Eric Faith

MTH 231-01

October 23, 2020

Mid-Semester Progress Report

Throughout the semester thus far, I have put much thought into what topic and dataset I wanted to pursue for my project. Since the first class, I had mostly been thinking about different topics related to marine science, since that is a field I am particularly passionate about, especially relating to either fisheries science or climate change. Given the fact that there are no shortage of marine science institutes in the area, from UMass Dartmouth's own School for Marine Science and Technology (SMAST) to the oceanography community in Woods Hole, I assumed that I wouldn't have much trouble with finding data related to this topic, since I would likely be able to message any researchers directly if I couldn't find any data online. After putting a bit more thought into which topic I should go with, I eventually decided to go with climate change, just because I assumed there would be a much wider variety of datasets online. Since there are so many different metrics involved in quantifying climate change, I decided that sea level rise would be a good one to use, since it is relatively straightforward but still very well-documented.

I then spent a decent amount of time searching through different datasets, primarily using the website catalog.data.gov, which houses an enormous variety of datasets from federal and state institutions across the country. I bookmarked over a dozen different datasets that seemed to be of interest, most of which involved historic sea level data and future projects for sea level rise in New York, Hawaii, or California. In addition to these, I also bookmarked a handful of other websites related to different topics, just in case I ended up needing or wanting to switch topics. Aside from catalog.data.gov, the other main website I used was fisheries.noaa.gov/inport, which is a data management website used primarily by the National Marine Fisheries Service (which, rather inconveniently, has a major research laboratory in Woods Hole). Once I had a hefty amount of different datasets bookmarked (around 18 total), I began to investigate and compare them, to evaluate which one would be best to go with.

Unfortunately, I soon ran into a problem. Almost all of the datasets obtained from catalog.data.gov were not accessible as raw data; most of them offered only metadata and summaries. Of the 18 total I bookmarked, only a handful of them offered the raw data available in CSV (or any other accessible format), and of that handful, they were all very small datasets, usually only containing a half-dozen rows and columns and only having a few dozen actual data points. These issues were present in essentially every dataset I bookmarked from this website, so as a result, most of them were unusable. I tried to search for a handful of other datasets from this website, but most of them still didn't offer any actual usable raw data, and the few that did were mostly either too small or, in some cases, didn't offer any guidance on what the different metrics or column-labels referred to, rendering the data essentially unusable.

All of these issues forced me to shift gears a bit. Fortunately, I still had a handful of other websites bookmarked, so I turned to fisheries.noaa.gov/inport to begin looking for a new dataset. Being forced to switch topics on such short notice wasn't particularly difficult or upsetting, since I still had the other websites bookmarked, and fisheries science is still a field that I am very interested in and passionate about. From NOAA's InPort website, I perused a handful of datasets, most of which originated from the Northeast Fisheries Science Center (NEFSC; the laboratory located in Woods hole) and therefore contained data obtained from New England waters, which I found particularly interesting. I eventually settled on utilizing the Food Habits Database (FHDBS), which is a database maintained by the NEFSC's Food Web Dynamics Program, and contains four individual datasets related to fishery food web dynamics.

Unfortunately, I wasn't immediately able to figure out if I was able to obtain the raw data using InPort, since it only seemed to display metadata; as a result, I decided to directly email one of the research staff within the Food Web Dynamics Program, Brian Smith. He responded rather quickly, and was able to offer instructions on how to obtain the raw data through InPort (which was actually fairly simply, since there was a link that I overlooked that allowed me to download it right on the InPort page). I downloaded the raw data, which consisted of four individual datasets available as CSV files: FHSPECIES, which documents the name and taxonomy of each species described; FHPD, which documents the habitat, physiology, and other data on predator species;

FHPY, which documents habitat, physiology, and other data on prey species, and FHPYL, which describes the length and sex of prey species, in a more compact file. The InPort page also offered a legend on what each attribute (row header) means for each dataset, since most of them used shortened abbreviations or acronyms; I made an additional Google Sheet file containing each of the abbreviations and their corresponding definitions, for my own convenience.

I soon, however, encountered a new minor problem. While the FHSPECIES and FHPYL datasets were easily able to be opened in Excel or Sheets, the FHPD and FHPY datasets were too large to be opened in either program. Fortunately, I am able to open them using a text editor, which allows me to actually view the attributes and data itself, however this isn't very suitable for direct analysis. At this point, my only two options were to either (1) utilize an external program, such as RStudio or JupyterLab, to perform analysis indirectly, or (2) only utilize the FHSPECIES and FHPYL datasets, which I can analyze and investigate in Google Sheets or Microsoft Excel. The option I use would depend on what questions I want to ask, which is something that I am still currently working on. Regardless of which option I select, I don't think I should have much trouble with the data; even if I go with the more limited option and only use FHSPECIES and FHPYL, they still offer a wide variety of data that could be used to investigate different questions related to length, sex, distribution, seasonal frequency, etc. of different prey species. As I begin investigating the datasets visually to compare the different attributes, I am hoping to develop some high-quality questions to investigate and analyze, which I am aiming to develop by the end of next week.

To view the raw data files, utilize the link below, which directs you to the InPort page for the Food Habits Database. Scroll down to where it says "Distribution Information), about ⅔ down the page, and click on the link where it says "Download URL". This will download a ZIP file, which can be extracted to view the four individual datasets as CSV files.

https://www.fisheries.noaa.gov/inport/item/8083