**Data Quality Dimensions**

Data quality dimensions help guide your thought process while assessing and also cleaning. The four main data quality dimensions are:

* **Completeness**: do we have all of the records that we should? Do we have missing records or not? Are there specific rows, columns, or cells missing?
* **Validity**: we have the records, but they're not valid, i.e., they don't conform to a defined schema. A schema is a defined set of rules for data. These rules can be real-world constraints (e.g. negative height is impossible) and table-specific constraints (e.g. unique key constraints in tables).
* **Accuracy**: inaccurate data is wrong data that is valid. It adheres to the defined schema, but it is still incorrect. Example: a patient's weight that is 5 lbs too heavy because the scale was faulty.
* **Consistency**: inconsistent data is both valid and accurate, but there are multiple *correct* ways of referring to the same thing. Consistency, i.e., a standard format, in columns that represent the same data across tables and/or within tables is desired.

Regarding the other data quality research mentioned in the video, the additional dimensions are super specific cases of these four dimensions listed above. Example: *currency*, defined as follows: *the degree to which data is current with the world that it models. Currency can measure how up-to-date data is.* Currency is a specific case of accuracy data in the sense that out-of-date data is (usually) valid but wrong. In other words, our definition of accuracy can include currency.