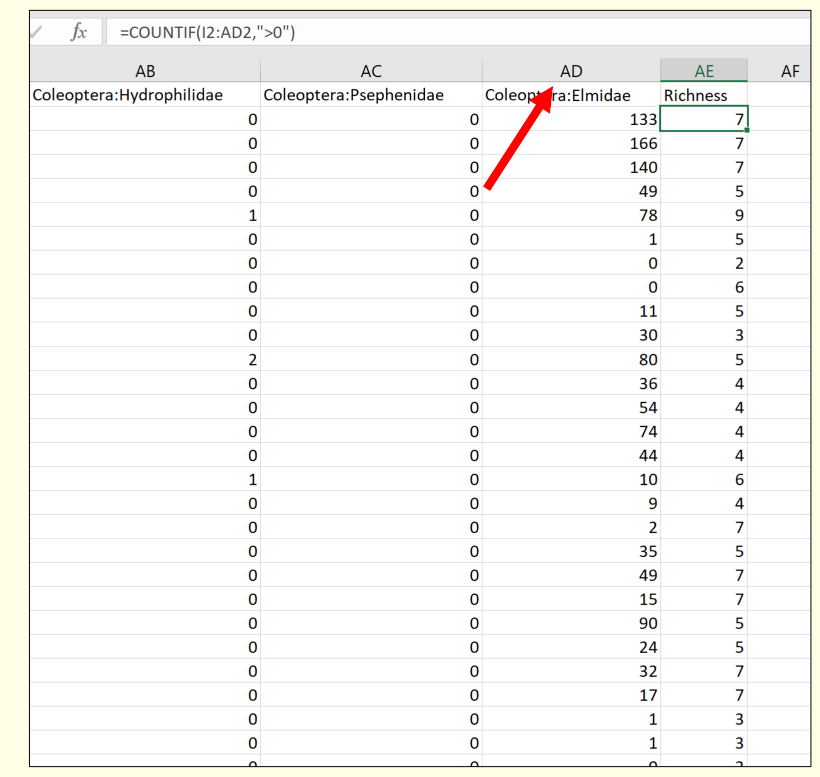

Figure 1. Screen capture of example demonstrating how to use an 'advanced' function to calculate an ecological index (taxon richness determined from the countif() function). In this case, the number of columns (i.e., taxa) is very large (red arrow), which prevents students from doing the calculations by hand.

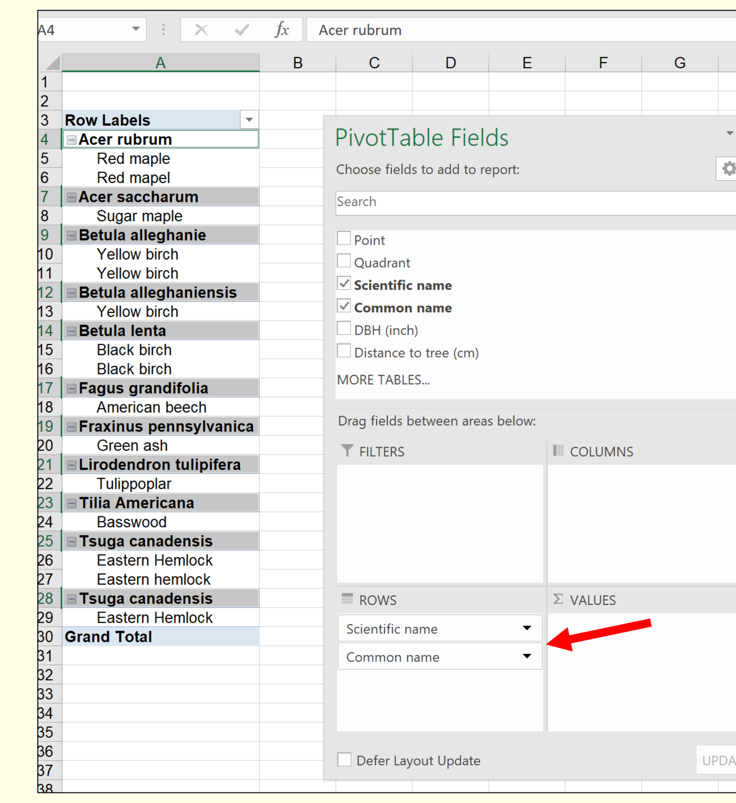

Figure 2. Screen capture of example using the pivot table to look for errors and to organize data. Common errors include spelling mistakes, (invisible) spaces, etc., and these errors must be corrected prior to analysis. Using only the rows (red arrows) can create nested lists of categorical variables

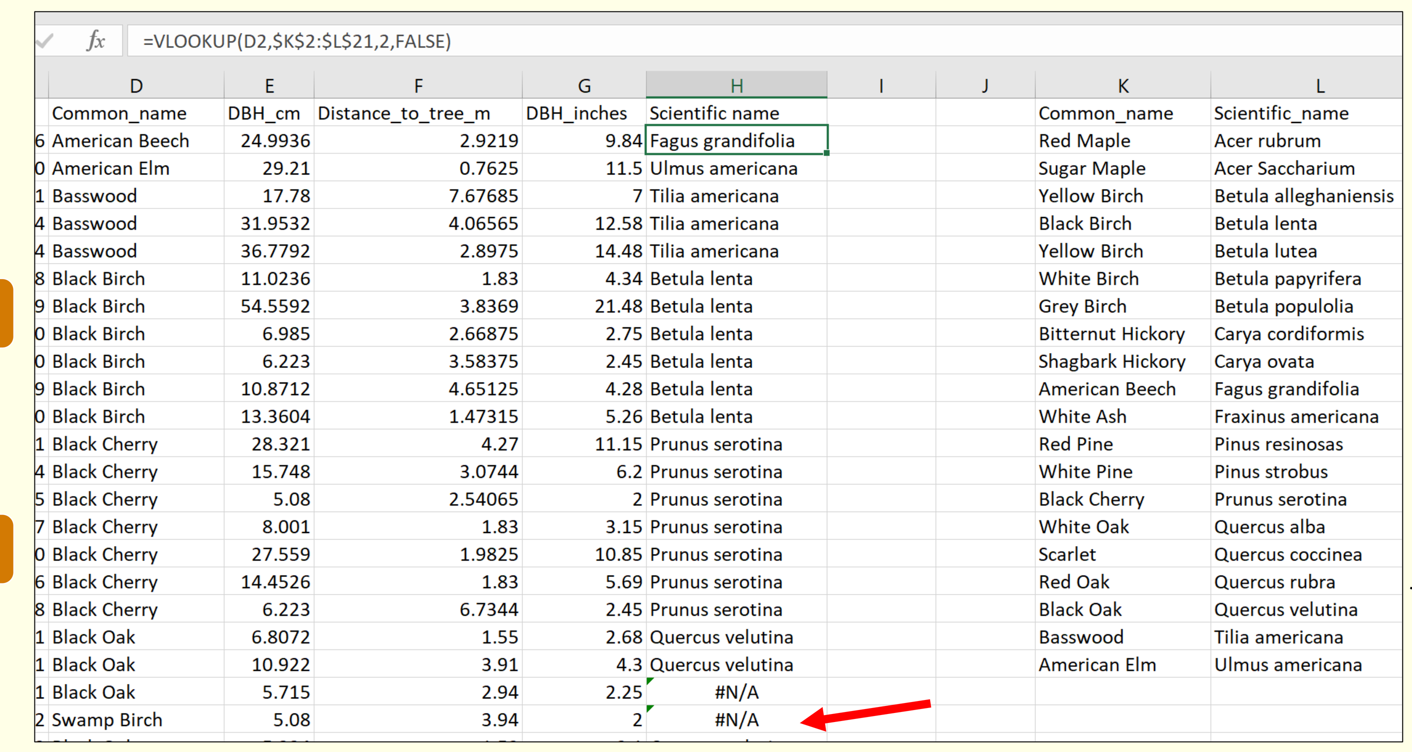

Figure 3. Screen capture of example using the vlookup() function to relate two datasets (i.e., add the scientific name to column H by matching column D and K) and to find potential errors in the data. As above, spelling mistakes, spaces, and other typographic errors will produce errors for analysis. This function produces a #N/A (red arrows) when a match can't be found rather than a zero or blank. #N/A are special cell entries.

The module (writeup and datasets) and future updates are available here: http://www.rfsmithecology.com/teaching.html

IMPLEMENTATION: COURSE/SEMESTER SCALE

APPLICATIONS TO OTHER LAB TOPICS

- Exploring logistic growth using Excel spreadsheets
- Census of forest tree communities
- Measuring species richness of stream fish communities
- Measuring forest vegetation characteristics to demonstrate succession
- Examining predator-prey interactions
- Investigating natural variability of samples of aquatic communities
- GIS analysis of landscape characteristics
- Semester-long project (long-term datasets from Harvard Forest)

DEMONSTRATING BROAD RELEVANCE

- Include a component that requires students to calculate grades
- Stress that proper data organization (i.e., 'quality control') leads to accurate analyses (and better grades)
- Stress throughout the semester that **distributions and variability in the data can be informative** for choosing analysis methods
- Incorporated common barriers for using Excel (e.g., spelling mistakes in data, improper formatting, spaces before/after text, etc.) helps generate some interest in learning the skills
- Explicit reminders in semester exercises about the use of these skills are needed to broaden relevance for the semester (requires improvement)
- **Primary outcome to date: difficulty generating student interest in applied concepts** beyond the basic use of the program

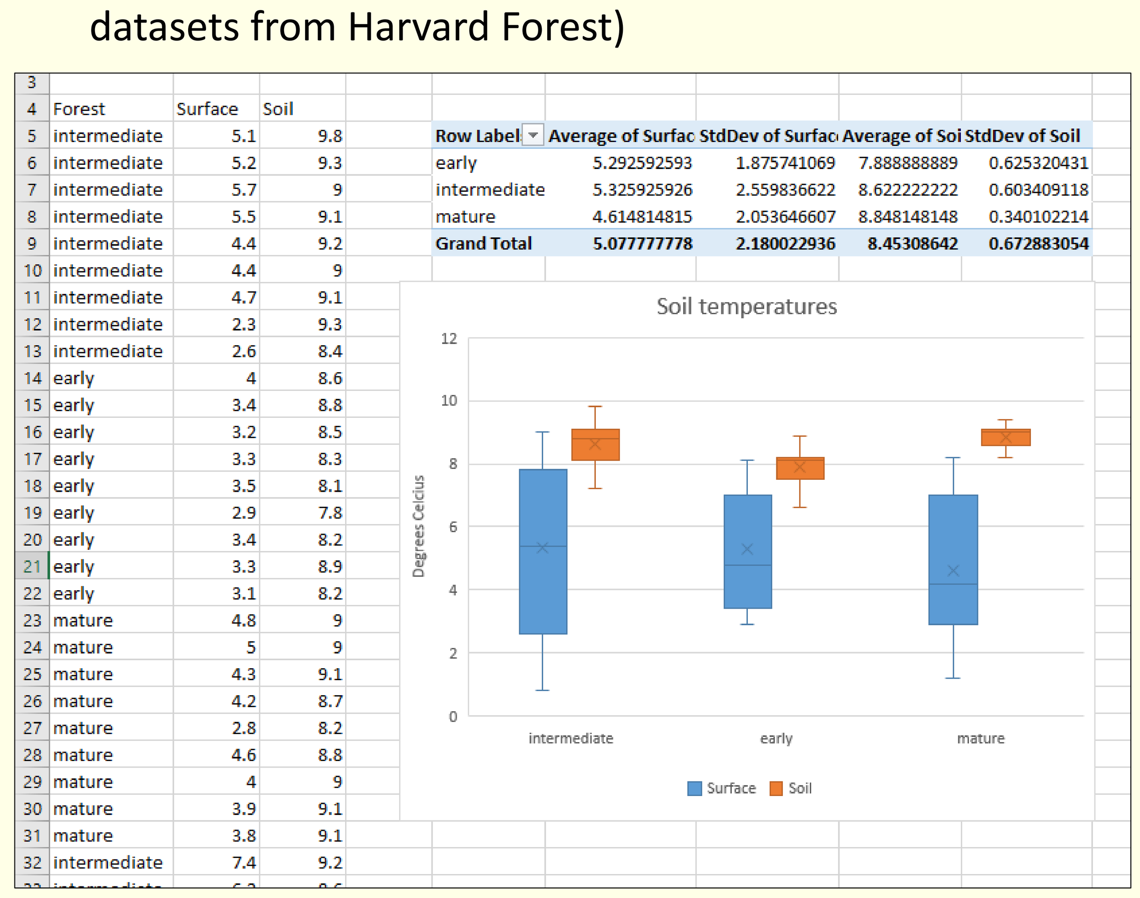

Figure 4. Screen capture of results from the forest succession exercise showing the use of pivot table and a box and whisker plot to investigate data structures.

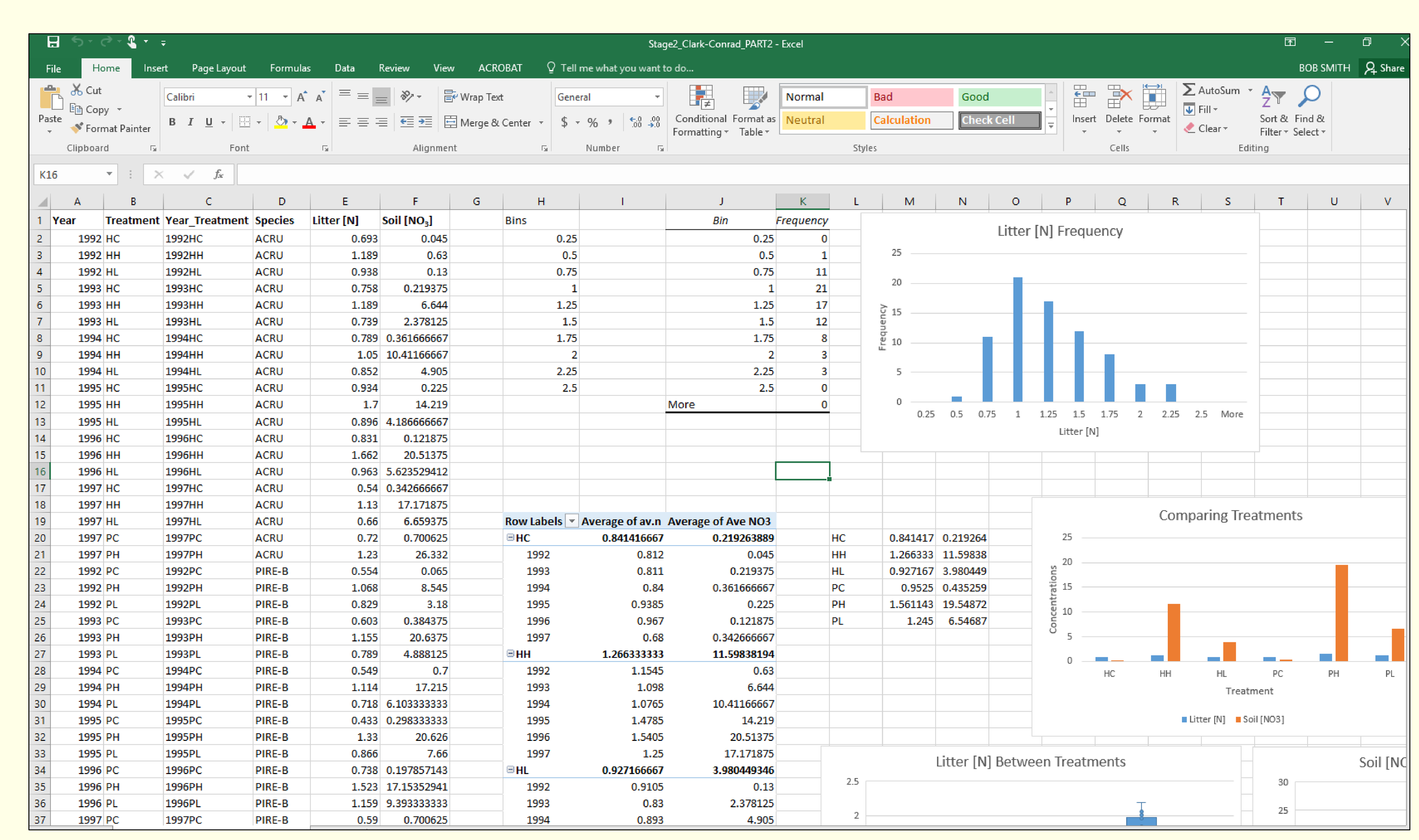

Figure 5. Screen capture of results from a student group's semester long project showing how students can learn to perform multiple summary analyses of the data (e.g., histogram and multiple uses of the pivot table.

CONCLUSIONS

- Teaching the use of MS Excel without a clear application to broader course materials or as a needed skill for employment is likely to be fail.
- A conceptual understanding of data management, summary, and analysis and knowledge of the MS Excel tools to perform this analysis is helpful for students who:
  - Are performing independent research outside of class (confirmed through informal student comments).
  - Are interested in field/lab technician jobs (especially in ecological fields).
- Students tend to avoid these techniques when working with small datasets even though they increase accuracy of analyses.
- Students often failed to carefully examine data organization results and made mistakes analyzing error-prone class aggregated datasets.