

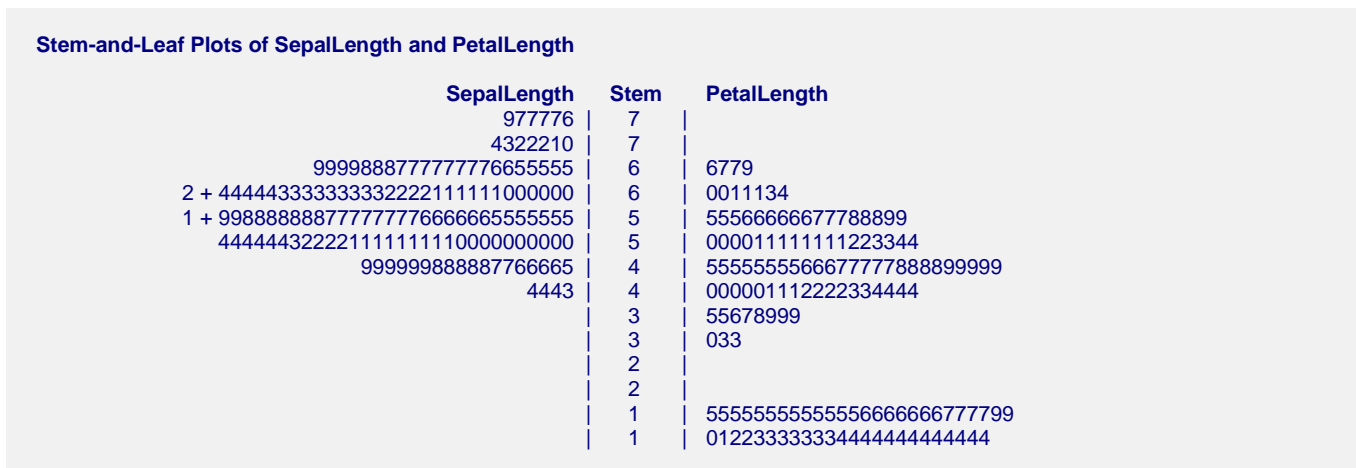
## Chapter 195

# Back-to-Back 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.



### Data Structure

The data are contained in a two columns or a single column with another grouping column.

#### Fruit Yied

Apple	Pear
75	69
63	78
64	59
77	68
55	64
	72

## 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.

#### Input Type

In this procedure, there are three ways to organize the data for constructing stem-and-leaf plots. Select the type that reflects the way your data is presented on the spreadsheet.

- **Response Variable(s) and Group Variable(s)**

In this scenario, the response data is in one column and the groups are defined in another column of the same length.

Response	Group
12	1
35	1
19	1
24	1
26	2
44	2
36	2
33	2

If multiple response variables are chosen, a separate plot is made for each response variable. If the group variable has more than two levels, a plot is made among each pair of levels. If multiple group variables are entered, a separate plot is made for each pair of each combination of group levels.

- **Two Variables with Response Data in each Variable**

In this selection, the data for each group is in a separate column. You are given two boxes to select the Group 1 data variable and the Group 2 data variable.

Group 1	Group 2
12	26
35	44
19	36
24	33

**Back-to-Back Stem-and-Leaf Plots**

- **More than Two Variables with Response Data in each Variable**

For this choice, the data for each group is in a separate column. Each variable listed in the variable box will be compared to every other variable in the box.

Group 1	Group 2	Group 3
12	26	25
35	44	29
19	36	41
24	33	28

---

### Variables – Input Type: Response Variable(s) and Group Variable(s)

Response	Group
12	1
35	1
19	1
24	1
26	2
44	2
36	2
33	2

**Response Variable(s)**

Specify the variable(s) containing the response data. For this input type, the response data is in this column and the groups are defined in another column of the same length.

If more than one response variable is entered, a separate plot is produced for each variable.

**Group Variable(s)**

Specify the variable(s) defining the grouping of the response data. For this input type, the group data is in this column and the response data is in another column of the same length.

If the group variable has more than two levels, a plot is made among each pair of levels. If multiple group variables are entered, a separate plot is made for each pair of each combination of group levels. The limit to the number of Group Variables is 5.

---

### Variables – Input Type: Two Variables with Response Data in each Variable

Group 1	Group 2
12	26
35	44
19	36
24	33
27	
32	

**Group 1 Variable**

Specify the variable that contains the Group 1 data. For this input type, the data for each group is in a separate column. The number of values in each column need not be the same.

**Back-to-Back Stem-and-Leaf Plots****Group 2 Variable**

Specify the variable that contains the Group 2 data.

---

**Variables – Input Type: More than Two Variables with Response Data in each Variable**

Group 1	Group 2	Group 3
12	26	25
35	44	29
19	36	41
24	33	28
32		

**Response Variables**

Specify the variable(s) containing the response data.

For this choice, the data for each group is in a separate column. Each variable listed in the variable box will be plotted against every other variable in the box.

---

**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 (Ua and La) 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 La and Ua. 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.

**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 35 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.

## Back-to-Back Stem-and-Leaf Plots

For example, the line (branches = 1)

```
2 | 000112222444444555677778899
```

could be split into two branches

```
2 | 000112222444444
```

```
2 | 555677778899
```

or five branches

```
2 | 00011
```

```
2 | 2222
```

```
2 | 444444555
```

```
2 | 67777
```

```
2 | 8899
```

### Max Leaves per Line

Only about 35 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 number truncated is displayed.

For example, the line

```
2 | 0001122224444445556 + 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
```

- **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.

## Back-to-Back Stem-and-Leaf Plots

### 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 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.

---

### Decimal Places

#### Decimals

This option allows the user to specify the number of decimal for outlier places directly or based on the significant digits. These decimals are used for output except the nonparametric tests and bootstrap results, where the decimal places are defined internally.

If one of the significant digits options is used, the ending zero digits are not shown. For example, if 'Significant Digits (Up to 7)' is chosen,

0.0500 is displayed as 0.05

1.314583689 is displayed as 1.314584

The output formatting system is not designed to accommodate 'Significant Digits (Up to 13)', and if chosen, this will likely lead to lines that run on to a second line. This option is included, however, for the rare case when a very large number of decimals is needed.

## Example 1 – Create a Back-to-Back Stem-and-Leaf Plot

This section presents an example of how to create a back-to-back 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 **Back-to-Back 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.

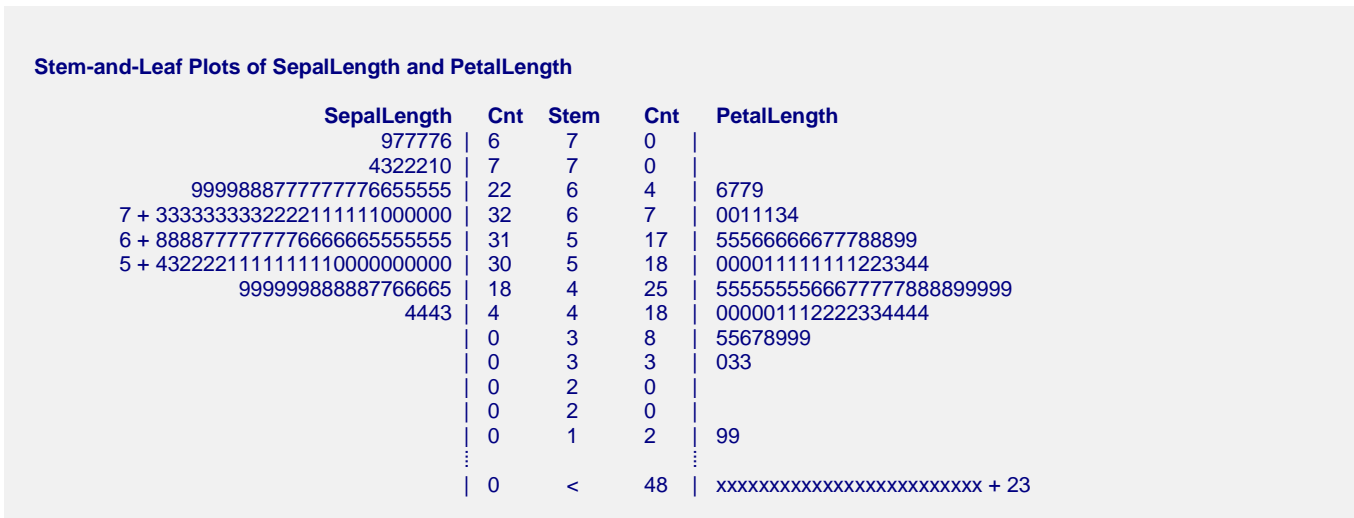
- On the Stem-and-Leaf window, select the **Variables tab**. (This is the default.)
- For **Input Type**, select **Two Variables with Response Data in each Variable**.
- Double-click in the **Group 1 Variable** text box. This will bring up the variable selection window.
- Select **SepalLength** from the list of variables and then click **Ok**. The word “SepalLength” will appear in the Data Variables box.
- Double-click in the **Group 2 Variable** text box. This will bring up the variable selection window.
- Select **PetalLength** from the list of variables and then click **Ok**. The word “PetalLength” will appear in the Data Variables box.

### 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



## Back-to-Back 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
Truncated to Nearest	1
Range Multiplier	1.5
Group 1	SepalLength
Low Outliers	None
High Outliers	None
Group 2	PetalLength
Low Outliers	10, 11, 12, 12, 13, 13, 13, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17
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.

### Counts

These are the number of leaves in each batch (variable or group) of data.

### 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 (Variable, or Group, Name)

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 shows various parameters used in constructing the plot.