SeaGlide_TEAM	Experimental Design & Causation	Lesson # 6 (65 minutes)

Description:

In experiments it is important to determine causation,i.e. the relationship between cause and effect. You are working for a global activist company as an analyst. Your boss wants you to evaluate global data then look for and report any present causation. Using GapMinder, you will explore causation and provide critical analysis of the data shown. You will also use GapMinder to create an original hypothesis, analyze data pertaining to your hypothesis, and determine causation.

Students will be able to:

- Explain what a controlled experiment is
- Use data to answer questions
- Look for relationships and causation within global data

Students will understand:

Students will understand the importance of experimental design and how a proper experiment is constructed and carried out. They will be able to define various aspects of experimental design and understand how samples and test groups are chosen for a given study. Students will practice answering questions using the data and statistical information that is provided for them through an online interface knows as GapMinder. Students will use this interactive tool to better explore data sets, correlation, and causation. They will be encouraged to propose hypotheses of their own and search for answers by using GapMinder.

Key Definitions & Concepts [1]:

- **Experimental Design**: how participants are allocated to the different conditions in an experiment
- **Simulation**: imitation of a situation or process.
- **Correlation**: a mutual relationship or connection between two or more things.
- **Causation**: the relationship between cause and effect

Standards[Copied from: 2]:

CC.2.4.HS.B.3 Analyze linear models to make interpretations based on the data.
CC.2.4.HS.B.4 Recognize and evaluate random processes underlying statistical experiments.

	Background Information	

Prior Knowledge:

- All previous *Measurement and Data Analysis* lessons
- Basic graphing
- Understanding of variables

Math Practices [Copied from: 3]:	**Core Ideas [Copied from: 4]:**	**Cross Cutting Concepts [Copied from: 5]:**
• Construct viable arguments and critique the reasoning of others. • Model with mathematics. • Use appropriate tools strategically.	• Defining and Delimiting Engineering Problems • Developing Possible Solutions	• Cause and Effect • Stability and Change • Patterns

Possible Preconceptions/Misconceptions:

Students may not realize how many different types of experimental design there are. They may also not understand the reasoning behind certain aspects of experiments such as controls. Students may also assume that the question they ask in the Investigation part of the explore has a correlation when it may not. When there is not a legitimate correlation, but when the data is used as if the correlation exists, then false interpretations have a greater probability of being widespread. Understanding the root cause of a problem is also important in rectifying an issue. Hence, assuming a false correlation is not going to help address any real problems.

Lesson Plan - 5E(+) Model

Engage [1]:

Students will watch the video Controlled Experiments: Crash Course Statistics #9. This video goes through the importance of experimental design and teaches the most important concepts surrounding experimental design in an engaging way. Students will fill out the *Crash Course: Statistics #9* worksheet while watching the video. This will allow them to easily retrieve and record the important information discussed in the video. It is recommended to warn the students that the questions are not directly in order with the video, nor are they explicitly answered. Hence, the students will have to pay close attention to the video in order to answer the questions on the worksheet. The students should complete their answers in under 15 minutes since they are working along with the video. Allow for an additional 5 minutes to conduct a whole class discussion geared towards correcting the worksheets. Hence, this activity should take 20 minutes to complete.

Explore [1], [6]:

Part I: Introduction

The teacher should hand out the *Answering Statistic Based Questions* worksheet and pair the students into groups of 2. Each group will need internet access via one tablet or one computer per group to complete this worksheet.

Part II: *Benchmark Lesson*: Answering Statistic Based Questions

Students can work in pairs to complete the *Answering Statistic Based Questions* worksheet. This worksheet guides students in working with GapMinder to answer statistical questions and to search for causation. This activity allows students use real world data and connect it to experimental design as

they are answering questions that have solutions developed through statistics. Students will graph various variables on the x- and y-axis to search for relationships between the data and determine causation. They will also be encouraged to propose hypotheses of their own and discover how answering one question may lead to the appearance of further questioning in order to develop a deeper understanding and solution to the proposed hypotheses. This may occur because students may wonder why their is a correlation but not a clear causation. This lends to the discover that correlation does not mean causation exists, and provides the opportunity for the instructor to introduces a discussion about how this phenomenon exists. Recall that correlation is a statistical technique showing how strongly a pair of variables are related, i.e. a value. Causation, however, provides reasoning for a change in one variable resulting in a change of another variable, i.e. cause and effect. This activity is designed to help students discover this important difference in an interactive, engaging way. This activity should take 25 minutes to complete.

Part III: *Investigation Lesson*: Developing a Question

Students will use the *Developing a Question* worksheet and GapMinder to develop and inquiry of their own and to search for the answer. Since they will be working on this activity individually, each student needs to have internet access via a tablet or computer. Each student should devise an individual and unique hypothesis, then utilize GapMinder to explore probable causation related to their hypothesis. The worksheet has guided steps to help the students use GapMinder as an effective tool. Some students may have difficulty finding a question that gives them a correlation. This is okay. In a way, this scenario answers the question by showing that there is no effect of the x-axis variable on the y-axis variable. This worksheet encourages students to ask questions and find answers by using common statistical techniques. This worksheet is designed to be individualized, so students should be encouraged to be creative. The purpose of this design was to enhance the activity by making it innovative and tailored for students to have a voice and choice in their statistical analysis. This activity should take 20 minutes to complete.

Explain:

Throughout the exploration of this lesson, students are constantly ask to explain why things are happening and to provide reasoning for their answers. They must explain their reasoning and support their answers with the statistical data from GapMinder. They also must search for answers as to why something is happening and explain the correlation and causation of their hypotheses. The teacher should also be asking probing questions during completion of the worksheets that prompt the students to explain what they are doing and why. This allows for students to learn through inquiry and discovery, which ensures that they gain deeper understanding of the learning goals of this lesson.

Elaborate:

Students can connect to real-world examples as they are using real, tangible data about the world that has been collected. This data has information about the past, the present and predictions about the future. The data is pertinent to current society, so students can relate to it more easily. Students can see the impact of statistics on the world and how good experimental design can help address real life issues facing our world. The lesson is set up in a fashion of having the students work as an analyst for a global activist company. Hence, students are exposed to an additional field in which statistics is

relevant in the real world. This helps to pull students' attention as they will begin to develop and deeper understanding of the uses of data and how it can affect them in their everyday lives.

Evaluate:

Students are evaluated both informally and formally throughout the entirety of this lesson. The informal evaluations occur during the open class discussions. When the teacher is circulating the classroom, they are able to check for surface level understanding and make sure that the class is all on the same page by listening to students' discussions and observing students' responses. The instructor can gauge student progress through observing whether or not they are struggling to complete the activities. The formal assessments of this lesson are the worksheets, by checking for correctness, especially for the *Answering Statistic Based Questions* worksheet. Teachers can see if the students have strong conceptual grasps on the concepts through the *Developing a Question* worksheet, depending on how successful the students are in developing a hypothesis and using GapMinder to search for a solution.

Enrich:

This can easily be differentiated into a World History class because GapMinder can be used to understand global issues. Students can analyze information, find causation that leads to new questions, and search for possible solutions. Students can also explore data from the past to see where the world was versus where it is today. Students can look at major events in world history to try to determine how it is that event influenced our current society while also being able to make predictions about the future.

All associated documents are attached below

Reference *Annotated Bibliography* on the very last page of this packet

Name:_____ Date:_____

Crash Course: Statistics #9 [1]

As you watch the video Controlled Experiments: Crash Course Statistics #9 answer the following questions.

1. What is a simulation?
2. What is random assignment? What does it prevent?
3. What are two types of bias?
4. What are treatments?
5. What is a control?
6. What is the purpose of a control?
7. What is a placebo?
8. What single blind study?
9. What is a double blind study?
10. What is repeated measure design?

Name:_____ Date:_____

Answering Statistic Based Questions [6]

In experiments, it is important to determine causation. You are working for a global activist company as an analyst. Your boss wants you to evaluate global data then look for and report any present causation. Using GapMinder, you will explore causation and provide critical analysis of the data shown.

There are 5 types of tools that can be used: Bubbles, Income, Maps, Ranks, and Ages. Look at each of the five.

1. Which of the tools can be used to determine possible causation? Why?

Go to the Bubbles tool. Notice that, by selecting the axis, it can be changed to a different data set. By changing the each axis, we can test to see how one factor may affect the other. Test the following scenario for correlation.

The first graph using the bubbles tool should show income vs life expectancy.

1. Record some basic observations about the graph and the bubbles tool.

2. Is there a relationship between life expectancy and income? If so, what is it?

3. Do see any causation from this data?

4. What other experiments might you set up to go deeper into this problem and answer possible further questions?

Now leave the x-axis as income. Click on the y-axis and select education, then mean years in school, then choose women 15-24.

1. Is there a relationship between mean years in school and income? If so, what is it?

2. Do see any causation from this data? Explain your answer.

3. Why do think this causation exists?

Next leave the y-axis mean years in school for women 15-24. Click on the x-axis and select babies per woman.

1. Is there a relationship between the mean years in school for young women and the number of babies per woman? If so, what is it?

2. Do see any causation from this data? Explain your answer

3. Why do think this causation exists?

4. What are some other questions you have from this data and how might you answer them?

Click on the x-axis and select infrastructure, then sanitation, then basic sanitation overall. Click on the y-axis then select environment, then forestry, then forrest coverage.

1. Do you see a relationship on this graph? Explain your answer.

2. Is there any causation? Why or why not?

Now leave the x-axis the same and change the y-axis to life expectancy.

1. How does sanitation affect life expectancy? Why do you think this is?

2. What other questions might you want to answer addressing this issue, and how might you use data to answer it?

Click on the x-axis, then select energy, then electricity, then choose electricity generation per person. Click on the y-axis, then select CO_2 emissions

1. What relationship do you see?

2. What about a causation? Why or why not?

Name:_____ Date:_____

Developing a Question [6]

Your boss wants you to create an original hypothesis, analyze data pertaining to your hypothesis, and determine possible causation by utilizing GapMinder. You can spend a few minutes looking at the different data sets before creating a question. Spend no more than 5 minutes doing so.

My question I want to answer using data and statistics:

_____

On the x-axis I will plot: _____

On the y-axis I will plot: _____

Why is it important to plot the correct variable on each axis?

My graph looks like….. (provide a sketch or detailed explanation)

The relationship on my graph is….. (discuss both correlation and causation)

This answers my question by telling me…..

Name:_____ANSWER KEY_____ Date:_____

Crash Course: Statistics #9 [1]

As you watch the video Controlled Experiments: Crash Course Statistics #9 answer the following questions.

1. What is a simulation?

   Imitation of a situation or process

2. What is random assignment? What does it prevent?

   Randomly choosing groups to test where there is equal chance of each subject being selected into a group.

3. What are two types of bias?

   Allocation bias
   Selection bias

4. What are treatments?

   Condition that we want to test

5. What is a control?

   When a group is given a treatment of nothing. (ie no treatment)

6. What is the purpose of a control?

   Helps us determine changes due to treatments and changes due to something else such as time

7. What is a placebo?

   This is when someone is being treated with a non-effective treatment (i.e. sugar pill versus medicinal pill).

8. What single blind study?

   When the the research participants do not know which group they are in or which treatment they are receiving

9. What is a double blind study?

   When both the research participants and the researchers do not know which group they are in or treatment they are receiving

10. What is repeated measure design?

    When the same subject is given multiple treatments one at a time to see how they react to each

Name:_____ANSWER KEY_____ Date:_____

Answering Statistic Based Questions [6]

In experiments, it is important to determine causation. You are working for a global activist company as an analyst. Your boss wants you to evaluate global data then look for and report any present causation. Using GapMinder, you will explore causation and provide critical analysis of the data shown.

There are 5 types of tools that can be used: Bubbles, Income, Maps, Ranks, and Ages. Look at each of the five.

1. Which of the tools can be used to determine possible causation? Why?

   The bubbles tool because it tests two variables and shows how inferences can be made about cause and effect. The bubble tool also shows observational relationships in the data The trend tool can be used but only if time is the only other variable that is being taken into consideration.

Go to the Bubbles tool. Notice that, by selecting either axis, it can be changed to a different data set. By changing the each axis, we can test to see how one factor may affect the other. Test the following scenario for correlation.

The first graph using the bubbles tool should show income vs life expectancy.

1. Record some basic observations about the graph and the bubbles tool.

   Students' answers may vary, but expect any of the following: The bubble sizes represent population of the given country. Time can be played to see data from different years. It looks like a scatter plot etc.

2. Is there a relationship between life expectancy and income? If so, what is it?

   It looks like there is a positive linear correlation.

3. Do see any causation from this data?

   Higher income leads to a higher life expectancy

4. What other experiments might you set up to go deeper into this problem and answer possible further questions?

   Why does higher income lead to a longer life?

Now leave the x-axis as income. Click on the y-axis and select education, then mean years in school, then choose women 15-24.

1. Is there a relationship between mean years in school and income? If so, what is it?

   Yes there is a positive linear correlation

2. Do see any causation from this data? Explain your answer.

   Countries with higher income keep girls in school longer

3. Why do think this causation exists?

   Students answers may vary, but some reasoning follows: countries with lower income either can't afford to keep students in school because it's too expensive or they have a lack of resources. Children may have to work to make more money. etc.

Next leave the y-axis mean years in school for women 15-24. Click on the x-axis and select babies per woman.

1. Is there a relationship between the mean years in school for young women and the number of babies per woman? If so, what is it?

   There is a negative linear correlation.

2. Do see any causation from this data? Explain your answer.

   The less time women spend in school, the more babies they are likely to have.

3. Why do think this causation exists?

   Countries that keep women in school less time may value women as homemaker over her having an education or career. (allow for creativity)

4. What are some other questions you have from this data and how might you answer them?

   What causes women to leave school early? (allow for creativity)

Click on the x-axis and select infrastructure, then sanitation, then basic sanitation overall. Click on the y-axis then select environment, then forestry, then forrest coverage.

1. Do you see a relationship on this graph? Explain your answer.

   No the point are scattered everywhere in no distinguishable pattern

2. Is there any causation? Why or why not?

   No there is no relationship between the two variables.

Now leave the x-axis the same and change the y-axis to life expectancy.

1. How does sanitation affect life expectancy? Why do you think this is?

   Better sanitation leads to higher life expectancy. This is most likely because sanitation prevents the spread of disease.

2. What other questions might you want to answer addressing this issue, and how might you use data to answer it?

   What diseases are spread from lack of sanitation? Search for reported cases within multiple countries etc. (allow for creativity)

Click on the x-axis, then select energy, then electricity, then choose electricity generation per person. Click on the y-axis, then select CO_2 emissions

1. What relationship do you see?

   Positive linear

2. What about a causation? Why or why not?

   Higher electricity use leads to more CO2 emissions.

Name:_____ANSWER KEY_____ Date:_____

Developing a Question [6]

Your boss wants you to create an original hypothesis, analyze data pertaining to your hypothesis, and determine possible causation by utilizing GapMinder. You can spend a few minutes looking at the different data sets before creating a question. Spend no more than 5 minutes doing so.

My question I want to answer using data and statistics:

EXAMPLE: Does income affect life expectancy (this should not be used by students since it is the example provided by GapMinder)

On the x-axis I will plot: _____Income_____

On the y-axis I will plot: _____Life Expectancy_____

Why is it important to plot the correct variable on each axis?

This is because the x-axis is the independent variable and the y-axis is the dependent variable. Hence, we are attempting to test whether or not the x-axis variable affects the y-axis variable.

My graph looks like…..(provide a sketch or detailed explanation)

Bubbles that have a positively increasing relationship. (This is dependent on students' individual graphs)

The relationship on my graph is…..(discuss both correlation and causation)

Positive linear trend. (some students may get a negative trend, exponential or no trend). The correlation shown is that a higher income results in a higher life expectancy. The causation is relative, but some possibilities are as follows: higher income means having more flux money for unexpected doctor's visits, car maintenance, higher education, etc.(This is dependent on students' individual hypothesis).

This answers my question by telling me…..

That higher income leads to a higher life expectancy. (This is dependent on students' individual hypothesis).

Annotated Bibliography

[1] CrashCourse. (2018, March 21). Controlled Experiments: Crash Course Statistics #9. Retrieved from https://www.youtube.com/watch?v=kkBDa-ICvyY

This video was used as an instructional for students to watch and complete a worksheet that has questions developed based on the video. Definitions were also taken directly from the video.

[2] Standards Aligned System. (n.d.). Retrieved from https://www.pdesas.org/

This website was used in each lesson in the Measurements and Data Analysis module to select proper Pennsylvania State standards, which are based in Common Core, that each lesson is centered around.

[3] Standards for Mathematical Practice. (n.d.). Retrieved from http://www.corestandards.org/Math/Practice/

This website used in every lesson in the Measurements and Data Analysis module to find Standards for Mathematical Practices that are applicable in each lesson.

[4] Nsta. (n.d.). Disciplinary Core Ideas. Retrieved from https://ngss.nsta.org/DisciplinaryCoreIdeasTop.aspx

This website was used in each lesson in the Measurements and Data Analysis module to select appropriate disciplinary core ideas set forth by the NSTA that are at the center of each lesson.

[5] Nsta. (n.d.). Crosscutting Concepts. Retrieved from https://ngss.nsta.org/CrosscuttingConceptsFull.aspx

This website was used in each lesson in the Measurements and Data Analysis module to selecting appropriate crosscutting concepts set forth by the NSTA that apply to each mathematics lesson.

[6] GapMinder. (n.d.). Retrieved from https://www.gapminder.org/

This is an online tool centered around data and statics that is used as an instructional aid and student exploration in the Experimental Design and Causation lesson in Measurements and Data Analysis module. A student activities Answering Statistic Based Questions and Developing a Question were developed based on the tools provided by GapMinder.