SOLUTIONS MANUAL

Second Edition

Analyzing Data and Making Decisions
Statistics for Business
Judith Skuce

Save Time, Improve Results. www.mathxl.com

Chapter 2 Solutions
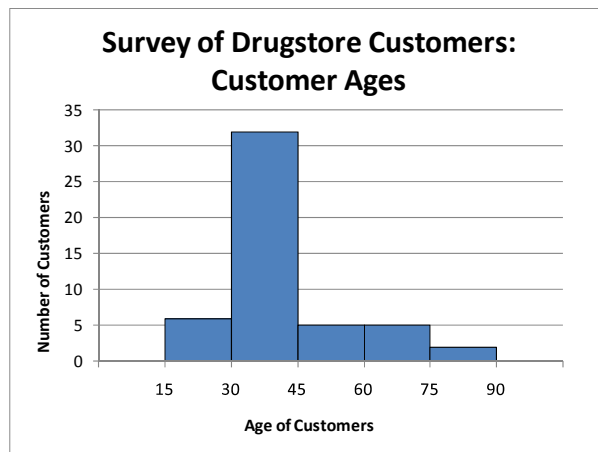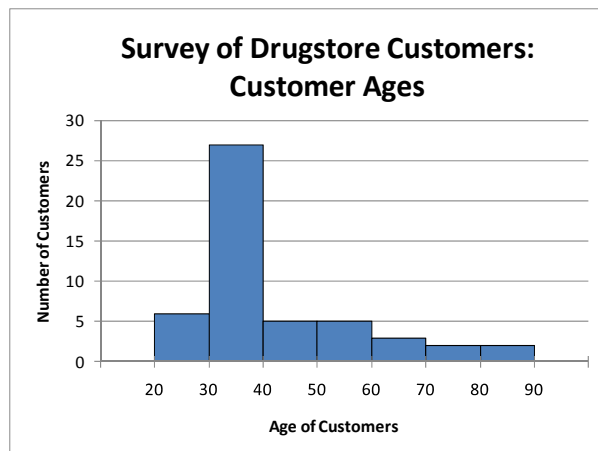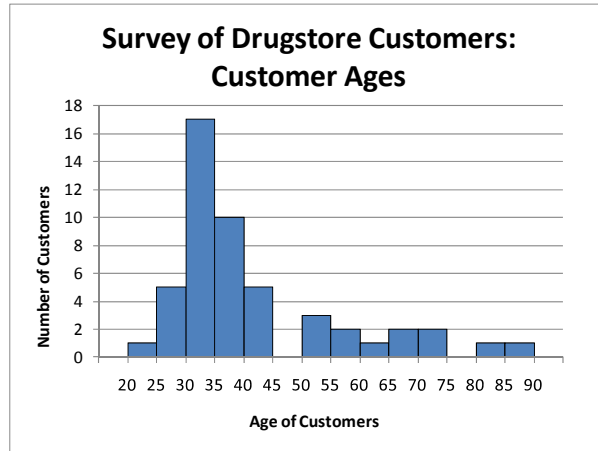
Develop Your Skills 2.1

1. The number of dented cans is a count of qualitative data (think of it this way—the original data might be recorded as "yes" or "no" to the question: is the can dented?). These are also time-series data, as they are collected over successive periods of time. The data are discrete.

2. Stock price data are quantitative data. These are also time-series data, as they were collected over three years. Prices are treated as continuous data.

3. The employees' final average grades from college are continuous quantitative data. The scores assigned by the supervisors are ranked data.

   Although the questionnaire will help, such a ranking is somewhat subjective. If different supervisors assign the ranks, they may not be comparable.

4. The price data are continuous quantitative data. They are also organized according to qualitative data on the size of the coffee. These are cross-sectional data, as they would be collected at around the same period in time.

5. Postal codes are qualitative data.

Develop Your Skills 2.2

6. The three different histograms are shown below.

Survey of Drugstore Customers: Customer Ages

(Histogram 1: x-axis "Age of Customers" from 20 to 90 in increments of 5; y-axis "Number of Customers" from 0 to 18)

Survey of Drugstore Customers: Customer Ages

(Histogram 2: x-axis "Age of Customers" from 20 to 90 in increments of 10; y-axis "Number of Customers" from 0 to 30)

Survey of Drugstore Customers: Customer Ages

(Histogram 3: x-axis "Age of Customers" from 15 to 90 in increments of 15; y-axis "Number of Customers" from 0 to 35)

All three histograms clearly show that the distribution of customer ages is skewed to the right, that is, while most customers are under 40 years old, there are some customers who are much older, in fact as old as 85. [Note that when you are describing the distribution, it is not sufficient to stop at "skewed to the right"—you should explain what this means, in the context of this particular data set.]

A class width of 5 is not a good choice for this data set. There are too many classes, many of which have only a very few data points. A class width of 10 or 15 would be a better choice.

7. A frequency distribution and histogram are shown below.

Survey of Drugstore Customers	
Customer Income	Number of Customers
$30,000 to <$35,000	2
$35,000 to <$40,000	9
$40,000 to <$45,000	14
$45,000 to <$50,000	7
$50,000 to <$55,000	5
$55,000 to <$60,000	6
$60,000 to <$65,000	3
$65,000 to <$70,000	4
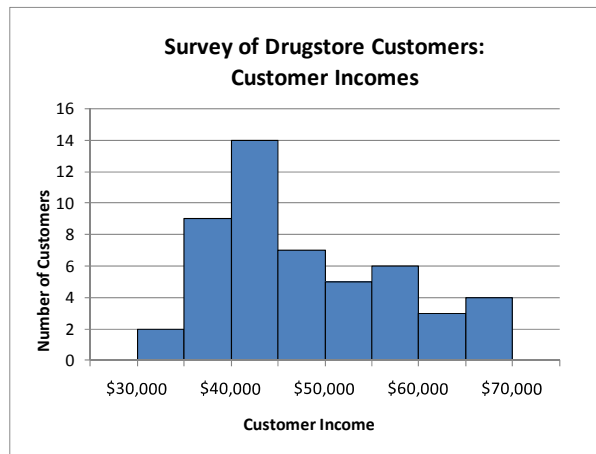
Choosing class widths is a bit tricky in this instance. The class width template suggests class widths of 5155, 8905, or 8170. None of these numbers is that comfortable for incomes. Class widths of $5,000 and $10,000 were considered. A class width of $10,000 was discarded because it would have resulted in only four classes (five is a good minimum number of classes).

The distribution of customer incomes in the drugstore survey is skewed to the right. Most customer incomes are in the $35,000 to $50,000, but there are a number of customers with higher incomes, the highest being $68,800.

8. The stem and leaf display is shown below.

1	9	7	4	8	2	7					
2	4	6	3	9	4	9	7	6	5	9	4	6
3	1	1	4	1	6						
4	0	1									

The lowest daily customer count in the random sample from Downtown Automotive is 12, and the largest is 41. Most days the shop deals with 20-some customers. It is unusual for the shop to deal with more than 35 customers.

9.
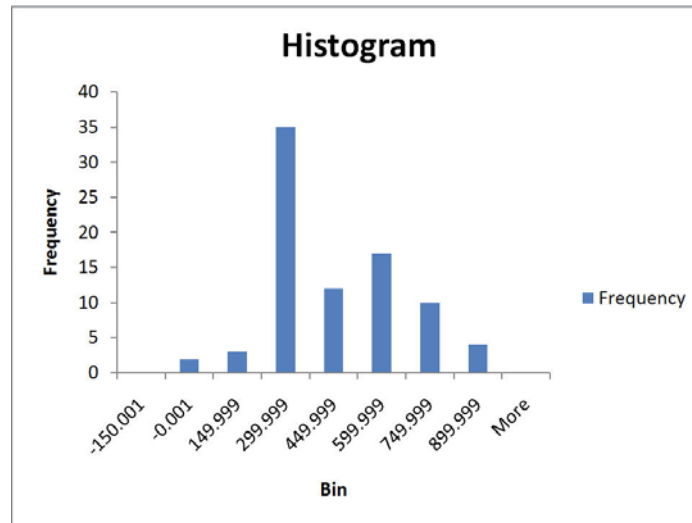
This histogram totally fails at its job of summarizing the accompanying data set.
1. The graph does not have meaningful titles or labels. It completely fails to communicate what it's about.
2. There are gaps between the bars, which there should not be.
3. It appears that the creator of this graph used bin numbers correctly, but s/he forgot to round them for presentation. The graph should show lower class limits along the x-axis in the proper location, that is, aligned under the left-hand side of each bar.

10.

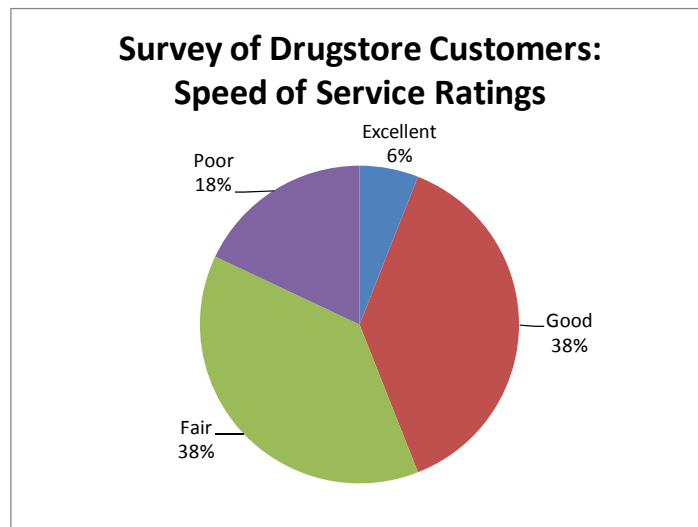**Ms. Nice's Marks**
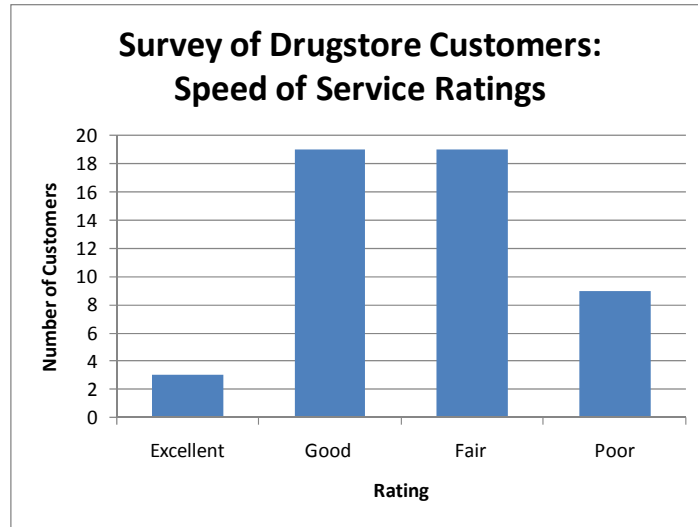
Mr. Mean's Marks
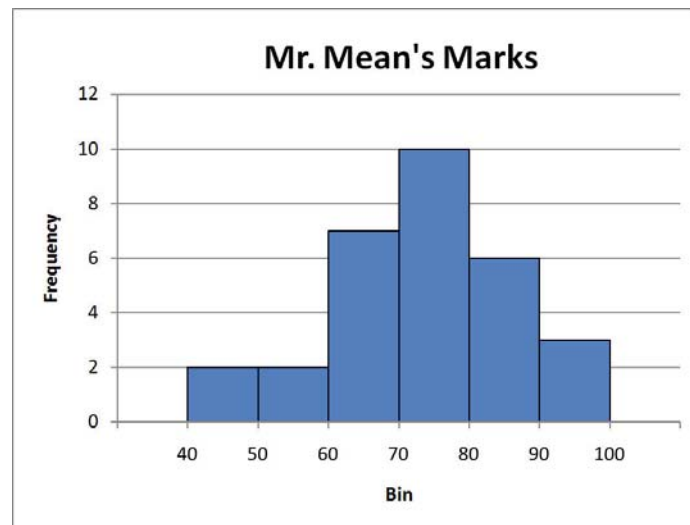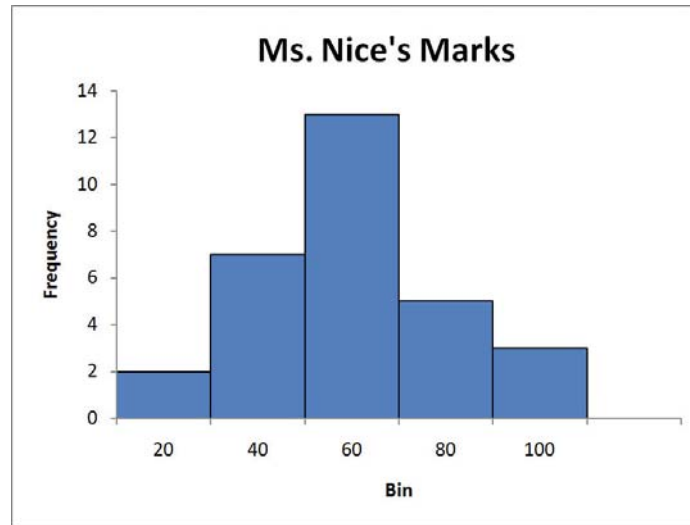
These graphs have not been properly set up, so Patty probably deserved her low mark in Statistics. The titles are not correct. They are not Ms. Nice's marks, they are marks from a random sample of students in Ms. Nice's statistics class, and the title should tell us that. The title for the marks from Mr. Mean's class should be similarly adjusted. In both graphs, the label on the x-axis should say something like "Final Grade". The label on the y-axis should say something like "Number of Students".

In the graph of the marks from Ms. Nice's class, the labels are at the centre of each class, and should be adjusted. Horizontal grid lines would also help the reader.

The graphs should be set up properly for comparison, with the same classes and scales on each axis. A quick glance at these graphs might lead you to think that the marks are lower in Mr. Mean's class, but in fact, the opposite is the case.

Develop Your Skills 2.3

11. Either a bar graph or a pie chart would be appropriate.

Survey of Drugstore Customers:
Speed of Service Ratings

(bar graph: Number of Customers vs. Rating — Excellent 3, Good 19, Fair 19, Poor 9)

Survey of Drugstore Customers:
Speed of Service Ratings

(pie chart: Excellent 6%, Good 38%, Fair 38%, Poor 18%)

The graphs indicate that over a third of customers (38%) rated the speed of service as good, but only 6% rated it as excellent. Over a third of customers (again, 38%) rated the speed of service as only fair, while 18% rated it as poor. These ratings indicate that there may be some room for improvement in speed of service at the drugstore.

12. The most effective graph would be a bar chart, showing actual and desired relative frequencies for each colour. First, create a table for the data, and then create the bar chart.

Actual and Observed Colours in Candy				
	Red	Green	Blue	Yellow
Observed sample values	305	265	201	96
Desired Percentage	40%	30%	20%	10%
Sample Percentage	35.179%	30.565%	23.183%	11.073%

Candy Colours, Random Sample After Reorganization of Production Process
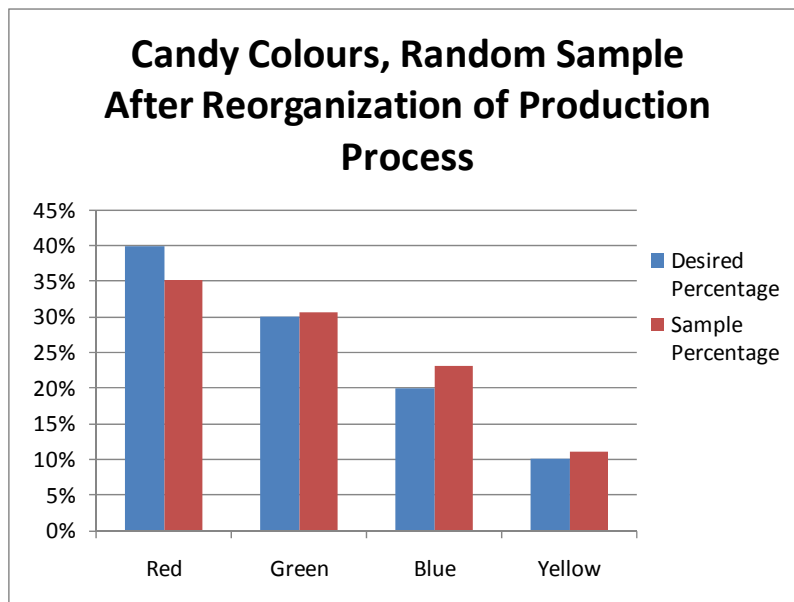
This graph makes it easy to see that the most important differences in the candy colour distribution are in the red candies (fewer than desired) and in the blue candies (more than desired).

13. Since we want to compare the number of defects by shift, it is appropriate to compare the categories for the number of defects across the horizontal axis. Since each shift produced a different number of items[1], it makes sense to use relative frequencies. For each shift, calculate the percentage of items with no defects, one minor defect, more than one minor defect, and then create a bar graph.
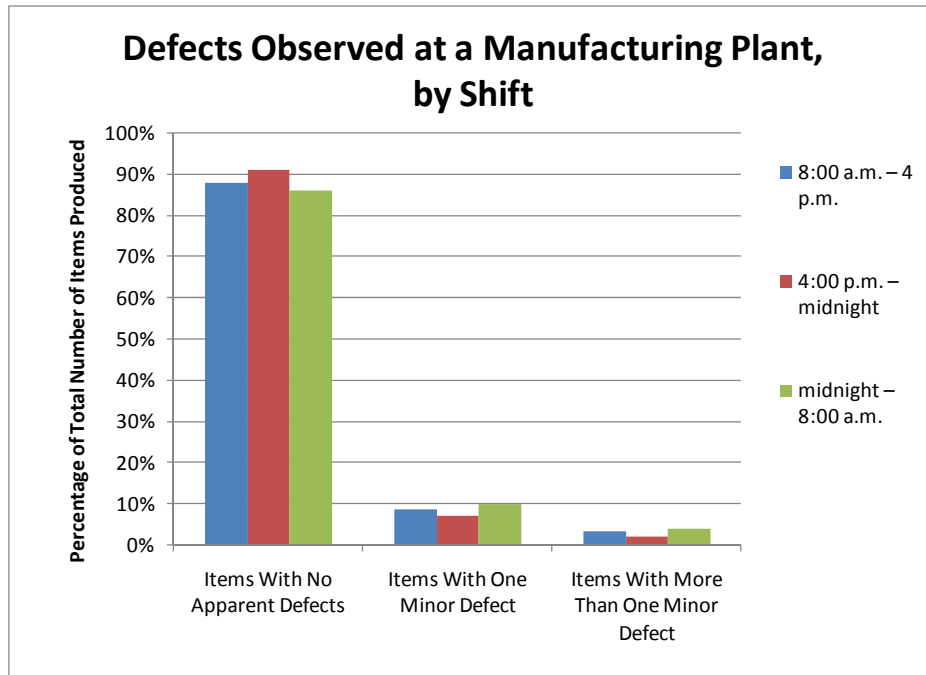
Across all three shifts, the percentage of items produced with more than one minor defect is small. For all three shifts, by far the greatest percentage of items produced has no apparent defects. The midnight-8:00 a.m. shift has the greatest percentage of defects, and the 4:00 p.m. – midnight shift has the lowest percentage of defects.
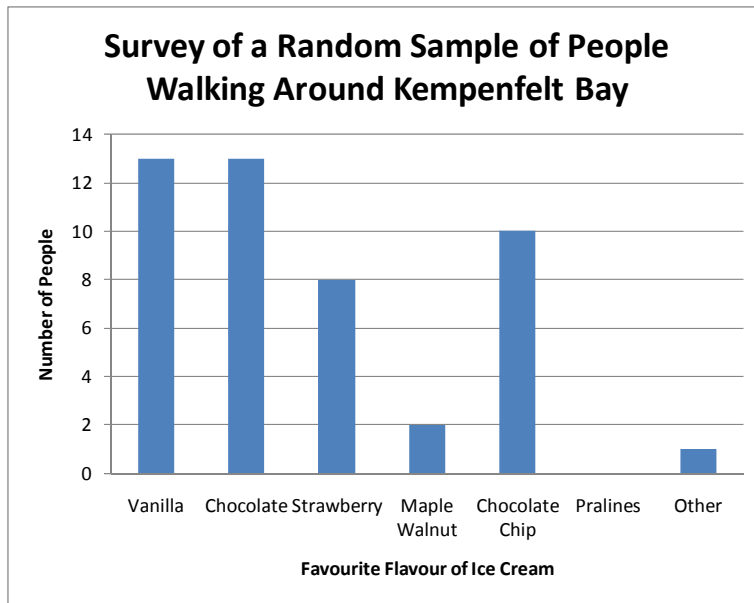
---

[1] This too is interesting, and the fact should be included in any accompanying report.

14.

Financial Services Company Customer Survey: "The staff at my local branch can provide me with good advice on my financial affairs."
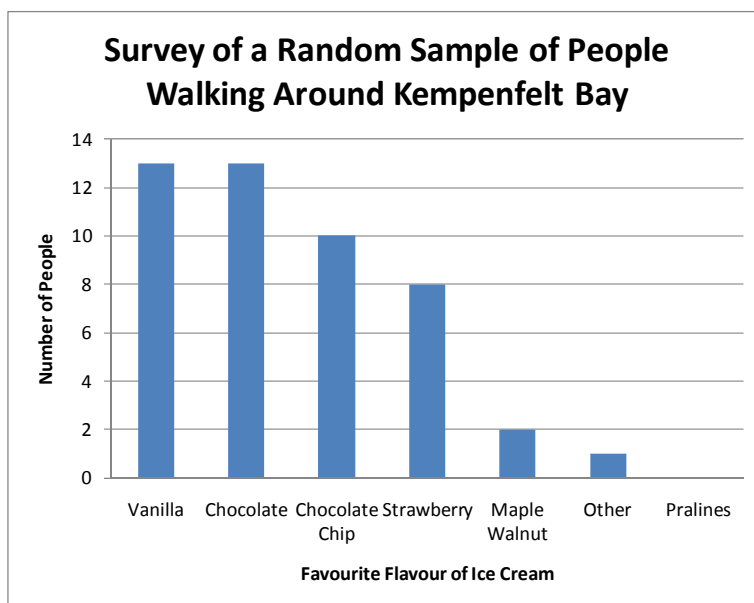

The majority of the customers surveyed agree with the statement that staff at the local branch can provide good advice on financial affairs. A significant number of the customers neither agreed nor disagreed with the statement, and it might be worthwhile to investigate why these customers appeared to have no opinion (was it lack of knowledge?). There were customers who disagreed or strongly disagreed. It might be worthwhile to investigate further (why was this the case? Were these customers disappointed in past advice, or do they just have an impression that local staff cannot provide good advice?).
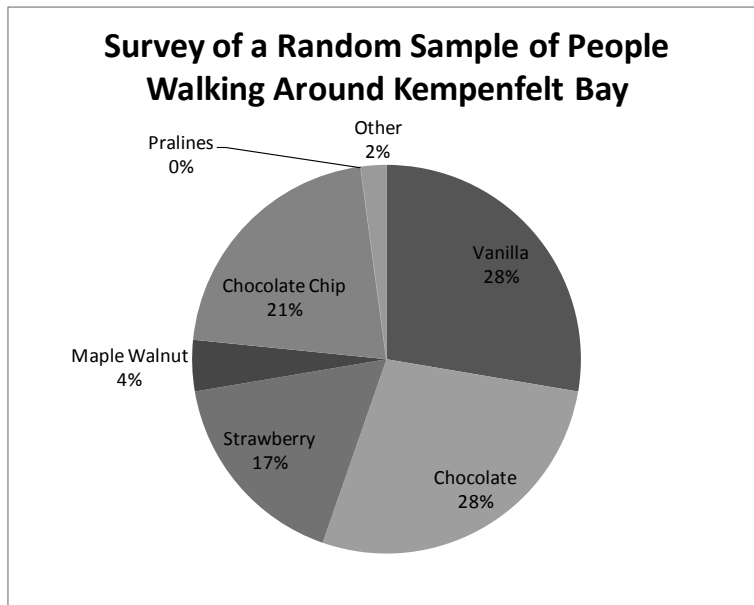
15.

Survey of a Random Sample of People Walking Around Kempenfelt Bay

Y-axis: Number of People (0 to 14)
X-axis: Favourite Flavour of Ice Cream

Vanilla: 13, Chocolate: 13, Strawberry: 8, Maple Walnut: 2, Chocolate Chip: 10, Pralines: 0, Other: 1

Vanilla and chocolate were tied as the most frequently-mentioned favourite flavours of ice cream among the people surveyed at Kempenfelt Bay, followed by chocolate chip. The fourth most popular flavour was strawberry. Only a few people cited maple walnut as their favourite flavour, and no one called pralines their favourite. One person had a favourite flavour other than the ones cited specifically in the survey.

Of course, there are other options for this graphical display. It might be helpful to arrange the categories from most preferred to least-preferred. As well, a pie graph is an option.
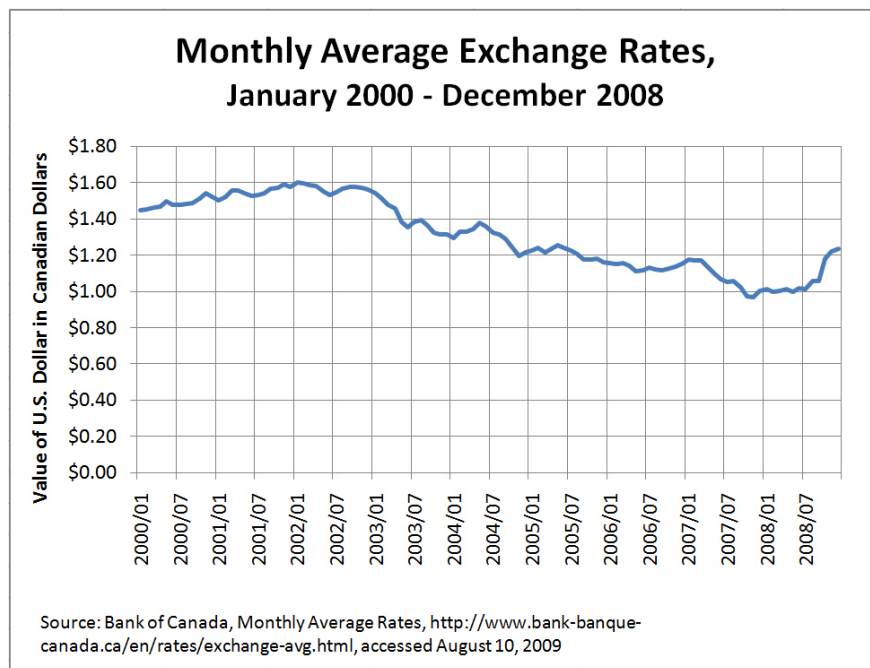
Survey of a Random Sample of People Walking Around Kempenfelt Bay

Y-axis: Number of People (0 to 14)
X-axis: Favourite Flavour of Ice Cream

Vanilla: 13, Chocolate: 13, Chocolate Chip: 10, Strawberry: 8, Maple Walnut: 2, Other: 1, Pralines: 0

Survey of a Random Sample of People Walking Around Kempenfelt Bay

- Other 2%
- Pralines 0%
- Vanilla 28%
- Chocolate Chip 21%
- Maple Walnut 4%
- Strawberry 17%
- Chocolate 28%

Develop Your Skills 2.4

16. Automobile sales are seasonal and cyclical, although this may not be as much the case as it once was. Sales tend to be higher when new models become available, and generally, auto sales are lower at year-end, when many people are focused on the holiday season. For these reasons, monthly sales data would be appropriate. Annual data would hide the month-to-month variations in sales.

17. Whatever your data source (as long as the data are accurate) you should see that over this period, the price of $1US in Canadian dollars was on an increasing trend from January 2000 until the beginning of 2002, with the highest exchange value of 1.599618 (monthly average) in January of 2002. From then until near the end of 2007, the exchange value of the US dollar in terms of Canadian dollars was on a declining trend, reaching 0.968 in November of 2007. The rate then stabilized around par, beginning to increase in the latter part of 2008, ending with a monthly average rate of 1.2343619 in December.

    The graph below shows the trends.

Monthly Average Exchange Rates,
January 2000 - December 2008

Source: Bank of Canada, Monthly Average Rates, http://www.bank-banque-canada.ca/en/rates/exchange-avg.html, accessed August 10, 2009

18. The Bank of Canada Bank Rate was at 4.5% in January 2007, and stayed there until July of 2007, when it increased to 4.75%. The rate declined to 4.5% in December of 2007, and continued to decline to 3.25% in April of 2008. The rate held steady at 3.25% until October of 2008, when it declined to 2.5%, further falling to 1.75% in December of 2008.

    Usually a line graph would be used for such a long time series. However, in this case, the movements in the Bank Rate are infrequent, and small, so the graph is not too cluttered. The advantage in using a bar chart is that it highlights the change in rates from one period to the next.

19. Your commentary should describe the data for the company you chose. Here is a checklist to help you:
    - Be sure to note the start and end dates for the data.
    - Comment on at least a couple of specific values in the data set (e.g., the high and low for the period).
    - Keep your language objective and descriptive. Do not leap to any conclusions about *why* the data might look the way they do.

20.
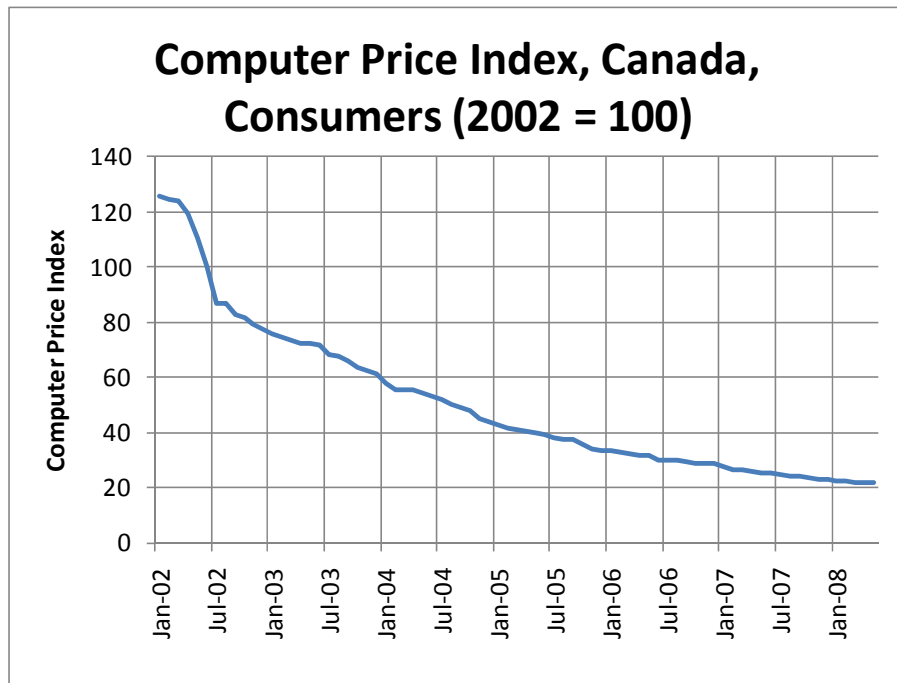
Your graph should look something like this. Be sure that it is labelled completely and correctly. For example, it is important to indicate the base year for any price index.

Your commentary should note that this price index has declined significantly since 2002, with the index hitting a low of 21.91 in May of 2008. The decline in the price index was most pronounced at the beginning of the period, in 2002. The rate of the decline in this price index has slowed somewhat at the end of the period (late 2007 and early 2008).

Develop Your Skills 2.5
21.
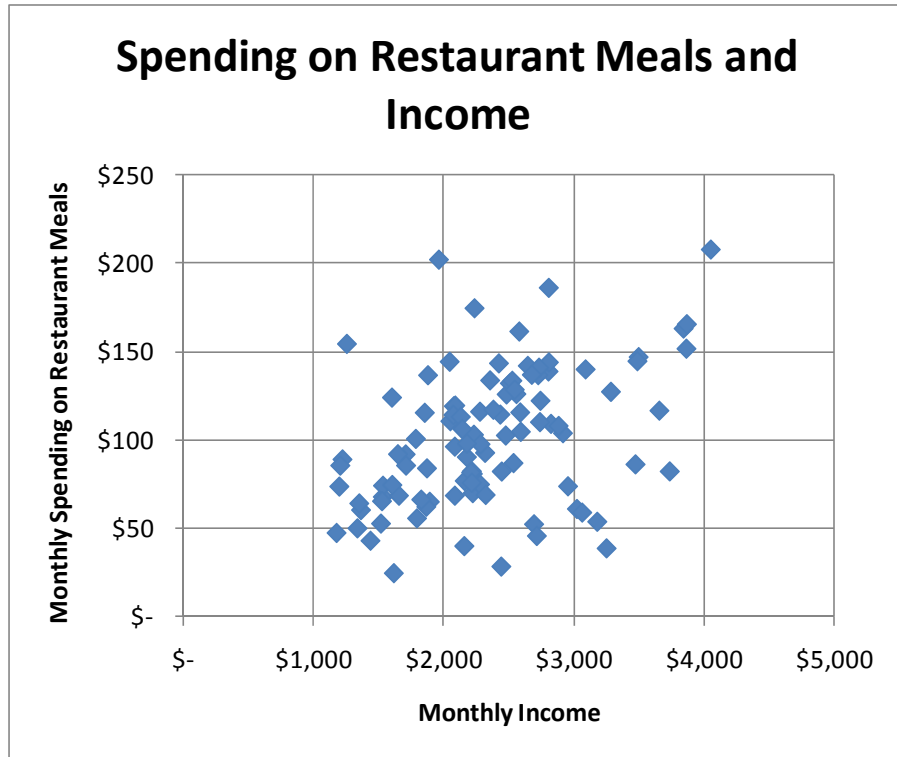
There appears to be a slight positive relationship between monthly income and monthly spending on restaurant meals, that is, the higher the monthly income, the greater the monthly spending on restaurant meals. However, there is a great deal of variability in the spending on restaurant meals, and the relationship is weak.

22.

Jack's Cookies, Daily Sales

Note this graph shows the data with the explanatory variable (price) on the x-axis, and the response variable (quantity sold) on the y-axis, which matches convention. However, if you have taken economics, you might recognize this as a demand curve. For historical reasons, a demand curve is normally graphed with price on the y-axis. So, it is also acceptable to graph these data as follows:

Jack's Cookies, Daily Sales
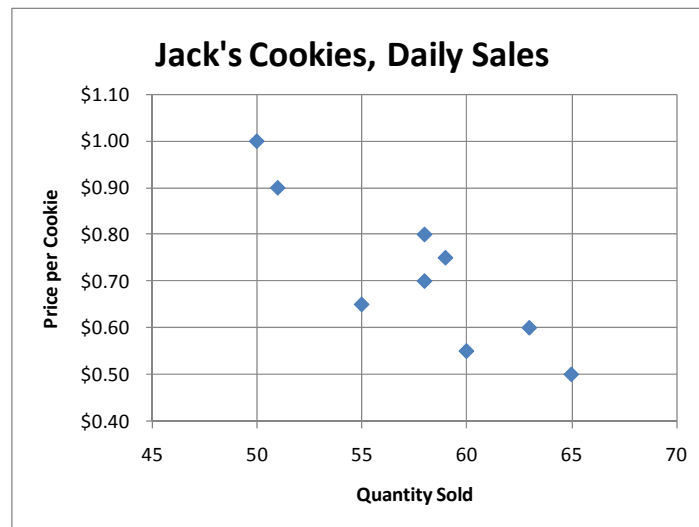
Notice that in both graphs, the axes do not begin at (0, 0). It is reasonable to scale the axes as shown, but this should always be clearly indicated.

The data for daily sales of Jack's Cookies show that the quantity sold and the price are negatively related, that is, the higher the price, the lower the quantity sold. This conforms to the Law of Demand.

23.

Survey of Drugstore Customers
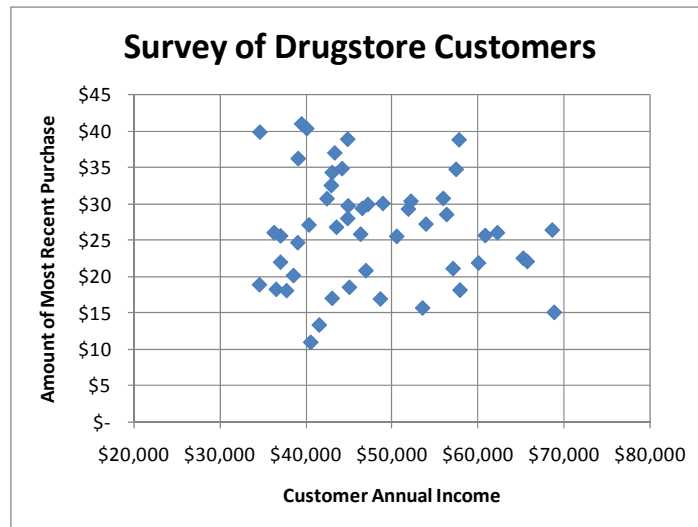
There does not appear to be a strong relationship between the customer's income and the amount of the most recent purchase. Note the scale on the x-axis does not start at zero.

If you think about it, this should not come as a surprise. While we might have expected those customers with greater incomes to have higher purchases, this effect is less likely to appear for a single purchase. There may be more of a positive relationship between annual income and total annual drugstore purchases.

24.

Hours of Work and Semester Marks, Random Sample of Students
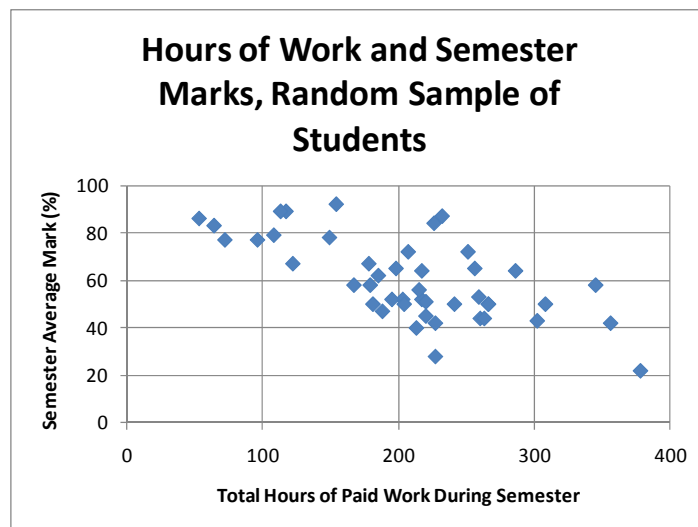
There appears to be a negative relationship between the total hours of paid work during the semester, and the semester average mark, that is, the greater the hours of work, the lower the semester average mark.

25. Exhibit 2.70c is not correct, because the explanatory variable is years of service, and it should be graphed on the x-axis. Exhibit 2.70b is probably not correct, because it depicts a negative relationship, that is, those with more years of service earn *lower* salaries. Exhibit 2.70a is the only possible choice, as it shows higher salaries associated with longer years of service.

Develop Your Skills 2.6
26. The first obvious problem with this graph is the 3-D aspect. It makes it hard to read the height of the bars. It is not clear if the bars for "good" and "fair" are the same height. The graph, with the 3-D aspect, is an appropriate way to represent these data. The graph would be improved if the 3-D aspect removed, as shown below.
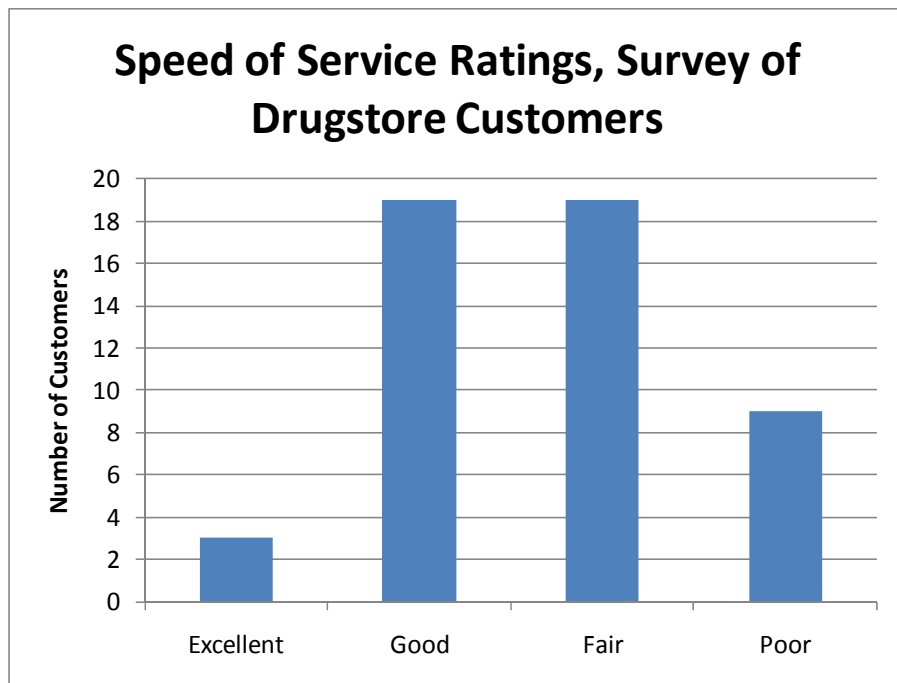
Another possible improvement would be to calculate relative frequencies for each rating category.

27. The pictograph looks as follows:

Figure 2. Purchasing power of the Canadian dollar, 1980 to 2000

1980 = $1.00

1985 = $0.70

1990 = $0.56

1995 = $0.50

2000 = $0.46

The year 2000 dollar is worth just under half of the value of the 1980 loonie, but the total area of the year 2000 loonie is only about a quarter of the area of the year 1980 loonie, so the pictograph misleads the viewer. When the loonie is shrunk, it shrinks not only in height (as a bar in a bar graph would), it is also shrunk in width, so the image is not distorted. This decreases the area disproportionately.

28. This is a good graph. The labels and titles are clear, and the graph can be understood without reference to anything else. We can see that there appears to be a positive relationship between the total monthly sales for Hendrick Software salespeople and the number of sales contacts during the previous month.

29. This graph cannot be interesting, because we have no clue what it is about. We can see that the distribution is skewed, but that's all. With no title, and no meaningful labels on the axes, the graph is useless.

30. There are quite a few categories in this data set, and a bar graph would be preferred. Also, the title is not correct. The survey was of favourite flavours of ice cream (not people). The labels on the pie slices contain the code for the flavour, which is unnecessary and only serves to clutter up the graph. There is also a spelling mistake in one of the labels: "maple walnut".

A better graph is shown below. It has the advantage of sorting the flavours from most favourite to least favourite.

Survey of Favourite Ice Cream Flavours, Random Sample of People Walking Around Kempenfelt Bay
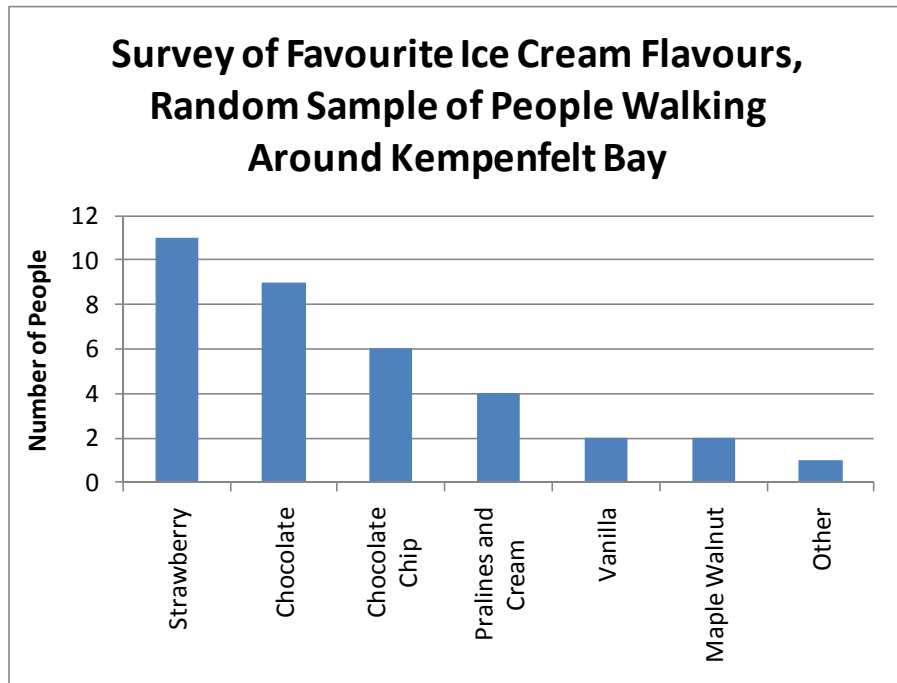
Chapter Review Exercises

1a. These data are qualitative, unranked, and cross-sectional.

1b. These data are qualitative, unranked, and cross-sectional.

1c. These data are quantitative, discrete, and cross-sectional.

1d. These are ranked qualitative data.

1e. These are time-series continuous quantitative data.

2a. A double bar graph could show males and females along the x-axis, with two bars above, one for those with fitness club membership, one bar for those without. Alternatively, categories of fitness club membership could show along the x-axis, with bars for males and females above each category.

2b. A bar graph could be organized with the four store locations along the x-axis, and bars above, each one corresponding to the type of payment. Alternatively, the payment types could show along the x-axis, with four bars above each, one for each store location.

2c. If the total number of pedestrians is recorded, there are only two data points, the number of people who passed by each location. A graph would not really add much to a simple table displaying these numbers, with a proper title and headings.

2d. A double bar graph could be used, with the ratings ("barely edible" to "absolutely delicious") showing along the x-axis, and two bars above each rating, one for each chef.

2e. It is likely that there is interest in the relationship between sales and advertising. A scatter diagram would be appropriate, with advertising along the x-axis, and sales on the y-axis.

3. These graphs are meant to be amusing and entertaining. Quirky images and bright colours make them attractive, but they are not good examples of graphs to summarize data.

4. The stem and leaf display is shown below. The order of the leaves in your display may be different, if you went through the data by rows instead of columns.

```
0 | 9  8  8  9
1 | 9  2  2  3
2 | 2  1  0  0
3 | 1  2  0
4 | 3  0
5 | 8  0
6 | 4
```
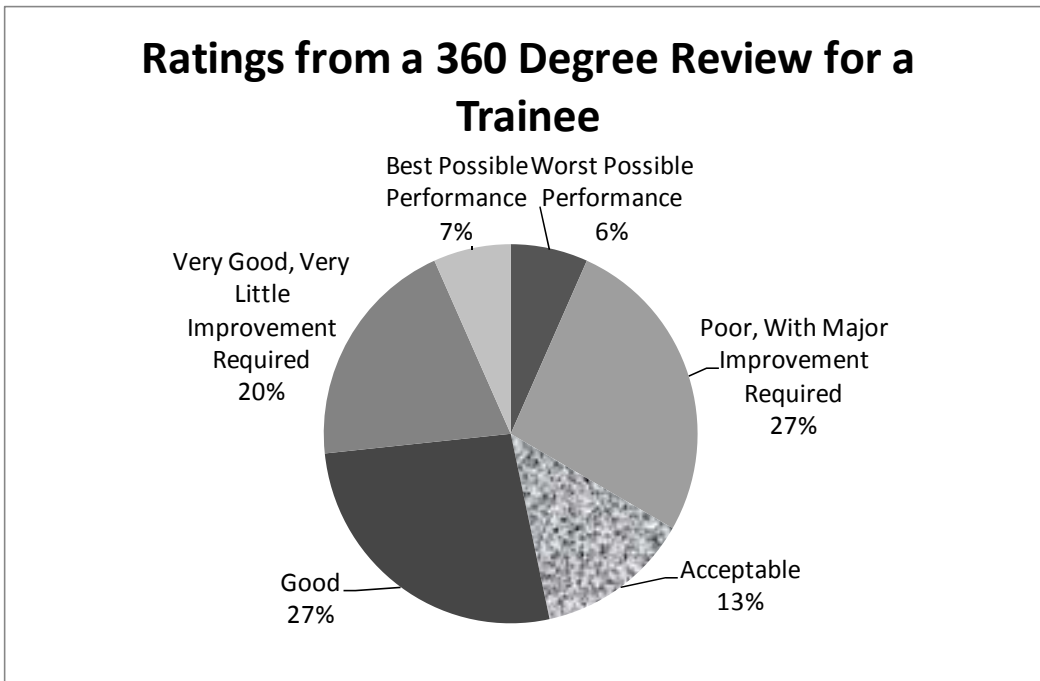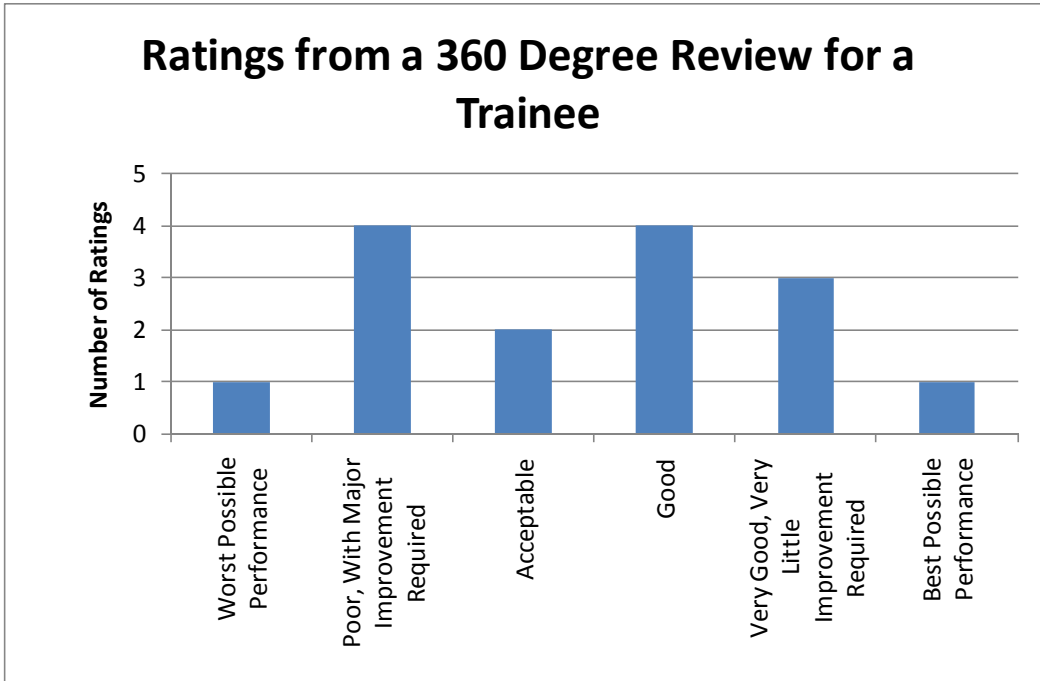
The data set is skewed to the right. Those who are under 30 years old are the largest age groups in the sample. There are only two people in their 40's and two in their 50's, and only one in the 60's.

5. It appears that there is a positive correlation between months of experience in the company, and salary, but the correlation does not appear to be particularly strong. In fact, the two observations with the highest salaries give the appearance of a positive correlation, and without them, there is no obvious relationship.

6. Since these are quantitative data, histograms are required. Since the data are quite different in range, it is a challenge to decide what class width to use. The graphs below show a class width of $10, which is probably too wide for the data for purchases by males, but allows comparison with the purchases by females. Note that it might have been wise to use relative frequencies, rather than frequencies, to make this comparison, since the data sets have different sizes. However, the sample sizes differ by only one, so it is not crucial to do this here.

Music Store Purchases by Females


Music Store Purchases by Males
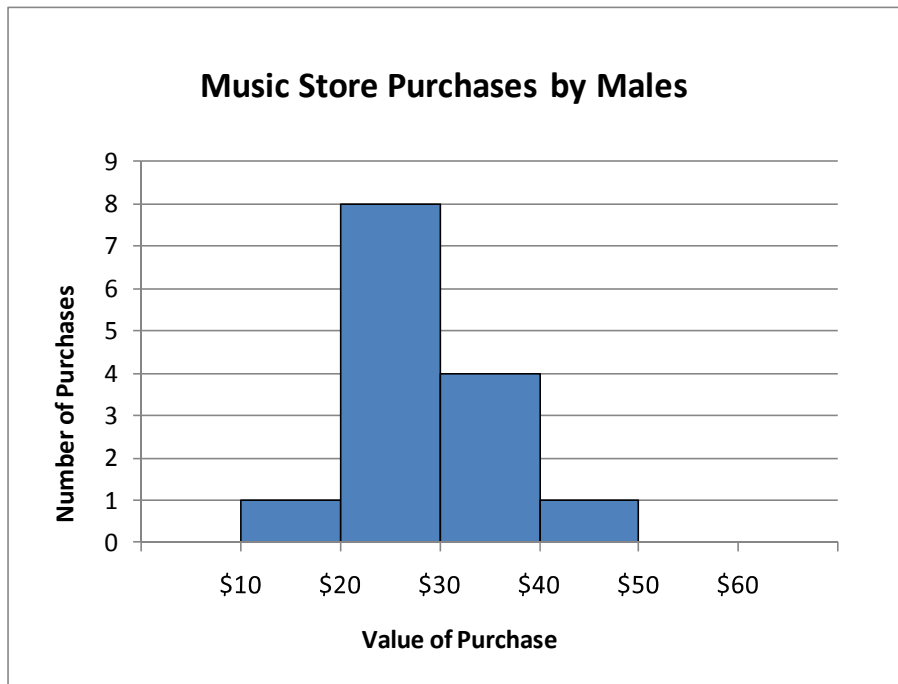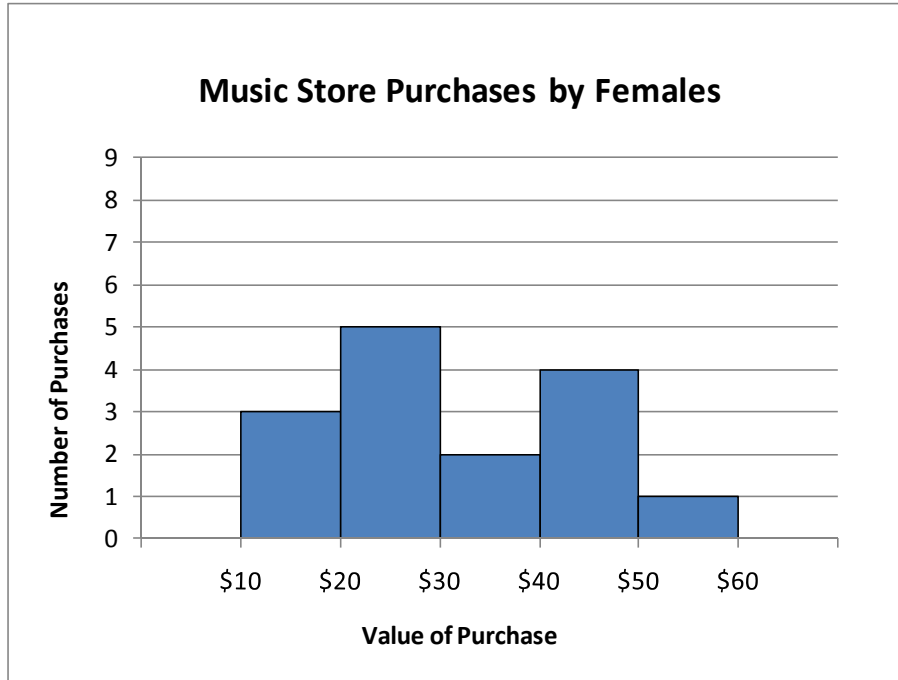
The histograms show that there is more variability in the music store purchases by females. As well, there are more purchases of higher value for females than males. The purchases by males are in the $10-$50 range, while the purchases by females are in the $10-$60 range.

7. There are two possible graphical displays, a bar chart or a pie chart. Both are shown below. The pie chart has been formatted for black and white printout.

Ratings from a 360 Degree Review for a Trainee

Ratings from a 360 Degree Review for a Trainee

Whichever graphical display is used, it is apparent that the trainee's ratings are not consistent. About 67% of raters indicated that the trainee's performance was acceptable or better. However 27% suggested that major improvement was required, and 6% rated the trainee's performance as the worst possible. Certainly, there seems to be a wide range of opinions about this trainee.

8. An appropriate graphical display is shown below.
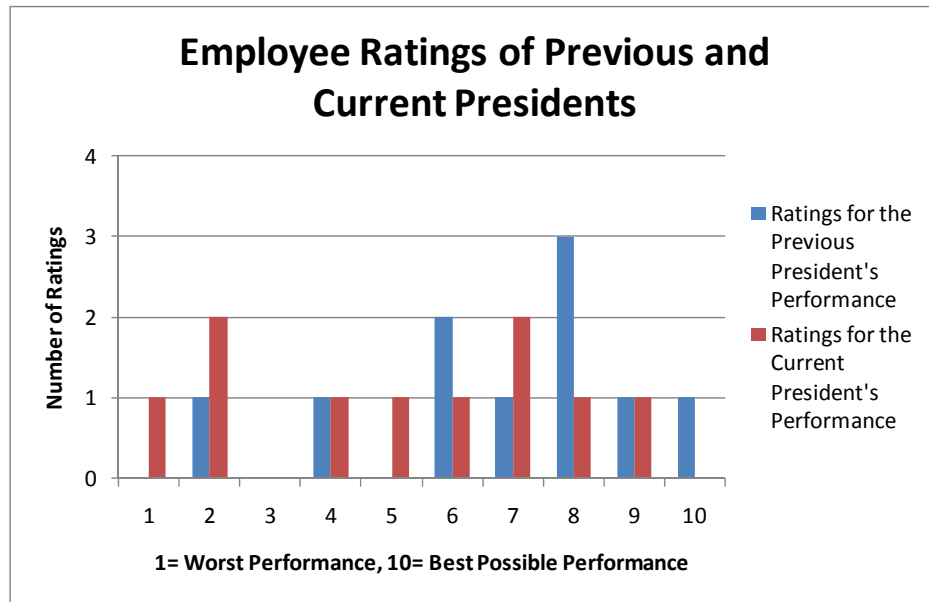
The performance ratings for the new president are generally lower than for the previous president. However, there seems to be great variability in the ratings for both presidents.

9. In this case, since the number of students in each sample is the same, it is appropriate to compare the number of students directly. An appropriate graph is shown below.

    The graph shows that the B.C. students were much more likely to rate this university as "excellent" than the Ontario students, with Ontario students much more likely to rate it as "poor". The Ontario and B.C. students have different opinions about this university.
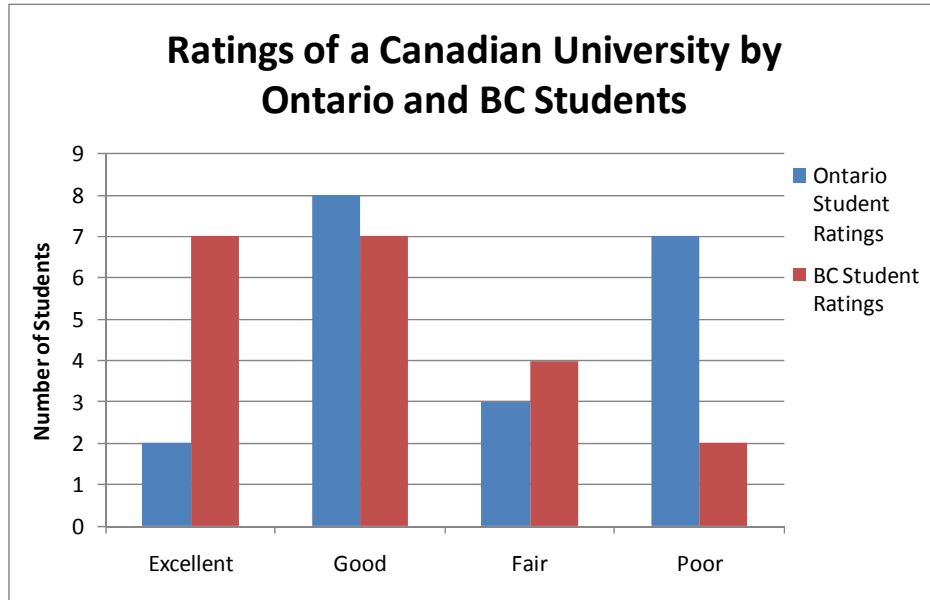
10. A graph to summarize the data is shown below.
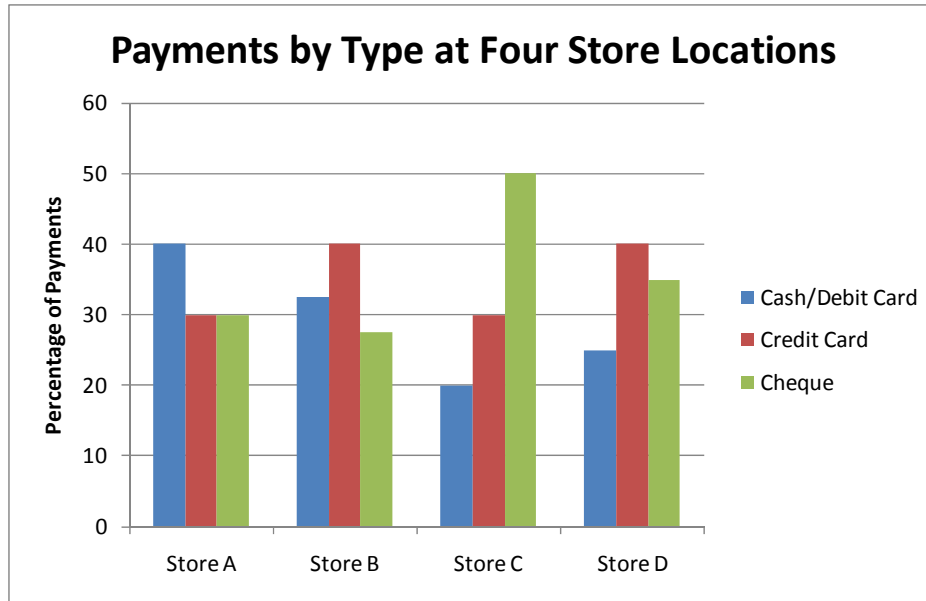
In this case, since the total number of payments is different at the stores, percentage of payments is displayed on the graph, so that the values are directly comparable.
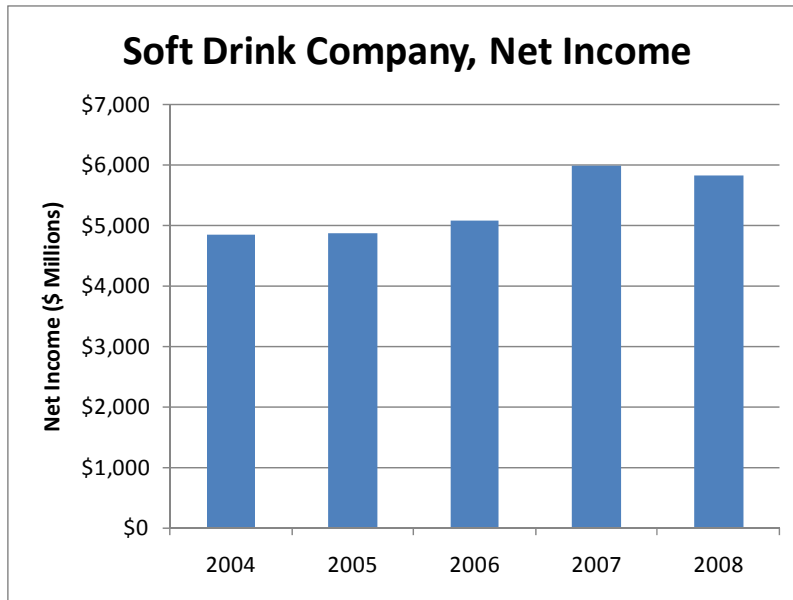
The graph shows that the percentage of payments by cash or debit card is highest at Store A, at 40% of payments, and lowest at Store C, accounting for only 20% of payments. The percentage of payments made by credit card is 30% at both Stores A and C, and is 40% at Stores B and D. Cheques account for 50% of the payments at Store C, which is higher than at any other store. The other three stores have a similar percentage of payments by cheque, from 27.5% to 35%.

11. Since the samples are different sizes, relative frequencies must be used to make the comparison.

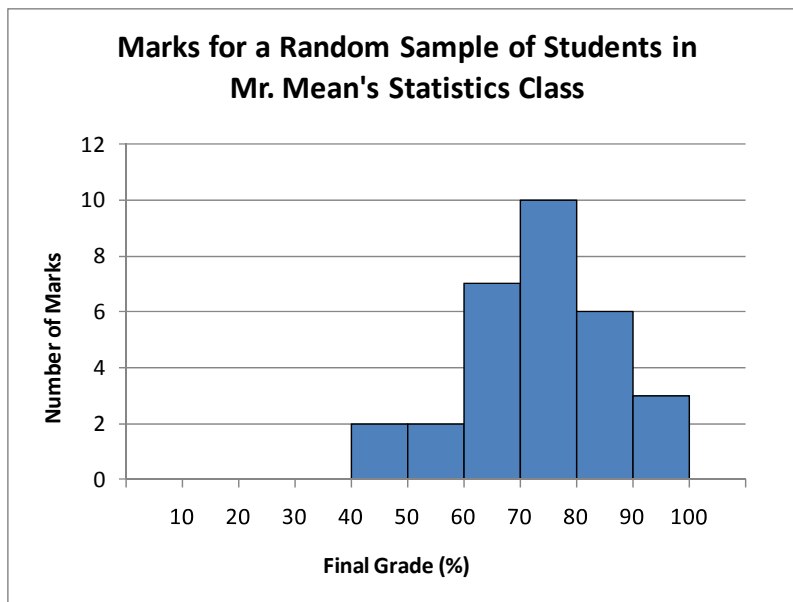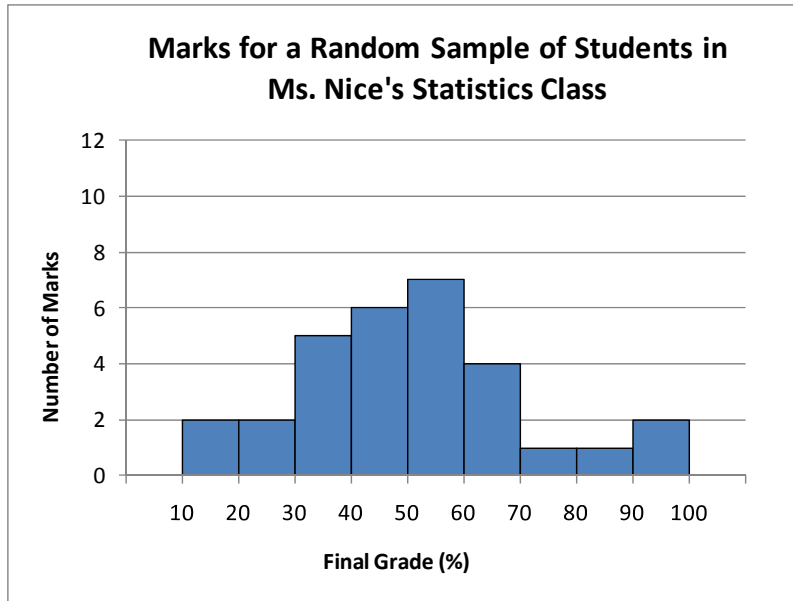**Origins of Students in Two College Programs**
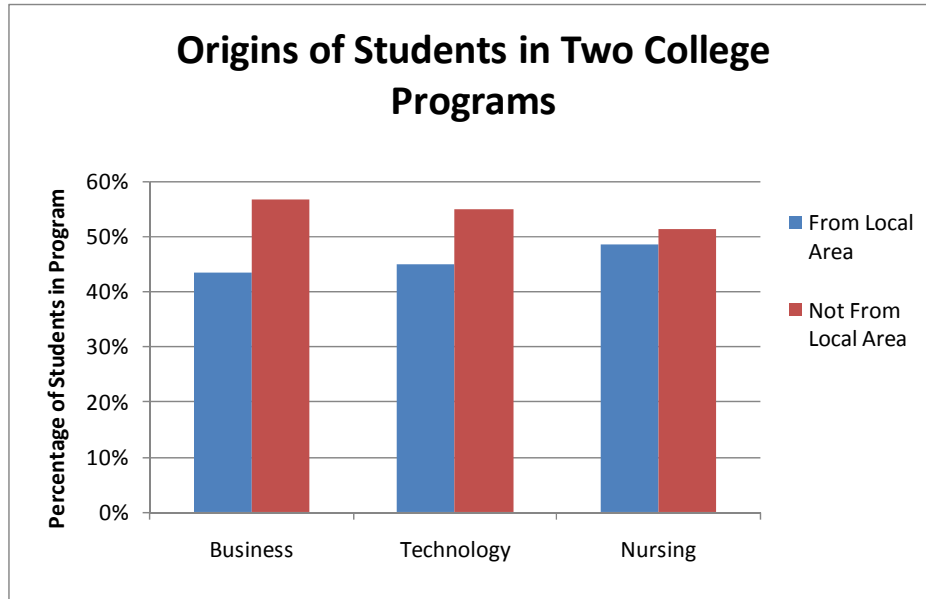
All three program areas draw a greater percentage of students from outside the local area, although the tendency is strongest for the Business program (about 57% of students not from the local area) and weakest for Nursing (about 51% of students not from the local area).

12. The use of the glass with a swizzle stick does not make the graph more interesting, it just makes it more difficult to read. It is quite difficult to judge the level of operating revenues from the pictures—is it the top of the glass or the top of the swizzle stick that we should read? A bar graph (or a line graph) would be a better choice to display these data, as shown below.

Soft Drink Company, Net Income

13. Two histograms, properly set up for comparison, are shown below.

Marks for a Random Sample of Students in Ms. Nice's Statistics Class

Number of Marks (y-axis, 0 to 12)

Final Grade (%) (x-axis, 10 to 100)

Marks for a Random Sample of Students in Mr. Mean's Statistics Class

Number of Marks (y-axis, 0 to 12)

Final Grade (%) (x-axis, 10 to 100)

Notice that the graphs are set up with the same x- and y-axis scales, for direct comparison. They are also similarly sized, so that it is possible to make a direct visual comparison. Class widths of 10 were used, because these are comfortable for marks data, and they allow us to make a distinction between passing and failing grades (assuming 50 is a pass).

The marks of the students from Mr. Mean's class are generally higher and less variable than the marks of the students from Ms. Nice's class. Half of the students from Ms. Nice's class failed the course, while only two of the students from Mr. Mean's class failed.

14. The two histograms are shown below.

Daily Pedestrian Traffic at Location 1

Daily Pedestrian Traffic at Location 2
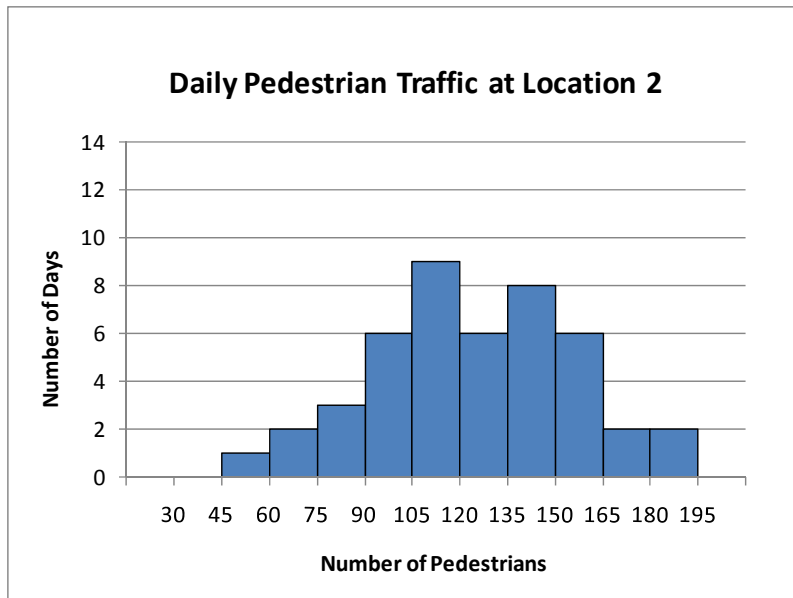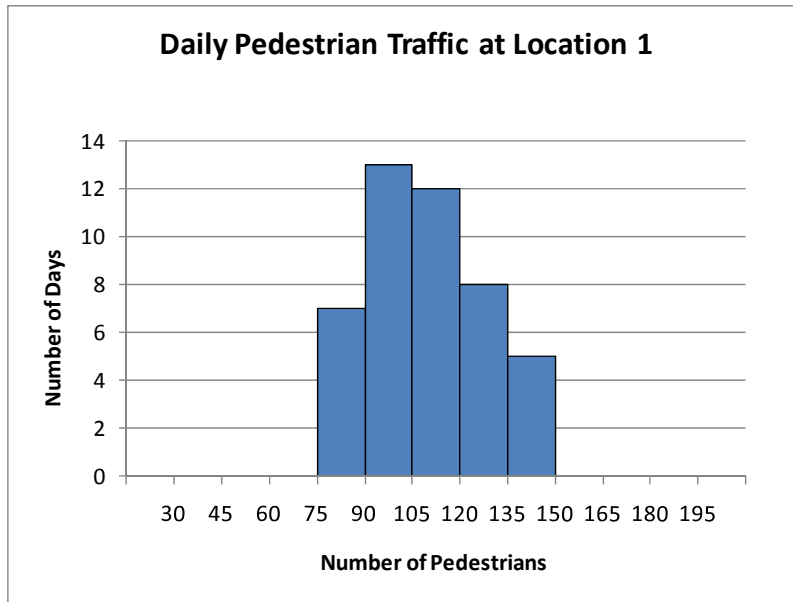
(Note that the histograms are set up with matching x- and y-axes, and are sized similarly, for ease of comparison. Because the two locations were surveyed for the same number of days, we can compare the numbers directly.)
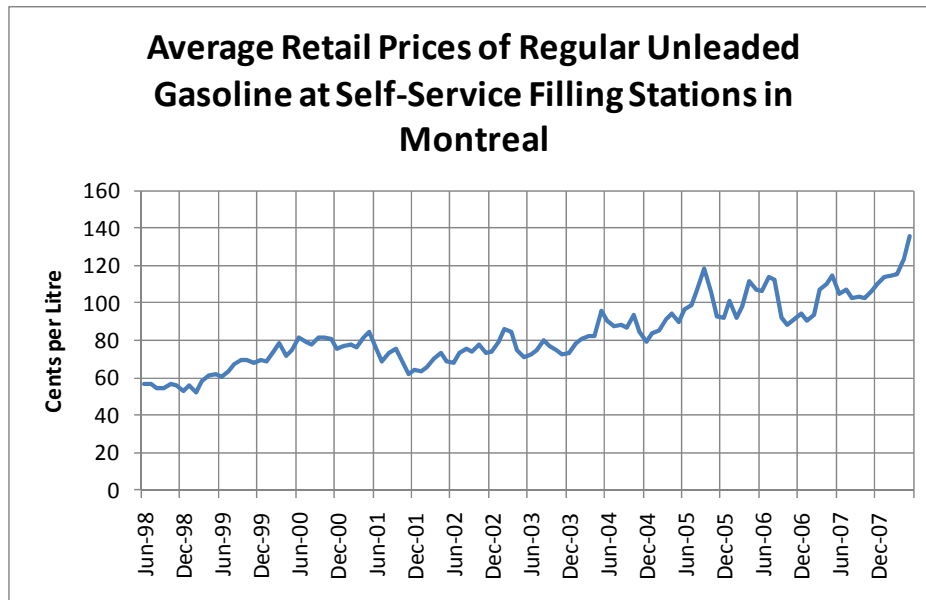
The histograms clearly show that daily pedestrian traffic is more variable at Location 2 than at Location 1. At Location 1, the daily traffic is in the 75-150 range, while at Location 2, it is in the 45-195 range. Generally, it appears the daily traffic at Location 1 is less than at Location 2. For both locations, the histograms are reasonably symmetric.

15. An appropriate histogram is shown below.

Downtown Automotive, Random Sample of Daily Sales
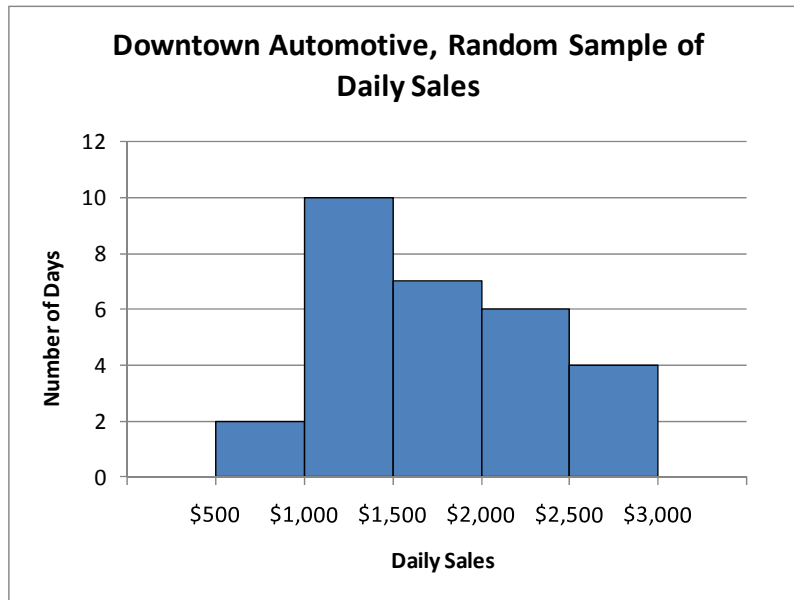
For Downtown Automotive, daily sales are usually above $1,000, with sales falling into the $1,000 to < $1,500 class on 10 of the 29 days in the sample. The distribution is somewhat skewed to the right, that is, there are a few days when sales are above $2,000. Daily sales range from $690 to $2,878.

16. The appropriate graph is a line graph, such as the one shown below. It covers the 10-year period ending in May 2008. You will have more recent data available.

Average Retail Prices of Regular Unleaded Gasoline at Self-Service Filling Stations in Montreal

Over the 10-year period from June 1998 to May 2008, retail gas prices have been rising. The lowest price over the period was 52.2¢ per litre, in February of 1999, and the highest price was $1.36 per litre, in May of 2008. Prices rose fairly rapidly over the end of 1999 and the beginning of 2000, and then stayed fairly steady until June of 2001. At that point, prices fell, from 84.5¢ per litre in May of 2001 to 61.9¢ in November of 2001. They then began to climb again, reaching a high of $1.185 in September of 2005. Retail gas prices in Montreal showed great variability in the range between 88.5¢ per litre to $1.145 per litre through 2006 and 2007, with a sharp increase in April and May of 2008.

17. In this case, while there may be an association between the two variables, the causality link would not be strong. It would not be correct to say that a high mark in Business Math *caused* a high mark in Statistics, because there is very little overlap between the content of the two courses. However, a student with good study habits and good class attendance might do better in both courses.

In this case, while the Business Math mark is not really the explanatory variable, since this course came first, we will put it on the x-axis. A graph of the data is shown below.
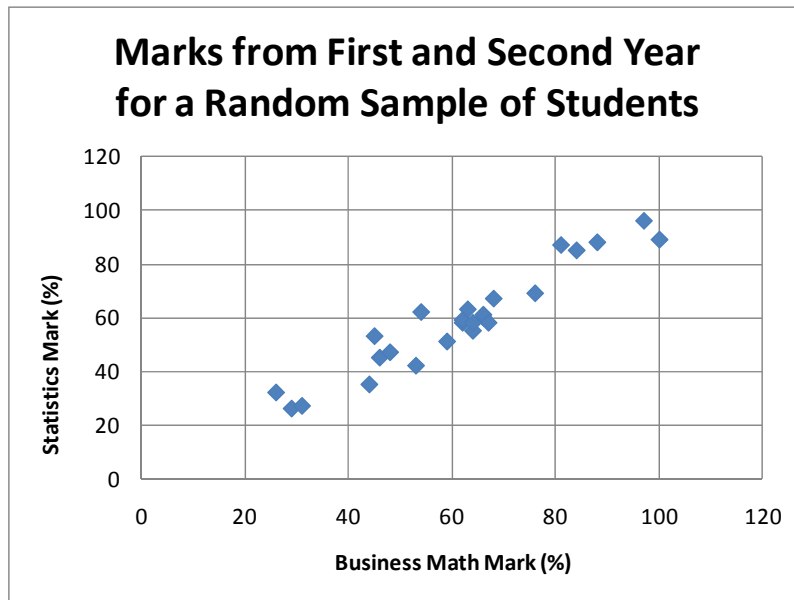
There does appear to be a positive relationship between the two marks.
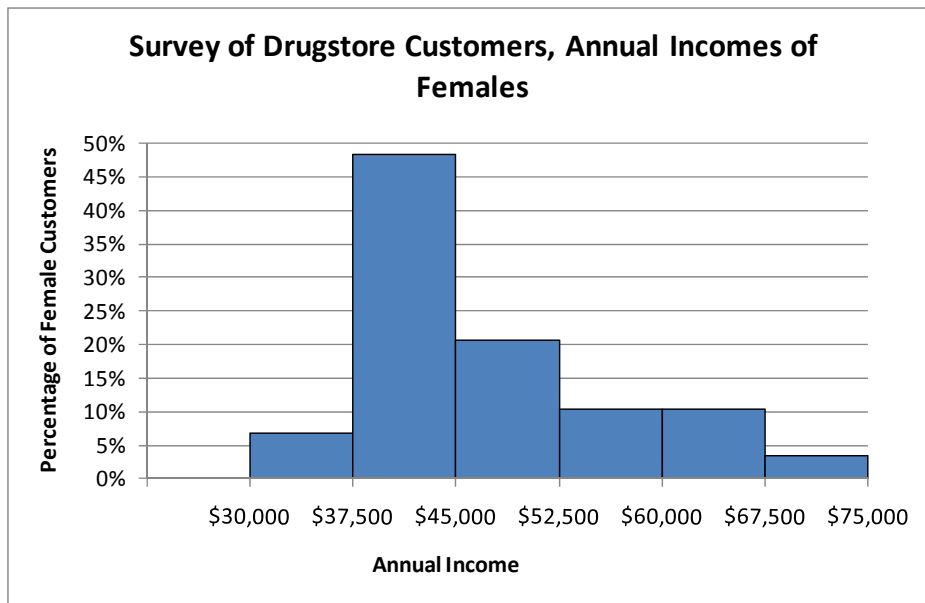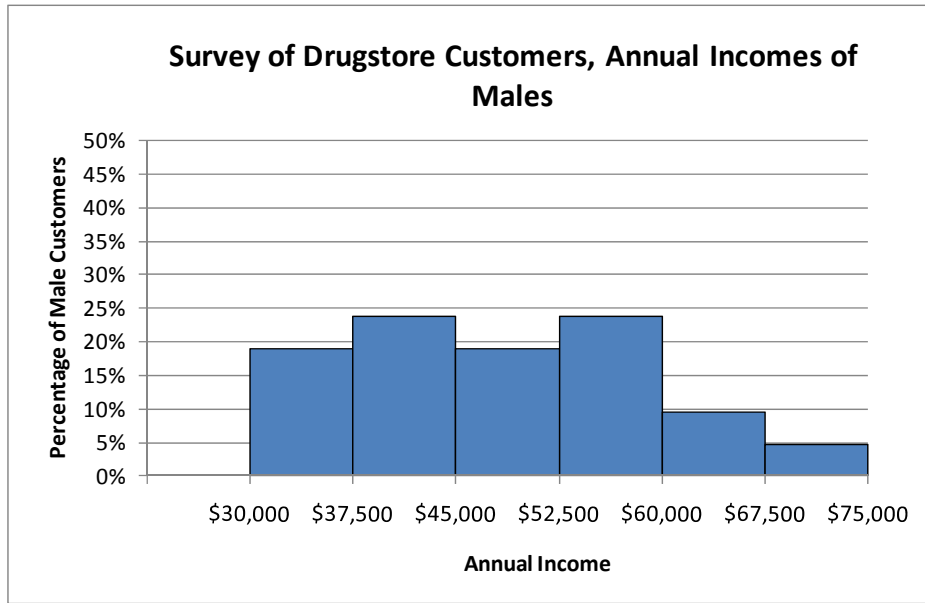
18. A graph of the data is shown below.
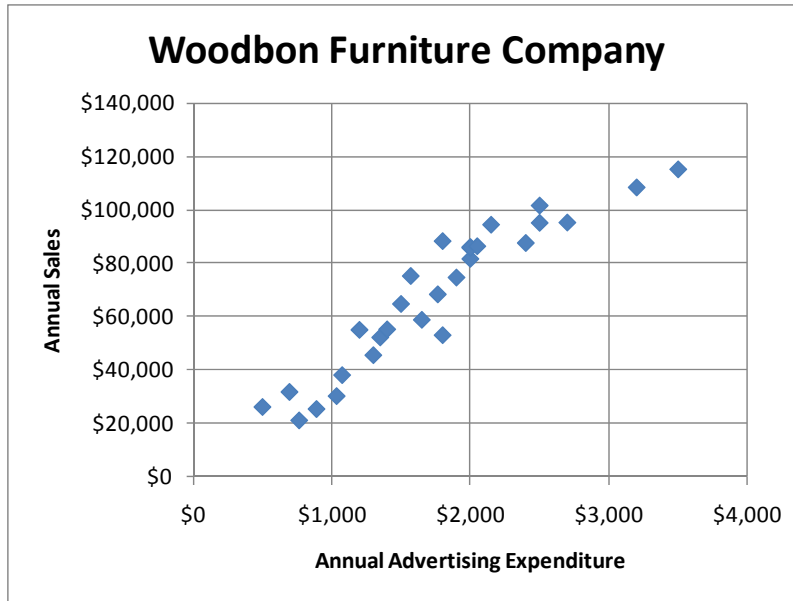
It appears there is a positive correlation between advertising and sales, that is, when advertising expenditure is higher, annual sales are also higher.

19. (Choosing an appropriate class width for comparison takes some thought. $10,000 is probably too wide (resulting in only 4 classes), and $5,000 is probably too narrow. A class width of $7,500 was used for the two histograms shown on the next page.

Because the samples are of different size, relative frequencies should be used for comparison.)

Survey of Drugstore Customers, Annual Incomes of Males

Survey of Drugstore Customers, Annual Incomes of Females

Annual incomes for female drugstore customers are generally in the $37,500 to < $45,000 class, which accounts for over 48% of female customers' incomes. Some incomes of female customers are higher, but this is unusual in the sample, so the distribution of female customers' incomes is skewed to the right.
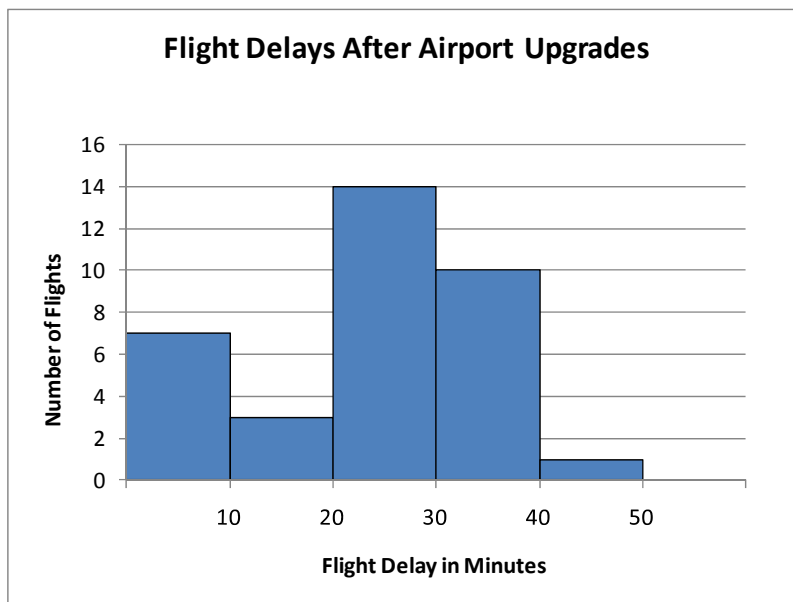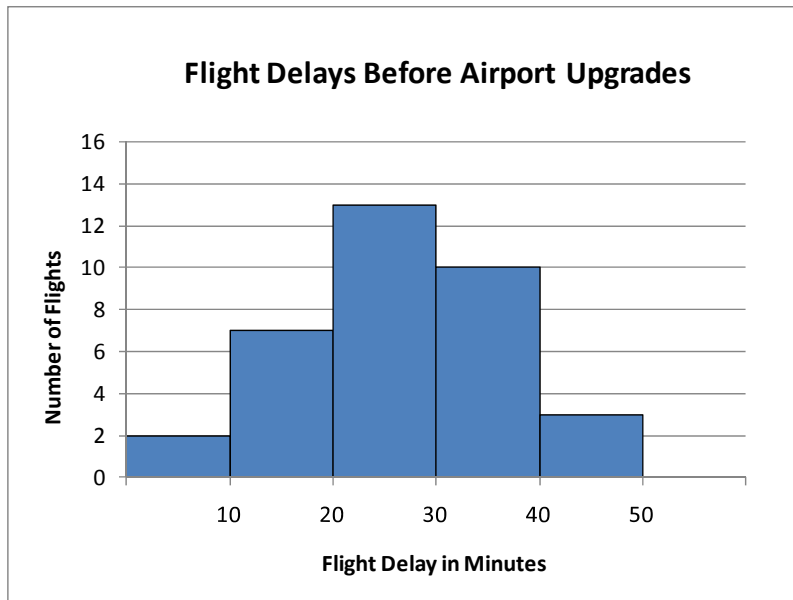
In contrast, the incomes of male drugstore customers are more variable. Incomes between $30,000 and < $60,000 account for over 86% of male customers' incomes, with incomes spread fairly evenly throughout this range. In general, greater percentages of male customers' incomes are in the higher classes.

20. The appropriate graph is shown below. Note that the graph shows percentages of males and females, because of the different sample sizes.

Drugstore Customer Survey, Speed of Service Ratings
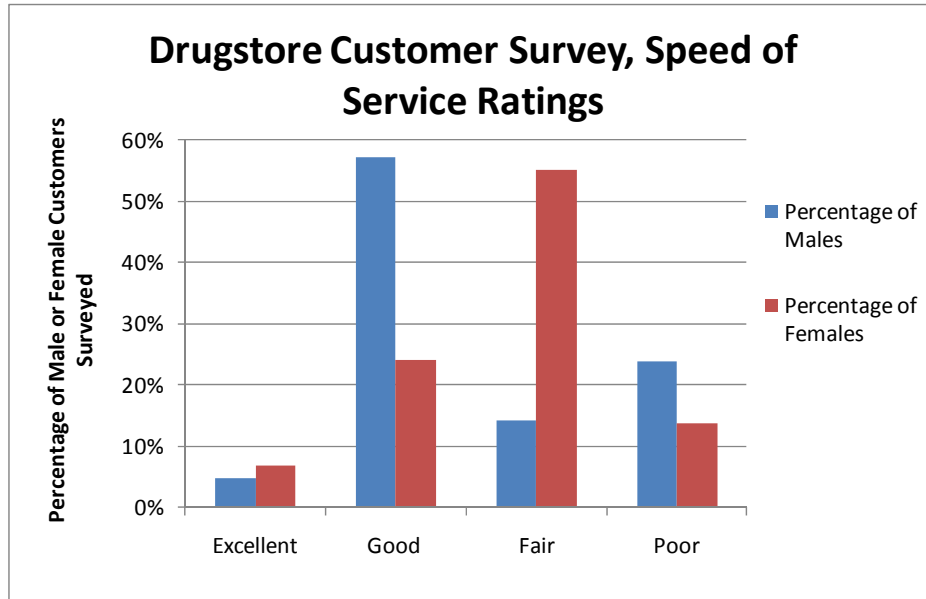
Approximately the same small percentage of male and female customers rated the speed of service at the drugstore as excellent (about 5% of male customers and about 7% of female customers). The largest group of female customers (about 55%) rated the speed of service as "fair", and the largest group of male customers (about 57%) rated the speed of service as "good". It appears that male and female customers rate the speed of service very differently at the drugstore.

21. Two histograms are shown below. Note that samples are the same size, so relative frequencies are not required.

Flight Delays Before Airport Upgrades

Number of Flights (y-axis: 0, 2, 4, 6, 8, 10, 12, 14, 16)

Flight Delay in Minutes (x-axis: 10, 20, 30, 40, 50)

Flight Delays After Airport Upgrades

Number of Flights (y-axis: 0, 2, 4, 6, 8, 10, 12, 14, 16)

Flight Delay in Minutes (x-axis: 10, 20, 30, 40, 50)

The histograms seem to indicate that flight delays have changed after the airport upgrade. Before the upgrade, flight delays were mainly in the 10 to < 40 minute range. Only two delays were less than 10 minutes, and three were more than 40 minutes (but less than 50 minutes). After the upgrades, there were seven delays less than 10 minutes, so a greater number of flights had shorter delays. As well, there was only one flight delayed more than 40 minutes. However, the number of flight delays
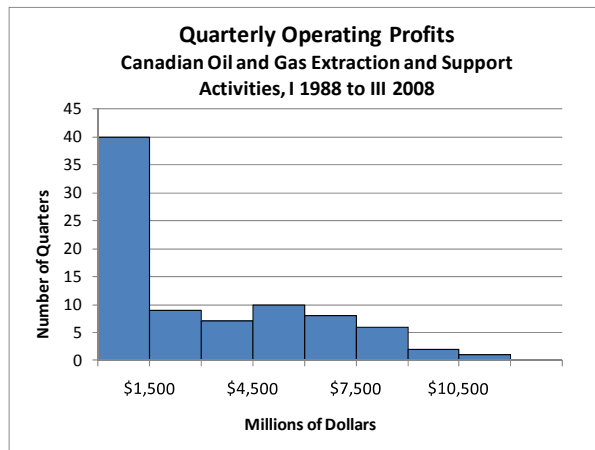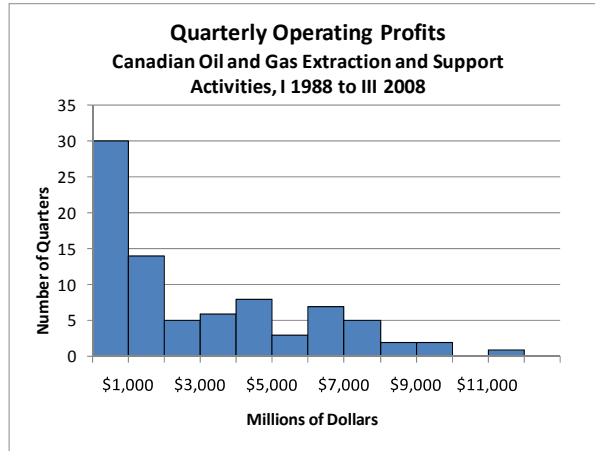
of 10 - < 20 minutes has been reduced from 7 to 3 after the upgrades, while the number of delays of 20 - < 30 minutes has increased from 13 to 14. The greater number of flights with delays less than 10 minutes indicates some reduction in delays, but results appear mixed.

22. The two graphs are shown below.
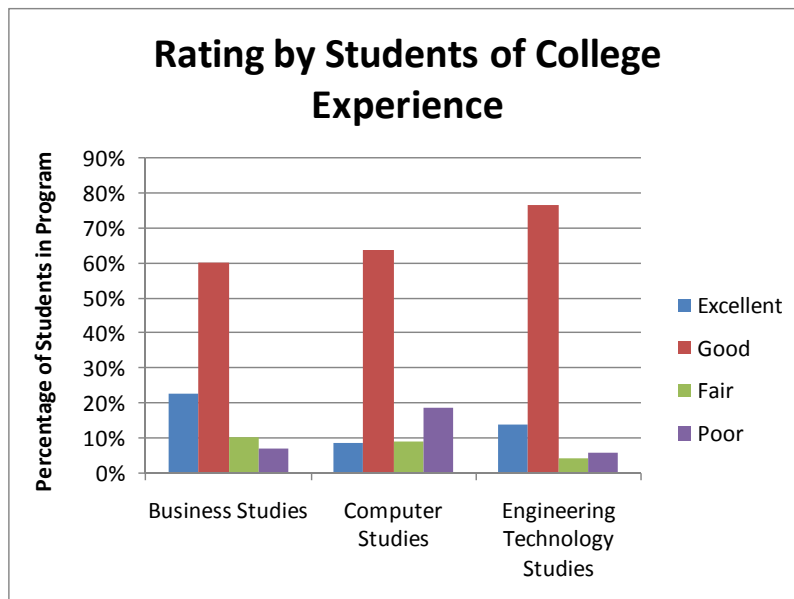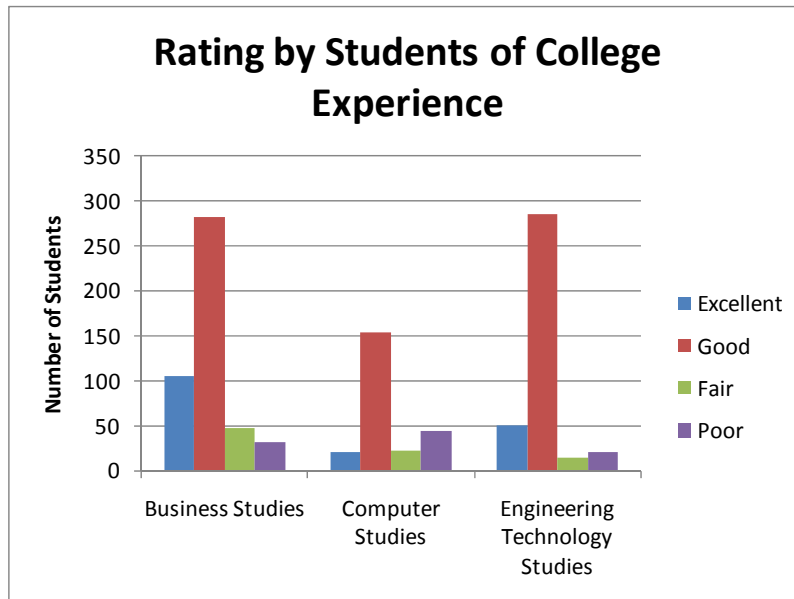
Because the programs have different numbers of students, the first graph using student numbers distorts the comparison. For example, it appears as if Business Studies and Engineering Technology Studies students choose a rating of "good" equally. However, the second graph reveals that a greater percentage of Engineering Technology Studies students rate their college experience as good. While the relative sizes of the ratings for each individual program remain the same, comparisons across programs are not valid unless relative frequencies are used.

23. The three histograms are shown below.

Quarterly Operating Profits
Canadian Oil and Gas Extraction and Support
Activities, I 1988 to III 2008

Quarterly Operating Profits
Canadian Oil and Gas Extraction and Support
Activities, I 1988 to III 2008

Quarterly Operating Profits
Canadian Oil and Gas Extraction and Support
Activities, I 1988 to III 2008

All three histograms show the same general shape, that is, the distribution is right-skewed. In most quarters, operating profits in the oil and gas sector were below $1.5 billion, but there were much higher profits in some quarters.

In the first histogram the classes may be too narrow, as there are very low frequencies in many of the classes. However, this histogram provides more information about the many quarters when operating profits were low, as there is a breakdown for below $1 billion, and from $1 billion to < $2 billion. This information is hidden in the histogram with the widest classes.

It can be a challenge to decide on appropriate class widths when the distribution is very skewed. In the histogram with the widest classes, a lot of data is contained in the first class (half of the data points are there), and so these classes may be a bit wide.

However, any one of these histograms would be acceptable. The particular choice depends on the focus of the analysis.

24. Because you will have more up-to-date data, we cannot provide the histograms for this question. However, you should use the histograms and commentary in the text as guidelines. Be sure to use the same class widths for your comparison, and size your histograms similarly. Choose a class width that works for both data sets (you will probably be able to use $2 billion as the class width, as in the text).

Remember that your commentary should simply describe the data sets. Do not get carried away with speculation about *why* the data look the way they do.