

Data Science in Global Health

Part I: Investigable Questions and Identifying Variables

Overview

For this lesson, the class will be divided into groups. Each group will be assigned a city in Washington State and given some data on air pollution for their city. By graphing the data, you will find out if there is a relationship between the time of year, or season, and the amount of air pollution for your city. Once all groups have determined this for their city, we will see if any trends or patterns can be observed across cities in Washington by compiling our data. You will then write a claim and support it with data to answer the investigative question.

Investigative Question #1

Is there a relationship between the time of year (season) and the amount of air pollution in Washington?

The city our group is investigating is: _____

- 1) With your group, discuss the following questions before writing your answers.
 - a) What is the *changed* variable in this investigative question? (also known as the independent variable or manipulated variable). In other words, the variable that is *affecting* the other variable.
 - b) What is the *measured* variable? (also known as the dependent or responding variable). In other words, the variable that is *being affected* by the changed variable.
 - c) What are some things that you think the scientists who measured the air pollution in your city had to control, or keep constant, in order to get data that is a fair test of the time of year's effect on the amount of air pollution?

Changed variable:

Measured variable:

What did the scientists control/keep constant:

2) What is your hypothesis, or prediction? Do you think that the time of year/season affects the amount of air pollution in Washington? Explain your thinking.

Hypothesis or Prediction:

Explain Your Thinking:

Names: _____ Date: _____ Period: _____

Part II: Graphing Data to See Patterns

Overview

It's hard to see any patterns, or relationships, from raw data. Graphing data makes it easier to see any patterns. Discuss what type of graph you think will best represent your city's data in order to answer the investigative question. You might choose to make three separate graphs or one graph to represent the data for all three years of data. Once you have agreed as a group, use the data table to make a graph of your data on the next page. Then use your graph to answer the graph analysis questions.

Things to remember when making your graph:

- The changed variable typically goes on the x-axis.
- The measured variable typically goes on the y-axis.
- Label your x-axis and y-axis.
- Use different colors to represent different years.
- Include a title for your graph.
- Include a legend (or a key) for your graph.
- If you have missing data in your data set (labeled N/A) just leave that month blank.

Plan it out!

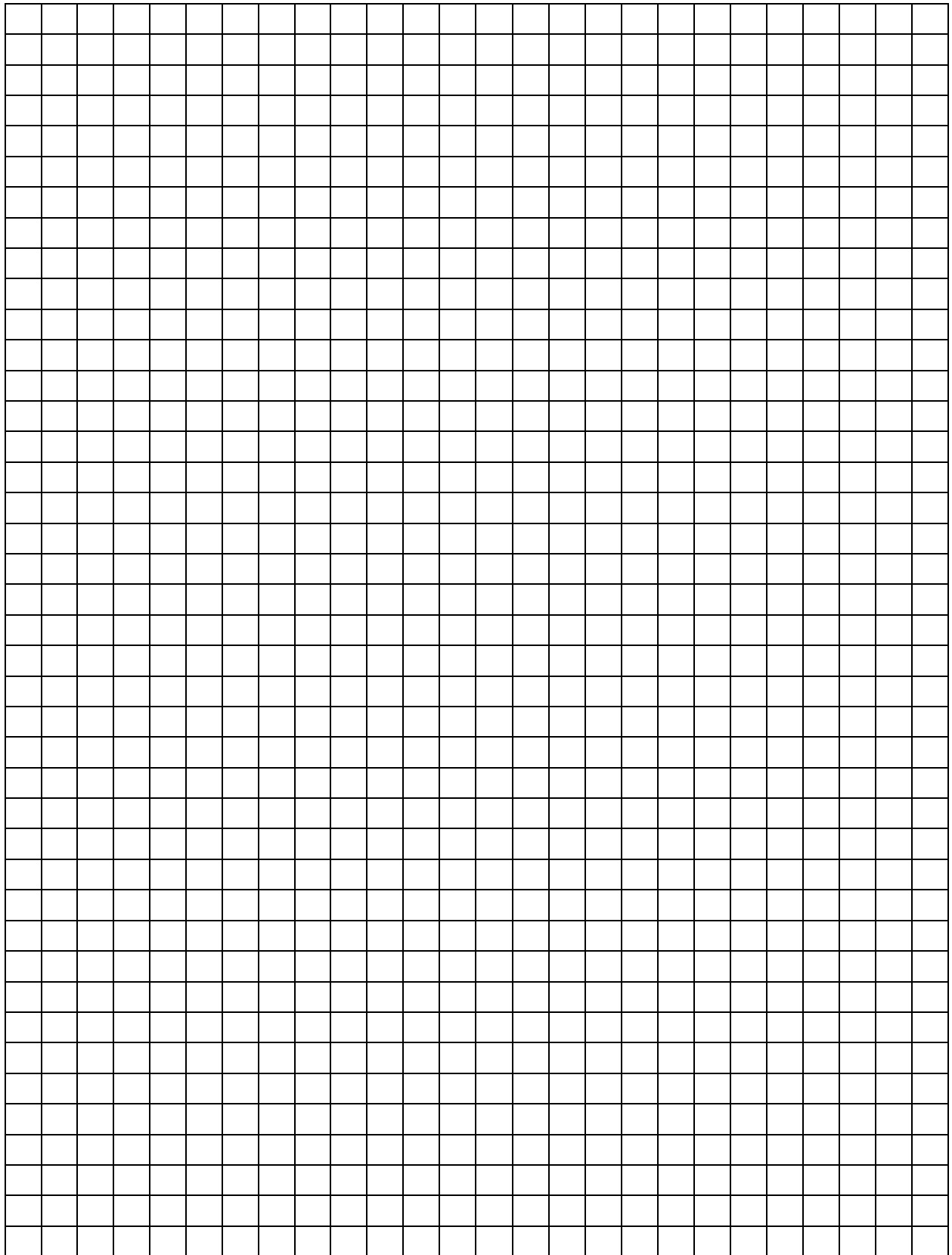
x-axis variable (changed variable):

y-axis variable (measured variable):

How I will represent the different years:

Graph title:

Names: _____ Date: _____ Period: _____



Names: _____ Date: _____ Period: _____

Graph Analysis

Discuss each question with your group before writing your response.

1) For your city, which month had the *highest* measured air pollution in (for the years available)?

2014?

2015?

2016?

2a) Based on the data from just your city, how confident are you claiming that there is a certain month, or season, in which the air pollution tends to be the *highest* in Washington? Explain why you are confident or not confident.

2b) What would make you more confident that there is a pattern in the time of year and the *highest* amount of pollution in Washington? (HINT: what kind of additional data might make you more confident?)

3) Which month had the *lowest* measured air pollution in (for the years available)....

2014?

2015?

2016?

4) Based on the data from just your city, how confident are you claiming that there is a certain month, or season, in which the air pollution tends to be the *lowest*? Explain why you are confident or not confident.

Names: _____ Date: _____ Period: _____

Part III: Making a Claim and Supporting it with Data

Overview

You are ready to write a claim that answers the investigative question. A claim is a one sentence answer to the investigative question that you believe to be true based on evidence. Scientists always support their claims with evidence. Support your claim using the table your teacher made of the compiled class data on the months with the highest and lowest air pollution in cities across Washington.

Investigative Question #1

Is there a relationship between the time of year (season) and the amount of air pollution in Washington?

Claim (a claim is a one sentence answer to the investigative question that you believe to be true based on any patterns observed in your data):

Evidence (the evidence should support your claim and should include a discussion of any patterns you observe in your data. Include quantitative details):

Go back and read your hypothesis on page 1. Does the data support or refute your hypothesis? (It doesn't matter if your hypothesis was correct or not. Scientists base their claims on evidence):

Names: _____ Date: _____ Period: _____

Part IV: Generating Basic Statistics to Answer a New Investigative Question

Overview

In addition to graphing data, another way that scientists organize and interpret data in order to find meaning is through statistical analysis. You will find some basic statistics for your city's air pollution and then compare those statistics with another group in order to answer Investigative Question #2:

Investigative Question #2

How does location (city) affect the amount of air pollution?

1a) In this question, what is the *changed* variable?

1b) What is the *measured* variable?

In order to compare air pollution levels in different cities, you will need to first calculate a few statistics: the minimum, maximum, median, and mean for your city. Write your city's statistics in column 1 of the table below. Once you have calculated your city's statistics, share your results with a different group and copy their statistics into column 2.

Definitions:

- **Minimum** – The smallest observation.
- **Maximum** – The largest observation.
- **Median** – The middle observation in a ranked list of observations.
- **Mean** – The sum of the observations divided by the number of observations.

STATISTIC	CITY 1 (YOUR CITY)	CITY 2
	NAME:	NAME:
MINIMUM		
MAXIMUM		
MEDIAN		
MEAN		

Names: _____ Date: _____ Period: _____

Use the table you just created to answer these questions:

- 2) What interesting differences between cities can you identify?

- 3) What might be some possible explanations for these differences?

- 4) Write a claim that answers Investigative Question #2 and support it with statistics from both of your cities. *Discuss with your group which statistics you think might be most relevant to include in your evidence statement.*

How does location (city) affect the amount of air pollution?

Claim:

Evidence:

Names: _____ Date: _____ Period: _____

Part V: Observing and Asking Questions about the State of Global Air Pollution

Air pollution is not just an issue in Washington; it's a global issue. In this last part, you will explore a map that shows data on global air pollution, make some observations, and write some investigable questions that you would like to further investigate.

- 1) Follow this link to State of Global Air map view: <http://www.stateofglobalair.org/air#PM>
 - Read the information on the webpage.
 - Explore the map: Move the cursor over countries to view levels of PM2.5.

- 2) Study the map and write down two interesting observations you have:

Observation #1:

Observation #2:

- 3) What are you curious about? What would you like to find out more about? Write down two investigable questions. Remember, investigable questions have a *changed* and a *measured* variable. Scientists' questions usually come out of observations they find interesting. Your questions might come out of the observations you wrote down above.

Investigable Question #1:

Investigable Question #2: