Danford Compton
CS 410
Explorations In Data Science
Midpoint Report

This is the midpoint report for my Data Science project of comparing obesity rates with the density of fast food restaurants in the United States. The topic itself has been pared down from my original lofty goals; now it is simply statewide obesity rates compared with fast food restaurant density and comparing the results with differences based on ethnicity.

This reduction is based on many factors; the starting goals of trying to associate this data with income is a bridge too far, and would require a closer analysis to be meaningful. For that to be effective it would probably necessitate going to a city-level analysis, so the scope of doing something like that is gigantic.

I have data sets for the locations of fast food restaurants and statewide obesity levels, both for 2018. This is the data I will be focusing on for any visualization creation. Both sets already had premade state-based graphics and this made me more certain that this is a worthwhile endeavor. I have found many research papers, though there is not a consensus other than "There very well could be something there, but we cannot pin it down." That being said, it seems that they were in search of an absolute definite result, whereas I believe that showing the correlation between the two will be acceptable for my purposes. I have two papers in particular that are created from aggregate data that strongly support my hypothesis. For my continuing work on this process, I intend to dive into their sources to get a better understanding of their data, and to ensure that it supports my paper.

As for team structure, Danford Compton has really pulled it all together, creating the needed deliverables after (nicely) being allowed a bit of extra time. Conversely, Danford Compton has really been slacking on the project, but it is due to his significant workload this term rather than disinterest.

For my Milestones, as previously mentioned, I have found a data set with the locations of fast food restaurants across the United States. I will simplify this data to be state-based, and it should be easily accomplished. I have also acquired data from the CDC that contains

obesity rates by state. More importantly, the CDC data also has the results based on ethnicity. This is crucial because the research papers that I have found (as per milestone) show a more pronounced correlation between density and obesity with black and hispanic people. Unfortunately I have not yet created the graphics that I will need to highlight the significance of my data, but that is probably the next step. At this point I have enough data that I believe the visuals will be quite striking and thus be quite helpful for my argument. The plan to cross the data with the median income of the regions has been scrapped. The final deliverables have not been created, but the plan is in place to have everything ready on time. There will need to be more research done on the sources for my aggregate sources, but I am hopeful that this will all center around the year 2018 so I can present the most coherent picture possible.

I am a little bit behind where I hoped to be at this point in the process, but the combination of condensing my plan as well as having a firmer understanding of what I hope to accomplish makes me believe that it is going to be alright. At this point in the research process I am really interested in an idea that occurred to me regarding fast food and obesity, and that is smell. Based on some research papers that I am not planning on using for this project, I think that smell is a critical factor in how much people in the area eat fast food thus leading to obesity. I do realize how difficult that would be to test in real life as you would have to take wind direction and speed into effect, but for some future research, I believe that it would prove fruitful.

My midpoint meeting was on Tuesday, July 21st at 3:30 PM and was helpful.