

Overview:

For this project I wanted to create a word cloud that would make the text size bigger for words that appear more frequently.

Implementation:

The main data structures that I used for this project were dictionaries and lists. I modeled most of my code after the histogram code from the text, where I would map each word to the amount of times that it showed up. I then took the n most common words and remapped the values to a value within the text size range I specified.

The most difficult decision that I had to make was about how to actually get a graphic representation of the fonts and the words. I originally started with a line, but after looking at it, was not satisfied with its aesthetic properties. So I talked with a ninja and we decided to try using a rotation matrix, which rotated the original vector that the words were located on around the origin. This was successful, and the words appeared in a circle, which is what I wanted. I then thought it would be really cool if I managed to make two circles, so I decided to divide the length of the vector by two after the first half of the terms. This worked really well and the final product looks really cool.

Results:

The final result looked really cool, and managed to really demonstrate the frequency of certain words. I realized early on that I would have to set a parameter so that I don't get "boring" words like "it, they, and, or". Some "boring" words slipped in, but generally the filter worked. I also enjoyed using matplotlib once I figured out how to do it.

Reflection:

I thought that matplotlib is very confusing, having only used matlab before. This was a great way to learn about it though. I tried bar graphs, pie charts, and finally just plotted with the matplotlib.pyplot.text feature. I also realized through working on the rotation matrix that while it has numpy, python kind of stinks when it comes to matrices. I think that matlab is much better for matrices, but I learned how to use numpy so I could do matrices in python if I need to.