1. This paper is about the data tidying component of data cleaning. It covers the importance of data tidying, and how to effectively carry out data tidying.
2. The purpose of a standard is to make initial data cleaning easier. The tidy data standard facilitates the exploration and analysis of data. By doing this, it simplifies the development of data analysis tools that work together.
3. The first sentence means a tidy dataset shares similarities with other tidy datasets because of how it is organized. However, a messy dataset can come in various forms, making every messy dataset unique. The second sentence points to how it is easy to identify individual observations, but being able to understand them on a general level is difficult.
4. A value comes in two forms. Quantitative values consist of numbers, and qualitative values consist of strings. A variable contains all values that measure the same attribute. Observations contain all values measured with the same unit across attributes.
5. In tidy data, each variable forms a column, each observation forms a row, and each type of observational unit forms a table.
6. The five most common problems are: column headers are values, not variable names, multiple variables are stored in one column, variables are stored in both rows and columns, multiple types of observational units are stories in the same table, a single observational unit is stored in multiple tables. Table 4 is messy because the column headers are values. Instead of values, they should be variable names. Melting a dataset means to stack it, which consists of turning the columns into rows.
7. Table 11 is messy because it is showing a lot of missing data points. Table 12 is tidy because it has dropped those missing values. The missing values are dropped to conserve space. In the molten weather dataset, the element column contains the names of variables.
8. The chicken and egg problem points out that if tidy data is only as useful as the tools that work with it, then tidy tools will be inextricably linked to tidy data. Wickham hopes that others will build on the framework and develop even better tools and data storage strategies.