

# 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



- 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  $\bullet$
  - Broadly applicable to students in numerous sub-disciplines in biology (Table 1)

## **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

leoptera:Hydrophilidae Coleoptera: Psephen

 $f_x$  =COUNTIF(I2:AD2,">0")

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 (rea arrow), which prevents students from doing the calculations by hand.

6		A		Ĩ.
CLEAN	WAI		3111	

Ö. -

	A	в				F	
1							
2							
3	Row Labels						
4	■Acer rubrum		PivotTa	ble Fiel	ds		
5	Red maple		Chasse field	a ta add ta w	o o o utu		
6	Red mapel		Choose field		epon:		
7	■Acer saccharum		Search				
8	Sugar maple						
9	🗏 Betula alleghanie		Point				
10	Yellow birch		Ouadrant				
11	Yellow birch		Scientific				
12	■Betula alleghaniensis			aname			
13	Yellow birch		Common	name			
14	<b>⊟Betula lenta</b>		DBH (inch	ו)			
15	Black birch		Distance 1	to tree (cm)			
16	Black birch		MORE TABLE	-5			
17	■Fagus grandifolia						
18	American beech		Drag fields h	atwoon area	c bolow:		
19	■Fraxinus pennsylvanica		Drag fields i	Jetween area	is below.		
20	Green ash		<b>T</b> FILTERS				
21	■Lirodendron tulipifera				_		
22	lulippoplar						
23	🗏 Tilia Americana						
24							
25	E Isuga canadensis						
26	Eastern Hemlock						
27							
20			≡ ROWS			∠ VALUES	
20	Grand Total		Scientific n	ame	-		
30 31	Grand Total		Common r				
21 22			Common r	lame	·		
ג גע							
34							
35							
36			_				
37			Defer Lay	out Update			
38							

- :  $\times$   $\checkmark$   $f_x$  Acer rubrum

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

• 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

- Mathematical operators (+, -, >, <, =,</li> 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 3. Screen capture
fx =VLOOF	(UP(D2,\$K\$2	2:\$L\$21,2,FALSE)							of example using the
D	E	F	G	Н	I	J	К	L	vlookun() function to
Common_name	DBH_cm	Distance_to_tree_m	DBH_inches	Scientific name			Common_name	Scientific_name	
S American Beech	24.9936	2.9219	9.84	Fagus grandifolia			Red Maple	Acer rubrum	relate two datasets
) American Elm	29.21	0.7625	11.5	Ulmus americana			Sugar Maple	Acer Saccharium	(i.e., add the scientific
Basswood	17.78	7.67685	7	Tilia americana			Yellow Birch	Betula alleghaniensis	
l Basswood	31.9532	4.06565	12.58	Tilia americana			Black Birch	Betula lenta	пате то соштп н бу
l Basswood	36.7792	2.8975	14.48	Tilia americana			Yellow Birch	Betula lutea	matching column D and
3 Black Birch	11.0236	1.83	4.34	Betula lenta			White Birch	Betula papyrifera	K) and to find notantial
Ə Black Birch	54.5592	3.8369	21.48	Betula lenta			Grey Birch	Betula populolia	K) unu to jina potentiar
) Black Birch	6.985	2.66875	2.75	Betula lenta			Bitternut Hickory	Carya cordiformis	errors in the data. As
) Black Birch	6.223	3.58375	2.45	Betula <mark>l</mark> enta			Shagbark Hickory	Carya ovata	ahove snelling
) Black Birch	10.8712	4.65125	4.28	Betula <mark>l</mark> enta			American Beech	Fagus grandifolia	ubove, spennig
) Black Birch	13.3604	1.47315	5.26	Betula lenta			White Ash	Fraxinus americana	mistakes, spaces, and
Black Cherry	28.321	4.27	11.15	Prunus serotina			Red Pine	Pinus resinosas	other typographic
Black Cherry	15.748	3.0744	6.2	Prunus serotina			White Pine	Pinus strobus	
Black Cherry	5.08	2.54065	2	Prunus serotina			Black Cherry	Prunus serotina	errors will produce
7 Black Cherry	8.001	1.83	3.15	Prunus serotina			White Oak	Quercus alba	errors for analysis. This
) Black Cherry	27.559	1.9825	10.85	Prunus serotina			Scarlet	Quercus coccinea	function and uses a
Black Cherry	14.4526	1.83	5.69	Prunus serotina			Red Oak	Quercus rubra	Junction produces a
Black Cherry	6.223	6.7344	2.45	Prunus serotina			Black Oak	Quercus velutina	#N/A (red arrows) when
Black Oak	6.8072	1.55	2.68	Quercus velutina			Basswood	Tilia americana	a match can't be found
L Black Oak	10.922	3.91	4.3	Quercus velutina			American Elm	Ulmus americana	a match can t be jound
L Black Oak	5.715	2.94	2.25	#N/A					rather than a zero or
2 Swamp Birch	5.08	3.94	2	#N/A					blank. #N/A are special

cell entries.

The module (writeup and datasets) and future updates are available here: <a href="http://www.rfsmithecology.com/teaching.html">http://www.rfsmithecology.com/teaching.html</a>

## GOALS AND INTENDED LEARNING OUTCOMES

General purpose of this module:

## **IMPLEMENTATION: COURSE/SEMESTER SCALE**

**APPLICATIONS TO OTHER LAB TOPICS** 

### DEMONSTRATING BROAD RELEVANCE

- 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')

- 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)

3										
4	Forest	Surface	Soil							
5	intermediate	5.1	9.8	Row La	abel 🔻	Average of Surfac	StdDev of Surfac	Average of Soi	StdDev of Soil	
6	intermediate	5.2	9.3	early		5.292592593	1.875741069	7.888888889	0.625320431	
7	intermediate	5.7	9	interm	ediate	5.325925926	2.559836622	8.622222222	0.603409118	
8	intermediate	5.5	9.1	mature	e	4.614814815	2.053646607	8.848148148	0.340102214	
9	intermediate	4.4	9.2	 Grand	Total	5.07777778	2.180022936	8.45308642	0.672883054	
10	intermediate	4.4	9							
11	intermediate	4.7	9.1			Soil	temperatures			
12	intermediate	2.3	9.3	12			comportaction			
13	intermediate	2.6	8.4	12						
14	early	4	8.6							
15	early	3.4	8.8	10		Т				
16	early	3.2	8.5		-		т		×	
17	early	3.3	8.3	8				T	[	
18	early	3.5	8.1							
19	early	2.9	7.8	5						
20	early	3.4	8.2		$\rightarrow$	<u> </u>	$\times$			
21	early	3.3	8.9	 ۳ ۲				$\times$		
22	early	3.1	8.2	4						
23	mature	4.8	9				$\perp$			
24	mature	5	9	2						
25	mature	4.3	9.1							
26	mature	4.2	8.7	0						
27	mature	2.8	8.2	0	int	termediate	early	r	nature	
28	mature	4.6	8.8				,	-		
29	mature	4	9				Surface 📕 Soil			
30	mature	3.9	9.1							
31	mature	3.8	9.1	 						
32	intermediate	7.4	9.2							
22	2 - A	<b>C</b> 2	0.0							

- 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

н	2.6.4							Sta	ge2_Clark-Conrad_PART2	- Excel										团		
File	Home In	isert Page Layout	Formul	as Data f	Review View	V ACRO	оват 🖓 те	ell me what you want	to do											во	B SMITH	₽ Sh
Paste	K Cut E Copy ← ✓ Format Painter Clipboard	Calibri B I U - E	• 11 • A		Alignmer	Wrap Text Merge &	t Ge Center - \$	eneral ▼ ▼ % ♪ €.0 .00 Number 5	Conditional Format a Formatting ▼ Table ▼	Normal s Neutral	Style	Bad Calculation	Good	l k Cell	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	ert Delete Cells	Format	∑ A ↓ F ◆ C	AutoSum → ïll <del>+</del> Clear + Ed	Sort & Find Filter - Sele	) 1& ct ▼	
К16		$\times \checkmark f_x$	D	E	F	G	н	I	L	K	L	м	N	0	р	0		R	S	т	U	v
1 Ye	ar Treatme	nt Year Treatment	Species	Litter [N]	Soil [NO <sub>2</sub> ]	-	Bins		Bin	Frequency	, _			_					-			
2	1992 HC	1992HC	ACRU	0.693	0.045		0.2	25	0.25	5 0				Litter	[N] Frequ	lency						_
3	1992 HH	1992HH	ACRU	1.189	0.63		0	.5	0.5	5 1		25										
4	1992 HL	1992HL	ACRU	0.938	0.13		0.7	75	0.75	5 11				_								
5	1993 HC	1993HC	ACRU	0.758	0.219375			1	1	21		20										
5	1993 HH	1993HH	ACRU	1.189	6.644		1.2	25	1.25	5 17	,	Č 15										
7	1993 HL	1993HL	ACRU	0.739	2.378125		1	.5	1.5	5 12		e D										
8	1994 HC	1994HC	ACRU	0.789	0.361666667		1.7	75	1.75	5 8		ยั้ 10	_									
9	1994 HH	1994HH	ACRU	1.05	10.41166667			2	2	2 3		LL.										
0	1994 HL	1994HL	ACRU	0.852	4.905		2.2	25	2.25	j 3		5										
11	1995 HC	1995HC	ACRU	0.934	0.225		2	.5	2.5	5 0			_									
12	1995 HH	1995HH	ACRU	1.7	14.219				More	0		0	0.5 0.7	75 1	125 15	1 75	2 2	25 21	5 More			
13	1995 HL	1995HL	ACRU	0.896	4.186666667							0.20	0.5 0.	/J 1	Litter [	1.75 N]	2 2.1	23 2	5 MIDIC			
14	1996 HC	1996HC	ACRU	0.831	0.121875										Litter [	ni]						
15	1996 HH	1996HH	ACRU	1.662	20.51375															Ţ		
16	1996 HL	1996HL	ACRU	0.963	5.623529412						<b>_</b>											
17	1997 HC	1997HC	ACRU	0.54	0.342666667																	
18	1997 HH	1997HH	ACRU	1.13	17.171875												C	omna	ring Tro	atments		
19	1997 HL	1997HL	ACRU	0.66	6.659375		Row Labels	<ul> <li>Average of av.n</li> </ul>	Average of Ave NO3								C	omha	ing ne	armenta		
20	1997 PC	1997PC	ACRU	0.72	0.700625		⊟HC	0.841416667	0.219263889	)	HC	0.841417	0.219264		25 —							
21	1997 PH	1997PH	ACRU	1.23	26.332		1992	0.812	0.045		нн	1.266333	11.59838		20 —							

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

I thank the students in my BIO224 (Ecology) classes for their feedback. I want to **ACKNOWLEDGEMENTS** 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.

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.

# CONCLUSIONS

0.4116666

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.

- 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.