PROJECT 4 – Letter Frequency
(A total of 50 points)

Did you know that the letter "e" is the most-often used letter in the English language? The different letters of the alphabet occur in words in differing amounts. The science of determining how often any letter occurs in an email, novel, textbook, or other set of text is called Frequency Analysis. Frequency analysis is often used to "crack" encrypted messages, but has other uses, too.

You will write a frequency analysis program in Python to count the number of occurrences of the letters a to z in a randomly selected set of 100 words. **Your program will do the following:**

1) Count the number of words N in the file **words.txt**.

2) Generate a list of 100 random numbers from the range of numbers 1 to N.

3) Read the **words.txt** file, and save into a list **only** words corresponding to the lines from your list of 100 random numbers. If a line in a file has more than one word, choose the first word for each line.You should write a file **random_words.txt** of these 100 words sorted by their line number in the input file with the following format (**Note: the numbers has to be RIGHT justified in their respective columns)**:

#	random number	word
1	25	able
2	301	catch
...		
100	110194	walking

4) For all the words in your list of 100 words, count the number of occurrences of each letter of the alphabet **a** to **z**. You should write a file **letter_count.txt** of these 100 words with the following format:

letter	count
a	17
b	9
c	15
...	

5) For every letter in your **letter_count.txt** file print a line of that many letters. You should write a file **letter_histogram.txt** of these letters **a** to **z** with the following format:

aaaaaaaaaaaaaaaa
bbbbbbbbb
ccccccccccccccc

6) There should be at least 4 functions **with parameters,** and one main function that calls most of the other functions.

Grading rubrics:

1)	Input validation and exception processing	(5 points)
2)	Counting the number of words in words.txt	(5 points)
3)	Generating the 100 random numbers	(3 points)
4)	Selecting the words from words.txt on the line corresponding to the 100 random numbers	(6 points)
5)	Counting the number of occurrences of each letter	(6 points)
6)	Creating a histogram for each letter per step 5	(6 points)
7)	Writing all 3 files in proper format	(6 points)
8)	Code reuse, algorithm implementation	(10 points)
9)	Comments	(3 points)

Due Dates:

Initial progress report (at least 1 page). Due **Friday, February 3, 2017**

The **project** and is **due Monday, February 13, 2017, at 07:00 AM**.

The Final Summary (at least two pages) due **Monday, February 13, 2017, at 07:00 AM**.

What did you learn? How your ideas change? How can be this project extended?

To submit your project, you must do the following:

1. The name of your program MUST begin with your first and last initials in upper case:
   GF_ LetterFrequency.py.
2. Create a folder within your Projects folder, called Project4.
3. Upload your Python program to be graded, and your 3 txt files into your Project 6 folder.
   Note: the files generated by your program must be named and format exactly as specified by the project requirements.

ATTENTION:

Please test your program extensively. User error handling and error checking to anticipate possible problems. Ensure that program doesn't generate syntax, runtime and semantic errors.