

## Comparing sets of Numerical data

The times taken for the students from two classes to travel to and from school are given below.

Class 9P: 19 49 25 25 22 55 26 38 54 22 33 44 15

86 31 18 33 67 34 42 49 29 45 65 29

Class 9W: 22 34 48 18 58 67 74 66 53 31 57 25 58

49 35 47 50 65 54 49 38 23 58 19 42

- Draw a back-to-back stem-and-leaf plot for this data.
- Comment on the shape of each distribution.
- Use the mean, median and range to compare the data.

	Solve	Think	Apply																											
<b>a</b>	<table border="1"> <thead> <tr> <th>Class 9P</th> <th>Stem</th> <th>Class 9W</th> </tr> </thead> <tbody> <tr> <td>9 8 5</td> <td>1</td> <td>8 9</td> </tr> <tr> <td>9 9 6 5 5 2 2</td> <td>2</td> <td>2 3 5</td> </tr> <tr> <td>8 4 3 3 1</td> <td>3</td> <td>1 4 5 8</td> </tr> <tr> <td>9 9 5 4 2</td> <td>4</td> <td>2 7 8 9 9</td> </tr> <tr> <td>5 4</td> <td>5</td> <td>0 3 4 7 8 8 8</td> </tr> <tr> <td>7 5</td> <td>6</td> <td>5 6 7</td> </tr> <tr> <td></td> <td>7</td> <td>4</td> </tr> <tr> <td></td> <td>8</td> <td></td> </tr> </tbody> </table>	Class 9P	Stem	Class 9W	9 8 5	1	8 9	9 9 6 5 5 2 2	2	2 3 5	8 4 3 3 1	3	1 4 5 8	9 9 5 4 2	4	2 7 8 9 9	5 4	5	0 3 4 7 8 8 8	7 5	6	5 6 7		7	4		8			Look for the tail to determine the shape of the distribution. The greater the mean and median, the longer the travel times. The greater the range, the greater the spread of times. The mean and range are greatly influenced by outliers, so it is often useful to ignore them.
Class 9P	Stem	Class 9W																												
9 8 5	1	8 9																												
9 9 6 5 5 2 2	2	2 3 5																												
8 4 3 3 1	3	1 4 5 8																												
9 9 5 4 2	4	2 7 8 9 9																												
5 4	5	0 3 4 7 8 8 8																												
7 5	6	5 6 7																												
	7	4																												
	8																													
<b>b</b>	<p>The distribution of times for Class 9P is positively skewed; that is, there are more shorter than longer times.</p> <p>The distribution of times for Class 9W is negatively skewed; that is, there are more longer than shorter times.</p>	For Class 9P the tail is to the right. For Class 9W the tail is to the left.																												
<b>c</b>	<p>For Class 9P: Mean = <math>\frac{955}{25} = 38.2</math></p> <p>Median = 33, Range = 71</p> <p>For Class 9W: Mean = <math>\frac{1140}{25} = 45.6</math></p> <p>Median = 49, Range = 56</p> <p><i>Comment:</i> As expected from the distribution shape, the mean and median of Class 9W are greater than those of Class 9P. In general, the students of Class 9W taker longer than the students of Class 9P to get to and from school. The range of times for Class 9P is much larger than for Class 9W, indicating a greater spread of times for Class 9P. But, if we ignore the outlier, 86, the range of Class 9P becomes 52, which is similar to that of Class 9W.</p>	<p>Mean = <math>\frac{\text{sum of scores}}{\text{number of scores}}</math></p> <p>Median = middle score</p> <p>Range = highest score – lowest score</p>																												

## Exercise 3C

- 1** Two groups of Year 9 students were asked to unscramble a seven-letter word. Their times, in seconds, are shown below.

Group 1: 11 16 39 23 51 24 31 4 29 16 27 40 13 23 30 29 6 22 34 38 13

Group 2: 12 27 46 17 26 32 18 15 21 41 37 36 23 8 25 43 34 7 36 12 7

- a** Draw a back-to-back stem-and-leaf plot for this data.  
**b** Comment on the shape of each distribution.  
**c** Use the mean, median and range to compare the data.
- 2** The scores for a class of 16 students on two tests are given below.
- Test 1: 22 42 34 30 19 39 46 41 38 35 47 39 24 45 27 32  
 Test 2: 13 18 21 6 40 16 26 24 35 12 20 26 31 13 15 19
- a** Draw a back-to-back stem-and-leaf plot for this data.  
**b** Comment on the shape of each distribution.  
**c** Use the mean, median and range to compare the data.  
**d** Suggest a possible reason for the skewness.
- 3 a** The table shows the January mean daily maximum temperatures for Sydney and Melbourne over 20 years. Draw a back-to-back stem-and-leaf plot for this data using stems 22 23 24 25 26 27 28 29.  
**b** Comment on the shape of each distribution.  
**c** Use the mean, median and range to compare the data.



Year	Sydney	Melbourne
1	26.7	26.9
2	25.4	28.6
3	26.2	23.9
4	25.0	25.0
5	27.2	25.3
6	26.2	24.0
7	26.9	28.0
8	26.9	24.3
9	26.3	29.5
10	26.4	27.4
11	26.4	24.2
12	25.4	23.9
13	26.8	23.8
14	25.9	23.1
15	27.0	24.4
16	26.8	28.1
17	25.4	25.7
18	25.5	26.0
19	28.9	26.0
20	25.3	22.4

## EXAMPLE 2

The scores of two groups of university students on a mechanical aptitude test are given at right.

- Draw a back-to-back histogram and a side-by-side histogram for this data.
- Comment on the shape of each distribution.
- Compare the mean, median and range for each distribution.
- Suggest a possible reason for the differences in these distributions.

1: Business students

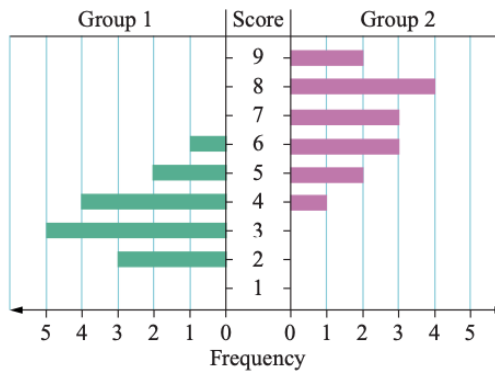
Score	Frequency
2	3
3	5
4	4
5	2
6	1

2: Engineering students

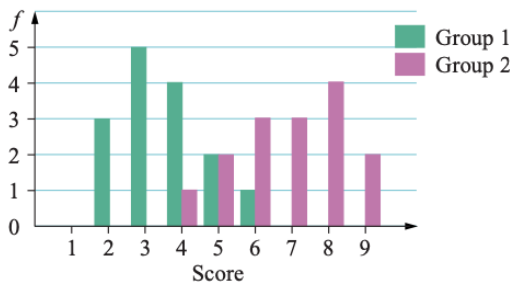
Score	Frequency
4	1
5	2
6	3
7	3
8	4
9	2

### Solve/Think

- For the back-to-back histogram, put columns for each group on the opposite sides of the score axis.



For the side-by-side histogram, draw columns next to each other on the same horizontal axis.



- Group 1 is positively skewed. (There are more lower than higher scores.)  
The group 1 tail is to the right.  
Group 2 is negatively skewed. (There are more higher than lower scores.)  
The group 2 tail is to the left.

### Apply

Back-to-back histograms are drawn on opposite sides of the score axis. The axis can be horizontal or vertical.

Side-by-side histograms are drawn separately next to each other using the same scale on both axes or combined on the same score axis. Use shading to distinguish groups.

- c** For group 1: Mean =  $\frac{53}{15} = 3.5$       For group 2: Mean =  $\frac{103}{15} = 6.9$   
 Median = 3      Median = 7  
 Range = 4      Range = 5
- Comment:* The mean and median for group 2 are higher than for group 1. Overall, group 2 students have scored much higher marks than group 1 students on the test.
- d** The engineering students would be expected to perform better, as shown by the negatively skewed distribution of their scores. The business students would not be expected to score well, as shown by the positively skewed distribution.

- 4** The scores of female and male students on a class English test are shown in the tables.

**Females**

Score	Frequency
4	1
5	1
6	1
7	6
8	4
9	2

**Males**

Score	Frequency
3	2
4	4
5	2
6	2
7	4
8	1

- a** For this data, draw:  
**i** a back-to-back histogram      **ii** a side-by-side histogram
- b** Describe the shape of each distribution.
- c** Compare the mean, median and range for each distribution.
- 5** The 25 students in a Year 9 class were given a test in term 1 and a test in term 2. The results are given below.
- Term 1: 8 9 10 10 10 9 8 7 9 8 10 9 10  
 9 8 9 3 4 9 9 8 5 5 8 9
- Term 2: 2 3 1 0 3 2 5 7 8 9 8 7 6  
 3 4 4 2 3 4 4 0 1 1 2 3

- a** Complete the following tables and draw a back-to-back histogram for this data.

**Term 1**

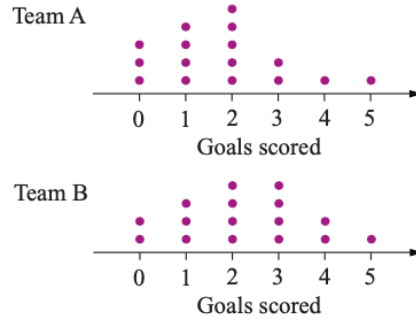
Mark	Frequency
3	
4	
5	
6	
7	
8	
9	
10	

**Term 2**

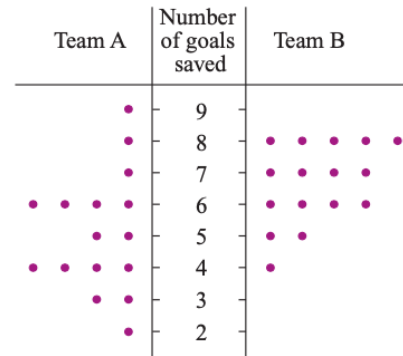
Mark	Frequency
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

- b** Describe the shape of each distribution.
- c** Suggest a possible reason for the skewness.
- d** Use the mean, median and range to compare the data.

- 6** The parallel dot plot shows the number of goals scored by two soccer teams in a 16 match competition.
- Describe the shape of each distribution.
  - Use the mean, median and range to compare the data.



- 7** This back-to-back dot plot shows the number of saves made by goalkeepers for the two teams in question 6.
- Comment on the shape of each distribution.
  - Use the mean, median and range to compare the data.



**Answers**

### Exercise 3C

1 a

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

b Group 1 is positively skewed, Group 2 is symmetric.

c Group 1: mean = 24.7, median = 24, range = 47

Group 2: mean = 24.9, median = 25, range = 39

The means and medians are approximately the same, indicating that both groups took about the same length of time. The range for group 1 is greater than that for group 2, indicating a greater spread of times. However, if we ignore the outlier 51, the range of group 1 becomes 37, which is almost the same as for group 2.

2 a

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

b Test 1 is negatively skewed, test 2 is positively skewed.

c Test 1: mean = 35, median = 36.5, range = 28

Test 2: mean = 20.9, median = 19.5, range = 34

The mean and median for test 1 were both considerably higher than those for test 2, indicating that, in general, the students scored much higher marks on test 1 than on test 2. The ranges indicate that the spread of marks on test 2 was slightly larger than on test 1.

d Test 2 was a lot harder than test 1.

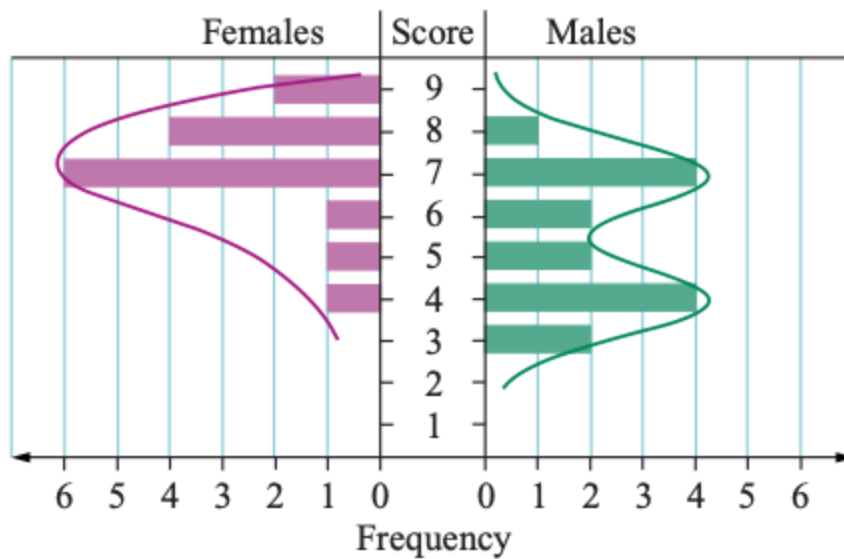
3 a

Sydney	Stem	Melbourne
	22	4
	23	1 8 9 9
	24	0 2 3 4
9 5 4 4 4 3 0	25	0 3 7
9 9 8 8 7 4 4 3 2 2	26	0 0 9
	27	4
2 0	28	0 1 6
9	29	5

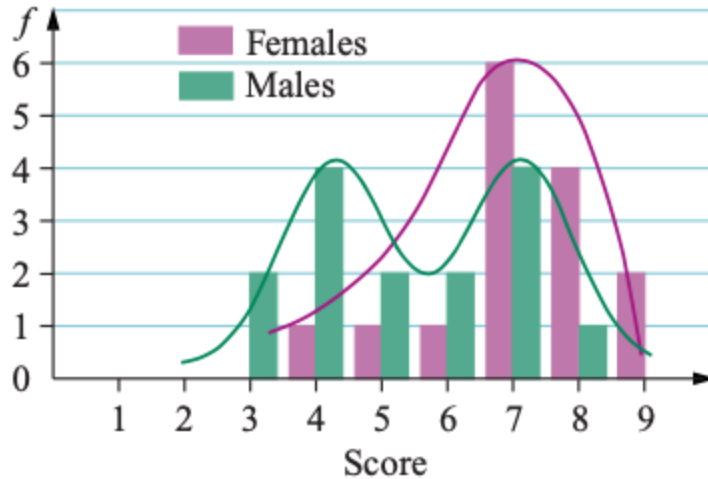
$22|4 = 22.4$

- b** Sydney is positively skewed; Melbourne is positively skewed and bimodal.
- c** Sydney: mean = 26.3, median = 26.35, range = 3.9  
Melbourne: mean = 25.5, median = 25.15, range = 7.1  
The mean and median for Sydney are both greater than for Melbourne indicating that, in general, Sydney's maximum temperatures are higher than Melbourne's. Melbourne has a greater spread of maximum temperatures than Sydney, where maximum temperatures are more consistent.

4 a i



ii



**b** Females: negatively skewed, males: bimodal

**c** Females: mean = 7.1, median = 7, range = 5

Males: mean = 5.3, median = 5, range = 5

The mean and median for females is greater than for males; that is, the females performed better than the males. The ranges are the same, but the female scores are skewed to the right whereas the male scores are more uniformly spread.



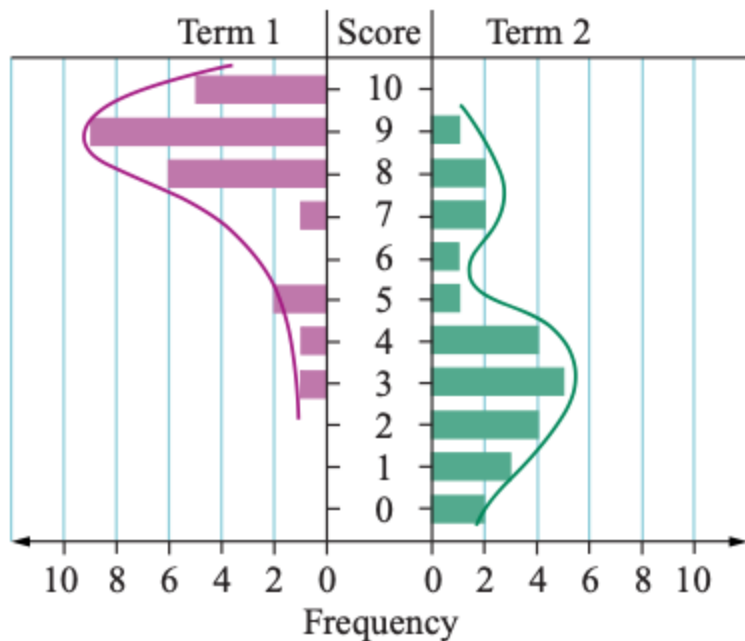
5 a

Term 1

Mark	Frequency
3	1
4	1
5	2
6	0
7	1
8	6
9	9
10	5

Term 2

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



- b** The term 1 test scores are negatively skewed.  
The term 2 test scores are bimodal and positively skewed.

- c** The term 2 test was harder than the term 1 test and perhaps there were some students who understood something the rest of the class did not understand.
- d** Term 1: mean = 8.1, median = 9, range = 7  
Term 2: mean = 3.7, median = 3, range = 9  
The mean and median of the term 2 test are much smaller than those for term 1, indicating that the class performance on this test was significantly less than on the term 1 test, possibly because the term 2 test was harder or because the students were not as well prepared. The spread of marks on the term 2 test was greater than on the term 1 test.
- 6 a** Team A: distribution is positively skewed.  
Team B: distribution is symmetric.
- b** Team A: mean = 1.8, median = 2, range = 5  
Team B: mean = 2.25, median = 2, range = 5  
The median score is the same for both teams, but the mean for team B is greater than that for team A, indicating that team B performed better. The ranges are the same, but as team A's scores are positively skewed, it has more low scores than team B.
- 7 a** Team A: bimodal, team B: negatively skewed
- b** Team A: mean = 5.1, median = 5, range = 7  
Team B: mean = 6.6, median = 7, range = 4  
The range for team A was greater than for team B, but the mean and median for team B were greater than those for team A, indicating that team B's goalkeeper made many more saves than team A's goalkeeper over the season.