

# Interpreting Statistical Summaries

## Lesson objectives

- I can interpret statistical summaries to describe a one-variable data set and to compare two related one-variable data sets
- I understand whether the data presented are valid and reliable
- I can describe how statistical summaries can misrepresent one-variable data
- I can make inferences as well as make and justify conclusions from statistical summaries of one-variable data
- I can interpret statistics in the media, assess the validity of conclusions made, and explain how statistics are used to promote a certain point of view

1.1

Lesson objectives

Teachers' notes

Lesson notes

MHR Page 298 #s 1 - 4, 6, 8, 13 & 16



Thoughts?

Agree or disagree?

What else would you like to know?

## Definitions

### Multiple Bar Graph

- Different quantities are represented by different colours and lengths of bars that are placed **side by side**

### Split Bar Graph

- Different quantities are represented by different colours and lengths of bars that are placed **one above the other**

### Relative Split Bar Graph

- Different percents, totalling 100, are represented by different colours and lengths of bars that are placed **one above the other**

### Reliable Data

- Results of a study that can be **duplicated** in another study

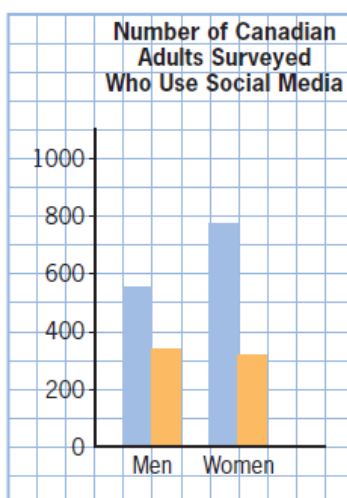
### Valid Data

- Results that accurately represent the **entire population**

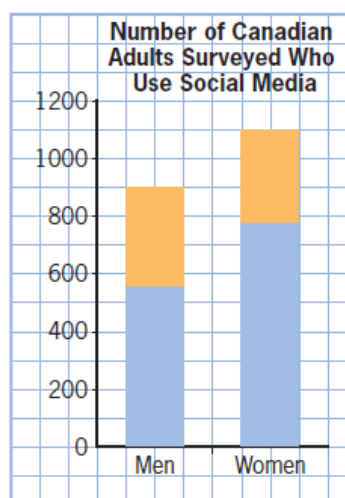
The headline “Significantly more women than men spend time using social media” came from an article about a survey of 900 men and 1100 women on their use of social media. The survey asked the question, “Do you use social media every day?”

The information was compared using a **multiple bar graph**, a **split bar graph**, and a **relative split bar graph**.

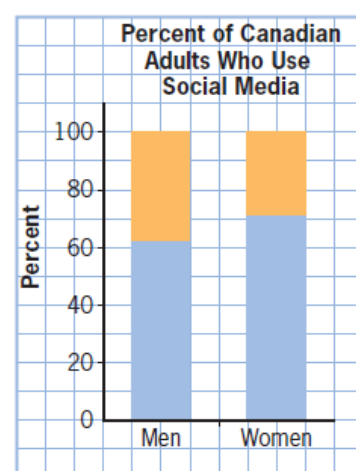
**multiple bar graph**



**split bar graph**



**relative split bar graph**



Investigate on Page 290

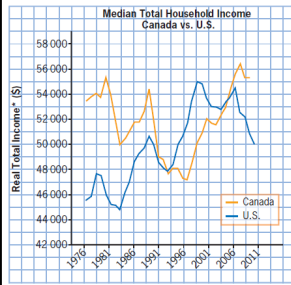


## Example 1

## Interpreting Measures of Central Tendency

A recent headline read, "Americans are significantly wealthier than Canadians." The article indicated that, in 2011, the income per capita in Canada was about \$38 000, and in the United States was about \$42 000.

Another publication showed the following graph. Both cited Statistics Canada and the US Census Bureau as their sources.



U.S. in 2011, Canada in 2010  
Source: U.S. Census Bureau, Statistics Canada

- Interpret these statements and explain what might cause the discrepancy.
- Compare the two countries' median household incomes over the years.
- Discuss the vertical scale of the graph and how it may influence the reader.

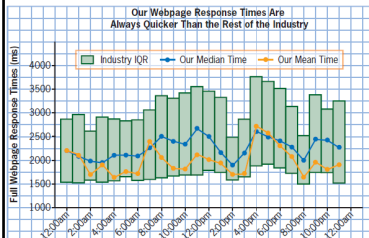
a) The graph indicates that the median income in Canada was about \$55,000, and in the USA it was about \$50,000. Income per capita represents the mean income per person. The mean is always influenced by outliers, so extremely wealthy people will increase the mean. This leads to the greater income per capita in the USA compared to Canada. Both articles are reliable, it is just the interpretation of the first article is inaccurate.

b) The Canadian median has stayed above the median for the USA for the period under review except for between 1992 - 2005. It doesn't state if these figures have been adjusted for exchange rates or not.

c) The vertical scale does not start at zero, so this will exaggerate the differences in the data.

## Your Turn

The graph below was included in a report by an Internet service provider.



- What statistical measures does the graph show?
- Which measures of central tendency are used in the Internet service provider's analysis?
- Discuss whether the Internet service provider's claim is accurate.
- Discuss the vertical scale of the graph and how it may influence the reader.

a) The graph shows the mean and median times for loading webpages for an ISP as well as the IQR for the industry as a whole.

b) When it comes to speed, they would prefer faster times. The ISP is using mean times as they are quicker than the median times. If the mean is larger than the median, then they will use the median in those instances.

c) The ISP's claims are inaccurate. Since the mean of ANY set of data is influenced by outliers, very fast times will reduce the size of the mean quite significantly. Also, the industry IQR does not state the median, so it is not possible to compare them accurately.

d) The vertical axis does not start at zero, so this will exaggerate the differences in the data.

**Creating a multiple bar graph by hand:**

Each bar represents the frequency.

**Creating a split bar graph by hand:**

Each bar represents the cumulative frequency.

**Creating a relative split bar graph by hand:**

Work out the percentage of each bar by comparing frequency to the overall frequency. Each bar should total 100%.

**Example 2**

**Comparing Groups Graphically**

The Athletic Council of Reliable College wants to know whether males or females play more tennis at their school. They decide to poll the grade 11 and 12 students to organize, display, and draw conclusions about their data.

The following table lists the data they gathered.

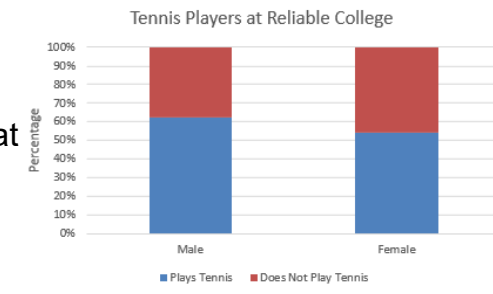
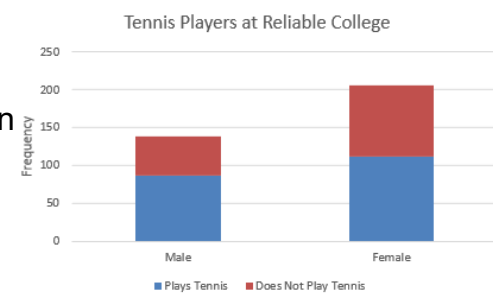
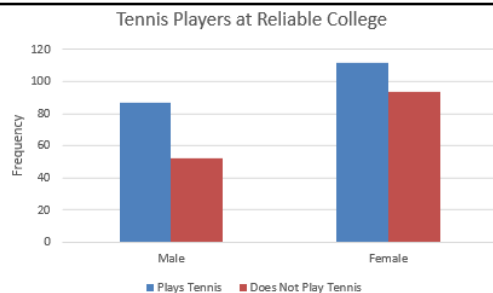
	Frequency	
	Male	Female
Plays tennis	87	112
Does not play tennis	52	94

- a) Create a multiple bar graph, a split bar graph, and a relative split bar graph to display the data.
- b) Which graph more clearly shows which gender plays tennis more often?
- c) Can you draw any other conclusions by looking at the visual representations of the data?

b) The multiple bar graph and split bar graph both show that more females than males play tennis. They also show that there were more females polled than males, so it is advisable to use the relative split bar graph. Knowing more about how the samples were selected would help.

c) The relative split bar graph shows that more than 50% of males and females plays tennis (blue sections).

a)



**Your Turn**

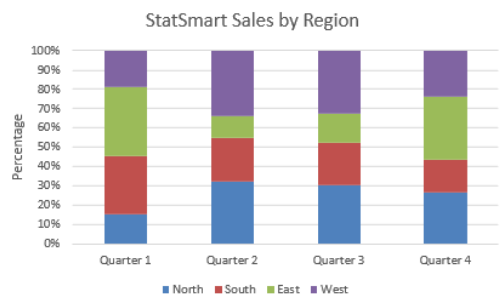
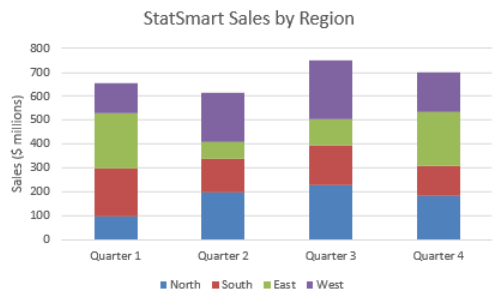
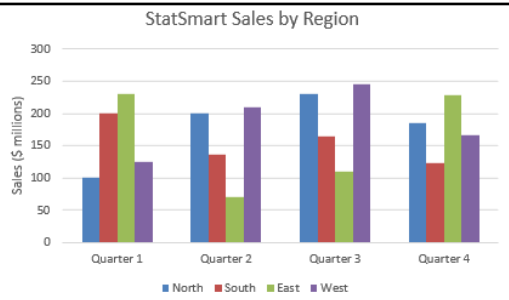
You work at StatSmart and your manager wants to determine how the store is performing on sales based on each quarter year, by region. The following sales data were gathered for the store during the year. Organize, display, and draw conclusions about the data to help show the manager how the store is performing.

Region	StatSmart Sales (\$ millions)			
	Quarter 1	Quarter 2	Quarter 3	Quarter 4
North	100	200	230	185
South	200	137	164	123
East	231	70	110	228
West	125	210	246	166

- a) Organize and display the data using a multiple bar graph, a split bar graph, and a relative split bar graph.
- b) What can you conclude from the data? Explain.

b) The multiple bar graph and split bar graph both show how each region's sales compared to one another by each quarter. As the sales are different, the relative split bar graph is useful for comparing the percentage of sales contributed by each region for each quarter.

For quarters 1 and 4, East did the best. For quarters 2 and 3, West did the best with North a close second.



**Example 3**

**Critical Analysis of Claims in the Media**

**Can You Recall The Last Time Your Phone Was *Not* Within Ear Shot?**

63% of smartphone owners keep their phone with them for all but an hour of their waking day. 79% keep it with them for all but two hours of their day.

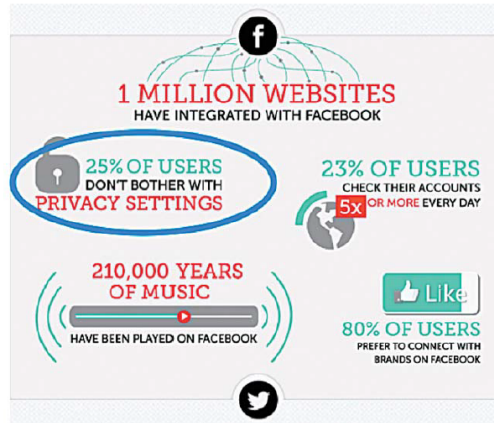
1 out of 4 of all respondents didn't recall a time in their day when their phone was not within reach or in the same room.

Whether it is a weekday or weekend, the amount of time away from one's phone didn't vary—it is a critical tool for connecting with friends, family and colleagues every day.

Source: WorkSmart: 10 Surprising Social Media Statistics That Will Make You Rethink Your Social Strategy; 25% of Smartphone Owners Ages 18–44 Say They Can't Recall the Last Time Their Smartphone Wasn't Next to Them.

- a) The message conveyed by the statistics is that a smartphone is a critical tool for communicating and being in close contact with people.
- b) Use of colour for highlighting. Using circles instead of bar graphs, so it is difficult to compare area sizes. Different sized time intervals. Phrasing infers the whole population was asked and not just 18-44 year olds.
- c) How large was the sample? How was it chosen and was it selected randomly? What questions did the survey ask?
- d) What was the source of the data? Was primary or secondary data used? Who sponsored (paid for) the survey?

## Your Turn



Source: WorkSmart: 10 Surprising Social Media Statistics That Will Make You Rethink Your Social Strategy; 25% of Facebook Users Don't Bother With Privacy Settings.

- What messages do these statistics convey?
- What techniques are used to influence the reader with statistics?
- What questions need to be answered to critically analyse the results of the survey?
- What questions need to be asked to check the reliability of the source?

a) As much as connectivity with brands and listening to music are important to users, their privacy settings are not.

b) Readers may be influenced by the use of graphics (lock, music player etc) as well as the font sizes and colours. One statistic was circled to highlight it.

c) How large was the sample? How was it chosen and was it selected randomly? What questions did the survey ask?

d) What was the source of the data? Was primary or secondary data used? Who sponsored (paid for) the survey?

## Key Concepts

- When you compare data values, it is possible to draw conclusions based on the data set results.
- There may or may not be a relationship between compared values.
- In some instances, graphs provide a stronger visual of the conclusion.
- You can use multiple bar graphs, split bar graphs, and relative split bar graphs to compare two similar data sets.
- Statistics are often used to represent certain points of view by manipulating graph axes, by citing only one measure of central tendency, or through measurement or sampling bias.
- It is key to perform a critical analysis of any statistical report.

R1. How are multiple bar graphs, split bar graphs, and relative split bar graphs different? How are they similar?

All three types of graph display the same data. The multiple bar graph and split bar graph show different quantities, where as a relative split bar graph shows different percentages. Multiple bar graphs display the data side by side where as the other two has the data stacked into one bar.

R2. a) Why is it important to critically analyse a statistical summary?

b) What are some important questions that would need to be asked?

a) It is important to critically analyse a statistical summary to determine whether you can make **valid** generalizations. Statistics are often used to represent certain points of view by manipulating graph axes, citing only one measure of central tendency, or through measurement or sampling bias.

b) Remember, data that is collected needs to be reliable and valid.

- How large was the sample?
- How was the sample chosen and was it selected randomly?
- What questions were asked in the survey?
- What was the source of the data?
- Were the data primary or secondary?
- Who sponsored (paid for) the survey?