Morgane Fiorentini

Astro 98 Decal - Final Project

28 April 2024

Slopes and Visualizations - an Experience

My utmost goal for this project was to create a program where I could take seemingly intangible data and turn it into something visual and manageable. Before beginning the code, I worked with thousands of compiled photographs taken nearly a year ago at Mono Lake to map the structure and formation of the lake's well-known sand tufa by first tracing out the orientation of individual tufa tubes. I did so through an outside platform on my research advisor's computer, and then analyzed and sorted through all of the data to have information I could extract from an excel spreadsheet and directly work with in a jupyter notebook.

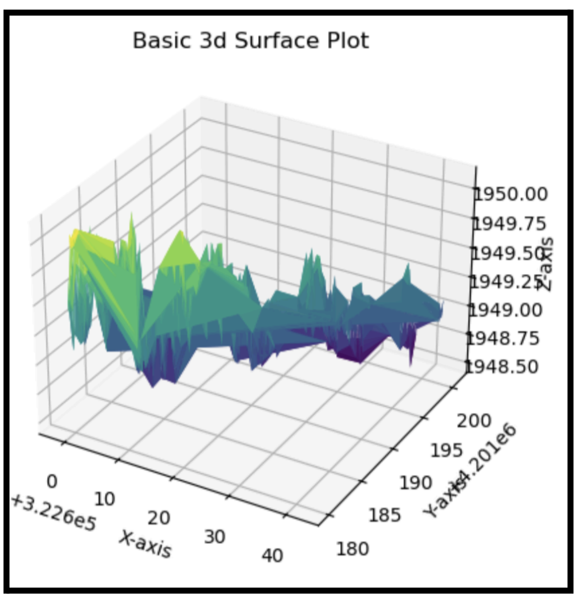
I analyzed roughly 742 collections of points, each containing x, y, and z coordinates, that would then allow me to calculate the equations of best fit and therefore the slopes and orientation to determine dip of each tube! Hopefully this will, after more research, allow me to understand what geologic abnormalities and processes enable such interesting structures to "grow." There is so much interest in Mono Lake because of the bizarre, almost otherworldly-like or unnatural characteristics and occurrences that have led it to become a hotspot for trying to predict or understand the processes taking place on planets beyond Earth.

I encountered many roadblocks when writing this code, specifically regarding the sheer amount of indexing I had to do and with having to take theoretical applications of physics and reconstruct them so that both the computer and I could understand what was going on. I do want to preface, my approach to this project may look very different from those of other students as the specific data I have worked with is not provided for online and doesn't directly involve subjects that are directly related to astronomy (exoplanets, HR diagram, etc). Nonetheless, I think the topic is really interesting and could provide for more overlap with astronomy in the future.

I am a visual learner, so I created multiple plots of my data to give more meaning to the project and add some color and excitement to my code! I made two scatter plots, one with all of the data and one that generates random points from the data pool I already have, a surface plot, and more.

In terms of collecting and being able to use random data for my project, I kind of manipulated this aspect into being able to collect random points from the large data frame I already had created. This was because not only is my project specific to the points I already have, but the other code I have written is designed specifically for my data. This is not to say that bringing in random data wouldn't work, but because the bulk of my code and data manipulation is in one large for loop, I thought it would be best to stick with my numbers but still bring in that aspect of randomness through what I already had. Furthermore, in terms of quality testing, it is important to note that, prior to generating my program, I repeatedly looked through all the points I had generated when I initially traced them out on my advisor's computer and eliminated all outliers. These outliers were in the form of inaccurately-traced lines that had one or more points that did not align with the positions of the line's other points. If I had left them on initially, there would have probably been very drastic changes in my values, because those points wouldn't accurately represent the orientation of the sand tufa. In looking at all the dips I generated, however, there are some that are a bit extreme compared to the rest, but I wouldn't know if they were outliers unless I could find that specific line I drew and see why the value was so different from the rest. These are the corresponding dip values that show up as "nan" when running the for loop block which has confused me given I filtered the data with the **try** and **except** lines of code in the sense that each i-value I look at in my for loop has a corresponding x, y, and z value, so there should be no reason as to why it is not running as expected. Most dip values do not show up as "nan," but I'm still trying to figure out why this is the case for some. Furthermore, I ran into the trouble of figuring out how to depict the changing theta values as, because the code that found a given theta value for a certain i was in a for loop, I couldn't create a graph outside of it but if I were to create one inside it would have made over 700 graphs. Obviously that is not what I intended for so I skipped over this aspect of mapping them out and moved on with my code. In the future, I would like to be able to display this because it would be a super neat way to visualize everything!

I used several libraries, from pandas to mpl_toolkits.mplot3d and beyond. Most of the libraries I implemented were so I could generate my graphs, like this:



By using pandas and numpy, I could really hone-in on my indexing and data-accessing skills, which have most definitely improved since the beginning of this project. Practicing with those libraries allowed me to model the data with more physics-based calculations by finding the magnitudes of the x, y, and x minimums and maximums for each given i-value (which there were 740-ish of). I did run into a bit of trouble when trying to display the dips I calculated because there were simply so many and creating a simple plot to show them didn't do them justice. The plot I made looked like jumbled points because of how many there were and how small the graph was, so I would most definitely like to fix that in the future. I would also like to learn how to create a graph that, from a 3D perspective, would show the differences in dips between each i-value, but unfortunately I couldn't figure it out in time!

In the future, I hope to take this project in multiple directions, from being able to utilize any data sets (not just specific to calculating the orientation for sand tufa tubes at Mono Lake) to calculating more than just the dip of each tube. I think it would be super cool if I could create a program that takes in all the data and the code I have written for this project and displays a 3D continuous model that depicts growth in real time (if that's even possible). Not only would this help me further understand the mechanisms of the specific area of the lake where my data was collected from, but it may also allow me to probe and predict the growth of sand tufa at other parts of the lake. As I stated above, the most difficult parts for me surrounded the topic of indexing, especially when trying to graph, but the most rewarding aspect was when, after hours of mistakes, the code finally came together, even if my graphs were not as extravagant as I had hoped them to be. Probably the only pieces of code that stayed with me from beginning to end in practically the exact format were the import statements, but even they underwent several changes as I kept on working on it. Every other line was rewritten or deleted entirely, especially when it came to that huge for loop I wrote. As someone who isn't great at physics, I am also really proud of that one section where I calculated the dip for each i! That took a lot of thought especially because it was difficult for me to take something semi-theoretical and tell my computer how to work with it. Regardless, I think this was a great experience in getting more time to do hands-on coding and I look forward to seeing where I can take the project in the future!