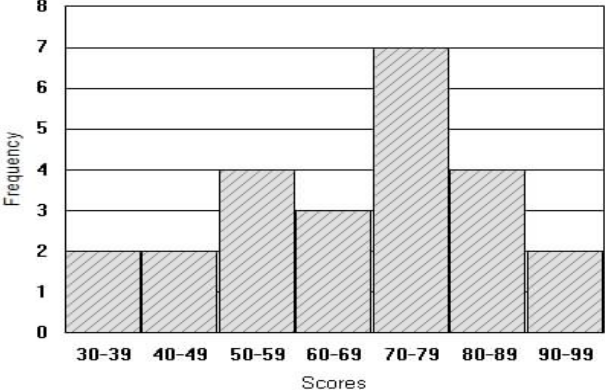# 2.3 Stem and Leaf Diagram

Another useful technique for displaying data that is an alternative to the frequency distribution is the *stem and leaf diagram*. The *stem and leaf* display, requires that each raw score or data value be separated into two parts: The first digit (or digits) is called the stem, and the last digit (or digits) is called the *leaf*. For example, the score $X = 65$ would have a stem of 6 and a leaf of 5. Similarly, $X = 97$ would have a stem of 9 and a leaf of 7.

> A ***stem-and-leaf diagram*** is a graphical display presenting the original data arranged into a histogram.
>
> The **stem**: the vertical axis of display containing the leading digit(s).
>
> The **leaves**: the horizontal axis of a display containing the trailing digits.

The following procedure is a guide to construct a stem-and-leaf display. Notice how the stem and leaf compare to a grouped frequency histogram.
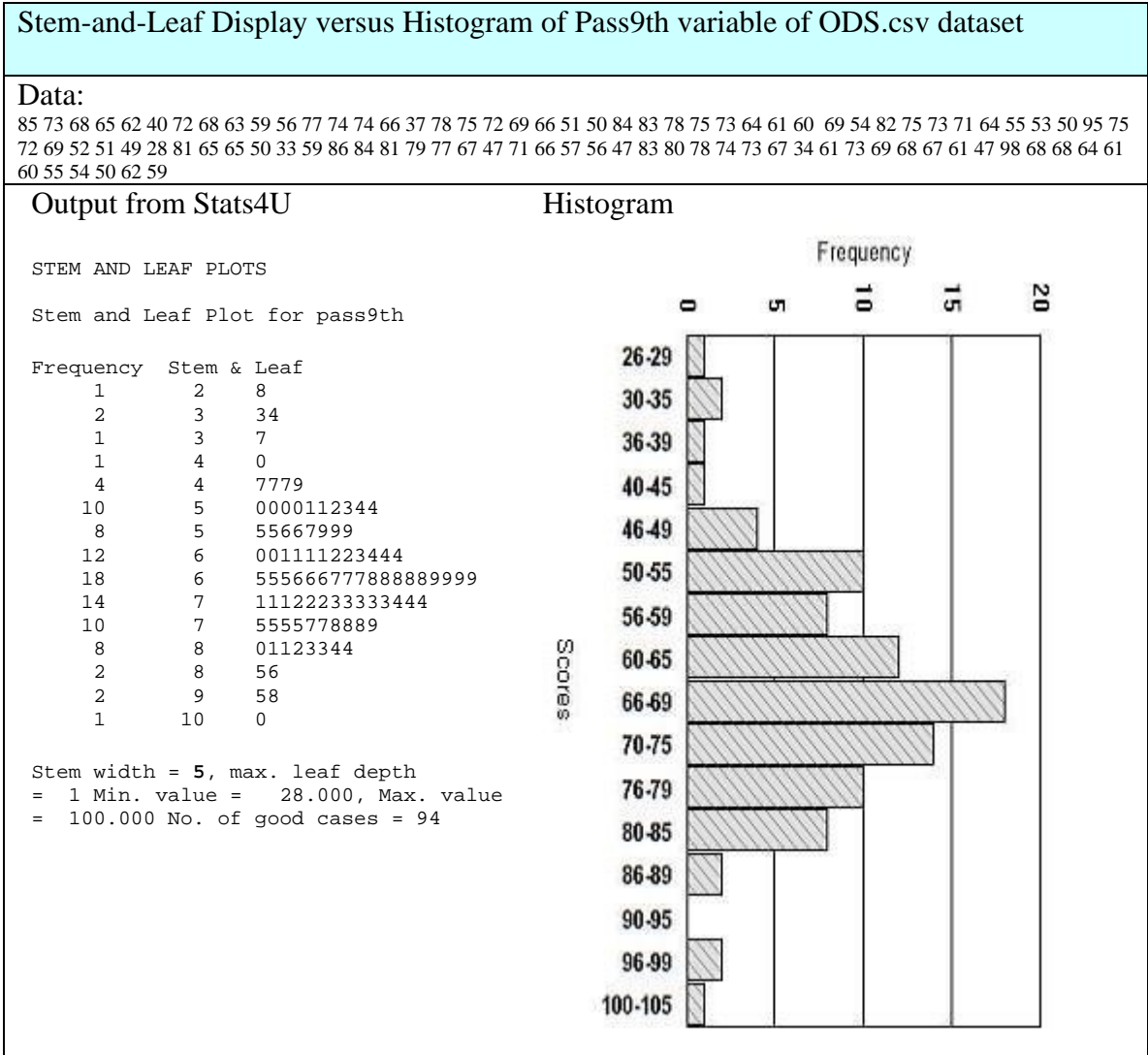
| Construction Guide: Stem-and-Leaf Display |
| --- |
| Data<br>83  62  71  76  85  32  56  74  82  93  68  52  42  57  73  81  63  78  33  97  46  59  74  76 |
| Step 1: List all stem in a column and<br>Step 2 Go thru each score and write the leaf next to its stem |

**Stem and Leaf Display**

3 | 23
4 | 26
5 | 6279
6 | 283
7 | 1643846
8 | 3521
9 | 37

**Frequency Histogram**

The stem and leaf display allows the converting of the display or diagram back to the original scores and frequencies. It maintains the original data values. Typically one group stems by 10 or intervals of 10; however, like a frequency distribution data can be grouped into any convenient interval size, say width of 5.

Below is the stem-and-leaf display for the *Pass9th* variable of the ODS.cs dataset:

| Stem-and-Leaf Display of Pass9th variable of ODS.csv dataset |
|---|
| Data:<br>85 73 68 65 62 40 72 68 63 59 56 77 74 74 66 37 78 75 72 69 66 51 50 84 83 78 75 73 64 61 60  69 54 82 75 73 71 64 55 53 50 95 75 72 69 52 51 49 28 81 65 65 50 33 59 86 84 81 79 77 67 47 71 66 57 56 47 83 80 78 74 73 67 34 61 73 69 68 67 61 47 98 68 68 64 61 60 55 54 50 62 59 |

```
Output from Stats4U

STEM AND LEAF PLOTS

Stem and Leaf Plot for variable: pass9th

Frequency   Stem & Leaf
     1          2     8
     2          3     34
     1          3     7
     1          4     0
     4          4     7779
    10          5     0000112344
     8          5     55667999
    12          6     001111223444
    18          6     555666777888889999
    14          7     11122233333444
    10          7     5555778889
     8          8     01123344
     2          8     56
     2          9     58
     1         10     0

Stem width = 5 (manually imputed), max. leaf depth =   1
Min. value =    28.000, Max. value =   100.000
No. of good cases = 94
```

Below is the stem-and-leaf display and corresponding histogram for the *Pass9th* variable of the ODS.cs dataset. The histogram was constructed using MS Excel.

| Stem-and-Leaf Display versus Histogram of Pass9th variable of ODS.csv dataset |
|---|

**Data:**
85 73 68 65 62 40 72 68 63 59 56 77 74 74 66 37 78 75 72 69 66 51 50 84 83 78 75 73 64 61 60  69 54 82 75 73 71 64 55 53 50 95 75 72 69 52 51 49 28 81 65 65 50 33 59 86 84 81 79 77 67 47 71 66 57 56 47 83 80 78 74 73 67 34 61 73 69 68 67 61 47 98 68 68 64 61 60 55 54 50 62 59

**Output from Stats4U**                                          **Histogram**

```
STEM AND LEAF PLOTS

Stem and Leaf Plot for pass9th

Frequency   Stem & Leaf
    1         2     8
    2         3     34
    1         3     7
    1         4     0
    4         4     7779
   10         5     0000112344
    8         5     55667999
   12         6     001111223444
   18         6     555666777888889999
   14         7     11122233333444
   10         7     5555778889
    8         8     01123344
    2         8     56
    2         9     58
    1        10     0


Stem width = 5, max. leaf depth
=  1 Min. value =    28.000, Max. value
=  100.000 No. of good cases = 94
```



Tony and Stem and Leaf Display

**Tony:** "It seems so simple to generate the stem and leaf display, so why bother with frequency and histogram distributions which then to hide the original data?"

**Rose:** "The stem and leaf display was conceived back in the late 1970's (by John Turkey in 1977) and was created long before we could draw more sophisticated graphs and diagrams with the aid of a computer; such computer generated outputs offers more insight into the data distributions than this crude method does."