**Q1.** Open the "tidy\_data.pdf" document in the repo, which is a paper called Tidy Data by Hadley Wickham.

1. Read the abstract. What is this paper about?
   1. The paper is about effectively cleaning data, specifically data tidying. This is a specific type of database that is easy to clean due to its structure. Tidy is structured so that each variable is a column, each data point is a row, and each type of data is in separate tables. The paper is about using case studies to demonstrate the advantages of consistent data structures and matching tools.
2. Read the introduction. What is the "tidy data standard" intended to accomplish?
   1. The “tidy data standard” makes the data-cleaning process easier as it aids with the initial data exploration and supports the creation of tools that work well together needed for data analysis; this takes away the long process of having to integrate tools together. It also makes the data-cleaning process easier, as individuals don’t have to create a new way of organizing and structuring data.
3. Read the intro to section 2. What does this sentence mean: "Like families, tidy datasets are all alike, but every messy dataset is messy in its own way." What does this sentence mean: "For a given dataset, it’s usually easy to figure out what are observations and what are variables, but it is surprisingly difficult to precisely define variables and observations in general."
   1. Like families, tidy datasets are all alike, but every messy dataset is messy in its own way." Tidy datasets are structured the same so that each variable is a column, each data point is a row, and each type of data is in separate tables and therefore are “all alike” but because every data set has unique data that have different meanings what is defined as variables, values, and types change. "For a given dataset, it’s usually easy to figure out what are observations and what are variables, but it is surprisingly difficult to precisely define variables and observations in general." This means that although it is normally easy to identify observations and variables in a specific dataset based on context, providing a precise and universally applicable definition for these terms across all possible datasets is difficult due to the variability in the data and the specific field of study. What is considered a variable in one dataset might be considered an observation in another, and vice versa.
4. Read Section 2.2. How does Wickham define values, variables, and observations?
   1. Wicham's definition of values in a dataset is that they represent the information collected and can typically fall into two categories: quantitative, which includes numerical data, or qualitative, which includes textual or categorical data. Values are associated with both a variable and an observation. A variable encompasses all the values that pertain to the same concept, measured using the same units, such as weight or temperature. On the other hand, an observation encompasses all the measurements taken for a specific unit across different characteristics or attributes, which could be, for instance, a person or a specific day.
5. How is "Tidy Data" defined in section 2.3?
   1. "Tidy data" is a structured and organized way of representing data that aims to make it easy to work with and understand. In tidy data, each variable is a column, each observation is a row, and each type of data is in separate tables.
6. Read the intro to Section 3 and Section 3.1. What are the 5 most common problems with messy datasets? Why are the data in Table 4 messy? What is "melting" a dataset?
   1. The five most common problems with messy data are when variable names are values, variables are in both rows and columns, multiple types of observation units are in the same table, and a single observational unit is in many tables. Data table four is messy because there are three variables; therefore, the dataset needs to melt. “Melting” a data set is when columns are turned into rows
7. Why, specifically, is table 11 messy but table 12 tidy and "molten"?
   1. In Table 11, variables are still being treated as columns making it messy. Table 12 is tidy and molten because table A (besides id and element); columns undergo a transformation where they are split into two new variables: one called "date," which stores the repeated column headings, and the other called "value," which contains the combined data values previously found in separate columns. However, 12a is not tidy because the element column still contains the names of variables. 12b is tidy because each row represents a signal observation, each column represents a variable, and all other variables are fixed.
8. Read Section 6. What is the "chicken-and-egg" problem with focusing on tidy data? What does Wickham hope happens in the future with further work on the subject of data wrangling?
   1. The chicken-and-egg dilemma in the context of tidy data revolves around the idea that the usefulness of tidy data is closely intertwined with the availability of appropriate tools for working with it. Tidy data and the tools designed to work with it go hand in hand. If people aren't interested in creating tools for tidy data, then there won't be a strong incentive for others to use tidy data. Conversely, if individuals aren't interested in using tidy data, they may not be motivated to develop tools for it. Wickham hopes that the community will continue to create frameworks for effective data organization and develop better tools. Ultimately, he encourages an ongoing discussion about how to best structure data.