Improving Your Graphs: an Exercise in Data Presentation

Resource Type: Curriculum: Laboratory

Publication Date: 9/28/2007

Authors

David Wessner
Davidson College
Davidson, North Carolina 28035
Email: dawessner@davidson.edu

Abstract

Too often, beginning science students do not fully understand the importance of data presentation. In their textbooks and during lectures, data generally are presented in a clear easy-to-understand format. In this exercise, students are presented with an introduction to graphs and tables, a short primer on Excel, and hypothetical raw data from several related experiments. Working together, pairs of students prepare figures displaying the data. They then show their figures to the class and, as a group, the class discusses the relative merits of different data presentation approaches.

Activity

Invitation for User Feedback. If you have used the activity and would like to provide feedback, please send an e-mail to MicrobeLibrary@asmusa.org. Feedback can include ideas which complement the activity and new approaches for implementing the activity. Your comments will be added to the activity under a separate section labeled "Feedback." Comments may be edited.

Learning Objectives.

At the completion of this exercise, students will:

·   Know the major components of a good graph and table

·   Be able to present data in graphical and tabular forms

·   Be familiar with graphing software

Background.

Several national science literacy initiatives, including Project 2061 and the National Science Education Standards (http://www.nap.edu/readingroom/books/nses/), have identified data presentation as a major goal for improved science literacy. Yet studies have shown that even college students often have a poor understanding of graphs (2). This exercise designed for beginning microbiology and biology students, or potentially nonscience majors, is intended to improve students' ability to construct graphs. We routinely complete this exercise during the third or fourth week of the semester in our introductory biology course for first-year students. To complete this exercise, students need to be familiar with the type(s) of experiment(s) depicted in the hypothetical scenarios. Because we begin our course with a unit on enzymes and enzyme kinetics, the students are familiar with the types of experiments depicted and spectrophotometry. The hypothetical experiments, though, easily can be altered to fit the knowledge base of the students. The only important feature that must be maintained is that the supplied data should be presented best in different formats (bar graph, line graph, table, etc.).

To complete these exercises, students should have access to computers with standard graphing software (e.g., Excel). Students also need a basic knowledge of the spreadsheet and graphing software available to them. In the Student Version, a short primer on the use of Excel and a brief introduction to graphs and tables are provided. Ideally, an overhead projection system also should be available to allow the entire class to easily see each group's figures.

PROCEDURE

Materials.

- Computers with appropriate spreadsheet and graphing software
- Overhead projection system

Student Version.

Introduction

As we have discussed in lecture, enzymes catalyze chemical reactions, usually resulting in the conversion of one substance, the substrate, into another substance, the product. And, as we have seen previously in the laboratory, we often can measure the conversion of the substrate to product using spectrophotometry. With relatively simple spectrophotometric assays, we can determine if an enzyme is present in a sample and examine the substrate specificity of the enzyme. And, if we measure the rate of substrate to product conversion over time, we can determine the activity of the enzyme.

In this exercise, we will explore ways of presenting data obtained from these types of experiments. Individually, read through the brief introduction to graphs and tables. This section provides you with a general format for figures and tables that is acceptable for many scientific journals. Then, working in pairs, read through and complete the following assignment. If you and your partner are not familiar with Excel, please consult the short primer provided at the end of the assignment when making your figures.

A. Graphs and Tables

Data that have been collected need to be presented clearly and succinctly. As a result, two forms of presentation are most commonly used in scientific papers: figures and tables. Which method to use depends somewhat on the data, but in general a picture (e.g., a graph or diagram) usually is more desirable than a table, because the reader can immediately see the trends in the data. In the paper itself, diagrams, photographs, and graphs are all referred to as "Figures" and are numbered sequentially in the order of presentation (Figure 1, Figure 2, etc.). Tables also are numbered sequentially in the order of presentation. Although figures and tables often are placed directly into the middle of scientific papers, you may include figures and tables within the text of your report or at the end of your report.

Graphs. Graphs can be made using a graphing program such as Excel. Remember to label each axis, including units of measurement, and clearly identify the data you are displaying (i.e., label each line in a graph). In addition, every graph must have a short description (legend) below it to tell the reader some basic information about the data and how they were obtained. The legend starts with the figure number, followed by a one-sentence title. The text of the legend should be a short paragraph. Below is an example of a graphic figure with legend.
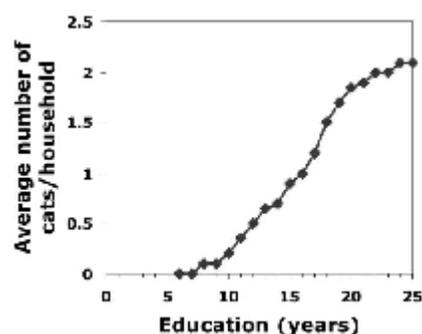

FIG. 1. Cat ownership is directly related to educational level. One hundred fifty-six Davidson, North Carolina, adult residents were surveyed to determine their education level and the number of cats in their household.

This graph was made using Excel. Notice how the axes are labeled, the figure is numbered and titled, the format is very simple, and the data are obvious. Avoid the temptation to add extra grid lines, 3-D features, shadows, backgrounds, etc. Such "chartjunk" distracts your audience from the data, your main means of conveying scientific evidence. The legend (paragraph below or next to the graph) explains how the data were obtained. You can look at any scientific paper for examples of legends. Note that all figures in your textbook also have legends.

Tables. Tables should be made using the same principles outlined for graphs, though the format is different. Tables can be created with Word or Excel. Tables are numbered, but this number usually appears at the top of the table. The title usually follows the table number.

Table 2. Number of pet cats and education levels for Davidson, North Carolina, residents

Subject's initials	Years of formal education	# of pet cats
AN	5	0
CD	12	1
CJL	22	2
CGM	9	1
ABH	7	0

Tables generally do not contain legends. Often, though, footnotes are included under a table to provide explanatory information. Of course, all column headings should be clearly labeled to describe the data listed below them.

When preparing your data for a presentation, think about the most effective way of showing your data to the audience. Some information can be conveyed most effectively in a table. Other information can be conveyed most effectively in a graph. If you do decide to use a graph, then consider what type of graph will be most effective (X-Y, bar, pie, etc.). Always make your graphs and tables as simple and clear as possible. Do not make your reader work hard to understand your point. More information about graphs can be found at http://www.mste.uiuc.edu/courses/ci330ms/youtsey/intro.html.

B. 16th Annual International Wildcatase Convention

The enzyme wildcatase catalyzes the conversion of wildcatin to catin. Previously, researchers determined that catin absorbs maximally at 605 nm, while wildcatin has a minimal absorbance at this wavelength. To characterize the properties of

wildcatase, you have conducted a series of experiments. The scientific community is extremely excited to hear about your results. In fact, the organizers of the 16th Annual International Wildcatase Convention have asked you to present your findings at their meeting in Cancun. Normally, you would decline this invitation, preferring to stay at school and spend some quality time in the library. Because your friend is one of the meeting organizers, though, you agree to attend and present your findings. Of course, now you must organize your data and make graphs and/or tables of your results for the presentation.

Brief descriptions of your experiments and the resulting data are provided below. For each experiment, determine the most effective means of presenting your data and generate a figure in Excel. When everyone is finished, we will look at the graphs each of you generated. Note: for this exercise, you do not need to write legends for your figures.

Experiment 1

After a series of experiments, you have concluded that two different enzymes can catalyze the conversion of wildcatin to catin. You have named these enzymes wildcatase A and wildcatase B. A study of 100 mammalian species has revealed that 50 of these species contain wildcatase A and B, 40 contain only wildcatase A, and the remaining 10 contain only wildcatase B.

Experiment 2

Your studies indicate that both deer mice and field mice contain only wildcatase A. To investigate the location of the enzyme within these organisms, you looked for the presence of wildcatase A in various tissues. You discovered that in the deer mouse, wildcatase A is present in the lung, brain, spleen, and liver tissues. It is not found in muscle, kidney, or heart tissues. In the field mouse, wildcatase A is present in lung, brain, and spleen tissues. It is not found in liver, muscle, kidney, or heart tissues.

Experiment 3

In examining the tissues of the deer mouse, you discovered that the different tissues containing wildcatase A contain different concentrations of the enzyme. Lung contains 350 units of wildcatase A per gram of tissue. Spleen contains 250 units per gram of tissue. Liver contains 100 units per gram of tissue. Brain contains 500 units per gram of tissue.

Experiment 4

To investigate the kinetics of wildcatase A and wildcatase B activity, you first combined 100 units of wildcatase A and 100 mM wildcatin. You then measured the absorbance at 605 nm at 30-second intervals for 2 minutes. This experiment produced the following absorbance data: 0 minutes: 0.010, 0.5 minutes: 0.100, 1.0 minute: 0.200, 1.5 minutes: 0.300, and 2.0 minutes: 0.400. You then did a similar experiment, but used wildcatase B instead. This experiment produced the following data: 0 minutes: 0.010, 0.5 minutes: 0.150, 1.0 minute: 0.300, 1.5 minutes: 0.450, and 2.0 minutes: 0.600.

C. A Brief Introduction to Graphing with Excel

Double click on the Excel icon in the tool bar. When Excel opens, you will see a spreadsheet. In cell A1, enter a name for your X axis values (e.g., "Time"). Enter the appropriate values in column A. In cell B1, enter a name for your dependent variables (e.g., "Abs."). Enter the appropriate values in separate cells of column B. In a similar fashion, enter the values for multiple replicates of the experimental conditions in columns C and D. In cell E1, enter the word "Average." At this point, you should have a spreadsheet that looks similar to the following image:

Using Excel, we can determine the average values for our replicates. Click on cell E2. Next, click on the Insert menu button. Select the Function option and Average from the Function Name selections. You should see a function box similar to the one below.

You might notice that Excel "wants" to calculate the average of the values in cells A2 to D2. Our experimental values, though, are only in cells B2 to D2. Using the mouse, make any necessary changes to these cell designations by highlighting the cells that you want to average (in this case B2, C2, and D2). Click the OK button and the average value of your

triplicates will be inserted into cell E2. To determine the averages for the other triplicate readings, grab the small square in the lower-right corner of cell E2 and drag it down to the end. When you let go, you should have averages in every cell.

To graph the data, highlight the averages you just calculated and click on the Chart Wizard icon on the top menu bar. Select the Chart Type option and then the XY (scatter) graph (step 1 of 4). Click the "next" button to get to step 2. You will then see a graph of your data, but you will notice that the X-axis values are not correct. In step 2 you should click on the "series" button in the top of the window and then click in the box next to "X values:" and then click on the small upward arrowhead to the right of this box. Clicking on the arrowhead will take you to your spreadsheet where you will highlight the cells in the A column (0.5 to 3) that belong on your X axis. When you have highlighted those cells, hit the return key on the keyboard.
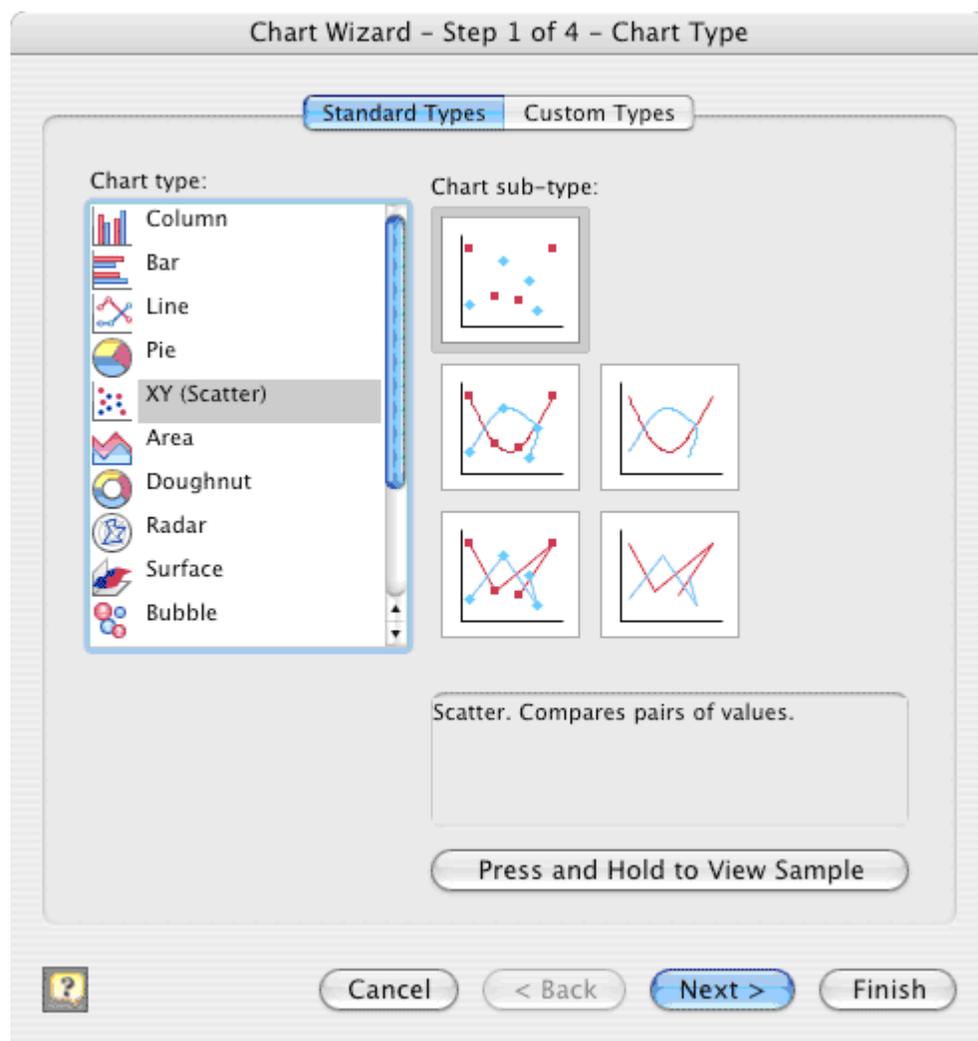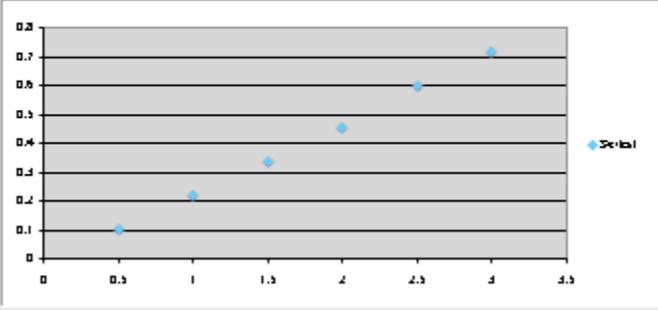
Step 3 of the chart wizard gives you several options to change the appearance of your graph through the buttons at the top labeled "Titles," "Axes," etc. Usually you will want to label your graph under the "titles" button, remove the distracting horizontal gridlines under the "gridlines" button, and remove the key (confusingly called a legend by Excel) if you have only one line graphed (you will want to keep and properly label the key if you have a graph of several lines) under the "legend" button.
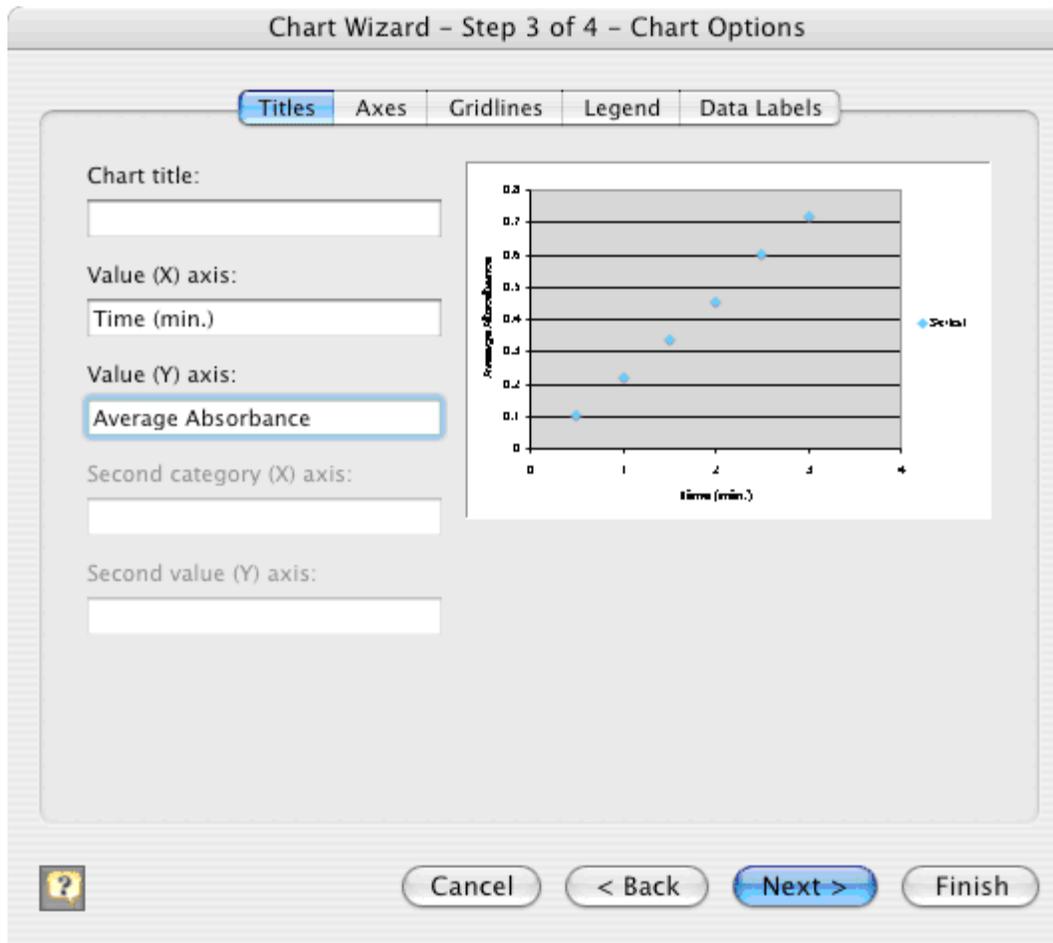
Step 4 of the chart wizard simply asks where you want to save your new graph. After you hit the "finish" button, you will see your graph.

You may notice that your graph might not look the way you would like it to appear. You can adjust many parameters simply by double clicking on specific parts of the graph. For example, to change the color, size, or shape of the data points simply double click on the diamonds and you will get a window that lets you change those parameters. To change the scale of either axis, simply double click on the line of the X or Y axis and you can change the minimum, maximum, major/minor units, etc. To change the font, size, or orientation of the labels simply double click on the text itself.
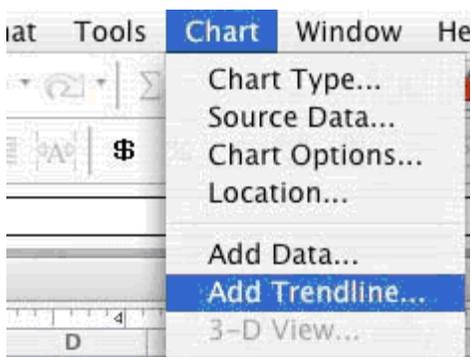
Finally, you can use Excel to determine a best-fit line of your data and the line's slope. After you have completed your graph, click on the Chart button in the top tool bar. Select "Add Trendline," select "linear" and "Series 1," and then click on the "Options" button. Click the boxes for "Display Equation on Chart," and "Display R-squared Value on Chart." Click the "OK" button. The equation for your trendline (in $y = mx + b$ format, where m is the slope and b is the value of the y-intercept) and an R^2 value will then appear on your graph. Briefly, the closer your R^2 value is to 1.0, the better your experimental data fit the trendline you just plotted. Save your completed graph and data worksheet in an appropriate folder with a descriptive name.

Add Trendline

Type   **Options**

Trendline name
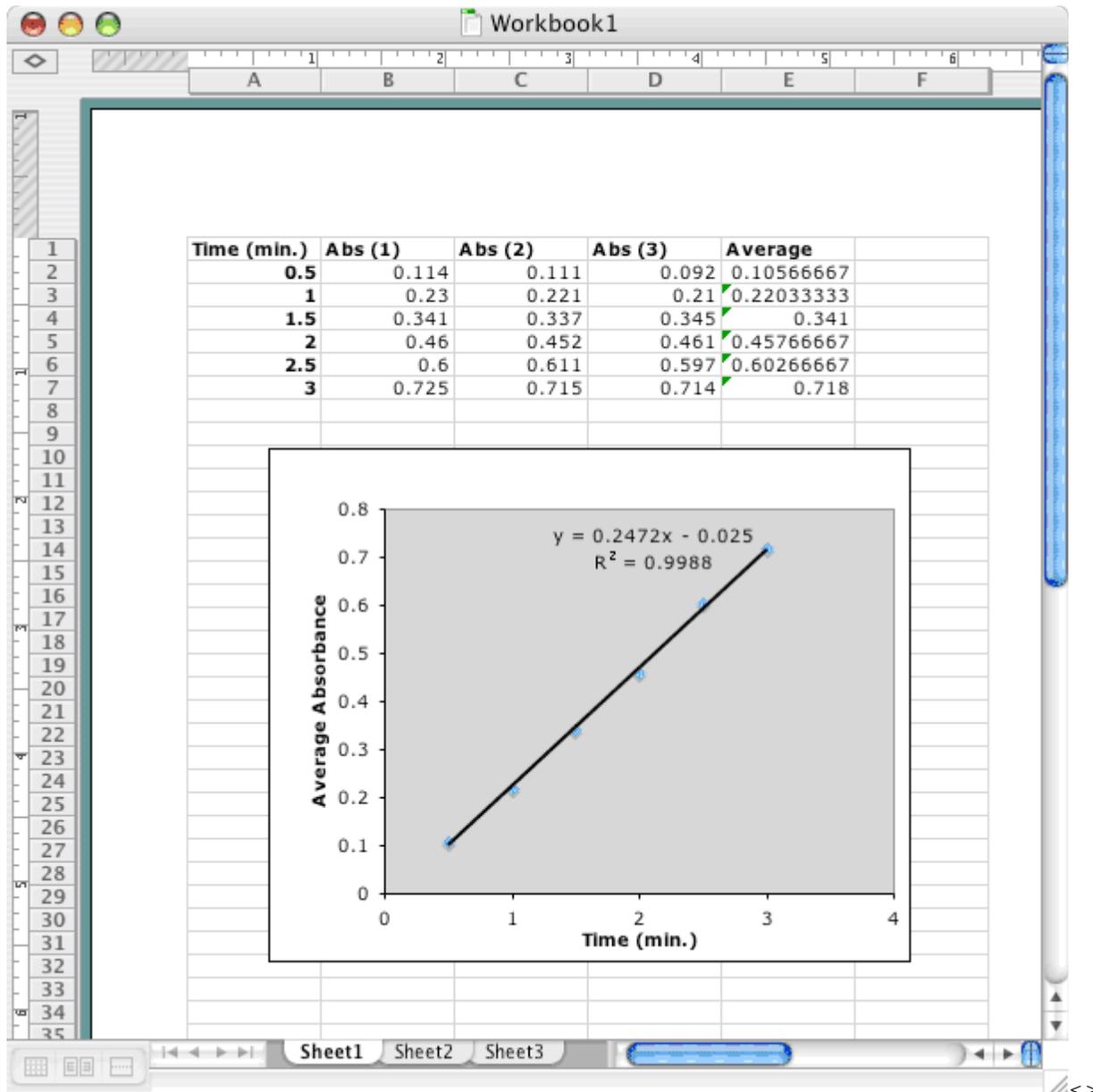
◉ Automatic: Linear (Series1)

○ Custom: [|]

Forecast

Forward: [0]   ▲▼   Units

Backward: [0]   ▲▼   Units

☐ Set intercept =   [0]

☑ Display equation on chart

☑ Display R-squared value on chart

(Cancel)   (OK)

Instructor Version.

In our introductory biology course, students complete this exercise during the third or fourth week of the semester, while completing a lecture and laboratory unit on enzymes and enzyme kinetics. The example scenarios presented here are designed to correspond with this topic, but can be modified to relate more appropriately to other topics and/or laboratory techniques. The exercise is most effective, though, if completed early in the semester, before students analyze or present their own experimental data. The students also should be familiar with the types of assays presented in the hypothetical experiments.

To complete this assignment, students work in pairs, with all students receiving the same experiment scenarios. Generally, a majority of our first-year students have at least a rudimentary knowledge of Excel, but a short guide to graphing with Excel is included in the student handout for students who are not familiar with this software. I explain to them that, when giving a presentation, the data must be presented clearly. No further directions are given. The students first are asked to read through the short introduction to graphs and tables. The students then have about 1 hour to develop four figures, one figure to depict the results of each experiment. After an hour, students are asked to show their figures to the class. Usually, I select a group to show their Figure 1. As a class, we then discuss the figure, focusing on its clarity, use of color, labels, etc. (see the Assessment and Outcomes section for a list of questions asked about each figure). I ask if any of the other groups has a significantly different figure for that experiment. If another group employed a different approach, then they show their figure, and we discuss the strengths and weaknesses of the differing approaches. We repeat this exercise for all four of the experiment scenarios.

Generally, very few problems are encountered during this exercise. The most common problems students encounter involve the use of Excel. If the Y-axis values are inserted in a row rather than a column, for instance, then it is difficult to generate an appropriate graph. When students follow the short primer on Excel, these problems should be minimal.

Safety Issues.
None.

ASSESSMENT and OUTCOMES

Suggestions for Determining Student Learning.

The assignment itself intentionally is not graded. It is designed to be a learning experience, where students can feel free to work together on the given problems without the added pressure associated with a grade. As a class, though, we do critique the graphs made by the students and discuss the strengths and weaknesses of the different graphs. For each graph presented by students, we ask the following questions:

·   Does the figure have a title?

·   Are the axes labeled?

·   Is the independent variable on the X axis and the dependent variable on the Y axis?

·   Are the units indicated?

·   Are the axes' scales reasonable?

   ·   Should the Y axis go from 0 to 1 or 0 to 5? (for instance)

   ·   Should the Y axis be linear or logarithmic? (for instance)

·   If there are multiple bars or lines on the graph, are they easily distinguishable?

·   Is the text easy to read?

·   Does the graph convey the most important aspects of the data?

In Appendix 1, representative student graphs and tables are presented. Following each figure is a list of comments made about the figure during our group discussion.

While this assignment is not graded, the ability to interpret data and present data in a clear manner is tested subsequently during the semester. In conjunction with this exercise, students conduct a series of experiments that, in some respects, mirror the hypothetical experiments described in this exercise. At the completion of this unit, students must hand in a written report and give an oral presentation to the class. Grading of these assignments is based, in part, on the quality of the figures. I pay special attention to the qualities that we addressed in class (see above and Appendix 1). Additionally, a question on the first test presents the students with a hypothetical experiment and resulting raw data and requires the students to produce an appropriate figure for their answer. I expect the students to be able to interpret the data and create an appropriate figure. Specifically, I grade each graph or table based on the questions listed above (e.g., Does the figure have a title? Are the axes labeled?). I deduct points if the students do not follow the format we discussed during this exercise. In these ways, the ability of each student to analyze data and construct appropriate figures is assessed.

Field Testing.

This exercise has been used in an introductory biology course for the past 4 years. Initially, I used it in my section of 32 students. In subsequent years, other faculty members teaching sections of this course also have adopted the assignment. Currently, approximately 90 students per year are exposed to this method of learning about data presentation. Most of these students are first-year undergraduates.

The exercise itself can be completed within 1 to 2 hours. Usually, I arrange the laboratory schedule such that we can conduct this introduction to graphing after the students have completed a series of enzyme kinetics experiments, but before they have written laboratory reports or given oral presentations. Immediately upon completion of the graphing exercise, the students have time in the lab to examine their real experimental data and begin to construct figures for their presentations. Their written laboratory reports and oral presentations are due the following week. By arranging the schedule in this manner, I have found that the students have fewer questions about their graphs, and the figures and presentations, in general, are better.

The students also seem to enjoy the exercise. And many of them comment later in the semester about the hypothetical wildcatase enzyme.

Student Data.

Generally, most groups construct reasonable figures for all four of the hypothetical experiments. The most common issues that we discuss are which variable goes on which axis, appropriate labeling of the axes, and the use of easily distinguishable markers or shading schemes when multiple lines or bars are on a graph.

For some groups, Experiments 1 and 2 present some problems. Students often attempt to use a Venn diagram, with limited success, to present the data in Experiment 1. Also, students often have difficulty constructing a good table to present the data in Experiment 2. Representative figures, good and bad, and selected criticisms are presented in the appendix.

Modifications and Extensions.

This exercise can be modified in many ways and integrated into many different courses. Most importantly, the hypothetical experiments can be altered to fit better into a specific course. It is important that the students are familiar with the types of assays described in the hypothetical experiments. It also is important that this exercise is introduced early in the semester, before students have written their first laboratory reports. Because we begin our course with a unit on enzymes, enzyme kinetics, and spectrophotometry, our hypothetical experiments involve using spectrophotometry to measure enzyme activity. Quite easily, though, different hypothetical experiments could be developed. If the first unit in a course involves bacterial replication, for instance, one could develop a series of experimental scenarios relevant to this topic. One set of data could examine growth rates of *Escherichia coli* in the presence of different concentrations of glucose, while another set of data could examine growth rates of different mutant strains of *E. coli*. If the first unit in a course involves microbial diversity and ecology, one set of data could examine the number of different bacterial species present in a particular habitat, while another set of data could examine the changes in species diversity over time. Let your imagination run wild!

In addition to modifying the hypothetical data to fit a specific course, the format of the exercise can be modified to fit your resources. Because we have computers in the laboratory, we have students complete the entire exercise during a regularly scheduled laboratory period. If a computer lab is not available, then the instructor could describe the assignment and discuss the basics of graphing in class. Students then could complete the graphs outside of class and email their graphs to the instructor and/or bring hard copies to the next class.

Several other extensions to this exercise could be implemented. First, an instructor could include a peer-review activity. Rather than having the whole class view the various graphs and comment on them, individual groups could exchange their graphs with another group. Each group, then, would be responsible for providing feedback to their peers on the quality of the graphs. Several studies have demonstrated the effectiveness of this type of peer review (1, 3). Second, the instructor could have students write figure legends for their graphs. This extension of the exercise certainly would require more time, most likely out of class time. It would, though, provide the students with great writing experience.

References.

1. **Blair, B. G., G. R. Cline, and W. R. Bowen.** 2007. NSF-style peer review for teaching undergraduate grant-writing. Am. Biol. Teacher **69:**34–37.

2. **d'Apollonia, S.** 2006. Science students' document literacy skills, p. 2465. *In* Proceedings of the 28th Annual Conference of the Cognitive Science Society. Cognitive Science Society, Wheatridge, CO. http://www.cogsci.rpi.edu/CSJarchive /Proceedings/2006/docs/p2465.pdf.

3. **Gratz, R. K.** 1990. Improving lab report quality by model analysis, peer review, and revision. J. Coll. Sci. Teaching **19**(5)**:**292–295.

SUPPLEMENTAL MATERIALS

APPENDIX 1.

Representative Student Data.

Following are several representative student figures generated as part of this exercise. For each, selected comments, positive and negative, are provided.

Experiment 1. Most students display these data in a bar graph (Fig. 1a) or a pie chart (Fig. 1b). Occasionally, students will attempt to display these data in a Venn diagram (Fig. 1c).

Experiment 2. Students generally realize that these data can be displayed most effectively in a table (Table 2). Some groups construct a bar graph (Fig. 2).

Experiment 3. Students generally present these data in a bar graph (Fig. 3).

Experiment 4. These data usually are presented as a line graph (Fig. 4). Because the hypothetical data are perfectly linear with respect to time, a "connect-the-dots" line graph works with these data. As a class, we talk about the imperfect nature of experimental data and the usefulness of an X-Y scatter plot and trendline.
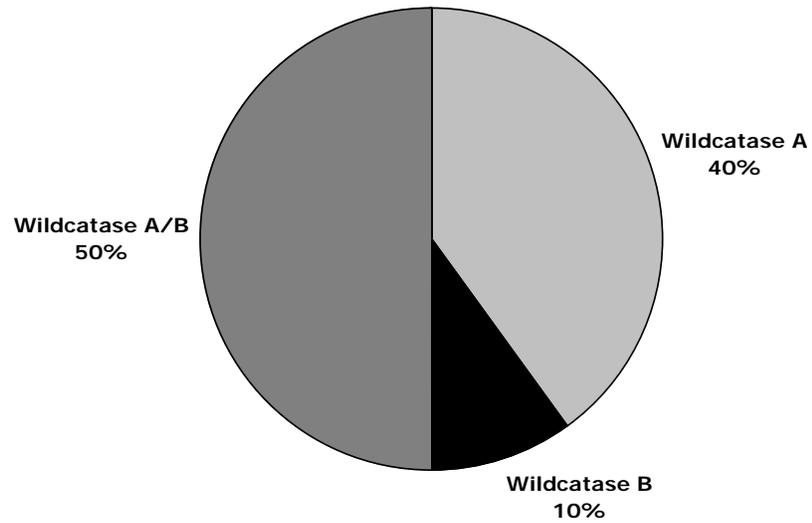
Figure 1a

Species Distribution of Wildcatase A and B
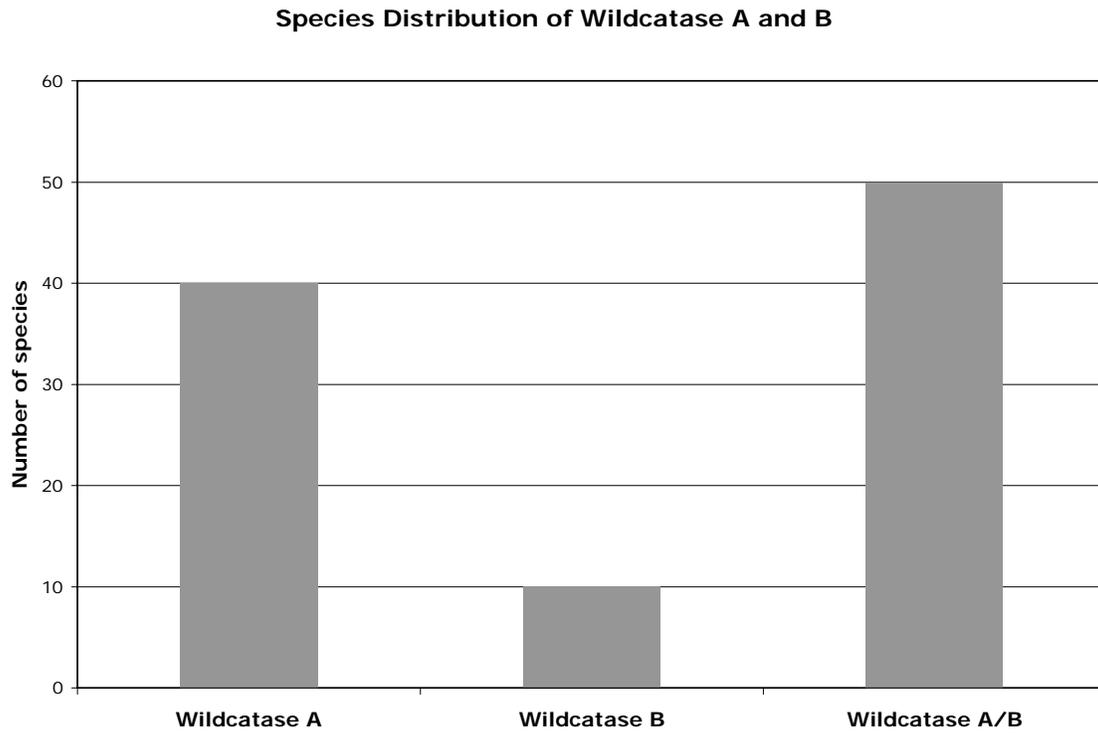
Selected comments:
- Clear and easy to interpret
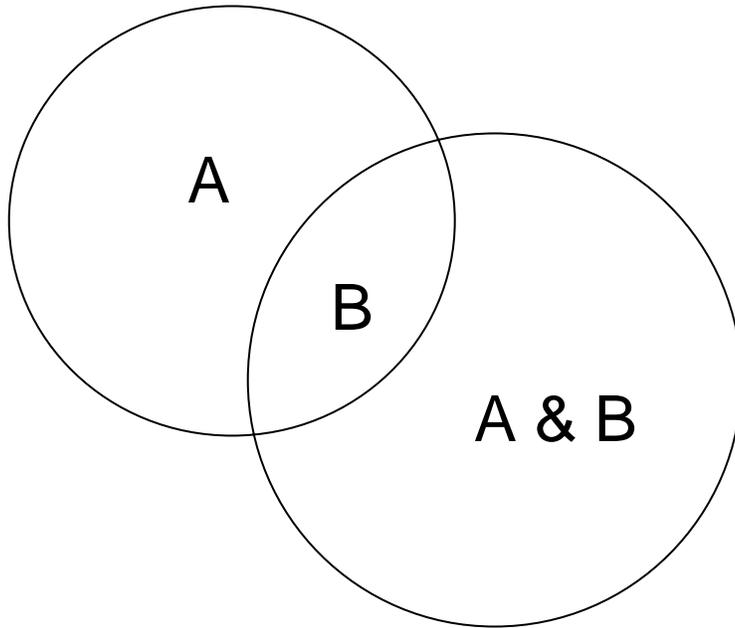- Numbers on Y axis are faint and should be bold

Figure 1b

Species Distribution of Wildcatase A and B

Selected comments:
- If the total number of species did not equal 100, a pie chart would be much less useful
- The shadings may be difficult to differentiate
- Generally not easy to quickly interpret

Figure 1c

Selected comments:
- No title
- No indication what "A", "B", and "A & B" refer to
- Does not accurately depict the data. Diagram indicates that "B" is present in "A" and "A & B".

TABLE 2.

Organism	Lung	Brain	Spleen	Liver	Muscle	Kidney	Heart
Deer mouse	+	+	+	+	-	-	-
Field mouse	+	+	+	-	-	-	-

Selected comments:
- Very clear and easy to read
- Could make column heads bold to emphasize them
- Should include a title
- Are (+) and (-) the most effective markers?

Figure 2

Tissue Distribution of Wildcatase A
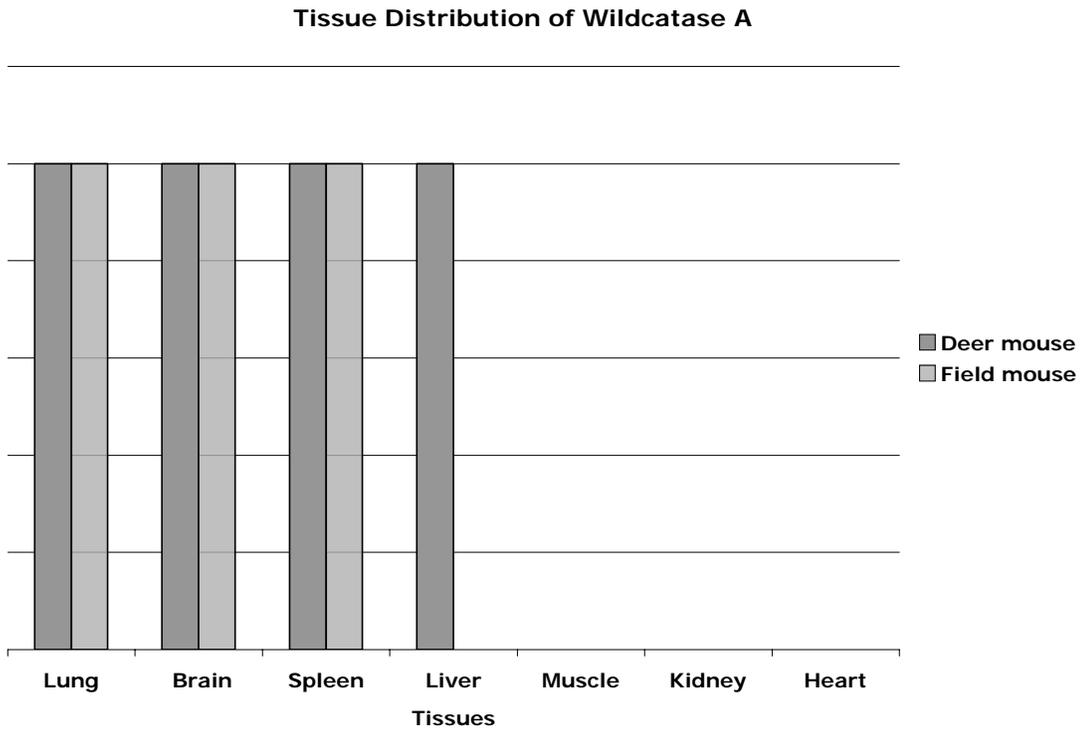
Selected comments:
- Shading may be difficult to differentiate
- Bar graph generally indicates a quantifiable measurement. As drawn, it appears that the same amount of wildcatase A is present in lung, brain, spleen, and liver. But the description only indicates if the enzyme was present in the tissues.
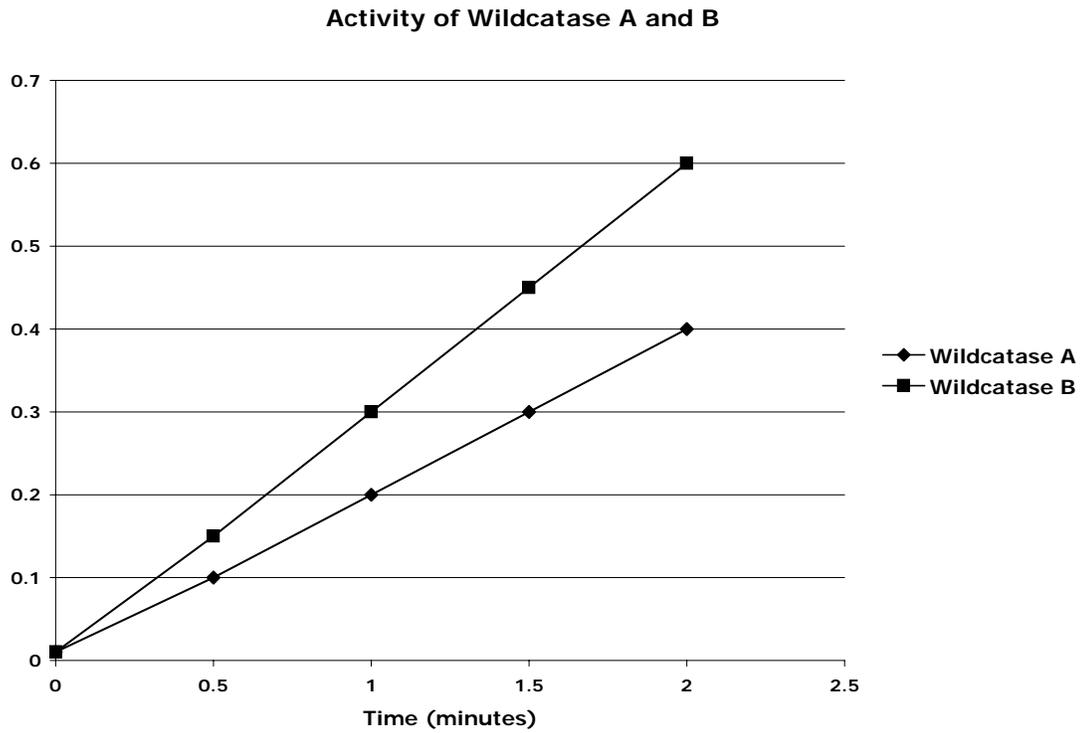- Is the title descriptive enough?

Figure 3

Wildcatase A in mouse tissues
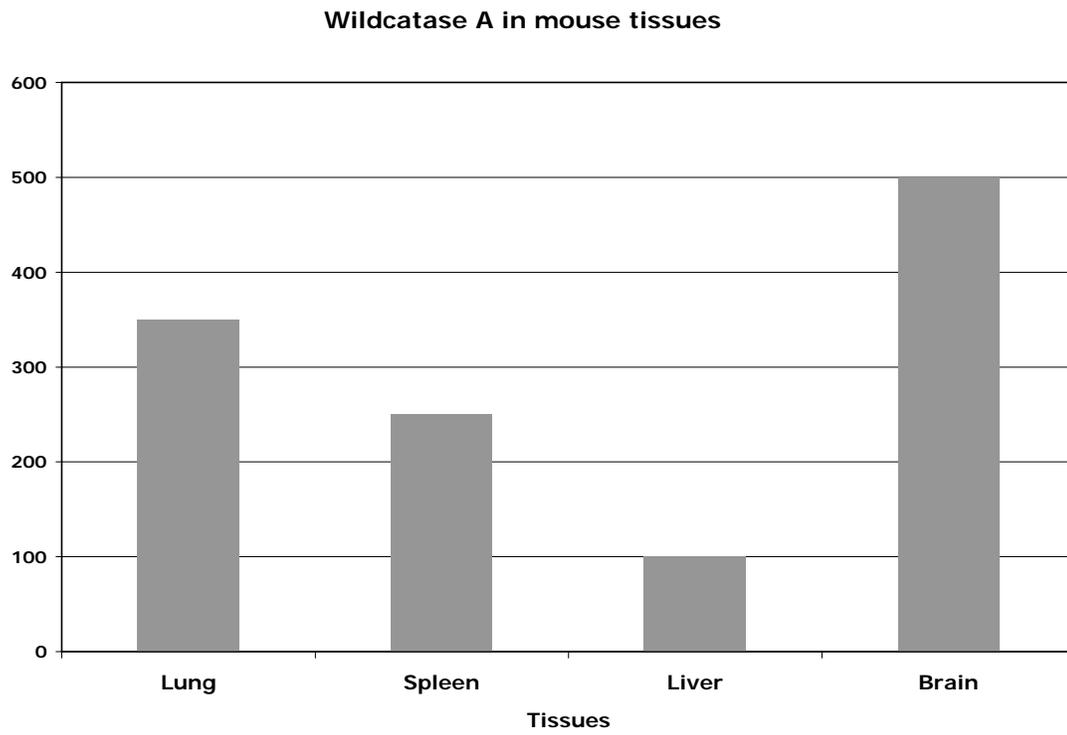
Selected comments:
- Clear and easy to interpret
- Y axis should indicate Units/gram

Figure 4

Activity of Wildcatase A and B

Selected comments:
- Very nice graph
- Easy to interpret
- Markers for "A" and "B" may be difficult to differentiate. Could use open versus closed markers or solid versus dashed lines