

Data Literacy Lesson Plan: Working with spreadsheets, displaying data and calculating r^2 value

Overview:

You've gone on a field trip and collected data. Now what can you do with it? Learn what a scientist does with data.

Materials/ Prep Needed:

- Data Literacy Homework.pdf (distribute a day or two before classroom visit)
- A laptop with Geogebra (free at www.geogebra.org), Excel, or Numbers installed
- Projector
- Data sheets (collected by students or sample data)
- Spreadsheet with column headers emailed to students (optional)
- Students need laptops

Activity	Time for 50-min class	Time for 90-min class
I Introduction	5	8
II Data Quality	5	7
III Demonstration of Data Entry Using Numbers or GeoGebra; tips; student data entry; Calculating sum, mean, average etc.	15	20
IV Determining type of graph to represent relationships/ going over homework sheet.	10	10
V Creating Bar Graph, Scatter Plot, Pie Chart in GeoGebra or Numbers	10	15
VI Creating Histograms in GeoGebra or Numbers	--	15
VII Calculating mean, r-value, what do these tell you?	5	5
VIII Vocab/ Review/ Questions	--	10

I Introduction: Project some data sets, and ask if they can think of questions that can be asked of them.

II Data Quality

“When you get back to the lab or classroom with data you collected on a field trip you will realize how important it is to have good quality data. In order to entering data into a spreadsheet, you'll notice how important it is to write legibly. Your notes may be clear to you, but confusing to someone else.”

Show example of hatchmarks/ totals interspersed on a data sheet. Someone would not know if two lines are hatch marks or an eleven. Ask students what they would do if they saw that on a page. Guessing would undermine the quality of the data. If written data is ambiguous, it can't be used in your analysis.

If applicable to the dataset, ask what “too numerous to count” means. Show what happens to graphs when TNTC is included. Try to find ways to improve the quality of the data, by defining TNTC as greater than x, and in the future, trying to get an accurate count by dividing the field of view into quadrants.

- Every data *sheet* should have a date.
- All original data sheets are valuable. Never let originals leave the lab.

III Data Entry / Using Spreadsheets

Email spreadsheet (Geogebra or Numbers format) with column names to all students (can be done beforehand). Project a laptop using Geogebra or Numbers. Demonstrate entering a row of data into a spreadsheet. Each student gets a page of data, enters it, and passes it down, until they have four of five rows filled in. Go over tips for using Geogebra or Numbers.

- Numbers: ctrl-click: add new column
- “Copy special” allows you to copy values rather than formula.
- Geogebra: Select cell or group of cells, click down on square at corner, drag across cells, release mouse- fills in rows or columns with the same numbers. This is called Relative Copy.

“Once you have the data entered, you can begin to use formulas in cells, columns or rows, to do calculations with the data.” Demonstrate calculating average and mean. Tips for calculations:

- Drag edge of cell with average formula all the way across a row and it calculates all averages.
- Averages can be misleading because of zeroes, but they're important for calculating density.

Vocab(project if they don't already know)

Mean: The mean is the average of the numbers: a calculated "central" value of a set of numbers. To calculate: Just add up all the numbers, then divide by how many numbers there are.

Median: The middle number (in a sorted list of numbers). To find the Median, place the numbers you are given in value order and find the middle number.

IV What kind of Graph should I use?

Students work in groups to answer these questions displayed on screen:

What is the question you would like to know about the data? _____

Is that question about relationship, comparison, distribution, or composition? _____

How many variables are involved? _____

Is one of the variables time? _____

Project the Chart: “choosing a good chart 09.pdf”

Discussion: Have students volunteer to share their answers to the above questions. (make note of what kind of questions they come up with). If the ones below didn't come up, ask what chart they would use to answer them:

- a) Percent of total organisms found on eelgrass: Pie Chart
- b) relationship between dissolved oxygen and numbers of phytoplankton: scatter plot.
- c) comparison of numbers of two types of phytoplankton over time: line graph.

V Demonstrate Creating Bar Graph/ Line Graph/ Scatter Plot/ Pie Chart and calculating r^2 -value in GeoGebra or Numbers.

Scatter Plot and r^2 -value instructions:

1. Arrange your data in an spreadsheet so that one column has data that you would like displayed on the x axis, and the column next to it has data you would like displayed on the y axis.
2. Select the two columns of data you would like to graph.
3. Under the scatter plot drop-down (on the Insert tab in Excel 2007), choose “Scatter with only Markers.” A scatter plot of the data will be generated.
4. Next, click a cell, and type in the function =CORREL(array1, array2). Once you type =CORREL(you can click and drag your mouse from the top cell in the column of your data that is displayed on the x-axis to the bottom cell in that column, and excel will automatically fill in the proper format of the array, C1:C9, for example. Next add a comma to the formula and then drag your mouse from the top cell of the column of your data that is displayed on the y-axis to the bottom cell of that column, and excel will automatically fill in the proper format of that array, D1:D9, for example. You can also type this in manually so it looks like this: “=CORREL(C1:C9,D1:D9)”.
5. Click enter and your r^2 value will be displayed in the cell.
6. R^2 values vary from -1 to 1. Generally, the closer the number is to either 1 or -1, the stronger the relationship between the two variables, either positive or negative.

VI Demonstrate creating a histogram using Excel.

Steps in Excel 2003

1. List bin limits (the upper value for each range of data you want to display) into a column in an empty area in the spreadsheet, with an empty cell to the right of each cell used.
2. Select the list of bin limits and the empty cells immediately to their right, as shown.
3. Click on Data Analysis under the Tools menu
4. Choose Histogram from the list.
5. Enter the range of cells containing the data to be sorted into the input field. You can do this by clicking on that field, and then clicking and dragging from the first cell to the last to highlight the data in the column you are analyzing.
6. Repeat this process for the bin range field, highlighting the range of cells that contain the bin limits.
7. Repeat this process for the output range, highlighting the empty cells to the right of the bin limits.
8. Select the radio button next to “New Worksheet”, and the checkbox next to “chart output”.
9. The results will be on a new sheet in the excel workbook.

Size	Histogram Data	
	Bin Limits	Frequency
	0.5	
	1	
	1.5	
	2	
	2.5	
	3	
	3.5	
	4	
	4.5	
	5	
	5.5	
	6	
	6.5	
	7	

Vocabulary: (project)

R²-value: R² is most often seen as a number between 0 and 1.0, used to describe how well a line fits a set of data.

Line Graph: A graph that uses points connected by lines to show how something changes in value (as time goes by, or as something else happens).

Pie Chart: A pie chart is a circular chart divided into sectors, each sector shows the relative size of each value.

Bar Graph: A graph drawn using rectangular bars to show how large each value is. The bars can be horizontal or vertical.

Histogram: A graphical display of data in which the data is grouped into ranges (such as "40 to 49"), and then plotted as bars. It is similar to a Bar Graph, but each bar represents a range of data.

Scatter Plot: A plot of data points in which each point represents the simultaneous value of two variables, usually with the independent variable on the x-axis and the dependent variable on the y-axis.

Frequency Plot: A plot of the data distribution. The data values of the variable being plotted are on the x-axis, a count or percentage is on the y-axis. Each point on the plot indicates the number or percentage of the data points that have that value. The Gaussian or bell-shaped "normal" curve is a frequency plot.