

Stem & Leaf Plots

Teacher Notes and Answers

7 8 9 10 11 12



Introduction

A Stem and Leaf plot is a device for representing quantitative data in a visual way without loss of information. A range of data sets have been included in this activity. The focus of this activity is on understanding how the Stem and Leaf plot displays information and how information can be gleaned from this representation.

Teacher:

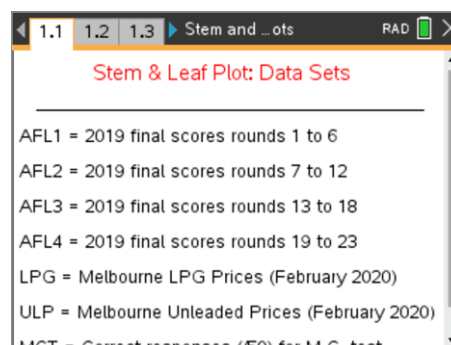
Note: This file will ONLY work on TI-nspire CX II and TI-nspire CX II CAS calculators. It uses a simple program that generates a Stem and Leaf Plot on the drawing canvas. Students must have the TI-nspire file “Stem and Leaf Plots” loaded on their calculators in order to complete the activity. The file contains the data and also the program to create a Stem and Leaf plot. It is possible to edit the program and have Quartiles highlighted, display other information or span the plot over multiple pages, however this is left up to the individual to edit. The program included in the file provides a basic Stem and Leaf plot with the aim for students to find key statistics from the plot and understand how it is formed essentially without any prior knowledge.

Open the TI-nspire file: “Stem and Leaf Plots”.

Page 1.1 contains information on the data sets.

Page 1.2 contains information about syntax and limitations.

Read through this information then navigate to page 1.3



The first data set to be used is the Multiple Choice Test (MCT) response data. This data contains the results for students completing a multiple choice test, the scores are marked out of 50.

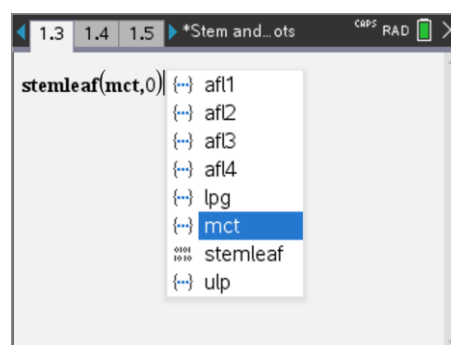
To generate a Stem and Leaf plot either type the command (Program name) or select StemLeaf from the Variable menu. [VAR]

Once this item has been selected, complete the syntax:

```
Stemleaf(mct,0)
```

Then press [enter] to see the Stem and Leaf plot and [esc] to exit.

Use the Stem and Leaf plot to answer the following questions.



The following questions relate to the Stem and Leaf plot for the Multiple Choice Test data.

Question: 1.

What value does each number in the Stem represent?

Answer: “Tens”.

Question: 2.

How many students obtained a perfect score of 50?

Answer: 2

Question: 3.

How many students attempted the test? (Received a score)

Answer: 30 students

Question: 4.

To allow for the fact that students could guess some of the answers (multiple choice), the teacher set the pass mark at 30. How many students passed the test?

Answer: 18 students passed (Scores of 30 or higher)

Question: 5.

What was the most common test score(s)? (Mode)

Answer: Most common score(s) are 11 and 48 (Three students in each group) making the data bimodal.

Question: 6.

What as the median study score?

Answer: 35.5 (Arithmetic mean of 35 and 36)

Question: 7.

The teacher decided to grade the papers according to the following table. Determine the quantity of students that received each grade:

Score Range	<25	25 – 29	30 – 34	35 - 39	40 – 44	45 – 50
Grade	F	E	D	C	B	A
Quantity of Students	8	4	2	6	1	9

Question: 8.

In relation to the grading scale provided in Question 7, discuss how the Stem and Leaf plot could be reconstructed so as to make it quicker and easier to extract the letter grade data.

Answer: The Stems could be set in increments of 5 instead of 10. Note that the A grade is slightly different.

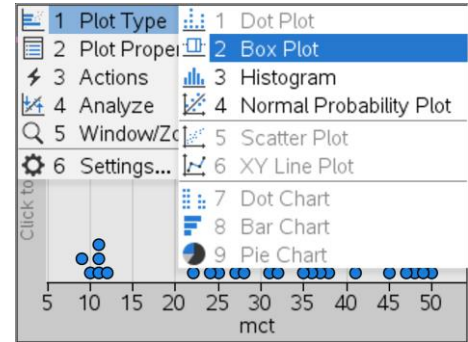
Teacher Notes

The purpose of this question is to have students think about different Stem structures. The program included in the file doesn't deal with Stem values other than magnitudes for 10. This forces students to think about what such a Stem and Leaf plot might look like and potentially construct one for themselves.

The Multiple Choice Test data can also be converted to a dot-plot.

Navigate to page 1.4 and put MCT as the variable on the x axis.

If a Dot Plot is not displayed, use the [menu] to change the Plot Type to Dot Plot or use the Window/Zoom option to zoom in on the data.



Question: 9.

Compare the information available from the dotplot with that contained in the Stem and Leaf plot.

Answer: The dot plot is missing exact numerical scores and the heights of the columns can depend somewhat on the amount of local points which can vary depending on the chosen scale. The dot plot however does show up that a group of six students are clumped around the score of 10 and there appears to be a jump up to scores in the low 20's. The Stem and Leaf plot makes the data appear to be more evenly distributed throughout.

A histogram is another way of displaying the data.

Use the **menu** to change the **Plot Type** to a **Histogram**.

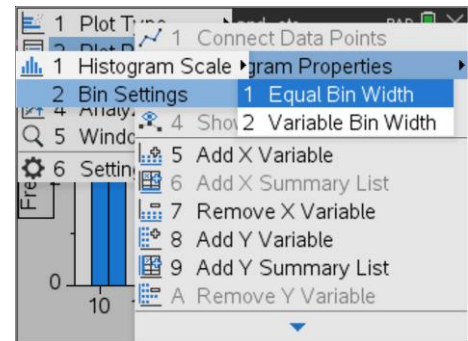
The histogram will use the default settings based on the data and window settings. Use the **menu** to change the 'bin width':

Menu > Plot Properties > Bin Settings > Equal Bin Width

Once the histogram is generated, change the window settings:

Menu > Window/Zoom > Window Settings

Xmin = 0, Xmax = 60, Ymax = 60



Question: 10.

Compare the histogram to the Stem and Leaf plot.

Answer: The histogram with a bin width of 10 units starting at 0 produces a purely graphical representation of the Stem and Leaf plot. It is visually simpler, but does not provide information about how the data is distributed within a single column.

Teacher Notes

How data is distributed within a single column is addressed in further detail when looking at the LPG and ULP fuel prices where figures generally end in a "0.9". Worthwhile discussing the common use of "0.99" on the cents of any price, or at another level where cars may sell for \$19,999.00. We tend to look at the first two digits to gain a level of significance, so a car for the price of \$19,999.00 can often be perceived as a \$19,000.00 car. Strangely, the '00' on the end (cents) can also assist in this perception.

We do this with words also:

"It deosn't mttar in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sill raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe."

Is this similar to our initial perception of numbers?

The following questions relate to the Stem and Leaf plot for the ULP data.

The ULP data represents pricing information collected for a series of service stations located within a 6km radius of a particular suburb in Melbourne in February 2020. Run the stemleaf program: stemleaf(ulp,0).

Question: 11.

What value do the numbers in the Stem represent for the ULP data?

Answer: Cents, 179 = \$1.79 or 179cents.

Question: 12.

What value do the number in the Leaf represent for the ULP data?

Answer: Tenths of a cent. Example: 179 | 9 = \$179.9 cents per litre of fuel.

Question: 13.

Why are there blank leaves adjacent to some of the Stem values? Why might this be required?

Answer: A blank leaf occurs when there is no corresponding data. For the ULP data, this means that there were no fuel stations charging 170.#, 176.#, 177.# or 178.#. As the Stem and Leaf plot provides a visual representation of the data, including the spread of the data, empty leaves must be included as to help show the spread.

Question: 14.

What do you notice about much of the Leaf data?

Answer: Most of the leaf values are equal to 9, which means 169.9 cents is really close to \$1.70

Question: 15.

What is the range of prices in this area during the data collection period?

Answer: Most prices are less than 175.9 and all are greater than 167.5. (Range = 8.4 cents per litre) One price however is an outlier at 179.2 which produces a range of 11.7 cents per litre.

Question: 16.

Discuss the distribution of the data and its practical implications.

Answer: One possible scenario for the practical interpretation of this 'bactrian' (two humped) distribution is that one collection of stores provides a common customer loyalty discount of 4 cents per litre. Other 'cheaper' outlets offer no loyalty discounts. The result would see a merging of the two 'humps'.

Another possibility is that one group of service stations is in the upside of the discounting cycle whilst the others have already commenced discounting. (Or the reverse)

The following questions relate to the Stem and Leaf plot for the LPG data.

The LPG data represents pricing information collected for a series of service stations located within a 6km radius of a particular suburb in Melbourne in February 2020. Run the stemleaf program: stemleaf(lpg,0).

Question: 17.

What do you notice about the distribution of LPG prices compared with ULP?

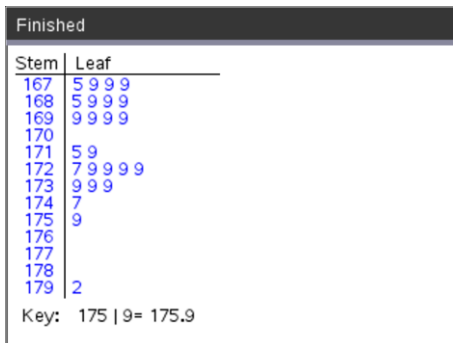
Answer: The distribution of LPG prices also has two 'humps' separated by a difference of approximately 4cents. The range is also very similar (11cents) and two relative (not quantified) 'outliers' of 68.9 and 69.9.

Teacher Note: The reference to outliers here is not quantified. If students use a boxplot or calculate $1.5 \times \text{IQR}$ they will see that these points are not quantified outliers, however they certainly stand out amongst the other service station prices and would be noticed by passing motorists.

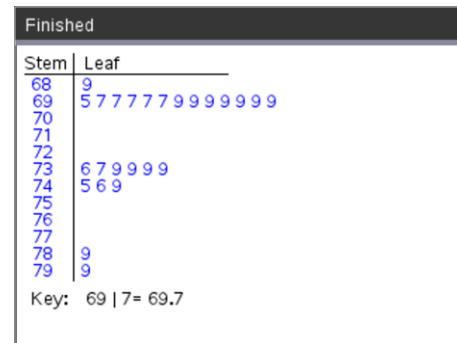
Question: 18.

Which fuel type offers the more consistent or predicible fuel price, LPG or ULP?

Answer: Statistically speaking the two distributions are very similar (Standard deviation, IQR and Range are all similar) however they hover around different median (and mean) prices. The Stem and Leaf plot for the LPG appears to indicate more consistency in pricing indicated by the longer leaves at 69 and 73cents per litre.



ULP Pricing – Stem and Leaf Plot



LPG Pricing – Stem and Leaf Plot

The following questions relate to the Stem and Leaf plot for the AFL data.

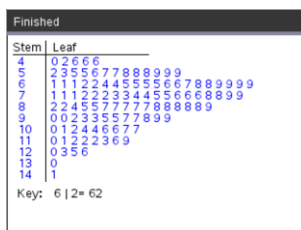
The AFL1, AFL2, AFL3 and AFL4 data sets relate to full time scores for every team in every match across the 2019 football season. AFL1 represents the first 6 rounds, AFL2 the second 6 rounds and so on.

Question: 19.

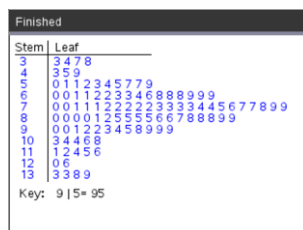
Explore all four data sets using a Stem and Leaf plot and **comment** on the spread of scores throughout the 2019 season. Explain why the entire season is not represented on a single Stem and Leaf plot and identify any other information from the data.

Suggested Answers and Discussion Points:

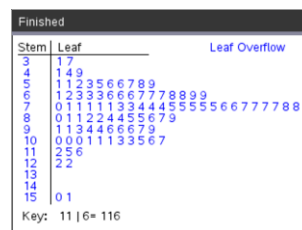
- **Too much data for a single Stem and Leaf plot.** Helps students identify the limitations of Stem and Leaf plots in comparison with histograms which would be better for an entire season's data.
- **The AFL3 data set includes an 'over-flow' of data in one leaf.** Another example of the calculator program teaching students that the size of the leaf should be limited and be used to inform the size of the Stem increments.
- **The four Stem and Leaf plots (below) show the score progression throughout the 2019 AFL season.** The most noticeable difference is the range of scores in the last quarter of the season. Scores ranged from a low of 14 to a high of 144, the spread (standard deviation) is also a little higher than for any other period.



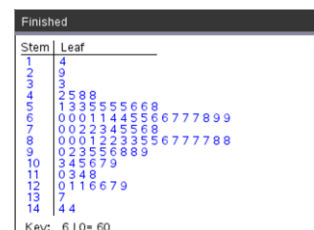
$\bar{x} \approx 81.5$
 $s_{n-1} \approx 21.6$
 $Q_1 \approx 65$
 $Q_2 \approx 78$
 $Q_3 \approx 97$



$\bar{x} \approx 79.2$
 $s_{n-1} \approx 22.6$
 $Q_1 \approx 64$
 $Q_2 \approx 76.5$
 $Q_3 \approx 91$



$\bar{x} \approx 79.6$
 $s_{n-1} \approx 21.4$
 $Q_1 \approx 67$
 $Q_2 \approx 76.5$
 $Q_3 \approx 94$



$\bar{x} \approx 81.2$
 $s_{n-1} \approx 26.2$
 $Q_1 \approx 64$
 $Q_2 \approx 80.0$
 $Q_3 \approx 98$

