**Student Worksheet

Name:**

**Members of your group:**

**Part 1: Goal A: Learn how to interpret data and make a scatterplot**

Examine the data in Table 1. Summarize the data by explaining any trends. What are these data points telling you?

 **Table 1:**

|  |  |
| --- | --- |
| Age of infected person | Number of 2019 flu cases in Anytown |
| 30 | 12 |
| 40 | 35 |
| 50 | 50 |
| 60 | 54 |
| 70 | 100 |

Using the image on the following page, create a graph using the data in Table 1. An appropriate style of graph for this data set is a scatterplot which uses one dot to represent each data point.

**All of these Items should be included in a graph:**
**X-axis:** The horizontal (left to right) line which usually holds the independent variable (the variable that stands alone).
**Y-axis:** The vertical (up and down) line which usually holds the dependent variable (the variable being measured).
**Axis labels**: Labels should be descriptive and include units of measurement.
**Proper scale:** Often, each axis will begin at zero and continue until just beyond the last data point. The axis can begin at a number above zero if the smallest data point is a large number. Tick marks designate each unit of measure.
**Graph description:** A good description should summarize the information (data points and axis labels) in the graph.
**Key/Legend**: If multiple data sets are graphed, add a visual picture (such as different colors or shapes) explaining the data sets on the graph.



After graphing the data, do you have further insight into the data?

**Part 1: Goal B: Understand Common Trend Curves and their use in interpreting data sets**

Now you will use trend curves to further explain the trend. A curve of best fit (or trend curve) is a curve that best connects data points on a scatter plot. The curves of best fit may pass through all of the points or just come near to the points. Trend curves are useful because they can help the reader interpret the data.

Review the following examples of common trend curves.

|  |  |
| --- | --- |
| **Linear trend curve:****Equation**: $y=mx+b$ **Description**: Growth rate is constant and doesn’t change. Graph is a straight line.**Example**: Money earned from a job, without a raise, over a year | **Linear Graph** |
| **Exponential trend curve:****Equation**: $y=Ae^{kx}$**Definition**: Growth rate increases over time**Example**: Growth rate of bacteria dividing by binary fission. | **Exponential Graph** |
| **Logarithmic trend curve:****Equation**: $y=A+Blnx$**Definition**: Growth rate decreases over time**Example**: Height increase of a person from a baby to an adult | **Logarithmic Graph** |
| **Logistic trend curve:****Equation**: $y=\frac{B}{1+Ae^{-kx}}$**Definition**: Growth rate increases then decreases**Example**: Number of oak trees in an acre, over time | **Logistic Graph** |

1. Which trend curve is the best fit for the data in Table 1? Why?
2. How does this trend curve help you interpret the data?
3. Choose a different trend curve (not the one you chose for Question 1) and explain how the data would have been interpreted differently if the data had fit that trend curve instead.
4. Which method (observing data in a table, observing data on a graph, comparing a graph to a trend curve) allowed for the most thorough explanation of the data in Table 1? Why?
5. Based on the trend curve, estimate the number of flu cases one would expect in 80-year olds living in Anytown in 2019?

**Part 2. Goal A. Interpreting complex data sets**

The data set in Table 1 was small enough to graph and visualize the overall trend fairly easily. But when we encounter more complex data, the grapher must spend more time planning out the graph so that a reader can easily interpret the data set. In Table 2, we can see the number of Covid-19 deaths in six countries near the beginning of the pandemic.

The data were pulled from Our World in Data. Current data of the pandemic can be found at the following link.

https://ourworldindata.org/coronavirus#how-can-we-make-progress-against-the-pandemic

**Use the data in Table 2 to answer these questions:**

1. Explain the data point in the top left corner (108). What does the data point mean? When and where was it recorded?

2.What is the largest data point in this set?

3. How do the number of deaths change over time?

4. If we want to graph this data so that it is easier to interpret, what challenges would we face when trying to graph this many data points on the same graph?

**Table 2: Number of Covid-19 Deaths**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Days after the 5th total confirmed death** | **20 days** | **40 days** | **60 days** | **80 days** | **100 days** |
| **United States** | 108 | 18,777 | 65,068 | 93,439 | 112,006 |
| **Brazil** | 299 | 5,017 | 15,633 | 37,134 | 57,622 |
| **Italy** | 2,158 | 15,889 | 25,549 | 31,908 | 32,229 |
| **Canada** | 380 | 2,465 | 5,912 | 7,897 | 8,484 |
| **Mexico** | 406 | 2,704 | 7,394 | 16,448 | 29,843 |
| **Australia** | 36 | 79 | 98 | 102 | 103 |

**Part 2. Goal B. Understand the benefits and drawbacks of various graph types**

There are many ways data sets can be graphed to tell different stories. And there are graph types other than scatterplots, such as the commonly used bar graphs, pie graphs, and line charts. However each type has its uses and drawbacks.

**Review these chart types and then answer the followup questions.

Pie Graph** (Pie Chart or Circle Graph): a subdivided circle that depicts categorical data as slices of a circle, in which the size of each slice is proportional to the frequency count for the category. **Image 1**


1. Summarize the data based on the pie chart (Image 1).
2. Why is this chart not ideal for this data set?
3. Name one other subset of data from Table 2 that could be represented in a pie chart.

**Bar Graph**: graphical display of categorical (qualitative) data. A Bar Graph uses bars of equal width to show the relative sizes of different categories. The bars may or may not be separated by gaps.

Image 3.


Image 4.

****

1. Why does Australia appear to have no deaths in Image 4?

1. Compare the bar graphs in Image 3 and Image 4. What happened to the number of deaths from the twentieth day to the one hundredth day?
2. What problems do you see with using these two bar graphs to observe trends over time? What would the graph look like if we tried to merge these two? Would we be able to see trends over time?
3. Name one other subset of data from Table 2 that could be represented in a bar graph?

**Line chart** is a chart in which adjacent data points are connected by a line. You may consider this line chart as a graph of function over a period of time.

Image 5 ****8. Why did the scale change on the y-axis (compared to Image 4)?

9. What problems do you see with displaying all of this data together?

10. Can we assume that the number of deaths 30, 50, and 70 days fell exactly on these straight lines?  **Part 2. Goal C. Graph data sets and compare resulting charts to trend lines

Based on the data in Table 2, create a scatter plot of this country’s data \_\_\_\_\_\_\_\_\_\_\_\_\_\_.** Create your graph on the image below. Add all of the essential components of a proper graph (from page 1).

****

 **When you have completed your graph, answer these questions.**

1. Compare your data to the trend curves from Part 1.B. Describe the trend curve that best fits your data set.
2. How does this trend curve help explain your data set?
3. Compare your graph with the graphs of the other countries. Were other trend curves seen? What does this mean about the timeline of Covid-19 spread in various countries?
4. Based on only these curves, which countries would be increasing most quickly after day 120?
5. Based on only these curves, which countries would be decreasing most quickly after day 120?
6. Based on only these curves, which countries would we expect to not change much after day 120?

**Part 3: Goal: Use various methods to make predictions.**

In this section, you will use the data and/or the scatter plots completed in Part 2 to predict the next data point (120 days after the 5th death). When you have completed that prediction, you will compare it to the actual value on day 120. You will then be asked to compare methods to discover which method most accurately predicted the future number of deaths.

There are six different methods that are appropriate to use on this data set to predict each country’s next data point. Your group is assigned this/these method(s): \_\_\_\_\_\_\_.

**1) Visual estimation using trend curves**
This method involves a simple comparison of your data to the trend curve that best fits your data set. Draw the curve of best fit over your data and extend it into the future.
 *How many deaths are indicated on the trend curve on day 120? \_\_\_\_\_\_\_\_\_\_\_\_

Show your teacher your prediction and you will be given the actual data that occurred on day 120. Enter that number here: \_\_\_\_\_\_\_\_\_\_\_\_

How accurate was this method in determining the number of deaths for your country on day 120? Explain the pros and cons of using this method.*

**2) Mathematical, based on average rate of change**This method involves finding the average rate of change using the known data points and then applying that rate of change to predict the next data point.

**Equation**: Average rate of change = (y2-y1) / (x2-x1)

x = number of days (independent)
y = number of deaths (dependent)

Find the average rate of change for each time period
20 - 40 days: \_\_\_\_\_\_\_\_
40 - 60 days: \_\_\_\_\_\_\_\_
60 - 80 days: \_\_\_\_\_\_\_\_
80 - 100 days: \_\_\_\_\_\_\_

Using the average rate of change sequence, look for a pattern to predict the rate of change for the time period from 100 - 120 days. \_\_\_\_\_\_\_\_\_

Now apply the predicted rate of change to the data point at day 100. *How many deaths do you predict occurred by day 120? \_\_\_\_\_\_\_\_\_\_\_\_

Show your teacher your prediction and you will be given the actual data that occurred by day 120. Enter that number here: \_\_\_\_\_\_\_\_\_\_\_\_

How accurate was this method in determining the number of deaths for your country by day 120? Explain the pros and cons of using this method.*

**3) Exponential function**
This method uses the exponential function to predict the next data point.
**Equation**: $y=Ae^{kx}$
Complete the equation where

y = cumulative deaths $(y>0)$
A = deaths at time zero (A= \_\_\_\_ , teacher provides)
e = natural base = approx. 2.7181828...
k = growth rate (k > 0 and teacher provides \_\_\_\_\_\_\_)
x = number of days

*How many deaths do you predict occurred on day 120? \_\_\_\_\_\_\_\_\_\_\_\_

Show your teacher your prediction and you will be given the actual data that occurred on day 120. Enter that number here: \_\_\_\_\_\_\_\_\_\_\_\_

How accurate was this method in determining the number of deaths for your country on day 120? Explain the pros and cons of using this method.*

**4)** **Natural** **logarithmic function**
This method uses the natural logarithmic function to predict the next data point.
**Equation**: $y=A+Blnx$
Complete the equation where:

$y$= Cumulative number of deaths

$A$= Cumulative number of deaths when $x=1$ ($A>0$)
$A+B$ = Cumulative number of deaths when $x=e^{}$

$x$ = Number of days
$ln x$ = Natural logarithm of $x$ ($x>0$)

*How many deaths do you predict occurred on day 120? \_\_\_\_\_\_\_\_\_\_\_\_

Show your teacher your prediction and you will be given the actual data that occurred on day 120. Enter that number here: \_\_\_\_\_\_\_\_\_\_\_\_

How accurate was this method in determining the number of deaths for your country on day 120? Explain the pros and cons of using this method.*

**5) Logistic growth function**

This method uses the logistic function to predict the next data point.

 **Equation**: $y=\frac{B}{1+Ae^{-kx}}$

USA: Logistic $y=\frac{111333.50376}{1 + 165.36564 e^{-0.08912x}}$

Brazil: Logistic $y=\frac{70711.55859}{1 + 218.49838 e^{-0.06866x}}$

Italy: Logistic $y=\frac{32220.62416}{1 + 48.45825 e^{-0.09209x}}$

Canada: Logistic $y=\frac{8544.46456}{1 + 88.35674 e^{-0.08835x}}$

Mexico: Logistic $y=\frac{50063.67314}{1 + 157.80241 e^{-0.05448x}}$

Australia: Logistic $y=\frac{102.96376}{1 + 11.35237 e^{-0.09049x}}$

*How many deaths do you predict occurred on day 120? \_\_\_\_\_\_\_\_\_\_\_\_

Show your teacher your prediction and you will be given the actual data that occurred on day 120. Enter that number here: \_\_\_\_\_\_\_\_\_\_\_\_

How accurate was this method in determining the number of deaths for your country on day 120? Explain the pros and cons of using this method.*

**Present your results to the class (or another group) and explain the prediction method(s) that you used.** **Final Questions:***Which method came closest to predicting the number of deaths on day 120? Would this method have worked for all of the data sets? Explain why or why not.*

*After 200 (or whatever days), how did the curve look? Would the class’ best method still have been able to predict far into the future?*

 *What other problems do you think data analysts encounter when trying to predict growth trends?*