Program 1
CSCE 557/MATH 587, Fall 2020
Due 10 September 2020

You are given ciphertext in the sample directory. You will each be given a different ciphertext; I will send you each the last two digits of the file name you are to use as your ciphertext.

What you know about this ciphertext:

- It is English, so you can use an English frequency count.
- It is all lowercase letters or numbers; all punctuation has been removed.
- I have left in the blank spaces between words.
- Each of the ciphertext files was encrypted with a substitution cipher, and each file used a different permutation of the alphanumeric letters and numbers.

This will be an exercise that is partly programming (to make the decryption less tedious and error prone) and partly analysis and guesswork.

You should be able to start with a frequency count and guess the most common letters.

You have dictionaries and lists of short English words, so you should be able to make guesses.

There may be proper names or other words in the plaintext but not in your dictionaries. If you insist on perfect matching of words and letter sequences, you may get stopped on such anomalies. You could fix this by writing a scoring function that counts the number of potential errors based on letter guesses, and then manually look at what are the best options. (In a real situation, you'd run tests automagically, but that's a programming exercise and not a cryptanalysis exercise.) Another way to deal with such anomalies is what I have done in my simple program: if I hack the ciphertext and place a word in quotation marks, my program doesn't check to see if the letter patterns of that word are legal for my list of English words.

Your Assignment

You should submit both code and a short paper on what you did to decrypt, since this isn't just programming.

Your code should come with a suitable README and must run on the Linux lab computers. Input should come from files named in command line arguments. You may program in Python, C++, or Java. If Python, then it must be Python 3.