Now there's a couple ways to do this exercise. I'll show you one.

Before any programming is done we have to download a couple packages. In the iPython terminal I'll type nltk.download\_gui(), this will open a menu with a bunch of download options, right now we only need stopwords, you'll see I already have it installed.

Then we have to install wordcloud, in the Anaconda prompt I'll type the command conda…, I already have it installed so I won't press enter right now.

So we're going to have a couple imports. We have re, nltk( from nltk.corpus we import stopwords), wordcloud, and matplotlin.pyplot

I'm gonna create an empty data container, then read our file into it.

Now I'm going to print the first 1000 characters of our file of gibberish

Our first step is to make every letter in this file lower-case, well do this with a for loop.

We'll print it again to take a look

Now we'll remove all of our special characters (this is where the import re is used), and print again

Next are the numbers, well print that again

Now we're going to create a variable with our stopwords, now stopwords are common words in the English language, and since this is a text file of gibberish I doubt there's going to be a lot taken out, but none the less we'll still do it.

So we're going to split the data into words with this forloop, and then we're going to remove the words that are also stop words. Let's print it again.

With that done let's create our wordcloud.

Awesome. In this video you learned how to read in a text file, process the data and create a wordcloud