Stem and Leaf Plots:

A stem and leaf plot uses the digits of data values to organize a data set. Stem and leaf plots have data placed into order from lowest to highest. The stem and leaf plot show how data are distributed.

Each data is broken into a stem (digit or digits on the left of the vertical line) and leaf (digit or digits on the right of the vertical line). The stems all represent tens place in stem and leaf plot. The leaves all represent one's place in stem and leaf plot.

Use the following steps to construct a stem and leaf plot:

Step 1: Order the data from least to greatest.

Step 2: Identify the stems and leaves.

Step 3: Order the stems from least to greatest.

Step 4: Write the leaves next to their stems.

Step 5: Order the leaves from least to greatest.

Step 6: Write the key. The key explains what the stems and leaves represent.

The numbers in **brackets** indicate how many values are in that class interval (These are not always included but can be useful when there is a large amount of data to display).

What are stem and leaf diagrams used for?

- The data is arranged into **classes** so at a glance it is possible to see the **modal class interval**
- As the data is in order the **median**, **quartiles**, **maximum** and **minimum** can be identified easily
    - Check you can do this – **find** the minimum, maximum, median and upper and lower quartiles from the at stem and leaf diagram the start of this revision note
    - Note that these five values are those needed in order to construct a **box-and-whisker diagram** (box plot)
- **Outliers** (A data value that is extreme or does not fit the pattern), once defined, can be easily identified and removed

Example-1: Draw a stem-and-leaf diagram for the following data

26 45 32 27 29 30 40 36 37

As the data is not in order, draw a **rough** diagram first to get the data values into the correct format:

Stem	Leaves			
2	6	7	9	
4	5	0		
3	2	0	6	7

Now put the stems and their leaves in order

Stem	Leaves			
2	6	7	9	
3	0	2	6	7
4	0	5		

Key: 2|6 means 26

Add a key so we know what the data is showing

Example-2: The stem and leaf plot show daily low temperatures (F°).

Stem	Leaf
1	5 6
2	1 5 7
3	0 6

Key: 1 | 5 means 15 F°

(i) Make an ordered list of the 7 values.
(ii) Find least value, greatest value, mean, median, mode and range.
(iii) What interval includes the most data?

Example-3:

Test Scores

Stem	Leaf
6	6
7	0 5 7 8
8	1 1 3 4 4 6 8 8 9
9	0 2 9
10	0

Key: 9|2 = 92 points

The stem-and-leaf plot shows student test scores. (a) How many students scored less than 80 points? (b) How many students scored at least 90 points? (c) How are the data distributed?

a. There are five scores less than 80 points: 66, 70, 75, 77, and 78.

∴ Five students scored less than 80 points.

b. There are four scores of at least 90 points: 90, 92, 99, and 100.

∴ Four students scored at least 90 points.

c. There are few low test scores and few high test scores. So, most of the scores are in the middle.

Making a Back-to-Back Stem and Leaf Plot

The back-to-back stem and leaf plots are used to compare two distributions side-by-side. This type of back-to-back stem and leaf plot contains three columns, each separated by a vertical line.

The center column contains the stems.

The first and third columns each contain the leaves of a different distribution. The numbers for the leaves of the distribution in the left most columns are aligned to the right and are listed in order.

These are used when it is helpful for the data to be split into two comparable categories such as boy/girl, child/adult, UK/non-UK. Etc

	BOYS					GIRLS					
(2)			4	2	0	1	3	4	4		(4)
(4)	9	6	3	1	1	0	2	6			(3)
(4)	6	5	4	0	2	3	3	3	5	8	(5)
(1)				2	3	1					(1)

KEY 6 | 2 | 3 REPRESENTS 2.6 FOR THE BOYS AND 2.3 FOR THE GIRLS

Note that the leaves on the left-hand side of the stems (Boys) increase from the center outward

Example-4: The following stem and leaf diagrams show the times taken by some children and adults to complete a level on a computer game.

	Children						Adults						
(3)			9	8	8	1							(0)
(4)		6	4	3	1	2	3	5	7	8	8		(5)
(5)	7	6	5	4	2	3	0	4	7				(3)
(3)			3	0	0	4	0	1	2				(3)

2 | 3 represents a time of 23 seconds

(a) Compare the times taken to complete the level between the children and the adults.

(b) It is later discovered two of the adults' times had been omitted from the diagram–times of 23 and 42 seconds.
Briefly explain whether adding these times would change the adults' median time

a) To compare, use an average – so medians – and a measure of spread – so interquartile range

	n	Q_1	Q_2	Q_3
Children	15	4^{th}	8^{th}	12^{th}
Adults	11	3^{rd}	6^{th}	9^{th}

Children					Adults					
(3)			9	8	8	1					(0)
(4)		6	4	3	(1)	2	3	5	(7)	8	8 (5)
(5)	(7)	6	5	4	(2)	3	(0)	4	7		(3)
(3)			3	0	0	4	(0)	1	2		(3)

	Q_1	Q_2	Q_3	IQR
Children	21	32	37	37 – 21 = 16
Adults	27	30	40	40 – 27 = 13

The children had a slightly higher median time of 32 seconds compared to the adults' 30 seconds
This means, on average, the adults completed the level on the computer game quicker than the children
The children also had a greater interquartile range of 16 seconds compared to the adults' 13 seconds
This means the adults' times to complete the level on the computer game were more consistent (less varied)

b) 23 seconds is less than the adults median of 30 seconds
42 seconds is greater
Therefore there will be no change to the adults' median time

Example-5:

A hospital is trying to compare two different medications that claim to reduce blood pressure. They give one set of patients "Drug 1" and a second set of patients "Drug 2" and three hours later record the amount the blood pressure of every patient is reduced by. The results for both groups are below.

Drug 1

12 31 24 18 21 34 40 19 23 17 16

Drug 2

24 18 29 27 32 36 34 31 28 31

(a) Draw a back-to-back stem-and-leaf diagram to show these results,

(b) Comment briefly on what drug you think is more effective, giving a reason why.

Solution:

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

Rough:

Notice how the Drug 2 leaves 'grow' from the centre outwards so when you do an ordered diagram the lowest values will be closest to the stems

Final:

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

Key: 4|2 means a blood pressure reduction of 42

Take your time to make sure you have all the leaves and that they are in order with the correct leaves (also in order)

Don't forget the key!

Example-6:

The number of 'how to vote' cards handed out by various Australian Labor Party and Liberal party volunteers during the course of a polling day is shown below.

Labor	180 193	233 202	246 210	252 222	263 257	270 247	229 234	238 226	226 214	211 204
Liberal	204 287	215 273	226 266	253 233	263 244	272 250	285 261	245 272	267 280	275 279

The stem and leaf plot becomes

Key: 18|0 = 180

Leaf Labor	Stem	Leaf Liberal
0	18	
3	19	
4 2	20	4
4 1 0	21	5
9 6 6 2	22	6
8 4 3	23	3
7 6	24	4 5
7 2	25	0 3
3	26	1 3 6 7
0	27	2 2 3 5 9
	28	0 5 7

For the Labor volunteers:
Mean = 227.9
Median = 227.5
Interquartile range = 36

For the Liberal volunteers:
Mean = 257.5
Median = 264.5
Interquartile range = 29.5

From the stem plot we see that the Labor distribution is symmetric and therefore the mean and the median are very close, whereas the Liberal distribution is skewed towards to high 200's.
Since the distribution is skewed, the median is a better indicator of the centre of the distribution than is the mean.
Comparing the medians therefore, we have the median number of cards handed out for Labor at 228 and for Liberal at 265, which is a big difference.
The data, as shown by the interquartile range, is slightly more spread for Labor as compared to the Liberals
In essence, the Liberal party volunteers handed out a lot more 'how to vote' cards than the Labor party volunteers did.

Note:

(i) The mean and the median (to measure the centre of the distributions), and
(ii) The interquartile range (to measure the spread of the distributions).