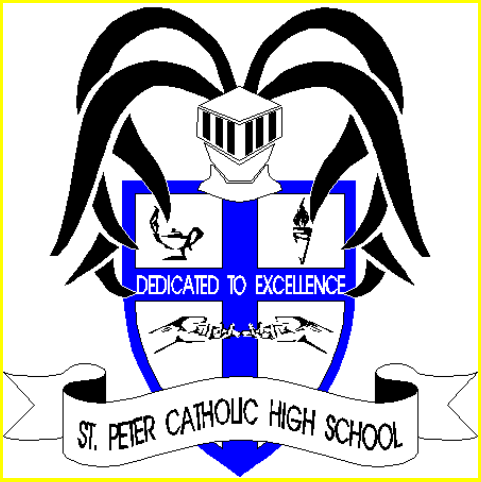
UNIT FOUR

Data Management & Relationships



Designed by Mark Couturier

Department Head at St. Peter Catholic High School

Website: www.habfanforever.com

YouTubeChannel: tangentsinecosine

Twitter: @tansinecosine

Important Terms and Terminology

Equivalent Ratios: A set of identical ratios which differ only as a result of a multiple.

Imperial System: A system of weights and measures built on the basic units of measure of the yard (length), the pound (mass), the gallon (capacity), and the second (time). Also called the British system.

Odds: A ratio comparing success against failure.

Percent Grade: The slope of a road in the form of a percentage comparing the rise or fall (accent or decent) against the run (or the horizontal distance). For example, a road with a 15% grade means that there is a rise or fall of 15 units for every 100 units of horizontal distance.

Pitch: The slope of a roof based on the truss in the form of a fraction comparing the height of the roof against half the base of the truss (roof).

Probability: A fraction or percentage comparing success against total outcomes.

Proportion: Two of more equivalent ratios linked together in the form of an equation.

Proportional Reasoning: Reasoning or problem solving based on the examination of equal ratios.

Rate: A fractional comparison between two concepts. For example, Black Forest ham is on sale for 4.50$/100g. The comparison is between cost (in dollars) against mass (grams).

Ratio: A numeric comparison between two of more concepts.

Respectively: A mathematical term to indicate that the order of a literal list is associated with the order of its numeric list. For example, in the sentence, "Josée, Dominique and Martine earn 5500$, 3300$ and 11500$ respectively" means that Josée earned 5500$, Dominique earned 3300$ and Martine earned 11500$. Without the word respectively, it cannot be assumed that literal list is the same as the numeric list.

Scale: A simplified numeric ratio comparison between a diagram's measurement against the actual life like measurement. Scales can represent magnifications such as biological microbes or representations of atoms and molecules in the form: Scale = 2,000 : 1 or they can represent demagnifications such as land masses like Québec or Luxembourg in the form of: Scale = 1 : 1,000,000.

Slope: A fractional comparison between the rise against the run of two concepts.

Unit Rate: A fractional comparison between two concepts, such that the latter concept is represented as single unit. For example, speed compares distance against time, hence 100km/hr. In this case, the speed is 100km per 1 hour.

Problem Solving Strategy
The Five-Step Process

LIST:	List all known and unknown variables in your problem.
FORMULA(E):	State any useful formulae that may be of use in your problem.
ALGEBRA:	Is your unknown isolated? If not, use algebra to isolate it.
PLUG-IN	Plug in the known variables into your formula(e).
EVALUATE:	Evaluate the problem and conclude with appropriate units.

Learning Goals in this Unit

By the end of this unit, you will be able to:

- interpret the meanings of points on scatter plots or graphs that represent linear relations, including scatter plots or graphs in more than one quadrant [e.g., on a scatter plot of height versus age, interpret the point (13, 150) as representing a student who is 13 years old and 150 cm tall; identify points on the graph that represent students who are taller and younger than this student] (***Sample problem:*** Given a graph that represents the relationship of the Celsius scale and the Fahrenheit scale, determine the Celsius equivalent of –5°F.);

- pose problems, identify variables, and formulate hypotheses associated with relationships between two variables (***Sample problem:*** Does the rebound height of a ball depend on the height from which it was dropped?);

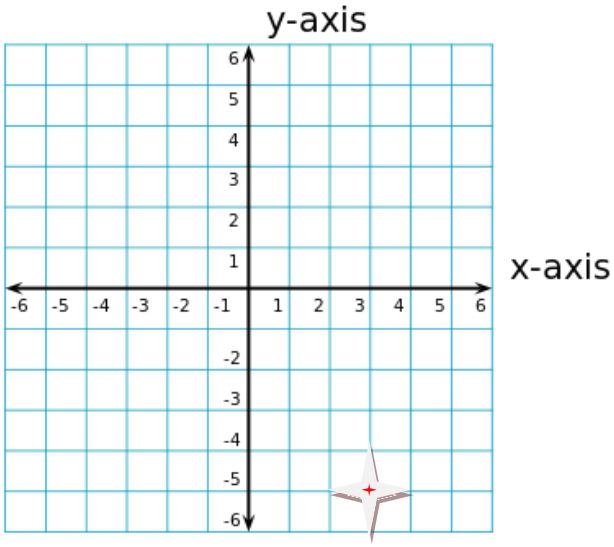
- carry out an investigation or experiment involving relationships between two variables, including the collection and organization of data, using appropriate methods, equipment, and/or technology (e.g., surveying; using measuring tools, scientific probes, the Internet) and techniques (e.g.,making tables, drawing graphs) (***Sample problem:*** Perform an experiment to measure and record the temperature of ice water in a plastic cup and ice water in a thermal mug over a 30 min period, for the purpose of comparison. What factors might affect the outcome of this experiment? How could you change the experiment to account for them?);

- describe trends and relationships observed in data, make inferences from data, compare the inferences with hypotheses about the data, and explain any differences between the inferences and the hypotheses (e.g., describe the trend observed in the data. Does a relationship seem to exist? Of what sort? Is the outcome consistent with your hypothesis? Identify and explain any outlying pieces of data. Suggest a formula that relates the variables. How might you vary this experiment to examine other relationships?) (***Sample problem:*** Hypothesize the effect of the length of a pendulum on the time required for the pendulum to make five full swings. Use data to make an inference. Compare the inference with the hypothesis. Are there other relationships you might investigate involving pendulums?).

INTRODUCTION TO SCATTERPLOTS

A **scatterplot** is a graph of x and y coordinates plotted to illustrate whether a relationship exists between two sets of data (the x and the y values).

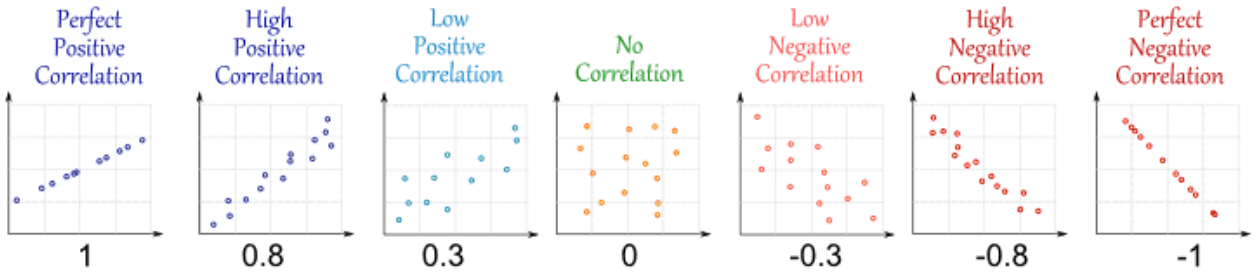
The **Cartesian Plane** (conceived by René desCartes, a Swiss mathematician) is made up of an x-axis (or the horizontal axis) and the y-axis (or the vertical axis). A point, (x,y) can be located by moving x units along the x-axis followed by y units along the y-axis. For example, the point (x,y) = (3,-5) can be found by moving three units right (positive) along the x-axis and then one unit down (negative) along the y-axis.



Can you plot the following points on the Cartesian Plane above?

(2,4) (4,1) (-1,5) (-3,-4) (6,-8)

A **correlation** defines the strength at with which two sets of data are related.



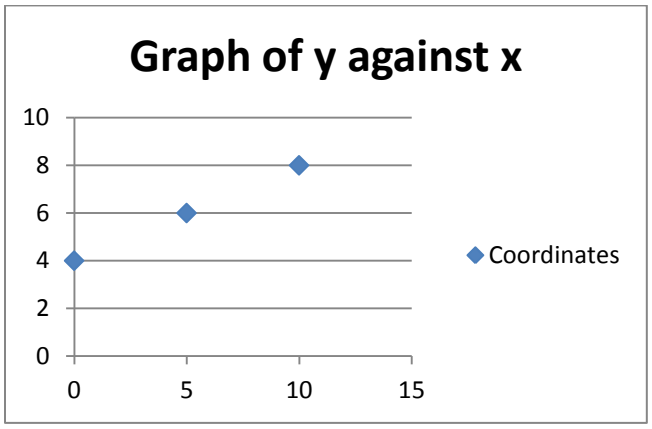
The scatterplot on the left is an example of a strong correlation because the data forms a near perfect line. When a line can be drawn across the data using a **line of best fit** we say that the data is **linear**. The scatterplot in the centre is an example of a correlation where the two data sets are not related. In this case, we cannot draw a **line of best fit**.

A correlation is said to be positive when **as the x-values rise so do the y-values**. In other words, if when the x-values **increase** so too do the y-values **increase**, then we have a **positive** correlation. If however, as the x-values **increase** but the y-values **decrease**, then we have a **negative** correlation.

A **line of best fit** is a line that the author must draw using their judgement in an attempt to quantify the linear relation. Correlations are deemed to be strong (very easy to draw the line), moderate (easy to draw but some judgement required and/or the data strays from the line), weak (difficult with a lot of judgement required and/or much data strays from the line) or none (it cannot be drawn despite judgement or it is not linear).

A **variable** is simply a representation of some concept's unknown quantity. The **independent variable** is located on the x-axis (time for example is **ALWAYS** on the x-axis) and is listed first on any table. This data is independent of the other variable. The **dependent variable** is located on the y-axis and is listed second on the table. This data is dependent on the independent variable.

Independent Variable x	Dependent Variable y
0	4
5	6
10	8



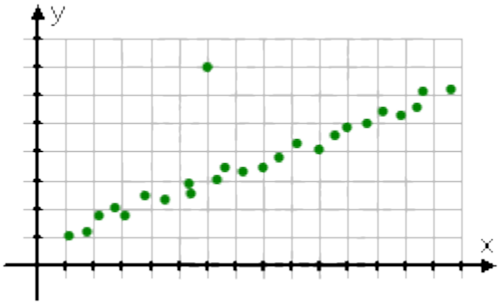
In statistics, we **interpolate** data when the author must make a prediction using a scatterplot when the result is **within** the data set. An author **extrapolates** data when a prediction is made using a scatterplot where the result is **outside** the data set. For example, if given data for every 10 years between the years 1950 to 2010, we would interpolate the data for 1975 but we would extrapolate the data for 1945 or 2014.

A **trend** is a description of how two data sets move in relation to each other. For example, as a child's age increases, so does the child's height.

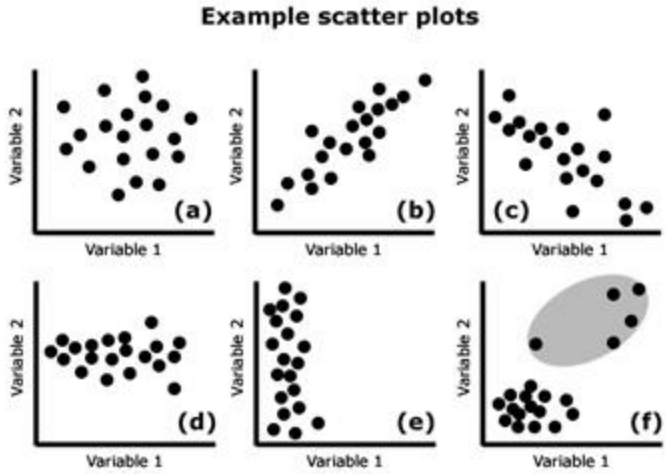
A **continuous** set of data is a description of data that exists for all values. For example, as time passes (increases) the distance that a jogger (either decreases, increases or remains). At any point in time, I can locate the jogger.

A **discrete** set of data is a description of data that only exists for specific values. For example, as time passes (increases) the number of holes that I dig (you can only have 0, 1, 2, …). In other words, you cannot dig half a hole.

An **outlier** is an anomaly point. This is a point that, for some reason, completely strays from the tendency of the correlation. In essence, we ignore this point when drawing the line of best fit and making our trend statement. In the case below, we would state that this scatterplot has a strong, positive correlation. We could further state, that an outlier exists at (6,7).



Examples:



In ALL examples, variable 1 is the independent variable because it lies on the x-axis, therefore, variable 2 is the dependent variable since it is on the y-axis.

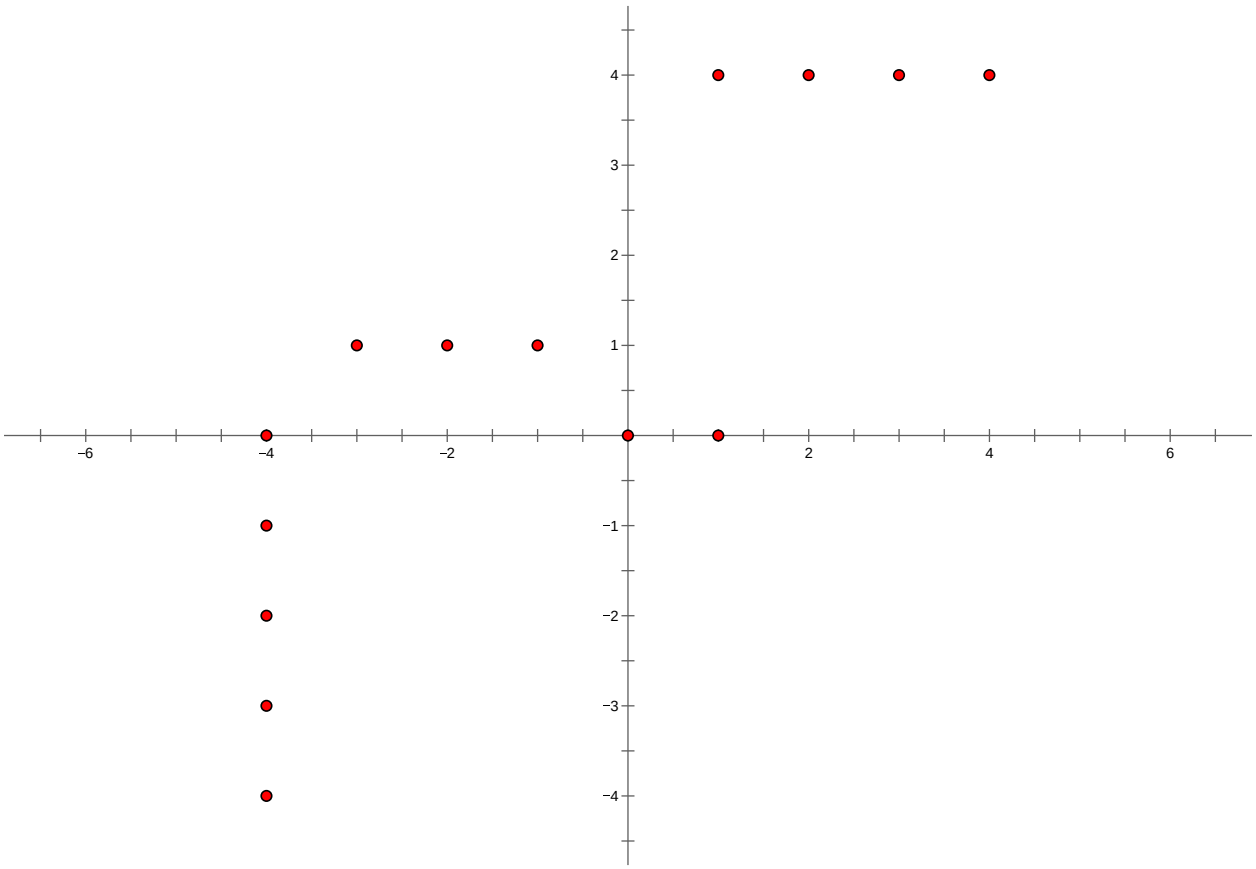
In example (a), there is no correlation as the data is too dispersed to make trend statement.

In example (b), there is a strong correlation as it would be easy to draw a line of best fit. The correlation is positive since the trend is that "as variable 1 increases, variable 2 also increases".

In example (c), there is a moderate correlation as it would be easy to draw a line of best fit, but the data does drift away from the line. The correlation is negative since the trend is that "as variable 1 increases, variable 2 decreases".

BATTLESHIP GAME – WARM UP

Let's play a short game of battleship to test our plotting skills. The red dots below are the locations of my ships on the Cartesian Plane. Give (x,y) points to see if you can sink my fleet (2 dots = patrol boat, 3 dots = frigate, 4 dots = destroyer and 5 dots = cruiser … since Canada no longer has any aircraft carriers since the HMCS Bonaventure).



Solution:

The Patrol Boat is located at the origin (0,0) and (1,0).

The frigate is located at (-3,1), (-2,1) and (-1,1).

The destroyer is located at (1,4), (2,4), (3,4) and (4,4).

The cruiser is located at (-4,0), (-4, -1), (-4, -2), (-4, -3) and (-4,-4).

GOLDEN OPPORTUNITY – WARM UP

You are given the following data in regards to some ore samples that you have been given.

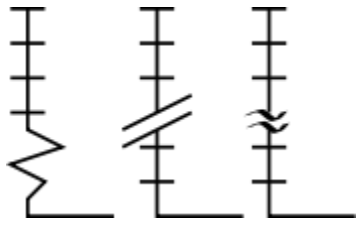
Volume (mL)	3.5	5.5	7.0	10.0	11.1	15.7
Mass (g)	67.6	106.3	135.2	193.2	55.5	303.3

Plot these points on the graph paper provided. Here are some helpful hints to get you started:

Since Volume (mL) is listed first, this becomes your independent variable which goes on your horizontal x-axis. Since the data starts at 3.5 and ends at 15.7 you may wish to start at 0 and end at 20 with equal increments in between.

This will also define the Title of the graph as being Mass against Volume (since titles should be y-variable against x-variable).

Since Mass (g) is listed second, this becomes your dependent variable which goes on your vertical y-axis. Since the data starts at 55.5 and ends at 303.3, you may wish to start at 0 and end at 350 (if possible) or you may wish to use the "break" symbol as shown below and start at 50 and end at 350.



What do you notice about the data? Use the Scatter Plot Reference for vocabulary.

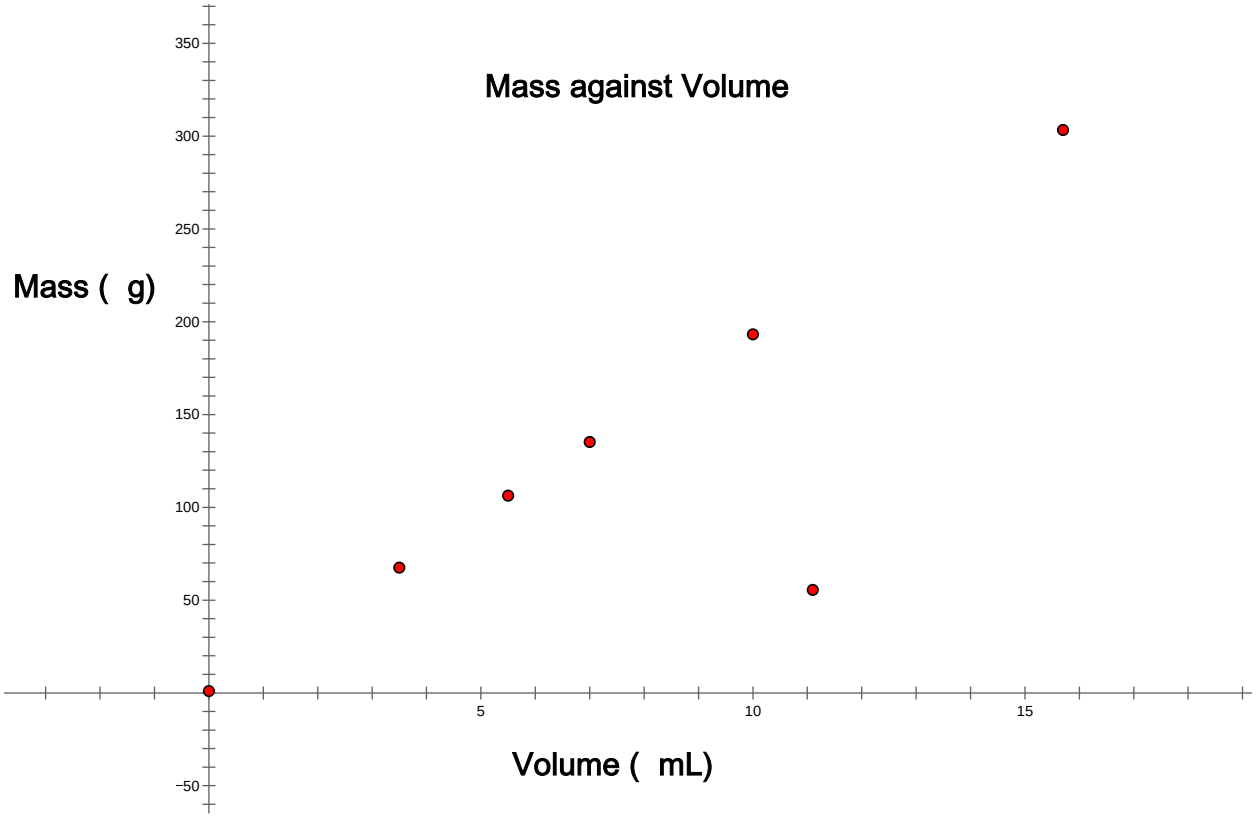
Does the data appear to be linear?

Can you draw a line of best fit?

Is there a piece of data that is out of place (an outlier)?

Solution:

Does you graph look like this in some way?



(Ignore the leftmost point as it is just indicating the origin (0,0))

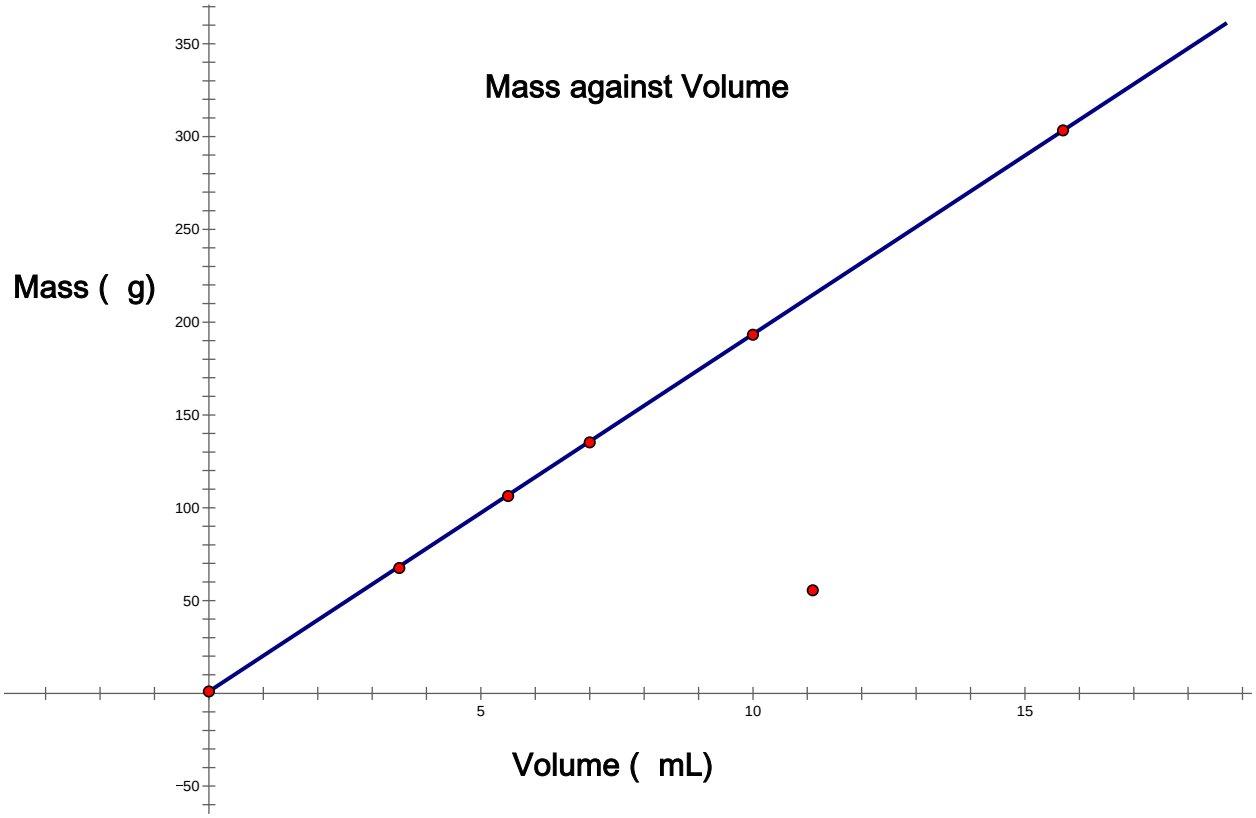
What do you notice about the data? Use the Scatter Plot Reference for vocabulary.

It appears that there is a relationship between Mass and Volume. Note that as the volume increases, the mass increases as well. This means that the relationship is positive.

Does the data appear to be linear?

There does appear to be a linear relationship as the points seem to all line up nicely with the exception of one point.

Can you draw a line of best fit?



We can draw a line of best fit and with great ease. Since it was so easy to do, we can say that the relationship is a strong, positive, linear relationship.

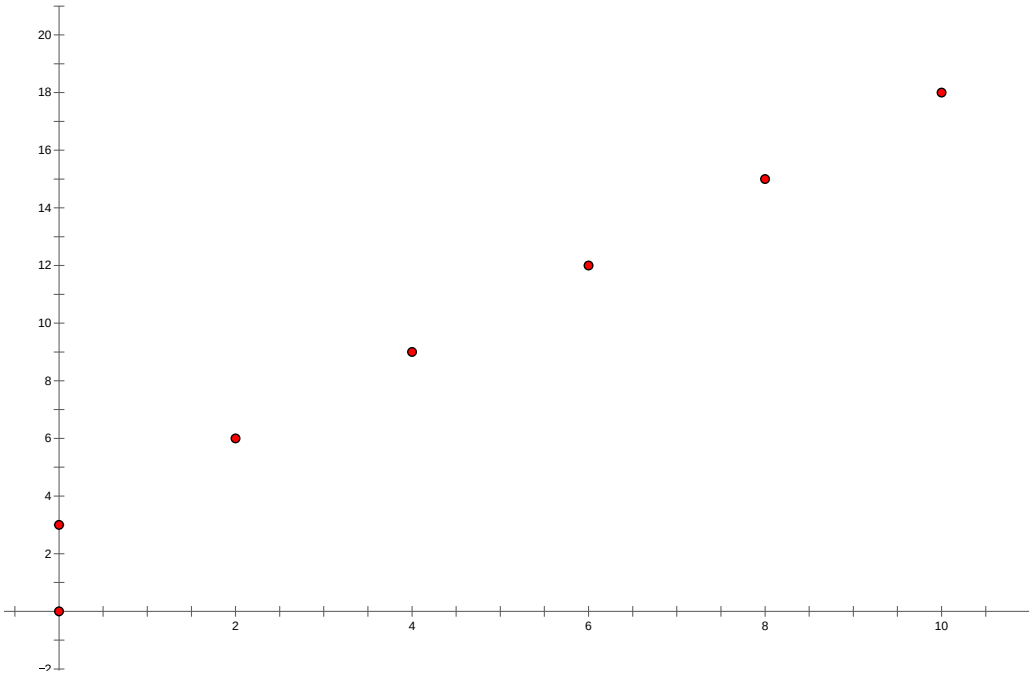
Is there a piece of data that is out of place (an outlier)?

There is an outlier. It appears to be the point (5.5, 11.1).

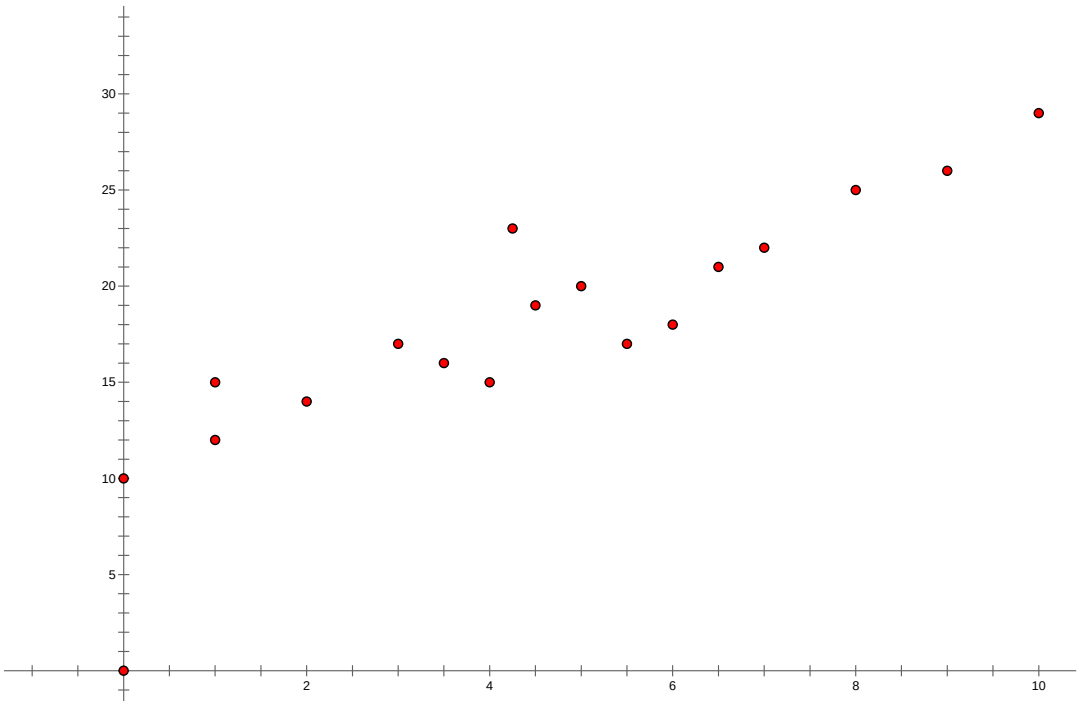
Science – Math Link : The data collected here was done so with sample of ore believed to be gold. Gold has a density of 19.32g/mL. If you take all the data and divide the Mass by the Volume (you get density) which will yield 19.32g/mL with the exception of the outlier. The outlier is actually pyrite (more commonly known as Fool's Gold) which has a density of 5.0g/mL.

MORE WARM-UP PRACTICE

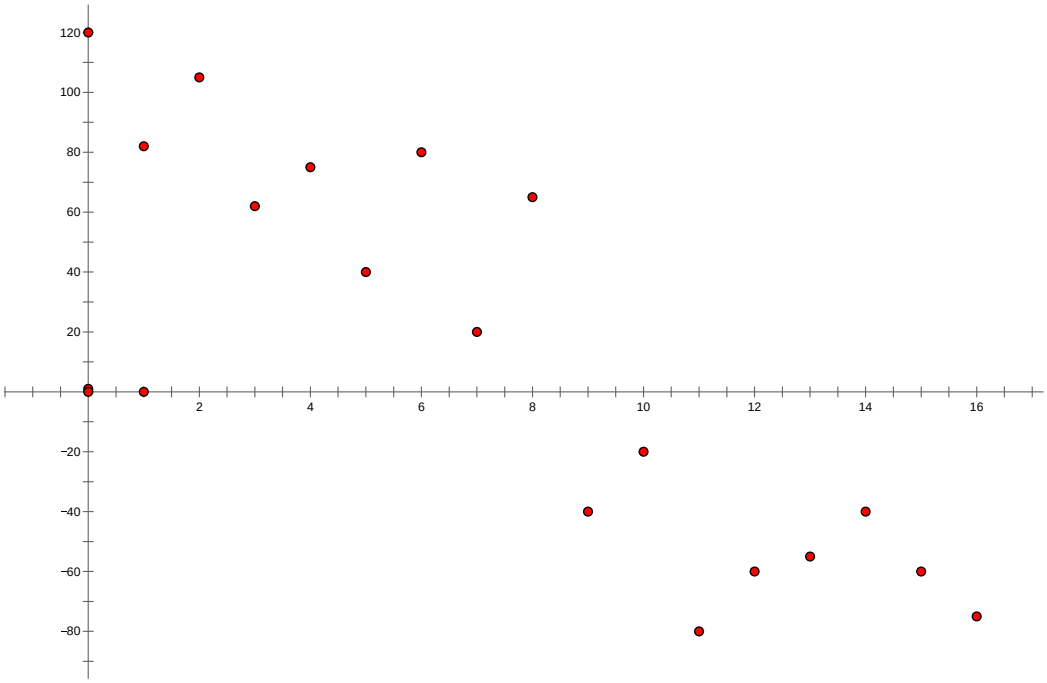
Given the following examples and using your Scatter Plot Reference Sheet, draw lines of best fit where possible. Recall that when drawing a line of best fit, attempt to split the data evenly such that there is almost an equal amount of data points on top and on the bottom, while following the trend.

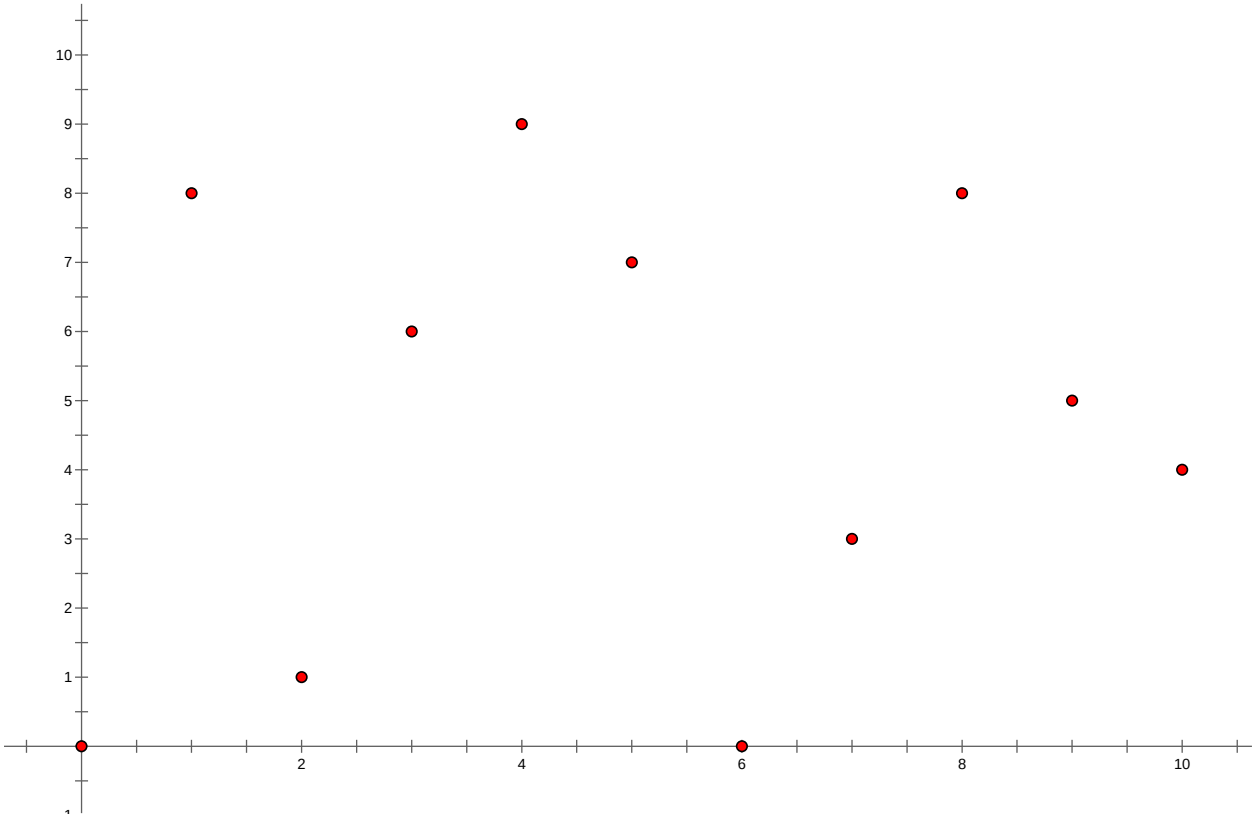


In each case, ignore the origin point as it is there for reference only.

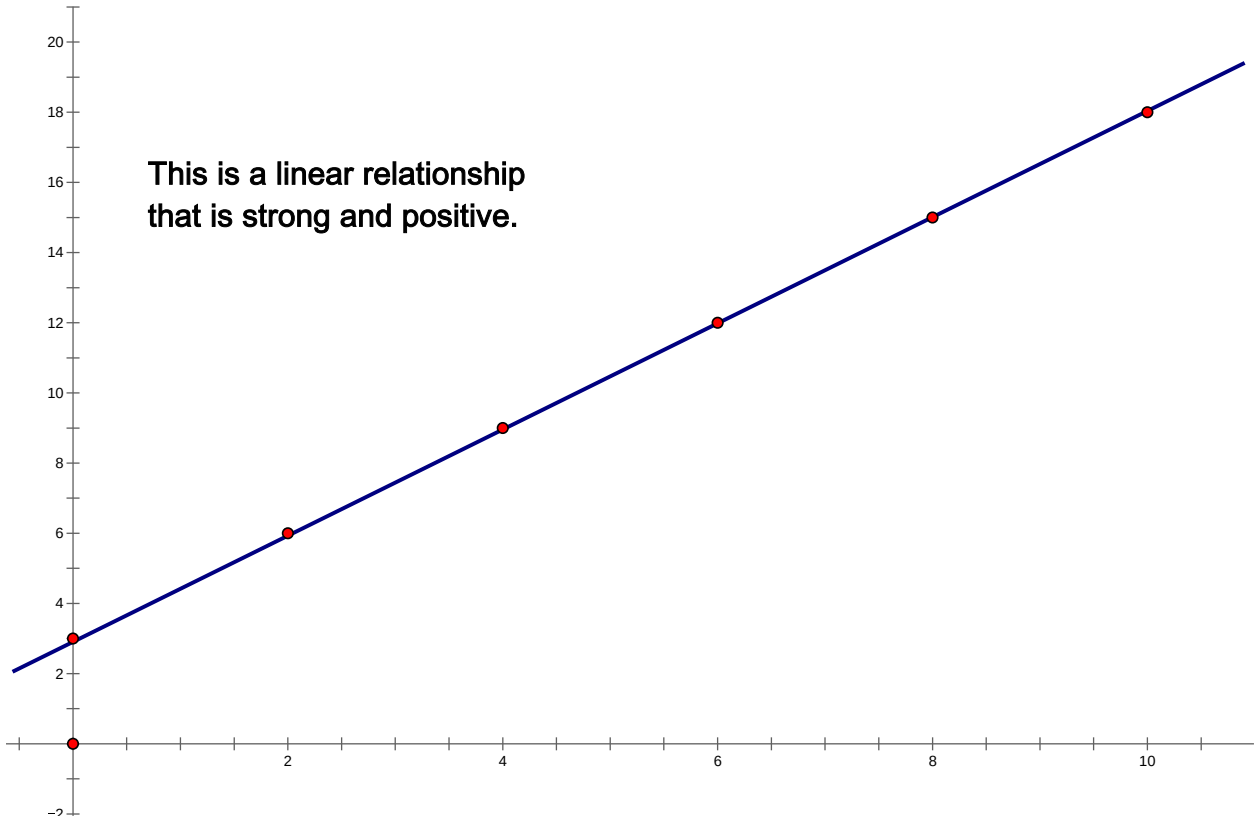


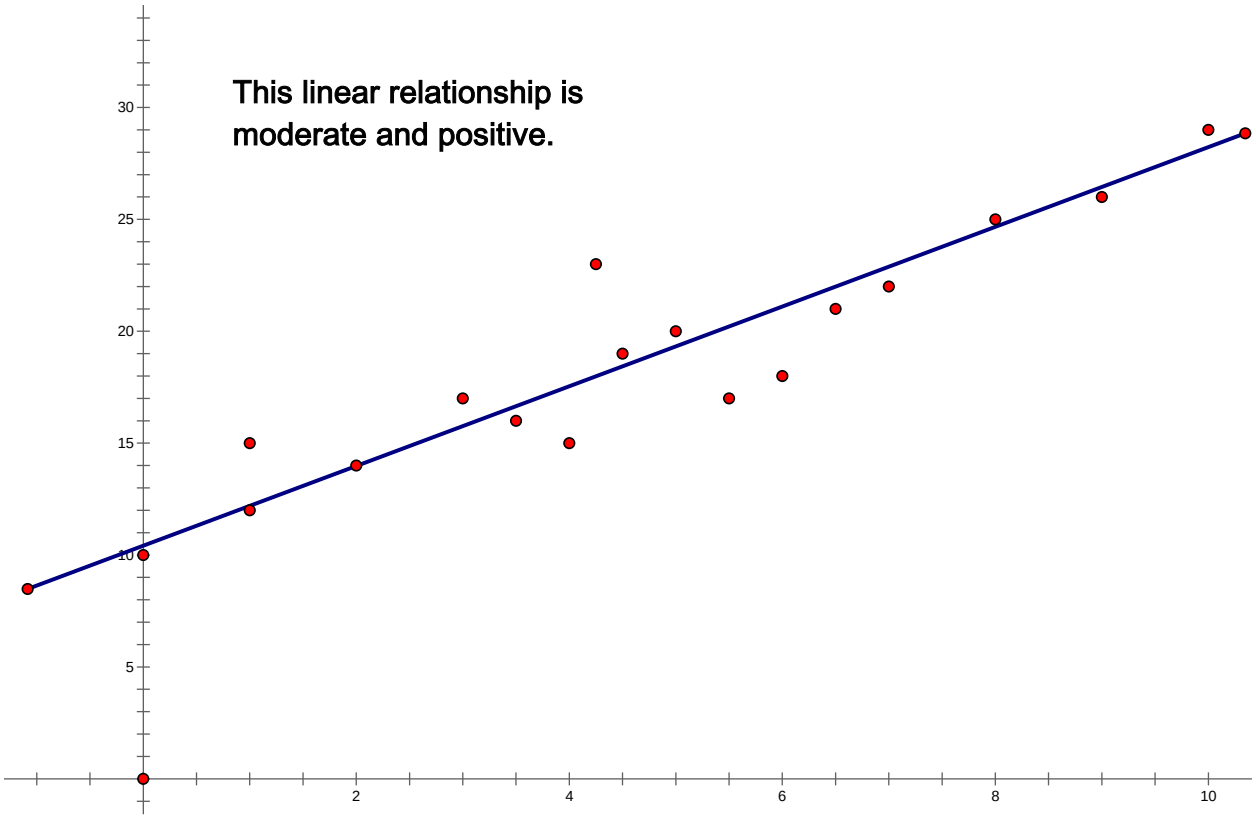
More Line of Best Fit Warm-up

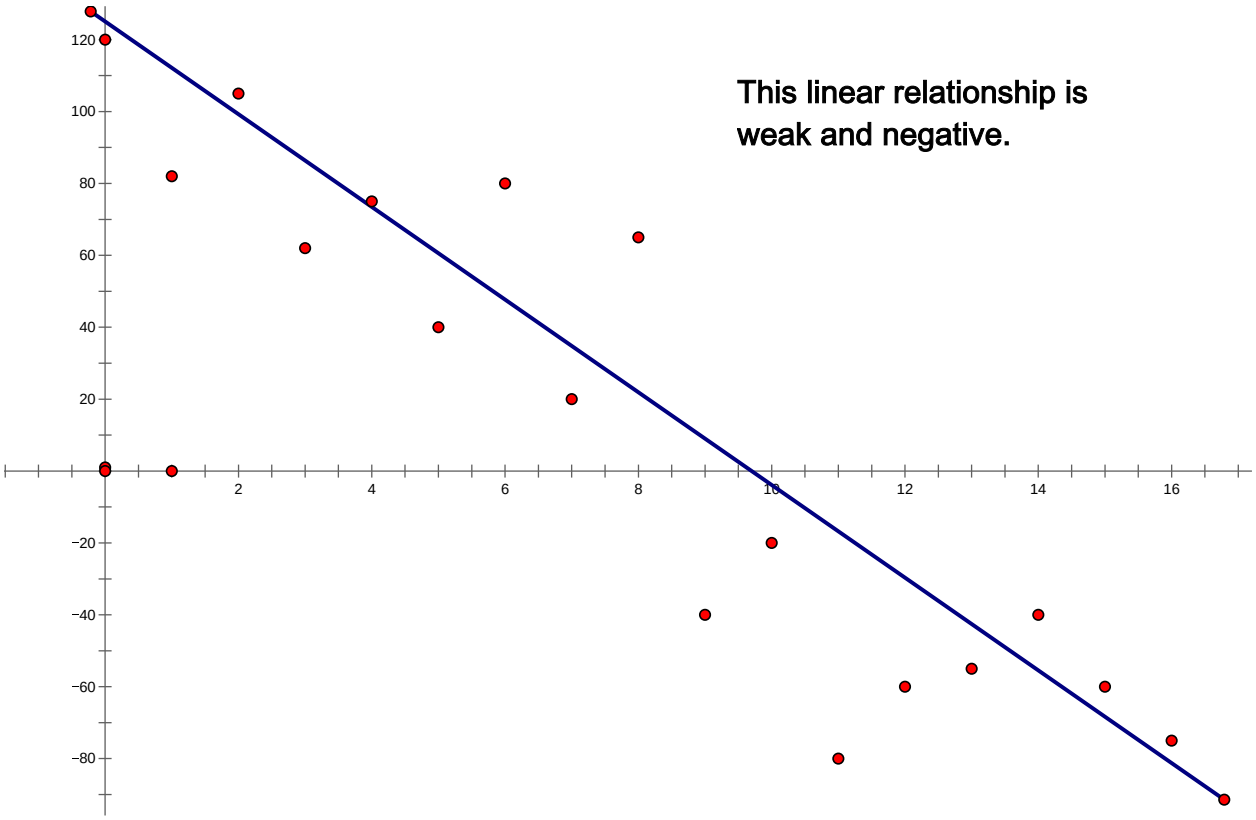


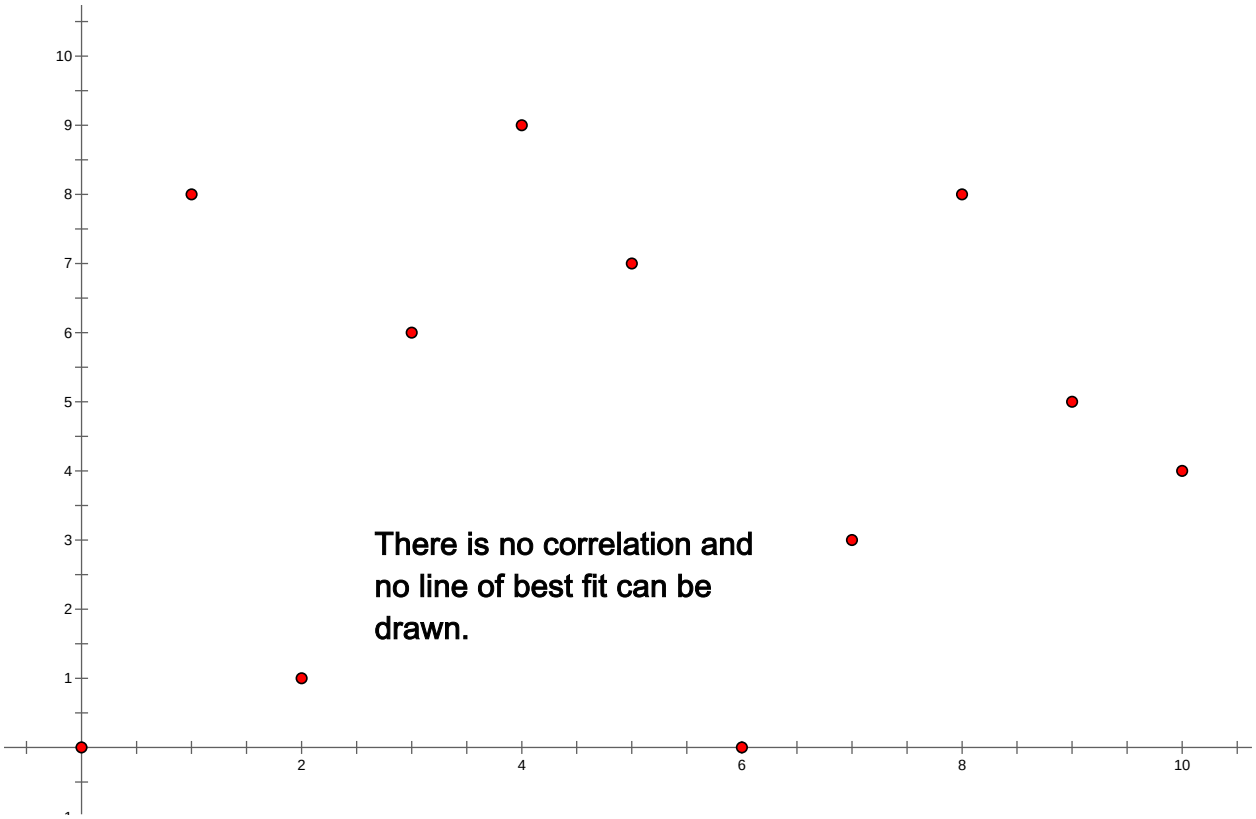


Solutions:


This is a linear relationship
that is strong and positive.
20
18
16
14
12
10
8
6
4
2
2
4
6
8
10


This linear relationship is
moderate and positive.
30
25
20
15
10
5
2
4
6
8
10


This linear relationship is
weak and negative.
120
100
80
60
40
20
2
4
6
8
12
14
16
−20
−40
−60
−80

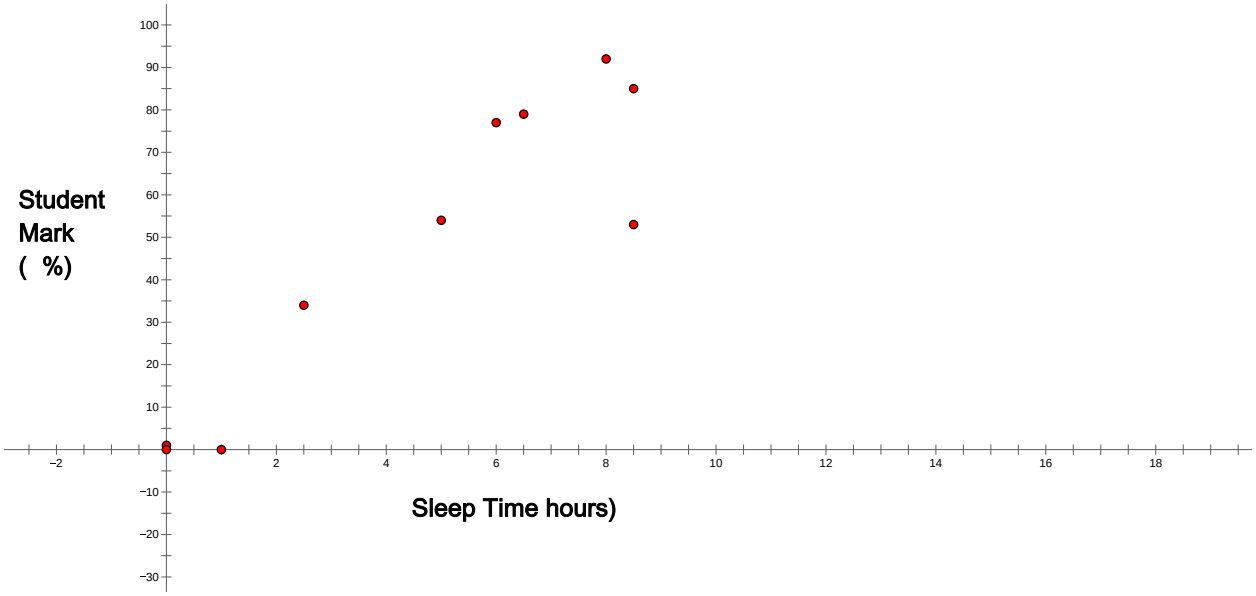

10
9
8
7
6
5
4
3
2
1
2
4
6
8
10
There is no correlation and no line of best fit can be drawn.

By the end of this class you will be able to interpret the meanings of points on scatter plots or graphs that represent linear relations.

Students in a math class have been analyzing the amount of time they sleep in hours against their overall mark in their class as a percent. They created a table of values and then made a scatterplot. (Note that units are always identified in the problem, in the table and in the graph)

Sleep time (hr)	5.0	8.5	8.0	2.5	6.0	8.5	6.5
Student Mark %	54	85	92	34	77	53	79

Note that the independent variable always comes first in the table of values.



Note that time is always placed along the x-axis as time is always the independent variable.

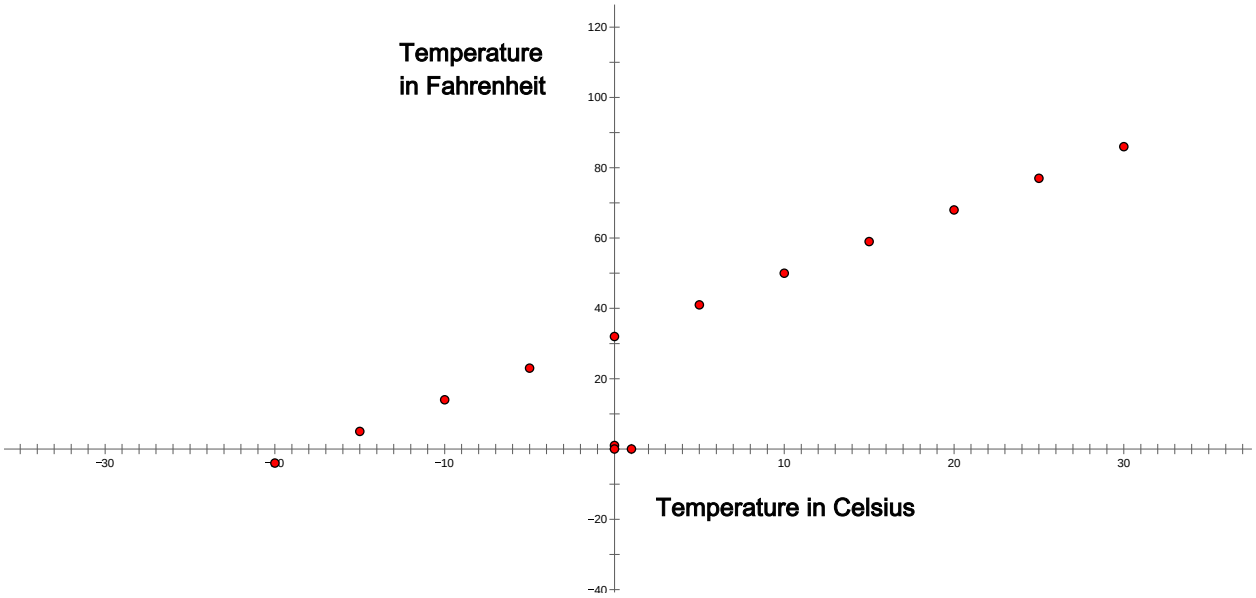
In your opinion, which students are doing better in class?

One student appears to be getting plenty of sleep but is not doing very well in class. Identify that student. What is that student's mark and how many hours of sleep are they getting?

Can you draw a line of best fit (remember, you can ignore outliers when doing so)?

Sample problem: Given a graph that represents the relationship of the Celsius scale and the Fahrenheit scale, determine the Celsius equivalent of –5°F.

A foreign exchange student from the US state of Ohio has been given the following graph to help him deal with the difference in Temperature measurements. His country uses degrees Fahrenheit while Canada uses the metric Celsius system.



Challenge:

What is the approximate outside Temperature in Fahrenheit, if the weather outside in Celsius is:

a) -4 ° C = _____ How did you obtain this? _____

b) 0 ° C = _____ How did you obtain this? _____

c) 33 ° C = _____ How did you obtain this? _____

What is the approximate outside Temperature in Celcius, if the weather outside in Fahrenheit is:

a) -5 ° F = _____ How did you obtain this? _____

b) 32 ° F = _____ How did you obtain this? _____

c) 90 ° F = _____ How did you obtain this? _____

Challenge:

What is the approximate outside Temperature in Fahrenheit, if the weather outside in Celsius is:

a) -4 ° C ≈ 24° F How did you obtain this? Interpolation

b) 0 ° C ≈ 32° F How did you obtain this? Interpolation

c) 33 ° C ≈ 91° F How did you obtain this? Extrapolation

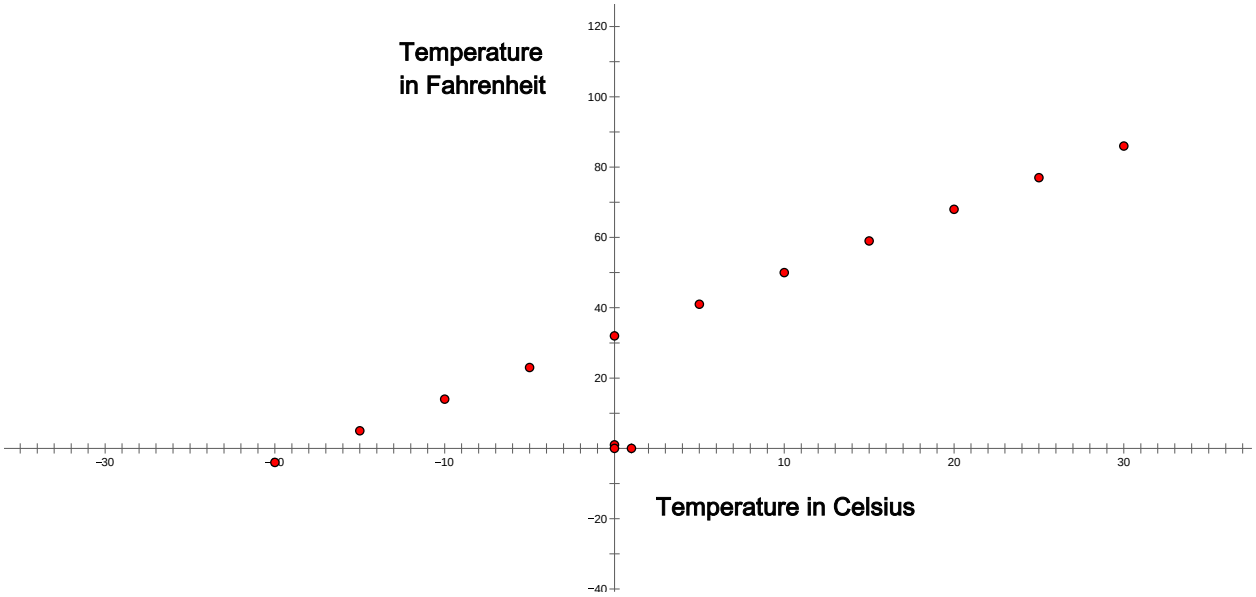
What is the approximate outside Temperature in Celcius, if the weather outside in Fahrenheit is:

a) -5 ° F ≈ -21° C How did you obtain this? Extrapolation

b) 32 ° F ≈ 0° C How did you obtain this? Interpolation

c) 90 ° F ≈ 32° C How did you obtain this? Extrapolation

What other things can we say about the graph between Fahrenheit against Celsius?



We can say that a relationship does indeed exist. We can further say that the relationship is strong, linear and positive. We can say this because it would be very easy to draw a "line of best fit", hence it is strong and linear. We can say that it is positive because as the Temperature in Celsius rises (from left to right) so does the Temperature in Fahrenheit (from bottom to top). Note that some of the Temperatures were not of the graph, yet we were able to "extrapolate" the approximate Temperatures rather easily.

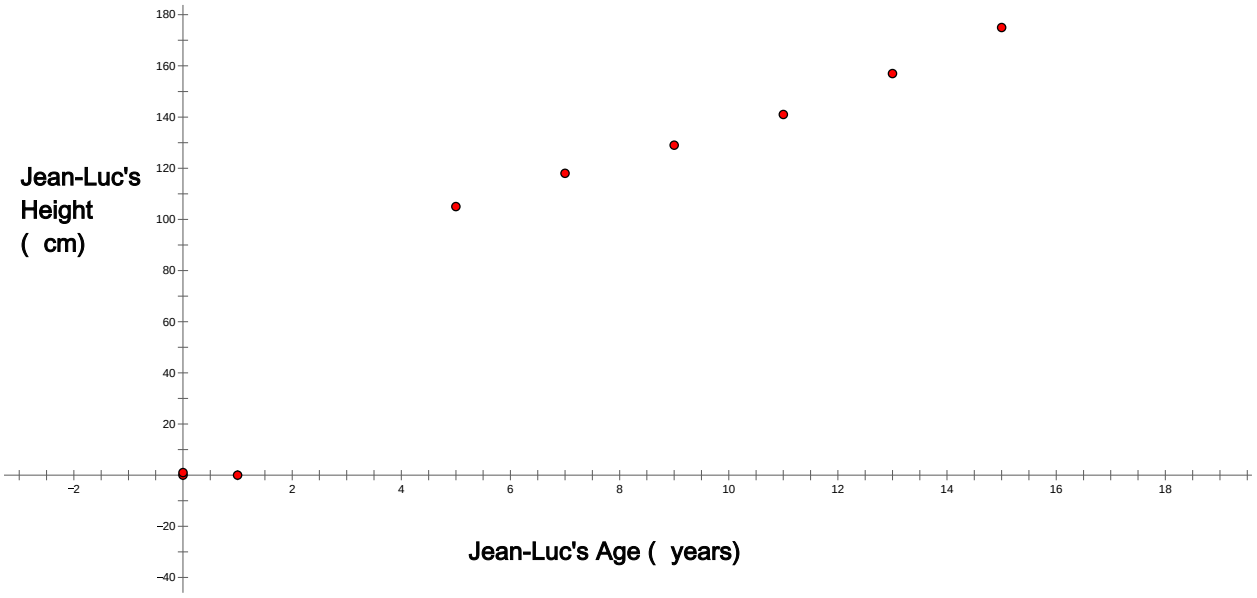
By the end of this activity you will be able to interpret the meanings of points on scatter plots or graphs that represent linear relations.

Male Height against Age

Two high school best friends are comparing heights as they grew up in school. One of the boys, Jean-Luc, creates a table and a graph of his height against his age starting at the age of four.

Age (years)							17
Height (cm)							

Male Height against Age



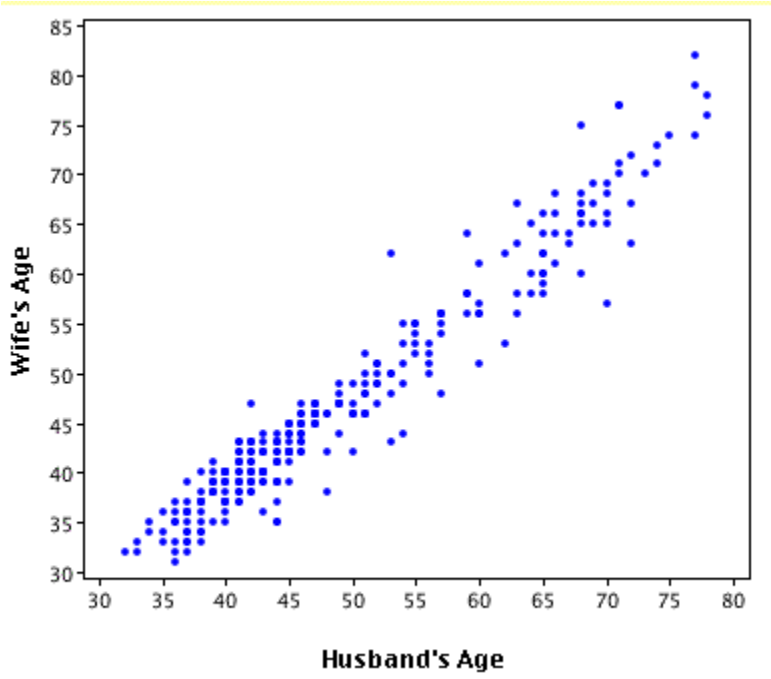
1) Complete the table above given the points from the table.

2) At about what age was Jean-Luc 140 cm tall? Explain your reasoning.

3) How tall was Jean-Luc at the age of 8 years? Explain your reasoning.

In class task Scatterplots Assignment

Given the graph below, answer the following questions:



Can a line of best fit be drawn? If so, draw the line of best fit. (2 marks)

Does a relationship exist given the graph? If so, describe the trend. (2 marks)

<u>Vocabulary Recall</u>: A **trend** is a description of how two data sets move in relation to each other. For example, as a child's age increases, so does the child's height.

If this is a relationship, is it a positive or negative relationship? Explain (1 mark)

If this is a relationship, is the relationship strong, moderate or weak? Explain. (1 mark)

COMMISSION BASES EARNINGS

You have just accepted a job working in a retail store. Under the conditions of your employment, you receive no salary but instead earn a commission. The following table illustrates the amount of product that you sold and the amount of commission that you earned.

Sales ($)	530.00	1254.00	845.00	0.00	3400.00	1578.00	2549.00
Commission earned ($)	58.30	137.94	92.95	0.00	374.00	173.58	280.39

Graph the data from the table. (3 marks)

Label the axis (2 marks)

Include a title (1 mark)

Draw a line of best fit if possible. (2 marks)

Does a relationship exist? If so, describe the trend. (1 mark)

If this is a relationship, is it a positive or negative relationship? Explain (1 mark)

If this is a relationship, is the relationship strong, moderate or weak? Explain. (1 mark)

EFFECTIVE STUDYING AND TEST SCORES

You decided to do a self experiment on yourself to see if effective studying actually helps you get better marks. You measure the amount of time that you study and then record the grade that you got. It yields the following table:

Time (hours)	0	0.5	1.0	1.5	2.0
Grade (%)	35	46	54	66	73

Graph the data from the table. (3 marks)

Label the axis (2 marks)

Include a title (1 mark)

Draw a line of best fit if possible. (2 marks)

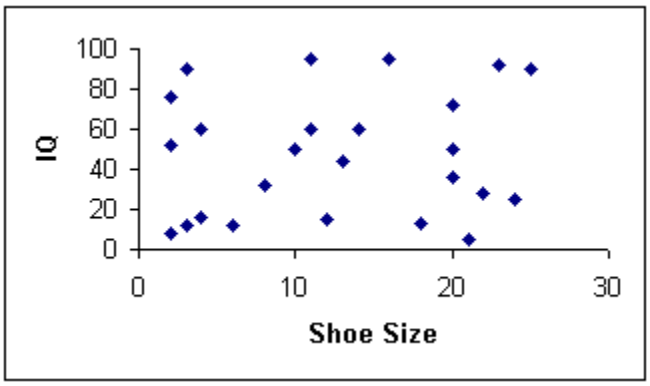
Does a relationship exist? If so, describe the trend. (1 mark)

If this is a relationship, is it a positive or negative relationship? Explain (1 mark)

If this is a relationship, is the relationship strong, moderate or weak? Explain. (1 mark)

YOUR SHOE SIZE & YOUR IQ

Given the graph below, answer the following questions:



Can a line of best fit be drawn? If so, draw the line of best fit. (1 marks)

Does a relationship exist given the graph? If so, describe the trend. If not, explain. (2 marks)

By the end of this class you will be able to carry out an investigation or experiment involving relationships between two variables, including the collection and organization of data, using appropriate methods, equipment, and/or technology and technique.

Sample problem: Perform an experiment to measure and record the temperature of ice water in a plastic cup and ice water in a thermal mug over a 30 min period, for the purpose of comparison. What factors might affect the outcome of this experiment? How could you change the experiment to account for them?);

Sample problem: Set-up

Question: What happens to the Temperature of the water over a period of time and does the container make a difference?

Hypothesis: I think that … I think this because ...

Using the online – timer – stopwatch at (http://www.online-stopwatch.com/), collect Temperature data (in °C or °F) using the provided thermometers.

Table of Values

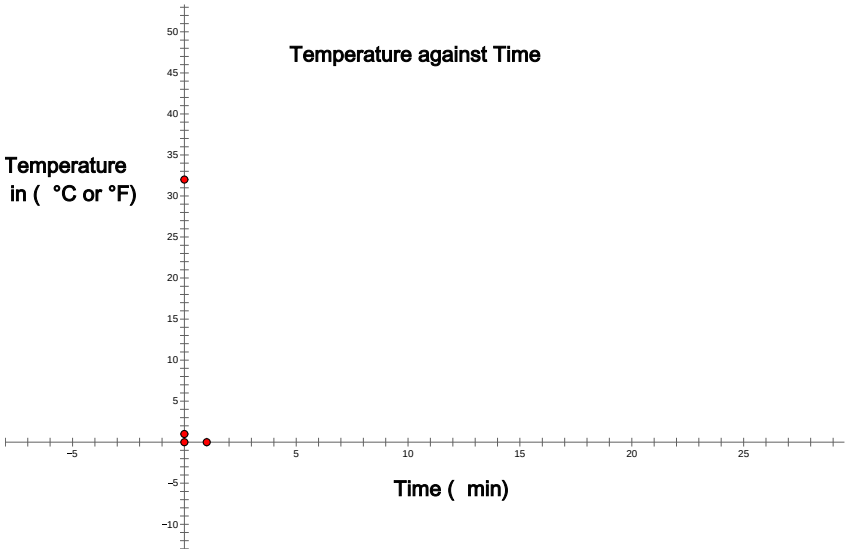
Recognize which data is the independent variable and needs to be placed along the x-axis. In this case, it will be time (min). Likewise, recognize which data is the dependent variable and needs to be placed along the y-axis. Here are two examples:

Time (min)	Temperature *in* (°C or °F)

OR

Time (min)								
Temperature *in* (°C or °F)								

Graph



By the end of this activity you will be able to carry out an investigation or experiment involving relationships between two variables, including the collection and organization of data, using appropriate methods, equipment, and/or technology and technique.

THE VITRUVIAN MAN LAB

The perfect human?

Leonardo da Vinci drew *Le proporzioni del corpo umano secondo Vitruvio*, more famously known as the Vitruvian Man, in 1490. The drawing is of a single man drawn twice in two positions such that we can see the correlations of the human body. da Vinci was referencing an architect by the name of Vetruvio who had noticed that the human body seemed to grow proportionately (rather than randomly) such that a palm equalled four fingers, a foot equals four palms, a cubit equals six palms, a pace equals four cubits and a man equals 24 palms. The drawing itself implies that the length of the outstretched arms equals the height of that man. Is this always true? Let's test this idea.

To make this lab very effective, while in your groups, each group member will measure you twice to ensure the accuracy of your readings. You are to measure each of the following items in cm and your partner will enter the measurement on your handout. You will then take an average of the two readings (add the two readings and divide by two).

My height: _____

My forearm: _____

My arm span: _____

My foot length:_____

Walking stride length: _____

My hand span:_____

My height is: _____

My forearm: _____

My arm span: _____

My foot length:_____

Walking stride length: _____

My hand span:_____

My averages are (ALL CALCULATIONS IN cm):

My height: _____

My forearm: _____

My arm span: _____

My foot length:_____

Walking stride length: _____

My hand span:_____

If you are done, find the indicated class computer and input your data in the table or in TINKERPLOTS. We will use this data later to attempt to make correlations.

THE TASK

Question

Once the data has been collected, you will now create a question that relates two of the measured items. For example, "Is the length of a person's leg related to the circumference of their neck?"

Hypothesis

Think about the question you created and make an educated statement relating two of the measured items. For example, the hypothesis should be phrased in the format, "I think that (concept A) is related to (concept B) in a non-linear (or linear) manner. I think this because …"

The Test

You will now test your statement against the data that we have collected as a class. You will do this by creating the appropriate labelled table. When the table is complete, you will then graph the two items against each other, labelling the axis and providing a title.

The Analysis

If possible, create a line of best fit and discuss its meaning. Identify any outliers. Analyze the graph using the terms discussed in class. Review the statement and modify it if necessary with your conclusions.

By the end of this activity you will be able to describe trends and relationships observed in data, make inferences from data, compare the inferences with hypotheses about the data and explain any differences between the inferences and the hypotheses (e.g., describe the trend observed in the data. Does a relationship seem to exist? Of what sort? Is the outcome consistent with your hypothesis? Identify and explain any outlying pieces of data. Suggest a formula that relates the variables. How might you vary this experiment to examine other relationships?)

Using the CIA Factbook pick a question (or make one up for yourself) and research the data. An example will be provided to you along with a short lesson on the CIA Factbook resource.

Do richer countries have better healthier children?

Hypothesis

I think that richer countries have … I think this because …

Research

What are the richest countries in the world in terms of GDP per capita (all income divided by the population)? To get a good distribution, we should look at the first, eleventh, twenty-first, etc.

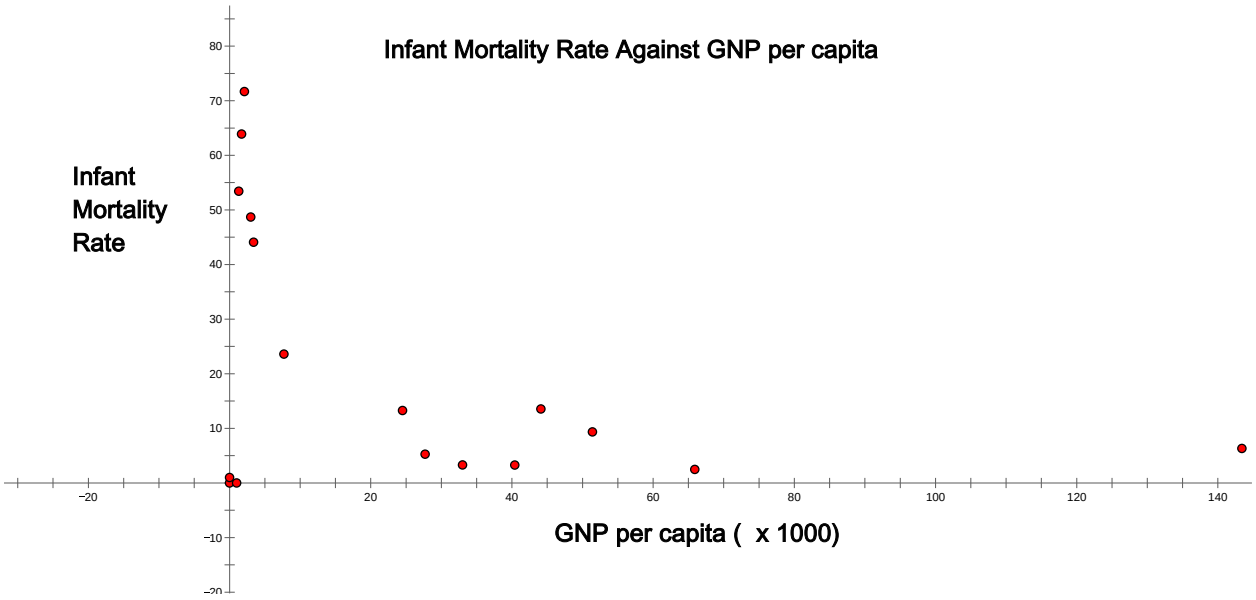
From this list, we should match these countries with the Child Mortality Rate. This indicator gives a statistic as per the number the children that do not live past a certain age.

Tabulate the Data

You now want to tabulate your data in preparation for the scatterplot that you will prepare.

Nation	GDP per capita ($)	Child Mortality Rate
Qatar	143,400	6.32
Norway	65,900	2.48
Bahrain	51,400	9.35
Oman	44,100	13.55
France	40,400	3.28
Spain	33,000	3.30
Slovakia	27,700	5.27
Malaysia	24,500	13.27
Morocco	7,700	23.60
Bangladesh	3,400	44.09
Djibouti	3,000	48.70
Sierra Leone	2,100	71.68
The Gambia	1,700	63.90
Guinea	1,300	53.43
Central African Republic	600	90.63

Graph the data



Analyze the Data and Draw Conclusions

Does a relationship seem to exist? Is it linear? Is it positive? Is it strong? Was your hypothesis correct? Are there any outliers? How would you do this project differently?

A relationship does exist but it is not linear. It is negative because as the GNP per capita increases so does the infant mortality rate.

Can I draw a CURVE OF BEST FIT?

What I would suggest doing differently is looking at the poorest countries only as these countries seem to follow a strong, linear negative relation. Then I would look at the middlemost nations as they appear to have a moderate, linear negative relation.

YOUR PROJECT

Here is a list of questions that you can research on the CIA Factbook website.

Do countries with the high rates of AIDS/HIV have small populations?

Do the countries with the highest military spending have the wealthiest people?

Do the countries with the highest per capita ratio of doctors have the lowest death rates?

Are the countries that are the most in debt per capita the poorest nations?

You can come up with your own question. These are just suggestions.

(***Sample problem:*** Hypothesize the effect of the length of a pendulum on the time required for the pendulum to make five full swings. Use data to make an inference. Compare the inference with the hypothesis. Are there other relationships you might investigate involving pendulums?).

Do this as a demo.

MOCK EQAO

UNIT REVIEW