**INFO 3401-Kexin Zhai**

**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?

If the actual relation of data occurs in other fields, people might only analyze the data they receive but are unable to interpret them. Because their only response is to analyze data. It people who analyze data don’t have related knowledge of specific field, they might not know how to interpret the data by using professional words in that field. Also, analyst might not know what the results of the data could be shown. Their direction to analyzing data might be vague. However, I think the reason Peter Naur choose to frame data science this way is if data is from other fields, the results of data analysis could be explained boardly. If people could work with other people who are from other fields when analyzing the dataset in that fields, analyst could have a better thought of what data actually refer to from field related expert. Or, if someone is a doctor and gives you medical data, you would be the person that could better interpret these medical data by using data analysis method. Then the results you get could provide doctor some clear ideas of these medical data and the doctor will have further interpretation based on the results of data.

**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.

In the 1970s, the definition of data science was “Data is a representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process”, “Datalogy, the science of data and of data processes and its place in education”. In early 2000s, data science was defined as descriptions of data systems, their publication on the internet, applications and legal issues. Almost everything has somewhat relation with data. The most important parts of data science is its applications. Data scientists are crucial to the successful management of a digital data collection. Because Social changes. More different source of data not only number and there are more people who deal with the data and data need to be analyzed and interpreted. Data science becomes a science that focus on application of data, solving problem with data.

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, as the development of digital technology, technology around us starts to record people’s data. It gathers all the data to gather and stores it. As people are doing their daily life activity, they are creating data for technology to gather.

Second, business company is increasing and more and more people start to focus on digital economy. They deal with company’s data, collecting other companies data and analyzing it.

Third, storage for data is cheaper than before. Data could be stored so it could grow without considering the storage.

B. Where is this new data coming from?

New data is coming from everywhere. For example, people are calling, texting, tweeting and browsing websites on mobile phones. Walmart handles more than 1 million customer transaction data in an hour.

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.

Interview, questionnaire and registration. These three methods all ask participants question and receive data from their answer. Interview is a method that researcher ask participants questions directly. It could be face-to-face, phone call or online meeting. This method takes more time to receive data but it’s better to ask complex question. Questionnaires are forms that are completed and return by participants. It requires high literacy and participants cooperations. Registration is a form that users enter information by themselves. It limited the information variability. For example, I am doing a research of the relationship between student GPA and music, exploring if learning musical instrument would influence students’ GPA. First, I need to collect students basic information such as gender, college in CU, year in university, GPA, have learnt musical instrument or not, how long do they learn, etc. I will collect these information by using questionnaire method. Next, I will pick the participants from CU student to do the interview to answer my further question. Interview method allows me to ask following questions if I find a point that I want to dig in. Also for the method of registration, if I am the manager of a new website and want to collect users data. Then I will set several question let users to fill such as birthday date, gender, area they from, etc.