1) This paper is about a study which aims to develop a “tidy” dataset which is easy to manipulate, model, and visualize by establishing each variable as a column, each observation as a row, and each type of observational unit as a table.

2) The “tidy data standard” is intended to provide a standard way to organize data values within a dataset so that during the initial phase of data analysis (data-cleaning), you don’t have to spend significant amounts of time just tidying or cleaning the data.

3) When we talk about datasets the sentence suggests that if datasets follow a format they have a consistent structure. On the other hand, datasets that are not tidy can be messy. Cause various issues, in different ways. In essence there is a way of organizing data (tidy). There are numerous ways in which data can be disorganized (messy).The next sentence implies that when you examine a dataset you can usually determine what constitutes observations (such as a single measurement or data point) and what comprises variables (like attributes or characteristics of the data). However if you attempt to establish a general definition of observations and variables to all datasets it becomes quite challenging. The reason for this challenge is that datasets can differ significantly in terms of their content, context and intricacy.

4) Wickham defines values as either numbers (quantitative) or strings (qualitative). A variable is classified as all values that measure the same underlying attribute across units, and an observation is defined as all values measured on the same unit across attributes.

5) Within section 2.3, “tidy data” is defined as the “standard way of mapping the meaning of a dataset to its structure.”

6) The 5 most common problems with messy datasets are the following:
    - 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 stored in the same table.
    - A single observational unit is stored in multiple tables.

The data in table 4 is messy because the table is a tabular data designed for presentation, where the variables from both the rows and the columns, and column headers are values not variable names. In simple terms, “melting” a dataset is simply turning columns into rows.

7) Table 11 is messy because it’s the original dataset containing even missing values. Table 12 is “tidy” and “molten” because missing values are dropped to conserve space, each row represents the meteorological measurements for a single day, and there are two measured variables and everything else is fixed.

8) The “chicken and egg” problem focusing on tidy data is that “if tidy data is only as useful as the tools that get to work with it, then tidy tools will be inextricably linked to tidy data.” He hopes that in the future, other will build on this framework and to develop even better data storage strategies and better tools.