# Unit 1 Glossary Terms

## Categorical data

Categorical data are data where the values are categories. For example, the breeds of 10 different dogs are categorical data.

Another example is the colors of 100 different flowers.

## Non-statistical question

A non-statistical question is a question which can be answered by a specific measurement or procedure where no variability is anticipated, for example:
- How high is that building?
- If I run at 2 meters per second, how long will it take me to run 100 meters?

## Numerical data

Numerical data, also called measurement or quantitative data, are data where the values are numbers, measurements, or quantities. For example, the weights of 10 different dogs are numerical data.

## Statistical question

A statistical question is a question that can only be answered by using data and where we expect the data to have variability, for example:
- Who is the most popular musical artist at your school?
- When do students in your class typically eat dinner?
- Which classroom in your school has the most books?

## Distribution

For a numerical or categorical data set, the distribution tells you how many of each value or each category there are in the data set.

## five - number summary

The five-number summary of a data set consists of the minimum, the three quartiles, and the maximum. It is often indicated by a box plot like the one shown, where the minimum is 2, the three quartiles are 4, 4.5, and 6.5, and the maximum is 9.
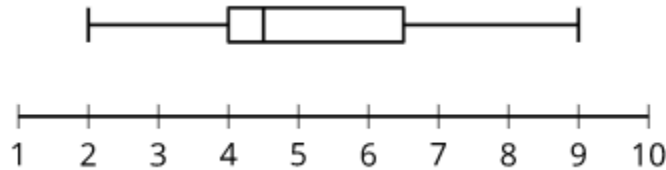
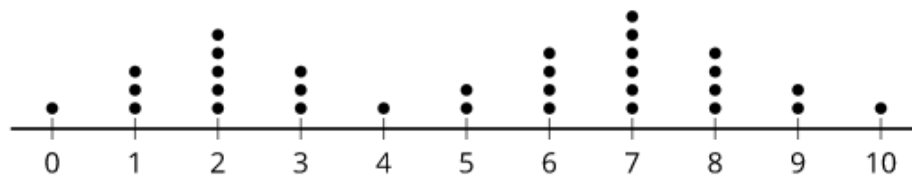## Bell-shaped                                                    distribution

A distribution whose dot plot or histogram takes the form of a bell with most of the data clustered near the center and fewer points farther from the center.
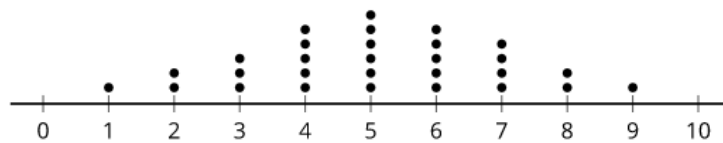
## Bimodal                                                        distribution

A distribution with two very common data values seen in a dot plot or histogram as distinct peaks. In the dot plot shown, the two common data values are 2 and 7,
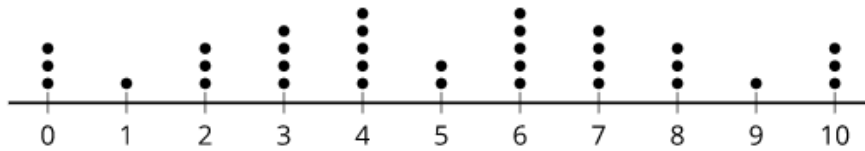
## Skewed distribution

A distribution where one side of the distribution has more values farther from the bulk of the data than the other side, so that the mean is not equal to the median. In the dot plot shown, the data values on the left, such as 1, 2, and 3, are further from the bulk of the data than the data values on the right.
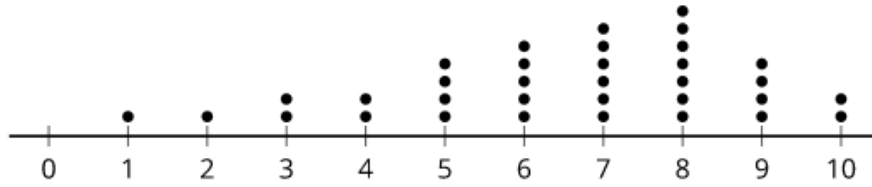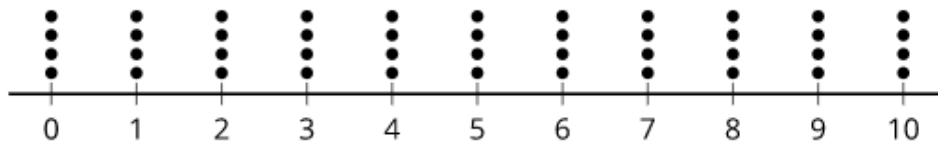
## Symmetric distribution

A distribution with a vertical line of symmetry in the center of the graphical representation, so that the mean is equal to the median. In the dot plot shown, the distribution is symmetric about the data value 5.



## Uniform distribution

A distribution which has the data values evenly distributed throughout the range of the data.



## statistic

A quantity that is calculated from sample data, such as mean, median, or MAD (mean absolute deviation).

## Standard deviation

A measure of the variability, or spread, of a distribution, calculated by a method similar to the method for calculating the MAD (mean absolute deviation). The exact method is studied in more advanced courses.

## outlier

A data value that is unusual in that it differs quite a bit from the other values in the data set. In the box plot shown, the minimum, 0, and the maximum, 44, are both outliers.