Section 3.1.8 – Back to back stem plots

VCAA "Dot Points"
Investigating associations between two variables, including:

- *back-to-back stem plots, parallel dot plots and boxplots and their use in identifying and describing associations between a numerical and a categorical variable*

Stem & Leaf Plot
A **Stem and Leaf Plot** is a type of graph that is similar to a histogram but shows more information. The Stem-and-Leaf Plot summarizes the shape of a set of data (the distribution) and provides extra detail regarding individual values.

A **five number summary** can also be calculated from a stem and leaf plot.

The data is arranged in **increasing order**. The digits in the largest place are referred to as the stem and the digits in the smallest place are referred to as the leaf (leaves). The leaves are displayed to the right of the stem. Stem and Leaf Plots are great organizers for large amounts of information.

Example.1
The following daily temperatures were recorded for the month of April:

12, 22, 32 ,18, 24, 28, 31, 14, 8, 20, 16, 15, 26, 24, 17, 30, 19, 23, 18, 19, 9, 21, 26, 11, 9, 13, 20, 27, 24, & 27

Task.1 Construct a stem and leaf plot for the above data

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

Key: 1|5 = 15

Task.2 Calculate a 5-number summary upon the data.
Min = 8°C
Q_1 = 15°C
Med = 20°C
Q_3 = 26°C
Max = 32°C

Task.3 Describe the shape/type of distribution
The distribution of the temperature is symmetrical.

Back to Back Stem & Leaf Plot

Back to back stem and leaf plots are used to display **two categorical sets of numerical data**. This type of plots are most useful for **comparing two sets of data**.

NB: Be sure that data is entered in ascending order starting with the smallest values placed nearest the stem and increasing in value as the leaf data spreads away from the stem.

Example.2

The following scores were recorded from a math class consisting of both boys and girls.

Boys Scores: 42, 44, 51, 52, 52, 53, 64, 65, 65, 67, 68, 69, 73, 76, 76, 78, 84, 87.
Girls Scores: 40, 51, 62, 63, 74, 74, 78, 79, 81, 83, 83, 87, 88, 89, 90, 96, 98, 99.

Task.1 Construct a back to back stem and leaf plot from the above data.

Leaf (Girls)	Stem	Leaf (Boys)
0	4	2 4
1	5	1 2 2 3
3 2	6	4 5 5 7 8 9
9 8 4 4	7	3 6 6 8
9 8 7 3 3 1	8	4 7
9 8 6 0	9	

Key: $1|5 = 15$

Task.2 Calculate a 5-number summary for both sets of data.

	Girls	Boys
Min	40	42
Q_1	74	52
Med	82	66
Q_3	89	76
Max	99	87

Task.3 Compare the distribution of the two sets of data. ie. girls and boys math scores

Comment.1 (about central tendencies)
The girls generally performed better than the boys. Med (girls) was 82% whilst med (boys) was only 66%.

Comment.2 (about spread)
The girls had less spread in scores than the boys in the middle 50% of data. The girl's IQR was 15%. Whereas the boy's IQR was 24%.

However, the girl's scores demonstrated a greater spread across the entire class. The girl's scores had a range of 59%. Whereas the boy's scores had a range of only 45%.
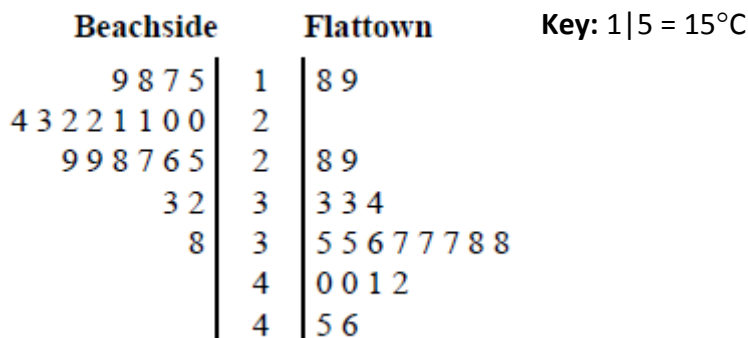
Comment.3 (shape)
The girl's scores appear to have a negative skew. Whereas the boys scores appear symmetrical.

NB: Comments are made about the central tendencies, spread and shape of distribution.
When commenting upon central tendencies, spread or shape you must back it up with a numerical comparison.

The following information relates to Examples 3, 4 & 5.

The back-to-back **ordered** stemplot below shows the distribution of maximum temperatures (in °Celsius) of two towns, Beachside and Flattown, over 21 days in January.

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

Key: 1|5 = 15°C

Example.3

The variables *temperature* (°Celsius) and *town* (Beachside or Flattown) are

A. both categorical variables.
B. both numerical variables.
C. categorical and numerical variables respectively.
D. numerical and categorical variables respectively.
E. neither categorical nor numerical variables.

Back to back stem plots always have a numerical variable and categorical variables (x2)

Temperature-numerical
Towns-categorical.

∴ Option D

<div style="border:1px solid black; display:inline-block; padding:4px 12px;">**D**</div>

Example.4

For **Beachside**, the range of maximum temperatures is

A. 3°C
B. 23°C
C. 32°C
D. 33°C
E. 38°C

Range = Max - Min
    = 38°C − 15°C
    = 23°C

∴ Option B

<div style="border:1px solid black; display:inline-block; padding:4px 12px;">**B**</div>

Example.5

The distribution of maximum temperatures for **Flattown** is best described as

A. negatively skewed.
B. positively skewed.
C. positively skewed with outliers.
D. approximately symmetric.
E. approximately symmetric with outliers.

Relatively symmetrical

Lower fence = $Q_1 − 1.5 \times IQR$
     = $33 − 1.5 \times (7)$
     = 22.5

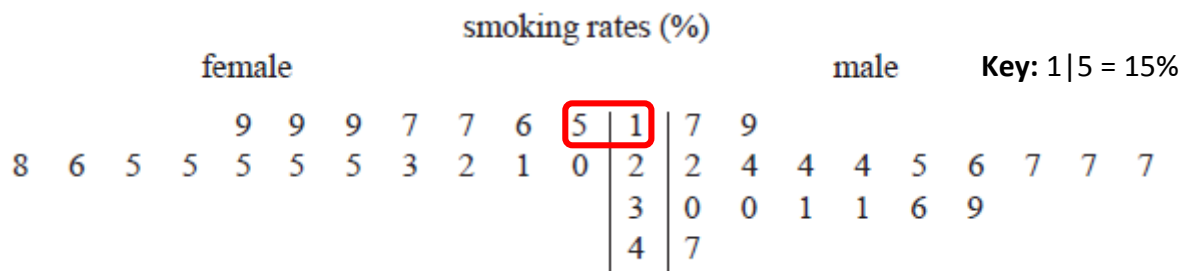∴ values 18 & 19 are both outliers.

∴ Option E

<div style="border:1px solid black; display:inline-block; padding:4px 12px;">**E**</div>

The following information relates to Examples 6 to 8.

The back-to-back ordered stem plot below shows the female and male smoking rates, expressed as a percentage, in 18 countries.

smoking rates (%)

| female | | male | **Key:** $1|5 = 15\%$ |

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

Example.6
For these 18 countries, the lowest female smoking rate is

A. 5%
B. 7%
C. 9%
D. 15%
E. 19%

Reading off the back to back stem plot it is easy to see that the lowest rate for female smokers was 15%.

$\therefore$ Option D

D

Example.7
For these 18 countries, the interquartile range (IQR) of the female smoking rates is

A. 4
B. 6
C. 19
D. 22
E. 23

$IQR = Q_3 - Q_1$
$\quad = 25 - 19$
$\quad = 6$
$\therefore$ Option B

B

Example.8
For these 18 countries, the smoking rates for females are generally

A. lower and less variable than the smoking rates for males.
B. lower and more variable than the smoking rates for males.
C. higher and less variable than the smoking rates for males.
D. higher and more variable than the smoking rates for males.
E. about the same as the smoking rates for males.

A

The min, median and max values for the females are all lower than the males, $\therefore$ female smoking rates are lower.
The IQR and range of females is lower than the males, $\therefore$ less variation also. $\therefore$ Option A.