**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?

While reading through this quote it brings me back to what we talked about in class about how data is useless without any meaning and any direction, for us to be given a dataset with no initial meaning and just numbers on a spreadsheet it is hard to change that into meaningful data. This statement is problematic because it shows that data science is nothing without meaning where data scientists can still make something out of nothing it may just not be meaningful.

1. 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.

The National Science Foundation sees data science as scientist who defines the data whereas the Data Science Journal defines it as a description of data. Nauer sees data science as a representation of facts or ideas in a formalized manner, Turkey sees it as more of science compared to mathematics.

1. Data continues to grow at an exponential rate today. List at least three technological factors that contribute to this growth and what role they play. List three major sources of data that contribute to this growth and at least one way they're being used.

Social Media – Advertisers and Marketers are able to go through and direct their advertising to specific groups of people with demographic and like information.

Mobile Phones – With recent discoveries in how much time is actually healthy for people to use on their phones we are now able to see how much time and what we are doing on our phones.

Sports – In the sporting industry data is full of information from how a player is performing to how other teams are attacking and defending and how data can help a team make decisions (MoneyBall)

**Part Two: Terminal Crash Course:**

4. How would you move into each of the following directories from the shell? Why do you think it is important to have shortcuts for each of these directories for navigating the file structure?

It is important to have   
A. Root

Cd/  
B. Home

Cd~  
C. Parent

cd..

5. Briefly describe what the following set of commands would achieve. What process would happen and what would be printed to the command line?  
  
cd ~ - *Go To* *Home directory*  
mkdir ./problem\_set\_1 – *Create new Directory in open directory*  
cd .. - *Goes back to home*  
pwd – *Print working Directory*

6. Assume all of the directories below are valid directories (e.g., they exist and we have access to them). What would the following print out?  
  
cd /usrs/INFO3401/../homework/./problem1  
pwd

7. What set of commands would you use to achieve the following:  
You've found yourself in a situation where your python program is trying to parse files in a directory and is throwing an error. You think it might be a permissions issue (i.e., you don't have permission to access the files in that directory) and need to verify the permissions on all files in that directory. The directory is in your home directory under the "datasets/activedata" directory. Make no assumptions about what directory you are currently working in.