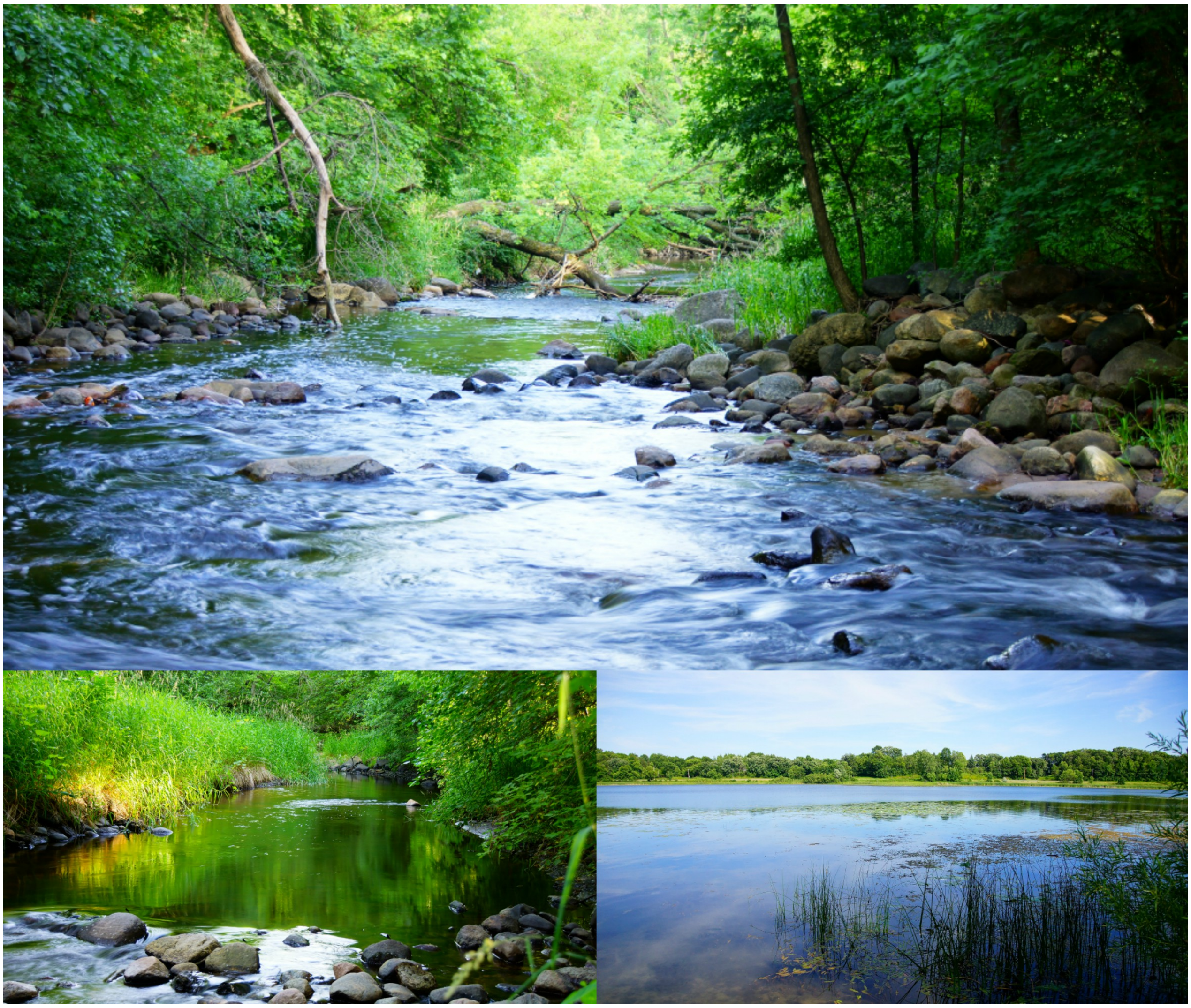
Teaching Manual for General Education Mathematics with Watershed Data

Engaging Mathematics Teaching Manual



Dr. Anthony Dunlop
Normandale Community College

Contents

License Information and Credit

This teaching manual was created for Engaging Mathematics with support from the National Science Foundation.

An initiative of the National Center for Science and Civic Engagement, Engaging Mathematics applies the well-established SENCER method to college level mathematics courses, with the goal of using civic issues to make math more relevant to students.

Engaging Mathematics will: (1) develop and deliver enhanced and new mathematics courses and course modules that engage students through meaningful civic applications, (2) draw upon state-of-the-art curriculum in mathematics, already developed through federal and other support programs, to complement and broaden the impact of the SENCER approach to course design, (3) create a wider community of mathematics scholars within SENCER capable of implementing and sustaining curricular reforms, (4) broaden project impacts beyond SENCER by offering national dissemination through workshops, online webinars, publications, presentations at local, regional, and national venues, and catalytic site visits, and (5) develop assessment tools to monitor students' perceptions of the usefulness of mathematics, interest and confidence in doing mathematics, growth in knowledge content, and ability to apply mathematics to better understand complex civic issues.

Updates and resources developed throughout the initiative will be available online at http://engagingmathematics.net/. Follow the initiative on Twitter: @MathEngaging.

Support for this work was provided by the National Science Foundation under grant DUE-1613211, formerly DUE- 1322883, to the National Center for Science and Civic Engagement. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Chapter 1

Introduction

This "Engaging Mathematics" course has been taught at Normandale Community College in Bloomington, Minnesota, USA as a section of Math 1020, *Mathematics for the Liberal Arts*, several times since Fall 2011. Other sections of Math 1020 are taught using a more conventional, textbook-based curriculum. I myself used Burger and Starbird's delightful *The Heart of Mathematics* for the course prior to this redesign.

Since the course makes extensive use of spreadsheets (the activities are written assuming Excel is being used, and the data are in Excel worksheets), the course is taught in a computer classroom. Due to space constraints, it was taught during one semester in a twice-a-week format in which we had computers for only one of the two days. This is reflected in the current state of the Activities, some of which do not assume a computer is readily available.

Math 1020 is intended for students who are seeking a Bachelor's degree in the social sciences or humanities, and who do not plan further study in mathematics or in any quantitatively based science. Therefore, the curriculum presented here was not designed with any specific "learning outcomes" in mind; the specific learning outcomes attached to each Activity arose spontaneously as suggested by the nature of the available data and materials. Instead of a predetermined list of topics to "cover," I wanted students to come away with some broad quantitative literacy, and familiarity with common analytical tools, which are brought to bear on issues that affect their immediate surroundings. The prerequisites for the class are a basic background in high school mathematics. No knowledge of trigonometry is assumed, and none is used. There is, in fact, very little algebra—in the sense of graphing functions and solving equations—in the class. I should specify that there is very little graphing *by hand*—the students do quite a lot of graphing of regression equations using spreadsheet software. At Normandale, there is a one-semester developmental mathematics course designed for students who do not plan further STEM coursework, which covers typical algebraic topics through quadratic functions and *very* elementary properties of logarithms. This course serves as a prerequisite for Math 1020. In practice, the level of skill and preparation of the students in Math 1020 varies widely, to say the least.

I made use of such data as I could find (for now almost exclusively from the Nine Mile Creek Watershed District) to raise questions, which were, hopefully, clearly motivated by the data and the literature that accompanied it. The students would then learn some mathematical tools that shed light on those questions. Of course any such tools needed to be accessible to students whose background does not go beyond the prerequisites for this course.

My desire was to have the course bring mathematical tools to bear on something students see every day. That way nobody could claim that these tools (such as ordinal-level ranking scales; linear and other regression graphs and equations; scatterplots and correlation coefficients) were somehow remote or abstract. Thus I fixated, early on, on Nine Mile Creek, which flows right past the Normandale campus, and most students cross a bridge over it on their way to and from campus. As we go through the course, I'll make suggestions about how the material could be adapted for other locations. The sort of data I use is widely collected anywhere there are lakes and streams. I would appreciate any suggestions from anyone reading these materials!

As I've taught the course, I've gradually come to realize that it can be divided, more-or-less naturally, into three or four "units," which are fairly coherent thematically, both in terms of the kinds of mathematical analysis done and (to a lesser extent) in the kinds of environmental issues addressed. Each Unit is introduced as it arises in the materials below.

The first two times I taught the course, I took the class to a Bloomington city park, called Moir Park, which is about two miles downstream from campus. There I broke the class into teams of two to four, depending on the task, and had them gather field data that was then incorporated into future class activities. We even borrowed a flow-meter and some hip-waders from the Geology/Geography department and had students measure the creek flow, from which they later estimated the creek discharge. Both of those classes had longer meeting times: The first time (Fall 2011) was a three hour, one evening per week class, and the second time (Spring 2012) met twice per week for an hour and a half. This time span was sufficient to get everyone to the park and back by the end of the scheduled class time. Since then the class has been offered in a thrice-per-week, 50-minute format, which does not lend itself to these trips. The instructions for the various data-gathering tasks in the field trips are in Appendix B.

The structure of the class is centered on daily "Activities," which are designed to be done in class with an instructor readily available to answer questions. The Activities are not written as outside-of-class homework, but are designed with student-instructor interaction in mind. I learned very early on that attendance must be *required*. The maximum number of absences I allow is equivalent to three weeks of class. I do post blank Activities on the course Web site after class, but do not accept late work for an absent student any later than the following class day unless the student has contacted me with a valid excuse. I require students to work in groups of two or, at most three, and to make sure they talk to each other I only hand out one copy of the day's Activity to each team. An instructor planning to use this material should be aware that, due to the aforementioned wide variation in skill and preparation level, many teams will finish the day's Activity early—sometimes very early. On the other hand, I always have at least one, and usually two or more, teams still working at the end of class. The Activities included here, in Appendix A, are based on three 50-minute class periods per week. There are a few instances in which two consecutive Activities could probably be combined without causing too much hardship for the slower students, but I consider it crucial in a course of this kind to do everything I can not to leave anybody "lost" and falling behind. As currently designed, the only way students get "homework" to be completed outside class is if they do not finish that day's Activity during the allotted class time. Thus there is no coursework that must be done outside class, apart from some reading assignments. I've learned to handle the variability in student "speed" by periodically (two or three times per semester) having a "catch-up day," when students who are finished with all current activities are excused from class and no new material is given. I also allow students to re-do some of the Activities for partial restored credit, which must be done outside class, but this is optional.

The first two days or so of the class have the most lecturing of the semester, as I begin to set the stage for the kinds of problems the students will be encountering. I have a standard harangue about the meaning of "liberal arts," referencing classical Greece and pointing out that at least two of the seven liberal arts were unambiguously mathematics. OK, so that's probably just me, and my background, and sadly most of the students seem pretty tuned out for that part. Then I get into a discussion about just what a watershed is, and I present a map showing the Nine Mile Creek Watershed, pointing out the location of our campus relative to the various "reaches" we'll be studying. I tell them, very broadly, what kinds of questions we'll be asking, largely about environmental quality and stormwater/flood control, and begin to suggest why these questions cannot be seriously and thoroughly addressed without mathematics: At the most basic level, we'll need to measure things; I'll then usually talk about things that can degrade water quality starting with goose poop (that usually gets their attention—it allows me to explain why mowing grass all the way up to the water's edge is such a bad idea, and this observation comes into play in the actual coursework pretty soon) and moving up to phosphorus, a key plant growth accelerant, which is why it's used in gardens, lawns, and golf courses—of which there are plenty nearby—but why we also have such a problem with algae blooms in our suburban lakes. This brings up the concept of a *relationship* between *variables*, such as phosphorus concentration and algae population, which we'll be developing tools to try to understand. (I never actually use the language of "function." I'm kind of afraid to with this audience—many of them are underchallenged by the course, but for quite a few of them it takes all they can muster to claw through with a C. Too much abstraction just might push them over the edge.)

Chapter 2

The Course: Daily Readings and Activities

The heart of this course is the daily Activities. What follows are brief descriptions of the four "Units" which divide the course, and of the actual individual Activities. It is hoped that the descriptions will be helpful for instructors wishing to include some of the materials, concepts, or lessons into their own courses. Many of the Activities could be used as stand-alone class projects in courses on elementary statistics, developmental algebra, or college algebra. The Activities and supporting materials (such as Excel spreadsheets) are available at the Engaging Mathematics Web site by following the links found at the beginning of each description.

2.1 UNIT 0: Introduction to Using Numbers and Quantitative Methods to Analyze and Classify Waterway "Quality"

The students are given a reading packet (see "Reading01 and Reading02," at the Engaging Mathematics Web site) from the Nine Mile Creek Watershed District. This packet contains a description of several ways in which environmental engineers and ecologists can assess, and make at least somewhat precise and non-arbitrary, the vague, intuitive notion of "waterway quality." The more inclusive term "waterway," rather than just "water quality," indicates an interest in a larger ecosystem, not just chemical impurities or other undesirable modifications to the water itself. In the readings, the students encounter the classification scheme for moving waterways (creeks and rivers) used by the local watershed district, which is hierarchical and classified on a scale of A through E, with A being the most pristine ranking and E being the most degraded. These rankings are called the "ecological use" of a waterway.

One of the interesting things about this classification scheme is that it can be approached from many directions. The reading refers to four of them, only three of which the students work with.

Existing Ecological Use refers to the condition of the waterway at one specific moment in time, and is assessed directly by looking at the specific kinds of fish currently living there. Historical Ecological Use is simply the arithmetic mean of Existing Ecological Uses of the same waterway over a specific period of time. Potential Ecological Use is an attempt to assess, based on many observations in and around the waterway, some of which are of a highly qualitative and somewhat subjective nature, what sort of organisms *should* be supportable by the waterway.

The students are told not to expect to have a thorough understanding of the reading by the next class, but to pay particular attention to the descriptions of how "ecological use" is defined and measured. The first couple of activities have them articulate this classification strategy as coherently and succinctly as possible, *and*, most importantly, from the point of view of mathematics and quantitative literacy, how numbers and arithmetic are indispensable to this process.

2.1.1 Activities 0 and 1: Ecological Use in General; Historical Ecological Use

Activity 0 Word pdf

Learning Outcomes:

- Converting a "qualitative" scale ("ordinal," in statistical parlance) to numerical
- Expressing a numerical process in one's own words, based on a written description
- Reading and interpreting graphical summaries of data
- Computing arithmetical mean from graphical data and interpreting the result in context

The overarching theme in these first couple of days of class, from a *mathematical* perspective, is the versatility of the concept of "number." We have a basic question: What, *exactly*, might one mean by the "quality," or even the "environmental quality," of a waterway? There are many specific attributes that might come to mind, some more relevant than others. But underlying any specific criterion—diversity of organisms supported, aesthetic value, chemical purity of the water, et cetera—is the notion of a *ranking*, or a hierarchy. Ultimately when one speaks of the "quality" of a waterway (or of a cake, musical performance, or . . .) one seeks to say things like, "This one is better than that one," or "This has improved since last year." There is comparison implied in any such hierarchy.

The main idea I intended students to get out of these first two Activities is that any such ranking is inherently mathematical, if not strictly quantitative, in nature, as it invokes one of the key properties of the number concept: namely ordering.

The first reading given to the students contains an extensive discussion of one way this ordering, or ranking, of a watershed "reach" can be done. (A "reach" is a geographically specific length of a stream, creek, or river.) (The ranking method introduced there has a citation as to its origin.) The ranking, whose nomenclature (coldwater sportfish, tolerant macroinvertebrates, et cetera) suggest ranking based on capacity to support various organisms, is also given an ordering in the lexical style, A - E (with A representing the top, E the bottom, of the hierarchy). The first thing students are asked is: How is that ranking numerical—or even mathematical?

This question follows an extensive discussion (actually usually mostly a monologue, unless I have an unusually talkative student or two) about three distinct, but related, aspects of the number concept:

- Order
- Counting
- Measuring.

So what the students are really being asked here is to identify which of these three properties of "number" is being invoked in discerning this hierarchy for waterway quality.

Once it is recognized that a *property* of number is being used, it's a fairly short trip (down what some would call a slippery slope) to actually assigning numerical values to the various levels of the ranking hierarchy. This brings us to Activity 1, and the three distinct kinds of "ecological use" that are described in the reading packet. All three use the same Ecological Use rankings, A - E (coldwater sportfish - very tolerant macroinvertebrates), and all three will be explored in this opening "Unit Zero."

Activity 1 Word pdf

The students should now be cognizant of the ranking hierarchy for Ecological Use. The Activity begins by reminding them that the environmental engineers and ecological scientists involved in this kind of reporting almost unconsciously switch into a very quantitative mode of thinking; in statistical terms, they switch from *ordinal* to *ratio scale* without batting an eye (hence the slippery slope mentioned earlier—I have chosen not to bring to their attention the fact that there is a school of thought that says this is unwise; clearly it is standard practice in many disciplines, and even in academia: cf. "grade point average").

The reading given prior to Activity 0 describes four distinct kinds of "ecological use" which are intended to give qualitatively different kinds of assessments. As mentioned above, this class studies three of the four, leaving out "attainable ecological use." The other three kinds of ecological use will be computed, using data supplied by Nine Mile Creek Watershed District (NMCWD) publications, and by the students in consecutive Activities, in increasing order of complexity of computation:

Activities 1 and 2: Historical Ecological Use
Activity 3: Existing Ecological Use
Activity 4: Potential Ecological Use

For Activity 1, the students are provided with graphs published by the NMCWD of what they will (much later) learn are called "time series." In addition, the graphs are shaded to indicate the regions pertaining to the five Ecological Use classes, and include horizontal lines showing the arithmetic mean of the Existing Ecological Uses plotted. After verifying some rudimentary graph-reading skills, and once again pointing out the ease with which our source material switches between ordering and "measuring," the students are invited to estimate a historical ecological use for one of the reaches with a fairly simple set of data points.

The Activity concludes with a mini-lecture (sometimes done at the beginning of the next class, as students' finishing times on the Activities have a very large standard deviation) focusing on the one reach in whose published time series the horizontal line, showing the calculated Historical Ecological

Use, has been calculated separately for two different subintervals of the overall time period, and graphed as "piecewise constant." In particular, that graph shows a slight improvement in calculated Ecological Use starting around 1990. NMCWD staff informed me that this coincides with some infrastructure improvements made along the creek, capped off by a dam at the downstream end of a large estuary through which the creek flows.

2.1.2 Activity 2

Word pdf

Learning Outcomes:

- Visual estimation of overall trends in time series graph
- Critical thinking: Recognizing that one can choose to emphasize trends in data, or de-emphasize them, depending on one's goals

In this Activity we are still considering the historical ecological use of a reach of Nine Mile Creek. This Activity follows from the "discussion break" at the end of Activity 1. Specifically, in Reach 7C, a wooded portion of the creek downstream from campus just before it empties into the Minnesota River, the historical ecological use is shown to have a slight uptick starting in 1990. This is in Figure EUC-16 in the supplementary materials for Activity 1.

This allowed me to do two things: Invite the students to sharpen their graph-reading skills, and point out how one can choose to emphasize certain features of one's data, or de-emphasize them, depending on the point one wishes to make. To me this kind of discernment is one of the key "meta-outcomes" of a general education quantitative literacy course: produce students who recognize that even when backed by numbers, graphs, statistics, et cetera, it is possible to be deceptive without actually being dishonest. I want critical consumers of quantitative information.

The students are then asked to "eyeball" the changes in Existing Ecological Use trends for a different, nearby reach—Reach 7A, a mile or so upstream from Reach 7C but still downstream from campus. (It so happens that Reach 7A passes right under a local busy road which many students use getting to school each day.) The graph is in Figure EUC-14. The break-point for the two periods is 1990, same as was used by the NMCWD for Reach 7C. It should be quite clear from the time series that the historical ecological use will be lower for the period 1990–2003 than for 1974–1990, as the data points drop visibly lower after 1990.

The students are then asked to "tighten up" their analysis. By saying that the historical ecological use will be lower, they are saying that, if they average the existing ecological uses displayed in Figure EUC-14 from 1990–2003, the resulting number will be *lower* than the corresponding average from 1974–1990. They are then asked to decide, *based on the visual display alone*, whether they think this reduction would be enough to actually decrease the reported ecological use *class* (A-E). This requires the students to try to get a visual sense of where the average value of the data points lies relative to the graph and the shaded areas, which represent use classes. One could go further than the Activity currently does with this line of reasoning: for example, one could ask them to draw a horizontal line where they think the historical ecological use for 1974–1990 ought to be, and similarly for 1990–2003. If one really wanted to get into a more gut-level understanding of "arithmetic mean/average," one could get into the "mass" metaphor for averages and ask them to draw a horizontal line that they think would balance the data points!

As a cautionary note, it's been my experience that about half the class attempts to answer Questions 2 and, especially, 3, by actually computing the historical ecological uses directly. Since they will be doing this in Questions 4 and 5, I try to catch them before they finish and get them to answer purely on the basis of visual information. Sometimes this requires reassuring them that they won't "lose points" for getting it "wrong"—it's important to assure them that they will be graded on their thought process.

Following this visual assessment, the students are invited to actually compute the two historical ecological uses separately. A table is provided for recording their estimate of the existing ecological use as represented by each data point. (Most of the cells in the tables for "date" are blank, because most data points do not have dates reported on the time axis.) The students are still getting used to the conversion between numerical values and the corresponding ecological use classes, so they may need to refer back to the table they made in question 12 of Activity 1.

The Activity concludes with something like a "critical thinking" question, going back to the meta-outcome of quantitative literacy. In particular, since the charge of the NMCWD is to recommend policies that improve waterway quality—and ecological use is the standard they have chosen to use—there is no incentive to present the data in a way that highlights a decrease in ecological use. It is important to realize that there is not actually any dishonesty or distortion here—merely a choice of emphasis or de-emphasis.

2.1.3 Activity 3: Historical Ecological Use

Word pdf

Learning Outcomes:

- Summation (Σ) notation
- Weighted averages and their interpretations
- Sets of data referenced with subscripted indices
- Formulas with multiple variables, some of which are indexed

This Activity has the students compute the second most complex form of ecological use: Existing Ecological Use. Existing Ecological Use is intended to provide a snapshot of the waterway, and its ecological condition, at a particular moment. Overall, ecological use is intended as a measure of the life support capacity of a waterway, and Existing Ecological Use does this quite directly, by actually surveying the fish living in the reach at a particular moment. The reading packet provided prior to Activity 0 describes the process. Each fish species has been assigned, by a reputable agency (the Wisconsin Department of Natural Resources, in this case), an ecological use value (A-E) that reflects that species' sensitivity to environmental degradation: The more sensitive, the lower (alphabetically speaking) its rating.

From a mathematical or quantitative literacy standpoint, these calculations, and especially the notation used (which is taken directly from the NMCWD publication), do stretch some of the students a bit, as the formula for calculating Existing Ecological Use makes use of subscripted indices and summation notation. I have written a handout attempting to illustrate both of these issues, as well as the whys and wherefores of weighted averages. This is given to students at the end of the preceding class.

Class begins, on the day this Activity is done, by briefly describing the way the data are collected, called "electrofishing." As it was described to me when I visited the NMCWD offices, an electrical shock is sent into the water. The resulting stunned fish are gathered, analyzed, and catalogued, and the fish are then released into the creek. The results of several of these surveys are in tables, published by the NMCWD, which are distributed to the students, along with a table showing the assigned ecological use values of the various fish species one might find in Upper Midwest streams and rivers. The students are then left to determine, in consultation with their partner(s) and, possibly, other teams around them, exactly which columns of data in the fish survey records are relevant to the computations they need to do in order to compute Existing Ecological Use. (Much to my surprise I've discovered that a large minority of students lack sufficient proficiency with modern calculators to successfully complete the computations even if they have them set up correctly. An instructor should plan to spend some time with individual teams on calculator troubleshooting.)

2.1.4 Activity 4: Potential Ecological Use

Word pdf

Learning Outcomes:

- More with Σ notation
- Careful organization of summary data
- Careful interpretation of organized summary data

Potential Ecological Use ranking of a waterway is an attempt to assess the current quality of a habitat based on observed environmental conditions in and around the waterway. A large number of factors are observed and assessed, all of which, no matter how qualitative, are ultimately assigned a numerical value for summary and ranking purposes. If one has time, it is a very nice exercise to bring students to the waterway being used as one's case study to actually gather as much of this data for oneself as possible. I have been able to do this when the class meets once a week for three hours, or twice a week for an hour and a half; recently we've been meeting thrice a week, for under an hour, which does not allow sufficient time to meet, observe, record, and get back to campus by the end of class time. Nevertheless, the instructions I gave to the student teams when we did go into the field are included in Appendix B, at the end of this Manual.

Many of the observations to be made, especially most of the qualitative ones, are reported on a "rating sheet," a blank copy of which is given to the students at the end of the preceding class. In class, the students are provided with filled-in rating sheets based on observations made and reported in the NMCWD literature. The numerical values associated with the observations made are combined, in a straightforward way, into a single integer value called the "habitat rating." There is a scale at the bottom of the rating sheet for interpreting this score as "Excellent," "Good," "Fair," or "Poor."

Once a Habitat Rating is assigned, this is included with other, strictly numerical (measured) data: average pH, maximum temperature, and minimum water flow/discharge. All of these are what the literature calls *limiting factors* in determining the Potential Ecological Use of a stream reach. The idea is that each of these factors can limit the kinds of organisms that can survive in the reach, regardless of how good other conditions are, so that the reported Potential Ecological Use will be determined by the *worst* of these factors, as recorded in Table EUC-4. For example, if all other factors point to Class B, but the water is too warm for any but Class C organisms, then the reported potential ecological use must be Class C.

The ranking sheet, Table EUC-4, can be confusing for some students, so be prepared to help some teams navigate through it the first time through.

Activity 4 concludes the introductory "Unit 0." This is a brief introduction to the use of numbers and quantitative thinking in environmental science and assessment of environmental quality.

2.2 UNIT 1 (Activities 5 – 11) Introduction to Data Analysis and Mathematical Modeling

This unit introduces the students to the concept of one particular kind of mathematical model: Namely, an equation representing a relationship between two variables, and the graph of that equation. After turning in their completed Activity 4, the students are given a one-page reading, which consists of small excerpts cut-and-pasted from a Wikipedia entry titled "mathematical modeling." The excerpts highlight those aspects of modeling that will be explored in the coming Activities: namely, scope, accuracy, and fit.

2.2.1 Activity 5: Introduction to Modeling from Paired Data

Word pdf

Learning Outcomes:

- Introduction to sets of paired data
- Deciphering and computing with complex formulas which include Σ-notation
- Generating a least-squares linear equation from paired data
- Data ranges and introduction to interpolation and extrapolation

The students are told to read the following brief introduction to regression lines, and work through the interactive practice problem: `http://www.zweigmedia.com/tuts/tutRegressionb.html?lang=en`.

The reading once again includes Σ notation. This helps to reinforce to the students that recognizing Σ notation, and understanding what it tells the reader to do, is a basic piece of mathematical literacy. Activity 5 has the students generate a least-squares line, using a pre-made template in Excel, which mimics the online tutorial linked above.

In Activity 5, we look at our first bivariate set of data: a collection of surface elevations (in feet above mean sea level) and estimated volumes (in millions of cubic meters) of Bush Lake, a popular fishing and swimming lake just northwest of the Normandale campus. Similar data should be available for any decently sized lake in the US or Canada. For the United States, the Environmental Protection Agency has a watershed clearinghouse for all fifty states here: `https://cfpub.epa.gov/surf/locate/index.cfm`.

The actual document from NMCWD from which the data were sourced reports volumes in cubic meters, but I converted to millions of m^3 so that the constants in the regression line do not come out in scientific notation when done in a calculator or on Excel.

It is important that the students understand that producing a model such as the one in this Activity, or in the online tutorial, is not just an academic exercise, but can help them to indirectly measure variables which are challenging, expensive, or both, to measure directly. In particular, measuring the surface elevation of a lake is very simple, but, due to the irregularity of a lake bed, measuring a volume *accurately* is complicated and requires a large number of observations and computations. I begin the class by briefly describing the process of estimating a volume by what a mathematician would call a three-dimensional Riemann sum, although of course I don't call it that in class. I do mention the intuitively obvious fact that the thinner one's columns of water are, i.e. the more individual depth observations one makes, the more accurate one's estimate will be—but of course at the cost of more work. Thus if we could find a relatively simple equation for computing volume from a *single* elevation observation, *and* if we had good reason to believe that the equation *accurately* predicted volumes, that would simplify volume estimation considerably!

Assuming that the students have done the online reading and interactive tutorial, they have already seen the rather complicated formulas for computing the slope and vertical intercept of a least-squares line. (By the way, I never use the expression "least squares" in class, nor do I spend any time explaining exactly why it works. All I tell them is that, of all the lines we could use to model the data, this one is the most accurate one there is. A little oversimplification, but close enough for our purposes.) The Excel spreadsheet that accompanies this Activity is pre-loaded with commands to automatically sum the relevant values, so students don't get *too* bogged down in the specifics of the spreadsheet software; at the same time, they will be using Excel extensively as an analytical tool throughout the rest of the semester, so this Activity gives them a chance to start getting comfortable with some of its features.

Most of the rest of the Activity involves the students producing, with the help of the spreadsheet software, the various terms and factors in the least-squares line formulas, and putting them together correctly. As was the case earlier, there is a lot of variability in these students' comfort level and facility with hand held calculators. Many of them do not seem to have used a scientific calculator before, and it increasingly seems like some of them have never used a calculator other than the one in their cell phone! So an instructor should expect to have to help out with getting all of the various pieces entered correctly in the students' calculators. Alternatively, students can use the computer's built-in calculator, in "scientific" mode. At the end of the Activity, they will have a linear model for predicting volumes of Bush Lake, in millions of cubic meters, from the surface elevation, in feet above mean sea level.

A word about units:

1. It is essential that students develop the habit of reporting the appropriate units of measure-

ment, every time, right away. The units of the two variables in this current model are less than completely straightforward, especially those for lake surface elevation, so it would be wise to spend some time carefully explaining exactly what those numbers are measuring.

2. A sizable minority of the students I've worked with seem completely baffled by units in millions of cubic meters. It would probably be wise to give an additional worksheet for practice converting back and forth between various ways of expressing large numbers (e.g. on the order of 10^3 or higher, so that it would be common to express the quantity with words and numbers together, such as "2.6 million." I do try to avoid scientific notation in the course, and have mostly succeeded. Perhaps this is a bad idea in a "quantitative literacy" class?). This difficulty and confusion will only increase when students start doing calculations that result in quantities such as 0.024. These students will need guidance to help them understand why "0.024 million" is the same number as "24,000."

A "postscript" for Activity 5: Later on in the course the students will be shown how to use built-in Excel commands to compute least-squares equations without explicitly using the formulas used here. It is important, from my view of a liberal arts education and of quantitative literacy, that students come away from the course with at least a notion that these computations come from a specific procedure having to do with the raw bivariate data and its relationship as expressed through a particular sequence of mathematical operations. More colloquially, I want to give them a look "under the hood" to see what makes the software kick out the equations it does.

2.2.2 Activity 6: Graphical Representation of Bivariate Data and the Relationship between a Scatterplot and a Least-Squares Regression Line

Word pdf

Learning Outcomes:

- Generate a scatterplot with spreadsheet software
- Interpret the scatterplot by relating points on the graph to pairs in the data list
- Read and identify trends in a scatterplot and interpret them in terms of the variables being observed
- Evaluate and interpret model equations

This Activity produces the first of many (!) scatterplots the students will make using spreadsheet software. The scatterplot will be made from data in this spreadsheet. This scatterplot, with lake surface elevation on the horizontal axis and estimated lake volume on the vertical, has the very nice property that the scatterplot forms an almost perfect straight line. Another skill the students will need throughout much of the rest of the course is to connect bivariate data in a table with its graphical display in a scatterplot, so that connection is made as soon as they have made their first scatterplot. (Note that they must carefully distinguish among the data points in the upper right corner of the plot for this particular data set. Any adaptation of this Activity from local data should do something similar.)

As mentioned, this Activity shows them how to have Excel produce the least-squares linear equation. Excel calls these "trendlines," and this terminology is used freely throughout the course. The Activity concludes with another "critical thinking" question, which addresses the surprise or concern that some students express when their intercept (b) in Activity 5 comes out as a negative number. There is an interpretation of this phenomenon which will be explored in a future Activity, in particular when thinking about the *scope* of our model, but for now I just ask the students to think in practical terms: Since x represents the surface elevation of Bush Lake in feet above sea level, $x = 0$ would mean that the surface of the lake is *at* sea level, over 800 feet below the surrounding land. That, in turn, would mean that Bush Lake is ordinarily over 800 feet deep, and no lake in Minnesota (other than Superior) is anywhere near that deep.

2.2.3 Activity 7: Further Analysis of a Linear Model from Data

Word pdf

Learning Outcomes"

- Modeling vocabulary: Determine whether a prediction is interpolated or extrapolated
- Unreasonable predictions; scope of a model
- Determining accuracy of a model's prediction: Error and relative error

In this Activity we deepen or extend our analysis of mathematical models constructed from paired data. The students are reminded of two key questions from the handout on "mathematical modeling:"

- How well does the model fit or match existing data? (the question of "fit" or "accuracy")
- Can the model be applied to inputs outside the given data range? If not, what kinds of things can go wrong? (the question of "scope")

To ensure that students understand, at minimum, when their predictions are instances of interpolation or extrapolation, they are asked to identify boundaries for the given input, ("x") values. This requires, among other things, that they understand the distinct roles of the input and output variables, as they have been designated for the current model.

Some confusion still arises at this point due to the fact that the least-squares line predicts volumes in *millions* of cubic meters. Some students are likely to need rather extensive guidance and attention before they can successfully translate these predicted volumes ("y values") into cubic meters.

After a couple of computations, the students are given a lake elevation, which leads to a *negative* volume prediction, to which they are invited to object. After mentioning how this distressing phenomenon ties into the idea of "scope," we then focus on what it might mean for our model to "fit" the *known*, i.e. *given* or observed, data pairs.

The rest of the Activity introduces the students to the notion of *interpolation errors* ("residuals" if you're a statistician; I do not use this term at any point in the course, but one certainly could). I give a brief lecture to motivate the computation of "relative error" as a more meaningful measure of "accuracy" by asking whether an accounting error of $10,000,000 is significant or not. Most of them say yes, but I tell them "Not if the background is the Federal budget" and conclude with the point that one wants to compare the error with the *scale* on which your variable is operating. For example, we find some prediction errors in the neighborhood of 25,000 cubic meters, which is a lot of water—but since we're dealing with a lake having over two *million* cubic meters of water in it, the error is not large at all. We thus characterize the "fit" or "accuracy" of a model in terms of the sizes of its relative errors when interpolating. I will later call this the "local" view of "fit," since it refers explicitly to individual points. This is in contrast to what I will call the "global" aspect of "fit," which has to do with capturing any overall trend such as increase/decrease or curvature.

2.2.4 Activity 8: Analyzing Problems with Extrapolation and Seeking to Improve Our Model

Word pdf

Learning Outcomes:

- Developing an intuitive feel for the scope of a model
- Carefully interpreting a model's predictions; identifying boundaries of its scope

The students begin by reviewing a key question about the least-squares linear model for predicting the volume of Bush Lake from its observed surface elevation; namely, they are reminded that the model *does* seem to accurately represent how the volume of Bush Lake is related to its surface elevation, *within* the range of actually observed lake surface elevations. So in that sense, we seem to have a very good model.

And yet—there are flaws that arise when extrapolating, and in fact we don't have to extrapolate very far to find one of them. This Activity is probably the most *conceptually* challenging so far, even though most of the computations aren't difficult. This is because we are trying to see what we are *actually* being told when, beyond a certain point, a model gives output that is clearly nonsense, such as predicting a negative volume of water.

To prepare students to conceptualize and contextualize this analysis, I begin class with a brief lecture and presentation at the whiteboard discussing what the two variables in the equation actually mean. I draw a schematic cross-section of a lake and draw a vertical line segment representing its depth, and also indicate visually the meaning of the variable x, via a line extending all the way down to sea level. I then point out that the depth of a lake can be thought of as the distance from the surface of the water to the lowest point on the lake bed, and thus can be thought of as the *difference* between the elevation at the surface and the elevation at the lowest point on the lake bed. Starting this way tends to reduce the number of students who are confused about how to compute the various differences that arise in the Activity.

Question 4 tends to generate a lot of questions, although it's been my experience that most students, when pressed, can intuit that the only sensible answer is "zero," but they "can't believe it's that easy." The follow-up question also tends to generate questions. I always have several students or teams who need quite a bit of guidance to get from "the volume is 0" to "the lake is empty or dry."

Question 8 is the first of only two times all semester when I have them actually solve an algebraic equation. How many teams can successfully solve it varies widely from semester to semester. I solve it at the board at the end of the class period, when the better mathematically prepared teams have already turned in their completed Activity and left. Then after a little more arithmetic we end up with the *predicted maximum depth* of Bush Lake, based on the least-squares line from the given data. According to documents from the NMCWD, this prediction is much too shallow. So in attempting to fix one flaw in the model and reduce the scope so that negative volumes are precluded, we encounter another limitation in the model's extrapolation power. We then proceed to look for a "fix" for this flaw, that the depth of the lake is inaccurately predicted.

The remainder of the Activity is devoted to recalibrating the model so that the independent ("x") variable is now the maximum depth of Bush Lake (in feet) rather than its surface elevation. The domain issue is clearer here, since clearly no lake can be fewer than zero feet deep! (I do not actually use the terms "domain" or "range" in class.) The analysis concludes by observing that the negative-volume issue has not gone away; that is, the new model predicts a negative volume for some positive lake depths. We now abandon the attempt to extend our linear model outside the range of given lake depths, and in Activity 9 will seek a more sophisticated model.

2.2.5 Activity 9: A Power-Function Model; Log-Log Data Transformations

Word pdf

Learning Outcomes:

- Length-to-volume relationships and power functions
- Log-Log plots and their relationship to power function models
- Logarithm base ten as a measure of the "size" ("order of magnitude") of a number
- Reinforcing the practical interpretation of variables in a modeling equation

We introduce the process of generating a power function model from linear least-squares regression on a log-log graph of a set of bivariate data. Usually most students will conjecture fairly quickly that a model for predicting a volume from a depth or length should have an exponent of three. Later they will see that the model produced from our depth-volume data very closely matches this conjecture!

The students are provided with an Excel data file (LakeLevelData.xls; see Activity 6), which already has the values of log(depth) and log(volume)—all logarithms are base 10—but these data are in a separate sheet, which they probably didn't know was there. By now the students are getting adept at generating least-squares regression lines and their equations using Excel's built-in commands. The students do *not* have to use any properties, themselves, to translate from the log-log linear equation to the power function model for volume vs. depth. The Activity provides them with a formula for this transformation. (The minimum prerequisite for this course at Normandale does not involve more than a very cursory glance at logarithms.) Once they have the power function model, they compute relative interpolation errors and compare them with the corresponding errors from the linear model. The errors are *considerably* smaller, often less than $\frac{1}{2}$ of 1%. The Activity concludes by checking that our new model will, at least, no longer predict negative volumes from nonnegative depths.

2.2.6 Activity 10: More Work with Power Function Models

Word pdf

Learning Outcomes:

- Compare and contrast linear and power function models
- Reinforce interpolation/extrapolation distinction

At the end of the class before this Activity is given, the students receive a reading packet from a NMCWD publication. This packet introduces the notion of "species richness," a measure of biodiversity of a patch of habitat. I chose to include this topic, even though the data are not from a local source, because the relationship between habitat area and species richness is an ecological relationship that seems to be well modeled by a power function relationship, as indicated in the reading packet.

The variables are: x = area of a given piece of a specific type of habitat, in hectares y = number of distinct species of a specific type of organism found living in that habitat.

The reading packet includes a bivariate set of data for tallgrass prairie and species of prairie songbird, and these data are in a spreadsheet provided to the students. The model measures how biodiversity depends on single, large, unbroken patches of habitat, and therefore emphasizes the negative impact habitat fragmentation has on biodiversity.

From the data in the spreadsheet, the students generate a scatterplot with the least-squares line. At this point I begin insisting that all graphs come with labeled axes, *including units of measurement*. I consider this part of basic quantitative literacy, so that readers of their work, including supporting graphs, know exactly what they are looking at, and on what scale.

This graph is the first (of many) examples of a bivariate data set for which the least-squares line is visibly *not* a particularly accurate model. After observing this, the students plot the log-log data, generate the least-squares line, and from that produce a power function model for the relationship. They do not yet have the tools to generate the graph of this model, so I do not yet ask them for a *visual* assessment of the model's fit. In the next Activity they will compute relative errors for both the linear and power function models, and compare the (local) fits of the two models.

2.2.7 Activity 11: Comparing Accuracy of Linear vs. Power Function Species Richness Models

Word pdf

Learning Outcome:

- Relative error as a measure of model accuracy for interpolated values

Using the two least-squares models, linear and power function, for the species richness vs. habitat area relationship, as developed in the previous Activity, the students analyze each model's accuracy when predicting species richness for area values that are included in the data set we have. The students have seen error analysis before; this is their second time through the process, so many of them may be quick about it. Make sure they are getting *negative* residuals when they should, and make sure they know what a negative (or positive) residual says about the relationship between predicted and observed species richness.

END of "Unit 1." I usually have the first exam cover Activities 0 through 11. The format is: One-third true/false questions on basic concepts; one-third multiple choice; and one-third free-response. The free-response questions have the students use Excel with some different data sets than were used in the Activities. The exams are open "book"—students may use their corrected Activities, and the readings, during the test.

Starting with Activity 12, the students start exploring less orderly data sets and working with correlation coefficients as numerical measures of "fit."

2.3 UNIT 2 (Activities 12 – 22)

Detailed analysis of bivariate data: correlation coefficient as goodness-of-fit; slope as rate of change; positive vs. negative correlation; linear vs. curved relationship; effects of outliers on perceived relationship; interpretation of all of this in the context of various measures of water quality, especially when the data are not a particularly good fit for *any* least-squares regression models.

2.3.1 Activities 12 and 13: Two Complementary Notions of Model Fit: "Local" vs. "Global"

Activity 12 Word pdf

Activity 13 Word pdf

Learning Outcomes:

- Using spreadsheet technology to compute a correlation coefficient
- Interpreting a correlation coefficient in terms of strength of a relationship as shown in a bivariate data set
- Assessing fit of a regression model two ways: Intuitive visual assessment; use of correlation coefficient

Activities 12 and 13 are closely related, and can almost be thought of as a single Activity in two parts. Before Activity 12 is given in class, the students are asked to read the following brief primer on correlation coefficients: `http://ecp.acponline.org/mayjun01/primer_correlcoeff.htm`.

I begin the class in which Activity 12 is done with a discussion of what the students have seen in the regression modeling they've done so far; in particular, we'll be focused only on interpolation and model accuracy, or "fit," from now on. I explain what I mean by a "local" criterion for a model being a good fit to bivariate data verbally—nothing written on the board, as I want them to listen and absorb what is said, not simply transcribe. I want them to listen carefully and report back, in their own words, what they think they heard, and *why* this interpretation of fit is called "local." The idea focuses on, to use statistical notation (which is not used in the course, although it could be—additional notation, in my experience, tends to get students at this level focused on getting the symbols "right" at the expense of actually understanding what they mean), how well $\hat{y}_1$ matches y_1 for each individual data point—in graphical terms, how close the trendline comes to individual ("local") data points. In other words, the local criterion for the fit of a model is a reflection of how closely the model's *predicted* output values match the *actual* values observed.

The "global" notion of fit then refers to how well the model seems to capture the behavior of the data set observed as a whole; things like:

- Is there an overall increasing or decreasing trend in y-values as x increases?
- Do the data points form a straight or curved pattern?
- Does the model reflect these observations about the data?

I don't yet tell them that these two criteria, "local" and "global" fit, are in some tension with each other, although this tension will be explored in a future Activity (#22, to be exact), after they have gotten a lot more comfortable with the correlation coefficient and its interpretations. In that future Activity, the students are shown a data set and model, which has an extremely *good* local fit but extremely *poor* global fit; after all, one can eliminate interpolation error by simply choosing a sufficiently high-degree polynomial, but then you get wild gyrations that have no relationship to any pattern in the data.

Speaking of correlation coefficients, it is in these two Activities that the concept is formally introduced. The Excel commands they've been using automatically print r^2 to the screen, but I've been telling them to ignore it until later in the course—well, that time has come. There is a new spreadsheet to go with Activity 12.

The first thing I point out is that the scatterplot and trendline package we've been using in Excel only gives r^2 , not r itself (and Excel uses a capital R for some reason), so they will have to find r for themselves. Earlier on in the course, students were invited to give their own judgement of the quality of the fit of the two least-squares linear models they've looked at so far, based purely on a visual inspection. They are now invited to compare their own intuitive judgements with the computed correlation coefficients.

The species richness-habitat area data, which was visibly nonlinear, turns out to have $r = 0.81$ for the linear model; the students tend to be all over the map as to whether this reflects a "good" fit or not, i.e. whether 0.81 is "close" to 1 or not. One reason for separating the two Activities is that Activity 13 introduces the standard statistics textbook criterion for a "good" fit, namely $r > 0.8$.

Activity 12 also introduces the students to the capacity of Excel to generate nonlinear regression models "directly." I "hid" this from them at first so they could at least get a sense of how some mathematics they've probably been exposed to at some point (namely, properties of logarithms) is an essential part of generating these models.

Lest the students get too focused on just pushing buttons and reporting numerical output, I insist on having them continue to provide their own analysis of the fit of a model to a scatterplot. I believe keeping the concepts always in their line of sight is essential. It is very tempting for students to simply resort to pushing buttons and reciting output, at the expense of actually *understanding* why they are doing what they are doing, and what the number they've written down is actually telling them.

Finally, towards the end of Activity 13, the students are given the standard ranges for $|r|$ that one finds in all the standard statistics books which correspond to "good," "fair," and "poor" fits. They also get their first real taste of the tension between local (as measured by r) and global fit, as the least-squares line rather blatantly fails to capture the global curvature of the species richness-habitat area data, even though its correlation coefficient puts it in the "good fit" range (barely).

2.3.2 Activity 14: Further Topics in Modeling from Data

Word pdf

Learning Outcomes:

- Slope of a line as rate of change
- Rate of change as a "multiplier" for predicting changes in "y"
- Introduction to positive and negative correlation
- Introduction to analysis of more irregular data

Activity 14 introduces a couple of new phenomena. First is the interpretation of the slope of a regression line as a *rate of change*. This idea will be returned to several times as the students look at different pairs of variables. (The various examples involved could be combined into a single module for illustrating this concept at the remedial algebra, college algebra, or even differential calculus level—the latter prior to introducing the derivative, when reviewing the slope-as-rate-of-change interpretation.)

Many, perhaps most, of the students at this level struggle with the concept of slope-as-multiplier. Conceptually this is probably understandable, as they have been trained to see any "formula" as something to "plug numbers into" without any consideration of what the various pieces of the formula might actually signify (if anything). I have recently begun to preface this Activity with a ten-to-20-minute lecture on slope as rate of change, using two examples commonly used in beginning algebra classes:

- Distance vs. time (at constant speed): the slope of the linear equation giving distance as a function of time is the rate at which distance is covered, i.e. the speed of the object; I illustrate with the question, "How much ground do you cover in 1 hour? 2 hours? 4.5 hours?"
- Cost of an electric bill as a function of kilowatt-hours used: the slope of the linear model gives the per-kwh charge; I illustrate with the question, "How much will next month's bill differ from this month's, if you use: One more kwh next month than this? 15 more kwh? 18 *fewer* kwh?"

Since I have been opening with these examples, the number of students incorrectly answering questions by simply evaluating the linear model seems to be reduced, although a disappointing number of students just want a rule: "When do I use the 'b' and when don't I?" Sigh?

Also, some students still struggle with converting the model output, which is in millions of cubic meters of water, into cubic meters.

The second half of this Activity begins a lengthy "story arc" within the course—probably the longest one, lasting until Activity 25. This deals with the following parallel themes:

i) It is often possible to make qualitative conjectures about a relationship between two variables, based either on one's own intuitive sense of the situation, or based on information provided by some source which one has reason to believe is authoritative or trustworthy;

ii) Data pairs consisting of actual observations of these variables often have weak, if any, patterns or correlations, and there is a myriad of tools or approaches one can take in attempting to make sense out of raw data. Furthermore, even if there *is* an identifiable pattern in the data, that pattern is not always consistent with what one might have predicted based on "common sense" or on readings from an authoritative source.

Eventually, the particular data I obtained from the NMCWD led to the fortuitous situation that I was able to have the students argue for seemingly contradictory assertions—but this comes later on. One of the things I have tried to do is to lift the students, in the next several Activities, through layers of Bloom's Taxonomy, until they are ready for at least rudimentary Synthesis. For a brief description of Bloom's Taxonomy: `http://www.edpsycinteractive.org/topics/cognition/bloom.html`.

The students begin with what I intend to be almost as much of a "no-brainer" as their first example—that when lake surface elevation increases, lake volume should also increase—namely, would they expect the observed surface elevation of a lake to increase, decrease, or stay the same when rainfall increases?

Again, this Activity uses data which will be readily available anywhere there's a lake. The scatterplot obtained from that data will, of course, vary widely, depending on local hydrological conditions. Since the example given here shows essentially no correlation, we can conclude that factors other than recent rainfall affect the volume of Bush Lake.

This is also the first time the students encounter a relationship that is assessed as negative by the least-squares line. Since Excel, at least in scatterplot mode, only gives r^2, the students need to have it emphasized that the *sign* of the correlation coefficient, r, is determined by whether the relationship between variables is positively or negatively correlated—or, as many of them like to say, they need to put a minus sign in front of their r value when the line goes down.

If your data show a stronger relationship, or a positive correlation, I don't think the questions should need to change, although of course the correct answers will!

2.3.3 Activity 15: Predicting the Nature of Relationships before Looking at Data

Word pdf

Learning Outcome:

- Interpret a verbal description of a relationship between two variables in quantitative terms, especially positive vs. negative correlation

At the end of the class before Activity 15 is handed out, the students are given another reading packet excerpted from NMCWD literature: "Bush Lake Use Attainability Analysis" (U.A.A.). This packet forms the basis for the data the students will be analyzing for several weeks, having to do with variables that are relevant to the environmental health of freshwater lakes. The class as taught at Normandale focuses on data from Bush Lake, near campus, but data on these variables should be available for any sizeable freshwater lake in the U.S. or Canada (as long as the lake is not too remote from population centers). The students are asked to at least skim the entire packet, but three or four sections are marked off for particular attention. Those sub-excerpts will be important for some specific questions the students will be exploring in the next several Activities. In particular, Activity 15 asks the students to make conjectures about whether various pairs of variables will have positive or negative relationships, based on information found in the reading. There is no actual calculation of any kind done in this Activity!

Similar documents should be available regarding the ecological status of local lakes from local, county, or state agencies. I acquired the document I hand out for free because it was a couple of years old; since it's the mathematical analysis I was interested in rather than the latest condition of Bush Lake, that was OK by me.

This Activity begins our study, which culminates in Activity 25, of the question, "Is excess phosphorus in the Bush Lake watershed a cause for concern?" The analysis will follow a careful study of some of the variables introduced in this Activity, with some detours to explore interesting mathematics along the way.

Question 4 in this Activity gives rise to more raised hands than most. Basically the students are being asked to figure out that the composition of two increasing functions will be an increasing function, although we never talk about "functions," let alone "composition," explicitly.

Question 6 also leads to some questions; it's worth lingering here, as it plays directly into what I call our "research question" stated above. The point here is that if we see evidence of a positive correlation between rainfall and phosphorus concentration, we'll take that as evidence that there is excess phosphorus in the surrounding watershed that is being washed into Bush Lake when it rains—which will, in turn, be interpreted as a "cause for concern."

2.3.4 Activity 16: Positive/Negative Correlation vs. "Strict Monotonicity"

Word pdf

Learning Outcomes:

- Core concepts behind positive and negative relationships between variables
- Distinction between positive correlation and increasing function, and parallel for negative correlation

The concept of positive and negative correlation has been explored in the last couple of Activities. However, when the students start looking at actual data in tabular or graphical form, and asked if they think they are looking at a positive or a negative relationship, it is important that they know what this *doesn't* mean. The purpose of this Activity is to point out that if one looks too closely at the changes between individual data points, one may not have much luck in discerning the nature of the relationship. Also, I want them to understand that positive and negative correlations are not the same thing as monotonicity (a word which I don't actually use, even in adjectival form, in the course); that is, not *every* point with a larger x-value must have a larger y-value in order for a correlation to be positive. Put another way, although it's not explicitly stated in the Activities—perhaps it should be—correlation is a *global*, not a local, feature of a bivariate data set. That is, it's determined by looking at the data in aggregate, as a cohesive whole, not at the behavior of individual pairs of pairs, in isolation from their larger context. This Activity uses a new data set, involving environmental factors such as rainfall and phosphorus concentration.

2.3.5 Activity 17: Some Data Sets with Less Well-Behaved Scatterplots

Part I: Word pdf

Part II: Word pdf

Learning Outcomes:

- Mostly reinforcement of earlier ideas: correlation, strength of relationship, linear or curved relationship

This Activity is in two parts. At this point it's worth pointing out that some of the Activities do not involve working directly with data sets in a spreadsheet. This was originally necessary because of a semester in which I was assigned a computer classroom for only one of our two weekly meetings, but a side benefit of that schedule has been that students must focus on the mathematical content of the data and its interpretation, without spending time distracted by Excel commands, which are, after all, only a tool and not part of the core course content.

The students are given a scatterplot, which is from data they analyzed at the "local" level in Activity 16. This is by far the most "scattered" scatterplot they have seen so far. The students are given an open-ended question about any trends or patterns they see. I tell them that the maximum possible score for this question will be determined by the team(s) that makes the greatest number of *valid and relevant* observations. Most semesters the question ends up being worth either 3 or 4 points. Some observations students have made include: There are two parallel groups of data points, each of which shows a downward or negative trend; Looking at the trend that goes from the lower group of four points to the upper group of two, one could see a positive trend; There are two points having the same phosphorus content but having different chlorophyll contents; there is no clearly identifiable overall trend.

To continue emphasizing the connection between data presented in a table, and the same data represented in a scatterplot, the students are asked to identify one particular data point, with reference to a variable (month of observation), which is in the table but not represented on the graph.

A note here: These are variables which will be observed and reported by any agency responsible for monitoring and reporting on the ecological status of a body of water, and the scatterplots obtained from other data are not likely to have a high degree of structure, so the questions in this Activity should be adjustable to data which is local to any institution that is near a substantial body of fresh water.

The least-squares line for the scatterplot is then provided. Students are asked to classify the relationship as positively or negatively correlated, which is trivial once the trendline is visible, but I *insist* that they explain their answer in terms of what the positive slope says about the relationship between the two particular variables being examined: When phosphorus concentration rises, chlorophyll concentration also tends to rise.

The Activity is divided into two parts. This is because Part II includes the coefficient of determination for the least-squares line, and I want the students to make their best guess about the strength of the relationship by visually examining the scatterplot and trendline. Therefore, they do not receive Part II until they have turned in their completed Part I. This is to emphasize the *meaning* of "strength of fit." Once the students have numerical criteria, they are tempted to just look at the r value and stop thinking.

In Part II, the students begin by seeing how well their own intuition or subjective judgement matches up with the cold, hard analysis of the correlation coefficient. We then take our first detailed look at a negatively correlated pair of variables: chlorophyll concentration and Secchi disk transparency. (Such data should be available from any watershed-district type agency, and should usually show a clear, negative correlation, so these questions should work just fine with local data.)

The students must articulate their interpretation of the negative correlation in terms of the variables: If the chlorophyll concentration increases, the Secchi disk transparency will tend to decrease. The correlation coefficient for the straight line least-squares model shows a (barely) strong linear relationship; some students will need to be reminded to report a negative number for r. Finally, the students are invited to notice a strong curvature to the data pattern. The curve they draw as capturing the overall trend will inform their choice of a nonlinear model for this relationship in a future Activity.

2.3.6 Activity 18: Further Analysis of Data Relationships

Word pdf

Learning Outcomes: Plotting points and identifying coordinates in a less-straightforward setting—The students now begin to look in detail at variables relevant to their "research question:" "Does data suggest that excess phosphorus in the Bush Lake watershed is a cause for concern?" The three variables to be explored in detail are:

- Monthly rainfall (inches)
- Phosphorus concentration (micrograms/liter; readings in our data set are from the *beginning* of each month, a fact we'll use later)
- Algae concentration (units/milliliter, also observed at the beginning of each month)

The reading from the Bush Lake U.A.A., handed out earlier, becomes relevant again here; the data from the table given earlier are now available in an Excel spreadsheet. The reading discussed some of the potential environmental threats faced by any body of surface water. I discussed this briefly in class when the reading was first given and now remind them of some major points:

Phosphorus loading in Bush Lake is, according to the U.A.A., mostly "external;" that is, most phosphorus added to the water comes from outside rather than inside the lake. (Internal loading, again according to the U.A.A., is mostly from decaying organic matter and is naturally occurring, while external loading can come from artificial fertilizer used by local property owners.) Since Bush Lake is in a highly developed suburban area, there is a lot of fertilizing going on in the surrounding watershed. Among other things, that means phosphorus and nitrogen. So we are going to look at our data to see if we have evidence of phosphorus washing into Bush Lake with rainfall. If so, we will take this as a cause for concern about excess phosphorus in its watershed.

One important aspect of extracting information from raw data is to present the data in a meaningful way. One of the basic quantitative literacy goals of the next couple of Activities is to get a sense of what a reasonable, informative comparison would be. For example, in this Activity our question is, "Do the data suggest that rainwater washes excess phosphorus into Bush Lake?" Since, say, the June phosphorus concentration is from the *beginning* of June, comparing June rainfall (which covers the entire month) with June phosphorus concentration doesn't tell us anything about whether rainfall in June contributed to the Bush Lake phosphorus load. Therefore, we compare June rainfall with July phosphorus concentration. We then have to be careful in aligning the data cells when setting Excel up to generate the appropriate scatterplot.

The first few questions are a check to see if the students have a rudimentary understanding of what the points in the scatterplot represent in terms of rainfall and phosphorus concentrations. The students then create a scatterplot showing the appropriately paired data. I have told them, and will continue to tell them, that labeling axes on all graphs, *including* units of measurement, is part of basic quantitative literacy and is required on all graphs, whether explicitly instructed to do so or not. The reader needs to know exactly what he/she is looking at, and on what scale.

Once the scatterplot is formed, the students answer some now-routine questions about the nature of the relationship. Some questions could be, and probably will be in future versions of the course, added about interpreting the slope as rate-of-change; this would allow us to explore questions about the *strength* of the relationship, which could be high—although in this case the strength is only moderate—and the *significance* of that relationship, as reflected in the slope of the regression line. A very strong positive relationship, but with a very small slope, could reasonably be interpreted as little cause for concern. But that's for the next "edition!"

2.3.7 Activity 19: Further Analysis of Variables Related to Water Quality

Word pdf

Learning Outcomes:

- No new topics; reinforcement of previous ideas with new variables

The students look at some other data points, looking for evidence—or lack thereof—for environmental threats against Bush Lake. Last time we compared rainfall to phosphorus concentration—the idea being that, as a plant growth accelerant, too much phosphorus can lead to a lake choked with weeds and covered with algae. This time we compare:

- Monthly rainfall with nitrogen concentration;
- Phosphorus concentration with algae concentration.

The first pairing is very much the same analysis we did with rainfall and phosphorus concentration in Activity 18. The main difference is that in this case there is almost *no* relationship. (After having taught the course a couple of times I was told by a chemistry teacher that nitrogen has almost no water solubility, so it is no surprise, in light of this, that rainfall doesn't seem to increase nitrogen loading.)

The next relationship, phosphorus concentration vs. algae concentration, will be our other main focus during this "story arc" (which has its *denouement* in Activity 25). The motivation for looking at this relationship is to check our conjecture, or "educated guess," based on the NMCWD literature, that these variables are positively correlated due to phosphorus being an essential nutrient for algae. If so, and if we are concerned about the potential for algae blooms as a result of too much phosphorus entering Bush Lake, then we will consider such a relationship to be "cause for concern."

Now I did *not* have the students "offset" the monthly data this time, which means that we are comparing phosphorus concentration at the beginning of June with algae concentration at the same time, before the algae population has had time to absorb and metabolize the phosphorus. Most likely in future "editions" of the course I will add such analysis to this Activity. At any rate, later on we will alter our approach to try to see "cause-and-effect" more directly.

The scatterplot that is generated from our data is far from clearly positively correlated. No reasonable conclusions can be drawn from the scatterplot or least-squares line. However, looking at the trendline shows the effect of the rather blatant outlier in the upper right of the graph. This motivates what may appear to be a tangential journey—exploring the effects of outliers on the perceived strength and nature of data relationships—in the next couple of activities. However, this is in the spirit in which I wanted to design this course—have the data drive the mathematics we study rather than the other way 'round.

Parts of a couple of upcoming Activities, in which outliers are examined, could also serve as a module on this topic in an introductory statistics course.

2.3.8 Activity 20: Looking More Closely at Data Relationships

Word pdf

Learning Outcomes:

- Identifying data outliers
- Exploring effects of outliers on perceived relationships between variables
- Changing perspective on data to seek cause-and-effect relationships

Before coming to the class in which Activity 20 is done, the students are asked to read the following brief Web-based primer on outliers in scatterplots: `http://ecp.acponline.org/mayjun01/`

primer_correlcoeff.htm starting at the section "Does the Coefficient Reflect a General Relationship or an Outlier?" and looking at Figure 3 for examples of various kinds of outliers.

The scatterplot of phosphorus concentration vs. algae concentration from Activity 19 has an obvious outlier. This Activity begins an exploration of outliers in paired data sets and the various effects they can have on the relationship between variables, as expressed in the equations, parameters, and correlation coefficients when doing least-squares regression analysis. The students will encounter outliers that:

- Make an otherwise negative relationship appear positive
- Make an otherwise strong relationship appear weak
- Make an otherwise weak relationship appear strong
- Have little effect on perceived strength, but change the regression equation in a way that has implications for predictions or interpretations based on that model
- Have negligible effect on a regression equation model for the data, either in strength or in equation.

It will be pointed out when the time comes that this last phenomenon, if $|r|$ is high, suggests that a very good model has been found for the relationship between variables under consideration.

If we delete the outlier from the phosphorus concentration—algae concentration data, we get a moderate, negative relationship. Since this is *not* what the students had been led to expect—phosphorus is, after all, a nutrient for algae and other plant life—we shift our focus to try to more directly see if any cause-and-effect is visible in the data. The idea in what follows is that elevated phosphorus concentration, if our initial conjecture was correct, ought to cause algae concentration to increase more rapidly than it otherwise would.

What we want to explore now is whether elevated concentrations of phosphorus really do have this growth-stimulant effect. "Growth" in *any* quantity is a type of change, and "change" means "difference." Therefore, what the students do next is pair data comparing phosphorus concentration at the beginning of a given month ("x") with the *change* in algae concentration during that month ("y"); that is, with the difference between algae concentration at the beginning of the following month and that at the beginning of the current month. Thus, if elevated phosphorus levels are leading to greater algae growth, it should show up here. A positive relationship would say, roughly, that increased phosphorus concentration is correlated with faster growth in algae, and a negative relationship would mean, roughly, that higher levels of phosphorus are correlated with *slower* algae growth. The former of these would be interpreted as "cause for concern."

The nature of these relationships is challenging for almost all students. The response variable is *already* a rate of change. Most students will need to be reminded multiple times that their "y" value is *not* "algae concentration," but is rather "*change* (over one month) in algae concentration."

The students compute the month-to-month differences directly, to emphasize what these differences mean. So, for example, we pair the phosphorus concentration at the beginning of June with the change in the concentration in algae from the beginning of June to the beginning of July. Some of these changes will come out negative, and some students will want to ignore the minus sign—don't let them, as this will throw off all their work for the rest of the Activity. Instead, make sure they understand what it means for a quantity to undergo a negative change. Also, students may need to be reminded that, for example, –250 is a greater number than –500 and that, say, 200 is greater than –500. This will be important when they are asked about "greater" or "lesser" changes in a couple of the variables they will be looking at.

Some analysis of these two variables is then done. As before, these are variables for which data should be readily available for just about any decently sized freshwater body. The pattern seen with local data will of course differ from what is seen here, so questions and exposition may need some modification. For the data used here, there is a very strong, *negative* correlation between phosphorus concentration and change in algae concentration. Even with a significant outlier removed the relationship remains clearly negative and fairly strong. The implications of this observation, and its relevance to our "research question" about whether excess phosphorus in the Bush Lake watershed is a cause for concern, will be explored in future Activities.

We then do a similar analysis for monthly rainfall and phosphorus concentration. In a previous Activity (18), the students found a moderate, positive relationship between these variables (paired in a way that seems sensible). As with phosphorus and algae, we now look at *changes* in phosphorus concentrations, during each calendar month, and compare those changes with rainfall from that month. The idea is to look for more direct evidence that watershed phosphorus was being washed into Bush Lake when it rained. This gives a very strong, *positive* relationship, the implications of which will also be explored in future Activities.

2.3.9 Activity 21: Further Analysis of Data Outliers

Word pdf

Learning Outcomes:

- Further analysis of possible effects of outliers regression models
- Correlation coefficients for nonlinear regression models
- Revisiting slope-as-rate-of-change and multiplier interpretation of slope; observing an effect of an outlier on this analysis

In plotting bivariate data, one commonly finds one or more data points that stand out from the rest in some way; the students have seen these points referred to as "outliers," and have seen one possible effect of such a data point. This Activity explores a situation in which an outlier actually *strengthens* the case for a particular model as a good representation of a relationship between variables. This is done by computing a least-squares equation (a power function, in this case) *without* the outlier, and observing:

1. The model predicts the dependent variable almost exactly at the outlier point (in notation not used in this class, a statistician would write $\hat{y}_i \approx y_i$)
2. The equation and its correlation coefficient are virtually identical when computed both with and without the outlier.

The conclusion is then that the model thus obtained is a good one, in that it successfully *extrapolates* values from the data set without the outlier, and it has $r \approx 1$.

The particular data set used here has been looked at before: species richness (y) vs. habitat area (x), for grassland habitats and prairie songbirds. The students have compared linear and power function models earlier, so now we focus on the effect of the outlier on these two types of models.

Before looking at the power function model, we consider the straight-line model. Visual inspection probably leads one to call it a "fair" fit, as it does not show the visible curvature in the scatterplot. The correlation coefficient, $r = 0.81$, is actually (barely) in the "good fit" range.

Now with the outlier removed, there is very little change in the strength of the relationship as measured by the correlation coefficient, but what *does* change substantially is the *slope* of the regression line. The practical significance of this difference is illustrated in some follow-up questions, which are a return to the rate-of-change and multiplier interpretation of a slope.

The Activity concludes with a look at the power function model, whose behavior is described in the opening paragraph above.

2.3.10 Activity 22: More on Outliers, AND When Is a Good Fit Not a Good Fit?

Word pdf

Learning Outcomes:

- Further effects of outliers
- Higher $|r|$ doesn't necessarily mean a better model for bivariate data
- Choosing explanatory and response variables for a graph
- Intelligently selecting a model for bivariate data

Activity 22 is in two parts. Students are not given Part II until they have completed and turned in Part I.

By now the students will have largely internalized the vocabulary and *basic* understanding of the concepts involved in analysis of bivariate data using scatterplots, least-squares equations, and correlation coefficients. In this Activity, we look at a data set that one might expect to have a strong, positive correlation: algae concentration vs. chlorophyll concentration. The students read in a packet from the NMCWD that chlorophyll is a pigment present in algae and other plants. When plotted, the data mostly lives up to this expectation, but with a catch: one (or two) prominent outlier(s).

The data set actually does have two points that could reasonably be called "outliers" due to their distance from the main body of data. However, one of them is what the Web-based primer called an "inline outlier" and the other is "offline." In other words, only one of the outliers lies well outside of the clear linear pattern formed by the others. By eliminating that outlier, the remaining pairs do show a strong, positive linear relationship. So here we have an example of an outlier which significantly reduced the reported strength of otherwise strongly-correlated variables.

We then reintroduce the outlier into the data set, and demonstrate most (if not all) of the other types of regression equations (polynomial up to degree four, exponential, power, and logarithm) that are built into Excel's scatterplot package. The point here is that by increasing the complication (or sophistication, although there's nothing sophisticated about going overboard with this process!) of the model, we can make the fit—in the "local" sense of closely matching the (x, y) values of the observed data—as good as we like, which will be reflected in $|r|$ being as close to one as we like. One must be very careful when doing this, as the resulting graph may actually bear no resemblance to the overall pattern of the scatterplot—that is, the global fit can be truly lousy. So there is a balancing act going on between the "local" and "global" notions of the "fit" of a model to data. One of the key ideas students should come away with is that one cannot choose a "good" data-based mathematical model solely by choosing one with $|r| \approx 1$.

In particular, the demonstration concludes with cubic and quartic polynomial models, which have superb correlation coefficients but are nevertheless egregiously bad models for the data, for a couple of reasons. In particular, the students are invited to notice that both models predict *negative* chlorophyll concentrations for certain algae concentrations.

In Part II, the students are provided with an Excel spreadsheet containing data on the variables "chlorophyll concentration" and "Secchi disk transparency," which have been looked at before. The relationship was predicted to be negative, and this prediction was borne out by the data. As a reminder, Secchi disk transparency is the depth (meters, in this case) to which a white ceramic plate can be lowered into a lake before it can no longer be seen: a simple, still-common measure of "murkiness" of a body of water.

Before making the graphs, however, the students are invited—for the only time in the course, as it's currently constituted—to decide which variable will be plotted on which axis. I introduce the statisticians' terminology of <u>explanatory</u> and <u>response variable</u>. They know that "x" is what "you plug in," and have had experience (before this course, one hopes) thinking about how to use a value of x to predict the corresponding value of y, but haven't thought at all so far in terms of "independent/dependent" or "cause/effect" or "explanatory/response." Of course we all know (all together now) "correlation does not imply causation," but using the language of cause and effect is still a useful heuristic to help students decide which variable is which.

I do not award full credit unless a student has shown understanding of what *underlies* a choice of explanatory ("cause") and response ("effect") variables. Thus, for example, "Chlorophyll concentration causes Secchi disk transparency to go down," while technically correct, does not show much in the way of understanding the reason for that relationship. Full credit requires something along the lines of, "Since chlorophyll is a pigment, it will cloud the water. This makes it harder to see into the lake, and will reduce the depth at which a ceramic plate can still be seen, reducing the Secchi disk transparency."

As before, these are variables that will commonly be measured for any lake, and the relationship here is simple enough that the basic form of a scatterplot from local data should not be too different from what is seen here. Adapted to local data, then, Part II of Activity 22 should adapt readily as a stand-alone exercise for exploring various regression models, in an introductory statistics or college algebra class, and analyzing their "local" and "global" fit.

The Activity concludes with an invitation for the students to explore various regression models for the chlorophyll concentration/Secchi disk transparency relationship, choose one that they think fits best, and defend their choice. The model they choose should match the graph they drew by hand on the scatterplot they printed; they should *not* simply select the model with the highest value of $|r|$! Even after all they've just read and done, I've had students do that!

2.3.11 Activities 23 – 25: Practical Interpretation of Data Analysis

Learning Outcomes:

- Interpretation of models involving complicated variables
- Writing about such an interpretation in terms of a civic issue

Activities 23 – 25 form a unit, which leads up to the only writing assignment (and it's a short one, but not exactly simple or completely straightforward) of the course as it currently exists. The actual numerical work is not particularly difficult, but the thinking involved is challenging even for the strongest students in the class. I've found that the key observations that need to be made in order to inform their analysis may need to be repeated, several times, for some students.

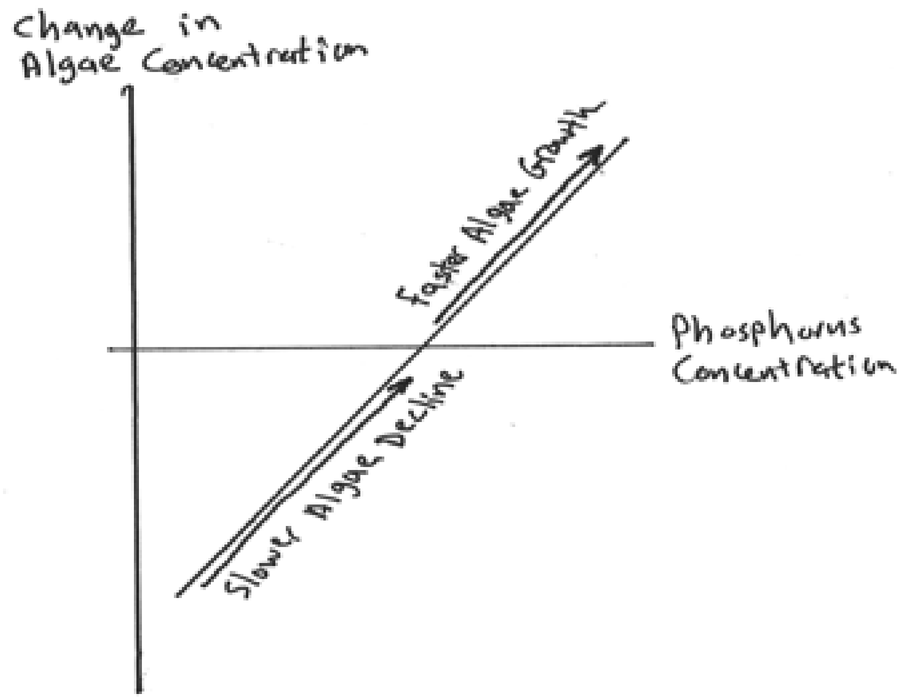
I begin the class in which Activity 23 is handed out with a mini-lecture in which I re-state the "research question" introduced several Activities ago, "Do the data suggest that excess phosphorus in the Bush Lake watershed is a cause for concern?" I then tell them they they will continue to approach the questions from two directions, which was actually begun back in Activity 20:

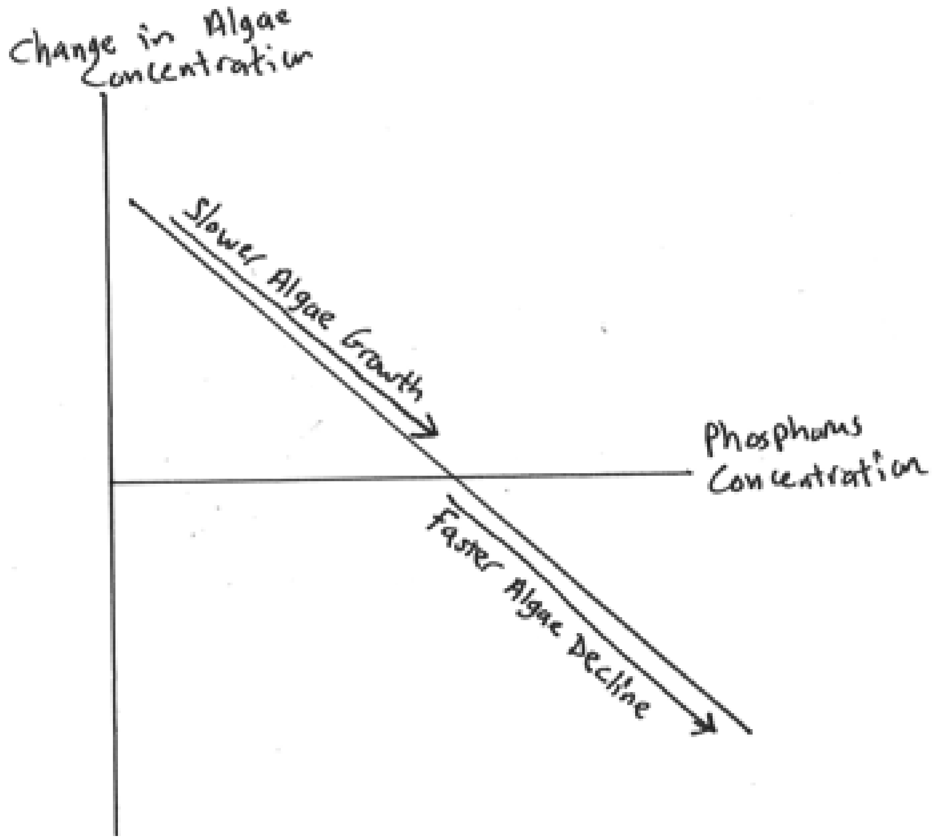
1. Look for a connection between increased phosphorus concentration and accelerated (or "faster" or "greater") growth in algae concentration. Since phosphorus is, as noted earlier, a necessary plant nutrient, a possible concern is that excess phosphorus from the watershed could lead to rapid algae growth, "choking" Bush Lake.
2. Look for a connection between increased rainfall and accelerated (or "faster" or "greater") growth in phosphorus concentration. Such a connection would suggest that excess phosphorus, possibly from artificial fertilizers, is being washed from the surrounding watershed into Bush Lake. It is a common concern in water quality management that such excess phosphorus can lead to plant overgrowth, including but not limited to algae blooms.

Activities 23 and 24 investigate the first and second issues, respectively.

Activity 23 Word pdf

This Activity begins by reminding the students that we changed our approach to the data when exploring the relationship between phosphorus concentration and algae concentration, as described in Activity 20. After the preliminary remarks outlined above, in class before handing out Activity 23, I draw two hypothetical graphs on the board: one showing a positive relationship between phosphorus concentration and change in algae concentration ("algae growth" for short), and one showing a negative relationship. One of the main causes of confusion, or difficulty, in looking at these data relationships is that when the change in a variable is negative, i.e. when a quantity is decreasing, an *increase* in the growth rate has a rather subtle interpretation: namely, that the *decline* in that variable is *slowing*. As mentioned earlier, this is a challenge for many of the students to internalize, but has important implications for the final analysis of the data sets. The graphs I draw on the board are labeled as shown below, first for a hypothetical positive relationship, then for a negative one:


Change in
Algae Concentration
Faster Algae Growth
Phosphorus
Concentration
Slower Algae Decline



Wrapping their minds around these relationships takes time and effort on the part of both students and instructor. A table, which the students were invited to generate in a previous Activity (not all of them do it correctly, which is why it's reproduced here), shows the relationship for the months of May through September (which is "fertilizing season" in Minnesota) of 2000. The data pairs show a generally negative relationship, which the students had already seen via a scatterplot in Activity 20. In this Activity, the students are invited to recognize this relationship *without* a scatterplot, which reinforces the actual *meaning* of a negative relationship.

Many students need help recognizing that a change of –5246 μg/L—that is, a decrease of 5246 μg/L—is a lower (smaller) value than all of the other changes. Many students will need to be reminded that a negative number is less (smaller) than *any* positive number, and that a negative number of larger magnitude is less than a negative number of lesser magnitude, in order to process the relationship as a negatively correlated one.

The students are then provided with Excel output for this data set, including the trendline and its equation. Before looking at some predicted values, the students are invited to recall that the graph is *not* showing a comparison of phosphorus concentration vs. algae concentration; the response variable is *change*, or *growth*, in algae concentration. We then have several questions asking the students to use the least-squares line's equation for predicting changes in algae concentrations for various phosphorus concentrations, and observe the downward trend in predicted values, *including* predicted declines, when y is negative. This last bit should be emphasized: The linear model predicts that, when the algae concentration is in decline, increased phosphorus concentration is associated with *greater decline* in algae.

The Activity concludes with a couple of what are intended as higher-order critical thinking questions, towards the "analysis" end of Bloom's Taxonomy. First, the students are asked to recognize that the model for our relationship predicts *decreasing* growth in algae concentrations as phosphorus concentration increases. Thus the students are invited to further recognize that, based on our earlier discussions about what various data patterns would say about our "research question," that these data do *not* suggest that excess phosphorus concentration in the Bush Lake watershed is a cause for concern, or does not appear to pose an environmental threat to Bush Lake, based on the rather narrow lens of being concerned about the potential for algae blooms. Ultimately Activity 25 will have the students write a paragraph explaining this conclusion and exactly how their data analysis supports it.

Activity 24 Word pdf

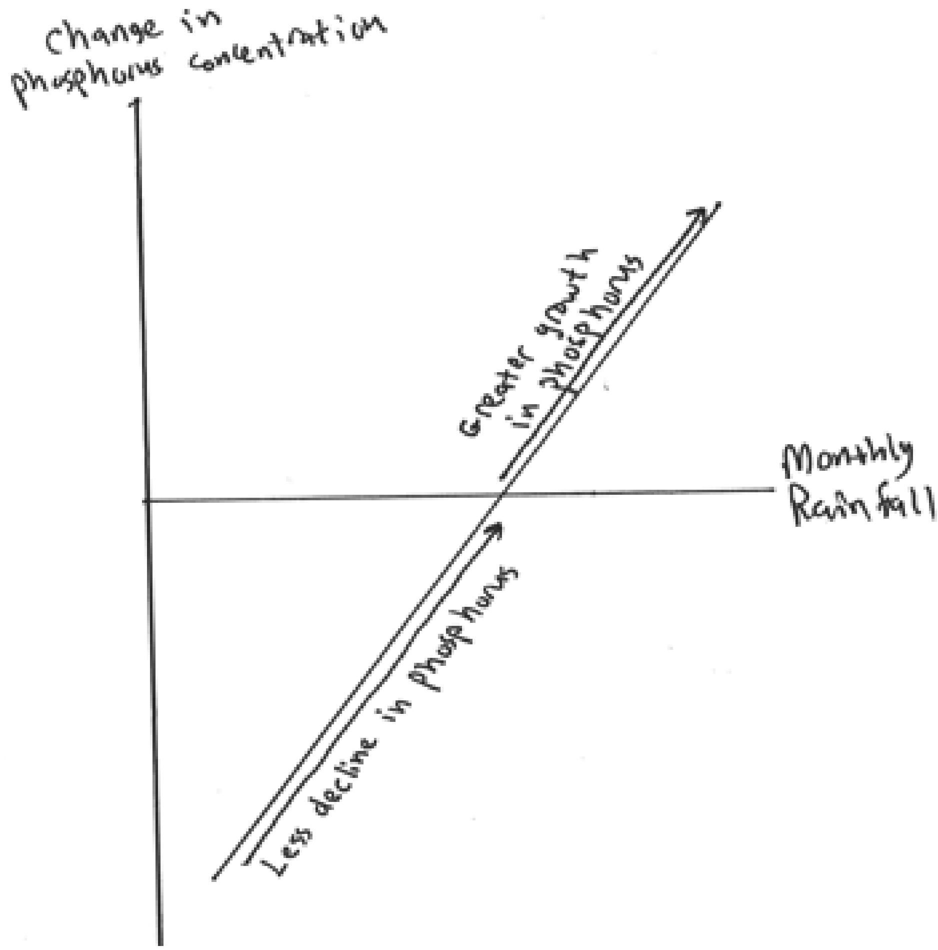
The results of this Activity should lead students to a conclusion that is in tension, if not flatly contradictory, to the conclusion of Activity 24. This was fortuitous, and may not happen with data from other lakes or from other time periods! The variables examined in this Activity are:

x: rainfall in one calendar month in the Bush Lake area

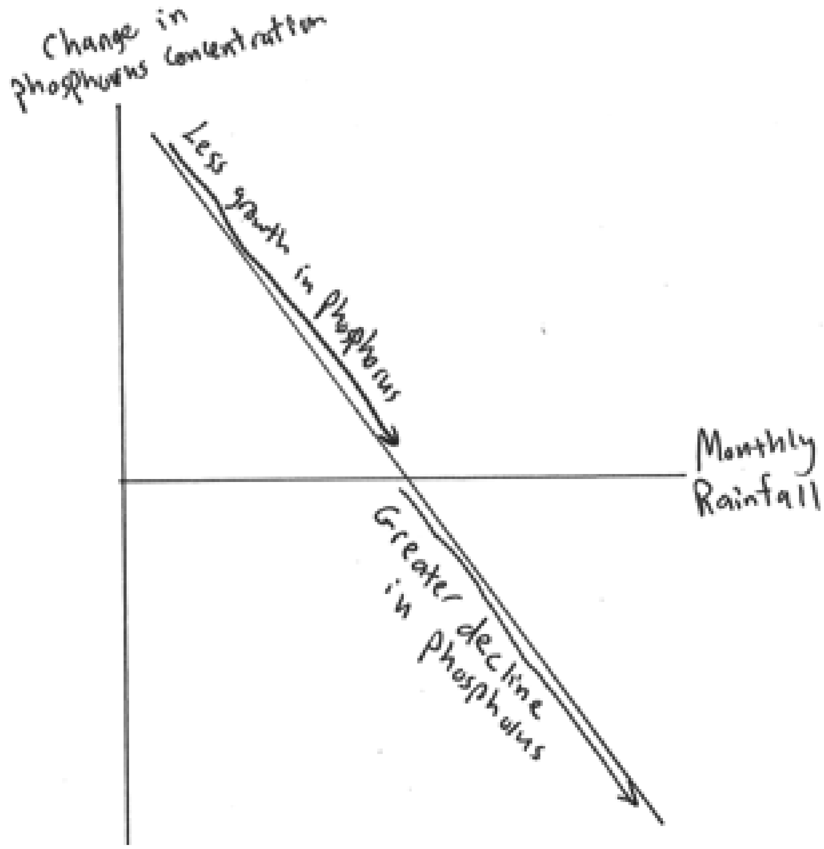
y: change in phosphorus concentration in Bush Lake during that same month.

This is written on the board for student reference during the class. The students are reminded that we took, as our response variable, the monthly change (increase or decrease) in phosphorus concentration, to get a more direct look at whether phosphorus was actually being brought into Bush Lake with rainfall. As at the beginning of Activity 23, I draw hypothetical graphs of the possible relationships between these variables, including a brief interpretation of what such a relationship would mean when phosphorus concentration is decreasing (negative y) and when it is increasing (positive y). The graphs are reproduced below.

For the hypothetical positive relationship:



For the hypothetical negative relationship:



As with Activity 23, a table of data pairs for these variables, as observed in fertilizing season in 2000, is provided. The data mostly show declining phosphorus concentration, so that the response variable values are mostly negative. This is another opportunity to remind students about the ordering of numbers, especially as this relates to negative numbers; if a quantity changes from –9 to –3, for example, that quantity has increased. Only when reminded of this ordering property will students recognize, from the data in tabular form, that they are looking at a positively correlated relationship. (In Minnesota it's not hard to get students to understand this ordering principle—all I need to do is ask, "Which temperature is lower, five below or fifteen below?" This might not work in a climate where it rarely or never gets below zero.)

The data are then processed by Excel and a least-squares line, along with its equation and coefficient of determination, are provided. (The students produced this output back in Activity 20, but not all of them do it correctly, so it is presented here so everyone is working from correct output.) The relationship is now visibly positively correlated, but due to the way we have chosen the response variable it's not obvious how to correctly interpret this correlation. The graphs drawn on the board at the beginning of class are intended to help them to organize their thoughts and to interpret what this relationship is telling them in terms of what seems to be happening to the phosphorus content of Bush Lake when it rains. By the end of this Activity, my goal is that the students have again been led close to the Analysis level of Bloom's Taxonomy, in that they have been invited to discern whether this data suggest that there is "cause for concern" about excess phosphorus in the Bush Lake watershed.

Following the least-squares regression equation, the students are invited to recognize that, as monthly rainfall rises, so does the predicted change in phosphorus concentration. As with the algae concentration, this has a subtle interpretation when phosphorus is in decline (so that y is negative in this context): at a time when phosphorus is declining, increased rainfall *reduces* (or *slows*) the rate of decline. The implication, which is not explicitly stated to the students, is that rainfall is, indeed, washing some phosphorus into Bush Lake. This implication will need to be coaxed out of most students; as a last resort, it will need to be explicitly stated to a few of them.

As with Activity 23, this Activity ends with two questions intended as a "critical thinking" summation of the thought process they have been led through. Namely, that higher monthly rainfall amounts are strongly positively correlated with greater monthly changes/growth in phosphorus concentrations, which implies that these data *do* suggest a cause for concern about excess phosphorus in the watershed, since they strongly imply that there is enough phosphorus being applied that not all of it is being absorbed by the plants being fertilized, so that some of it is washing into Bush Lake when it rains.

Activity 25: Summarizing and Presenting What Has Been Observed, With "Spin" Word pdf

Learning Outcomes:

- Interpreting data relationships in written form from a specific perspective
- How "cherry-picking" of data can be done; selecting data that support one's point and ignoring data that do not; becoming a skeptical consumer of conclusions from quantitative arguments

This Activity spent quite a bit of time in the "development" stages before taking its current form. Its motivation comes from the fact that the data analysis culmination in the conclusions of Activities 23 and 24 was motivated by a question about the existence or nonexistence of a specific environmental threat to Bush Lake: Namely, excess phosphorus in the watershed and, through this, the potential for algae blooms, which sometimes plague once-pristine suburban lakes. The students' final responses to this question will come in the form of two brief essays, of one to two paragraphs, based on the work done exploring the relationships between variables done in Activities 23 and 24 (but begun back in Activity 20).

Now, it just happened that the way the relationships worked out, in Bush Lake in spring through autumn of 2000, allowed me to include a further "quantitative literacy" point into the course at this point. The two analyses seem to lead in opposite directions. The result of Activity 23 was that, even under conditions of heavier phosphorus loading, we saw a *reduction*, or decrease, in the growth rate of algae concentration. Since in this context "cause for concern" would manifest itself in algae accelerating its growth under heavier phosphorus loading, one can reasonably appeal to this analysis to claim *no* cause for concern.

On the other hand, the result of Activity 24 was to conclude that heavier rainfall is strongly correlated with *increased* changes in phosphorus concentrations, suggesting that there *is* sufficient excess phosphorus in the watershed to be washed in with rainfall, so one could reasonably appeal to this analysis to claim that there *is* a cause for concern.

Therefore, I can ask the students to give *opposing* answers to the research question, "Do the data suggest that excess phosphorus is a cause for concern in the Bush Lake watershed?" by appealing to one data set and ignoring the other. Thus the students have an illustration of what is sometimes referred to as "cherry-picking" of data to support a particular point one wishes to make. It is pointed out that there is no need, when doing this, to actually resort to falsehood—at no time are they required to misrepresent the data.

I do not accept completed essays on the day the Activity is given in class. The students must spend the remainder of the class discussing their interpretations with others and writing draft paragraphs. I am willing to give extensive commentary on their drafts before they turn them in (and do *not* grade on grammar, spelling, et cetera). The earliest I'll accept a completed Activity 25 is the beginning of the following class, making this the only time they must do at least some of the work outside of class time. The latest I accept the essays is the day of Exam 2, which is typically a couple of class sessions later.

The two questions in this Activity are graded based on the following five-point rubric:

Question 1:

- Selecting phosphorus/algae data to support their conclusion: 1 pt.
- Pointing out the negative correlation between the relevant variables: 1 pt.
- Correct interpretation of this negative correlation, *including* its implication for both declining *and* for growing algae concentrations: 2 pts.
- Conclusion stated clearly and correctly: 1 pt.

Question 2:

- Selecting rainfall/phosphorus data to support their conclusion: 1 pt.
- Observing the positive correlation between the relevant variables: 1 pt.
- Correctly interpreting this correlation, including its implications for both declining *and* for growing phosphorus concentrations: 2 pts.
- Clear and correct articulation of the conclusion: 1 pt.

EXAM 2 is given shortly after these activities, and covers Activities 12 – 22. The format and policies are the same as exam 1.

2.3.12 Activity 26: Time Series and Their Interpretation

Word pdf

Learning Outcomes:

- Introduction to Time Series graphs
- Recognizing and interpreting regression analysis in public policy-related literature
- More interpretation of slope as rate of change

- More on prediction with a model and relative error

This Activity introduces the concept of a *time series* explicitly. The students have already seen one, when they looked at historical ecological use way back in Activity 1, but the term "time series" was not introduced at that time; it would have had no broader context. Here a time series can be described as just a special kind of scatterplot, in which the horizontal axis measures the passage of time.

This entire Activity was motivated by a brief statement in the Bush Lake U.A.A, handed out several weeks earlier as a supplemental reading. This statement refers to the rate of decline in the lake's "epilimneal phosphorus concentration." At the beginning of class, I tell the students that "epilimneal" is a hydrological term simply meaning "upper layer" of a body of water, and that they are not responsible for this piece of vocabulary, but they need to recognize it when they see it in the reading material. Thus the Activity focuses on forming a "time series," which (they are told in opening remarks for this class session) allows one to track any changes in a quantity over time. In particular, a time series allows one to see at a glance whether, for example, water quality is improving, declining, or staying the same.

The students are provided with a spreadsheet containing the mean phosphorus concentration, as reported by the NMCWD, over the course of several years, and make a scatterplot—thus the time series for phosphorus concentration. Now as with any scatterplot, it is possible to create a least-squares regression line. That model has a slope, which can be interpreted as an (average) rate of change, per year, of Bush Lake's phosphorus concentration. Equivalently, the slope of the model can be used to predict the change in phosphorus concentration over any (reasonable) given time period, using its "multiplier" property (change in y = slope*(change in x)), which has been explored in earlier Activities in other contexts. Thus the claim found in the reading, about the rate of annual decline in phosphorus concentration, can be understood as a claim about the slope of a regression line!

Once the trendline, along with its equation and correlation of determination, have been produced, it's clear that the correlation is negative (if not particularly strong); in other words, the overall trend is, indeed, for reduced phosphorus concentration over time. However, the slope of the trendline does *not* match the claimed reduction rate from the literature! Now, there is a rather obvious outlier, showing a year with unusually high phosphorus concentration. This will be investigated later on in the Activity.

The students go through several instances of interpreting the slope as a rate of change, from a slightly different perspective than before. Starting with an actual, observed mean phosphorus concentration, the students use the slope of the trendline to predict the concentration one or more years later. Since the correlation coefficient is quite low, it should not be surprising that some of the predictions miss the mark rather badly, as computed in their relative errors.

After using the slope for a few predictions of this type, the students are reminded of the rate of change, now interpreted as a slope, asserted by the NMCWD publication, and are invited to compare that rate to the rate they have been using in the preceding calculations. Finally, the students remove from the data list the pair corresponding to the outlier, re-generate the scatterplot, and re-compute the regression line, with equation and r^2 value. The resulting relationship is still negative, but is now both significantly stronger with the magnitude of the slope fairly substantially reduced. However, the new slope *also* fails to match the rate of decrease reported in the literature.

I noticed a curious thing: When rounded to the nearest tenth (of a microgram/liter per year), the two slopes obtained, by including and then excluding the high-phosphorus outlier, *average* to the reported rate! The Activity concludes with a brief series of questions leading the students to notice this.

2.4 UNIT 3

The last handful of Activities forms the last more-or-less cohesive "unit" of the course. The focus shifts, in two ways:

1. Instead of developing mathematical models by fitting functions to raw data points, we work out equations, or formulas, from general considerations about the particular circumstances being considered. What I intend the students to come away with is the following experience: They notice that they are performing essentially the same operation, repeatedly, just using different numerical values of some variable each time. (Of course they may not think of it using this vocabulary.) This recognition hopefully motivates them to see that they can express that process in a single equation, or formula, with letters for the variables. Thus the course concludes with a completely different approach to forming a mathematical model, one based on theoretical considerations rather than on observed data. We then use such a model to predict the behavior of Nine Mile Creek's capacity to absorb stormwater runoff before overflowing its banks.
2. In non-mathematical terms, we shift our focus away from the biochemical ecological state of local waterways onto another responsibility of most watershed district authorities: managing stormwater and reducing the risk of flooding. The computations done here, specific to Nine Mile Creek, its watershed, and the Normandale Community College campus, could be adapted for use with data sourced and collected locally. We use some data that were actually collected by students in past sections of this course, using equipment borrowed from the geology department.

2.4.1 Activity 27: Modeling Rainwater Runoff

Word pdf

Learning Outcomes:

- Mathematical modeling from theoretical considerations
- Forming algebraic expressions from verbal summary of a procedure
- Complex unit conversion with unfamiliar or cumbersome units

A basic geometric relationship gives rise to the fundamental constant underlying the essential calculation done, *ad nauseum*, in this Activity. As mentioned in the general introduction to this Unit of the course, the first few exercises are geared towards having the students repeat a computational process to the point where they can recognize a pattern; then, with the variables they've been working with pointed out to them, hopefully the students can arrive at a formula or equation representing the process in its generality. In practice some students arrive at a correct formula immediately while others need considerable guidance.

One feature that I've learned needs considerable commentary from the instructor is the business of producing formulas in which percentages or proportions play a crucial role. Thus I illustrate the common notation, involving "$100p$ %," in a "sale/discount" context. The $100p$ percent notation throws almost all of the students for a loop when they first encounter it, even *with* a detailed couple of examples at the beginning of class.

One aspect of the general (definitely overly-simple) model for stormwater runoff used in this Activity that should probably be emphasized strongly, since it is the reason students must convert watershed areas, reported in NMCWD literature in acres, to square feet, is that the equation developed for runoff is valid only when area is measured in square feet, due to the way the constant in the equation was defined.

The rest of this Activity is intended to give the students a sense of just *how much* water there is to be managed during a rainfall event. I point out to them that *all* of this water must *go somewhere*, and that poor management can lead to various kinds of otherwise avoidable flooding. Gallons are used here since students have a pretty clear, intuitive/experiential sense of how large a gallon is. Later we switch to cubic feet, since our publication reports waterway discharges in cubic feet/second (cfs). Most students have little sense of the magnitude of the volume units most commonly found in hydrological literature, cubic feet or (!) acre-feet.

2.4.2 Activity 28: Estimating Overflow Capacity of a Waterway

Word pdf

Learning Outcomes:

- Further development of mathematical models from theoretical considerations
- Approaching a difficult problem by looking at a simple case of that problem
- Return of sigma notation and subscripted indices
- Approximation of complex regions with simple geometric shapes (basic Riemann sums, really)

This Activity continues, and expands upon, the idea of building mathematical models for observable phenomena from general, "theoretical" considerations rather than from data. It introduces a common procedure in mathematics: Namely, approach a difficult or complicated problem by first solving a simpler case of the same problem. In this case, the problem is to calculate the flow rate of water, or "discharge rate" in hydrological nomenclature, of a moving waterway such as a stream, creek, or river. Since the shape of such a waterway is almost always highly irregular, we begin by computing discharge in a highly idealized setting: a waterway which is perfectly rectangular, and in which the water flows past at a uniform velocity. (The complication that water flows more slowly near the bottom and sides of a creek is not directly addressed in this course. Flow readings in the field were taken about $\frac{1}{3}$ of the way from the bottom, on the advice of a professional hydrologist!)

Using the model thus constructed, students use field data from Nine Mile Creek depths and flow (velocity) rates, taken at regular intervals across the creek at one location near campus, to build a discharge estimate, using the simplification that models the creek-bed using rectangles concatenated together. (Yes, they are calculating a Riemann sum! I've actually done this exact same process with a first-year calculus class.)

The ultimate goal of this Activity is to make use of a simple algebraic model of creek discharge, in terms of water depth, water velocity, and creek width, to estimate the "bankfull capacity" of Nine Mile Creek at one specific location near campus. "Bankfull capacity" means the maximum discharge a waterway can carry without overflowing its banks. Thus the students apply the model they develop for waterway discharge, along with data collected by an earlier class, to answer the question, "How much water can Nine Mile Creek carry without flooding?" It is part of the Watershed District charge to study and recommend water management strategies that will minimize the chances of the creek frequently exceeding that discharge rate.

Along the way it is pointed out, by direct computation, that one can arrive at the same discharge estimate as is obtained by the Riemann sum (of course that term is never used!) by doing the "rectangle" computation using the average depths and velocities, and total width, of the creek. This could, of course, be proved algebraically, but that would be well beyond the scope of a course at this level.

The field teams were assigned different data-gathering jobs (see Appendix B). One team gathered the data that is used in this Activity to estimate the creek discharge on the field trip day, namely depths and velocities at regularly spaced points, and bank-to-bank width. A different team, at a *different* nearby point on the creek, collected data which are relevant to bankfull capacity: they estimated the point at which the water would spill over the banks of the creek into the park, and then measured the bank-to-bank width at that water level, and measured the distance from that level to the creek bed at several locations across the creek. These latter measurements would, of course, be creek depths if the creek were bankfull. A problem remains: For estimating creek discharge, we also need an average water velocity, which this second team did not measure, since we had only one flow meter. Velocities can vary quite a bit in a creek, from location to location, depending on the local topography, mostly based on the depths at that point. Thus we need, somehow, to estimate the average water velocity at this second location. Algebra comes to the rescue! Thus we encounter only the second, and final, time in which the students need to actually *solve an equation* for an unknown variable. Many of them, perhaps half in a typical class, can solve the relevant equation on their own. I do need to at least give hints on the solution process for many of the teams. The equation the students derive takes advantage of the fact that, while creek *velocity* varies from location to location, the *discharge* rate does not, assuming there are no tributaries, springs, or storm drains adding water between the two locations.

Once the students have their velocity estimate, it's a simple matter to use the estimated bankfull dimensions to predict the bankfull capacity discharge rate. The class concludes by revising this estimate considering the observable fact that, as water rises, its velocity tends to increase.

2.4.3 Activity 29: Estimating Complicated Areas

Word pdf

Learning Outcomes:

- Using collections of simple shapes (rectangles & triangles) to estimate complicated areas
- Using scale maps to compute areas
- Area unit conversion

I was graciously provided by the Normandale building and grounds department with detailed, scale maps of the entire college property. Since our final goal for the course is to determine whether stormwater runoff from Normandale alone would be enough to flood Nine Mile Creek, we are going to need a reasonably accurate estimate, in square feet, of the total impervious area of campus. I had the students include all parking lots, plazas, and sidewalks, but not rooftops. This could easily be adapted to any local campus or other facility for which such scale maps exist. The computation provides at least three challenges:

1. The impervious surfaces tend to be highly irregularly shaped. Students are encouraged to approximate them with rectangles and triangles, and are provided with rulers and reminded how to compute areas of non-right triangles by dropping a perpendicular.
2. All areas must be reported in square feet. There are various ways of carrying out the conversion—they'll be measuring map areas in inches, and using the given scales—each team is left to its own devices as far as deciding how to do the conversion.
3. There can be very large discrepancies among computed areas reported by different teams working on the same campus subdivision. Part of the mathematical point of this exercise is to understand that there is inherent inaccuracy built into any large scale, complex estimation of this kind, which arises not only from measurement or calculation errors, but also from different decisions in how to produce one's approximation even if no actual "mistakes" are made. (Thus, for example, widely divergent estimates in crowd sizes in "marches on Washington.")

Each team of two or, at most, three students is provided with a scale map of one subdivision of campus, along with a ruler or two. The subdivisions are numbered (numbers are printed on the maps) and the numbers are written on the whiteboards surrounding the classroom. Each team is to write their estimated area, in square feet, on the board as they finish it and agree amongst themselves. Each subdivision is assigned to at least two teams, and as estimates invariably differ, the teams are to meet, sort out their differences, and arrive at a common, agreed upon estimate. Depending on the size of the class, this process can easily take two class sessions to complete the whole campus. It is definitely the case that some subdivisions are simpler and go more quickly than others.

Once the teams arrive at area estimates for each relevant subdivision (some subdivisions have no paved surfaces and hence are not considered), the areas are added together to obtain our class estimate of the total impervious area of the Normandale campus. This area will be used in the final Activity of class to estimate the water discharge from campus under various rainfall conditions, and compare those rates with the maximum creek capacity calculated in Activity 28.

2.4.4 Activity 30: Estimating Runoff from Campus

Word pdf

Learning Outcome:

- Unit conversion of rates

In the previous Activity, the students arrived at an estimate of the total impervious area of our campus—as a commuter campus, there's a lot of it, since the buildings are surrounded by acres of parking lots. It is pointed out at the beginning of class that there are stormdrains at various locations, to carry water away from these surfaces, and that all of that water is going . . . *somewhere*. Once upon a time it would have been standard practice to simply build stormdrains to the nearest body of water, which in our case is Nine Mile Creek. In addition to water pollution issues, doing this would add a lot more water to the creek than it would naturally receive from a rainfall event, and so would increase the risk of flooding downstream. So the question addressed in this Activity is: Would the discharge from campus be enough, by itself, to overflow Nine Mile Creek?

This question is answered by looking at rainfall events with a constant rainfall rate. Since this is highly unlikely in real life, one can interpret these rates as hourly averages. The students make runoff calculations very similar to those done in Activity 27; the only difference is that rainfall *amounts*, in inches, are replaced by rainfall *rates*, in inches per hour. This gives rates of water discharge from Normandale's impervious surfaces in cubic feet per hour. Since our estimated bankfull discharge was given in cubic feet per second, we need to do some unit conversion. I do not give help here unless the students request it; about half of them typically can work it out on their own. Finally, the students compare their computed campus discharge rates, in cfs, with their earlier estimate of the creek's bankfull discharge rate, and determine under what sorts of rainfall conditions our campus' runoff would be enough to flood Nine Mile Creek.

At this point the students take their third regular exam, whose format is the same as the first two. I conclude with one or two (50-minute) class sessions for review, make-up work, and student questions. The final exam is cumulative, with the same format as the three regular term exams (but twice as many questions in each of the three parts) and with very similar questions. The students may use all of their completed, corrected Activities, any reading material handed out, and their three corrected exams when taking the final. The final uses different data sets, but asks similar questions about those data sets, as the three regular exams.

2.5 APPENDIX A: Sample Syllabus

MATH 1020, MATHEMATICS FOR LIBERAL ARTS, 3 CREDITS

FALL SEMESTER, 2015

MWF 11:00 – 11:50 am, C3060

Instructor: Tony Dunlop

Office: C3124

Telephone: (952) 358 – 8487

Electronic Mail Address: tony.dunlop@normandale.edu

Office Hours: T 11:00 – 11:50 (C3124), 1:00 – 1:50 (Tutoring Center); W 3:00 – 3:50 (C3124); F 1:00 – 1:50 (C3124), 2:00 – 2:50 (Tutoring Center); **or by appointment**. Or just stop by if you're on campus; if I'm too busy to chat, I'll politely say so. You may also call or email with questions. I don't answer the phone if I'm with a student, so leave a message if I don't answer during an office hour.

Textbook: None required. Course materials, mostly written by the instructor, will be distributed in class. Additional reading from newspapers, government or industry documents, and the World Wide Web will be assigned as needed.

About the Course: Hopefully, this will be unlike any math class you've ever taken. We will not learn a bunch of formulas or rote algorithms. Instead, we will look at some environmental and infrastructure topics, from the point of view of mathematics. In particular, we will see how mathematical thinking can inform our understanding of water and energy issues. Much of the mathematics we use will be familiar to you (such as equation solving, graphing, and logarithms), some of it will be new, and we will apply the math you do know in ways you might not have thought of.

Topics covered from the following list.

- Using numbers and measurements to classify stream types and determine level of ecological degradation, if any.
- Stormwater runoff and flood potential. Effect of permeable vs. impermeable surfaces. Effect of detention ponds, dams, and wetlands on flood potential from heavy rainfall. Impact of development on flood potential. Quantitative justification of watershed development/redevelopment regulations.
- Pollutants in Nine Mile Creek and surrounding lakes, especially excess phosphorus and nitrogen, and de-icing salt. Relationship between excess phosphorus and algae blooms. Growth/decline of native and sensitive species.

Mathematics used will be from the following list:

- Numerical summaries of data
- Applications of ratios
- Careful measurement of field data
- Elementary descriptive statistics, including regression and correlation
- Strength of a relationship between tow variables
- Fitting data to a model; interpolation vs. extrapolation
- Analysis of graphs of data for overall trends and anomalous readings
- Positive vs. negative relationships
- "Outliers" and their effects
- Manipulation of graphs to over- or underemphasize data trends
- Absolute vs. relative change and rate of change. Linear vs. exponential and other models.
- Relationship between rate and accumulation

- Analysis of units of measurement; unit conversion
- Scale, order of magnitude, scientific notation. Units appropriate for different scales.
- Discrete vs. continuous variables
- Weighted averages
- Estimation; upper and lower bounds
- Iterative processes
- Testing a model against empirical data
- Dependent vs. independent variables

The class will be a combination of whole class discussion, small group work, and, if it can't be avoided, short lectures to try to clarify difficult concepts. Since there is no textbook, and your grade is largely based on your in-class activities, attendance is essential.

Your grade will be based on the following components:

Class Activities/Homework..40%

3 examinations, 13 $\frac{1}{3}$ % each.......................................40%

Final Examination...20%

Class Activities/Homework: Each week, you will work with one or two other students on activities based on that week's topic. It is the results of these activities that you will turn in for this portion of your grade. It may happen that you finish the activity before the end of class. If you do not, it is the responsibility of you and your partner(s) to finish it outside of class, and turn it in no later than the beginning of the next class period.

The *examinations* will consist of questions which are similar to the in class/homework problems, basic definitions, multiple choice, and "true/false" questions. Exam dates will be announced at least one week in advance. The final will be cumulative, of the same format, and about twice as long as a regular exam. All exams are open "book"/open notes.

Advice: Through the activities mentioned above, you will get to know several other students. I urge you to get together outside of class to discuss the activities and readings, even if you finished your activity in class. If I'm doing my job properly, you will be frustrated from time to time. This is what happens when we encounter new ideas. It will also give you a tiny taste of what the great discoverers of history go through *before* they make the discoveries for which they become famous. Having others to work with is the single best way to get from frustration to understanding.

The Fine Print: Rules and regulations.

Missing Class. Since there is no textbook, class attendance is vital.

Missing more than 9 classes will automatically result in a grade of F.

Late work will be accepted only if you notify me in advance (if possible), or as promptly as possible. If you miss a class, then show up at the following class without having contacted me, your work will not be accepted for credit. If you are sick, caring for a sick person, stuck in traffic, etc. please call and leave a voice mail message *as soon as possible* (but please do not call while driving). I'm a reasonable guy; just let me know what's going on and you'll be OK.

Make – up exams: Similar rules apply. Let me know in advance, if possible, if you cannot make it to an examination. You must have a compelling and documentable reason. It will then be your responsibility to arrange to make it up outside of class, in the make – up testing room at the NCC library. To be fair to students who take their tests on schedule, make-up tests are slightly harder.

Conduct in the classroom: Inconsiderate behavior is rarely an issue, but for the record, all students at Normandale have the right to a dignified and respectful environment that is conducive to learning. It hopefully goes without saying that *all telephones and other such devices should be turned off during class.* If your phone rings in class, you'll have to watch me dance until it stops. Oh, the horror. Disruptive behavior, such as habitually arriving late or leaving early, rude or insulting language, talking while another person has the floor, or other activity not directly related to that class day's activity will not be tolerated, and I reserve the right to ask such a student to leave. **Anyone asked to leave class more than once during the semester will have his/her grade reduced by one letter.**

Withdrawals must be initiated by the student. If you realize you are not going to complete the course, but forget to formally withdraw, I must report a grade of F. I cannot enter a grade of W.

Incompletes are very difficult to come by; you must have completed all but a small amount of the course work at a C level or better, and there must be some severe, unexpected, and documentable situation preventing you from completing the course on time. Incompletes which are not finished by the end of the following semester automatically become Fs. *You cannot receive an incomplete simply because you are not satisfied with your performance and want more time.*

Important Dates:

NO CLASS: Monday, Sept. 7; Friday, Oct. 16; Wednesday, Nov. 11; Friday, Nov. 27.

LAST DAY TO WITHDRAW: Wednesday, November 25

FINAL EXAM DATE/TIME: **Monday, Dec 14, 11:00 am – 1:50 pm.**

The Common Course Outline for this class is on the reverse side of this page, and is posted in the Content section of this course's Desire "2" Learn Web site. This is an official document of the Normandale math department, giving an overview of the prerequisites and goals of the course. It is mostly useful as documentation for transfer schools. **This course satisfies Goal 4 of the Minnesota Transfer Curriculum** and will transfer to any college in the MnSCU system.

2.6 APPENDIX B: Tasks for Data-Gathering Field Trip

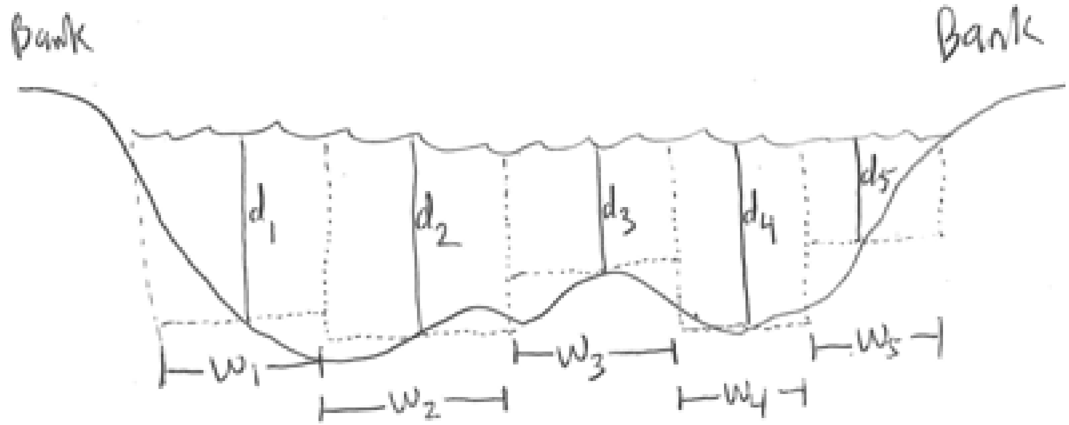
Math 1020, Mathematics for the Liberal Arts
Task 1: Measuring current speed and depth at a creek cross-section
Tools Needed: Current meter; yard/meter stick; boots/waders; string

Goal: Determine creek flow, in cubic feet per second (cfs).

Data collection: Find a point where the creek is deep enough that you can get readings from the current meter but not so deep that you can't wade in. You'll also need to make sure that the water is flowing at a decent pace.

Measure the current speed (hold the current meter about $\frac{2}{3}$ of the way down) and the depth at 5 or 6 roughly equally spaced points across the creek. **Make sure you accurately measure the distance between measuring points.** Carefully record all of these measurements in a notebook.

Calculation: For the above measurements, calculate $\sum v_i d_i w_i$, where v is the current velocity (feet/second), d is the creek depth (in feet), and w is the width (in feet) (see diagram below). The result is an estimate of the total creek flow, in cfs ($\frac{ft^3}{sec}$).



Why it works: If water is flowing at a constant rate v feet/sec, through a rectangular passage of depth d feet and width w feet, then the amount of water flowing through in 1 second is $v \cdot d \cdot w$. What you are doing is *approximating* the flow of the irregularly shaped creekbed by "modeling" it as 5 or 6 rectangles of varying depth (see diagram above). If you wanted to get a more accurate approximation, you would take a larger number of readings, closer together.

Table for Data Collected: Task 1

Location (i)	Creek depth, inches (d_i)	Width, inches (w_i)	Creek velocity, feet/sec (v_i)	Estimated discharge (ft^3/sec) ($d_i w_i v_i$)
1				
2				
3				
4				
5				
6				

Fill in the first three columns for as many readings as you take (see diagram on original handout), up to six. You do not need to fill in the last column right away. CAUTION: You'll need to convert your depths and widths to FEET before multiplying, so that your discharges are in the correct units!

To get your total estimated creek discharge ($\sum d_i w_i v_i$), add all the entries in the last column. The class will be using your estimate in a future activity!

Math 1020, Mathematics for the Liberal Arts
Task 2: Survey watershed (i.e. area draining directly into the creek); assessment of erosion or erosion potential. (See first line of the survey table for things to look for.) Discuss as a team until you agree which of the four categories (Excellent, Good, Fair, Poor) to put this reach into.

Equipment needed: None, other than a sharp eye. Look for loose soil, places where it appears that soil/sand/gravel has washed into the creek, etc.

Goal: Assess the erosion potential of the area immediately surrounding the reach. Classify it as Excellent, Good, Fair, or Poor, using the guidelines in the first row of the evaluation table. Be aware that the "watershed" is wider than just the immediate creek banks.

Note: You are not trained ecologists, so you won't be graded on getting it "right." Do write down some observations that led you to your conclusion below. Make your conclusion, Excellent, Good, Fair, or Poor, clear.

Math 1020, Mathematics for the Liberal Arts
Task 3: Nonpoint Pollution Source survey.
Tools Needed: Car/Driver; Satellite photo of region (from, say, Google Maps)

Goal: Assess the wider region around the reach for possible pollution sources, such as storm drains, roads/parking lots, industrial sites, chemically treated and/or fertilized lawns, wetlands draining into creek, etc. Use criteria from second row in evaluation table for guidance.

Procedure: Take a short drive around the neighborhood surrounding the reach. (NOTE: The driver must **not** be involved in any of the observations; his/her only job is to drive safely.) Make a note of any possible pollution sources, as mentioned above. Also look at satellite photo for further observations. Discuss what you observed as a group until you agree on a category (Excellent, Good, Fair, or Poor), as described in the second row of the evaluation table.

Note: You are not trained ecologists, so you won't be graded on getting it "right." Do write down some observations that led you to your conclusion below. Make your conclusion, Excellent, Good, Fair, or Poor, clear.

Math 1020, Mathematics for the Liberal Arts
Task 4: Creek bank survey, for erosion and vegetative cover.
Tools needed: None, other than a sharp eye and good observation skills.

Goal: Assess the condition of the immediate creek bank, as far as erosion level and amount of plant/vegetation cover. Determine category (Excellent, Good, Fair or Poor) for the third and fourth rows of the evaluation table.

Procedure: Look up and down the immediate bank of the reach as much as possible, observing the extent of eroded spots and the level of vegetative covering. Assess the quality of the bank according to the rankings given in the third and fourth rows of the evaluation table. Discuss the observations among yourselves until you agree on which classification to use for each of those two rows.

Note: You are not trained ecologists, so you won't be graded on getting it "right." Do write down some observations that led you to your conclusion below. Make your conclusions, Excellent, Good, Fair, or Poor, clear.

Math 1020, Mathematics for the Liberal Arts
Task 5: Assessing creek bottom as habitat; assessing "aesthetics" of the reach
Tools needed: None, other than keen observation and a sharp eye

Goal: Classify the creek bottom according to the row "Bottom Substrate;" classify the surroundings according to the row "Aesthetics" in the evaluation table.

Procedure: Look into the creekbed (i.e. underwater bottom) for quality of habitat. Look for rocks, gravel, or plants ("other stable habitat") as a sign of good habitat; less of these things (and more sand or mud) means lower-quality habitat. Classify the habitat as Excellent, Good, Fair, or Poor, according to the "Bottom Substrate" row. Discuss in your group until you agree on the classification.

Look at the physical surroundings of the reach, in light of the categories in the "Aesthetics" row of the evaluation table. Classify the surroundings as Excellent, Good, Fair, or Poor, according to the categories in this row. Discuss in your group until you agree on the classification.

Note: You are not trained ecologists, so you won't be graded on getting it "right." Do write down some observations that led you to your conclusion below. Make your conclusion, Excellent, Good, Fair, or Poor, clear.

Math 1020, Mathematics for the Liberal Arts
Task 6: Measuring depths of riffles and pools.
Tools Needed: Boots/Waders, Yard/Meter Stick.

Goal: Determine the average depth of the two main kinds of regions of the creek: riffles (i.e. rapids, where the creek is fairly turbulent) and pools (where the creek is fairly calm).

Procedure: Find at least two riffles in the reach, and at least two pools. For each, take five or six regularly spaced depth readings across a cross-section. Record these depths in a notebook, being careful to keep riffle readings separate from pool readings.

For the riffle readings: Calculate the average depth at each cross section, then average these averages. This will be the figure you use to classify the reach according to the "Avg. Depth Riffles and Runs" row of the evaluation table. (Nine Mile Creek is considered a warm water creek.)

For the pool readings: Calculate the average depth at each cross section, then average these averages. This will be the figure you use to classify the reach according to the "Avg. Depth of Pools" row of the evaluation table. (Nine Mile Creek is considered a warm water creek.)

Show your calculations below, and give you classification (Excellent, Good, Fair, or Poor) for the "Avg. Depth Riffles and Runs" row, and the "Avg. Depth of Pools" row, in the space below.

Table for Data Collected: Task 6

First Riffle: Depth at five or six regularly spaced intervals (in inches).

Average depth, in inches, of first riffle (average the readings above):

__________ inches.

Second Riffle: Depth at five or six regularly spaced intervals (in inches).

Average depth, in inches, of second riffle (average the readings above):

__________ inches.

Average riffle depth (average the 2 averages): __________ inches.

Classification (warm water creek), according to evaluation sheet:

Excellent Good Fair Poor

(see reverse for pool data recording)

First Pool: Depth at 5 or 6 regularly spaced intervals (in inches).

Average depth, in inches, of first pool (average the readings above):

__________ inches.

Second Pool: Depth at 5 or 6 regularly spaced intervals (in inches).

Average depth, in inches, of second pool (average the readings above):

__________ inches.

Average pool depth (average the 2 averages): __________ inches.

Classification (warm water creek), according to evaluation sheet:

Excellent Good Fair Poor

Math 1020, Mathematics for the Liberal Arts
Task 7: Determine "Lower Bank Channel Capacity"
Tools needed: Boots/Waders, Yard/Meter Stick, String, Tape measure
Follow-Up to Task 6; uses data from that task.

Goal: Measure the width-to-depth ("W/D") ratio of the creek in this reach. This will allow us to classify the reach according to the "Lower Bank Channel Capacity" row of the evaluation table.

Procedure: Using the average depth from the four or more locations used in Task 6, you'll need to find the "bankfull depth." To do this, your team will need to decide, by looking at the creek bank, the point at which the creek will be considered "full." Then you'll need to stretch a string or tape measure across the width of the creek. You need two measurements: (1) The width of the creek at that point; (2) The difference between the current creek level and the "bankfull" level.

The measurement (1) will be the "bankfull width," W; to get the bankfull depth, D, add measurement (2) to the average depth you found for that point. (See diagram below.) Finally, calculate the ratio $\frac{W}{D}$. This value tells you in which category (Excellent, Good, Fair, or Poor) to classify the reach in the "Lower Bank Channel Capacity" row of the evaluation table.

Show your calculations on the bottom and, if necessary, reverse of this page.

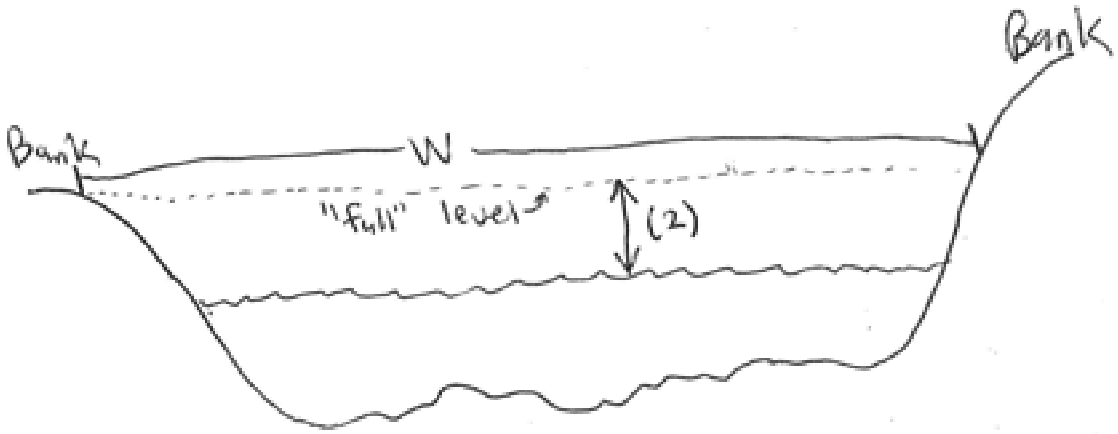


Table for Data Collected and Ratio Calculation: Task 7

Bankfull width, in inches (see diagram on original handout):

W = ___________

Difference between today's water level and bankfull level (labeled "(2)" on diagram), in inches:

Average creek depth, in inches (average the avg. riffle depth and avg. pool depth from Task 6):

Add the preceding two measurements; this will be your estimated "bankfull depth," in inches:

D = ___________

Bankfull Width-to-Depth Ratio: Divide W/D, using values given above:

W/D =

Use the criteria in the "Lower Bank Channel Capacity" row of the evaluation sheet to classify the creek:

Excellent Good Fair Poor

Math 1020, Mathematics for the Liberal Arts
Task 8: Measuring and calculating "pool/riffle" ratio
Tools needed: Boots, Waders; String; Tape measure

Goal: Assess "variety of habitat" by measuring the ratio: $\frac{dist.betweenriffles}{streamwidth}$. Fits with 2^{nd} to last row in evaluation table.

Procedure: Find at least two places in the reach where two riffles are separated by a pool. (The boundaries between riffles and pools are not always clearly defined; try to agree as a team where to measure from.) Use the string to follow the stream from the beginning to the end of that pool (that is, from the end of one riffle to the beginning of the other). This may take some creativity! Then measure the length of the string to estimate the distance, in feet, between riffles. Then calculate the average of the two (or more) distances you found; this will be the numerator of your ratio.

Measure the width of the creek, using string and/or tape measure, in a few different spots (in the same part of the reach where you were doing the above-mentioned measurements). Average these widths, and use that figure as the denominator of the ratio.

Record your measurements in a notebook, and report the results in the space below, and (if necessary) on the back of the page. Calculate the ratio $\frac{dist.betweenriffles}{streamwidth}$ using the numerator and denominator as described above.