

Data Analytics

ECON 1008, Semester 1, 2019

Giulio Zanella
University of Adelaide
School of Economics

Chapter 3

Graphical descriptive techniques -
Nominal data

Remember from last lecture...

2. Nominal Data

The values of **nominal data** are *categories*.

E.g. Responses to questions about marital status are categories, coded as:

Single = 1, Married = 2, Divorced = 3, Widowed = 4

These data are **categorical** in nature; arithmetic operations don't make any sense (e.g. does Married \div 2 = Divorced?!)

All we can calculate is the proportion of data that falls into each category.

Nominal data are also called **qualitative** or **categorical**.

Graphical techniques to describe nominal data

The only allowable calculation on nominal data is to count the frequency of each value of the variable.

We can summarise the data in a table that presents the categories and their counts called a *frequency distribution*.

A *relative frequency distribution* lists the categories and the proportion with which each occurs.

The methods presented apply to both

- the entire population, and
- a sample selected from the population.

Bar charts

The **bar chart** is mainly used for *nominal data*.

A bar chart graphically represents the frequency of each category as a bar rising vertically from the horizontal axis.

The height of each bar is proportional to the frequency of the corresponding category.

Pie charts

Another useful chart to present nominal data is the pie chart.

The pie chart is a very popular tool used to represent the proportions of appearance for nominal data.

A pie chart is a circle that is subdivided into slices whose areas are proportional to the frequencies (or relative frequencies), thereby displaying the proportion of occurrences of each category.

Example 1

(Example 3.1, page 45)

XM03-01 A magazine readership survey carried out in New Zealand (Roy Morgan, March 2012) shows that women's magazines are the most popular magazines, having the largest readership and increasing yearly sales. The survey results of 300 readers were recorded and are given below in coded form. The top six magazines considered here are (1) *Australian Women's Weekly (NZ Edition)*; (2) *NZ Woman's Weekly*; (3) *NZ Woman's Day*; (4) *New Idea*; (5) *Next*; and (6) *That's Life*. The data, using the codes 1, 2, 3, 4, 5 and 6, are listed below. Create the frequency distribution table and construct a bar chart and a pie chart to summarise the data.

Example 1...

Women's Magazine Readership, New Zealand, 2015

1	1	5	3	2	4	3	5	1	2	5	3	5	1	3	1	4	1	3	5	2	1	4	3	2	1	3	1	1	2
5	3	1	4	3	2	4	3	5	6	3	1	1	1	4	5	2	3	4	3	1	1	1	3	3	2	1	3	3	5
3	3	3	2	1	1	2	4	3	1	4	6	2	3	1	3	3	1	3	2	3	1	3	2	3	1	2	3	2	2
4	6	3	6	5	5	1	2	4	5	2	5	3	3	1	1	1	2	2	1	5	1	2	3	2	3	3	5	1	3
1	3	2	1	1	3	1	6	2	3	5	3	4	4	5	3	3	3	2	3	3	3	2	1	2	3	3	4	3	3
6	3	2	5	3	5	3	5	6	3	4	4	2	2	3	3	2	2	2	4	3	5	4	6	1	3	2	2	3	2
3	4	5	3	5	4	1	3	1	4	2	3	6	6	2	3	4	2	4	1	1	2	3	1	3	2	3	3	6	4
4	2	3	4	5	3	5	1	6	2	5	4	1	4	3	4	2	3	2	4	3	4	3	3	1	2	6	4	3	5
2	5	4	2	1	5	2	5	3	1	3	2	5	2	1	6	6	4	1	3	1	1	3	1	1	2	2	2	4	4
6	1	2	6	3	1	6	3	4	5	1	6	5	6	5	4	3	2	5	4	3	2	2	3	1	1	6	3	3	3

Example 1: Solution

Identifying the technique

As the data are nominal, the only technique is to count the number of occurrences (also known as **frequencies**) of each value and then convert these counts to proportions.

The frequencies are presented using a *bar chart*. The proportions (market share of each magazine) are presented in a *pie chart*.

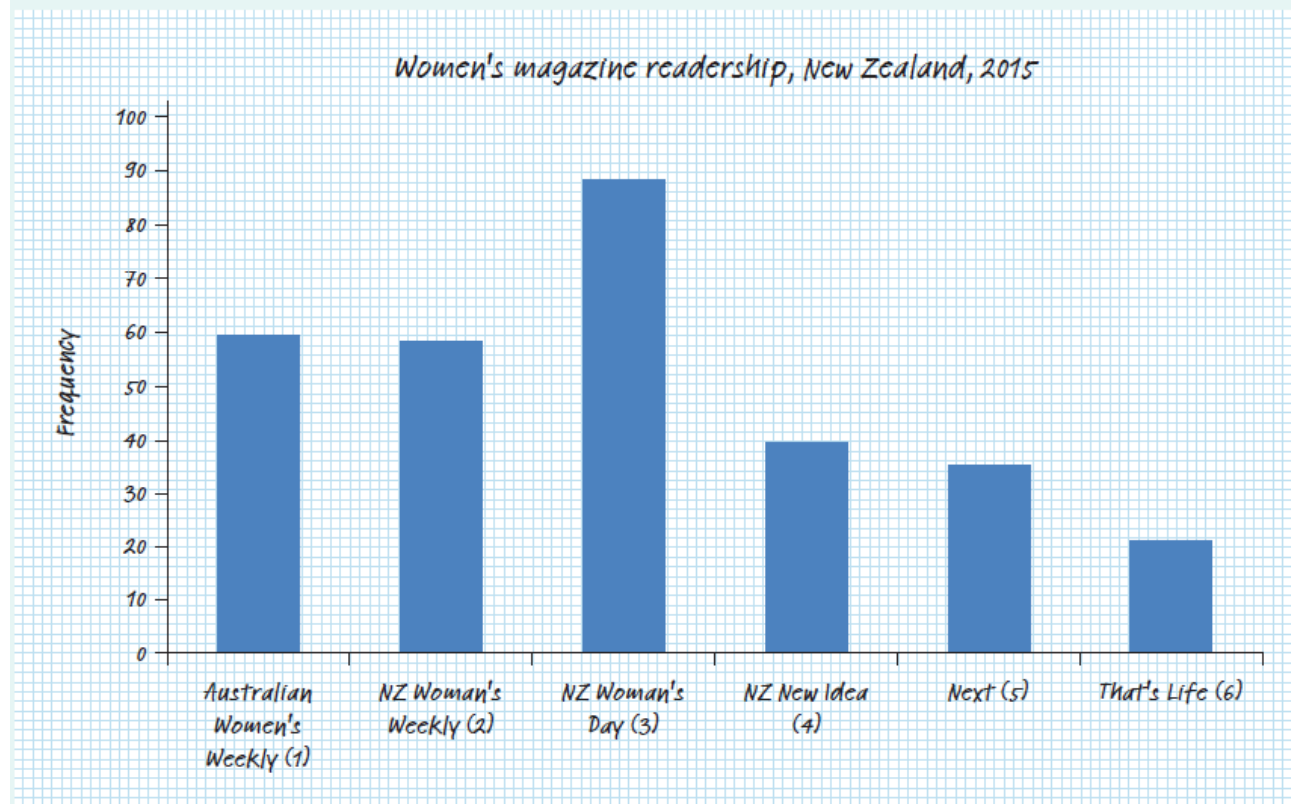
Example 1: Solution...

Magazine	Number of readers	Proportion of readers (%)
<i>Australian Women's Weekly</i> (1)	59	19.7
<i>NZ Woman's Weekly</i> (2)	58	19.3
<i>NZ Woman's Day</i> (3)	88	29.3
<i>NZ New Idea</i> (4)	39	13.0
<i>Next</i> (5)	35	11.7
<i>That's Life</i> (6)	21	7.0
Total	300	100.0

Example 1: Solution...

Bar chart - Using Excel

Figure 3.1 Bar chart for Example 3.1



Example 1: Solution...

Pie chart - Calculating manually

In constructing a pie chart, the size of a slice of the circle is proportional to the percentage corresponding to that category.

For example, the angle between the lines demarcating the *Australian Women's Weekly (NZ Edition)* readers is $19.7 \times 3.6 = 70.8^\circ$.

The angles of the pie chart for the other five categories are calculated similarly, as shown in the table below:

Example 1: Solution...

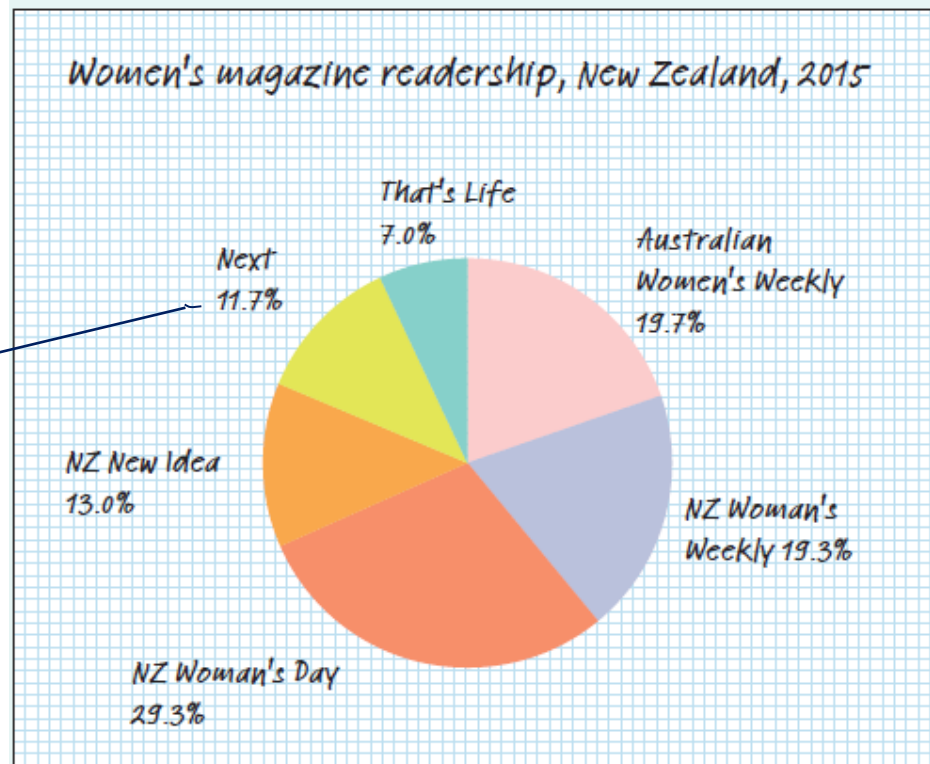
Pie chart - Calculating manually

Magazine	Proportion of readers (in percentages)	Angle of the slice
<i>Australian Women's Weekly</i> (1)	19.7	$19.7 \times 3.6 = 70.8^\circ$
<i>NZ Woman's Weekly</i> (2)	19.3	$19.3 \times 3.6 = 69.6^\circ$
<i>NZ Woman's Day</i> (3)	29.3	$29.3 \times 3.6 = 105.6^\circ$
<i>NZ New Idea</i> (4)	13.0	$13.0 \times 3.6 = 46.8^\circ$
<i>Next</i> (5)	11.7	$11.7 \times 3.6 = 42.0^\circ$
<i>That's Life</i> (6)	7.0	$7.0 \times 3.6 = 25.2^\circ$
Total	100.0	360°

Example 1: Solution...

Pie chart - Manually/Using Excel

Figure 3.2 Pie chart for Example 3.1



Angle = $(11.7/100)(360^\circ) = 42^\circ$

Bar or Pie?

Which chart is best - bar or pie chart?

- Depends on what you want to emphasize
- If the focus is to compare the size or frequency of various categories, a bar chart may be appropriate
- If the focus is on the distribution (or share) of each category, use a pie chart

Example 2

(Table 3.5, page 53)

Table 3.5 Top 10 best-selling new passenger vehicle sales by manufacturers in Australia, 2013 and 2014

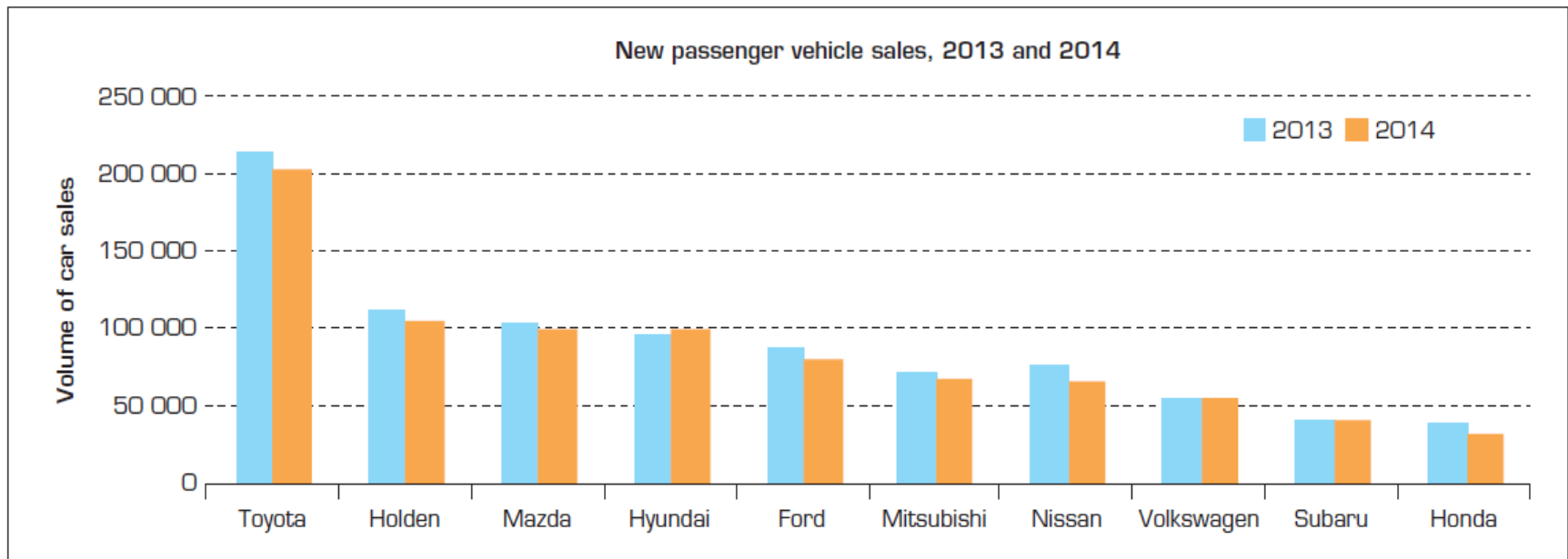
Rank	Model	Volume of sales		Percentage change (%)	Market share	
		2013	2014		2013	2014
1	Toyota	214 630	203 501	−5.2	23.9	23.9
2	Holden	112 059	106 092	−5.3	12.5	12.4
3	Mazda	103 144	100 704	−2.4	11.5	11.8
4	Hyundai	97 006	100 011	3.1	10.8	11.7
5	Ford	87 236	79 703	−8.6	9.7	9.3
6	Mitsubishi	71 528	68 637	−4.0	8.0	8.0
7	Nissan	76 733	66 025	−14.0	8.6	7.7
8	Volkswagen	54 892	54 801	−0.2	6.1	6.4
9	Subaru	40 200	40 502	0.8	4.5	4.7
10	Honda	39 258	32 998	−15.9	4.4	3.9
	Total	896 686	852 974		100.0	100.0

Source: Federal Chamber of Automotive Industries, <http://www.fcai.com.au/news/index/index/year/all/month/all/article/379>, 6 January 2015

Example 2...

If the focus is to compare the size of frequency of various categories, a bar chart is more appropriate.

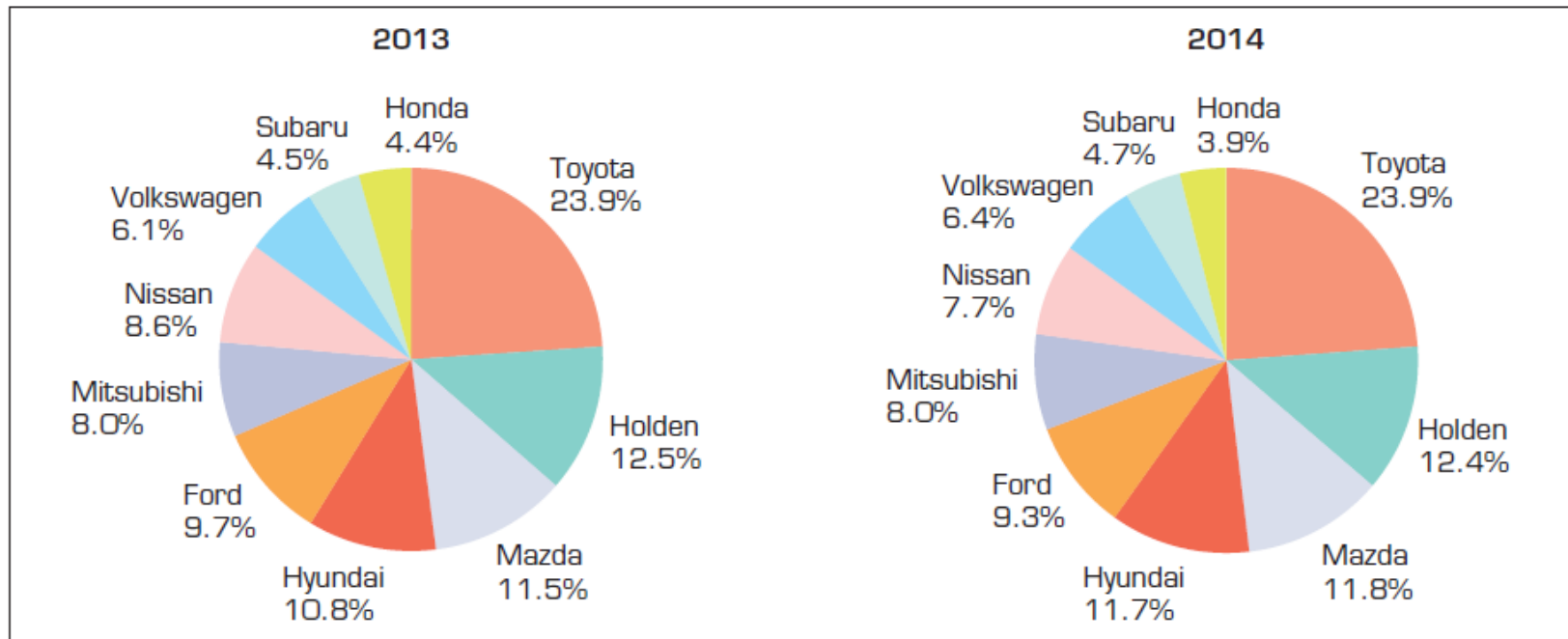
Figure 3.6 Bar chart of new passenger vehicle sales, 10 best-selling manufacturers, 2013 and 2014



Example 2...

If the focus is on the distribution of each category either pie or bar chart may be appropriate.

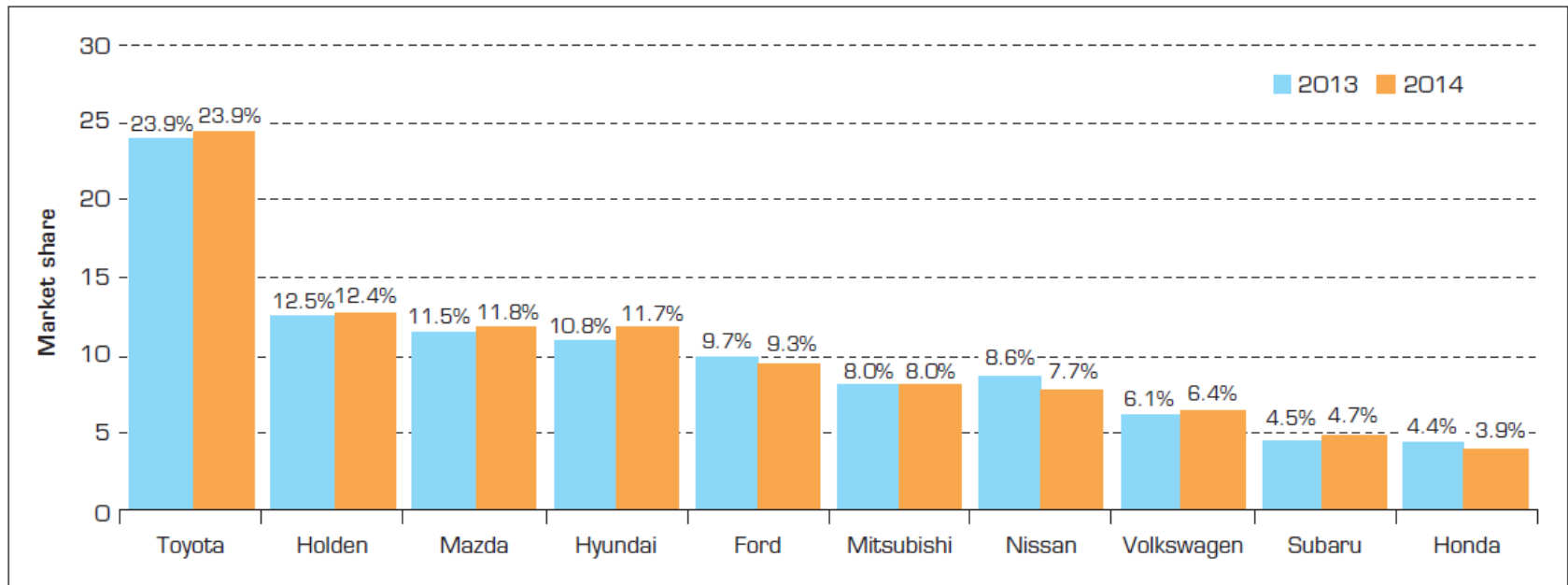
Figure 3.9a Pie charts emphasising change in market share, 2013 vs 2014



Example 2...

If the focus is on the change in market share between the two years, a bar chart may be appropriate.

Figure 3.9b Bar chart emphasising change in market shares, 2013 vs 2014



Component bar chart

A **component bar chart** represents all categories within a single bar.

The bar is partitioned into components, with the height of each component proportional to the frequency of the category that it represents.

Component bar charts offer a good alternative to using two pie charts, when a comparison of two breakdowns is desired.

Example 3

(Table 3.8, page 59)

Table 3.8 Components of likelihood of buying shares in the next 12 months (in percentages), 2008, 2010 and 2012

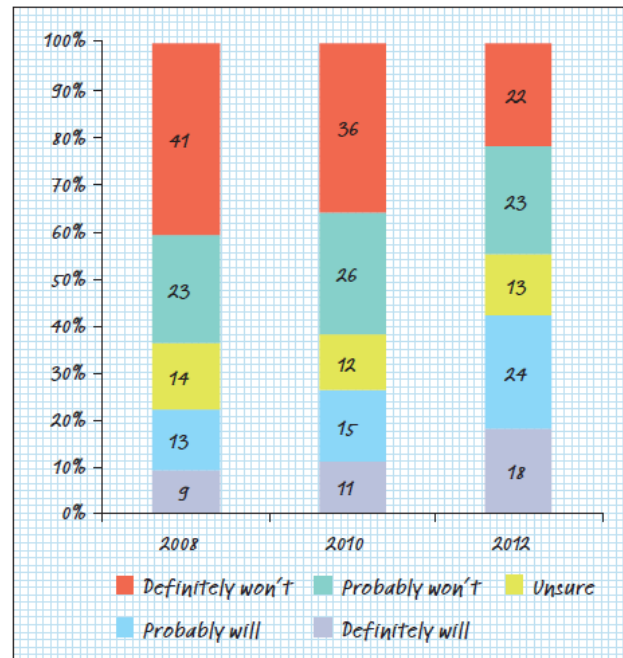
Buying shares	2008	2010	2012
Definitely will	9	11	18
Probably will	13	15	24
Unsure	14	12	13
Probably won't	23	26	23
Definitely won't	41	36	22

Source: *Australian Share Ownership Study*, Australian Stock Exchange Ltd, 2013

Example 3...

The shares of likelihood in 2008, 2010 and 2012 can be compared by displaying them in a component bar chart.

Figure 3.13 Likelihood of investors buying shares in the next 12 months, 2008, 2010 and 2012



3.3 Graphical techniques to describe ordinal data

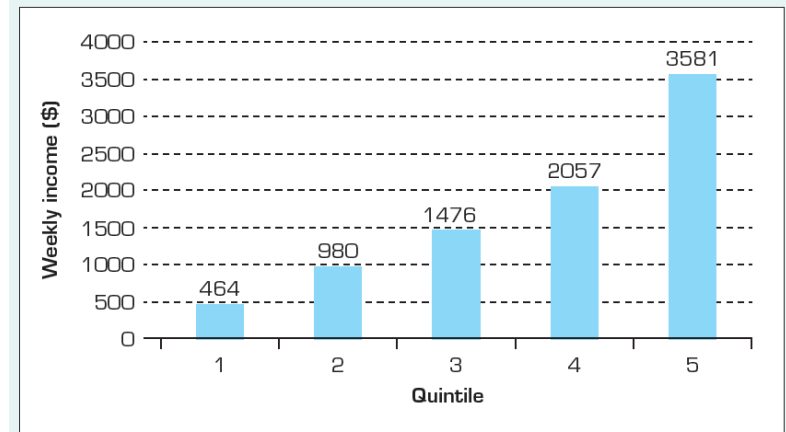
When the data are **ordinal (or ranked)**, treat the data as nominal and use a bar or pie chart.

For example, consider the average weekly household income for the 5 income quintiles (Example 3.6, page 59).

Australian household income by quintile

Quintile	Weekly average income (\$)
1	464
2	980
3	1476
4	2057
5	3581

Figure 3.14 Gross weekly average household income by quintiles, Australia, 2009–2010



Describing the relationship between two nominal variables...

Two ways to describe

- A cross-classification table (or contingency table or cross-tabulation table)
- A variation of a bar chart

Describing the relationship between two nominal variables...

A ***cross-classification table*** is used to describe the relationship between *two* nominal variables.

A cross-classification table lists the ***frequency of each combination*** of the values of the two nominal variables.

Example 4: Newspaper Readership Survey

(Example 3.7, page 67)

XM03-07 In a major Australian city there are four competing newspapers: N1, N2, N3 and N4.

To help design advertising campaigns, the managers of the newspapers need to know which segments of the market are reading their newspapers.

A survey was conducted to determine whether the newspaper readership and occupation are related.

A sample of newspaper readers was asked to report which newspaper they read: N1, N2, N3, N4, and to indicate whether they were blue-collar worker (1), white-collar worker (2), or professional (3).

Example 4...

Newspaper readership and occupation

Reader	Occupation	Newspaper
1	2	N2
2	1	N4
3	2	N1
...
...
352	3	N2
353	1	N3
354	2	N3

Example 4: Solution

By counting the number of times each of the 12 combinations occurs, we produce Table 3.9.

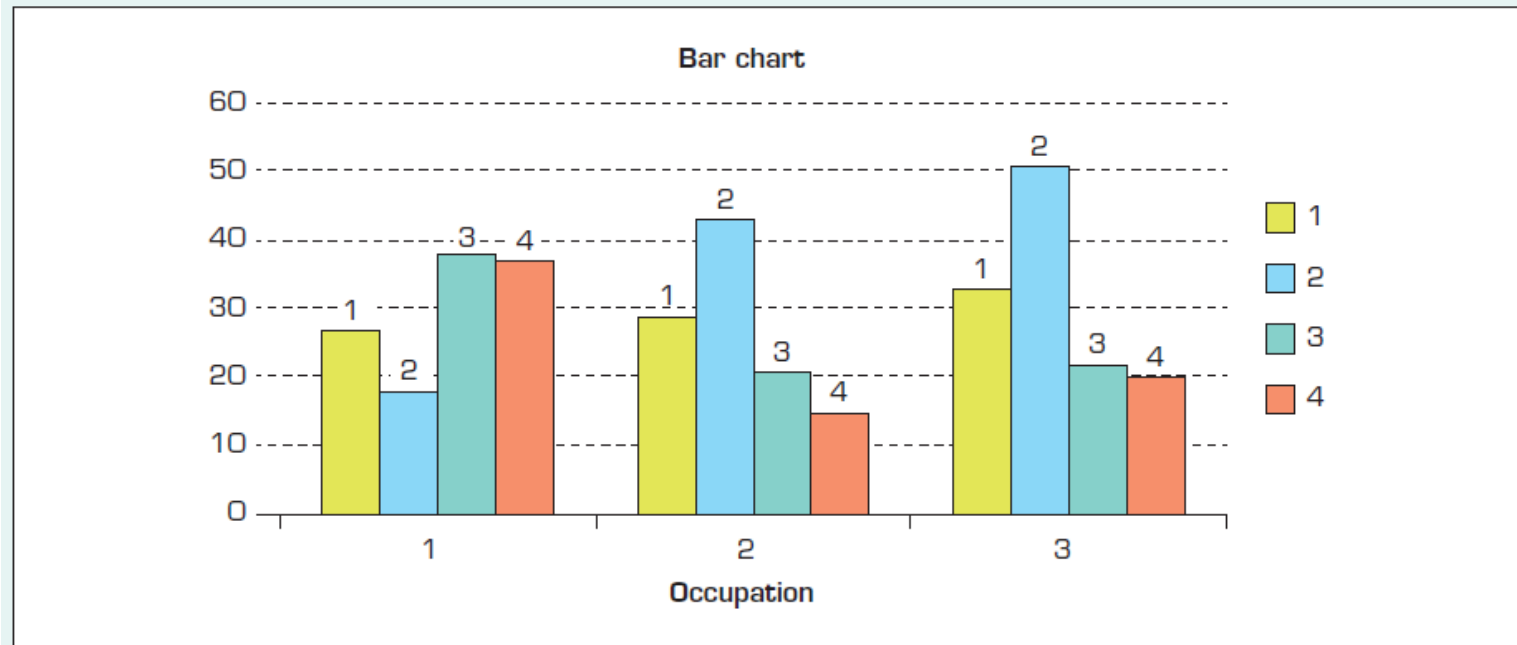
Table 3.9 Cross-classification table of frequencies for Example 3.7

Occupation	Newspaper				Total
	N1	N2	N3	N4	
Blue collar (1)	27	18	38	37	120
White collar (2)	29	43	21	15	108
Professional (3)	33	51	22	20	126
Total	89	112	81	72	354

Example 4: Solution...

The frequencies can be depicted in graphical form using a bar chart.

Figure 3.15 Excel bar chart showing readership of the four newspapers by occupation



Example 4: Solution...

If occupation and newspaper readership are related, then there will be differences in the newspapers read among the occupations.

An easy way to see this is to convert the frequencies in each row (or column) to relative frequencies using each row (or column) total.

That is, compute the row (or column) totals and divide each frequency by its row (or column) total.

Example 4: Solution...

Table 3.10 Row relative frequencies for Example 3.7

Occupation	Newspaper				Total
	N1	N2	N3	N4	
Blue collar (1)	0.23	0.15	0.32	0.31	1.00
White collar (2)	0.27	0.40	0.19	0.14	1.00
Professional (3)	0.26	0.40	0.17	0.16	1.00
Total	0.25	0.32	0.23	0.20	1.00

Example 4: Solution...

Interpretation:

Notice that the relative frequencies in the rows 2 (white-collar) and 3 (Professionals) are similar and that there are large differences between row 1 (blue-collar) and rows 2 and 3.

This tells us that:

- blue collar workers tend to read different newspapers from both white-collar workers and professionals; and
- white-collar workers and professionals are quite similar in their newspaper choices.