Part One: Histories

1. Peter Naur is famously quoted as saying data science "deals with the data, while the actual relation of data to what they represent should occur in other fields." Why do you think he'd choose to frame data science this way? What might be problematic in this statement?
   A. I find that Naur decided to frame data science this way because he finds that the job of a data scientist should ultimately only focus on compiling, cleaning, aggregation, etc. of data and the analysis should be conducted in other fields. This could be problematic because the one dealing with the data is often responsible for cleaning and compiling it (organizing the columns, dropping cells, etc.). If the person dealing with the data does not know the outcome or how it is being used, then there may be some issues with the way the data has aggregated. It is important for data scientists to understand the full story of the data set, scope and goals in order for resulted meaningful work.

2. In 2002, data science began to gain momentum as its own dedicated subfield. Compare and contrast the definitions of data science at that time, exemplified by the National Science Foundation, Data Science Journal, and Journal of Data Science, to those from Tukey & Naur in the 1970s.
   A. The rise in popularity of data science started to get redefined by different groups who have different definitions. The Data Science Journal shared similarities with Naur between the definitions of data science. The Data Science Journal defines Data Science as "Descriptions of data systems, their publication on the Internet, applications and their legal issues." This definition is similar to Naur's because they both place the focus of data science on more of the technical components, rather than more general definitions that are relevant such as those of the Journal of Data Science and the National Science Foundation. The National Science Foundation and the Journal of Data Science both define data science as not only the data/ technical components, but they also include a broader range of what it can consist of. Over time, the meaning of data has evolved form not only focusing on the actual technical aspects of data to also considering the additional factors that are relevant to the data.

3. Data continues to grow at an exponential rate today. List at least three technological factors that contribute to this growth. List three major sources of data that contribute to this growth

   A. One technological factor that has led to the growth of data growing at an exponential rate is the rate that the scalability of ubiquitous technology such as personal computers and smartphones, many people have access to them and the platforms available on

them. Another technological improvement has been hardware. Computers have extremely developed over the last 20 years, and will continue to in the future. For instance, they have become much more powerful, cheaper to produce and purchase, smaller, enhanced UI/UX features, etc. Lastly, another technological factor that has led to the rapid growth of data is the internet and the rise of online advertising/ecommerce. The ability to make a profit off of available data on users has created a whole new marketplace for collecting personal data for the benefit of companies.

B. One example of a major source of data that has contributed to this growth is payment systems such as venmo and online banking. Transactional data is continuing to increase due to the rise of these new digital payment methods. Another example of a major source of data that has contributed to this growth is third party platforms that collect additional points of data so that companies can target audiences better. These platforms constantly collect and track users movements to identify patterns to become more profitable, for instance google analytics, fullstory, hotjar, segment, persistIQ, etc. Social media platforms such as facebook, twitter, and instagram are another major source that have contributed to the exponential growth of data. All movements from these platforms and interactions are captured by these social platforms, stored, and often sold.

Part 2

a.Root: cd/
b.Home: cd ~
c.Parent: cd ..

5.
a.cd~: home directory
b.mkdir ./problem_set_1: go to problem_set_1 from mkdir
c.cd.. : change current dictionary to parent dictionary
d.pwd: print working dictionary