ALGEBRA I MATH MODULE: OUTLINE

Rationale

The purpose of this module is to introduce students to basic concepts within data science while also providing an introductory activity for the instruction of related topics contained in the Missouri Learning Standards. This module will allow students to construct and interpret scatter plots from real-world data. Students will have an opportunity to work with linear models, residuals, and correlation. This module serves as a starting point for instruction related to the following Missouri Learning Standards:

Math:

- A1.DS.A.5 - Construct a scatter plot of bivariate quantitative data describing how the variables are related; determine and use a function that models the relationship. a. Construct a linear function to model bivariate data represented on a scatter plot that minimizes residuals. b. Construct an exponential function to model bivariate data represented on a scatter plot that minimizes residuals.
- A1.DS.A.6 - Interpret the slope (rate of change) and the y-intercept (constant term) of a linear model in the context of the data.
- A1.DS.A.7 - Determine and interpret the correlation coefficient for a linear association.
- A1.DS.A.8 - Distinguish between correlation and causation.

Science:

- 9-12.ESS2.A.2 - Analyze geoscientific data to make the claim that one change to Earth's surface can create changes to other Earth Systems.

MoExcel Data Science Standards

- MoExc1: **Identify** issues, problems, questions, or claims that can be addressed using large datasets.
  *The expectation is that students be able to **identify** statements, claims, or questions that can be refined into testable hypotheses.*
- MoExc2: **State** data-driven investigative questions.
  *The expectation is that students be able to **state** investigative questions based on quantitative data.*
- MoExc3: **Construct** visual representations of real-life data from publicly available datasets and **describe** patterns observed.
  *The expectation is that students are familiar with large datasets of publicly available data that allow users simple but rich manipulation of bivariate data and **describe** patterns that result from purposeful manipulation of the information.*
- MoExc4: **Suggest** and **discuss** the possible interactions among data.
  The expectation is that students can provide and consider alternative explanations to the relationships (or lack thereof) among data.
- MoExc5: **Identify** and **discuss** potential factors that can influence relationships between the independent and dependent variables.
  *The expectation is that students reflect on the complexity of real-life problems and consider it when attempting analyses or problem-solving. This includes identifying and accounting for different forms of control variables (intervening, confounding, or antecedent). Discussion of the differences among control variables is **not** expected.*
- MoExc6: **Interpret** real-life data by using patterns and relationships among data.
  The expectation is that students are able to construct stories that provide plausible explanations for relationships that have been identified among data.

Standards for Mathematical Practice

Standard#:	Standard:
MP1	Making sense of problems and persevere in solving them.
MP2	Reason abstractly and quantitatively.
MP3	Construct viable arguments and critique the reasoning of others.
MP4	Model with mathematics.
MP5	Use appropriate tools strategically.
MP6	Attend to precision.
MP7	Look for and make use of structure.
MP8	Look for and express regularity in repeated reasoning.

Prior Knowledge & Possible Misconceptions:

Prior Knowledge:

This module assumes that previous instruction has covered the Algebra I standards, specifically:

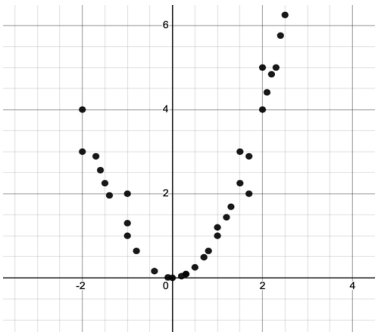
A1.LQE.A.1 Distinguish between situations that can be modeled with linear or exponential functions. a. Determine that linear functions change by equal differences over equal intervals. b. Recognize exponential situations in which a quantity grows or decays by a constant percent rate per unit interval.

A1.LQE.A.3 Construct linear, quadratic and exponential equations given graphs, verbal descriptions or tables.

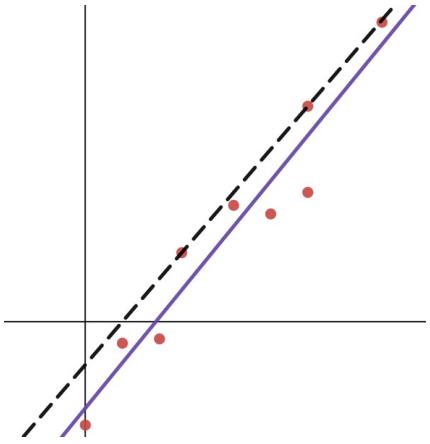
A1.DS.A.1 Analyze and interpret graphical displays of data.

Possible Misconceptions:

1. Students may confuse nonlinear data with data that has "no association." For example, a linear function will not provide a good model for the scatter plot shown, but there is still a clear trend in the data, it is just a nonlinear relationship.



2. Students may believe that the most accurate trend line is the one that passes through the most data points. For example, they may decide the black dashed line is a better model for this data because it passes through several data points, while the purple (and more accurate) line of best fit does not pass through any data points.



Algebra I Data Science Math Module

Example: Global Temperature, Ice Sheets, and Ocean Levels

Question: Do rising temperatures affect sea levels?

Data: Temperature, Ice Loss, Sea Level from NASA, JPL, and University of Colorado

This can be a teacher demonstration or students can explore the visualization on their own device.

Discussion Outline:

Initial intuition questions before we consider any data:

1. What effects do rising temperatures have on the earth? What effects do rising temperatures have on humans?

   If students are unaware of this phenomenon the following website is a good resource with charts and visualizations: https://climate.nasa.gov/vital-signs/global-temperature/. Show students the map visualization and encourage them to look for where on earth the temperature has risen the most and where it is currently warmest. This ties into the rest of the lesson.

2. Do you think that rising temperatures affect sea levels? If so, how?

   Students may identify that increased global temperatures cause glaciers/ice sheets to melt which in turn contributes to rising sea levels. Depending on their previous knowledge, they may talk about specific glaciers or general Arctic/Antarctic melt, but this lesson will focus on the melting of the Greenland Ice Sheet. As you have these discussions with students, it will be important to point out that ice melt is not the only reason for rising sea levels. The thermal expansion of water (also caused by warming temperatures) is another contributing factor.

To introduce the data we will be looking at, show students the following news report on Greenland (3:43): https://youtu.be/cgblPFbLxAE

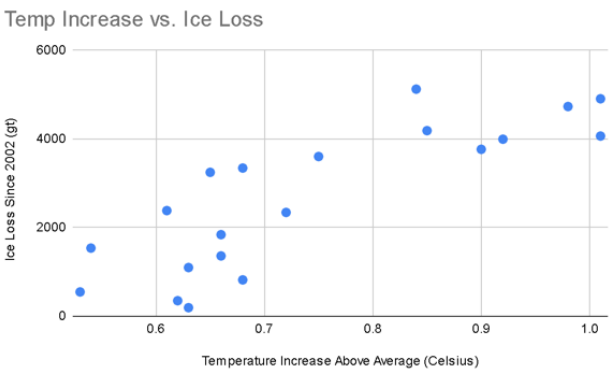
After showing the video, have students open the data file (or open yourself if you are doing this as a demonstration). The data contains three variables:

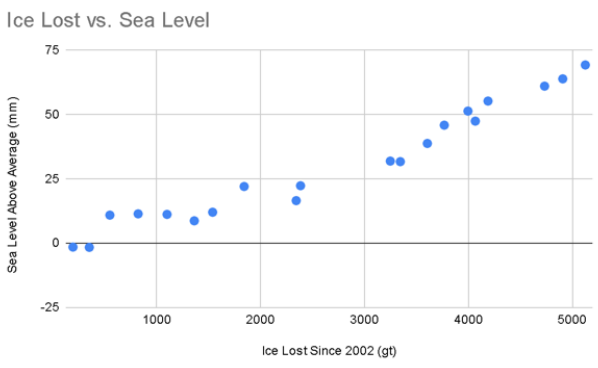
- Global temperature change (in degrees Celsius) above the 1951-1980 average
- Cumulative ice loss (in gigatonnes) of the Greenland ice sheet compared to initial measurements taken in 2002
- Global sea level change (in millimeters) compared to the CSOL01 mean sea surface

  (Note: for the ice loss and sea level data, only the last measurement taken during the given year is included to simplify the creation of scatter plots and best fit lines. If interested, the entire datasets are linked at the bottom of the spreadsheet.)

Ask students if they notice any trends in the data. Do the observed trends match what the class discussed initially?

Have Google Sheets create two scatter plots: temperature vs. ice loss and ice loss vs. sea level. Consult the first page of this Google Sheets Tutorial for assistance. The scatter plots should look like this:





Using the graphs, have students "tell the story" of the data. Based on the data, is it reasonable to infer that rising temperatures have an effect on sea level? Is it safe to say that temperature increases alone are causing the ice to melt? Is it safe to say that the ice melt alone is causing the sea to rise? (The answer is "no." It is safe to say there is an association between the variables, but not that temperature alone is causing sea levels to rise as there are other contributing factors.) What other factors may be contributing to what this data is showing us?

After discussing the questions, show this video (6:08) that describes many of the additional variables that are causing sea levels to rise: **Sea level rise is so much more than melting ice**

Option for after the discussion:

Have students create a scatter plot for the temperature data vs. sea level. Does it tell the same story? Brainstorm on what humans can do to slow sea level rise.

Suggestions for Unit Integration

Throughout the unit, refer back to the dataset to reinforce and relate to the concepts that are being taught. Resources and ideas are given below.

1. Use the graphs to explain the purpose of scatter plots and how they can be used to show trends in bivariate data. Ask students to describe what shape they see in both graphs. Talk about linear relationships in bivariate data and how a linear function could be used to model the data. Discuss what linear function would best fit the data in the graphs. Introduce the concept of a residual and talk about how whatever line we use we want the distance between the line and the actual data point to be as small as possible. Have students approximate the equation of the line for both plots. This would also be an appropriate time to teach students that while both of these plots have positive linear relationships, bivariate data may also have a negative linear relationship, no linear relationship, or even an exponential relationship. Be sure to show them what scatter plots with those relationships would look like.

2. Utilize Google Sheets to insert the trendline into the plots. Have students use the trendline to predict outcomes. For example, how much ice could we expect to melt in Greenland if the average temperature this year is 0.9 degrees Celsius above average? How far above average could we expect sea level to be if 6000 gigatonnes of ice melt in Greenland this year? If students are interested in knowing what would happen if all of the ice on Greenland melts (8:25): **What if GREENLAND Melted?**

   Also, use Google Sheets to calculate the line of best fit for both plots. Consult the second page of this **Google Sheets Tutorial** for assistance. The line of best fit for temperature vs. ice lost should be y=8729.91x-3817.62. The line of best fit for ice lost vs. sea level should be y=0.0135x-5.53. Be sure to explain what the slope and y-intercept represent for both lines. For example, in the first line for every degree above average the temperature is, roughly 8,730 gigatonnes of ice on Greenland melts. The y-intercept in this instance represents the amount of ice "lost" (an approximate 3,818 gigaton increase since the y-intercept is negative) if the temperature for a year was 0 degrees above average.

3. Explain to students that the correlation coefficient (r) is a measure of the strength of relationship between two variables. While the formula to calculate this value by hand is complicated, Google Sheets can be used to find it by using the =CORREL() formula in an empty cell. Make sure you put the response variable range first and then the explanatory variable range. Teach students how to estimate the correlation coefficient from the two graphs and point out that the second graph has a correlation coefficient closer to r=1 than the first graph because it more closely resembles a perfect line/has smaller residuals. Also show students how to identify bivariate data with a negative or zero correlation coefficient. It would be good to point out that bivariate data with an exponential relationship would not

4. As stated in the Missouri Learning Standards, students must understand that correlation does not necessarily imply causation. It is important to emphasize that while the data suggests a significant trend between temperature and ice loss as well as ice loss and sea level, the trend alone does not demonstrate a causal relationship between the variables. Additional studies that account and/or control for the multitude of other variables affecting the response variable would need to be conducted prior to asserting that one thing causes another. This can further be demonstrated via the website https://www.tylervigen.com/spurious-correlations. Another resource to emphasize this point is the following Crash Course video: https://www.youtube.com/watch?v=GtV-VYdNt_g&list=PL8dPuuaLjXtNM_Y-bUAhblSAdWRnmBUcr&index=10