Thomas Girmay

Milestone 1

For this final project we are tasked with analyzing data and creating a presentation with our results. After shortening my list of datasets to use, I am left with three: “The geography of language on Twitter”, Wordbank’s “Open Database of Children's Vocabulary Development”, and the US Department of Labour’s O\*NET database. The Twitter dataset keeps track of word usage by county as well as year so I could analyze words such as character names from popular movies or locations of disasters to see if the there exists a significant trend between when a word becomes popular in a county and the locations of those counties to see if this popularity moves in a general direction. To analyze this trend, I could choose a specific event, such as the release of a movie, and see how the usage a select group of words changed over time and if there is an apparent direction to this increase. If there is no real trend with this direction, then I could see if certain states react faster to news via tweeting about it compared to other states using the same approach. Another idea for this project was to use the children’s’ vocabulary development dataset and see if there exists a trend between certain countries with comparable cultures and the first words the children in those countries learn. We might share a lot of similar words learned such as “mother”, “father”, “food”, or “water”, but maybe in the US words such as “car” or “phone” might be learned earlier than in third world countries. Lastly, I could use the Department of Labour’s dataset to calculate which jobs have the highest payrate per level of education and really compare certain jobs everyone seems to desire as a child or teenager such as doctor or lawyer. I could also bring into light other well-paying jobs that might not be as well known and compare the skills required and the years in school or training required to attain them. This could be used to inform college or high school students the options they have and the benefits they will receive after attaining their desired position before working towards a degree or certification. This could also be used to decide if it is worth pursuing higher education past college in your desired field by showing if a masters is worth earning in terms of pay increase. This could also have certain trends such as individuals earning their degrees in math or English get a larger percentage increase in payment than engineers or doctors when earning a Ph.D. or Masters. After further analysis of these datasets and consultation with Dr. Mattei, I will choose one of these goals to be the focus of this final project.