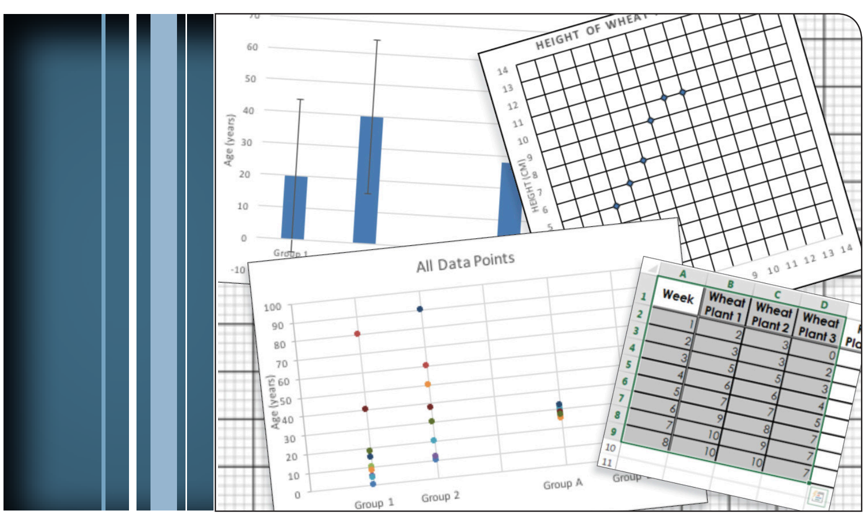
Introduction to Graphing



Investigation Manual

CAROLINA
DISTANCE LEARNING

INTRODUCTION TO GRAPHING

Table of Contents

Overview

Scientific investigation requires the analysis and interpretation of data. Knowing how to graph and what the different components mean allow for an accurate analysis and understanding of data. In this investigation you will practice creating graphs and use some simple statistical tools to analyze graphs and datasets.

Objectives

- Create graphs from datasets, both by hand and electronically.
- Analyze the data in the graphs.
- Compare the slope of trendlines to interpret the results of an experiment.

Time Requirements

Activity 1: Graphing by Hand .... 20 minutes
Activity 2: Computer Graphing...................................... 20 minutes
Activity 3: Linear Regression .... 20 minutes

Key

Personal protective equipment (PPE)

goggles gloves apron

follow link to video

photograph results and submit

stopwatch required

warning corrosion flammable toxic environment health hazard

Made ADA compliant by NetCentric Technologies using the CommonLook® software

Background

Science requires the collection of data to test hypotheses in order to see if it supports or does not support ideas behind the experiment. Collecting data creates a record of observations from experiments that is needed to ensure the ideas in a hypothesis are accurate. This allows the scientist to better understand the processes they are investigating. Sharing data is critical since it allows other scientists to examine the experimental setting and draw conclusions based on the data obtained. It also allows for the replication and comparison of data obtained in the experiment to confirm results and conclusions. This will aid in the understanding of a scientific principle.

Table 1, shows data from a study of plants. Two types of plants, wheat and rye, were grown over 8 weeks, and the height of the plants were measured in centimeters (cm).

The aim of this experiment was to examine growth rates of the two plant types in comparison with each other in order to find out which grows under a certain set of environmental circumstances.

When looking at an experiment, the experimenter is typically looking at variables that will impact the result. A **variable** is something that can be changed within an experiment. An **independent variable** is something the experimenter has control over and is able to change in the experiment. Time can be a common independent variable as the total duration of the experiment can be changed or the intervals at which data is collected can be changed. **A dependent variable** changes based on its association with an independent variable. In the data from Table 1, the measured height of the plant was the dependent variable. The aim of

Table 1.

	Height in cm					
Week	**Wheat Plant 1**	**Wheat Plant 2**	**Wheat Plant 3**	**Rye Plant 1**	**Rye Plant 2**	**Rye Plant 3**
1	2.0	3.0	0.0	0.0	1.0	0.0
2	3.0	3.0	2.0	1.0	2.0	1.0
3	5.0	5.0	3.0	1.0	2.0	2.0
4	6.0	6.0	4.0	2.0	3.0	3.0
5	7.0	7.0	5.0	3.0	4.0	3.0
6	9.0	8.0	7.0	3.0	4.0	3.0
7	10.0	9.0	7.0	4.0	5.0	4.0
8	10.0	10.0	7.0	5.0	6.0	5.0

continued on next page

Background continued

experiments is to determine how an independent variable impacts the dependent variable. This data can then be used to test the hypothesis which has been made at the beginning of the experiment.

Data can be presented in different ways. One way is to organize it into a table as it is being collected. When working with a limited amount of data points, this can be the best option; for larger studies, the data in data tables can be overwhelming and difficult to interpret. To help see the trends in large data sets, a scientist may rely on summary statistics and graphical representations of the data.

Summary Statistics

Summary statistics are methods of taking many data points and combining them into just a few numbers. The most common summary statistic is an **average,** or **arithmetic mean. An average is the sum of a group of numbers, divided by how many numbers were in the set.** To find the arithmetic mean you find the sum of the data to be averaged and divide by the number of data points. For instance: If we wanted to find the average wheat plant height in week 8 from the above data we would perform the following calculations:

Equation 1:

$$\text{average} = \frac{x_1 + x_2 \ldots + x_n}{n}$$

In this equation, x_1 indicates the first number in a data set, x_2 would be the second number, and so on. x_n is the last number in the set. The "n" is the number of items in the set. So a dataset with 8 numbers would go up to x_8. This is the same "n" that the sum of the numbers is divided by. Using equation 1 for the wheat plant height in week 8 would give the following equation.

$$\text{average} = \frac{x_1 + x_2 + x_3}{3}$$

Since there are 3 wheat plants in week 8, there are 3 numbers that would be added together (x_1 x_2 x_3) divided by the number of plants (3).

$$\text{average} = \frac{10 + 10 + 7}{3} = 9 \text{ cm}$$

In science it is important to know how much variation is found in the data collected. The most common measurement of variation is the **standard deviation**. To calculate the standard deviation:

1. Calculate the average of a data set.
2. Calculate the difference between each data point and the average.
3. Square the result.
4. Find the average of these squares. This yields the variance (σ^2).
5. Taking the square root of the variance gives the standard deviation (SD) as seen in Table 2.

The standard deviation is an indication of the distribution of your data. In the example above, the average height of the plants was 9 cm. The standard deviation was 1.4 cm. Statistically this indicates that 68% of the data was within 1.4 cm of the average. In this way it is a useful tool to gauge how close the results on an experiment are to each other.

continued on next page

Table 2.

	Height at Week 8 (cm)	Difference from Average	Difference Squared
Wheat Plant 1	10	$10 - 9 = 1$	$(1)^2 = 1$
Wheat Plant 2	10	$10 - 9 = 1$	$(1)^2 = 1$
Wheat Plant 3	7	$7 - 9 = -2$	$(-2)^2 = 4$
Average	$\frac{(10 + 10 + 7)}{3} = 9$ cm	Variance	$= \frac{(1 + 1 + 4)}{3} = 2$
		Standard Deviation	$= \sqrt{variance} = \sqrt{2} = 1.4$

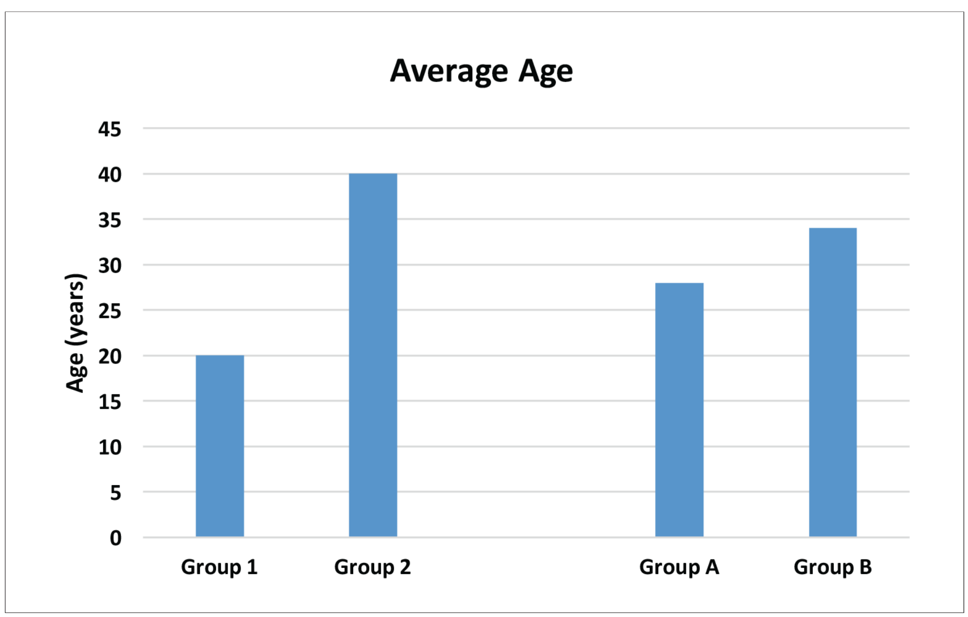
Interpreting Graphs in Scientific Literature and Popular Press

Graphs are an excellent way to summarize and easily visualize data. Care must be taken when interpreting data from a graph or chart. Information can be lost in summarization and this may be critical to our interpretation. For example, in **Figure 1**, the average age of 4 groups of people was graphed using a bar graph. A bar graph is most useful when directly comparing data as it allows for differences to be more easily seen at a glance. Looking at **Figure 1**, it is tempting to conclude that the difference between Groups 1 and 2 is much greater than between Groups A and B.

However, if we look at a graph of all the data that went into the average age, we can see that the variance in Groups 1 and 2 is much greater than in Groups A and B (**Figure 2**).

This information can be conveyed in the graph by the use of **error bars**. Error bars are a graphical representation of the variance in a dataset.

Figure 1.



continued on next page

Background continued

The chart below uses the standard deviations from the data to show the variance of the data. There are multiple ways to represent variance, so it is important that the caption of the figure tells the reader what measure is being represented by the error bars (**Figure 3**).

Standard deviation is highly influenced by **outliers**, or data points that are highly unusual compared to the rest of the data, so scientists frequently use **confidence intervals** to represent variance on graphs. Confidence intervals express the probability that a data point will fall within the error bars, so error bars with a 99% confidence interval say that 99% of the data will fall between the error bars. Confidence intervals are typically published at 99%, 95%, or 90%. The main point is that when error bars overlap, as they do when comparing Group 1 with Group 2, it is not strong evidence that there is a difference between the two groups, even if the averages are far apart. A real difference is more likely between Group A and Group B.

Figure 2.

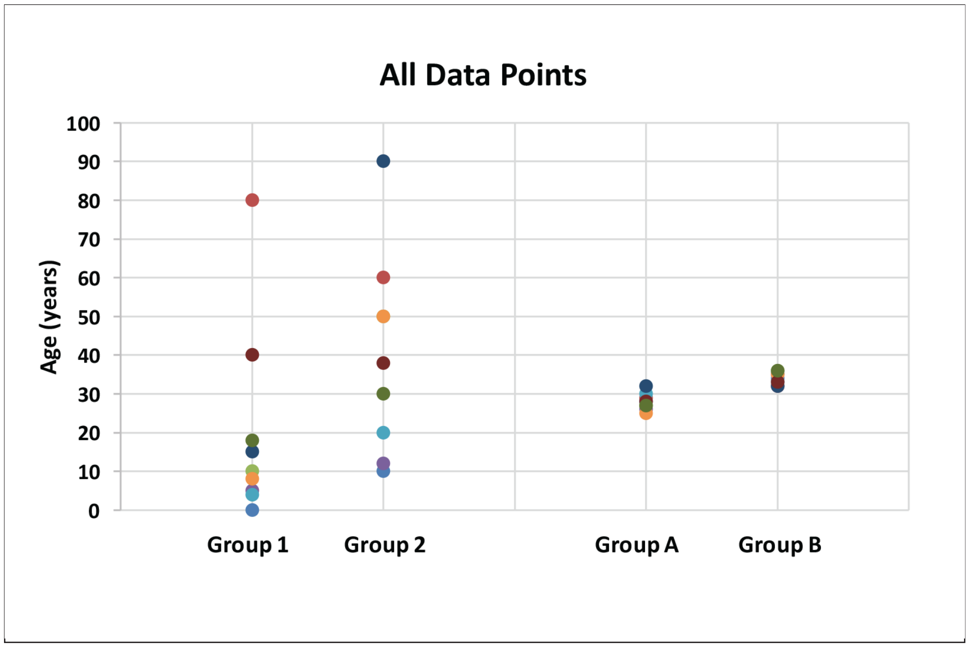
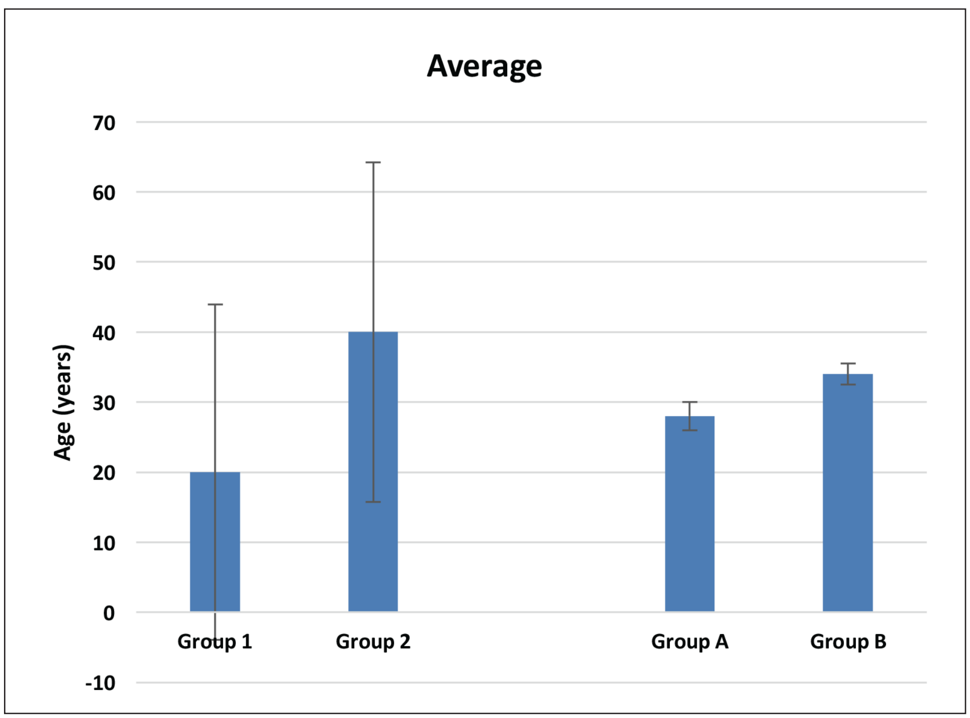


Figure 3.



ACTIVITY

Materials

Needed but not supplied:
- Graphing Software (Excel®, Open Office®, etc.)
- Printer to print graphing paper

Safety

There are no safety concerns for this lab.

ACTIVITY 1

A Graphing by Hand

A common method to look at data is to create an x,y scatter graph. In this first activity, you will create two graphs of the data from Table 1.

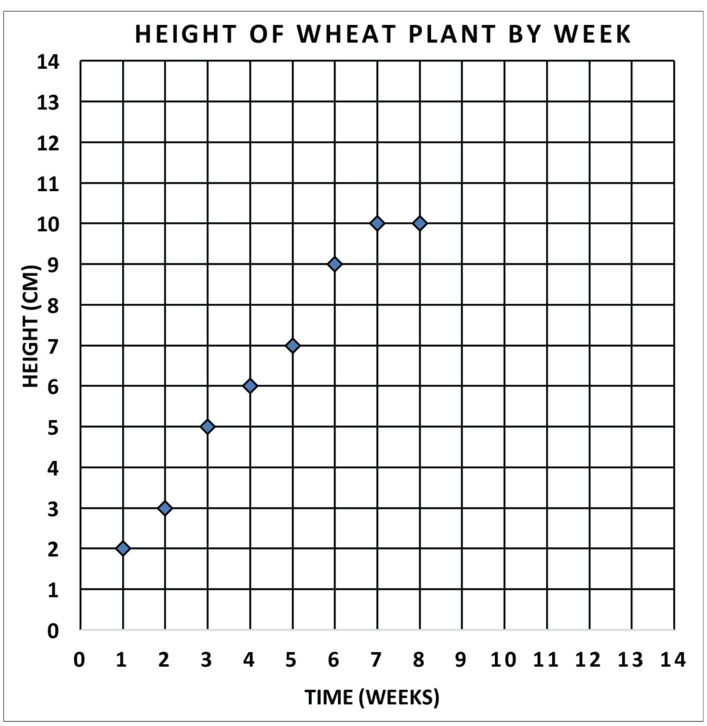
1. Print 2 copies of the graphing sheet found on page 13.
2. Title the first graph "Wheat plant height by week."
3. Title the second graph "Rye plant height by week" and set aside for later.
4. At the bottom of the graph there is a space to label the x-axis. The x-axis runs from left to right, with smaller numbers starting on the left and the numbers increasing as you move to the right.
5. At the left of the graph there is a space to label the y-axis. The y-axis runs from the bottom to top of the graph, with smaller numbers starting at the bottom and the size of the numbers increasing as you move up.
6. You will now label each axis and decide which pieces of data will be our x-values and our y-values, respectively.
7. One method to determine which data should be your x versus y axis is to think about the goal of the experiment. The y-axis should be for data that you measured for, the dependent variable. In the data set in Table 1, the scientists were measuring the height each week. This means that the height is the dependent variable.
8. Label the y-axis "Height (cm)." It is important to always include the unit of measurement on the axis. In this case the unit is centimeters (cm).

continued on next page

ACTIVITY 1 continued

9. The x-axis is the independent variable, the parameter of the experiment that can be controlled. In this experiment the scientists were controlling when they measured the height.
10. Label the x-axis “Time (weeks).” This indicates that a measurement was taken each week.
11. Locate the lower left corner of the graph. This will be the origin of your graph. The origin on a graph is where both the values of x and the values of y are 0. If the numbers in a data set are all positive (i.e. there are no negative numbers) it is a best practice to set the origin in the lower left corner. This allows the view of the data to be maximized.
12. The axes then need to be numerically labeled. Referring to **Figure 4**, label each axis from 0 to 14 along the darker lines.

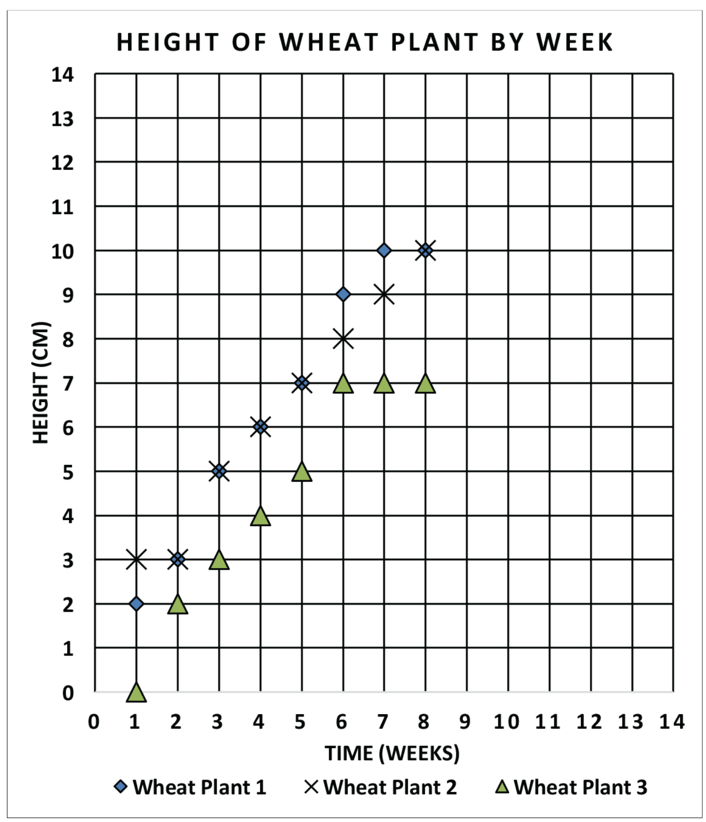
Figure 4.



13. Starting with the “Wheat Plant 1” data in **Table 1**, count over 1 (for week 1) on the x-axis for time, then count up to 2 from there to indicate 2 cm. Place a dot at this point
14. Repeat this process for the remaining data points for “Wheat Plant 1.” Your graph should now look like **Figure 4**.
15. Using this same process, graph the data for “Wheat Plant 2” and “Wheat Plant 3” on the same graph. You will need to be able to distinguish the data from each set from each other. Use different colors, or symbols to make this differentiation.
16. When complete, compare your graph to **Figure 5**. Your exact colors or symbols may be different, but the data should be in the same locations.

continued on next page

Figure 5.



ACTIVITY 1 continued

17. You can now create a legend. The legend is what shows another person what the points on your graph represent. Refer to **Figure 5** for an example legend for this graph.
18. Create your legend. It is below the x-axis label as in **Figure 5**. The legend can be anywhere on the graph, so long as it does not interfere with the reading of the graph.
19. Create your own graph of the data for "Rye plant height by week." Use the process outlined in this activity to graph all of the data for each plant.

ACTIVITY 2

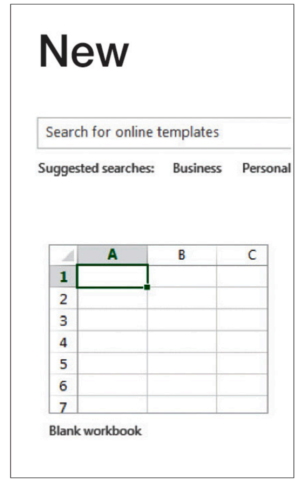
A Computer Graphing

Graphing by hand can be useful for observing trends in small data sets. However, as the quantity of the data grows it can be useful to graph using a computer. This activity will give a general outline of how to graph on a computer. **Please note Microsoft Excel® was used to generate the figures for this activity. Your exact software may look different or have slightly different labeling than what you will see here. You may need to refer to the documentation of your exact program to determine how to perform a particular step.**

In this activity you will graph the data from Table 1 into your computer.

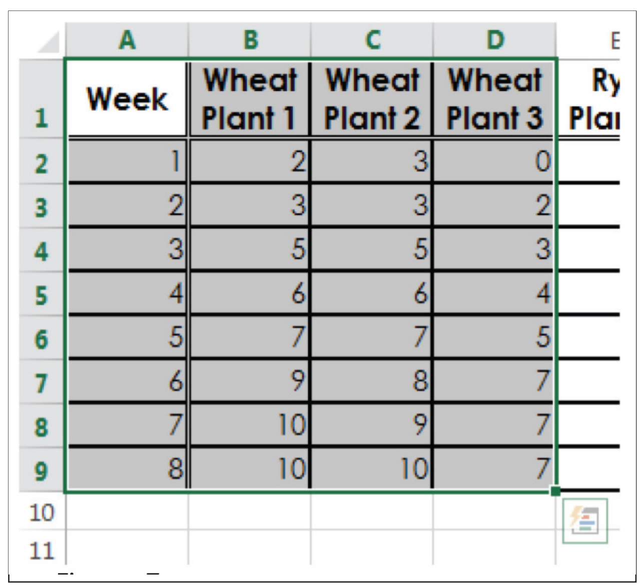
1. Open a new workbook. This will open a new sheet (**Figure 6**).
2. You will see a large sheet with lettered columns and numbered rows. These letters and numbers can be used to refer to a specific cell (the box where information can be typed.) For example the upper left cell is A1 representing column A, row 1.
3. Starting in cell A1 type "Week." In cell B1 type "Wheat Plant 1." Continue across putting each title in a new cell in the first row.

Figure 6.



4. Move to row 2. Type the corresponding numbers under the correct column.
5. Continue until your table looks like Table 1.
6. Select the data for Week thru Wheat Plant 3. You can do this by clicking on cell A1 and then dragging down and over to cell D9. All of the data and titles should be selected for the wheat plant (**Figure 7**).

Figure 7.



	A	B	C	D	E
1	Week	Wheat Plant 1	Wheat Plant 2	Wheat Plant 3	Ry Plai
2	1	2	3	0	
3	2	3	3	2	
4	3	5	5	3	
5	4	6	6	4	
6	5	7	7	5	
7	6	9	8	7	
8	7	10	9	7	
9	8	10	10	7	
10					
11					

continued on next page

ACTIVITY 2 continued

NOTE: The next several steps may vary greatly depending on the exact software you are using, but the goal is the same.

7. Find the menu labeled "Insert."
8. Among the "Charts" find "Scatter," or "x,y Scatter," and click it.
9. A basic graph similar to **Figure 8** should appear.
10. Edit the chart title so that it matches the one created in Activity 1. This can usually be accomplished by clicking (or double clicking) on the title and then typing.
11. You can then add a label to each axis. This step in particular is very different depending on your software. You will typically be looking for a menu option titled "Axis Title." You will need to do this twice, once for each axis. Your graph should now look like **Figure 9**. You will use this graph again in Activity 3.

Figure 8.

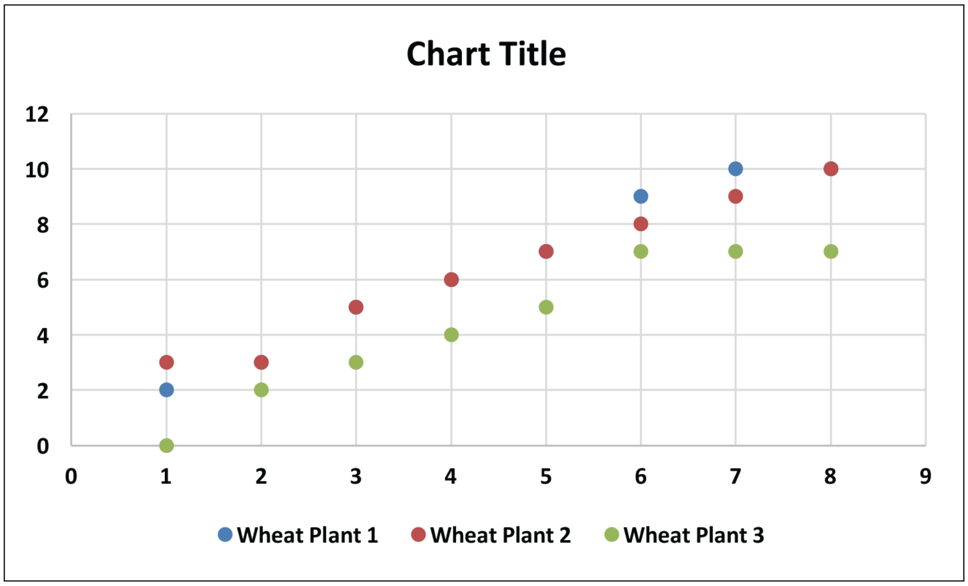
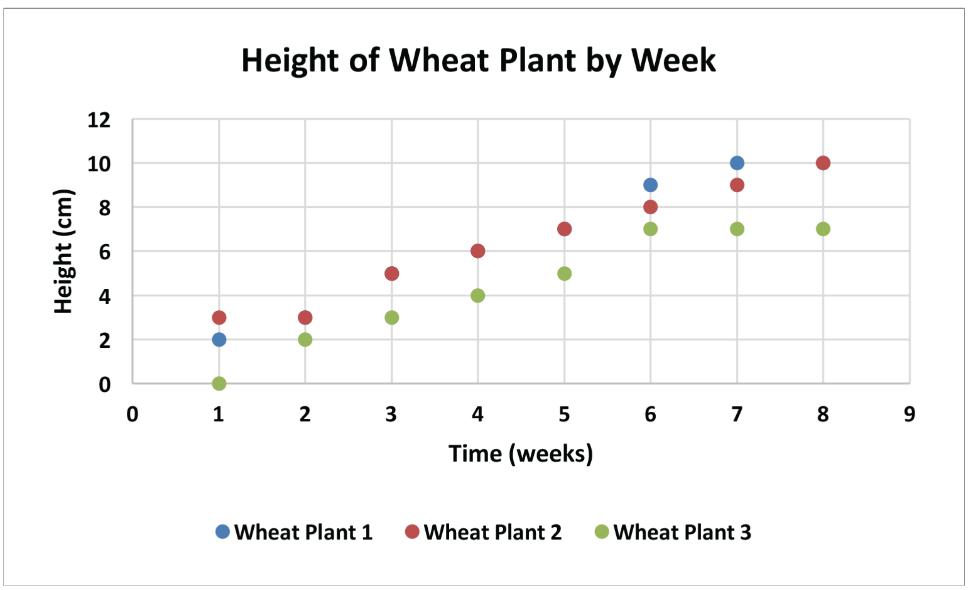


Figure 9.



ACTIVITY 3

A Linear Regression

Typically if you are graphing using an x,y scatter plot you are looking for trends (a recognizable pattern) in your data. In this activity you are looking to see if there is a trend in height of the plants over time. More specifically, you are looking for the rate at which the plants grew. This rate can be determined from the graph produced in Activity 2.

1. In your graph from Activity 2, click on a point from the Wheat Plant 1 dataset.
2. Right-click on the data point and select "Add Trendline."
3. Select "Linear."
4. Select "Display Equation on chart."
5. The equation displayed on the graph should read $y = 1.2381x + 0.9286$. Write this in "Wheat Plant 1 trendline equation" in the Data Table.

This is the equation of the line. In its general form is $\mathbf{y = mx + b}$. The "m" symbol stands for the slope of the line. The slope is how far the line rises (y) over a certain distance (x.) The "b" is called the **y-intercept**; this is the point at which the line crosses the y-axis. For the equation from step 6, this would mean that "1.2381" would be the slope and "0.9286" would be the y-intercept. This equation allows you to find the length of a plant at a certain time. For example, if you wanted to determine the height of the plant in week 9, based on this equation the estimated height would be 12.0715 cm.

$Y = 1.2381 * 9 + 0.9286$

$Y = 12.0715$ cm

Since the slope is calculated from $\frac{\text{change in y}}{\text{change in x}}$ it uses the same units as the dataset. In this case, this means that the slope has units of $\frac{\text{cm}}{\text{week}}$. The slope then means that on average, Wheat Plant 1 grew 1.2381 centimeters per week.

The y-intercept indicates that at week 0 the plant was likely 0.9286 cm tall. However, in this experiment the plants were all grown from seeds, so at week 0 they should have a height of 0. This information can be added to a trendline without having to add to a dataset.

6. Right click on the trendline and select "Format Trendline."
7. Select "Set Intercept" and set the number to 0. This is setting the y-intercept to 0. You can do this whenever you know the exact value of your dependent variable at the 0 for the x-axis.
8. Write the new trendline in "Wheat Plant 1 trendline corrected" in the Data Table.
9. Using the same procedure, create a corrected trendline for each additional wheat plant on the graph. Write the corrected equation for each in the data table.
10. Based on the corrected trend lines, which wheat plant grew fastest? Record your answer in the data table.

continued on next page

ACTIVITY 3 continued

Data Table.

Wheat Plant 1 trendline equation	
Wheat Plant 1 trendline corrected	
Wheat Plant 2 trendline corrected	
Wheat Plant 3 trendline corrected	
Wheat plant with fastest growth	

continued on next page

Title: ____

Label (y-axis): ____

Label (x-axis): ____

NOTES

Introduction to Graphing
Investigation Manual

www.carolina.com/distancelearning
866.332.4478

Carolina Biological Supply Company
www.carolina.com • 800.334.5551

CB781021610