Brooke Russell

Problem Set One

1. Naur is essentially describing the difference between data and information. Data just refers to the raw numbers, figures, or facts, whereas information conveys their meaning. One can be able to manipulate data fields, tables, etc. without understanding what they mean or why and particular data relationships are important. The problem with this view is that data is essentially useless without interpretation.
2. The Journal of Data Science makes a point of defining data science as inclusive of data applications. Both the Data Science Journal and the National Science Foundation do the same, albeit with slightly different terminology. What all three have in common is that they either explicitly or implicitly view data science as an applied field. Tukey’s definition of data analysis (putting it into the category of science rather than math) implies that there is something demonstrative about it, which also draws attention to the practical and usable aspects of data science rather than the raw, statistical manipulation of figures (which he rightly points out is a mathematical, not scientific, enterprise). Naur’s definition of data, on the other hand, is more focused on data statistically or mathematically speaking, without significant attention or devotion to its applicability.
3. The first reason for the growth of data is the Internet of Things. Constant connectivity between devices all collecting high volumes of highly penetrative data simply did not exist before the interconnectivity of smartphones, PCs, networks, etc., all of which feed into each other. The second driver is the monetization of data; firms see data as an asset, and are therefore highly incentivized to develop new products, services, etc. to capture and analyze it. Thirdly, we have the business world’s need for data. As data becomes more popular, businesses realize that it is needed for marketing and decision making. This makes the use of data absolutely essential for any business that wants to capture a competitive advantage.
4. In order to navigate to (A) root, (B) home, and (C) parent directories, a user would execute the following commands: A. cd / B. cd ~ C. cd ..

It is important to have shortcuts to these directories so that users always have a reference point that is easy to remember when navigating their computer’s file structure.

1. These commands will first move into the user’s home directory, create a “problem\_set\_1” directory there, navigate up to the parent directory, and then print the working directory. It will print the overall user’s directory since the only previous navigation was to the specific user’s home directory: “/Users/”