**Monday:**

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.”* What might be problematic in this statement? Why do you think he’d choose to frame data science this way?

I personally think his statement is problematic because while data science does deal with the data it may not be true that relations of that data should occur in other fields. Data collection is important but those who collect specific data should be the ones who provide the interesting way of taking these numbers and findings then making them meaningful. I believe that the people who collect data are the ones who are most familiar with issues or practical use for its collection. He may think that data science and data collection may be completely separated from the domains the data was originally collected form.

I think Naur choose to frame data science this way because he wanted to convey the importance for data collection and production but further discoveries and fundamental uses should be handled in other fields. Data science takes a question and tracks data to prove or disprove the idea of the hypothesis. It may be harder for others to produce unique or influential research questions from other data science collections. Overall he thinks relation of data to what it represents should occur in separate fields.

**Wednesday:**

2. There was a substantial shift in the ways we define data science between the 1970s and the early 2000s. Describe this shift and why it may have emerged.

This shift in the ways we define data science between 1970’s and the early 2000’s stems from new research, theories, and technology. The data science world was becoming a popular area of study. This study evolved the way researches defined data science and how it is much different than just the original collection of the data resources itself. The shift emerged due to the new complex theories, different forms of data and the fields and the increase in advancing technology that captures mass data. Also ideals of distribution and legal ramifications further shift the emerged differences in defining data science and if collection is public or private.

3. The idea of "big data" dominates much of modern data science. However, data is still growing at an exponential rate.

A. What factors do you think may have led to this growth? Mention at least three and describe why they have contributed to recent explosions in data volume.

1. The Internet - User interaction data collection is endless which is why the internet use had lead to the growth in data rates. Virtual data collection is happening in the background and seems less intrusive than physical data collection. The mass amount of data volumes gathered gives creators/advertisers a better understandings of their target users, audiences, habits, needs, and wealth. Basically being able to build knowledge from raw data collection.
2. Advanced computer technologies - Data mining methods and storage rates have greatly increased the mass amounts of data being collected. This rapid collection is due to the need to serve innovation and influence of mass data collection. Mass data collection is needed more and more for researchers to present influential patterns from data analysis.
3. Mass business/organization success - It was shown in class lecture that businesses are told to collect all their data but they are not sure why because they do not utilize most of the collection storage. While major businesses operate at mass levels many details, transactions or procedures would be hard to document without data collection. Major business success has influenced others to produce and collect all their data weather it seems useful or not. I think the idea of having data for backup is the main reason everything is collected and not just whatever the users think should be collected and stored for later use or analysis.

B. Where is this new data coming from?

Overall this mass amount of new data revolves around user generation that is captured by rapid data scraping or data collection technology sources. Coded data programs that capture and store or categorize data are essential to business and society success. Humans have created machines who can now do the data collection work at more than double the speed of human collection. The shift from physical data collection to digital has been a key factor to new and mass data.

**Friday:**

1. Name three different data collection methods. How are they similar? How are they different? Consider using specific scenarios where you may need to collect data to ground your responses.

Three different data collection methods include interviews, surveys, and observations. Interviews, surveys, and observations all qualify as qualitative data collection. All three of these methods give valuable behavior understandings. Since all these methods don't rely on numerical data it’s qualitative. Collecting, analyzing, and interpreting data from these methods may provide the same findings/results but the way they are conducted, documented and analyzed differ. When conducting interviews you must draft participants, transcribe data then analyze it for trends. While in surveys you implement the questions, launch it, then wait and collect user submitted data. From that user data you can clean and make sense of research outcomes. Observations are much different than both of the previous collection methods. Observations deal with physical deployment in order to capture interaction for collection. The observations can be classified or just random interactions that take place.