## Tabular/Columnar Relationship

This is the simplest data structure and is one with which most people are familiar thanks to the ubiquity of Microsoft’s “Excel” application (which is widely regarded as the first “killer” app.) The figure below, which is a snapshot weather data collected from National Oceanic and Atmospheric Administration (NOAA) station 46237 just outside the entrance to San Francisco Bay, provides a good example of tabular data relationship.

![](`r url`)

Source: [National Oceanographic and Atmospheric Administration]( https://www.ndbc.noaa.gov/)

The organizing principal behind this data is quite simple. All of the data contained in the table came from the same source and each observation (each row of the table) is an observation of the same set of variables (each column). Additionally, note that the values for each variable are all continuous. The are no categorical values in the table, such as assigning each observation to a category like “morning, afternoon, evening, or night.” It is common for people to include categorical variables in data that is stored in tabular form (such as, for example, City names, etc.). This is not incorrect from a data storage perspective, but from a conceptual, data structure perspective, this means that the data is relational and not tabular. (We’ll discuss relational data in a moment). Purely tabular defines no categorial or other relationships between individual points of data in the data itself. The table structure itself provides the only relational information.

## Tree Data

Tree data is data that has an inherent tree structure. As noted earlier, data on a family tree is a good example of tree data.

![](`r url2`)

Source: Lobsterthermidor (talk) 17:38, 4 December 2018 (UTC), CC BY-SA 3.0, via [Wikimedia Commons](https://creativecommons.org/licenses/by-sa/3.0).

By definition, trees have a root (the earliest known ancestor) level and branches that grow in an hierarchical manner with additional, identifiable levels. Note that nested data, with levels that move out from a comment center is also tree data, with the central node serving as the root.

## Graph Data

Graph data is similar to tree data, in that it represents a collection of entities that are related to each other in a particular way. What differentiates graphs from trees is that, whereas trees are hierarchical in nature, beginning with a root and adding elements across layers of connection, graphs are \*rhizomatic\* in nature, meaning there is no root element, and elements, or nodes as they are known in graph theory, can connect to other nodes without implying any hierarchy or levels of connection. This does not mean that there can be no temporality or direction to the connections (or “edges” as they are called in graph theory) between nodes, only that they are not baked into the data structure itself and exist simply as additional data in the dataset.

![](`r url3`)

Source: Mssemantics, CC BY 4.0, via [Wikimedia Commons](https://creativecommons.org/licenses/by/4.0).

## Relational Data

From one perspective, nearly all data is relational, in that some relationship exists between the various observations that are captured in the dataset. When we speak of relational data, however, we are speaking of data where the categorial relationships between individual observations is integral both to how the things being observed exist in the real world and how we, as humans, understand them. A good example of relational data is a grocery store (real or virtual) inventory. The inventory, of course, contains individual, discrete items. By they are always arranged in the storefront according to categorical associations such as produce, canned food, etc. In this case, the category to which a product belongs is frequently more important than the actual item. You don’t go the “beans” section of the grocery store to find every type of bean. You go to the canned food section, then the bean section, then find your particular bean. Additionally, in these datasets, the categorical variables are frequently more important than the individual observations. A shopper may care if the store carried canned black beans but almost never cares about a particular can of black beans. With relational data the categorical relationships between items are the central, organizing principal of the dataset.

![](`r url4`)

Source: [iNetTutor.com]( https://www.inettutor.com/source-code/online-grocery-shopping-system/).

The above image presents an Entity Relationship Diagram (ERD) for a simplified grocery store inventory data store that is relational in nature. Note that the entire design of the database is built around the idea of creating separate tables for each category of information that will be collected and managed and then defining the relationships between categories of information at the data structure level. These relationships are then inherited automatically by any data that is placed into the data store. You can determine if your data is inherently relational by considering how many categorical variables you have across the dataset. If either the majority or a collection of the most significant variables in the data are categorical, then the dataset as a whole is inherently relational.

## Bags of Data

Sometimes we make decisions for computational reasons to discard any structure inherent in our data prior to storage and analysis. A good example of this the frequently employed Bag of Words approach in various forms of Natural Language Processing. By definition, readable texts are highly structured. They following a complex set of grammatical rules and words are also presented in a particular order. Many natural language processes, however, completely ignore this structure and instead treat each text as bag filled with random word tokens.

![](`r url4`)

Source: [Force of LSTM and GRU Blog]( https://dudeperf3ct.github.io/lstm/gru/nlp/2019/01/28/Force-of-LSTM-and-GRU/)

It is worth noting that, prior to computation, even these bags must be given some structure in order to store them, load them into computer memory, and act upon them. But these structures are completely unrelated to the structure that was inherent in the original data. Thus, we consider to the data be initially structureless from the perspective of the data itself.