Problem Set #4

Due: 9.23.18

**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?
   1. Peter Naur’s quote, “deals with the data, while the actual relation of data to what they represent should occur in other fields” has both pros and cons. If you don’t know the problem, you may not know what to do with the data once you get that. Naur might have framed it in this way due to the social structures of the time. If he’s living in a time where people are very specialized, and there’s not a lot of cross-over in fields, then it’s quite logical to say, “if I have information on this one subject, maybe I should pass it to a person who knows that field really well.” At the time, it was likely that there wasn’t a common language or dialogue yet as most people were only specialized in one specific area.
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.
   1. The amount of data was growing rapidly between the 1970s and the early 2000s. Because of this, the shift of how we defined data science could have been caused by the a) influx of information and b) the social ways we approached data and c) the amount of people who began to understand the power of (big) 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.

Factors that may have led to this growth include the transformation of technology that have grown in the past decade, allowing people to become more digital, businesses learning that they could use their customer’s data to improve their sales and profit (even if they didn’t know how to use the tools quite yet), and

B. Where is this new data coming from?

This new data is coming from everywhere!! It’s coming from our fingertips, through our cars, laptops, and anything else humans touch. We’re at a point where we now have so much data that we can more or less easily recognize patterns in our environment, whether modern or in the backcountry. All sorts of documentation is a piece of data, and everything that we post on the web can be considered data- meaning we’re producing it at a ridiculous rate!!!

Wednesday