Uses and Misuses of Data

Lesson objectives

- I can recognize that the same data can be presented in different ways
- I can see that the way data are presented can have an impact on how the data are interpreted
- I can recognize when, how, and why data are deliberately distorted in order to influence the perception of the reader

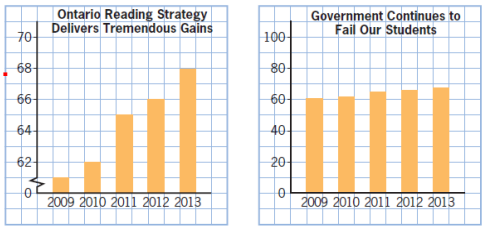
1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 422 #s 1 - 8

Investigate on Page 416

1. The numerical information is the same in both graphs, but they are displayed differently.



Source: Education Quality and Accountability Office

2. There is measurement bias in the way the data has been displayed. The truncation of y-axis has visually exaggerated the "gains" made. The titles of both graphs are also leading statements (not neutral).

3. a) The title is leading the reader to think that there have been big improvements in Grade 3 reading levels due to a particular strategy.

b) The title is leading the reader to think that the government are failing the students.

c) A more appropriate title would be "Ontario's Grade 3 Reading Scores".

How you choose to display data can have a significant impact on how a reader is likely to interpret them. By changing the vertical scale on a graph, a relatively small trend can be made to appear much larger, and vice versa. The intentional use of biased words can influence the interpretation of data. As a critical user of data, it is important to recognize when, how, and why data are being distorted.

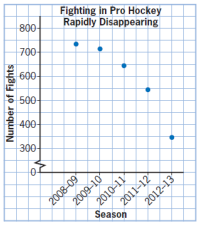
Example 1

Distorting Data to Sway Opinion

Is fighting in hockey on its way out in the NHL? The graph and table illustrate the total number of fights per season in NHL hockey games over time.

Season	Games	Fights
2012–2013	720	347
2011–2012	1230	546
2010–2011	1230	645
2009–2010	1230	714
2008–2009	1230	734

Source: NHL Fight Stats Table, hockeyfights



a) Identify the sources of bias in the graph.

b) Present the data in an unbiased way and reinterpret the data.

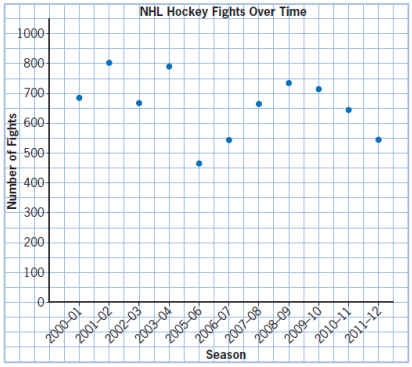
a) Graph title - Not a neutral statement. It's trying to imply that fighting in pro hockey is on the decline.

Vertical scale - It doesn't start at zero, leading to a distortion of the visual rate of decrease.

Outlier - the 2012-13 season only had 720 games compared to the other seasons that had 1230. The proportion of fights per game has actually increased in this season.

Sample size - Only five data points were used. This could be trying to hide a less dramatic rate of decrease.

b) Change the title, change the scale on the y-axis, remove the outlier, use more data points. The graph still shows a slight down trend although not as extreme as before. So despite there being a reduction in the number of fights in the NHL, it doesn't look like fighting is going away anytime soon.



Your Turn

Is the number of multiple-fight NHL games on the decline? The table shows the number of games per season with more than one fight over time.

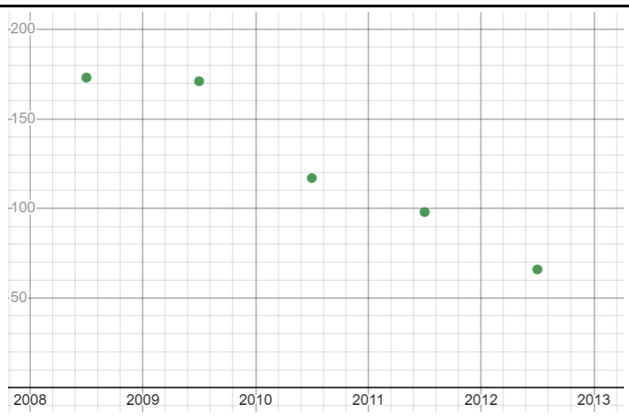
Season	Games With More Than One Fight
2012–13	66
2011–12	98
2010–11	117
2009–10	171
2008–09	173

Source: NHL Fight Stats Table, hockeyfights

a) Create a scatter plot that shows the number of games with more than one fight as a time series.

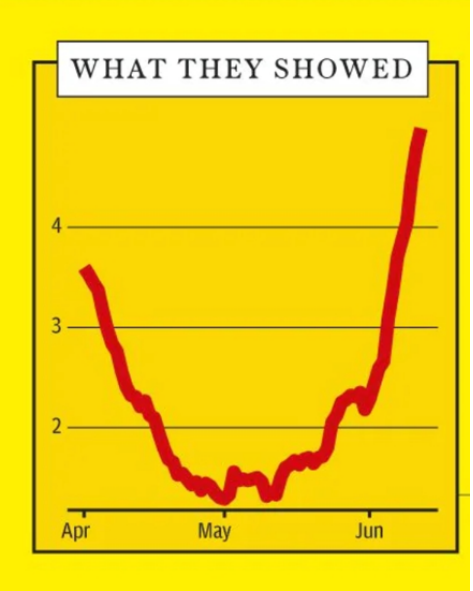
b) Is there an outlier in the data? If so, should it be removed? Explain.

c) Identify any other sources of bias that could be distorting the linear correlation.



b) As with the example there is an outlier for the 2012-13 season... there were only 720 games, instead of 1230. This should be removed as it exaggerates the rate of decrease.

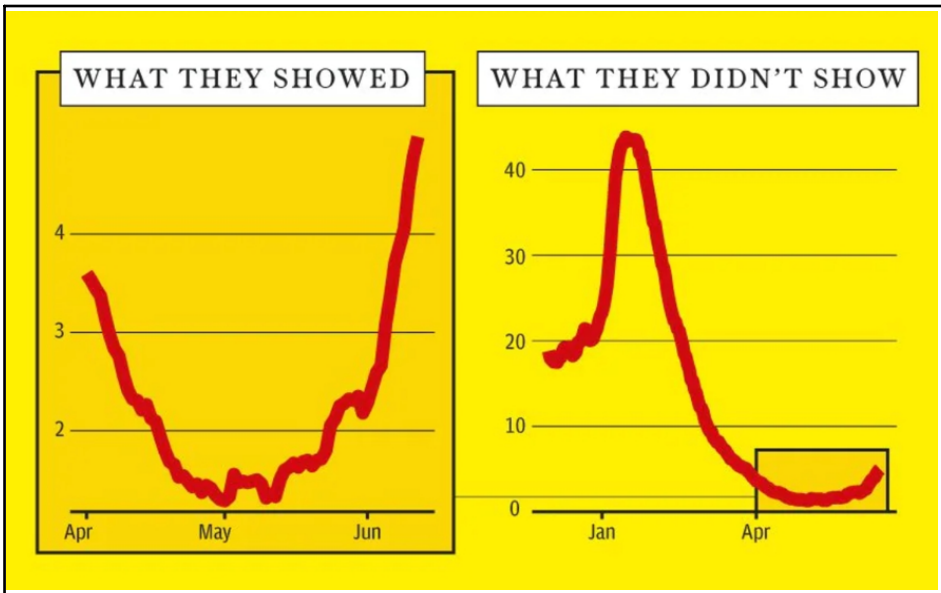
c) Again, there are only five data points which is a form of sampling bias. It could be disguising a less severe downward trend.



Graph representing Covid-19 cases in UK during 2021.

What are your interpretations of this graph?

What else might you like to know?



And then a more wholistic view of the situation!

Example 2

Sensational Use of Data

The following article was taken from an Internet news blog.

a) Discuss how data are used in this article in a questionable way.

b) What inappropriate conclusions are drawn?

c) What is the probable intent of the article?

The aliens are getting braver: UFO sightings in Canada doubled last year

By Lindsay Jolivet | Daily Buzz

The alien invasion has arrived at Earth's doorstep.

People look at the night sky using night vision goggles during a UFO tour (Reuters)

Canadian reports of UFO sightings more than doubled between 2011 and 2012, according to a report released this week by Ufology Research.

Sure, the skies have become more and more crowded with space junk and debris. And yes, it's possible that we're all getting just a little bit crazier.

However it's safest for everyone if we interpret the data as irrefutable evidence of an imminent attack from extra terrestrials that are swarming our skies in greater numbers than ever.

Canadians reported 1,981 UFO sightings in 2012, according to the report, which analyzed data of reported sightings from researchers and websites that monitor UFOs.

About 40 per cent of last year's reported sightings took place in Ontario.

Library and Archives Canada keeps a database of historical UFO sightings, including a case in 1969 when residents of Prince George, B.C. saw a round, glowing object ascending into the sky. An official RCMP investigation, detailed in a report, found the object was a plastic laundry bag and candles converted into a makeshift hot air balloon.

Well, you can't be too careful with these things.

The report says only a tiny fraction of cases involved close encounters with UFOs and in fact, most were merely sightings of lights in the sky. The report speculates on several explanations, including an increase of secret military exercises or a lack of knowledge about objects frequenting our skies that are not exactly "unidentified."

Chris Rutkowski, Ufology's research director, told the *Winnipeg Sun* last year there was no definitive evidence that any of the people reporting sightings have been in contact with aliens.

But what if he's one of *them*?

Source: Yahoo! News Canada, October 12, 2013

a) We are not told how the data was collected or what a constitutes a "UFO sighting". Observations of sightings are more likely to be made by UFO believers who have misinterpreted normal events. There may be a voluntary response bias.

b) The title of the article is implying that as there have been more sightings, aliens are closer than ever to us. The fact that the author states that an attack is imminent, is also a false claim.

c) The likely intent was satire. The sensational tone of the article is evident right from the very beginning with the title and opening sentence.

Your Turn

Consider the following statement taken from the article:

> About 40 per cent of last year's reported sightings took place in Ontario.

Does this suggest that aliens are more interested in Ontario than they are in other provinces? Explain your thinking.

Not necessarily. It is more likely that 40% of the sightings took place in Ontario because about 40% of the Canadian population live in Ontario.

Key Concepts

- You can display data in multiple ways. Sometimes data are deliberately distorted to make an argument more convincing.
- The media often sensationalize data to generate public interest.

R1. **a)** In what ways can data be deliberately distorted?

b) List some motives someone could have to deliberately distort data.

a) Data can be deliberately distorted graphically by choice of scales, titles, sample size, and outliers or sensationalized by using data in a questionable way or by making inappropriate conclusions.

b) Some possible motives are: to grab the attention of the viewer, for humour, to sway opinion, or exaggerate a point.

R2. A juice company shares data in a television commercial that suggest its brand of juice is twice as popular as a competing brand.

a) Why would this company want to share this information?

b) What information would you like to know before making an informed interpretation of the study's findings?

a) They want to share this information in the hope that it will result in even more sales.

b) I would like to know how the data were collected, the sample size used, and the meaning of "twice as popular".

R3. **a)** What is meant by the sensational use of data?

b) What is the purpose of using data in a sensational way?

a) Sensationalism is a type of bias where a piece of work uses tactics such as over-exaggeration to provoke an emotional response.

b) The purpose could be to gain attention by provoking controversy.