

Statistics is the collection, organization and study of numerical data.

Data may be collected in a unorganized or random manner. For example, we may want to record how many times each student in a class of 18 students rode the bus to school last week. The data may have been recorded as follows:

5, 0, 3, 4, 4, 5, 5, 3, 3, 1, 2, 5, 2, 3, 5, 1, 5, 5

The above data could be organized into a table as follows:

| Number of Bus Rides | Tally | Frequency |
|---------------------|-------|-----------|
| 0                   |       | 1         |
| 1                   |       | 2         |
| 2                   |       | 2         |
| 3                   |       | 4         |
| 4                   |       | 2         |
| 5                   |       | 7         |
|                     |       |           |

The tally column (column 2) in the table above contains a tally mark corresponding to the **number of bus rides** (column 1) each of the 18 students recorded for last week. For example, four students recorded they rode the bus to school three times last week. The frequency column (column 3) is just the count of tally marks in each of the rows of column 2.

With the data organized in this manner, we can more readily answer questions such as the following:

1. How many students rode the bus at least 3 times last week?
2. How many students did not ride the bus at all last week?
3. What percent of the students rode the bus all five days last week?

## Stem-and-Leaf Diagram

A Stem-and-Leaf diagram is another way of organizing collected data and still be able to preserve the individual data values. To illustrate this type of organization, a class of 21 students was surveyed to find out how many minutes each student spent surfing the internet yesterday. The data was recorded in a random fashion as follows:

29, 10, 35, 24, 30, 10, 5, 22, 31, 8, 40, 45, 41, 32, 16, 7, 39, 15, 8, 17, 0

Since the data values are one or two digit numbers, the tens position of the numbers was chosen to be the stem and the units position of the numbers are the attached leaves. So, the stem-and-leaf diagram would look as follows:

| Stems | Leafs     |
|-------|-----------|
| 0     | 0 5 7 8 8 |
| 1     | 0 0 5 6 7 |
| 2     | 2 4 9     |
| 3     | 0 1 2 5 9 |
| 4     | 0 1 5     |

Notice that by putting the stem and leafs together in the first row of the above table, we have the numbers: 00 05 07 08 08. When we put the stem and leafs together in the second row of the table, we have the numbers: 10 10 15 16 17. By continuing the process for the rest of the rows of the table, we see that all of the original data values have been preserved. The data in the table has been organized from smallest value (00) to highest value (45). Thus, the table lends itself to easy analysis.

## Grouped Data

When there is a large number of data values collected, grouping may be the best option for organizing and analyzing the data. To group data, non-overlapping intervals are created to cover the full range of the data values. To illustrate this type of organization of data, a teacher recorded the following test grades of 30 students:

50, 65, 40, 85, 100, 92, 38, 66, 81, 90, 95, 55, 90, 78, 100, 80, 80, 70, 55, 42, 77, 20, 63, 75, 85, 90, 30, 58, 82, 100

Since the smallest grade recorded is a 20 and the largest grade is 100, the intervals must cover the range from 20 to 100. The original data value must fall within only one interval. Thus, the intervals should be non-overlapping and of equal width. For this data, an interval width of 10 would be a good choice. The table below shows the data organized by interval grouping.

| Test Score Intervals | Tally | Frequency |
|----------------------|-------|-----------|
| 91-100               |       | 5         |
| 81-90                |       | 7         |
| 71-80                |       | 5         |
| 61-70                |       | 4         |
| 51-60                |       | 3         |
| 41-50                |       | 2         |
| 31-40                |       | 2         |
| 21-30                |       | 1         |
| 11-20                |       | 1         |

The data in the above table can also be displayed by a bar graph called a histogram.

