

As Tamara Munzner describes, data semantics and data types are important. In short, these are vital for humans' ability to understand the data. The semantics of the data is what the data represent in the real world. A table with a row containing 5 different numeric values is useless unless the person working on it knows what these numbers represent. The type of the data is equally important, as this is vital in understanding how to work with the given data. If it is a number, is it a quantity or an ID? This information is key, as adding two numeric values together makes sense for quantities, but would be useless for ID or code.

There are five basic data types: Items, attributes, links, positions and grids. An item is an individual entity, like a row in a table, since a single row in a table only describes one unique item in the given table. An attribute is a measurable property, such as the base salary in different professions, a person's height and so on. A link is the relationship between items in networks, an example being the relationships between people in a city, such that a link describes the relationship between the citizens (Neighbours, family, acquaintances, friends...). A position is spatial data, describing a location in two-dimensional or three-dimensional space. Here, a great example is the earth, which is divided into latitude and longitude coordinates, such that every point on earth has positional data, or a coordinate. Finally, a grid describes the strategy for sampling continuous data in geometric and topological relationship between its cells.

The four basic dataset types are tables, networks, fields and geometry. Other possibilities include clusters, sets and lists. However, it is common for these basic types to be combined in real-world situations. Following will be a short description of each dataset type. In a table, each row represents an item of data, and each column is an attribute of the dataset. Each individual cell contains a value of the pair of the item and the attribute.

Networks and Trees specify the relationship between two or more items. An item in a network is commonly referred to as a node, and a link is the relation between two nodes. A network with a hierarchical structure is called a Tree. Unlike general networks, trees don't have cycles, as every child has only one node pointing to it, referred to as its parent node.

The field dataset type consists of cells. It is similar to tables, as it is build by keys and attributes. The main difference between fields and tables is that fields are used for continuous data. Each cell contains measurements from a continuous domain. In a way, there is an infinite amount of values to measure. Examples of this in the real world would be measuring the temperature. Finally, the geometry dataset type describes information about the shape of items. The items can range from one-dimensional lines to three-dimensional reconstructions of buildings and landmarks.