SSIE-500: Homework 2
Playing with Python

Hiroki Sayama
sayama@binghamton.edu

Deadline: September 23, 2020 (1:00pm)

Work on the following tasks:

1. Prepare some kind of plain text data as a .txt file. The text should be reasonably long (say, several thousand characters to fully capture the statistical nature of the text).

2. Write a Python code that does the following:

   (a) Read the above text file.

   (b) Count the frequencies of characters in the text, and plot a histogram.

   (c) Count the frequencies of transitions from one character to another in the text. For example, if the content of the text file is "`peter piper picked a peck of pickled peppers`", then the frequencies of transitions should be like:

      i. `p -> e` : 5
      ii. `e -> t` : 1
      iii. `t -> e` : 1
      iv. `e -> r` : 3
      v. `r -> ' '` (space) : 2
      ...

   (d) Visualize the results of (c) somehow.

   (e) (Optional) Using the result of 2c above as a Markov process model (i.e., a process in which the probabilities of next characters are determined by the current character and are proportional to the frequencies of such transitions), generate a random gibberish of made 3,000 characters starting with letter 'T'.

3. Produce a PDF file of your report using LaTeX, showing your Python code, the results and any comments/discussions, and submit it to the Assignment page in MyCourses by the deadline. To include program codes in LaTeX, you can use either the `listings` environment or the `pdfpages` package.