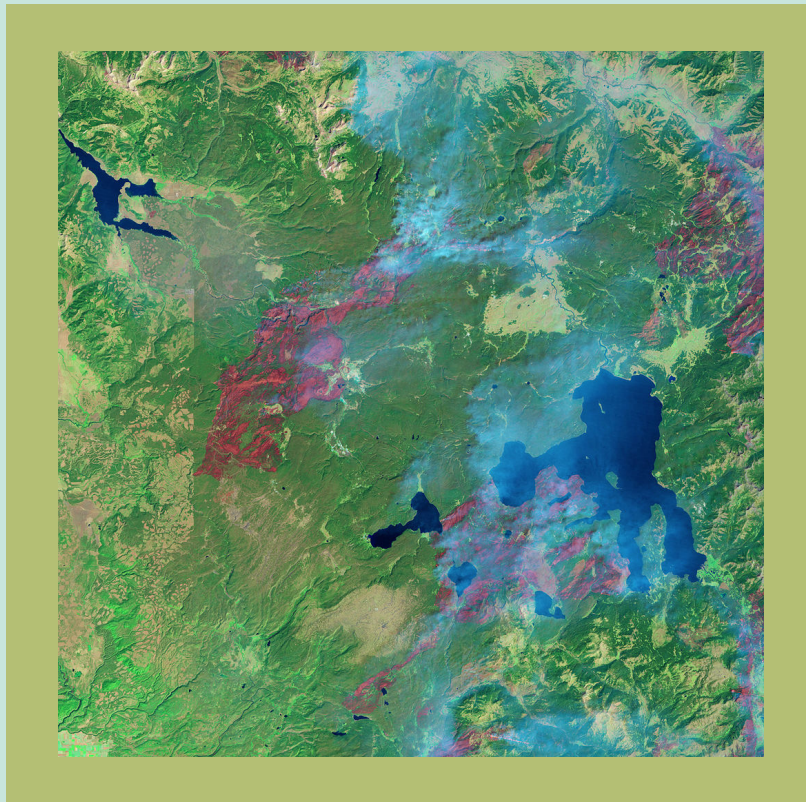


STEM *Sims*™

Data Visualization

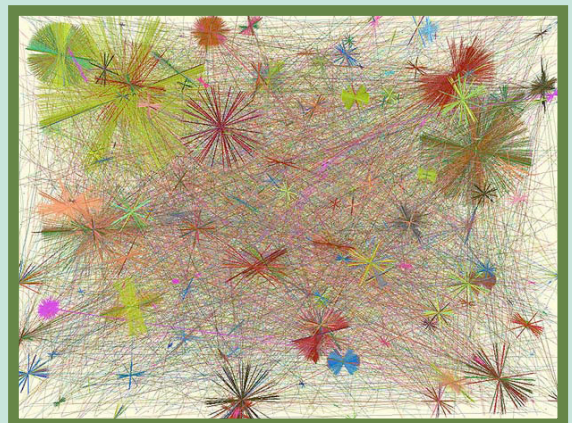


Data Visualization

**Do you need an idea for a scientific study?
Try out one of our ideas or make one of your own.**

Start learning right now about data visualization. Take the following brief quiz to see how much you already know about how data can be represented in a variety of useful and deceptive forms. See the bottom of page 4 to check your answers.

1. In the mid-1850's, an outbreak of cholera in London was halted by removing pump handles from wells.
 - a. true
 - b. false
2. Data visualization is closely related to all of the following *except*:
 - a. norm sampling.
 - b. information visualization.
 - c. statistical graphics.
 - d. scientific visualization.
3. Which of the following software is *not* commonly used for data visualization?
 - a. TinkerPlots
 - b. Tulip
 - c. Tuffy
 - d. Smile
4. TwittEarth data visualization software shows the locations of:
 - a. birds.
 - b. bees.
 - c. Travis Tritt.
 - d. tweets.
5. The data visualization to the right shows a map of the people in Cambridge, England in terms of what:
 - a. song they are listening to.
 - b. color they are wearing.
 - c. television show they are watching.
 - d. time they left for work.



Lie Factor

Edward Tufte discussed how some use data graphics inappropriately in his book called Visual Display of Quantitative Information. He defined the phrase, Lie Factor, to “describe the relation between the size of effect shown in a graphic and the size of effect shown in the data.” In other words, the images presented in a chart or graph can mislead a viewer to make inaccurate conclusions about the data. The “lie factor” of a graphic can be determined by finding the ratio of the difference in the proportion of the graphic elements versus the difference in the quantities they represent. So a graphic representing data with a lie factor = 1 would be the most “truthful” picture of that data. Consider the graphic below that shows how inflation impacted spending power from the years 1958 to 1978.



1958



1978

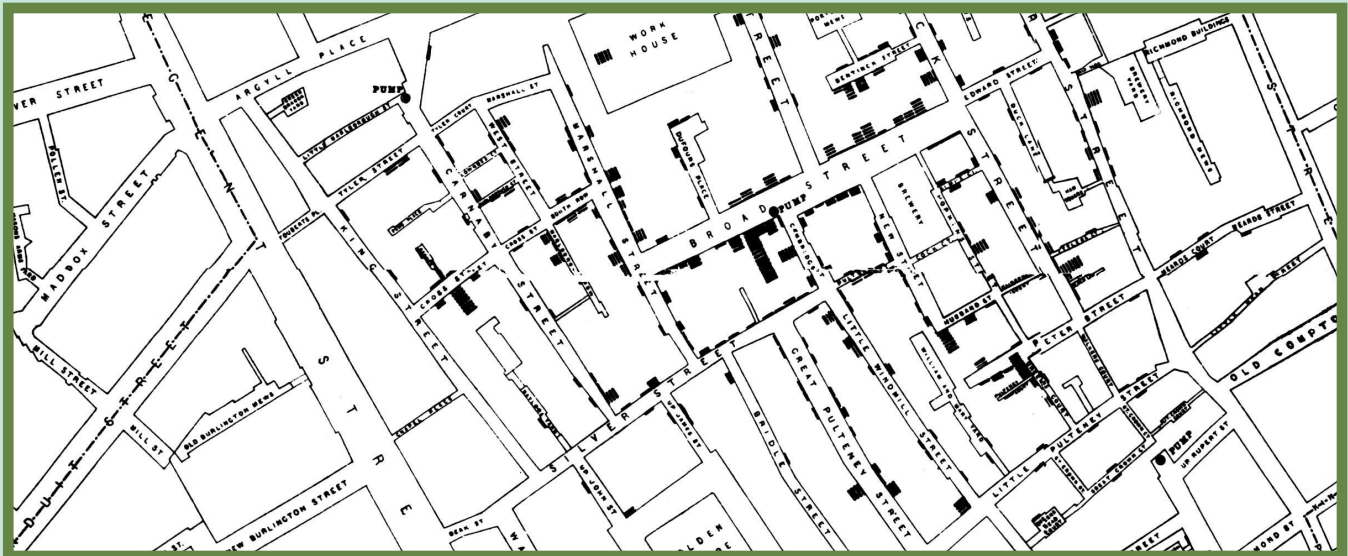
- Use a centimeter ruler and measure the lengths of the two dollar bills.
 Year 1978 dollar bill length = _____
 Year 1958 dollar bill length = _____
 Percent change in length from 1958 to 1978 = _____
- Use a centimeter ruler and measure the areas of the two dollar bills.
 Year 1978 dollar bill area = _____
 Year 1958 dollar bill area = _____
 Percent change in area from 1958 to 1978 = _____
- What is the “lie factor” of the dollar bill inflation graphic?

Data Visualization

Data Visualization History

The use of data visualization has been around for quite some time. One of the earliest applications of this tool was when a local doctor used data visualization to find the source of an outbreak of cholera in London. The water-borne bacteria, *Vibrio cholera*, attacks a human's small intestine and causes the illness. This leads to severe diarrhea and vomiting. Before antibiotics, this led to an often fatal outcome of the disease. Even today, around 100,000 people die each year worldwide from the infection.

Dr. John Snow created a map of the cholera outbreak similar to the one shown below. He studied the map and started interviewing people. One of his questions was about where the infected people got their drinking water. He noticed infection clusters around certain wells, leading him to correctly believe that the well water was spreading the illness. However, some infected people lived far from the "hot spots" wells. During his interview process he found that those people preferred the taste of the contaminated wells and would travel great distances to draw that water. Pretty amazing when you consider that human waste in the water was the main factor for the presence of the bacteria. Yuck!



Please visit our site for more helpful information:
STEMsims.com

Answers: Page 2 Answers: 1) a. 2) a. 3) c. 4) d. It shows where the most Twitter tweets are being sent from. 5) b. The site/map is called the Micro-Fashion Network. Page 3 Lie Factor Answers: 1) 1958 = 14 cm, 1978 = 6 cm (the percent change in length is about 55%, which is close to the change in currency worth during those years). 2) 1958 = 84 cm², 1978 = 15 cm² (the percent change in area is about 82%, not 55% as it should be). 3) Since most people interpret area as the data representation instead of length, the Lie Factor = 2.5.

The Science Fair Kits project was funded in part under the Department of Homeland Security Science and Technology Directorate grant contract #N10PC20003. Its contents are solely the responsibilities of the authors and do not necessarily represent the official views of the Department of Homeland Security.

© 2015 STEM Sims. All rights reserved. Reproduction in whole or in part is prohibited without the prior written consent of the copyright owner. The information presented in this document does not form part of any quotation or contract, is believed to be accurate and reliable, and may be changed without notice. No liability will be accepted by the publisher for any consequence of its use. Publication thereof does not convey nor imply any license under patent or other industrial or intellectual property rights.