Bias

Lesson objectives

- I can distinguish among types of bias when sampling data
- I can analyse and interpret statistics presented by the media to judge their validity
- I can identify different ways that graphical data can be misleading

1.1   Lesson objectives   Teachers' notes   Lesson notes

MHR Page 240 #s 1, 2, 4, 6, 7, 12 & 13abcd

Definitions

Response Bias
- When respondents change their answers to influence the results, to avoid embarrassment, or to give the answer they think the questioner wants

Sampling Bias
- When the sample does not closely represent the population

Measurement Bias
- When the collection method is such that the characteristics are consistently over- or under-represented

Non-Response Bias
- When the opinions of respondents differ in meaningful ways from those of non-respondents

Type of Bias	Example
response bias	A teacher asks students to raise their hand if they cheated on last week's test. Students will not want to admit to cheating on a test so it is unlikely that many will raise their hand.
sampling bias	A politician goes back to the farming community she grew up in to ask for opinions on her latest initiative for the agriculture industry. It is likely that a larger proportion of the people she speaks to would support the initiative, both because it would benefit them and because she grew up in the area. This would not accurately represent the entire population.
measurement bias	A survey question asks, "A lot of people do not like math. How would you feel being referred to as a math geek?" This is a leading question; the wording of the question can affect the outcome by influencing someone's answer. Other types of measurement bias can occur when the collection method affects the results, for example when the options in a multiple choice question are too limited for an honest response.
non-response bias	A mail-in survey asked respondents about their drinking habits. Only 3% of the surveys were returned. Such a small return rate would likely not yield a representative sample. In fact, those who respond often have very strong opinions about the subject matter and so the results could easily over- or under-estimate the feelings of the population.

Example 1 ⊶

Identifying Bias

A large sample does not guarantee good data. This is especially true if the method of data collection has some bias associated with it. Identify the type of bias that may occur in each situation.

a) Families in a neighbourhood are told they are part of a study about healthy eating habits.

b) You ask only students on sports teams how to spend the school fundraising money.

c) A survey question asks, "Who is the best basketball player of all time, Michael Jordan or LeBron James?"

d) A radio call-in show asks callers to answer the question, "Are you in favour of a law that would ban pitbulls from the city?"

a) Response bias - families are likely to start to eat better if they are being monitored.

b) Sampling bias - students on the sports teams are more likely to suggest spending the money on sports related things.

c) Measurement bias - there are only two options.

d) Non-repsonse bias - only people who have a strong opinion tend to call in. Most people don't bother. So in that respect, it is also an example of sampling bias as most of the population are unrepresented.

Your Turn

Identify the type of bias that may occur in the following situations.

a) A survey question asks, "How many words per minute can you read?"

b) A survey is sent to parents of school-age children that asks whether bus safety lanes should be installed.

c) A phone company surveys its customers via text message about which services people like the best.

d) A survey asks, "Now that the city is in debt, do you think the current mayor will win the next election?"

a) Response bias - people are likely to give a high number to give a good impression.

b) Sampling bias - the survey is only sent to parents of school age children and they are likely to be in favour of this idea.

c) Non-response bias - generally only people with extreme views will respond to this.

d) Measurement bias - this is a leading question.
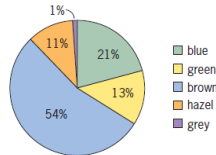

Example 2

Misleading Statistics

Explain how each case uses misleading or questionable statistics.

a) The *Daily Mirror* and *The Times* retweeted this tweet about a British high-speed rail project (HS2) that is expected to take 20 years to complete.

> Interesting stat: 93% of today's UK population will be dead by the time HS2 is finished. Wow. #bbcqt

b) In 2005, the British Cheese Board conducted a study to show that eating cheese does not give you nightmares.

c) Classmates' Eye Colour

Pie chart showing Classmates' Eye Colour: 21% green, 13% (yellow), 54% blue, 11% hazel, 1% grey. Legend: blue, green, brown, hazel, grey.

d) A newscaster reports about a survey that asked, "Did scientists falsify research to support their own theories on climate change?" and shows the following summary:

59% Somewhat likely
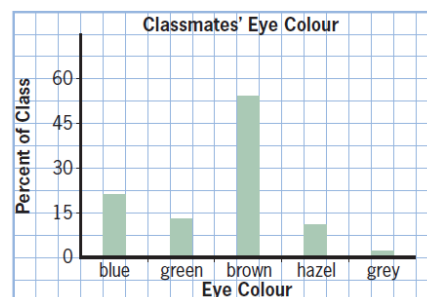35% Very likely
26% Not very likely

e) Research shows that if you eat a sausage a day, your risk of a certain type of cancer goes up by 20%.

a) The tweet is highly unlikely to be true. This should have been fact checked before being published.

b) The validity of this survey is questionable: 1) The Cheese Board are funding it and 2) How do you quantify a nightmare?

c) A bar graph would have been better to allow the reader to compare the different colours:

Bar graph titled "Classmates' Eye Colour" with Percent of Class on the y-axis (0 to 60) and Eye Colour on the x-axis (blue, green, brown, hazel, grey).

d) The percentages add up to 120%, so something is wrong.

e) We need to know more information about the actual prevalence of this type of cancer. If about 5 in 100 have it, then a 20% increase, which sounds a lot, means we now have 6 in 100, which doesn't sound as bad.
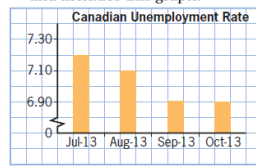
Your Turn

Explain how each case uses misleading or questionable statistics.

a) A news report on jobs lost shows this graph.

Jobs Lost

15 mil
13.5 mil
9 mil
7 mil

Dec '07 Sept '08 Mar '09 Jun '10
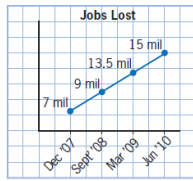
b) A newspaper headline states "Unemployment Dropping" and includes this graph.

Canadian Unemployment Rate

7.30
7.10
6.90
0

Jul-13 Aug-13 Sep-13 Oct-13

c) A yogurt manufacturer states that 67% of customers surveyed claimed they lost weight while eating its brand of yogurt.

a) The graph gives the impression of a linear relation, but the "rises" are different and the "runs" are different. There should be consistent gaps to represent time as well as a scale on the y-axis.

b) The unemployment is dropping, but because there are numbers missing on the y-axis it makes it appear that it is dropping quicker than it actually is. If the vertical scale started at zero, you would see this more accurately. Also, there is no drop from Sept to Oct.

c) This may be true, but it can not be verified that it was due to the customers eating the yogurt. Also there is no data on how much was lost, if those surveyed were on a strict diet or if they had increased their amount of exercise being taken.

Example 3

Some Numbers Are Bigger Than Others

Sometimes numbers that seem big are really small, and other times numbers that seem big truly are. When you see big numbers, try to put them into perspective.

a) A recent settlement of $52.5 million was awarded to 970 000 households from Québec affected by the 1998 ice storm. Is this a big settlement?

b) Switzerland has the highest consumption of chocolate per person. The population is 8.0 million people and they eat about 82 million kg of chocolate each year. Is that a lot of chocolate?

a) Whilst $52.5 million may **seem** like a large amount for a settlement, it only works out at $54.12 per household. (52,500,000 ÷ 970,000)

b) 82,000,000 ÷ 8,000,000 = 10.25 kg per person, per year.

A bar is about 50g, so that is 10250 ÷ 50 = 205 bars per year.

This works out to be 205 ÷ 52 = 3.94 bars/week.

For someone who rarely eats chocolate, this would seem like a lot,

but for someone who eats it everyday, it probably wouldn't.

Your Turn

a) Every year, Canadians and Americans spend over $8 billion on Halloween. Do you think this is a lot of money? Justify your response.

b) In 2011, Apple Corporation's profit grew by almost $19 000 000 000 from 2010. Do you think this is a large increase? Justify your response.

a) In 2011 there are about 138 million households in the USA and 13 million in Canada.

Spend per household = $8,000,000,000 ÷ 151,000,000 = $52.98

To those in Canada and the USA this won't sound like very much but those who don't celebrate Halloween, it might sound like a waste of money.

https://www.msn.com/en-us/money/news/how-much-did-americans-spend-on-halloween-in-2019/ar-AAJG3vL

b) Using data found from an internet search, Apple's gross profit in 2010 was $25.684 billion. In 2011 it was $43.818 billion.

In terms of a percentage increase = (43.818 - 25.684) ÷ 25.684

= 70.6%

In this case, a $19 billion increase in profit is significant.

https://www.macrotrends.net/stocks/charts/AAPL/apple/gross-profit

---

Example 4

Interpreting Infographics

Infographics combine both visual and text-based representations of data. In some cases the representations are standard, while others are more unique or stylish.

a) Describe how this infographic was constructed.

b) Why do you think this representation is called a heat map?

c) The scale represents the ranking. How does this show that ranking is a problem (especially if that is all you know about the data)?

d) What kind of story is told by the data?

Which Birth Dates Are Most Common?

U.S. BIRTHDAY RANK

Source: NYTimes, Amitabh Chandra, Harvard University. Photo from the Daily Viz.

a) Each rectangle represents a day of the year. The day with the most birthdays is the darkest and is ranked first. The day with the fewest is the lightest and ranked last (366th).

b) It is called a heat map because the colour coding is often used to represent temperature. In this case the different shading is used to represent the rank of data. The darker it is, the more often it occurs.

c) Ranking is a problem because you only know the order of the data, not its size or how spread out it is.

d) Despite not knowing the actual data, we can see that the highest number of birthdays occur in September.

Your Turn

Consider the infographic from CBC's *Hockey Night in Canada*.

a) How are Canadian teams distinguished from other teams?

b) How could you determine which team had the fewest North American players?

c) How could you use the data to argue that hockey is Canada's pastime, but not the pastime of the United States?

d) What information is difficult to determine from this representation?

e) Is this infographic free of bias? Explain.

f) Write a tweet (140 characters or less) that is related to this infographic.

WHICH IS THE MOST CANADIAN NHL TEAM?
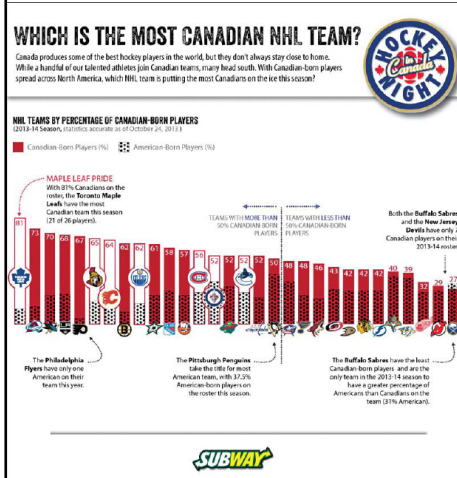
Source: "Infographic: The Most Canadian NHL Team," CBC, October 25, 2013

a) The Canadian teams have their logo displayed on a white bar.

b) The team with the fewest North Americans would be the team with the smallest percentage of Canadian and American players.

c) You could argue that hockey is Canada's pastime and not that of the USA based on the number of Canadian-born players in the NHL and on each team.

d) It is difficult to determine the actual percentage of American-born players on each team.

e) No, it is not free of bias. Based upon the presentation of the information it is slanted towards Canadian-born players.

Key Concepts

- For data to be valid, collection methods must be free from bias.
- The data can be affected if the collection methods suffer from sampling, measurement, response, or non-response bias.
- Different ways of displaying data can distort it and make it biased.
- Large numbers should always be put in context.
- Infographics can be dense with information or convey an idea with unique methods.

R1. Companies often want to back up their product with scientific claims. Is it possible to make misleading claims even if the data are free from bias?

Yes. Misleading claims can be made from data that is free from bias by displaying the data in a distorted manner.

R2. How is it possible for someone to "lie" with statistics yet still be telling the truth?

You can "lie" with statistics by omitting some of the information. Studies can also be based upon faulty studies or experiments which will give misleading results.

R3. In recent years, many soft drink companies have developed vitamin-enriched drinks, claiming they have health benefits. When the owners of vitaminwater® were sued for making false health claims, part of their defence was that

"no consumer could reasonably be misled into thinking vitaminwater was a healthy beverage or was composed only of vitamins and water because the sweet taste of vitamin water puts consumers on notice that the product contains sugar."

Source: Gleeson, John, "Memorandum and Order," CV-09-0395 (JG) (RML).

What is the company claiming and why should it make you wary as a consumer?

The company is trying to blame the consumer. Consumers should not be misled by the name of the product. They should check to see what is actually **IN** the product.