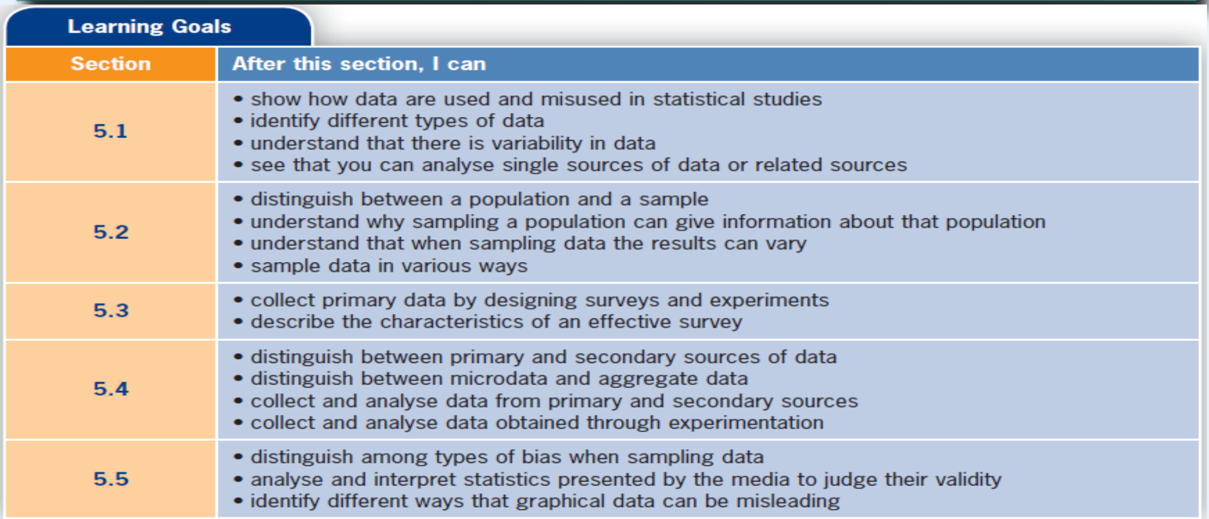
Organization of Data for Analysis Review

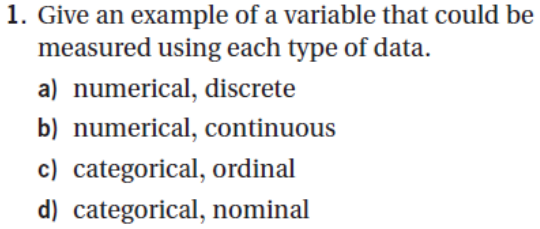


Learning Goals

Section	After this section, I can
5.1	• show how data are used and misused in statistical studies • identify different types of data • understand that there is variability in data • see that you can analyse single sources of data or related sources
5.2	• distinguish between a population and a sample • understand why sampling a population can give information about that population • understand that when sampling data the results can vary • sample data in various ways
5.3	• collect primary data by designing surveys and experiments • describe the characteristics of an effective survey
5.4	• distinguish between primary and secondary sources of data • distinguish between microdata and aggregate data • collect and analyse data from primary and secondary sources • collect and analyse data obtained through experimentation
5.5	• distinguish among types of bias when sampling data • analyse and interpret statistics presented by the media to judge their validity • identify different ways that graphical data can be misleading

MHR Page 244 #s 1 - 8

Solutions



1. Give an example of a variable that could be measured using each type of data.

   a) numerical, discrete

   b) numerical, continuous

   c) categorical, ordinal

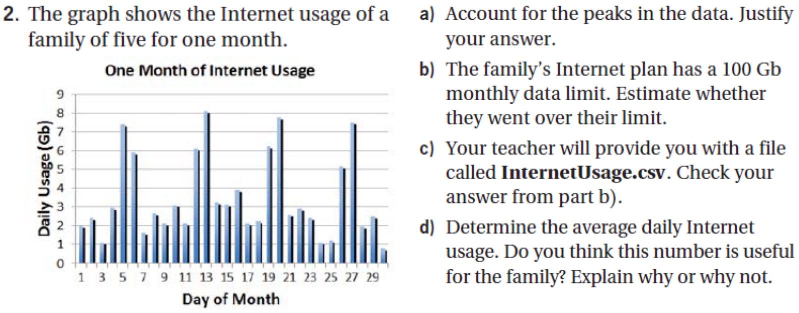
   d) categorical, nominal

a) Has to be "countable" - number of students in a class

b) Has to be "measurable" - how long it takes you to run 100m

c) Have to be able to be "ranked" - answers could be strongly agree, agree, no opinion, disagree, strongly disagree

d) Have a "quality" as an answer - hair colour



2. The graph shows the Internet usage of a family of five for one month.

a) Account for the peaks in the data. Justify your answer.

b) The family's Internet plan has a 100 Gb monthly data limit. Estimate whether they went over their limit.

c) Your teacher will provide you with a file called **InternetUsage.csv**. Check your answer from part b).

d) Determine the average daily Internet usage. Do you think this number is useful for the family? Explain why or why not.

a) Peaks in the data are likely to be on the weekends. There is a repeating pattern of two high columns followed by five lower ones.

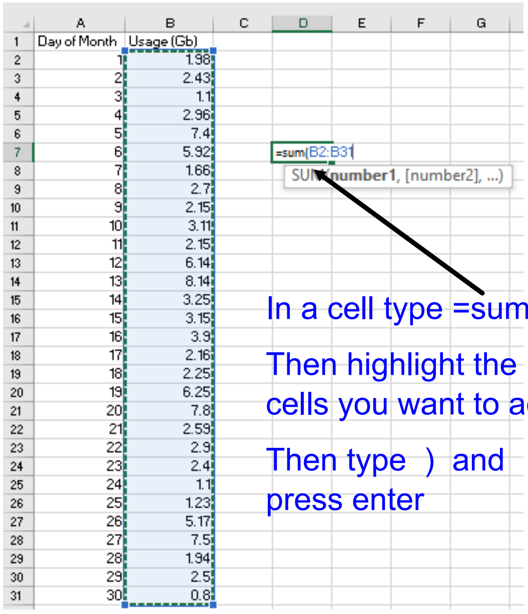
b) Estimate at about 97GB, so no just under their limit.

c) Using the sum function in the spreadsheet, the total is 104.73GB.

d) Using the average function in the spreadsheet, the daily average usage is 3.49GB.

This is not particularly useful as they use less than average on a weekday, but significantly more than average at the weekends.

c) Explained



	A	B
1	Day of Month	Usage (Gb)
2	1	1.98
3	2	2.43
4	3	1.1
5	4	2.96
6	5	7.4
7	6	5.92
8	7	1.66
9	8	2.7
10	9	2.15
11	10	3.11
12	11	2.15
13	12	6.14
14	13	8.14
15	14	3.25
16	15	3.15
17	16	3.9
18	17	2.16
19	18	2.25
20	19	6.25
21	20	7.8
22	21	2.59
23	22	2.9
24	23	2.4
25	24	1.1
26	25	1.23
27	26	5.17
28	27	7.5
29	28	1.94
30	29	2.5
31	30	0.8

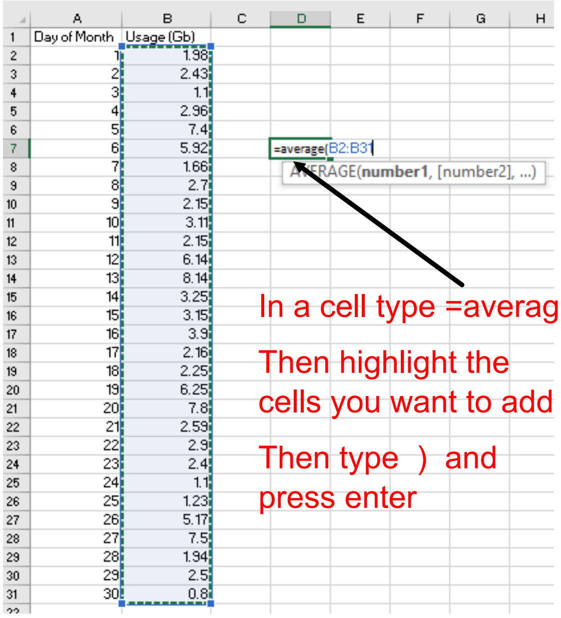
=sum(B2:B31

In a cell type =sum(

Then highlight the cells you want to add

Then type) and press enter

d) Explained



	A	B
1	Day of Month	Usage (Gb)
2	1	1.98
3	2	2.43
4	3	1.1
5	4	2.96
6	5	7.4
7	6	5.92
8	7	1.66
9	8	2.7
10	9	2.15
11	10	3.11
12	11	2.15
13	12	6.14
14	13	8.14
15	14	3.25
16	15	3.15
17	16	3.9
18	17	2.16
19	18	2.25
20	19	6.25
21	20	7.8
22	21	2.59
23	22	2.9
24	23	2.4
25	24	1.1
26	25	1.23
27	26	5.17
28	27	7.5
29	28	1.94
30	29	2.5
31	30	0.8

=average(B2:B31

In a cell type =average(

Then highlight the cells you want to add

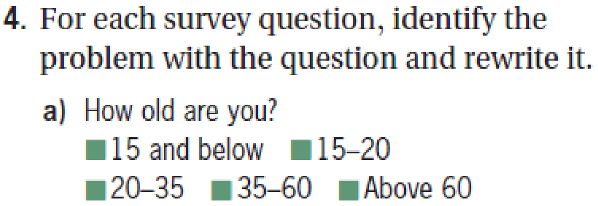
Then type) and press enter

3. Determine the type of sampling.

a) Your favourite social networking site asks what your favourite band is.

b) The Pelee Island Bird Observatory does its annual bird count every December. Researchers set up traps in the trees in three locations to gather the birds.

a) Convenience sampling - each user chooses whether they will respond to the survey or not.
This could also be considered a voluntary sample as only those who repsonded will be part of the sample.
b) Convenience sampling - only three trees were picked to sample, likely because they were easy to access. There is no other information about where the trees were located or how or why they were selected.
This could also have a second answer in that it could be considered to be a cluster sample. The area could be divided into sections and then three of those sections are chosen for sampling.

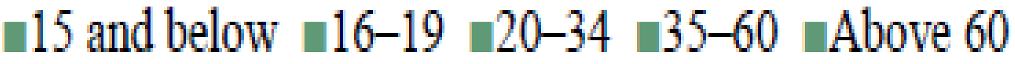


4. For each survey question, identify the problem with the question and rewrite it.

a) How old are you?
□ 15 and below □ 15–20
□ 20–35 □ 35–60 □ Above 60

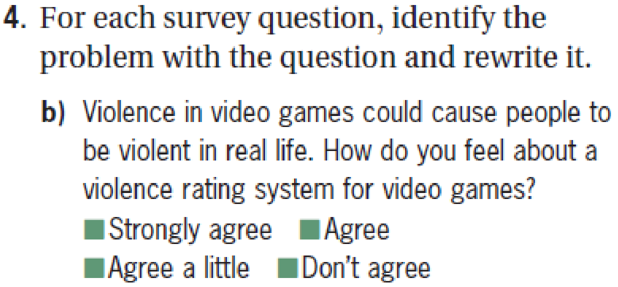
a) There are two options for those aged 15, 20 or 35. Rewrite as:

How old are you?



□ 15 and below □ 16–19 □ 20–34 □ 35–60 □ Above 60

You could also consider making the option boxes have a similar age range.

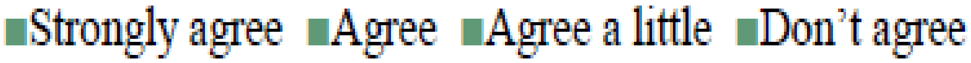


4. For each survey question, identify the problem with the question and rewrite it.

b) Violence in video games could cause people to be violent in real life. How do you feel about a violence rating system for video games?
□ Strongly agree □ Agree
□ Agree a little □ Don't agree

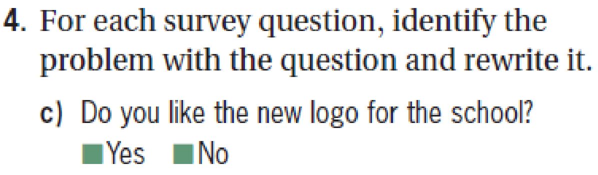
b) The initial statement is leading and therefore biased. Rewrite it so that it is neutral:

There should be a violence rating system for video games.



□ Strongly agree □ Agree □ Agree a little □ Don't agree

You also could have an equal number of positive and negative options.



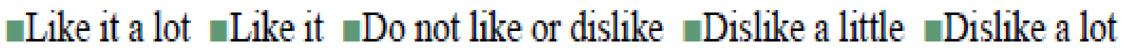
4. For each survey question, identify the problem with the question and rewrite it.

c) Do you like the new logo for the school?

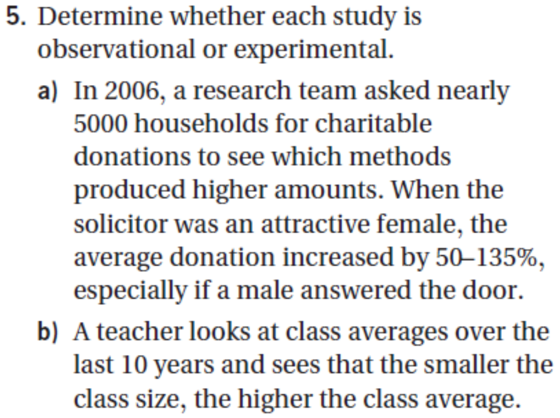
🟩Yes 🟩No

c) There is not enough options. Rewrite it so there are more of them including a neutral option:

Do you like the new school logo?



🟩Like it a lot 🟩Like it 🟩Do not like or dislike 🟩Dislike a little 🟩Dislike a lot



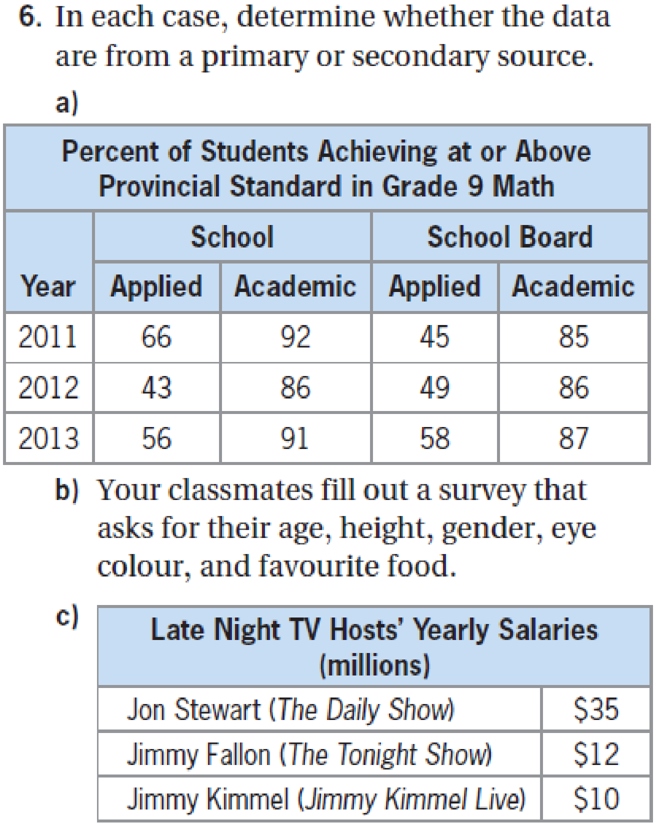
5. Determine whether each study is observational or experimental.

a) In 2006, a research team asked nearly 5000 households for charitable donations to see which methods produced higher amounts. When the solicitor was an attractive female, the average donation increased by 50–135%, especially if a male answered the door.

b) A teacher looks at class averages over the last 10 years and sees that the smaller the class size, the higher the class average.

a) Experimental - The research team is controlling the method of collection of donations and making inferences from the results.

b) Observational - The marks have already been collected and there is nothing that can be done to change them. Inferences are then made based upon the data collected.



6. In each case, determine whether the data are from a primary or secondary source.

a)

Percent of Students Achieving at or Above Provincial Standard in Grade 9 Math				
	School		School Board	
Year	Applied	Academic	Applied	Academic
2011	66	92	45	85
2012	43	86	49	86
2013	56	91	58	87

b) Your classmates fill out a survey that asks for their age, height, gender, eye colour, and favourite food.

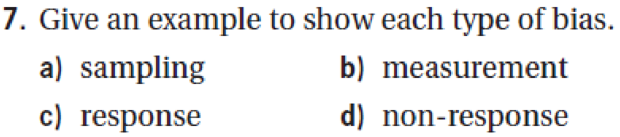
c)

Late Night TV Hosts' Yearly Salaries (millions)	
Jon Stewart (*The Daily Show*)	$35
Jimmy Fallon (*The Tonight Show*)	$12
Jimmy Kimmel (*Jimmy Kimmel Live*)	$10

a) It depends...
Primary - if used by those that collected the data.
Secondary - for anyone who uses the data.

b) Primary - you are collecting this information.

c) Secondary - this is collected publicly available information.



7. Give an example to show each type of bias.

a) sampling b) measurement

c) response d) non-response

a) Choosing a small group that is not representative of the population.

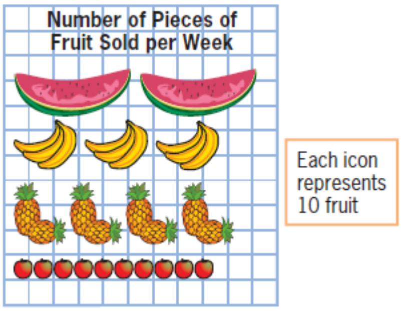
b) If there is a leading question or statement that precedes a question, or there is a limited number of response options available.

c) A scenario that could lead to someone not answering truthfully to avoid embarrassment or to make the questioner happy.

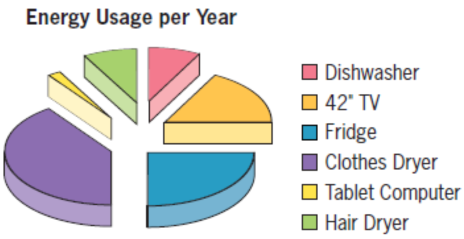
d) Any situation that may lead to a voluntary response.

8. In each case, indicate how the representation is misleading.

a)



b)



a) The icon sizes are inconsistent to their numerical value. The row lengths are similar but they do not total the same.

b) No percentage data is given for the sizes of each segment and we have no idea of the total number of people surveyed to see if it is a representative sample.