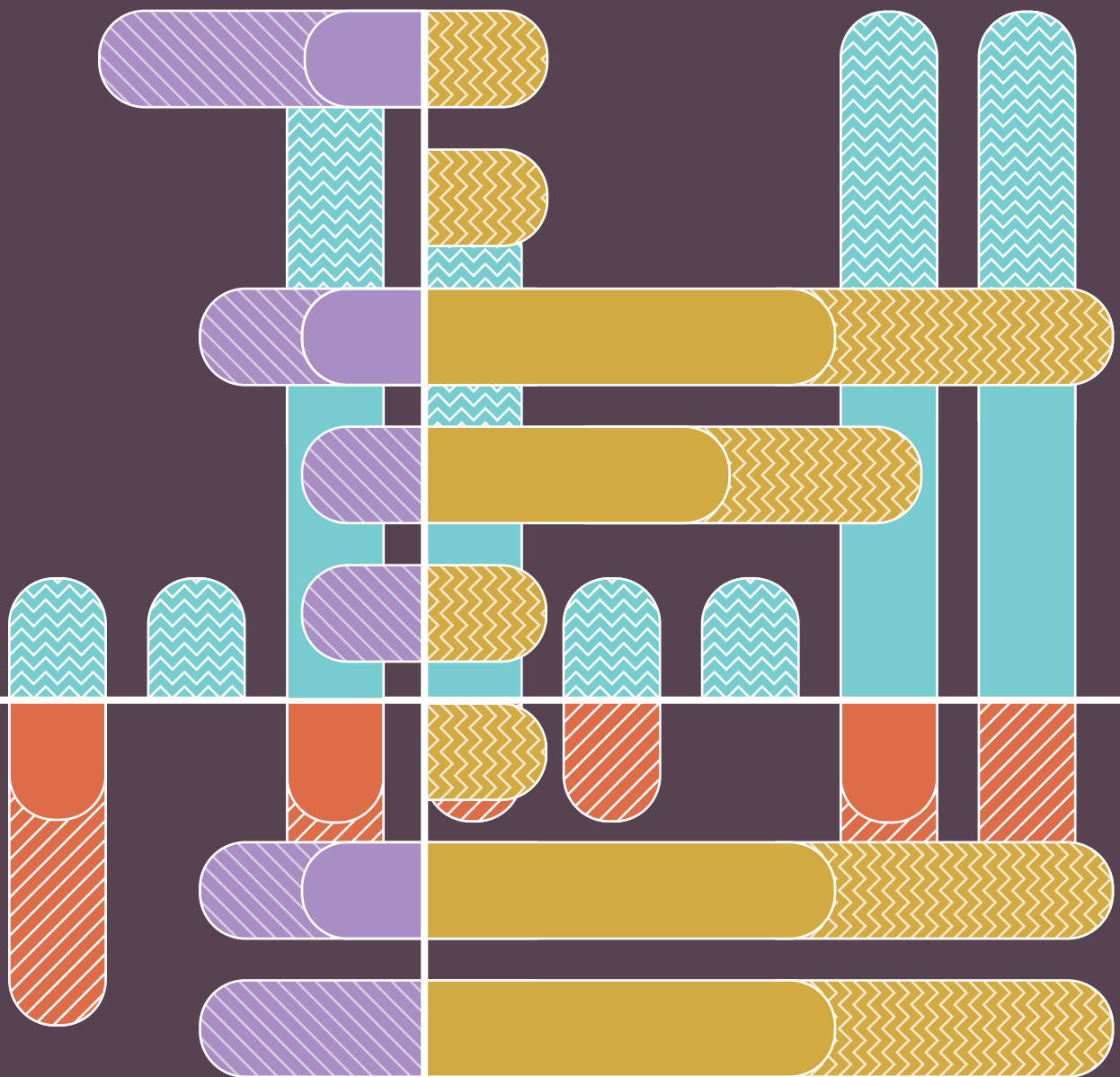


Business Statistics

Australia / New Zealand
8th edition

Solutions Manual



Eliyathamby A. Selvanathan

Saroja Selvanathan

Gerald Keller

1 What is statistics?

- 1.1 a Population:** The collection of all measurements of interest in a statistical problem; e.g., heights of all Australians.
- b Sample:** Any subset of measurements from a population; e.g., heights of 100 selected Australians.
- c Parameter:** A descriptive measure of the measurements in a population; e.g., average height of all Australians.
- d Statistic:** A descriptive measure of the measurements in a sample; e.g., average height of 100 selected Australians.
- e Statistical inference:** A conclusion about a characteristic of a population, based on the information provided by a sample drawn from the population; e.g., testing whether the average height of all Australian adults is greater than 150cm using the sample information based on the heights of 100 randomly selected Australian adults.
- 1.2** Descriptive statistics consists of graphical and numerical methods used to describe sets of data, both populations and samples. Inferential statistics consists of a body of methods used for drawing conclusions about characteristics of a population, based on information available in a sample drawn from the population.
- 1.3 a** Views on internet banking of the 12 000 customers
- b** Views on internet banking of the 300 customers surveyed
- c** Statistic.
- 1.4 a** The complete production run of light bulbs
- b** 1000 bulbs selected
- c** The proportion of the light bulbs that are defective in the whole production run.
- d** The proportion of bulbs that are defective in the sample of 1000 bulbs selected.
- e** Parameter
- f** Statistic
- g** Because the sample proportion (1%) is much less than the claimed 5%, we can conclude with the confidence that there is evidence to support the claim.
- 1.5** Select a number of graduates (say 100) from each group (Business group and Arts and Science group) and workout the sample mean salaries of the two groups. Compare the two mean values to see whether there is some support for the claim that in general, average salary of business graduates is greater than the average salary of arts and science graduates.

- 1.6**
- a** Flip the coin 100 times and count the number of heads and tails.
 - b** Outcomes of unlimited number of repeated flips of the coin.
 - c** Outcomes of the 100 flips.
 - d** The (population) proportion p of heads in unlimited number of flips is expected to be 0.5.
 - e** The sample proportion $\hat{p}$ of heads in the 100 flips.
 - f** If the sample proportion $\hat{p}$ is close to the population proportion $p = 0.5$, we conclude that there is some support for the claim that the coin is a fair coin.
- 1.7**
- a** The coin is not a fair coin as the sample proportion $\hat{p} = 0.95$ is not close to $p = 0.50$ (for fair coins).
 - b** The coin may be a fair coin as $\hat{p} = 0.55$, close to $p = 0.50$ for fair coins. Need more trials (say, 1000) to confirm this.
 - c** If it is not a fair coin, the answer is no. The number of heads and tails can be anywhere between 0 to 100, out of 100 trials (for example, 60 heads and 40 tails, or 30 heads and 70 tails). If it is a fair coin, one would expect the number of heads and tails to be close to 50, out of 100 trials.

2 Types of data, data collection and sampling

- 2.1 Numerical:** **a** Kilometres commuted to work
 b Age of students in a statistics class.
- Ordinal:** **a** Fortnightly Australian family income
 i Under \$1500
 ii \$1500–2000
 iii \$2000–2500
 iv \$2500 or over.
 b Patient’s condition: excellent, good, fair or poor.
- Nominal:** **a** Country of origin of Australians
 b Brand of car owned.
- 2.2** **a** Numerical
 b Numerical
 c Ordinal
 d Numerical
 e Numerical.
- 2.3** **a** Numerical
 b Nominal
 c Numerical
 d Ordinal
 e Numerical.
- 2.4** **a** Numerical
 b Nominal
 c Nominal
 d Ordinal
 e Numerical
 f Ordinal.
- 2.5** **a** Ordinal
 b Numerical
 c Nominal
 d Numerical
 e Numerical.
- 2.6** **a** Numerical
 b Ordinal
 c Nominal
 d Numerical
 e Ordinal
 f Nominal.

- 2.7**
- a** Nominal
 - b** Numerical
 - c** Nominal
 - d** Ordinal
 - e** Numerical.
- 2.8**
- a** Numerical
 - b** Ordinal
 - c** Nominal
 - d** Numerical
 - e** Nominal
 - f** Ordinal.
- 2.9**
- a** Numerical
 - b** Numerical
 - c** Nominal
 - d** Ordinal
 - e** Numerical.
- 2.10**
- a** Ordinal
 - b** Ordinal
 - c** Ordinal
 - d** Numerical.
- 2.11** Primary data are published by the original source. Secondary data are published by someone other than whoever originally collected and published the data. Secondary data sources often summarise much of the original data, resulting in a loss of some information.
- 2.12**
- a** *Australian Bureau of Statistics*; Year Book, Australia (annual); rate of unemployment, population
 - b** *Reserve Bank Bulletin* (monthly); interest rate, exchange rate
 - c** *CIA Fact Book* (annual); electricity consumption, flags of the world
- Note: The two specific pieces of information contained in the latest issue of these publications will, of course, vary considerably, unless the instructor is more specific about the information requested.
- 2.13** In an observational study, there is no attempt to control factors that might influence the variable of interest. In an experimental study, a factor (such as regular use of a fitness centre) is controlled by randomly selecting who is exposed to that factor, thereby reducing the influence of other factors on the variable of interest.
- 2.14**
- a** This is an observational study, because no attempt is made to control factors that might influence cola sales, such as store location or store type.
 - b** Randomly select which stores (both grocery and convenience) receive cola in bottles to reduce the influence of factors like location. Separately analyse the two types of stores in order to reduce the influence of store type.

- 2.15** a Randomly select 4000 people over the age of 50. Compare the proportion of smokers who have lung cancer with the proportion of non-smokers who have lung cancer.
 b The study described in part a is observational, because we haven't controlled who smoked.
- 2.16** a A survey can be conducted, for example, by means of a personal interview, a telephone interview, or a self-administered questionnaire.
 b A personal interview has a high response rate relative to other survey methods, but is expensive because of the need to hire well-trained interviewers and possibly pay travel-related costs if the survey is conducted over a large geographical area. A personal interview will also probably result in fewer incorrect responses arising from respondents misunderstanding some questions. A telephone interview is less expensive, but will probably result in a lower response rate. A self-administered questionnaire is least expensive, but suffers from lower response rates and accuracy than personal interviews.
- 2.17** Five important points to consider when designing a questionnaire are as follows:
- The questionnaire should be short.
 - Questions should be clearly worded and unambiguous.
 - Consider using dichotomous or multiple-choice questions, but take care that respondents needn't make unspecified assumptions before answering the questions.
 - Avoid using leading questions.
 - When preparing the questions, think about how you intend to tabulate and analyse the responses.
- 2.18** a The sampled population will exclude those who avoid large department stores in favour of smaller shops, as well as those who consider their time too valuable to spend participating in a survey. The sampled population will therefore differ from the target population of all customers who regularly shop at the mall.
 b The sampled population will contain a disproportionate number of thick books, because of the manner in which the sample is selected.
 c The sampled population consists of those eligible voters who are at home in the afternoon, thereby excluding most of those with full-time jobs (or at school).
- 2.19** We used Excel to generate 40 three-digit random numbers. Because we will ignore all randomly generated numbers over 800, we can expect to ignore about 20% (or about 8 to 10) of the randomly generated numbers. We will also ignore any duplications. We therefore chose to generate 40 three-digit random numbers, and will use the first 25 unique random numbers less than 801 to select our sample. The 40 numbers generated are shown below, with a stroke through those to be ignored.

6	357	456	449	862	154	55	412	475	430
999	912	60	207	717	651	10	294	327	165
576	871	990	354	390	540	893	181	496	870
738	820	32	963	160	32	231	86	970	46

- 2.20** We used Excel to generate 30 three-digit random numbers. Because we will ignore any duplicate numbers generated, we generated 30 three-digit random numbers and will use the first 20 unique random numbers to select our sample. The 30 numbers generated are shown below.

169	470	744	530	554	918
318	858	698	203	383	938
836	116	123	936	539	154
110	630	856	380	145	692
909	269	811	274	553	749

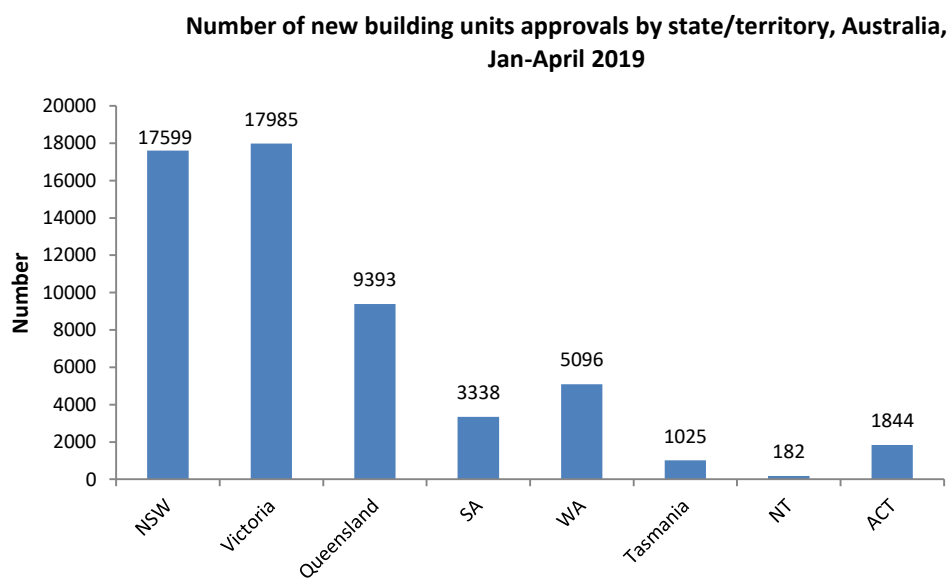
- 2.21** The operations manager can select stratified random samples where the strata are the four departments. Simple random sampling can be conducted in each department.
- 2.22** Stratified random sampling is recommended. The strata are the school of business, the faculty of arts, the graduate school and the all the other schools and faculties would be the fourth stratum. The data can be used to acquire information about the entire campus but also compare the four strata.
- 2.23** A stratified random sampling plan accomplishes the president's goals. The strata are the four areas enabling the statistics practitioner to learn about the entire population but also compare the four areas.
- 2.24 a** Sampling error refers to an inaccuracy in a statement about a population that arises because the statement is based only on sample data. We expect this type of error to occur because we are making a statement based on incomplete information. Nonsampling error refers to mistakes made in the acquisition of data or due to the sample observations being selected improperly.
- b** Nonsampling error is more serious because, unlike sampling error, it cannot be diminished by taking a larger sample.
- 2.25** Three types of nonsampling errors:
- Error due to incorrect responses
 - Nonresponse error, which refers to error introduced when responses are not obtained from some members of the sample. This may result in the sample being unrepresentative of the target population.
 - Error due to selection bias, which arises when the sampling plan is such that some members of the target population cannot possibly be selected for inclusion in the sample.
- 2.26** Yes. A census will probably contain significantly more nonsampling errors than a carefully conducted sample survey.

Part 1

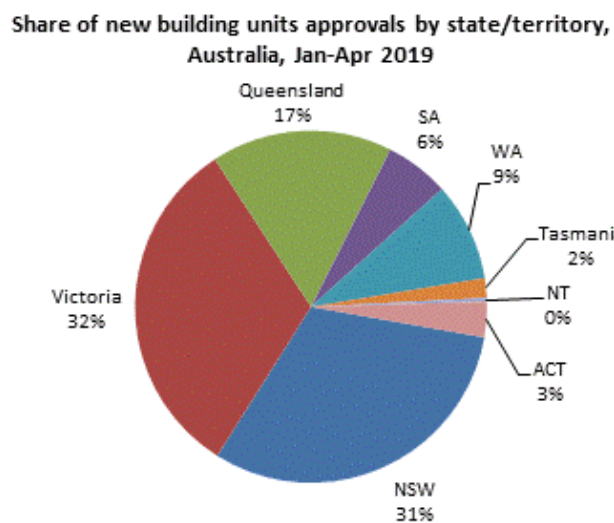
Descriptive measures and probability

3 Graphical descriptive techniques – nominal data

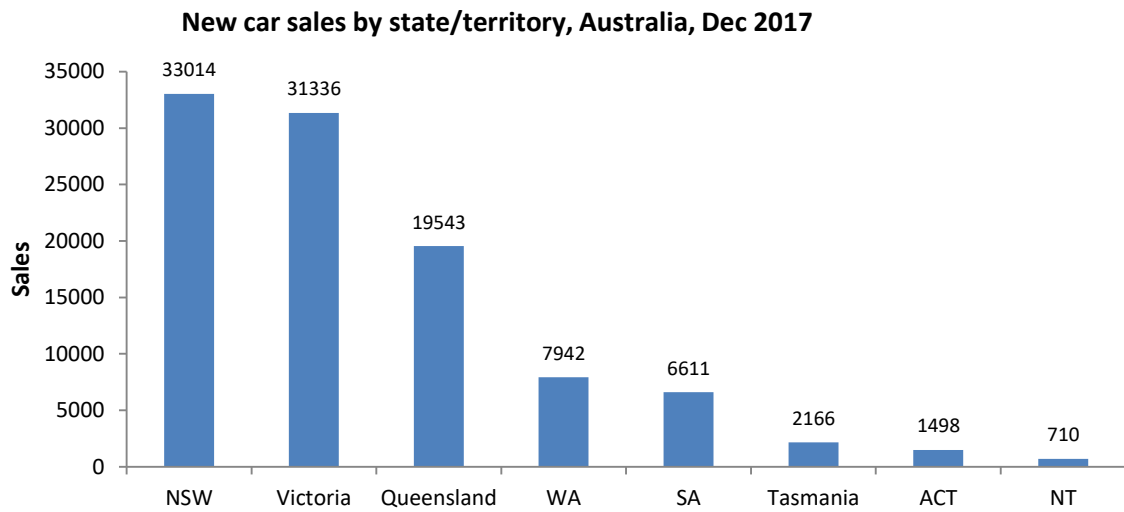
3.1 a



b

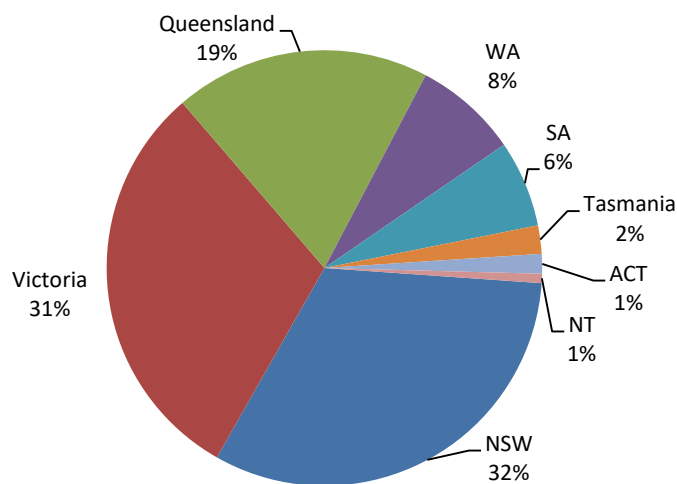


3.2 a A bar chart would be appropriate.

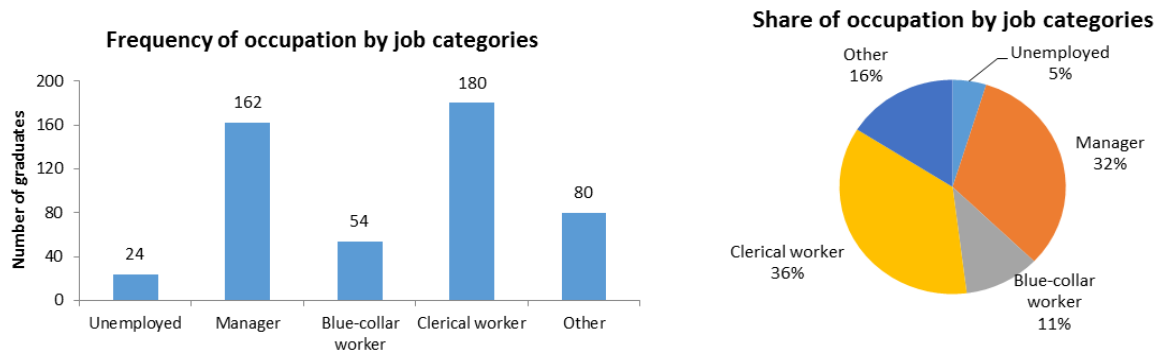


b A pie chart would be appropriate.

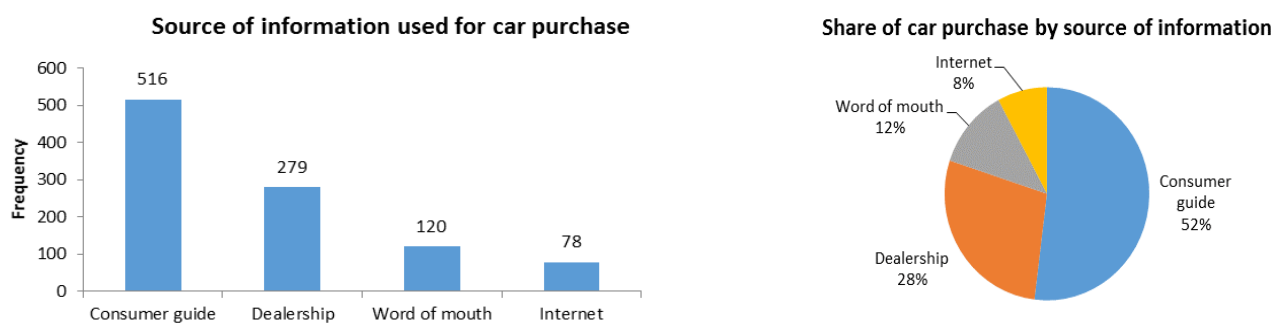
Share of new car sales by state/territory, Australia, Dec 2017



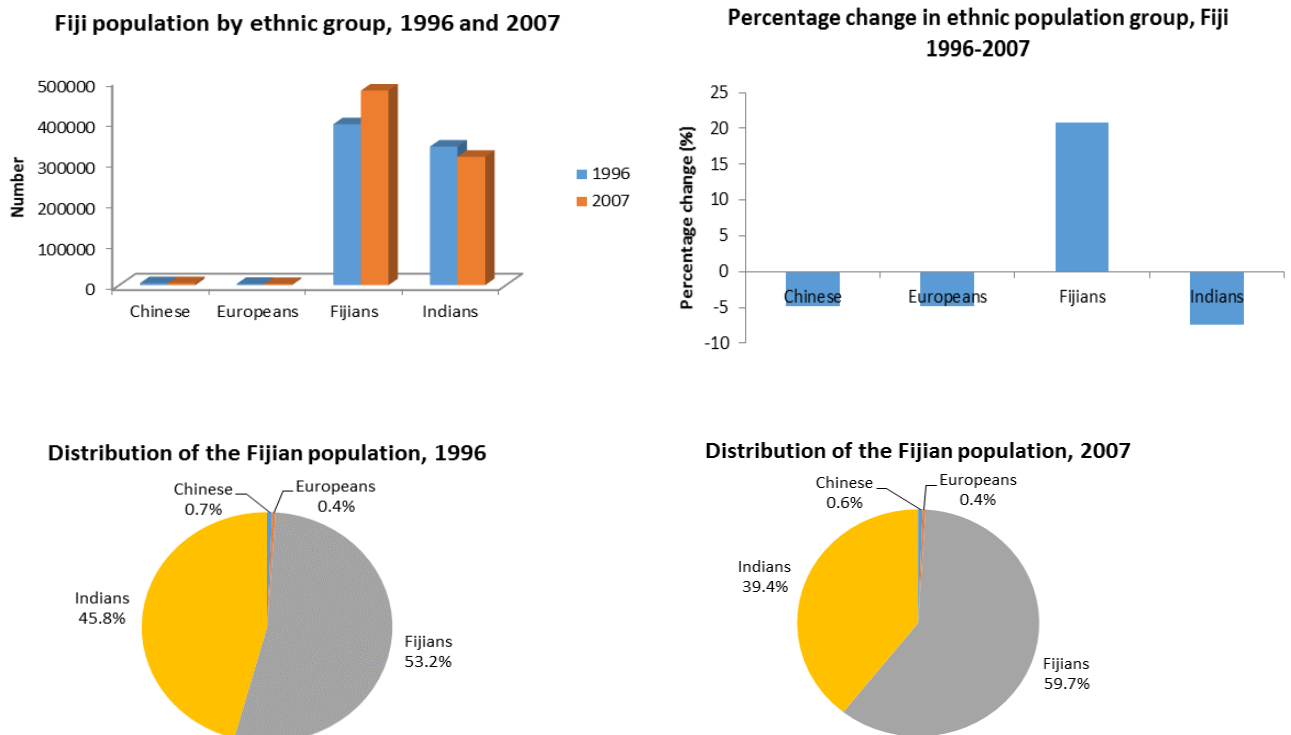
3.3 A pie or bar chart can be used; however, a pie chart is more suitable here.



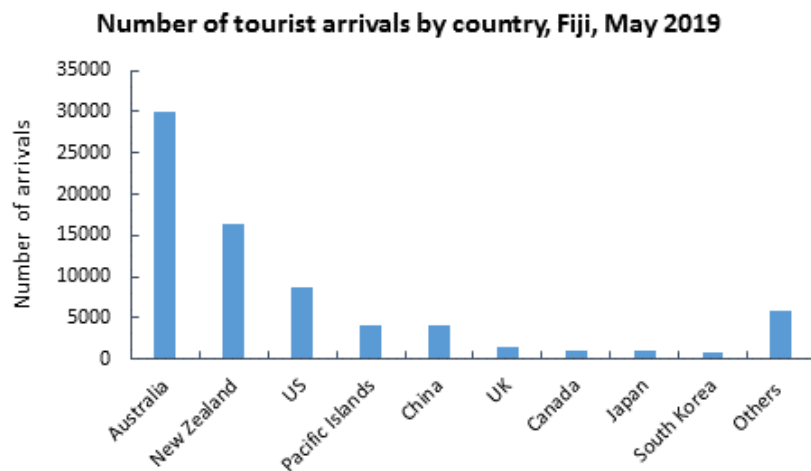
3.4 A pie or bar chart can be used; however, a pie chart is more suitable here.



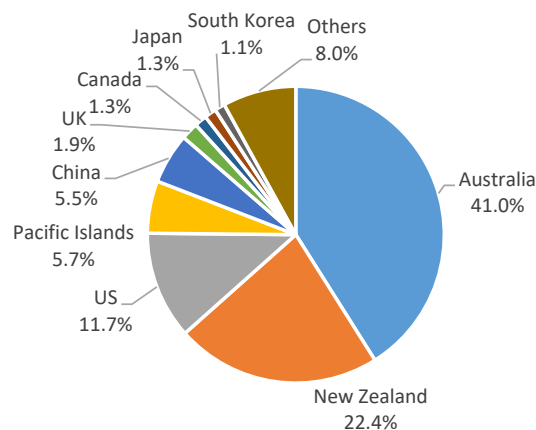
3.5



- 3.6 To compare the number of tourist arrivals by country to Fiji, a bar chart would be appropriate. On the other hand, to compare the share of the tourism market from each country, a pie chart would be appropriate.

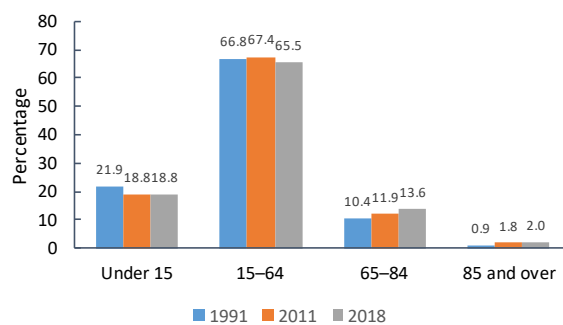


Share of tourist arrivals by country, Fiji, May 2019

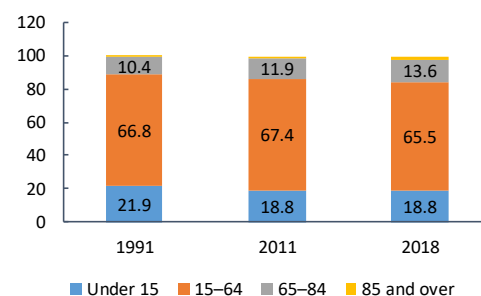


3.7

**Share of Australian Population by age group
1991, 2011 and 2018**

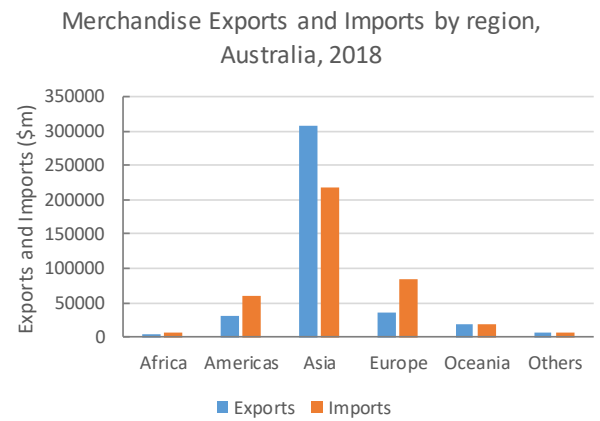
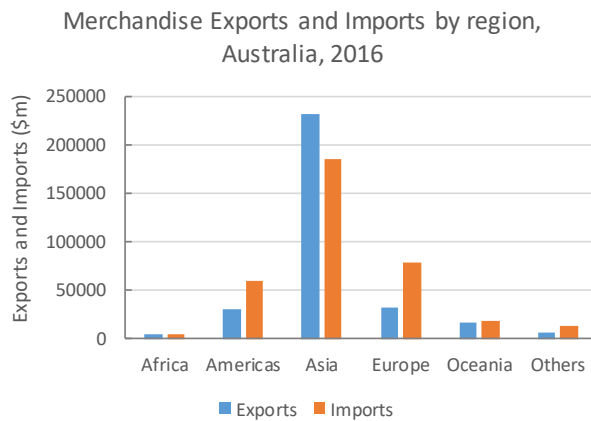


**Share of Australian Population by
age group 1991, 2011 and 2018**

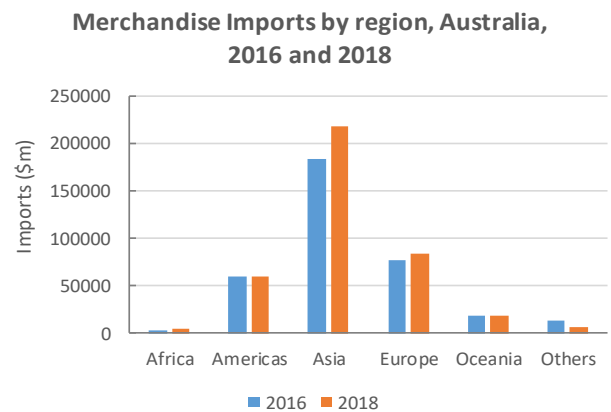
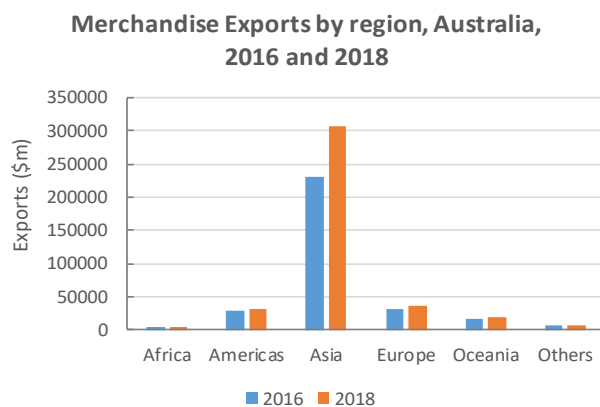


3.8 To compare the exports and imports in a particular year to and from the 6 trading partner regions, bar charts in (i) and (ii) would be useful. To compare the exports during 2016 and 2018, bar chart (iii) would be useful. Similarly, to compare the imports during 2016 and 2018, bar chart (iv) would be useful.

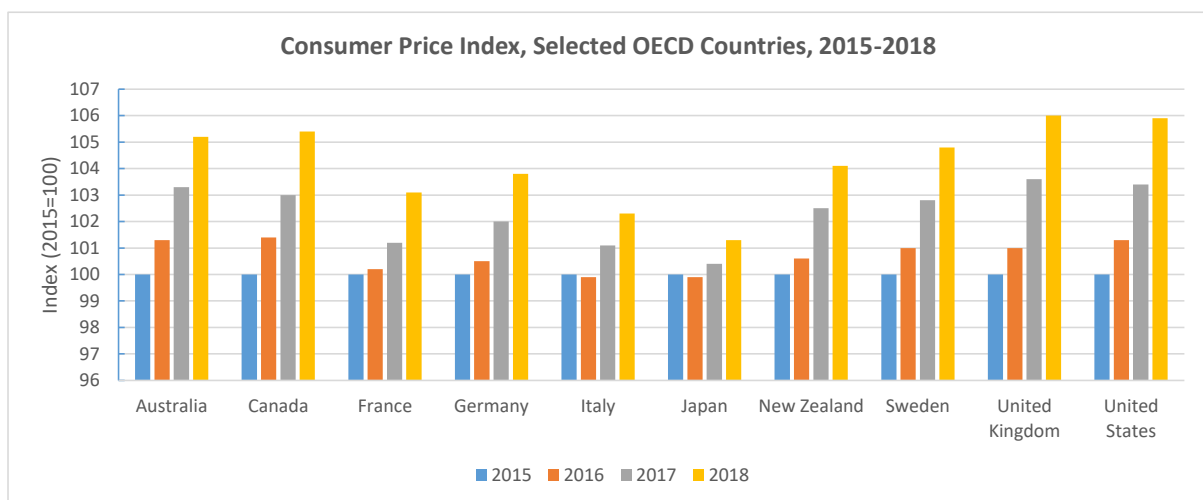
(i) & (ii)



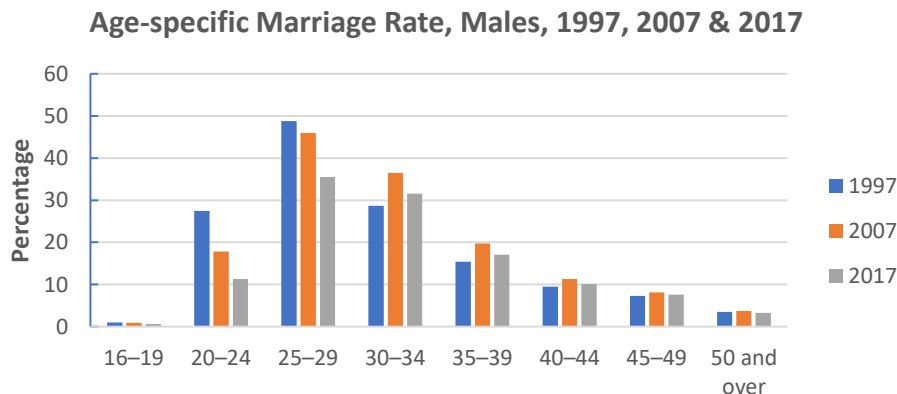
(iii) & (iv)



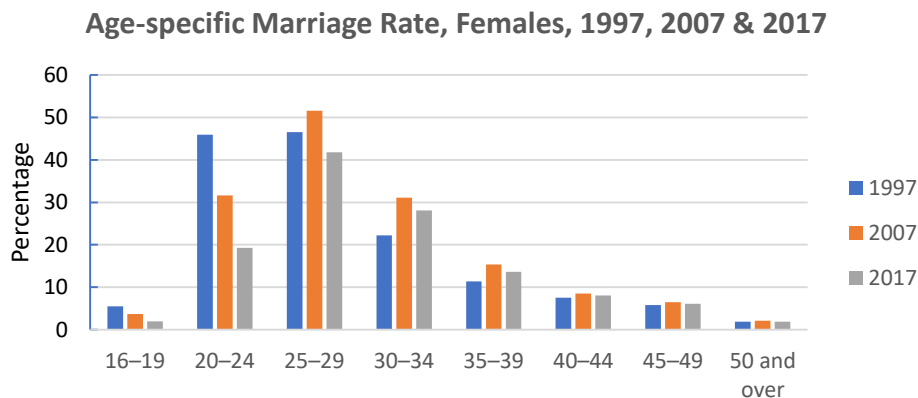
3.9 A bar chart would be appropriate.



- 3.10 a** Appropriate graph to compare the marriage rates for males would be a bar chart using data for Males only.

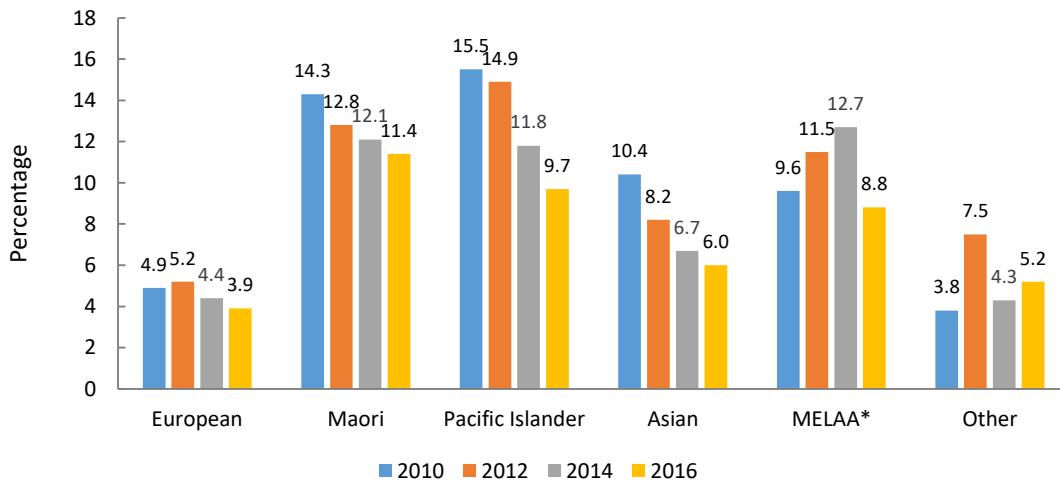


- b** Appropriate graph to compare the marriage rates for females would be a bar chart using data for Females only.

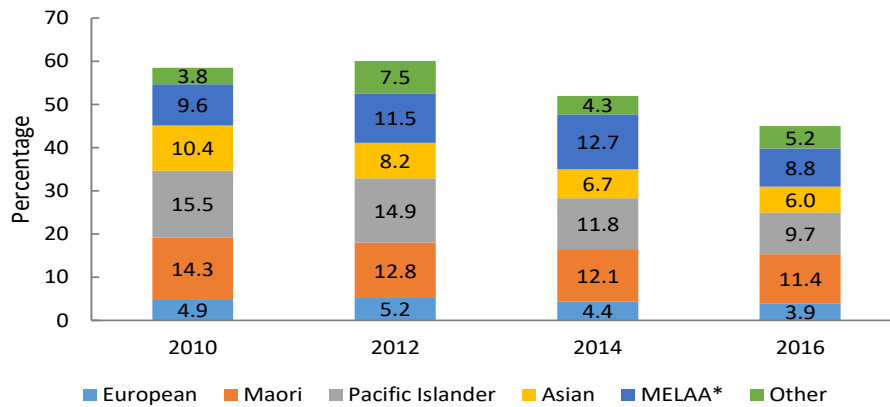


- c** Among the age groups for males, the marriage rate has fallen for the 16-19, 20-24, and 25-29 age groups, while it has increased for all the age groups 30 years and over between 1997, 2007 to 2017. For females, the marriage rate has fallen for the 16-19 and 20-24 age groups, while it has increased for all other groups between 1997 and 2007. For both males and females, for 30 years and over groups, the marriage rate has decreased slightly between 2007 and 2017, but still the rate is higher than the 1997 levels. This shows that the age at marriage has increased between 1997 and 2017 for both males and females. That is, more and more males and females are waiting longer to get married.
- d** A bar chart is more appropriate as the aim is to directly compare the marriage rates between the years 1997 and 2017, among males and females separately.
- 3.11** A bar chart would be appropriate to compare the rate of unemployment during the three years, 2010, 2012, 2014 and 2017 for each ethnic group. A component bar chart would be appropriate to compare the level of unemployment among the ethnic groups for each year.

**Rate of unemployment by ethnic group, New Zealand
2010, 2012, 2014 and 2016**

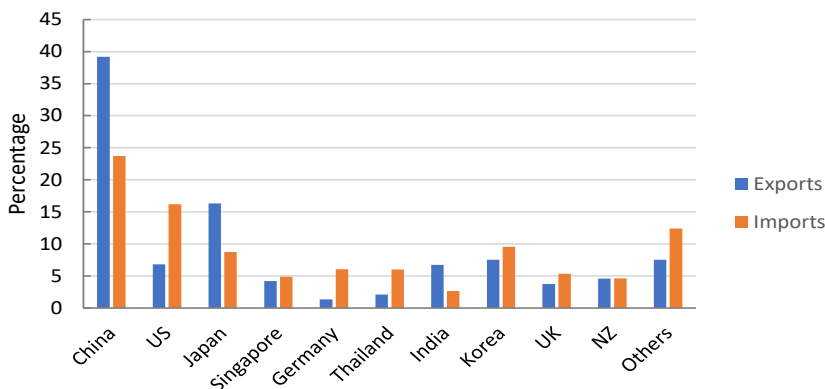


**Rate of unemployment by ethnic group, New Zealand,
2010, 2012, 2014 and 2016**

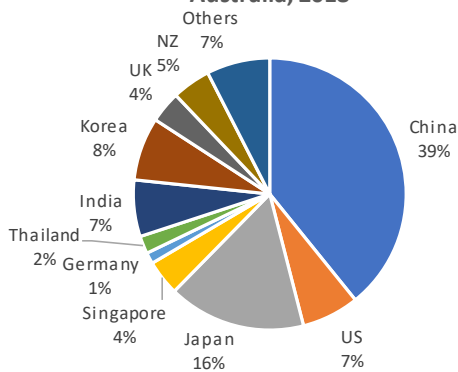


3.12 a

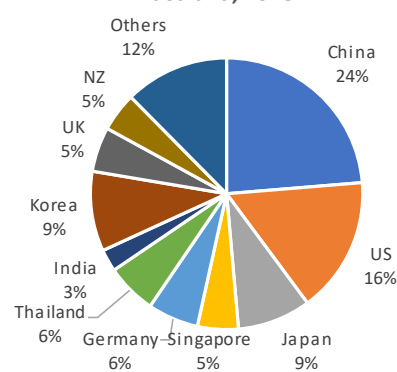
**Share of Australian exports and imports,
10 major trading partners, Australia, 2018**



Share of exports, 10 major trading partners, Australia, 2018



Share of imports, 10 major trading partners, Australia, 2018



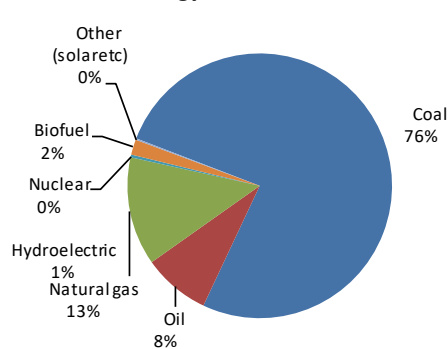
As can be seen from the bar chart, the share of the value of exports is higher than that of imports for China, Japan and India, while the share of the value of imports is higher than that of exports for all other countries. China is the top trading partner of Australia in terms of both imports and exports followed by Japan and then the US.

- b** Since the information is given in percentage shares, either a bar chart or pie chart would be appropriate for comparison. However, a pie chart would be more appropriate to easily see the major contributors to Australian exports and imports. To compare the exports and imports for each country the bar chart is easier to visualise.

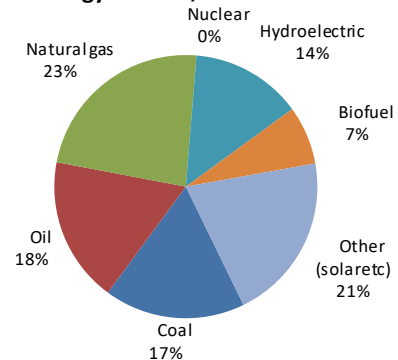
- 3.13** Pie charts (a) are helpful to compare the share of all forms of energy sources between and within Australia and New Zealand. Pie charts (b) for the non-renewable energy sources for Australia and New Zealand would be appropriate for comparison of the contribution of non-renewable energy sources within and between the two countries. Separate pie charts (c) for the renewable energy sources for Australia and New Zealand would be appropriate for comparison of the contribution of renewable energy sources within and between the two countries.

(a)

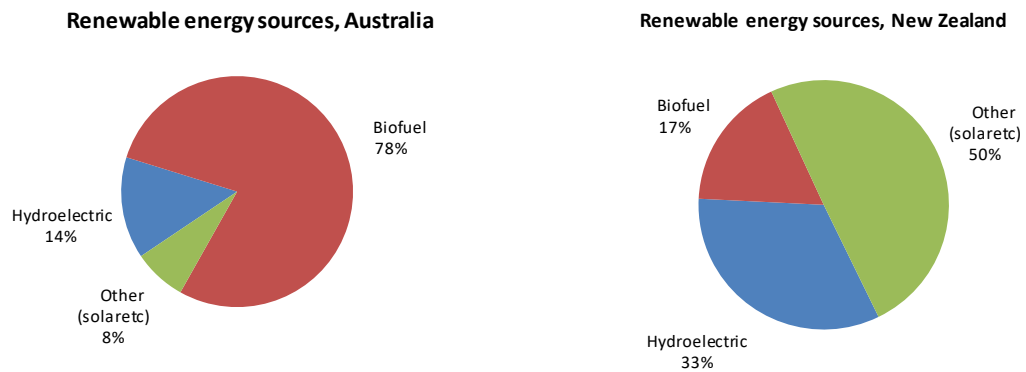
Energy sources, Australia



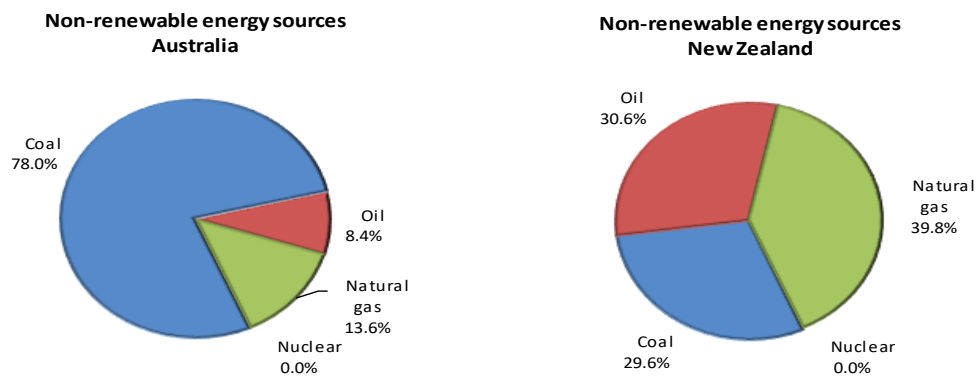
Energy sources, New Zealand



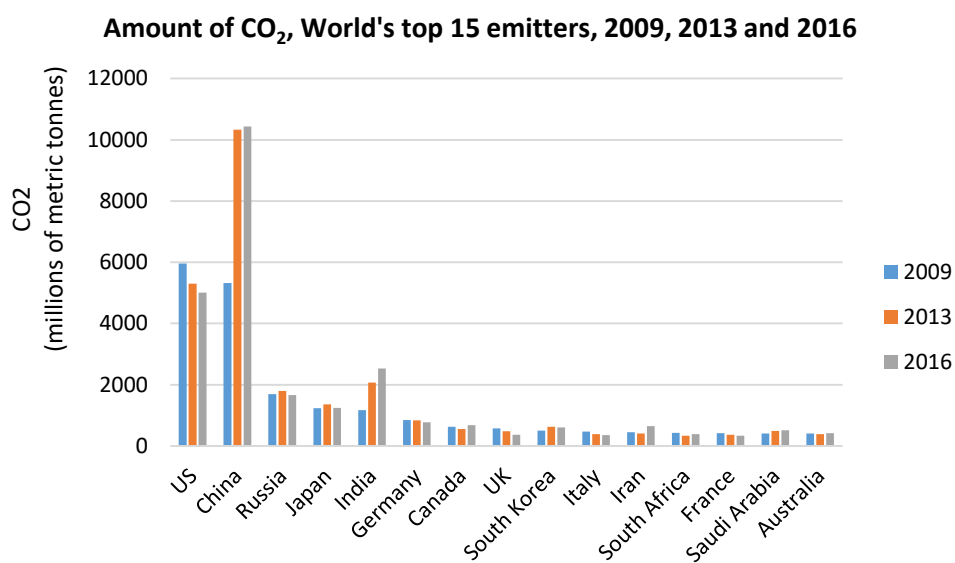
(b)



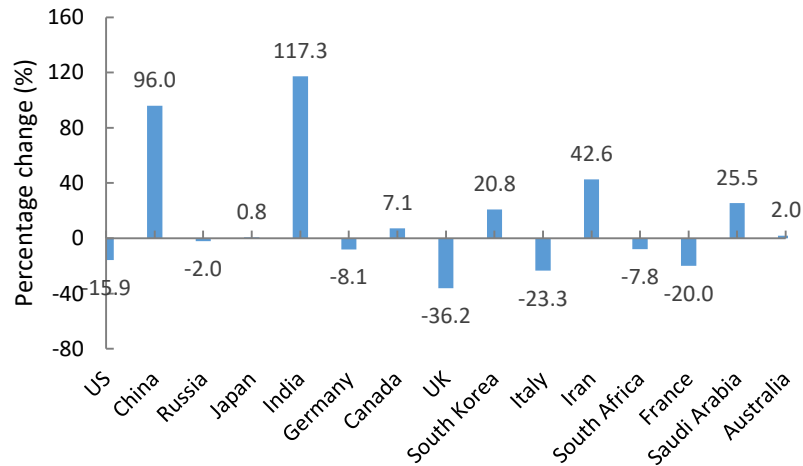
(c)



3.14 A bar chart would be appropriate. To see how countries are performing in terms of reducing CO₂ emissions, a bar chart of percentage change in emissions from 2009 to 2016 for the 15 countries would be appropriate.



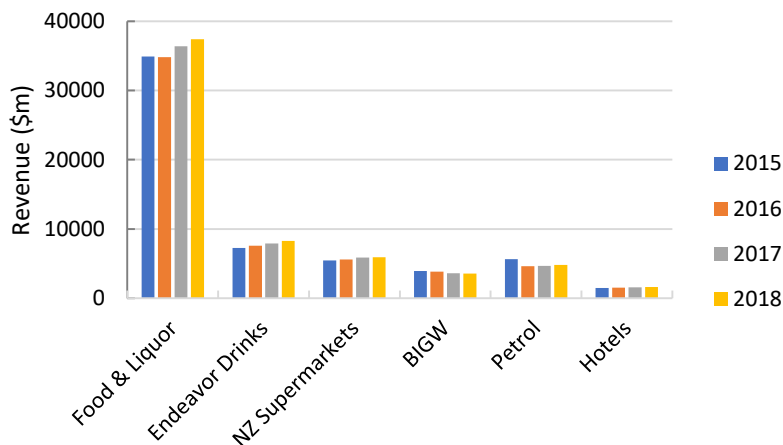
World's top 15 CO₂ emitters, Percentage change in emissions between 2009 and 2016

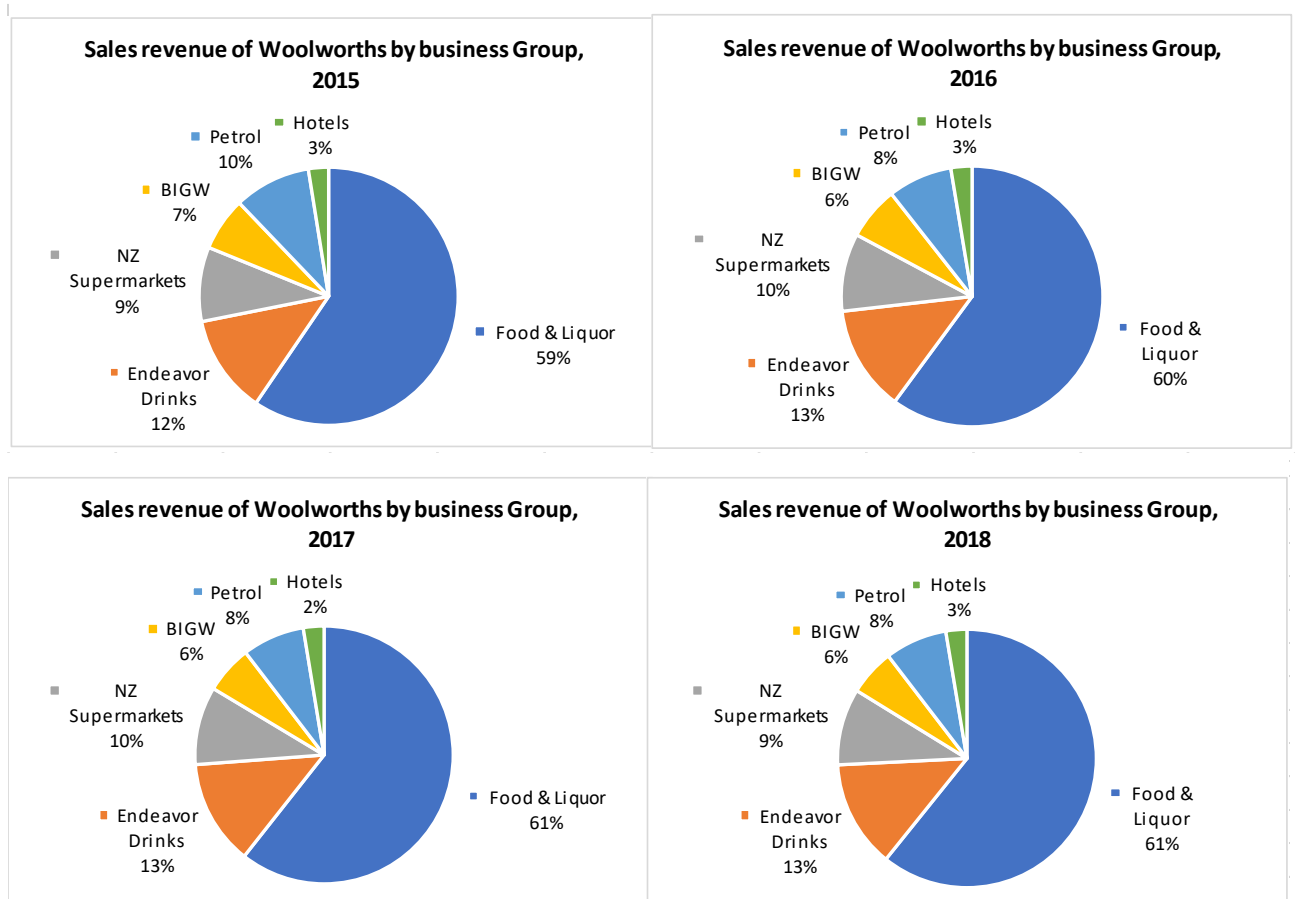


The top CO₂ emitters in the world in 2016 are China followed by the US, India, Russia and Japan in that order. Between 2009 and 2016, India, China, Iran, Saudi Arabia, South Korea, Canada, Australia and Japan have increased their CO₂ emissions while Russia, South Africa, Germany, US, France, Italy and the UK have reduced their CO₂ emissions.

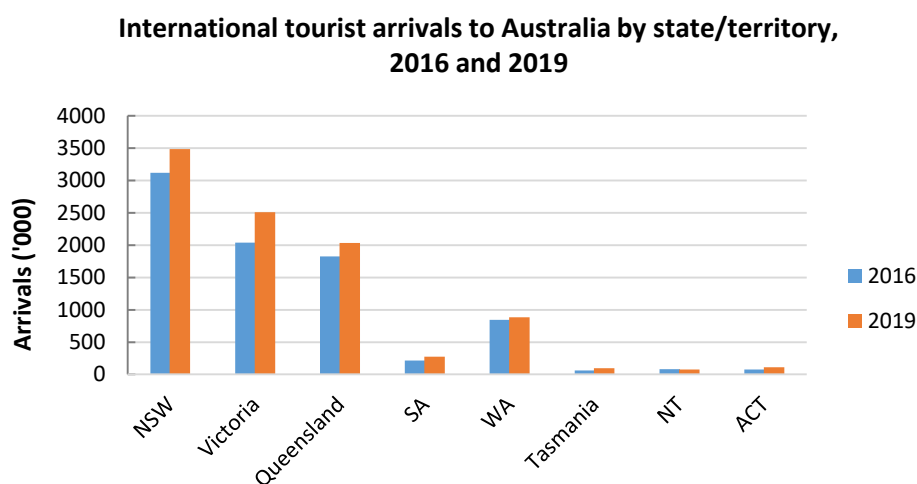
- 3.15** A bar chart would be useful to compare the values of each Woolworths business group across the four years, 2015, 2016, 2017 and 2018. As can be seen, there has been steady but slow increase in revenue for all business groups, except BIGW. Pie charts, one for each year, are useful for comparison of the share contribution of revenues from various business groups during the four years. As can be see, the share of revenue across the various business groups have remained similar over the four years.

Sales revenue of Woolworths by business group, 2015-2018

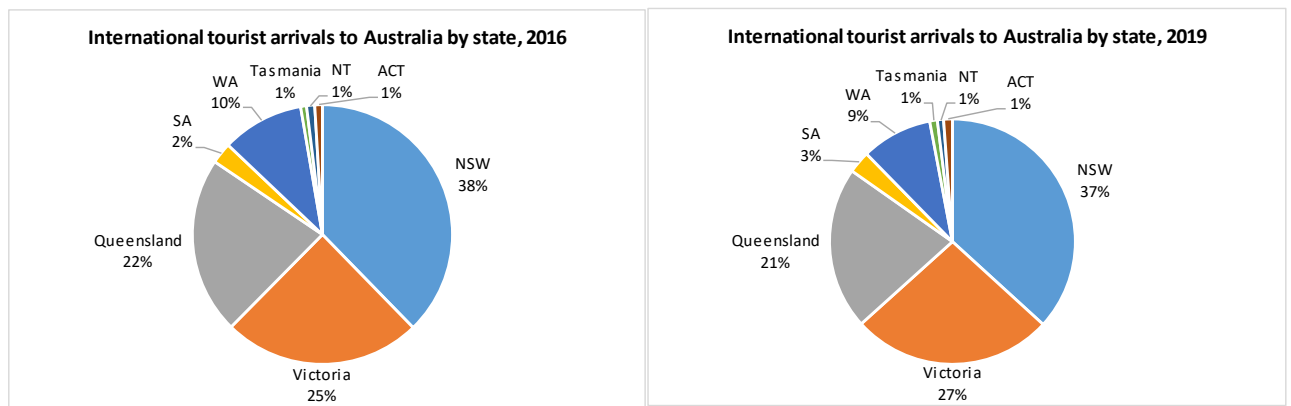




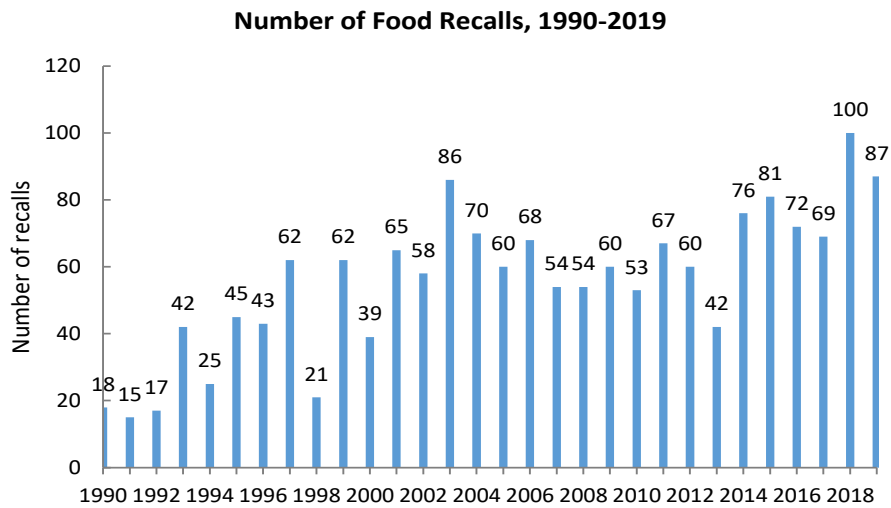
3.16 A bar chart would be appropriate to compare the change in the number of tourist arrivals from 2016 to 2019. The tourist arrivals to NSW, Victoria and Queensland has increased considerably between 2016 and 2019. Tourist arrivals to other states and territories have increased slightly or remained unchanged. A pie chart, one for each year, would be useful to analyse the share of tourist arrivals to the different states and territories in Australia. The share of tourist arrivals to the different states and territories have remained nearly the same in 2016 and 2019.



b

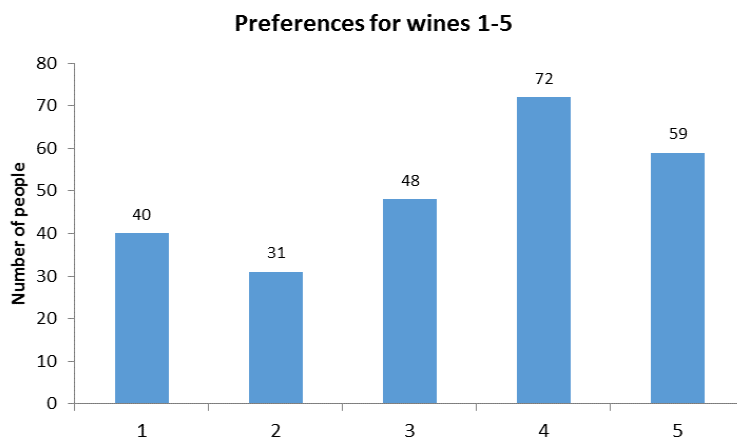


3.17



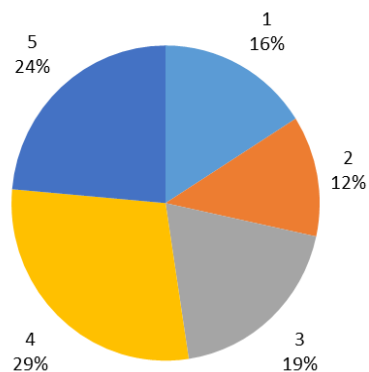
3.18

- a If the focus is on the actual numbers who prefer each type of wine, then a bar chart would be useful.

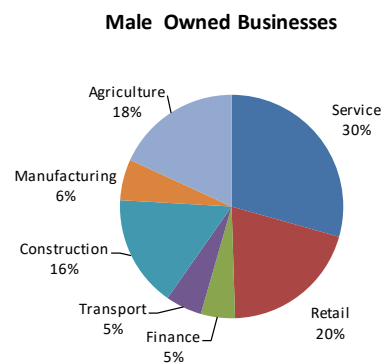
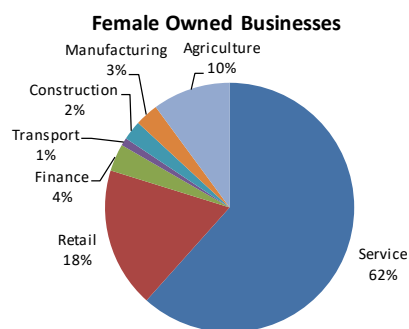


- b** Pie chart would be useful to show the share of the preferences for each type of wine among all types of wine.

Share of preferences for Wines 1-5



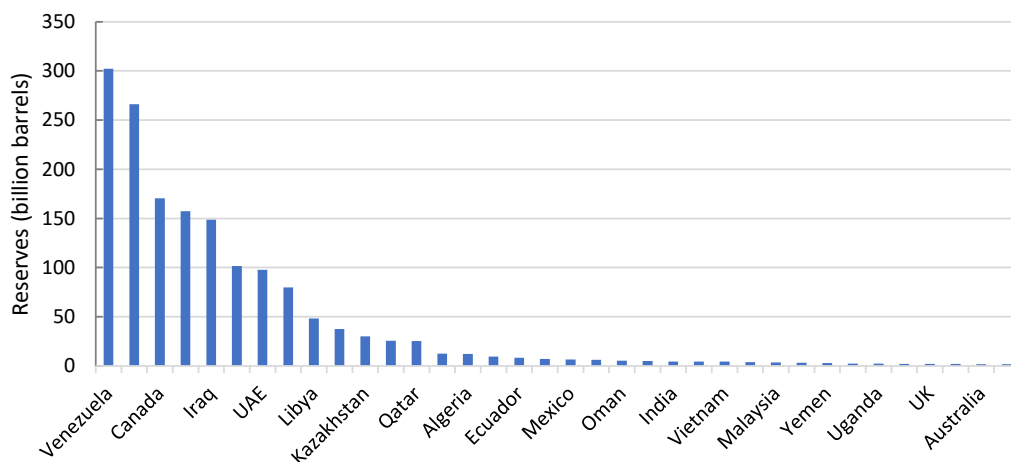
- 3.19** Pie charts of female-owned and male-owned businesses would provide the share of each business type for comparison.



The biggest difference is that 62% of female-owned businesses are in the services sector, compared with 30% of male-owned businesses. Only 3% of female-owned businesses are in the construction sector, compared with 16% for male-owned businesses.

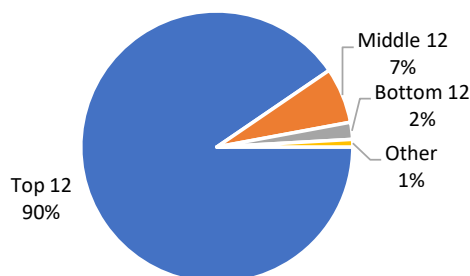
3.20

World crude oil reserves , Top 36 Countries, 2018



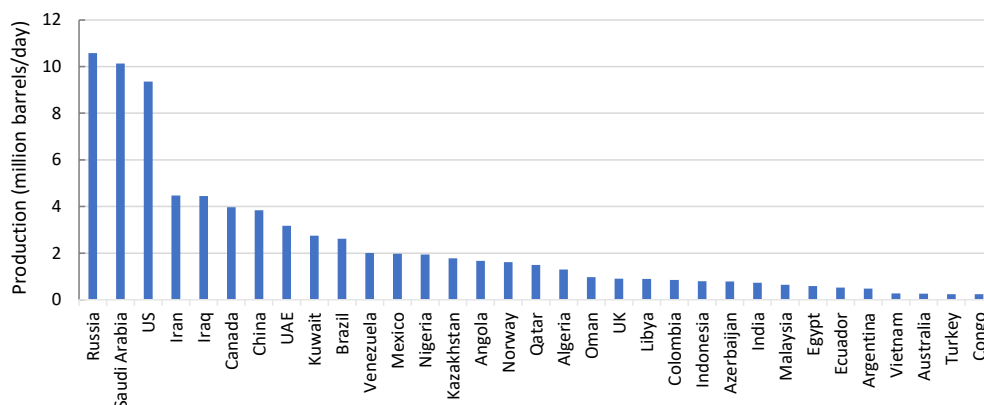
- 3.21** A pie chart for the four groups of top 12, middle 12, bottom 12, and other countries would emphasise the breakdown of oil reserves among the four groups of countries. As can be seen, the top 12 countries have 90% of the world oil reserves. Only 1% of the oil reserves are in countries other than the top 36 countries.

Crude oil reserve shares of groups of the top 36 countries, 2018



3.22

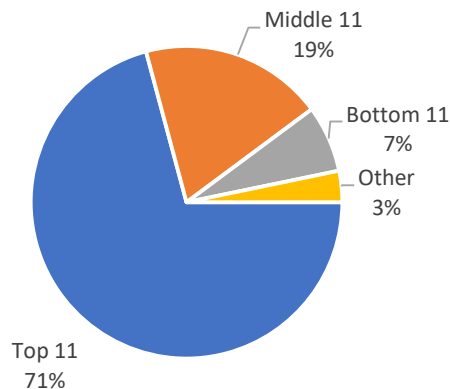
World oil production, Top 33 countries, 2018



Russia, Saudi Arabia and the US are the world's largest oil producers, followed by Iran, Iraq, Canada, China, UAE, Kuwait and Brazil.

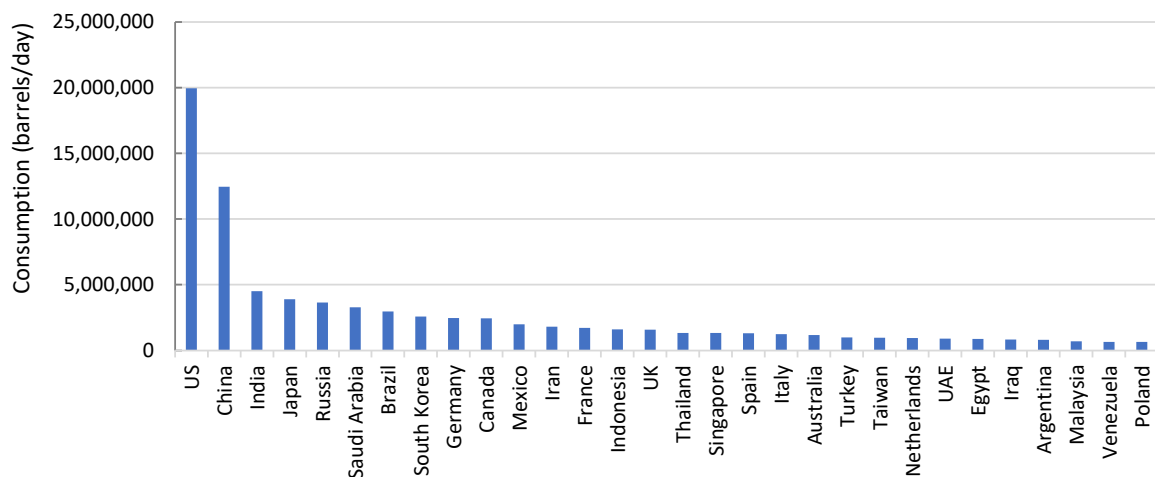
- 3.23** A pie chart for the four groups, top 11, middle 11, bottom 11, and other countries, would emphasise the breakdown of oil production among the four groups of countries. As can be seen, the top 11 countries produce 71% of the world's oil production and the middle 11 countries produce only 19%. Only 3% of the world's oil production occurs outside the top 33 countries.

Share of oil production by groups of the top 33 countries, 2018



3.24

Average oil consumption, Top 30 Countries, 2018

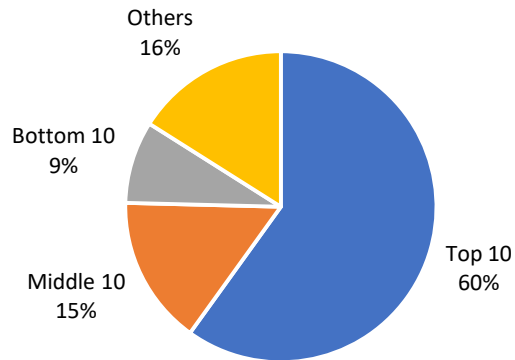


The US is the largest oil consumer, China is the second largest, followed by India, Japan, Russia, Saudi Arabia, Brazil, South Korea Germany and Canada.

- 3.25** A pie chart for the three groups of top 10, middle 10 and bottom 10 countries would emphasise the breakdown of oil consumption across the 30 countries. As can be seen, the top 10 countries consume about 60% of the world's oil production and the middle

and bottom 20 countries consume about 24%. Only 16% of the world's oil consumption occurs outside the top 30 countries.

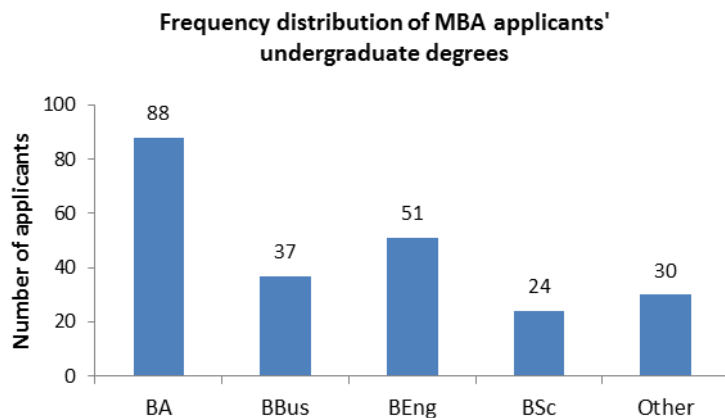
Share of oil consumption by groups of the top 30 countries, 2018



3.26 a Frequency distribution of undergraduate degrees of MBA applicants.

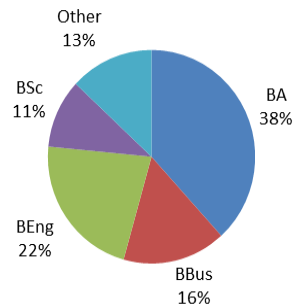
Degree	Frequency
BA	88
BBus	37
BEng	51
BSc	24
Other	30
Total	230

b



c

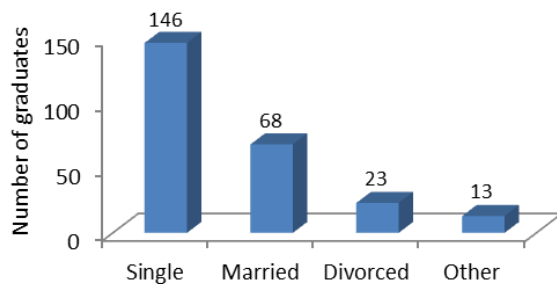
MBA applicants by type of undergraduate degree



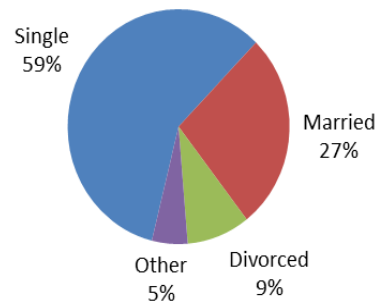
d. The charts in b and c show that a majority of applicants are BA graduates, capturing 88 (38%) of the applicants, followed by BEng 51 (22%), then BBus 37(16%) and BSc 24 (11%).

3.27 A bar or pie chart could be used. The graphs show that a majority of recent graduates are single (59%). About 27% of them are married and 9% are divorced.

Frequency distribution of recent graduates by their marital status

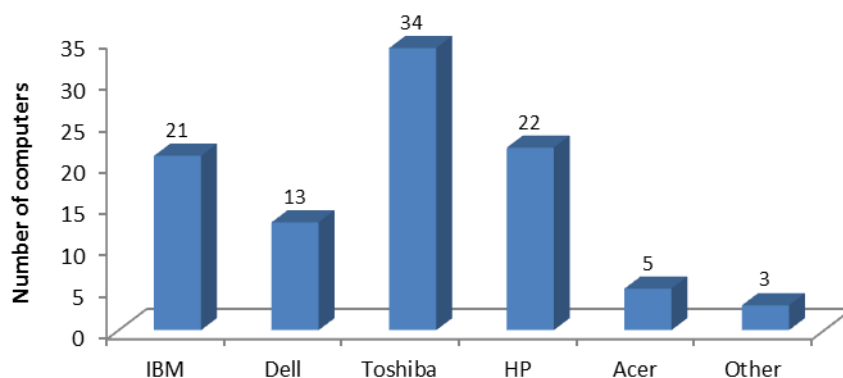


Share of recent graduates by marital status



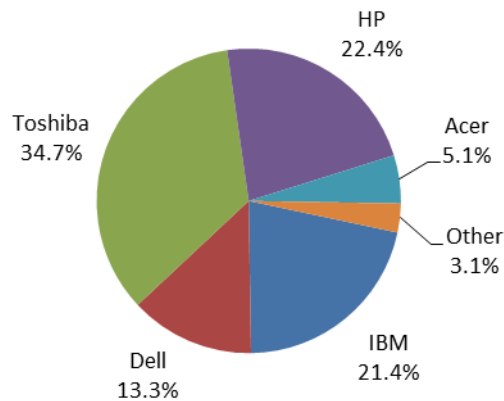
3.28 a A bar chart would be appropriate to depict the frequency distribution.

Frequency distribution of types of computers purchased



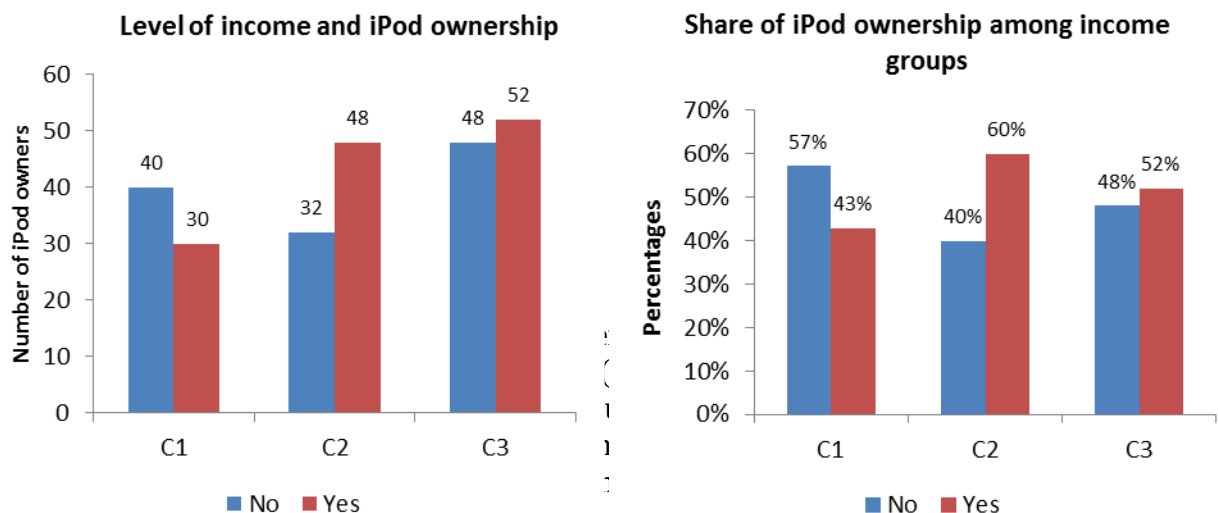
- b** A pie chart would be appropriate to depict the proportions.

Share of computers purchased by type of computer



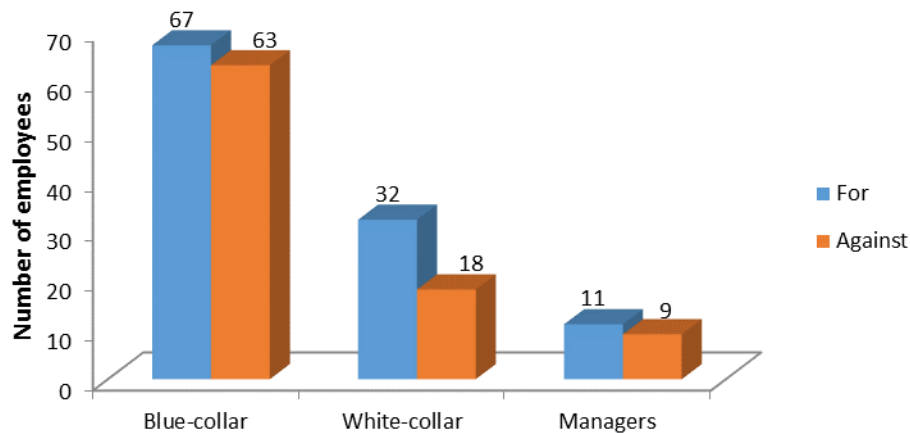
- c** Based on the sample data provided, the charts show that Toshiba is the most popular brand followed by HP and IBM. Dell, Acer and other brands are the least popular brands among university students when they make a computer purchase.

- 3.29** The frequency distributions are presented as bar charts in frequencies as well as proportions (in percentages). As income level increases, ownership of iPod also increases. However, the share of iPod owners and non-owners in the high income group (C3) is nearly equal, whereas in the middle income group (C2), the share of iPod owners is much higher.

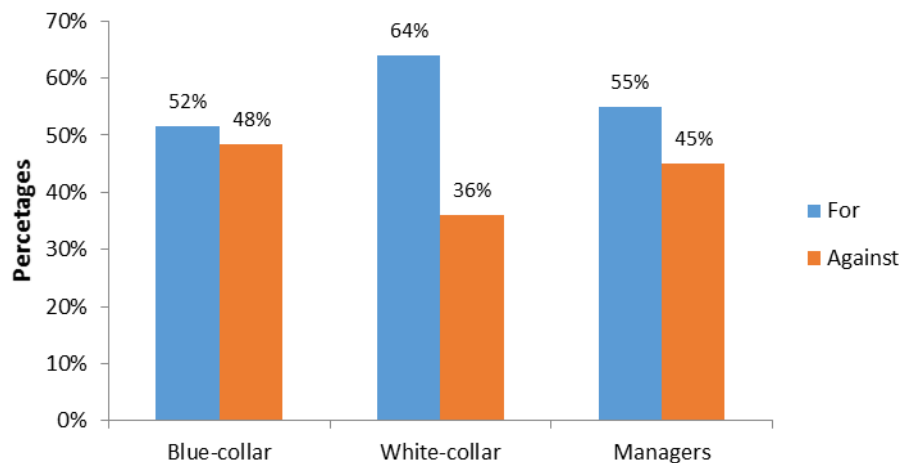


- 3.30** A pivot bar chart by type of worker would be more appropriate. Within each employee category, the proportion of 'For' response is greater than that of 'Against' response. However, the proportion of various types of workers 'For' the revision of the scheme is not similar. Similar result is shown for 'Against' as well. Therefore, the responses differ among the 3 groups.

Responses by Employee categories

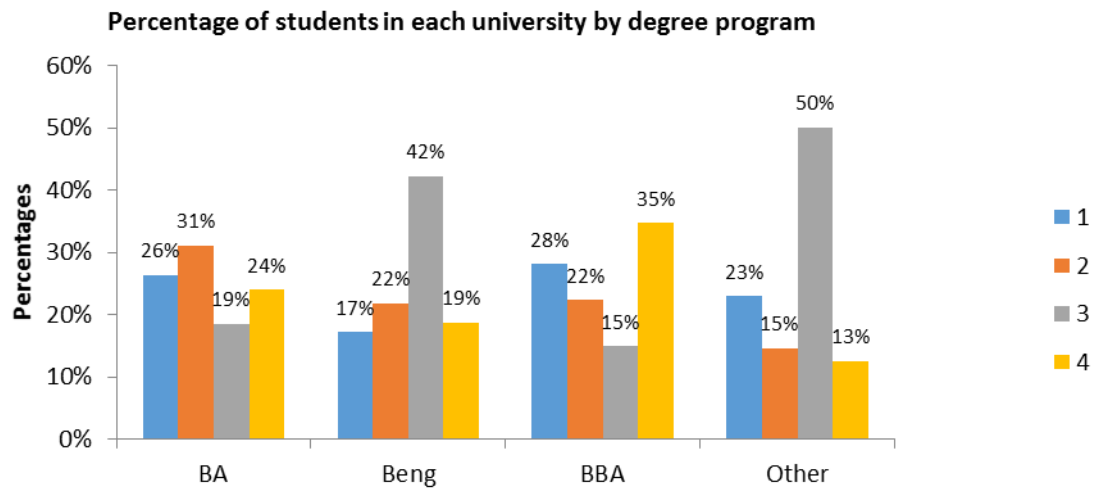


Share of Responses by Employee categories

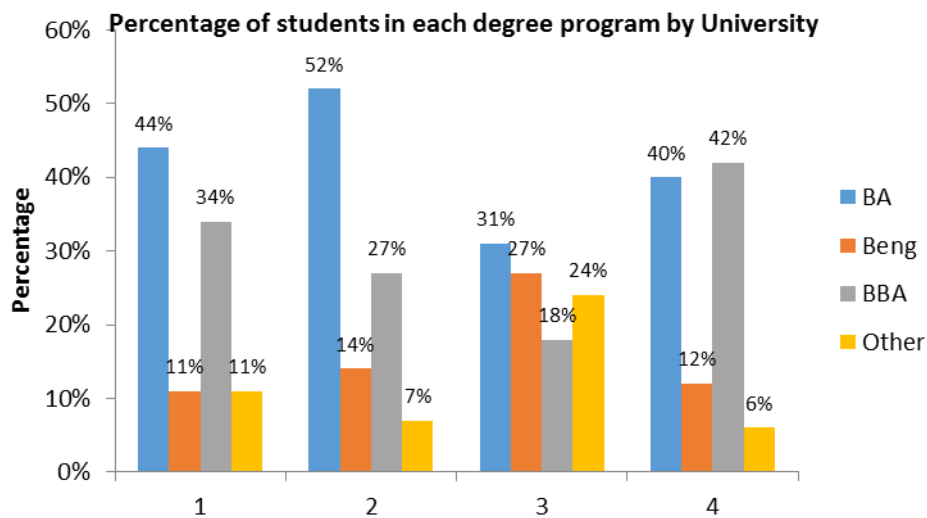


3.31 A pivot table and pivot chart would be appropriate to determine whether the undergraduate degree program and the university each person applied to are related. As can be seen, for each degree program, the proportion of students applying for a particular university is not similar. Similarly, among the universities, the distribution of students applying for a particular degree is not similar. Universities 1 and 2 are similar and quite dissimilar from universities 3 and 4, which also differ. The two nominal variables appear to be related.

Count of Student	University				
Degree	1	2	3	4	Grand Total
BA	26%	31%	19%	24%	100%
Beng	17%	22%	42%	19%	100%
BBA	28%	22%	15%	35%	100%
Other	23%	15%	50%	13%	100%
Grand Total	25%	25%	25%	25%	100%



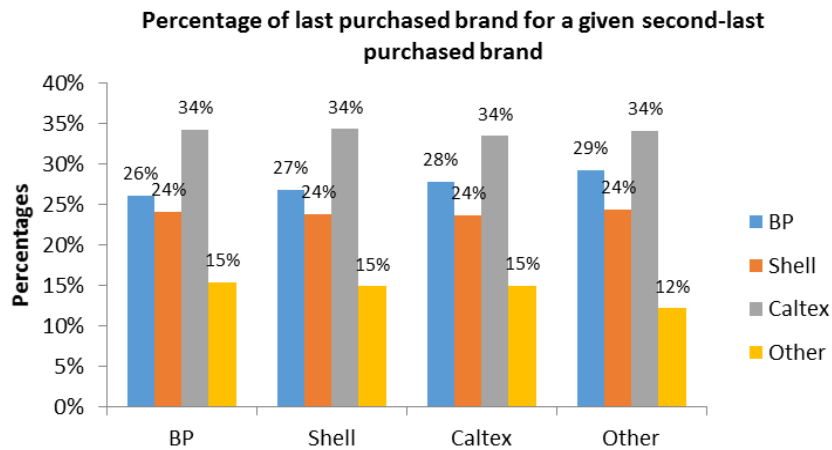
Count of Student	Degree				
University	BA	BEng	BBA	Other	Grand Total
1	44%	11%	34%	11%	100%
2	52%	14%	27%	7%	100%
3	31%	27%	18%	24%	100%
4	40%	12%	42%	6%	100%
Grand Total	42%	16%	30%	12%	100%



3.32 Constructing a pivot table (in percentages) and pivot chart would give the information required to compare and conclude whether there is brand loyalty among car owners.

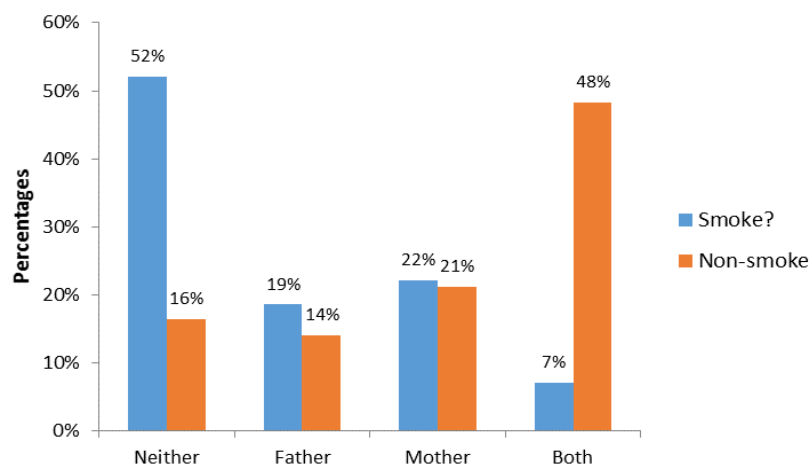
As can be seen, the column proportions are similar; the two nominal variables appear to be unrelated. Similar distributions can be seen from the pivot chart as well, confirming that there does not appear to be any brand loyalty.

Count of Owner	Last				
Second-last	BP	Shell	Caltex	Other	Grand Total
BP	26%	24%	34%	15%	100%
Shell	27%	24%	34%	15%	100%
Caltex	28%	24%	34%	15%	100%
Other	29%	24%	34%	12%	100%
Grand Total	27%	24%	34%	15%	100%



3.33

Count of ID	Smoke?		
Parent	Smoker	Non-smoker	Grand Total
Neither	52%	16%	39%
Father	19%	14%	17%
Mother	22%	21%	22%
Both	7%	48%	23%
Grand Total	100%	100%	100%

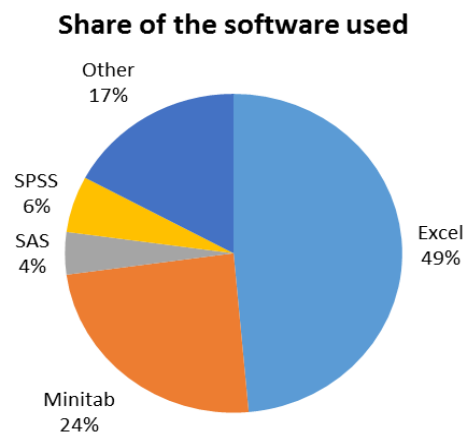


The two variables are related.

3.34 a Use Excel to count the number of courses using each software.

Software	Frequency
Excel	34
Minitab	17
SAS	3
SPSS	4
Other	12

b A pie chart would be appropriate to depict the proportions.



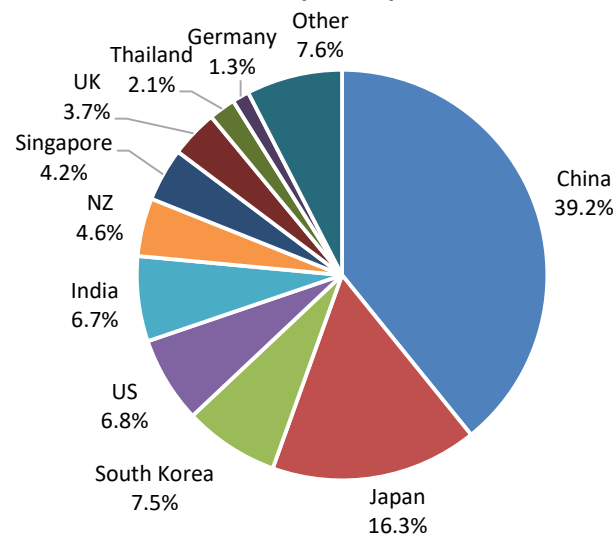
c Excel is the choice of about half the sample, one-quarter have opted for Minitab, and a small fraction chose SAS and SPSS.

3.35 a Its appropriate to depict the amount of exports in a bar chart.



b The export market shares can be depicted using a pie chart.

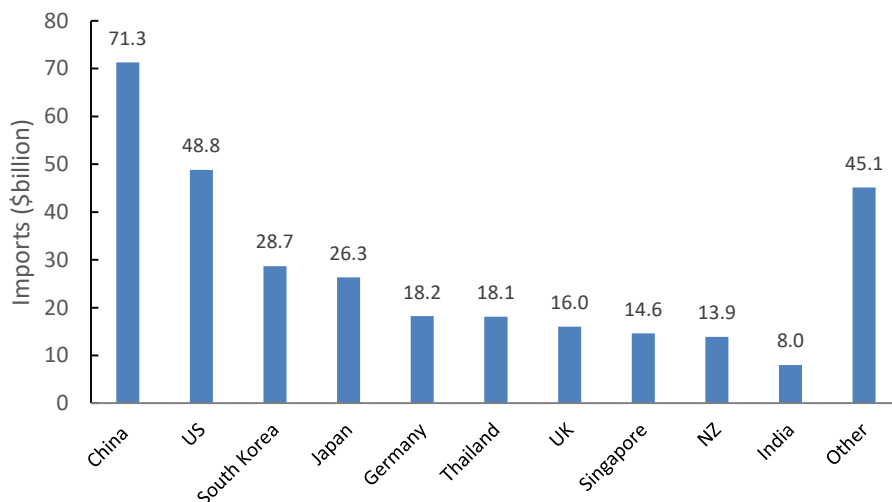
Shares of Australian top 10 export markets, 2017-18



c As can be seen, among the top 10 export countries, China is the largest export market (39.2%), followed by Japan (16.3%), South Korea (7.5%), the US (6.8%) and India (6.7%).

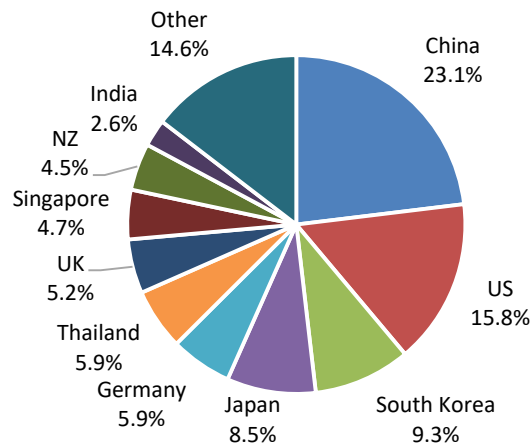
3.36 a Its appropriate to depict the amount of imports in a bar chart.

Australian imports, Top 10 markets, 2017-18



b The import market shares can be depicted using a pie chart.

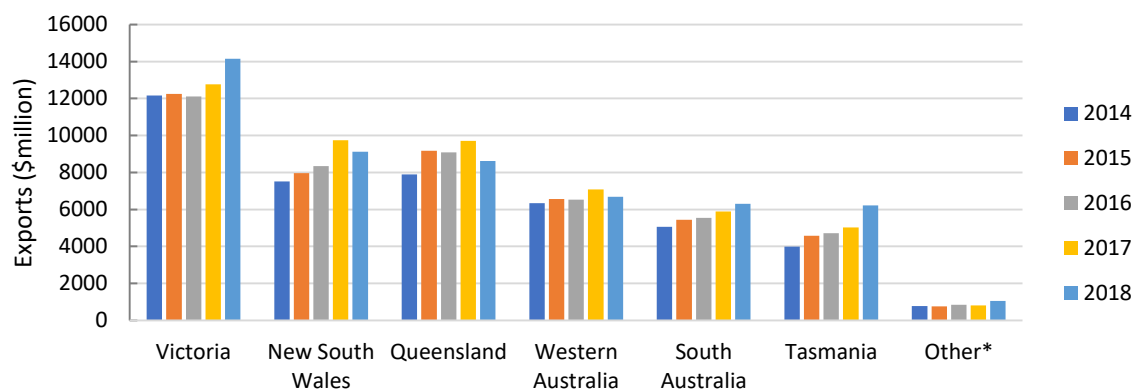
Shares of Australian top 10 import markets, 2017-18



c As can be seen, among the top 10 import markets, China is the largest (23.1%) followed by the US (15.8%), South Korea (9.3%) and Japan (8.5%).

3.37 A combined bar chart by state would be more appropriate. Australian food and fibre exports have generally been on the increase in all states and territories (Other* denotes the Australian territories) between 2014 and 2018.

Australian food and fibre exports by state, 2014-2018

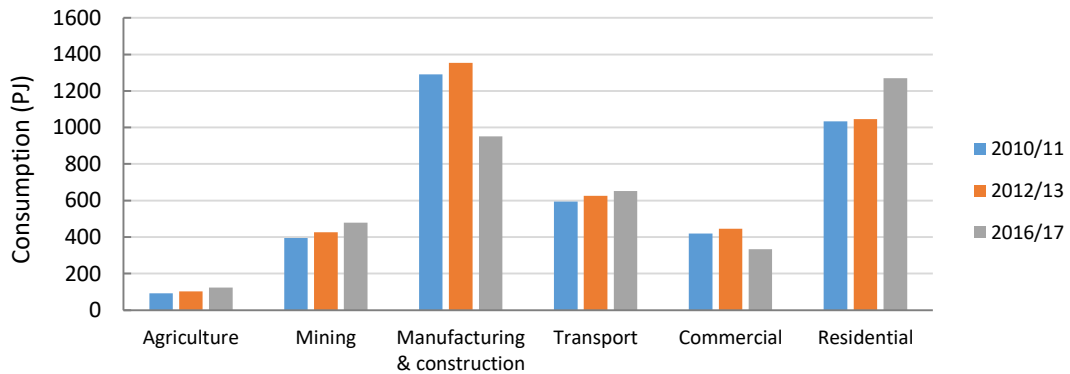


3.38 A combined bar chart by sector would be appropriate to compare the trend in energy consumption over the three time periods by sector. As can be seen, energy consumption is highest in the manufacturing and construction sector followed by residential, transport, commercial and mining. The agriculture sector consumes the least. As can be seen, energy consumption has continued increase from 2010/11 to 2016/17 for the agriculture, mining, transport and residential sectors. However, energy consumption in the manufacturing and construction, and commercial sectors have increased between 2010/11 and 2012/13, but has declined between 2012/13 and 2016/17.

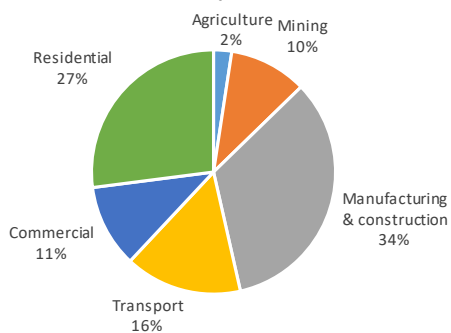
Pie charts for the two years (2010/11 and 2016/2017) would provide information regarding how the shares of energy consumption of each sector has changed over the six-

year period. During this period, manufacturing and construction and commercial sectors' energy consumption shares have decreased, while the shares of residential, mining, transport and agriculture have increased.

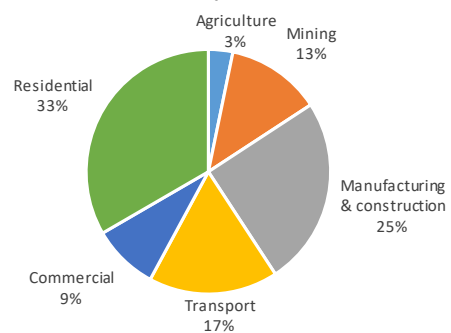
Energy consumption by sector, Australia, 2010/11-2016/17



Energy consumption share by sector, Australia, 2010/11

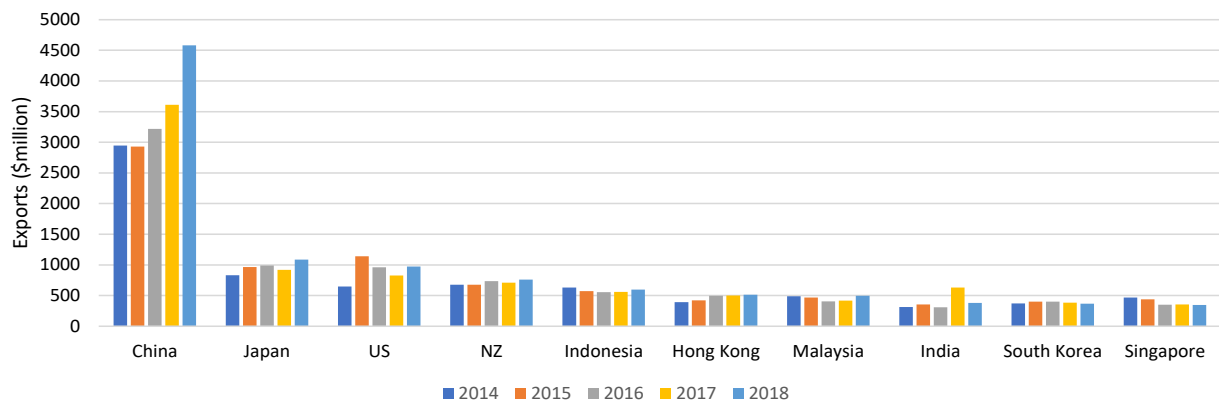


Energy consumption share by sector, Australia, 2016/17



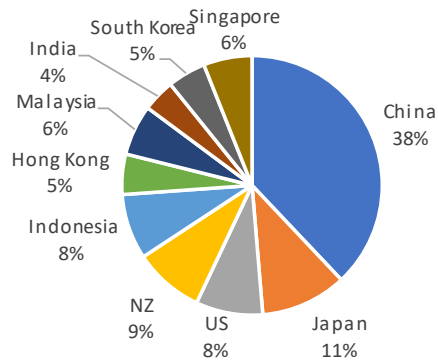
3.39 A combined bar chart by destination would be appropriate. Pie charts for the five years would also provide information regarding how the share of exports to each destination has changed from 2014 to 2018.

Top 10 Victorian food and fibre exports by destination, 2014-2018

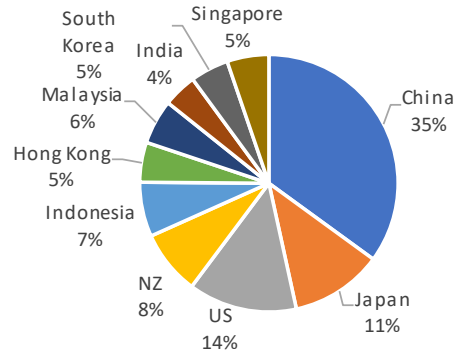


Food and fibre exports from Victoria has been generally increasing during 2014-2018. China is Victoria's major food and fibre export destination followed by Japan, the US and New Zealand.

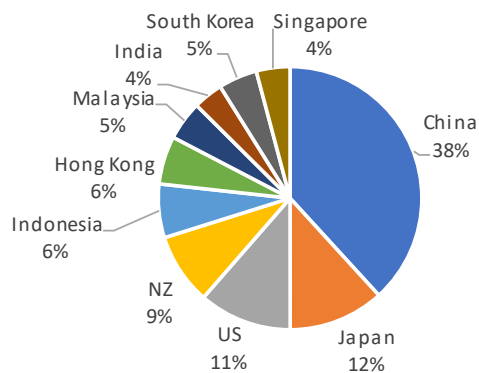
Top 10 Victorian food and fibre exports by destination, 2014



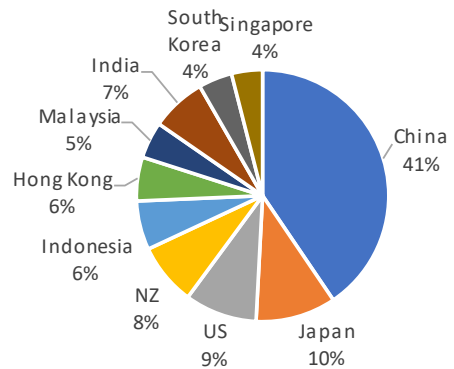
Top 10 Victorian food and fibre exports by destination, 2015



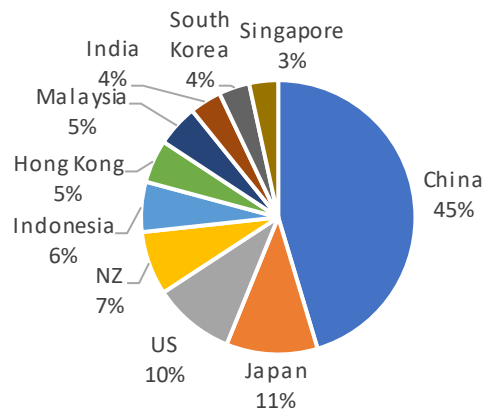
Top 10 Victorian food and fibre exports by destination, 2016



Top 10 Victorian food and fibre exports by destination, 2017



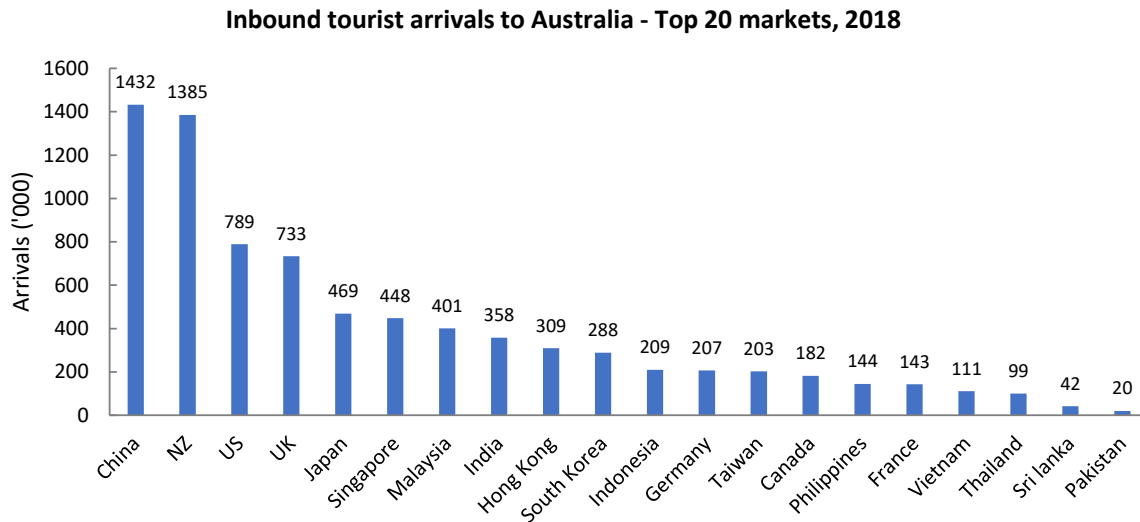
Top 10 Victorian food and fibre exports by destination, 2018



As can be seen from the five pie charts, food and fibre export share to China has

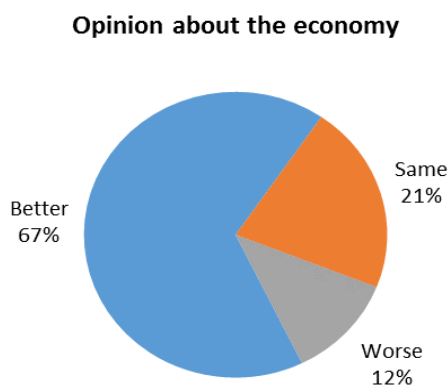
increased from 38% 2014 to 45% in 2018. Share of export to Japan and other countries have changed only slightly over the years.

3.40 a A bar chart would be appropriate to compare the arrivals from the top 20 markets.



b The graph allows us to easily gauge visually the level of differences in the overseas tourist arrivals from the top 10 countries. China and New Zealand are the major contributors to Australian tourism followed by the US and the UK.

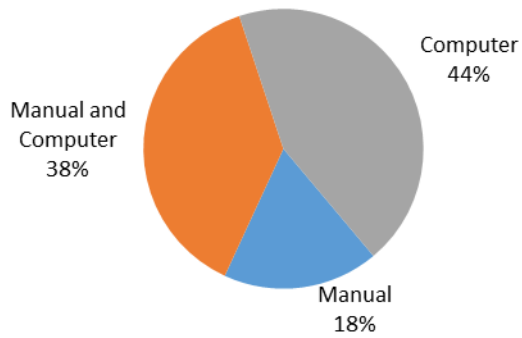
3.41 As the data are nominal survey data, we obtain the frequency for each category and consider the proportions that fall in each category, which can be depicted using a pie chart.



As can be seen, more than two-thirds of those surveyed said that the economy will be better next year, and only 12% said that the economy will be worse.

- 3.42** As the data are nominal survey data, we obtain the frequency for each category and consider the proportions that fall in each category. A pie chart would be more appropriate.

Applied statistics: Teaching approaches



4 Graphical descriptive techniques – Numerical data

Throughout this chapter,

- In all frequency distributions, histograms and ogive, the class intervals contain observations greater than their lower limits (except for the first class) and less than or equal to their upper limits.
- In all histograms and ogive output using Excel, the upper limits of the class intervals are printed in the centre of the classes.

- 4.1 Using Sturge's formula for $n = 125$, the approximate number of classes for a histogram would be

$$K = 1 + 3.3 \log_{10} 125 = 7.92 \approx 7 \text{ or } 8 \text{ classes}$$

- 4.2 Using Sturge's formula for $n = 1500$, the approximate number of classes for a histogram would be

$$K = 1 + 3.3 \log_{10} 1500 = 11.48 \approx 11 \text{ or } 12 \text{ classes}$$

- 4.3 a Using Sturge's formula for $n = 300$, the approximate number of classes for a histogram would be $K = 1 + 3.3 \log_{10} 300 = 9.17 \approx 9 \text{ or } 10 \text{ classes}$.

- b Approximate class width using $K = 10$ classes can be calculated as

$$d = \frac{(\text{largest} - \text{smallest})}{K} = \frac{(239 - 147)}{10} = 9.2 \approx 10$$

Therefore, the 10 class intervals would be

$$\begin{array}{llll} 140 \leq X \leq 150, & 150 < X \leq 160, & 160 < X \leq 170, & 170 < X \leq 180, \\ 180 < X \leq 190, & 190 < X \leq 200, & 200 < X \leq 210, & 210 < X \leq 220, \\ 220 < X \leq 230, & 230 < X \leq 240 & & \end{array}$$

- 4.4 a Using Sturge's formula for $n = 40$, the approximate number of classes for a histogram would be $K = 1 + 3.3 \log_{10} 40 = 6.29 \approx 6 \text{ or } 7 \text{ classes}$.

- b Approximate class width using $K = 6$ classes can be calculated as

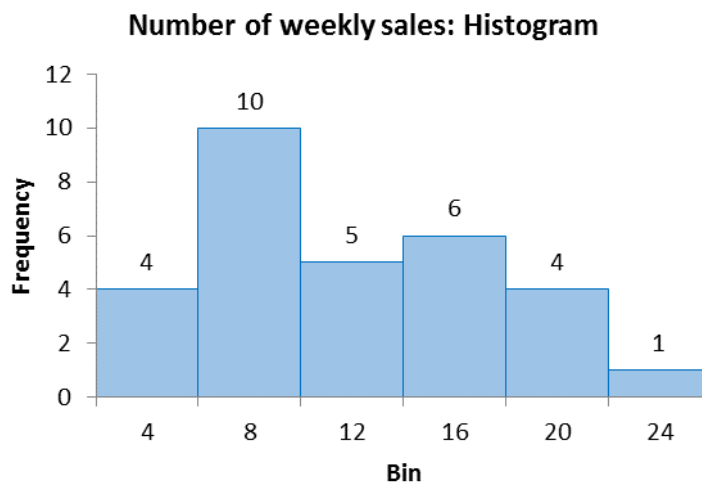
$$d = \frac{(\text{largest} - \text{smallest})}{K} = \frac{(6.1 - 5.2)}{6} = 0.2$$

Therefore, the 6 class intervals would be

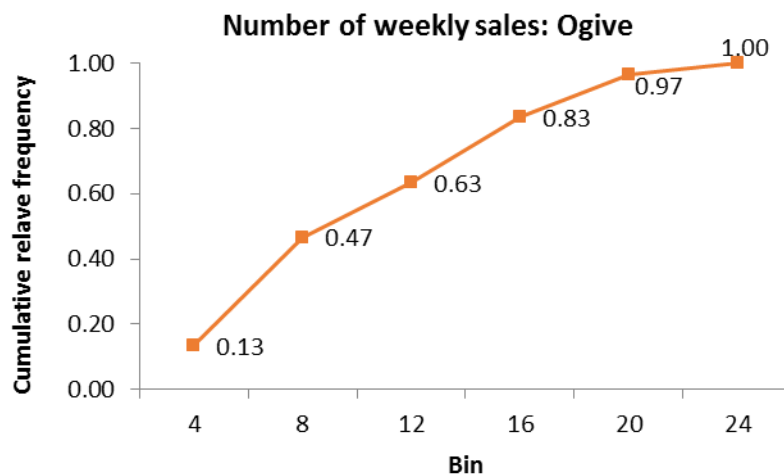
$$\begin{array}{lll} 5.0 \leq X \leq 5.2, & 5.2 < X \leq 5.4, & 5.4 < X \leq 5.6 \\ 5.6 < X \leq 5.8, & 5.8 < X \leq 6.0, & 6.0 < X \leq 6.2 \end{array}$$

4.5 a

Class interval	Tally	Frequency	Cumulative Frequency	Relative cumulative frequency
0 up to 4		4	4	0.13
4 up to 8		10	14	0.47
8 up to 12		5	19	0.63
12 up to 16		6	25	0.83
16 up to 20		4	29	0.97
20 up to 24		1	30	1.00



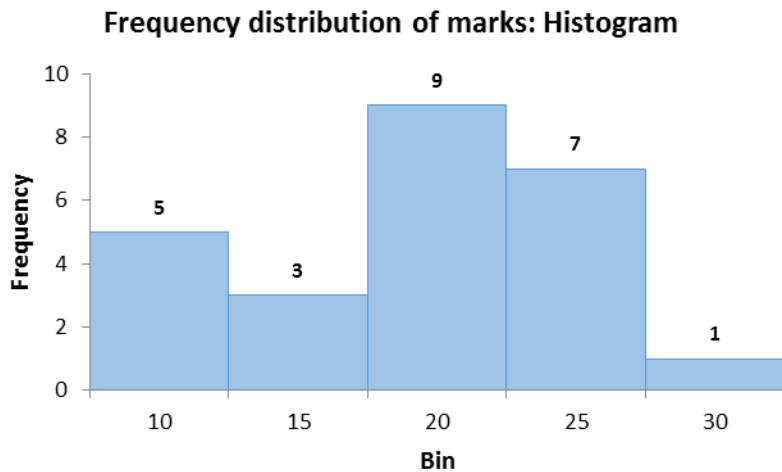
b



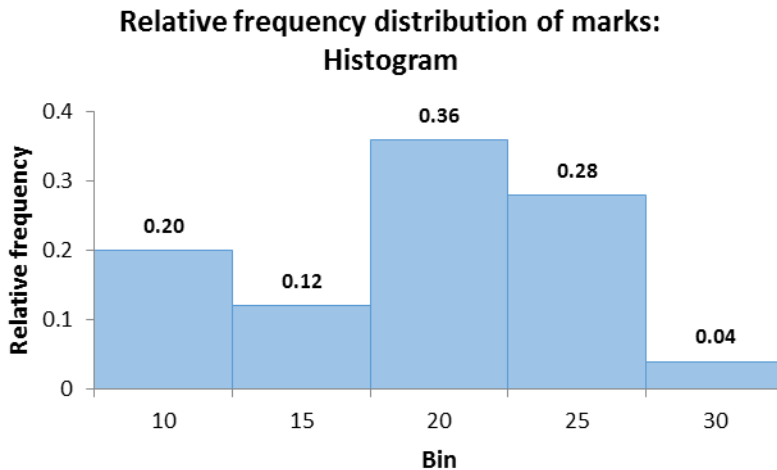
c The distribution is slightly skewed to the right.

4.6 a Frequency distribution table

Class interval	Bin	Tally	Frequency (f)	Relative frequency
$5 \leq X \leq 10$	10		5	0.20
$10 < X \leq 15$	15		3	0.12
$15 < X \leq 20$	20		9	0.36
$20 < X \leq 25$	25		7	0.28
$25 < X \leq 30$	30		1	0.04



b



- c** The area of each rectangular strip is proportional to the relative frequency of that class. As the class widths are all the same the height is represented by the relative frequency of that class.

- 4.7 Using Sturge's formula for $n = 28$, the approximate number of classes for a histogram would be $K = 1 + 3.3 \log_{10} 28 = 5.78 \approx 5$ or 6 classes.

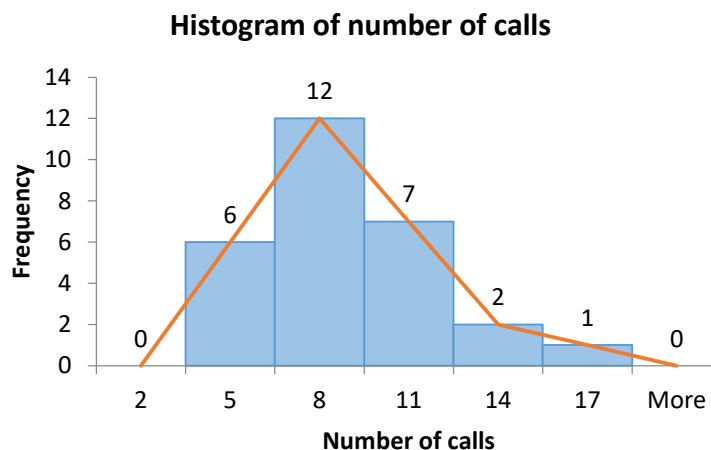
Largest = 17, Smallest = 3.

Approximate class width using $K = 5$ classes can be calculated as

$$d = \frac{(\text{largest} - \text{smallest})}{K} = \frac{(17 - 3)}{5} = 2.8 \approx 3$$

Therefore, the frequency distribution table is

Class interval	Bin	Tally	Frequency (f)	Relative frequency
$2 \leq X \leq 5$	5		6	0.21
$5 < X \leq 8$	8		12	0.43
$8 < X \leq 11$	11		7	0.25
$11 < X \leq 14$	14		2	0.07
$14 < X \leq 17$	17		1	0.04
			28	1.00



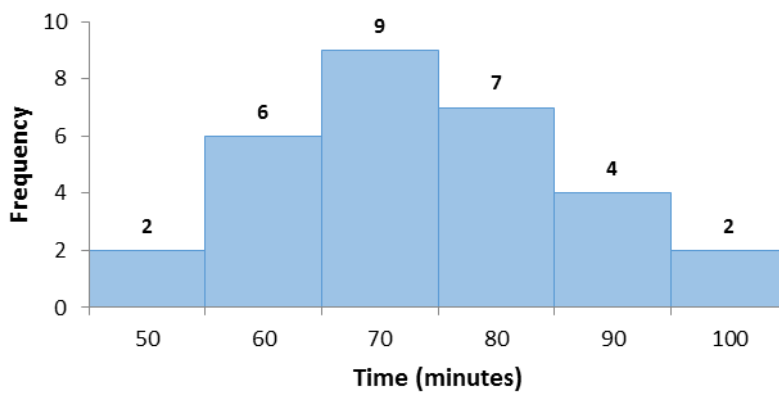
4.8 a

Stem	Leaf
4	58
5	245889
6	11245667
7	0357789
8	02366
9	14

b

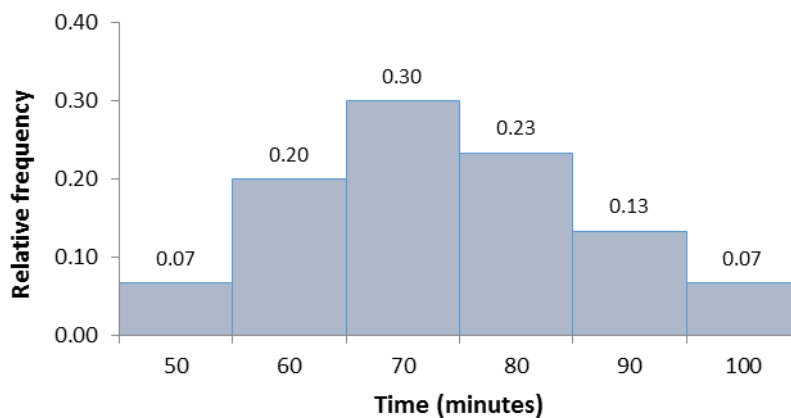
Class interval	Bin	Tally	Frequency	Relative Frequency
$40 \leq X \leq 50$	50		2	0.07
$50 < X \leq 60$	60		6	0.20
$60 < X \leq 70$	70		9	0.30
$70 < X \leq 80$	80		7	0.23
$80 < X \leq 90$	90		4	0.13
$90 < X \leq 100$	100		2	0.07
Total			30	1.00

Frequency distribution of Time: Histogram



c

Relative frequency distribution of Time: Histogram



d The stem and leaf display and the histograms show a near symmetrical distribution of the time taken to complete the quiz.