**Problem Set # 4**

**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?

Peter Naur falsely says that data science is restricted to the data alone and none of the inferences and patterns that come along with it. We are able to connect this information to other fields via visualizations and efforts in communication. He might have chosen to frame data science this way because at the time of this statement data was purely quantitative and not able to be seen as much more.

**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 substantial shift in data science was due to the emergence of the internet. With online connectability, come recorded interactions. Data started piling up at a rate previously unfathomable.

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.

First off, there are millions on new data sources emerging everywhere because of our new and increasingly digital society. Second, we are still finding out new ways to decipher the data and leverage it, which makes this information that much more valuable and sought after by everyone. Lastly, the cost of storing this data is getting really cheap. Companies can store incredible amounts of data in warehouses and have it easily at their disposal.

B. Where is this new data coming from?

This new data is coming from new data driven sources that are being discovered daily based on our societies digital needs. The internet of things is a great example. Technologies in this sphere are connected to the internet and have the ability to be recorded informationally.

**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.

**Interviews:** This type of data collection is important because it allows a researcher a personal account into what they are trying to answer. It affords the opportunitiy to get a story behind an experience. It differs from quantitative research in the way that it is less numbers and more interpretation. This could be useful for understanding open-ended questions.

**Surveys:** Surveys give a researcher a good mix between qualitative and quantitative data. There could be questions based on personal experience and there could also be questions that look to gather numbers and statistics. This type of data collection is often quick and cost effective.

**Census Scraping:** This specific type of quantitative data collection would be useful if you are wanting numbers data on a target demographic. Would not provide much more than statistics, but if combined with one of the other two data collection methods would give you a comprehensive amount of information.