TASK ONE:
CRYPTANALYSIS
REPORT

Table of Contents

Mono Alphabetic Cipher

A mono-alphabetic cipher is a simple substitution cipher in which each letter of the plaintext is replaced with a fixed letter or symbol. It is also known as a "static" or "non-polyalphabetic" cipher. The Caesar Cipher, which replaces each letter with the letter three places down the alphabet, is an example of a mono-alphabetic cipher. These types of ciphers are relatively easy to break, especially with frequency analysis.

Encrypted Text (Ctext-1.txt)

snmy fr rnketcflb wcfhc fr jnmbem meeojy fl nuq casft nd tcnubct we aqe ahhurtnkem duqtceq tn qebaqm tcqee onfltr ar seflb rftuatem nl a rtqafbct jfle fd tcefq aooaqelt onrftfnlr hal se kame tn hnflhfme dnq nsreqvatfnl wftc nle eye ulmeq ruftasje hcnfhe nd nuq ojahe nd nsreqvatfnl fd fl ouqrualhe nd nuq casft nd tcnubct we lnw ruoojekelt tce oqnonrftfnlr nd euhjfmeal benketqy sy tce rflbje oqnonrftfnl tcat twn onfltr nl a oqahtfhajjy qfbfm snmy ajwayr hnqqeronlm tn tce rake ln bqeat frrue oeqcaor eveq culb uonl tcere jnlejy rea hnksatr sut

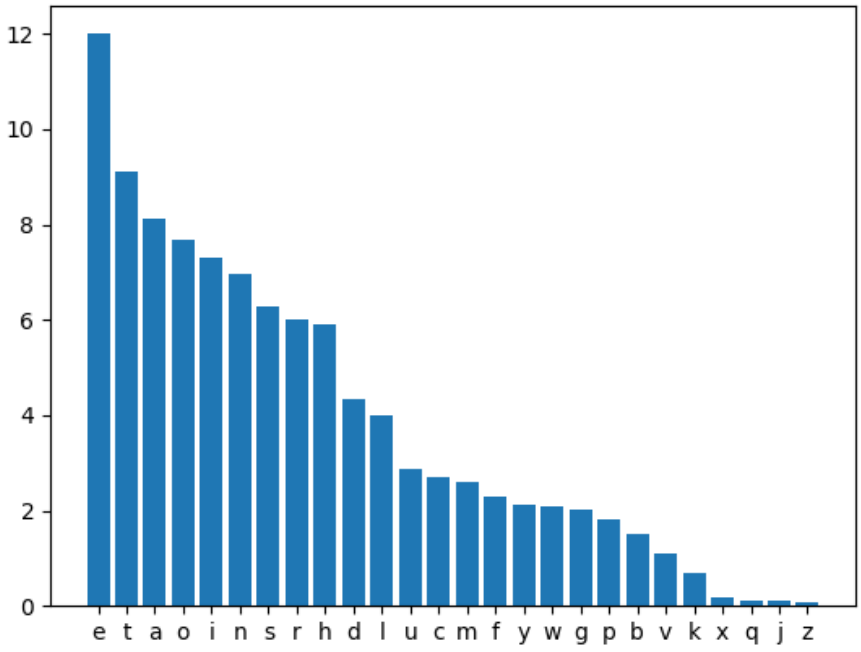
Decryption Process

Frequency Analysis

The frequency distribution of the English Alphabet used in daily life is given below:

E	T	A	O	I	N	S	H	R	D	L	C	U	M	W	F	G	Y	P	B	V	K	J	X	Q	Z
11.7	9.1	8.2	7.5	7.0	6.7	6.3	6.1	6.0	4.3	4.0	2.8	2.8	2.4	2.4	2.2	2.0	2.0	1.9	1.5	1.0	0.8	0.15	0.15	0.10	0.07

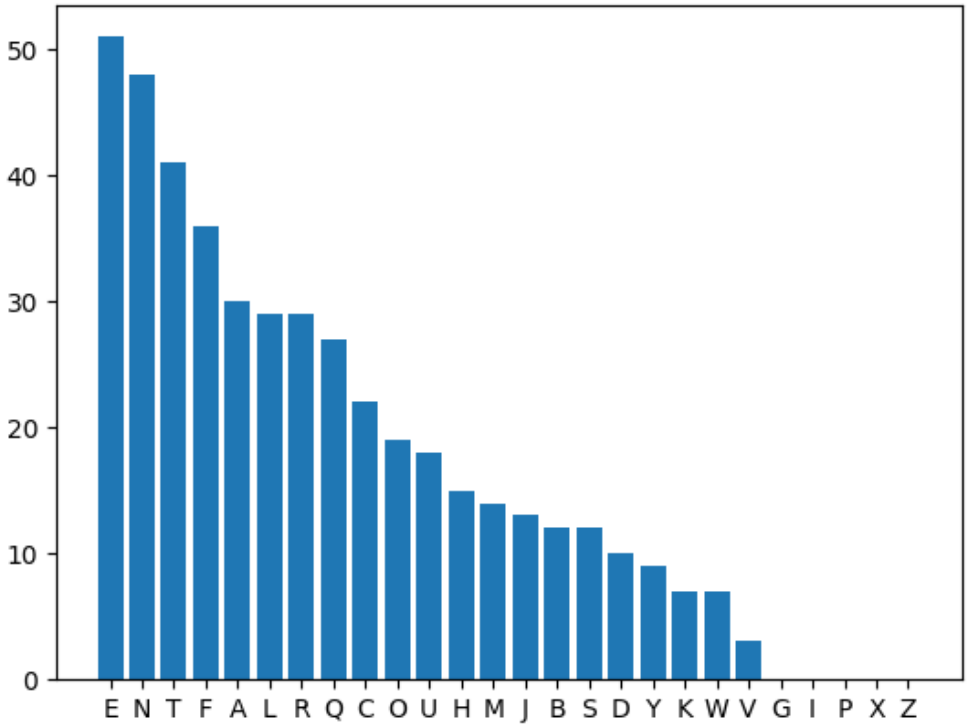
The graph for the above is as follows:



The frequency distribution of the alphabet provided in the cipher text is as follows:

E	N	T	F	A	L	R	Q	C	O	U	H	M	J	B	S	D	Y	K	W	V	G	I	P	X	Z
51	48	41	36	30	29	29	27	22	19	18	15	14	13	12	12	10	9	7	7	3	0	0	0	0	0
11.3	10.6	9.1	8.0	6.6	6.4	6.4	6.0	4.9	4.2	4.0	3.3	3.1	2.9	2.7	2.7	2.2	2.0	1.5	1.5	0.7	0.0	0.0	0.0	0.0	0.0

Similarly, the graph for the above distribution is as follows:



Comparisons and Substitutions

Just by replacing the alphabets according to the frequency distribution does not correctly decrypt the cipher. Rather, different comparisons based on other statistics and some hit-and-trial methods are used.

Bi-graphs, Tri-graphs and Double Letters

By comparing the most occurring bi graphs and tri graphs in English Language to the most occurring bi graphs and tri graphs in the cypher text, we can come across some useful substitutions.

The most common digraphs in the English language are:

TH, HE, AN, IN, ER, ON, RE, ED, ND, HA, AT, EN

The most common digraphs in the message are:

TC, NL, FL, FT, ON, ND, QE, CE, EQ, AT, TF, KE, ME

The most common tri-graphs in the English language are:

THE, AND, THA, ENT, ION, TIO, FOR, NDE, HAS, NCE, TIS, OFT, MEN

The most common tri-graphs in the message are:

TCE, TFN, FNL, RFT, NKE, FLB, NUQ, BCT, NFL, ONR, NRF, FTF, NON

Final Substitutions

Once most of the substitutions are made correctly using previously mentioned techniques, we can easily decrypt the rest of the cipher by observing words.

Decrypted Text (Ptext-1.txt)

body is something which is lodged deeply in our habit of thought we are accustomed further to regard three points as being situated on a straight line if their apparent positions can be made to coincide for observation with one eye under suitable choice of our place of observation if in pursuance of our habit of thought we now supplement the propositions of Euclidean geometry by the single proposition that two points on a practically rigid body always correspond to the same no great issue perhaps ever hung upon these lonely sea combats but

Substitution Key (Key-1.txt)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	S	H	M	E	D	B	C	F			J	K	L	N	O		Q	R	T	U	V	W		Y	

The empty boxes denote the Substitutions that cannot be determined from the given text as the letters J, K, Q, X and Z are never used in the original text.

Vigenère Cipher

The Vigenère Cipher is a polyalphabetic substitution cipher that uses a repeating key to encrypt the plaintext. The key is a sequence of letters that are used to encrypt the corresponding letters of the plaintext. The Vigenère Cipher was first described by Giovan Battista Bellaso in 1553, but it was later misattributed to Blaise de Vigenère in the 19th century.

The Vigenère Cipher uses a table called the Vigenère square or Vigenère table, which consists of the alphabets written out 26 times in different rows, each alphabet shifted cyclically to the left compared to the previous alphabet, corresponding to a Caesar Cipher. To encrypt a plaintext message, the key is added to each letter of the plaintext message to produce the cipher text. To decrypt the cipher text, the key is subtracted from each letter of the cipher text.

The Vigenère Cipher is considered to be more secure than mono alphabetic ciphers because it uses a different substitution for each letter of the plaintext, making it more difficult to break using frequency analysis. However, it can still be broken using methods such as the Kasiski examination and the Index of Coincidence.

Encrypted Text (Ctext-2.txt)

bj itqftn dltjifz je nzl rludpuzyk iz nbaji kbw lngcjl irm tlfe qgu bex tf xycuo jk gmzu sy ehjenspovx lofjy sudzyfa tvu xphynk obmy jlbc ufk qvlehovhl cbcow dirn tlukyj lyrghsfj ix jpff zhsuczvpu ix jizpsssfok sppudaz ki lof wfsn pw mwsg iydpben wufiaq ufvx tl jduypovx gy evmayfu nzl hvhwybkcgu uyul jbilalt kbw ofrpq ivixwu pw nzl fdjayf ki vhz tufupk uxmpix lv gflylu kbw abcy gm tlwz lygfgpuj xazurhul mzhw pokyjcbc cfkfgyfkfendf pw uff dyufnfj cf wpjclppe ng dizwz df duq zvsdwju kbw ipus lof glgwpjclppem gm flwdpevuf nffgwasp

Kasiski Analysis

Kasiski examination is a technique used to determine the key length of a Vigenère Cipher or a similar polyalphabetic substitution cipher. The method is based on the observation that, in a piece of cipher text that has been encrypted using a repeating key, identical sequences of letters will appear at regular intervals.

The Kasiski examination involves finding repeating sequences of letters in the cipher text and then finding the greatest common divisor (GCD) of the distances between the repeating sequences. This GCD is likely to be a factor of the key length. Once the likely key length has been found, the cipher text can be divided into columns, one for each letter of the key. Each column can then be treated as a mono alphabetic substitution cipher and can be attacked using frequency analysis.

It's important to note that Kasiski examination is not always successful, and it's not always possible to determine the key length and decrypt the cipher text.

Decryption Process

Step 1: Determining Repeated Sequences

We first determine certain phrases and alphabets that are being repeated after certain gaps.

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
POVX	140	Y		Y	Y		Y			Y				Y						Y				
UKBW	95				Y														Y					
WPJC	40	Y		Y	Y			Y		Y										Y				
PJCL	40	Y		Y	Y			Y		Y										Y				
JCLP	40	Y		Y	Y			Y		Y										Y				
CLPP	40	Y		Y	Y			Y		Y										Y				
LPPE	40	Y		Y	Y			Y		Y										Y				
UFNF	65				Y								Y											
JEN	60	Y	Y	Y	Y	Y				Y		Y			Y					Y				
YJL	30	Y	Y		Y	Y				Y					Y									
IXJ	15		Y		Y										Y									
LOF	100	Y		Y	Y					Y										Y				
POV	140	Y		Y	Y		Y			Y				Y						Y				
OVX	140	Y		Y	Y		Y			Y				Y						Y				
NZL	215				Y																			
KBW	220	Y		Y	Y					Y	Y									Y		Y		
NZL	255		Y		Y										Y		Y							
AYF	50	Y			Y					Y														
LUK	170	Y			Y					Y							Y							
KBW	270	Y	Y		Y	Y			Y	Y					Y			Y						
MTL	267		Y																					
KYJ	205				Y																			
CBC	220	Y		Y	Y					Y	Y									Y		Y		
YFK	73																							
FKF	5				Y																			

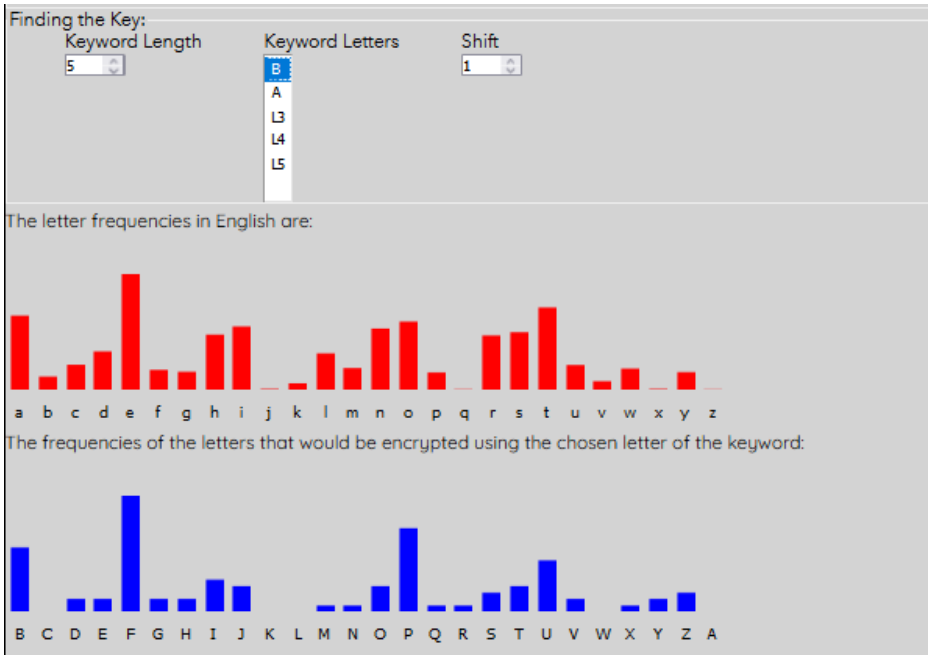
Step 2: Determining Keyword Length

For the different number of gaps found for each sequence, we compute the possible factors for each gap length. For example, for a gap of 16, possible factors are 2, 4, 8 and 16, each of which can be used as a keyword length. Once the factors are determined for all repeated phrases, we determine the factor that most commonly occurs. This number serves as the keyword length. As shown in the figure, in our case the keyword length is 5.

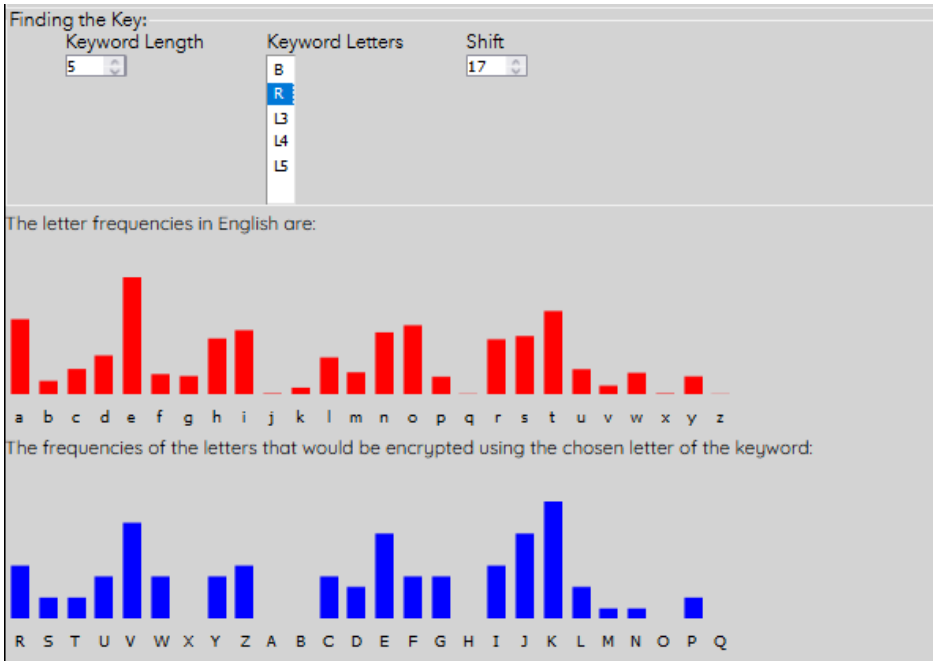
Step 3: Evaluating the Keyword

After determining the length of the keyword, we look up frequency distributions and compare the values for each letter to evaluate the letters of the keyword. In the figures that follow, the red frequency distribution determines the frequency of the English Alphabets while the blue one represents the frequency distribution for the letter of the keyword. We need to shift the blue graph such that both the graphs are almost similar. In this way we can determine each letter of the keyword.

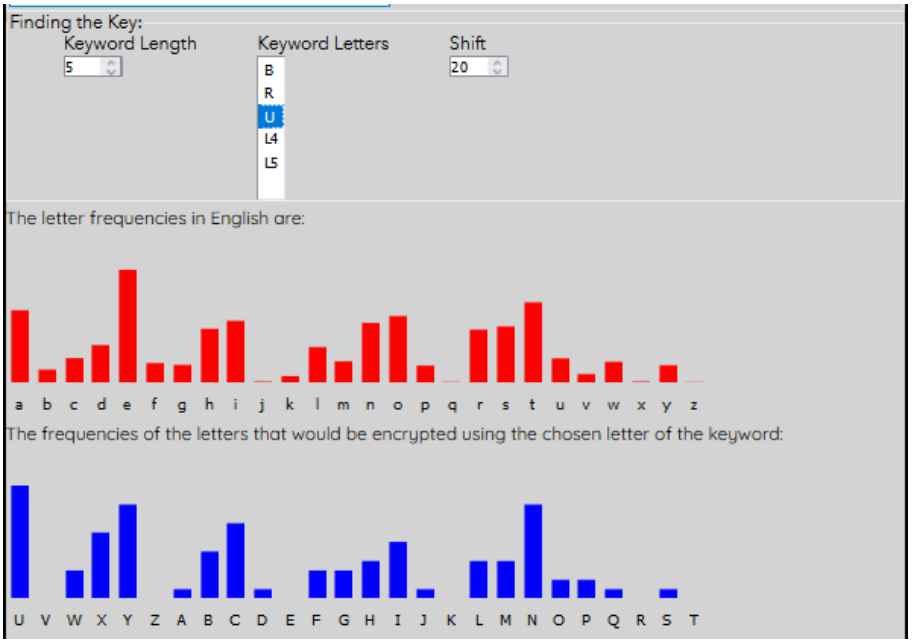
For the first letter:



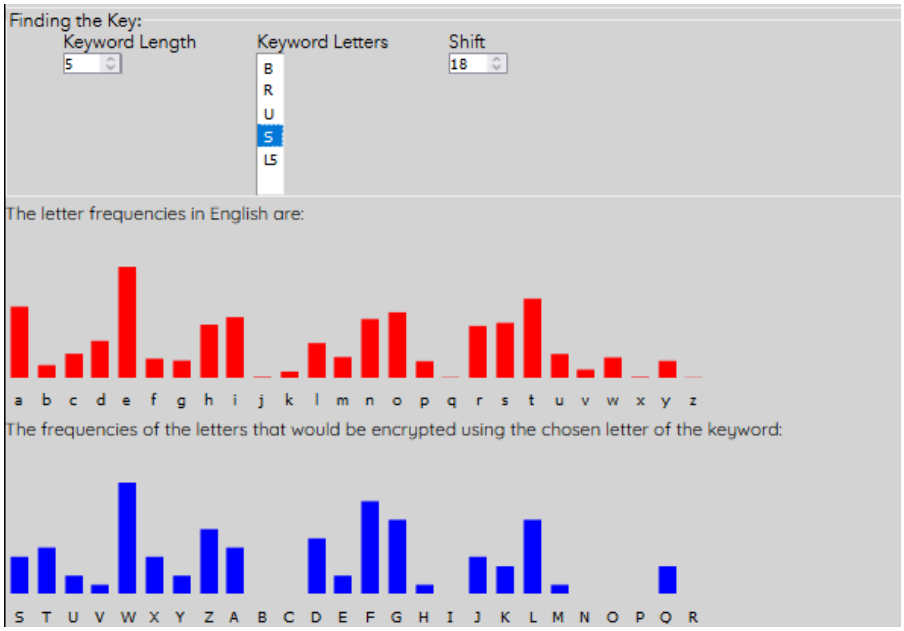
The graph for letter ‘B’ matches the graph, so we select it. Similarly, for the second letter, we have ‘R’:



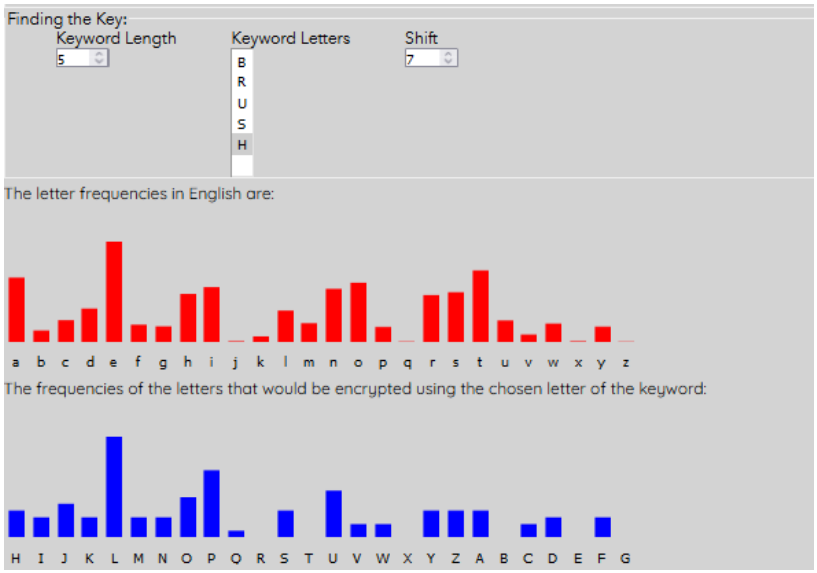
Similarly, for the third letter, we have ‘U’:



Similarly, for the fourth letter, we have ‘S’:



And finally, for the fifth and last letter, we have ‘H’:



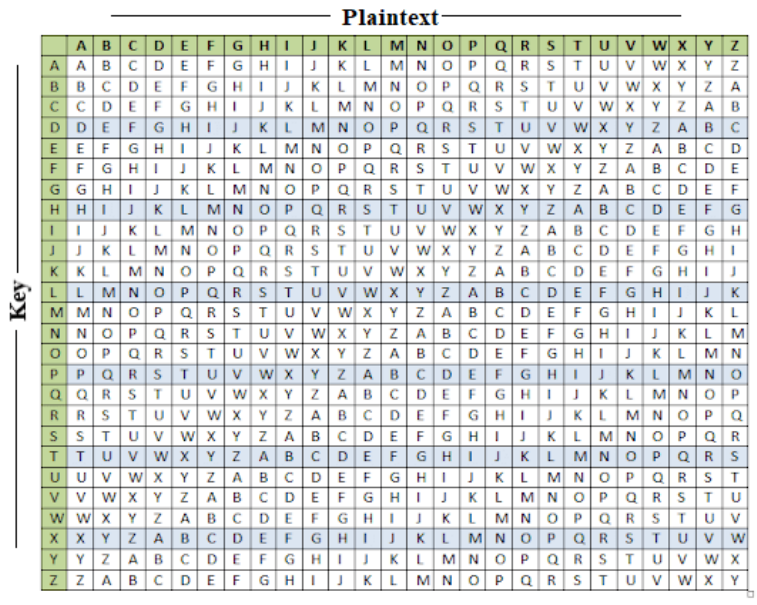
Thus, the keyword for this cipher is ‘***BRUSH***’.

Step 4: Using Vigenère Square

After obtaining the keyword, it is cyclically appended to itself to match the length of the encrypted text.

brushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrush
brushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrush
brushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrush
brushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrush
brushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrushbrush
brushbr

The cipher is the decrypted using the key above and the Vigenère Square as given below.



For each letter in the key, the corresponding letter in the cipher text is looked up in the Vigenère Square to get the plain text. The rows represent the key, whereas, each column represents the plain text that is to be decrypted.

Decrypted Text (Ptext-2.txt)

> as object lessons in the qualities by which the empire has been won and by which it must be maintained these ancient sea fights have real and permanent value what better examples of cool hardihood of chivalrous loyalty to the flag of self reliant energy need be imagined or desired the generation that carries the heavy burden of the empire to day cannot afford to forget the tale of such exploits distance line interval independently of any changes in position to which we may subject the body the propositions of euclidean geometry

Substitution Key (Key-2.txt)

B R U S H

Tools & Software Used

For Mono Alphabetic Cipher, the following link was used to draw results:

Frequency Analysis: Breaking the Code – Crypto Corner

https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html

For Vigenère Cipher, the following link was used to draw results:

Kasiski Analysis: Breaking the Code – Crypto Corner

https://crypto.interactive-maths.com/kasiski-analysis-breaking-the-code.html