8TH GRADE 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 examine scatter plots and interpret data. They will be able to understand different patterns of association and lines of best fit. They will use linear models of two variable data to explain and model real life situations. This module serves as a starting point for instruction related to the following Missouri Learning Standards:

Math:

- 8.DSP.A.1 - Construct and interpret scatter plots of bivariate measurement data to investigate patterns of association between two quantities.
- 8.DSP.A.2 - Generate and use a trend line for bivariate data, and informally assess the fit of the line.
- 8.DSP.A.3 - Interpret the parameters of a linear model of bivariate measurement data to solve problems.

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 8th grade functions standards, including:

8.F.A.1c Graph a function.

8.F.A.3a Interpret the equation y = mx + b as defining a linear function, whose parameters are the slope (m) and the y-intercept (b).

8.F.A.3b Recognize that the graph of a linear function has a constant rate of change.

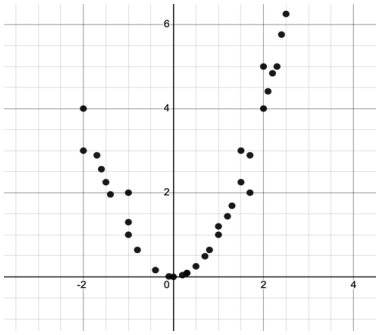
8.F.A.3c Give examples of nonlinear functions.

8.F.B.4a Explain the parameters of a linear function based on the context of a problem.

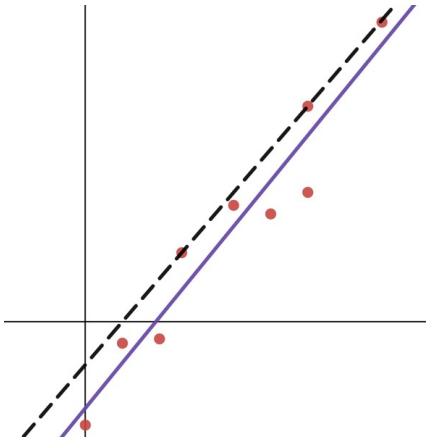
8.F.B.4b Determine the parameters of a linear function.

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.



8th Grade Data Science Math Module

Example: Life Expectancy and Income

Question: How does income relate to life expectancy?

Data: Gapminder Website

Goal: Students will see the need for sampling.

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 is life expectancy? What is income?
2. Do you think it is generally true that rich people live longer? Why or why not?
3. If you live in a rich country like America or Germany, are you more likely to live longer?

After you talk for a few minutes, write the general sense on the board about each question.
If the room doesn't agree, maybe write some reasons about why they think yes or no.

Show students the Gapminder graph (and give some time to look at the graph) and ask:

1. What does each bubble represent?
2. What are the two variables on the graph? What is the X-axis and what is the Y-axis?
3. Some countries are toward the right-hand side of the graph and some are on the left-hand side. What does that mean?
4. Some countries are on the higher (North) part of the graph and others are on the lower (South) part of the graph. What does that mean?
5. Do you see any pattern related to income and life expectancy? If so, what shape does the relationship have? Write what you think about their relationship in a sentence or two.
6. Look at what you wrote at the beginning. Does the data match your intuition? Why do you think income and life expectancy are or aren't related?

Students might ask:

7. Why are the bubbles different sizes?
8. Why are the bubbles different colors?

If not yet discussed, the teacher might point out:

1. The size of the bubble is the population of each country.
    a. Notice two big red bubbles - China and India
    b. Where is the United States?
    c. If you move mouse close to bubble, it will tell which country the bubble represents.
2. The color of the bubble is the continent the country is in.
3. Life expectancy and income change over time. Try the triangle button at the right bottom corner.
4. Often, two variables are not enough to tell a complete data story. But, it might seem tricky to make a scatterplot with more than two variables, especially when some variables are categories, not numbers. Here, we use the size and color of the bubbles to consider additional variables. This is called "multivariate thinking."
5. If it is a computer lab students can play with different buttons on the web comparing different countries such as the United States and China, Russia and Ukraine, France and Germany, South Korea and North Korea, etc.

Option for after the discussion:

Watch the **Hans Rosling** 2-minute video.

Questions about the video:

1) Did he explain anything on the graph that didn't make sense before?
2) Does he think there is a relationship?

Now, let's focus on two variables and learn some mathematics behind them.

An alternative approach to this discussion is to first show students a simplified version of the graph without the context of income and life expectancy. Then slowly scaffold the visualization, revealing more information as factors are discussed one-by-one. This **Google Slides presentation** uses the "**slow reveal graph**" instructional routine to guide the discussion.

Suggestions for Unit Integration

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

Scatter Plots

Types of Association

After you have discussed different types of association (positive, negative, and no association), go back to the Gapminder Income vs. Life Expectancy visualization. Ask students to describe what type of association they see. They should observe that the data has a positive correlation.

Extension: consider discussing income vs. other variables to show different types of association. Examples for discussion are provided in this **presentation**.

Creating a Scatter Plot with Technology

After students have practiced creating a scatter plot by hand, have students use Google sheets to create a scatter plot of the Income vs. Life Expectancy data. Here is a **Google Sheets Tutorial** describing how to create a scatter plot. Note: this document also includes directions for creating a trendline, but you will only want to follow the directions for creating a scatter plot on the first page.

You can have students start with this **Income vs. Life Expectancy dataset**. The first sheet includes income and GDP per capita for 2018. If you would like to extend your exploration (perhaps looking at life expectancy over time), the second sheet includes historical data as well as population. Note: You may want to consider discussing Qatar as an outlier.

Consider asking students to use their scatter plot to make predictions. Possibilities include:

- How long would you expect to live if your home country's average income was $50,000? $150,000? Which of these questions is easier to answer? Why?
- What income would you expect a country to have if the life expectancy is 80 years? 50 years?

Discuss the wide range of responses that different students may give. It is difficult to make accurate predictions based solely on looking at the scatter plot.

Line of Best Fit

Approximating a Line of Best Fit

Print the scatter plot that students created using the Income vs. Life Expectancy dataset. A completed scatter plot is also available **here**. Have students approximate a line of best fit for the data.

Ask students to make predictions based on their line. Revisit the questions listed above.

In addition, ask:

- Are these questions easier to answer now than they were with just the scatter plot?
- Are our answers closer to one another now?

Using Technology to Generate a Regression Equation

Have students use Google sheets and the Income vs. Life Expectancy scatter plot that they created earlier to make a trendline and compare it to their hand-drawn approximation. Use the regression feature of Google sheets to generate the equation for the line of best fit. See page two of the **Google Sheets Tutorial** for directions. A completed version is available **here**.

Discuss with students what the slope and y-intercept represent in this context. The slope represents years of life expectancy gained per dollar of income, and the y-intercept represents the baseline life expectancy if the GDP per capita was $0.

Now have students use the regression equation to answer the questions that they were asked to make predictions for earlier. How do the results compare?

Extension/Summary Activity

Have students select a country from sheet two of the **dataset** (or any other dataset of interest). Ask them to make their own copy of the sheet and delete the information for the countries they are not using. Ask students to complete the following:

- Create a scatter plot of the life expectancy of their country over time.
- Generate the regression equation and add the trendline to their scatter plot.
- Write several sentences describing their data. What type of association does the data have? What story does it tell? Does anything unexpected happen? Are there any outliers?
- Have students generate several questions about their data. For example: What was the life expectancy in 1952? What would you expect the life expectancy to be in 2030? In what year would you expect the life expectancy of the country to be 60? Have students answer these questions or trade with a classmate and answer one another's questions.
- You may want to have students go back to the Gapminder page and reconsider the answers to their questions from the first class discussion.

Sources and Links

Bubbles Tool. (n.d.). Gapminder. Retrieved August 16, 2022, from **https://www.gapminder.org/**

Life expectancy vs. GDP per capita. Our World in Data. (n.d.). Retrieved August 16, 2022, from **https://ourworldindata.org/grapher/life-expectancy-vs-gdp-per-capita**

Introduction. Slow Reveal Graphs. (n.d.). Retrieved August 16, 2022, from **https://slowrevealgraphs.com/introduction/**