Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date \_\_\_\_\_\_\_\_

Class \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Mrs. Garrido/Science

Analyzing Data

**Essential Question**: How can science provide answers to your questions about the world around you?

**Learning Outcome**: I can determine the different features of a graph by analyzing data.

**Kick start**: Analyze the graph “MTA Subway Ridership (weekday average). Answer the questions that follow.

1. What do you see?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What do you think?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. What do you wonder?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Introduction:** We have been studying how science provides answers to our questions about the world around us. We have applied the scientific method and the engineering design process to solve problems in the classroom. In these processes, the collection of data and the interpretation of that data is a key element. What we do with that data gives us information and leads us to our next steps.

**Based on the article provided, answer the following questions:**

1. The purpose of the author writing this article is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What is the authors source? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. What does “ridership” mean? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. What year was the “Average Subway Weekday Dataset” collected? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. Why is the author creating a “visualization”? Which type of chart did the author use? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
6. Why did the author choose colors in figure 27? How are these colors indicated on the chart? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
7. In the “Averaged Total Ridership in Weekday” chart,:
	1. The independent variable is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
	2. The dependent variable is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
8. What should the horizontal axis title be? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
9. Why is this information useful for the MTA? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
10. Ridership in Manhattan, in the year 2014 is around \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Exit Ticket:** Name three features a graph should have. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**MTA Subway Ridership**

June 9, 2015

The New York City subway system is the pulse of this metropolis. 24 subway routes and 468 subway stations spread out on the five boroughs of NYC, carrying residents, visitors and tourists to their destinations on a 24-hour schedule everyday. Having been living in the city for nearly 5 years, I’m curious as to how significant the subway service is to NYC?

I found the MTA ridership dataset from [MTA website](http://web.mta.info/nyct/facts/ridership/). Ridership “consists of all passengers (other than NYC Transit employees) who enter the subway system, including passengers who transfer from buses. Ridership does not include passengers who exit the subway or passengers who transfer from other subway lines, with the exception of out-of-system transfers” ([MTA](http://web.mta.info/nyct/facts/ridership/)). Therefore, how many passenger does the subway serve within a day? Which station is the busiest in the system? How do these numbers change from year to year?

The entire dataset include ridership data of weekday average and weekend average from 2009 to 2014, including bus and subway. I chose the [Average Subway Weekday Dataset](http://web.mta.info/nyct/facts/ridership/ridership_sub.htm) from 2010 to 2014, which cannot be directly downloaded in a neat format. So I used Microsoft Excel and Google Refine to organize the data, and visualized data with Tableau.

Main purpose of my visualization is to show the magnitude of subway ridership and to compare difference among boroughs and stations. I tried different types of charts and found that bar chart best illustrate my point.

I used basic barcharts and adjust them them following examples from the textbook. One particular charts that inspired the first chart is Figure S. 19 from “Now You See It” (p. 110):

Figure 22 from article “Effectively Communicating Numbers” gave me idea on how to present rankings.



Also it is a good idea to use color to specify variables like Figure 27 from “Effectively Communicating Numbers”



In the “Averaged Total Ridership in Weekday” chart, I put *accumulative* data of all boroughs as dependent variable and year as independent variable. Total ridership of all boroughs each year were present in one bar, in order to give readers a big picture of the total amount of ridership. Data component of each borough are color-coded, giving readers a rough idea of portion of each borough. (I thought about assigning colors that are directly associated with boroughs, but that depends on the viewer, so I just assigned colors based on personal preference.) Borough ridership was ranked within each bar, with largest ridership at the bottom. I also added label of each borough on the bars so readers know the borough name without looking away from the chart. Time series information can be obtained from the year label, so I removed the horizontal axis title in order to simply the chart.

These visualizations revealed new information that cannot be easily read from the original dataset.

From the Averaged Total Ridership chart, we can see that the total ridership is gradually increasing. By 2014, averaged total ridership in a weekday is reaching 6,000,000. Ridership in Manhattan, which is around 3,000,000, takes about half of the total ridership, followed by ridership in Brooklyn, which is around 1,000,000, slightly more than ridership in Queens. The largest ridership increase happened in 2014, with ridership in Manhattan contributing nearly half of the total increase. Apparently ridership increase was relatively low in 2012 and 2013, which may be incurred by Hurricane Sandy and following constructions.

For practical purpose, this visualization could be useful for the MTA construction department for reconstruction plans or construction priorities. Also, urban planning people may find this helpful.

Next step with this data is to add this visualization on a subway map, and combine some stations nearby. For example, the 34 St-Penn Station stop on 1/2/3 train and 34 St-Penn Station stop on A/C/E train may be combined, also nearby stops around World Trade Center can also be combined.

http://research.prattsils.org/blog/coursework/information-visualization/mta-subway-ridership/#