# Structures Implied by Research Questions

The structures inherent in a dateset itself are just one thing a researcher should consider when determining the correct data storage strategy. Different research questions are facilitated by different data structures.

Imagine, for example, that you are a sociologist who is primarily concerned with how different class associations affect human behavior. In this case, a data structure that captures the various class associations of individuals observed in a structured manner (relational) would facilitate your research better than one that is individual focused (tabular) or hierarchical (tree). If, on the other hand, you are an economist studying the productivity of individual workers at different echelons of the corporate management structure, an hierarchical data structure (tree) would better suit your needs.

It is important that you, as a researcher, understand that the nature of your research question(s) should inform your data storage decisions as much, if not more, than any other factor. At times, the nature of your research question may be the most significant factor in your choice data store. For example, researchers who employ computational Topic Modeling frequently discard the native structure of the texts they study in favor of a Bag of Data storage model because it facilitates computation. 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/)

The “bags of words” must, however, 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 and are instead organized to facilitate the computational algorithms that will act upon them.