Lesson Eighteen

Math 6080 (for the Masters Teaching Program), Summer 2020

18. Frequencies. A message encoded by a cipher that uses a "random" function

$$f : \{\text{letters of the alphabet} \to \{\text{letters of the alphabet}\}$$

might seem hard to decode. Using frequency tables, however, we can do it if the encoded message is long enough. There is bit of guesswork involved, though.

The (English) frequency table for the letters of the alphabet records the percent of the time a letter chosen at random from a text will be the given letter. Thus:

$$e \text{ occurs with frequency } 11.162\%$$

$$t \text{ occurs with frequency } 9.356\% \text{ etc}$$

Exercise. Maka a Python table using the Wikipedia page on letter frequency:

a	b	c	$\cdots$
.08497	.01492	.02202	$\cdots$

(here we convert the percentages to floating point numbers).
Check that these numbers add to (approximately) 1.

Exercise. Prompt the user for some text, which you store as a string. Step through the string, and record the frequencies of the letters by counting the number of times each letter occurs (lower or upper case), and dividing by the total number of letters (discarding the non-letters in the string from your count). Compare your frequencies with Wikipedia's list in a bar graph.

Strategy for decoding a coded message. Enter the coded message into the previous exercise. With some luck (and a long enough coded message), you should be able to determine much of the function f by matching your observed frequencies of letters with those in the Wikipedia table. There will probably not be a perfect match, but by looking at small words the frequency matches should give you enough of a hint to allow you to complete the function f and decode the message.

Let this strategy aid you in decoding:

Yr. Jrduibt tbztrv zet jtz vefj zf rtaivztr d lfyjudibz dhfcz d ptdp Bfrktaidb Huct jdrrfz mcvz dv zet vefjqttjtr iv jrtjdriba zf lufvt zet tvzdhuiveytbz xfr ucble. Ptvjizt htiba zfup zedz zet hirp iv ptltdvtp dbp zedz iz edp httb bdiutp zf izv jtrle, zet jrfjritzfr ibvivzv zedz iz iv "jibiba xfr zet xmfrpv" fr viyjus "vzcbbtp".Dv zet todvjtrdztp Jrduibt dzztyjzv zf kdqt cj zet jdrrfz, zet vefjqttjtr zritv zf ydqt zet hirp yfgt hs eizziba zet ldat, dbp Jrduibt trcjzv ibzf d rdat dxztr hdbaiba "Jfuus Jdrrfz" fb zet lfcbztr. Dxztr uivziba vtgtrdu tcjetyivyv xfr ptdze ("iv bf yfrt", "edv ltdvtp zf ht", "htrtxz fx uixt, iz rtvzv ib jtdlt", dbp "zeiv iv db to-jdrrfz") et iv zfup zf af zf zet jtz vefj rcb hs zet vefjqttjtr'v hrfzetr ib Hfuzfb xfr d rtxcbp. Zedz jrfgtv pixxilcuz, dv zet jrfjritzfr fx zedz vzfrt (kef iv rtduus zet vefjqttjtr, vdgt xfr d xdqt yfcvzdlet) ludiyv zeiv iv Ijvkile, ketrtdv zet rdiukds vzdzifb dzztbpdbz ludiyv et iv dlzcduus ib Hfuzfb dxztr duu. Lfbxrfbziba zet vefjqttjtr'v "hrfzetr" xfr usiba, zet vefjqttjtr ludiyv et kdv judsiba d jrdbq fb Jrduibt hs vtbpiba eiy zf Ijvkile, keile kdv d jduibprfyt xfr Hfuzfb; Jrduibt jfibzv fcz zedz zet vefjqttjtr kdv krfba htldcvt d jduibprfyt xfr Hfuzfb kfcup edgt httb "Bfzufh". Mcvz dv Jrduibt edv ptliptp zedz "zeiv iv atzziba zff viuus", Ardedy Ledjydb'v bf-bfbvtbvt Lfufbtu hcrvzv ib dbp frptrv zet vqtzle vzfjjtp.