

Chapter 196

Stem-and-Leaf Plots

Introduction

This procedure generates a stem-and-leaf plot of a batch of data. The stem-and-leaf plot is similar to a histogram and its main purpose is to show the data distribution while retaining the uniqueness of each data value. It is recommended for batches of data containing between 15 and 150 data points.

If the underlying data values have more digits than can be displayed on the chart, the values are truncated, not rounded. For example, 29.87 would be truncated to 29 if two digits were needed or 29.8 if three digits were needed.

Stem-and-Leaf Plot of SepalLength

Count	Stem	Leaves
6	7 (5-9)	677779
7	7 (0-4)	0122234
22	6 (5-9)	555556677777778889999
32	6 (0-4)	0000001111112222333333334444444
31	5 (5-9)	55555556666667777777888888999
30	5 (0-4)	00000000001111111122223444444
18	4 (5-9)	566667788888999999
4	4 (0-4)	3444

Data Structure

The data are contained in a single column.

Height dataset (subset)

Height
64
63
67
.
.
.

Procedure Options

This section describes the options available in this procedure. To find out more about using a procedure, turn to the Procedures chapter.

Following is a list of the procedure's options.

Variables Tab

The options on this panel specify which variables to use.

Data

Data Variable(s)

Specify a list of one or more variables for which stem-and-leaf plots are to be generated. You can double-click the field or single click the button on the right of the field to bring up the Variable Selection window.

Frequency Variable

This optional variable specifies the number of observations that each row represents. When omitted, each row represents a single observation. If your data is the result of a previous summarization, you may want certain rows to represent several observations. Note that negative values are treated as a zero weight and are omitted. This is one way of weighting your data.

Break Variables

Break Variables

You can specify up to five categorical-break variables. When one or more of these are specified, a separate plot is generated for each unique set of values for these variables.

Stem-and-Leaf Options

Range Multiplier

The data values are split into two groups: regular values which shown in the stem-and-leaf plot and outliers which are shown at the beginning and end of the plot. This value determines how many of the data values are classified as regular values. This is accomplished as follows.

1. Find the upper and lower hinges (L and U) which are roughly the 25th and 75th percentiles.
2. Calculate $IQR = U - L$.
3. Calculate the upper and lower adjacent values (U_a and L_a) which are the data values just less than $U + R \times IQR$ and just greater than $L - R \times IQR$.
4. The stem-and-leaf plot is constructed on those values between L_a and U_a . Any other values are treated separately as outliers.

Common values are 1.5 to 3. If you want to include most all of the data, set this value to 9 or 10.

Stem-and-Leaf Plots

Number of Plot Lines

This is the number of lines in the stem-and-leaf plot. That is, it is the number of print lines.

You can vary this number to lengthen or shrink the plot.

Auto

If this value is set to 'Auto', the number of lines is calculated based on the number of observations.

Input Range

Typical values other than 'Auto' range from 10 to 50.

Branches per Stem

The leaves of the plot are formed from the integers from 0 to 9. When a line becomes too long to display (e.g. if there are more than 70 leaves), the stem may be split into branches. The first branch contains leaves made up of the digits 0 through 4 and the second branch contains leaves made up of the digits 5-9. This option controls the number of branches that are displayed for each stem.

For example, the line (branches = 1)

```
2 | 000112222444444555677778899
```

could be split into two branches

```
2 | 0001122224444444
2 | 555677778899
```

or five branches

```
2 | 00011
2 | 2222
2 | 444444555
2 | 67777
2 | 8899
```

Max Leaves per Line

Only about 70 leaves can be displayed on a single line. If more than this number occur, the line is truncated to this number of leaves and the truncated is displayed.

For example, the line

```
2 | 00011222244444445556 + 5
```

means that the last five leaves were not displayed.

Branch Label

You can include a special label when stems are split into 2 or 5 branches. These labels the digits contained on that line. The following options are available.

- **Omit**

The label is not shown.

- **Range (X-Y)**

The range of digits is shown. For example,

```
2 (0-4) | 000112222444444
2 (5-9) | 555677778899
```

Stem-and-Leaf Plots

- **Symbol Set 1**

A single character is used. The first letter of the first digit is used. For example, 'Z' for zero or 'F' for five.

2 Z | 000112222444444

2 F | 555677778899

- **Symbol Set 2**

A single character is used. The character may be the first letter of the all digits in line or a symbol such as '!' or

'*'.
.

Orientation

Specify the whether the plot goes from low to high (Low on top) or from high to low (Low on bottom).

Bar Symbol

Specify the symbol used to separate the stem from the leaves. Usually, the vertical bar '|' is used.

Outlier Symbol

Specify the symbol used to represent outliers.

Show Counts

Choose whether or not to display the 'Counts' column on the report. The count is the number of leaves contained on that line.

Show Depth Values

Choose whether or not to display the 'Depth' column on the report. The depth is the number cumulative total of the counts from the nearest edge to the center.

Show Outliers

Choose whether or not to display the 'Outliers' rows on the report. The outliers are displayed on rows labeled '<' and '>'.

Show Rows with No Leaves

Choose whether or not to display lines that contain no data (no leaves). This is sometimes called pruning.

Reports Options Tab

The options on this panel control the format of the report.

Report Options

Value Labels

This option applies to the *Break Variable(s)*. It lets you select whether to display data values, value labels, or both. Use this option if you want the output to automatically attach labels to the values (like 1=Yes, 2=No, etc.). See the section on specifying *Value Labels* elsewhere in this manual.

Variable Names

This option lets you select whether to display only variable names, variable labels, or both.

Example 1 – Create a Stem-and-Leaf Plot

This section presents an example of how to create a stem-and-leaf plot on the *SepalLength* variable in the *Fisher* dataset. To run this example, do the following or simply load the completed template **Example 1** by clicking on Open Example Template from the File menu of the procedure window.

1 Open the Fisher dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Fisher.NCSS**.
- Click **Open**.

2 Open the Stem-and-Leaf window.

- Using the Graphics menu or the Procedure Navigator, find and select the **Stem-and-Leaf Plots** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the SepalLength variable.

- On the Stem-and-Leaf window, select the **Variables tab**. (This is the default.)
- Double-click in the **Data Variable(s)** text box. This will bring up the variable selection window.
- Select **SepalLength** from the list of variables and then click **Ok**. The word “SepalLength” will appear in the Data Variables box. Note that you could have entered a “1” here signifying the first (left-most) variable on the dataset.

4 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

The following reports and charts will be displayed in the Output window.

Stem-and-Leaf Plot

Stem-and-Leaf Plot of SepalLength

Count	Stem	Leaves
1	7	9
5	7	67777
1	7	4
4	7	2223
2	7	01
7	6	8889999
10	6	667777777
12	6	444444455555
13	6	2222333333333
12	6	00000111111
10	5	8888888999
14	5	6666667777777
13	5	4444444555555
5	5	22223
19	5	00000000011111111
11	4	88888999999
6	4	666677
4	4	4445
1	4	3

Stem-and-Leaf Plots

Stem-and-Leaf Parameters

Parameter	Value
Example	'5 36' represents the values 53 and 56.
Stem Units	10
Leaf Units	1
Truncation to Nearest	1
Range Multiplier	1.5
Low Outliers	None
High Outliers	None

The stem-and-leaf plot is a type of histogram which retains much of the identity of the original data. It is useful for finding data-entry errors as well as for studying the distribution of a variable.

Count

This is the number of leaves on the line.

Depth

This is the cumulative number of leaves, counting in from the each end.

Stem

The stem is the first digit of the actual number. For example, the stem of the number 523 is 5 and the stem of 0.0325 is 3. This is modified appropriately if the batch contains numbers of different orders of magnitude. The largest order of magnitude is used in determining the stem. Depending upon the number of leaves, a stem may be divided into two or more branches. A special set of branch labels is then used to mark the branches.

Leaf

The leaf is the second digit of the actual number. For example, the leaf of the number 523 is 2 and the leaf of 0.0324 is 2. This is modified appropriately if the batch contains numbers of different orders of magnitude. The largest order of magnitude is used in determining the leaf.

Stem-and-Leaf Parameters

This report at the bottom shows various parameters used in constructing the plot.