Fall 2019
Chris Christensen
MAT/CSC 483
Section 14

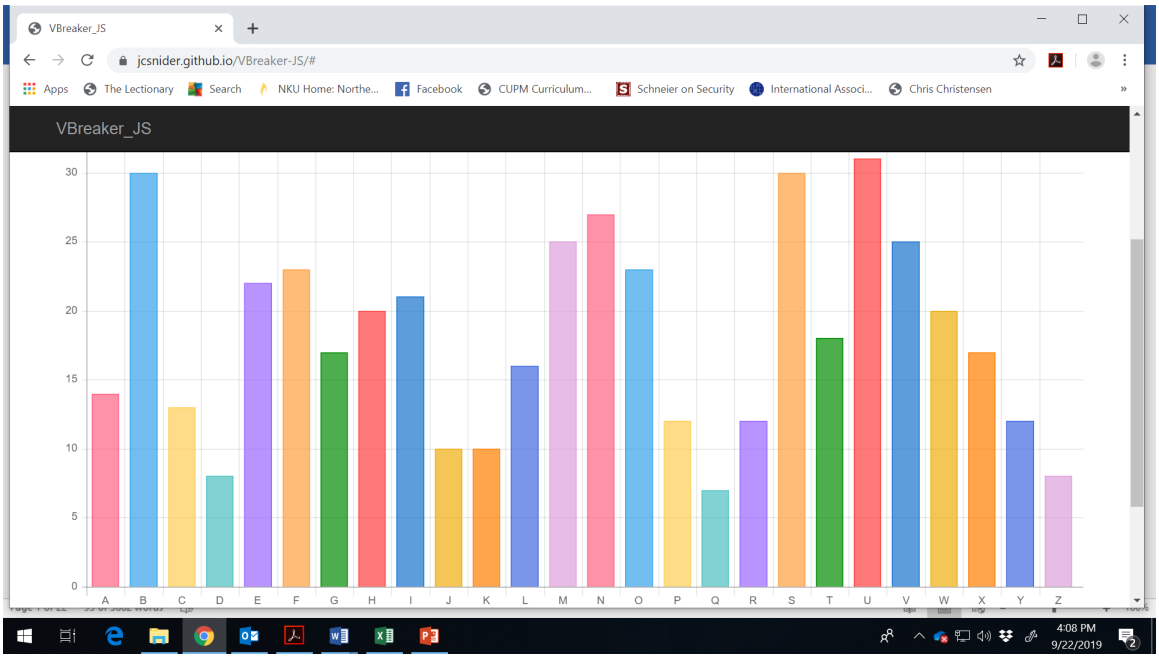
Cryptanalysis of the Vigenère Cipher

The keyword of a Vigenère cipher describes the rotation among the Caesar cipher alphabets that are used. That rotation leads to patterns that can be exploited by a cryptanalyst. If we know the length of the keyword, we can often determine the keyword and, hence, decrypt all messages encrypted with that keyword.

Here is a ciphertext message that has been encrypted with a Vigenère cipher.

```
nifon aicum niswt luvet vxshk nissx wsstb husle chsnv ytsro
cdsoy nisgx lnona chvch gnonw yndlh sfrnh npblr yowgf unoca
cossu ouoll iuvef issoe xgosa cpbew uormh lftaf cmwak bbbdv
cqvek muvil qbgnh ntiri ljgig atwnv yuvev iorim cpbsb hxviv
buvet vxshk uorim mjbdb pjrut fbueg ntgof yuwmx miodm ipdek
uuswx lfjek sewfy yssnm zscmm bpgeb huvez ysaag usaew mffvb
wfgim qpilw bbjeu yfbef vbfrt mtwnz uorig wpbvx hjsnm zpfag
uhsnm npglb jbqrh mttrh huwek mpfak ljjen hbbnh ooqew vzdak
udvum yucbx yoquf vffew vzonx hjumt lfgef vmwnz uxsiz bumag
xbbtb kvotx xumpx qswtx l
```

Here are the ciphertext frequencies:

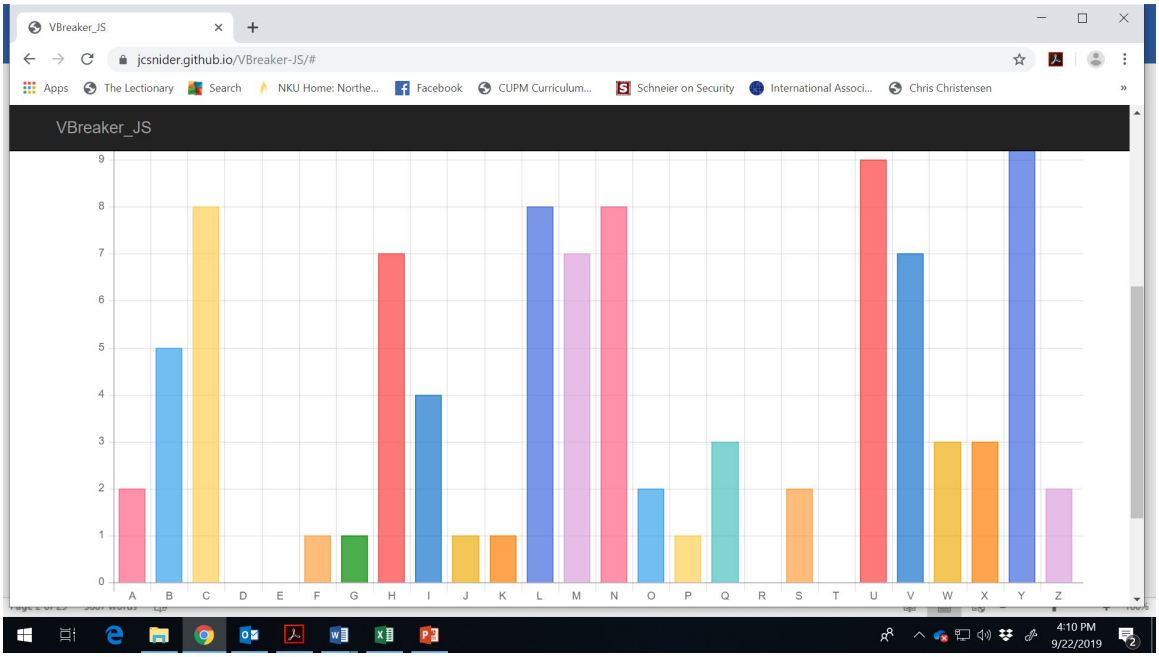


Assume that, somehow, we have discovered that the keyword has length five. Then the first letter of each block is encrypted with the same row of the Vigenère square – they are encrypted with the same Caesar cipher. Similarly, the second letter of each block is encrypted with the same row – the same Caesar cipher. The third letters with the same Caesar cipher. The fourth letters with the same Caesar cipher. And, the fifth letters with the same Caesar cipher.

Because Caesar ciphers are easily broken by frequency analysis, we can discover the letters of the keyword. Here is how we can proceed.

Strip off the first letters of each block and do a frequency analysis on the result. They should have all been encrypted with the same Caesar cipher.

Alphabet number one – first letters of each block

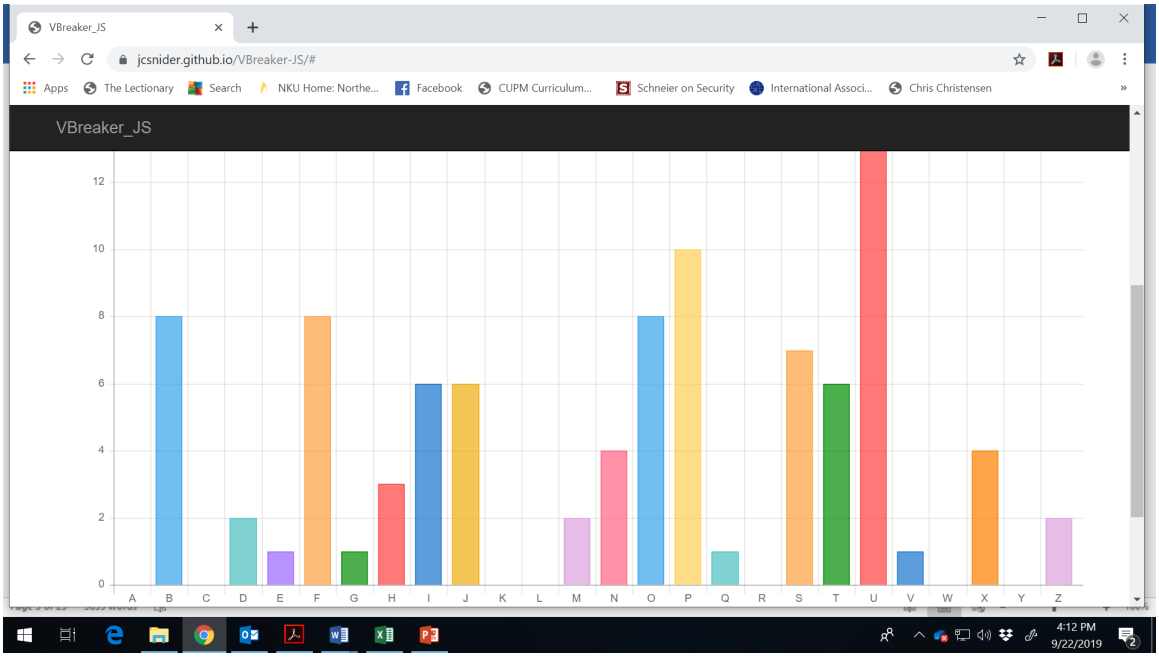


It appears that ciphertext `Y` corresponds to plaintext `e`. (Not just because it is the most frequent letter but because all the high frequency letter patterns fit – `U` would correspond to `a`; `C` would correspond to `i`; `H` and `I` would correspond to `n` and `o`; and `L`, `M`, and `N` would correspond to `r`, `s`, and `t`.)

Now recall that when we are encrypting using a Vigenère square plaintext `a` corresponds to the first letter of the row being used – the letter of the keyword being used. So, it appears that (because `U` corresponds to `a`) the first letter of the keyword is `u`.

The keyword is `u _ _ _ _`.

Alphabet number two – second letters of each block

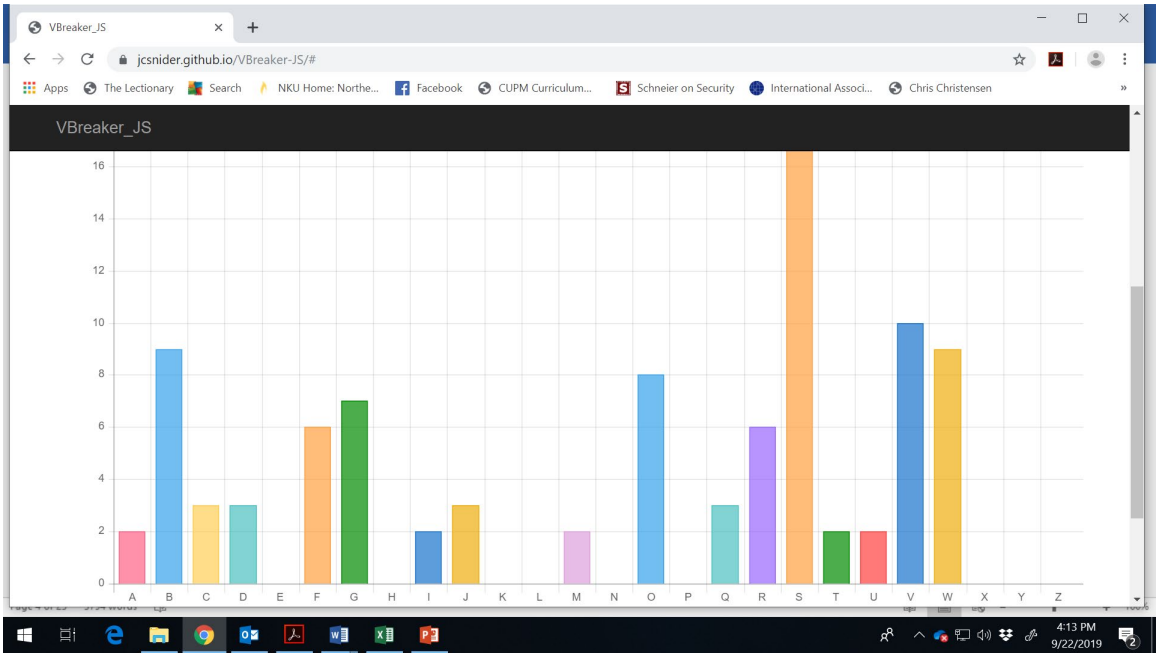


It appears that ciphertext `F` corresponds to plaintext `e`.

So, it appears that (because `B` corresponds to `a`) the second letter of the keyword is `b`.

The keyword is `u b _ _ _`.

Alphabet number three – third letters of each block

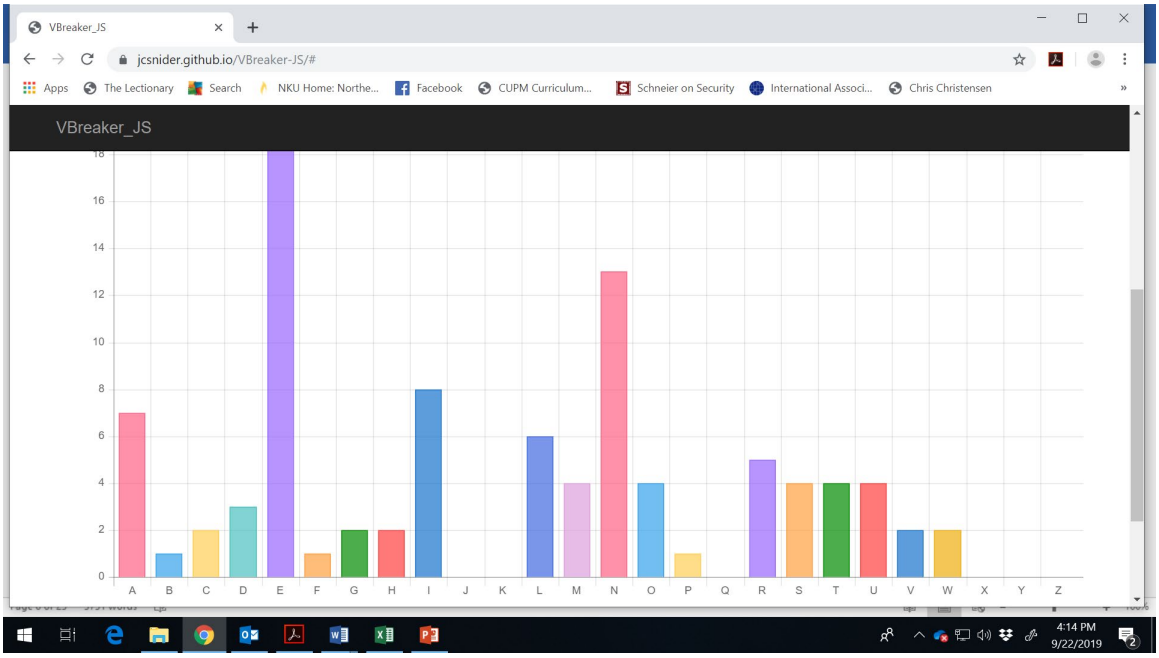


It appears that ciphertext `S` corresponds to plaintext `e`.

So, it appears that the third letter of the keyword is `o`.

The keyword is `u b o _ _`. (Perhaps, you can already guess the keyword.)

Alphabet number four – fourth letters of each block

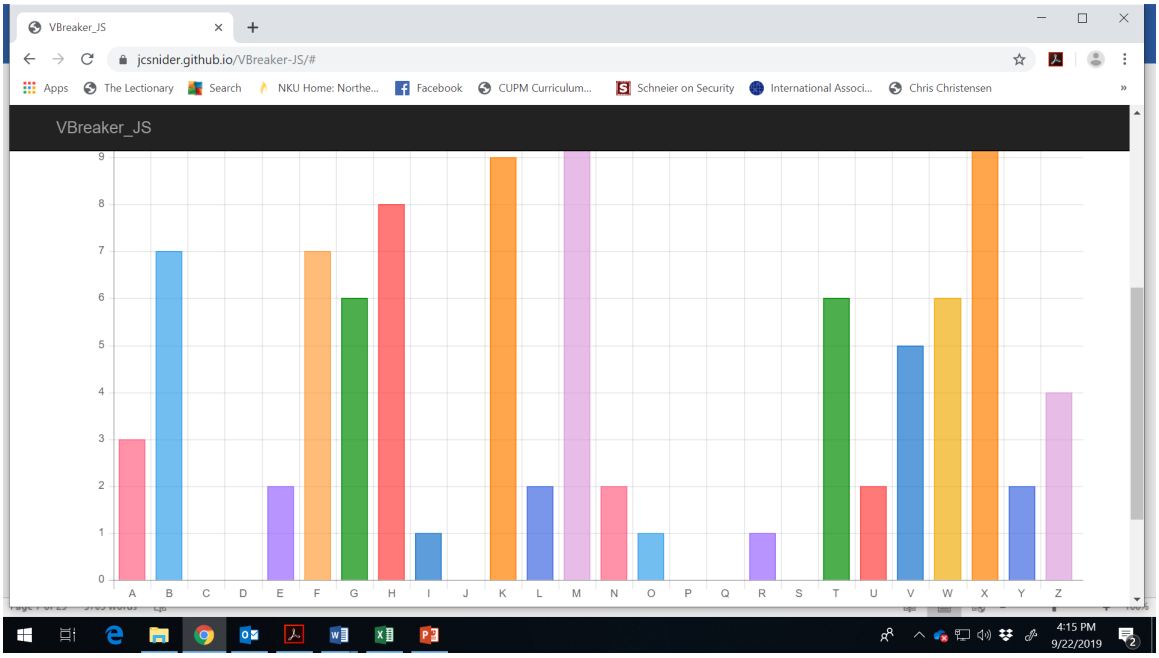


It appears that ciphertext `E` corresponds to plaintext `e`.

So, it appears that the third letter of the keyword is `a`.

The keyword is `u b o a _`.

Alphabet number five – fifth letters of each block



It appears that ciphertext `X` corresponds to plaintext `e`.

So, it appears that the last letter of the keyword is `t`.

The keyword is `u b o a t`.

So, knowing just the length of the keyword, we were able to determine the keyword.

So, how can we determine the length of the keyword?

Soon, but first, some history.

The Cryptanalysts: Charles Babbage and Friedrich Kasiski

The Vigenère cipher might first have been broken by the English mathematician Charles Babbage (1792 – 1871); Kahn quotes Babbage as saying "an indistinct glimpse of defeating it presented itself vaguely to my imagination." But, if Babbage had a solution, he never published it. Babbage apparently had the tendency to never be satisfied with a work and to continue to refine things; so, he might never have been satisfied enough with his solution to publish it.

Friedrich Kasiski (1805 – 1881) is credited with breaking the Vigenère cipher in 1863. From the Sixteenth Century until the Nineteenth Century the cipher was generally considered to be secure.

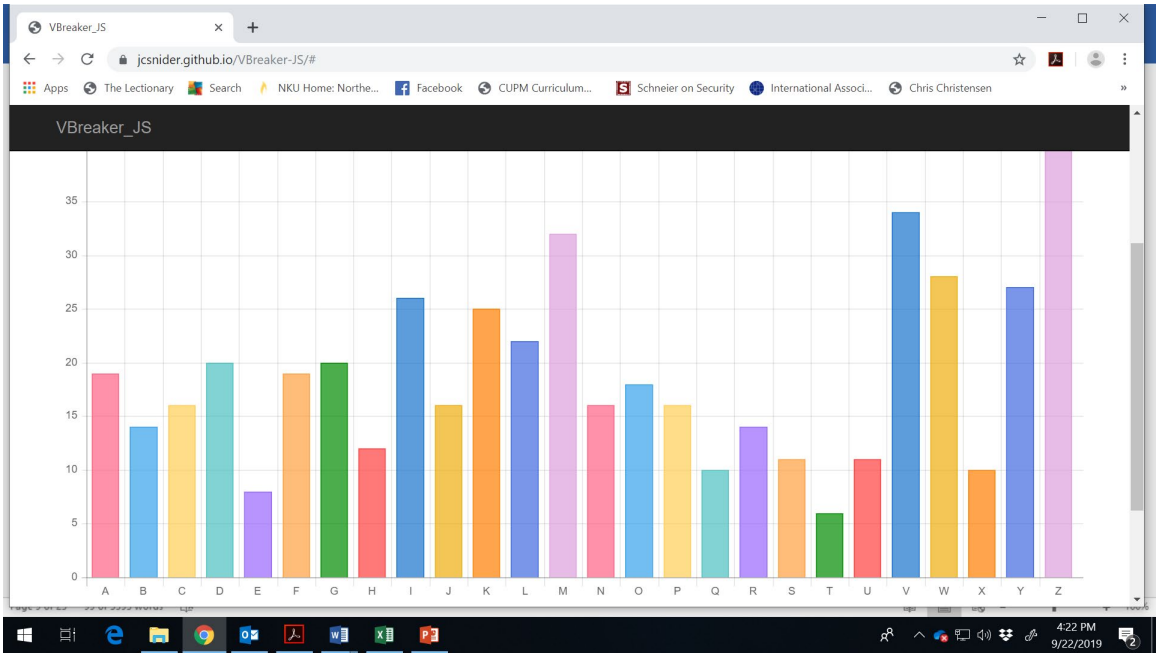
> Friedrich Kasiski was born in November 1805 in a western Prussian town and enlisted in an East Prussian infantry regiment at the age of 17.
>
> He moved up through the ranks to become a company commander and retired in 1852 as a major. Although he had become interested in cryptology during his military career, it was not until the 1860s that he put his ideas on paper. In 1863 his 95-page text *Die Geheimschriften und die Dechiffrirkunst* (*Secret Writing and the Art of Deciphering*) was published. A large part of its contents addressed the solution of polyalphabetic ciphers with repeating keywords, a problem that had tormented cryptanalysts for centuries.
>
> Disappointed by the lack of interest in his findings, Kasiski turned his attention to other activities including anthropology. He took part in artifacts searches and excavations and wrote numerous archeological articles for scholarly journals. He died in May 1881 not realizing the significance of his cryptanalytic findings." Wrixon, Fred B., *Codes, Ciphers & other Cryptic & Clandestine Communication: Making and breaking secret messages from hieroglyphs to the internet*, Black Dog & Leventhal Publishers.

Kasiski's Method

Here is a message enciphered with a Vigenère cipher. (It is taken from: Beutelspacher, Albrecht, *Cryptology: An introduction to the Art and Science of Enciphering, Encrypting, Concealing, Hiding and Safeguarding Described Without any Arcane Skullduggery but not Without Cunning Waggery for the Delectation and Instruction of the General Public*, Mathematical Association of America, 1996.)

```
DBZMG   AOIYS   OPVFH   OWKBW   XZPJL   VVRFG   NBKIX
DVUIM   OPFQL   VVPUD   KPRVW   OARLW   DVLMW   AWINZ
DAKBW   MMRLW   QIICG   PAKYU   CVZKM   ZARPS   DTRVD
ZWEYG   ABYYE   YMGYF   YAFHL   CMWLW   LCVHL   MMGYL
DBZIF   JNCYL   OMIAJ   JCGMA   IBVRL   OPVFW   OBVLK
OPVUJ   ZDVLQ   XWDGG   IQEYF   BTZMZ   DVRMM   ANZWA
ZVKFQ   GWEAL   ZFKNZ   ZZVCK   VDVLQ   BWFXU   CIEWW
OPRMU   JZIYK   KWEXA   IOIYH   ZIKYV   GMKNW   MOIIM
KADUQ   WMWIM   ILZHL   CMTCH   CMINW   SBRHV   OPVSO
DTCMG   HMKCE   ZASYD   JKRNW   YIKCF   OMIPS   GAFZK
JUVGM   GBZJD   ZWWNZ   ZVLGT   ZZFZS   GXYUT   ZBJCF
PAVNZ   ZAVWS   IJVZG   PVUVQ   NKRHF   DVXNZ   ZKZJZ
ZZKYP   OIEXX   MWDNZ   ZQIMH   VKZHY   DVKYD   GQXYF
OOLYK   NMJGS   YMRML   JBYYF   PUSYJ   JNRFH   CISYL
N
```

We begin the attack by frequency analysis.



The "relatively equal" frequencies suggest multiple alphabets – a polyalphabetic cipher, which, for us, would suggest a Vigenère cipher.

Here's the idea behind the Kasiski test. Consider a Vigenère cipher with keyword *Galois* (a nineteenth century mathematician).

```
abcdefghijklmnopqrstuvwxyz
GHIJKLMNOPQRSTUVWXYZABCDEF
ABCDEFGHIJKLMNOPQRSTUVWXYZ
LMNOPQRSTUVWXYZABCDEFGHIJK
OPQRSTUVWXYZABCDEFGHIJKLMN
IJKLMNOPQRSTUVWXYZABCDEFGH
STUVWXYZABCDEFGHIJKLMNOPQR
```

Think of a common trigraph – say, `the`. There are six possible encryptions of `the`.

```
GALOIS  GALOIS  GALOIS  GALOIS  GALOISG  GALOISGA
the      the      the      the      the       the
ZHP      TSS      EVM      HQW      BZK       LNE
```

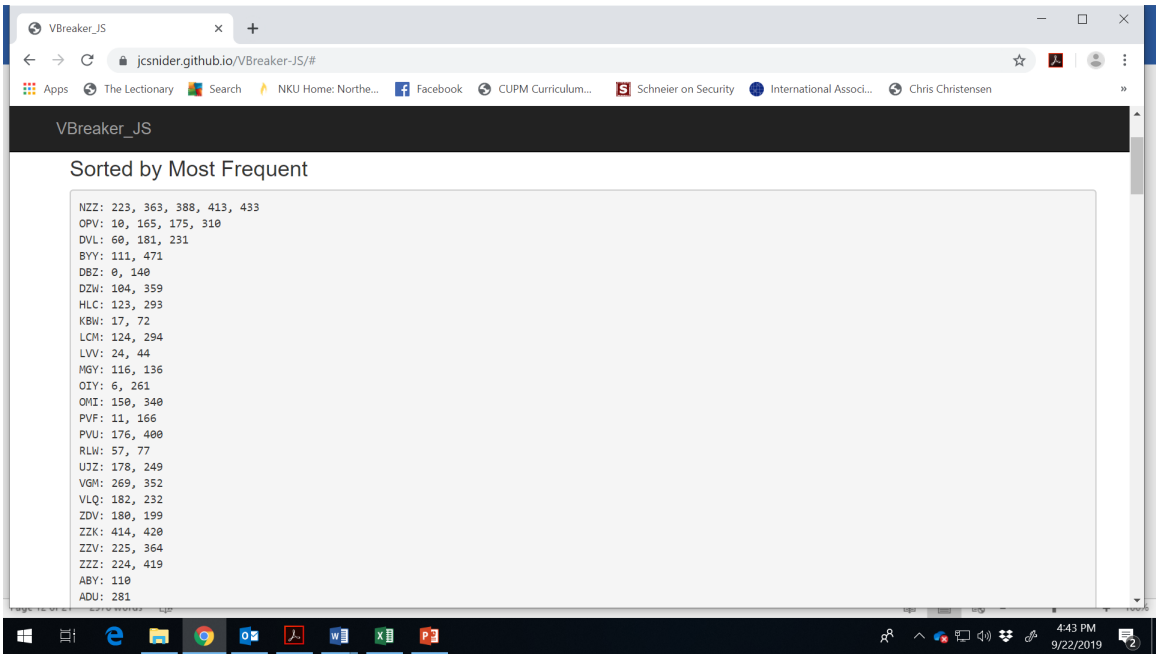
If there are more than six `the`'s in the plaintext, duplicate trigraphs will appear. Even if there were fewer than six `the`'s, duplicates might happen.

```
GALOISGALOISGALOISGALOISGALOIS...GALOISGALOISGALOIS
  the                                    the
  EVM                                    EVM
```

Notice that when such duplicates occur, the distance between the duplicates is a multiple of the length of the keyword.

So, we search through the ciphertext for repeated trigraphs (or strings of other lengths). Sometimes, of course, the repetitions are just accidental – two different strings of three letters are encrypted into the same three-letter string by different alphabets, but when the repetitions correspond to the same three-letter string being encrypted by the same three alphabets, we learn something about the length of the keyword. The latter are the occurrences that we would like to discover.

Here are the most frequent trigraphs in the ciphertext and their positions in the ciphertext:



Remember that we are looking for a length that is a common divisor of "all" of these lengths – well, not “all” because some repetitions are accidental – but most.

NZZ	363	388	413	433
	223	363	388	413
	140	25	25	20
OPV	165	175	310	
	10	165	175	
	155	10	135	
DVL	181	231		
	60	181		
	121	50		

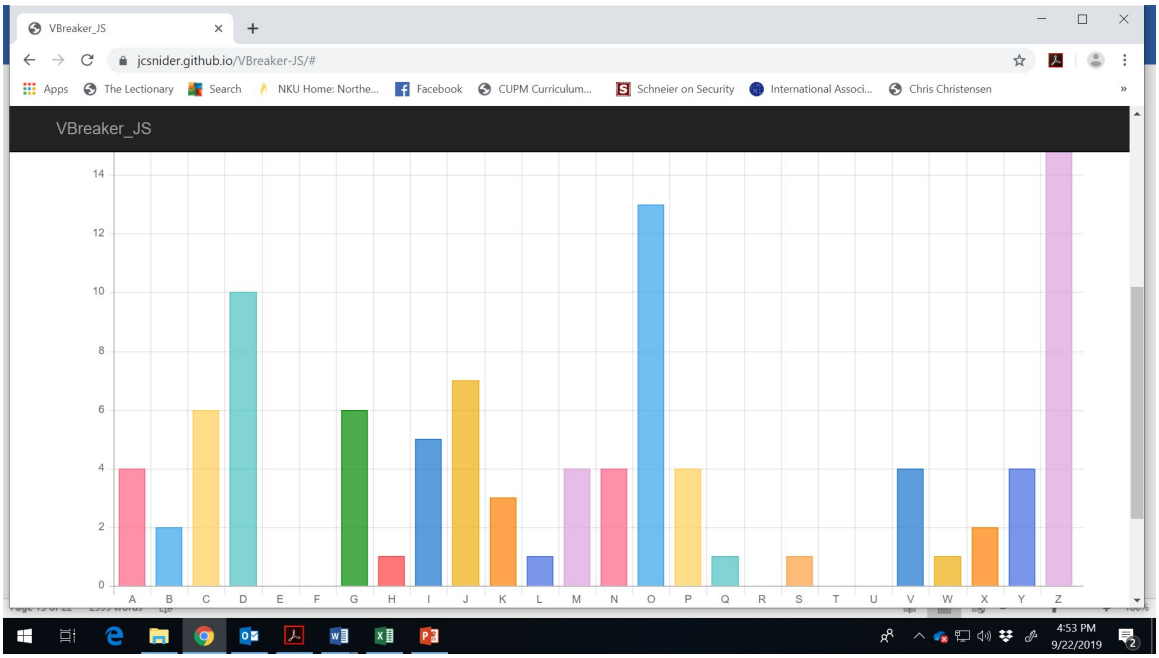
5 seems to be the most likely length for the keyword.

Now we return to the ciphertext and separate it into its five alphabets. We begin with the first letter and take every fifth letter after it. Then take the second letter and every fifth letter after it. Then take the third letter and every fifth after it. Etc.

If we determined the length of the keyword correctly, we should have partitioned the ciphertext into five sets of ciphertext letters each of which was encrypted with a Caesar cipher. Then we proceed as we did for the first example of this section.

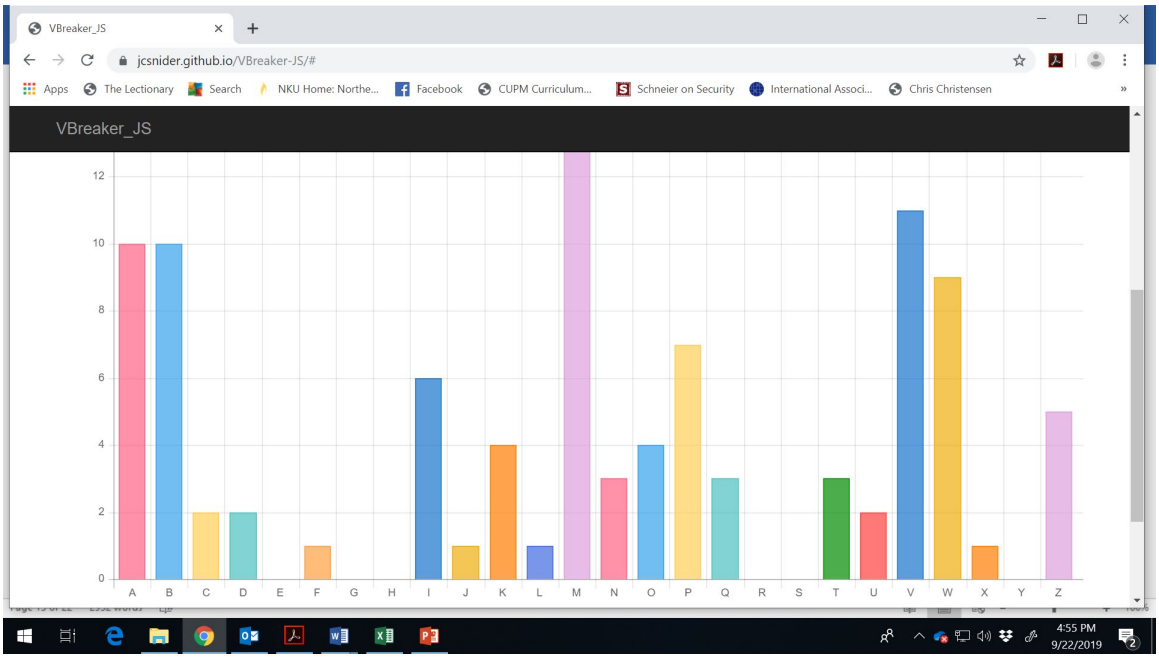
Let us look at each alphabet separately.

Alphabet number one



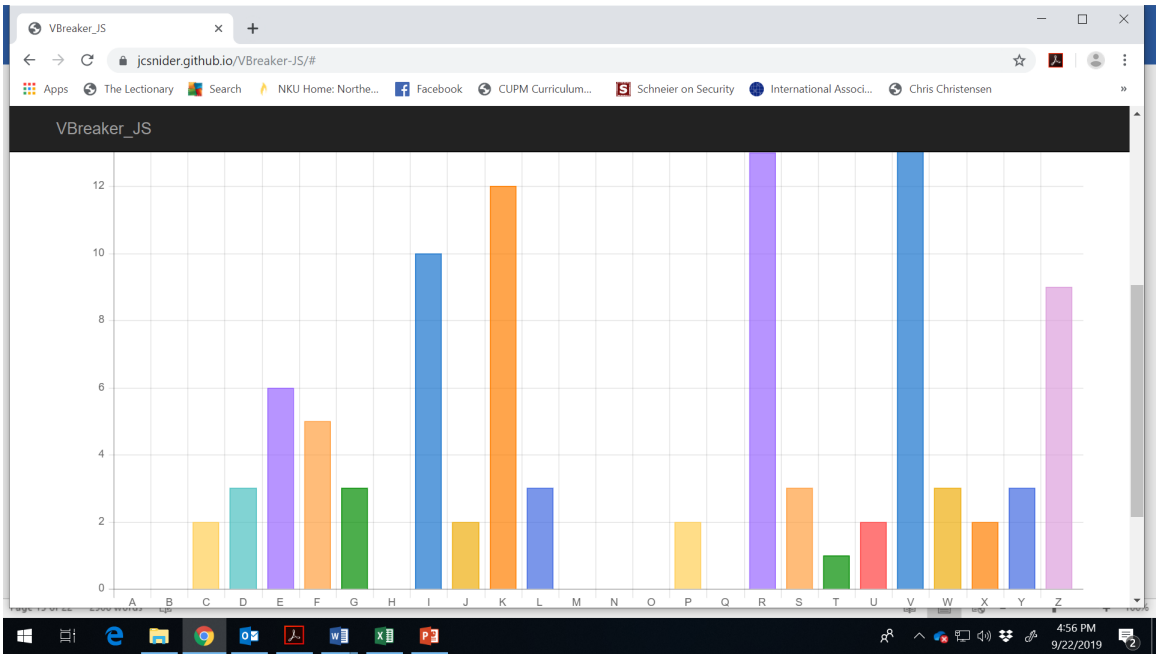
It appears that `Z` might correspond to `e`; that would make `V` correspond to `a`. The first letter of the keyword would be `v`. `v _ _ _ _`

Alphabet number two



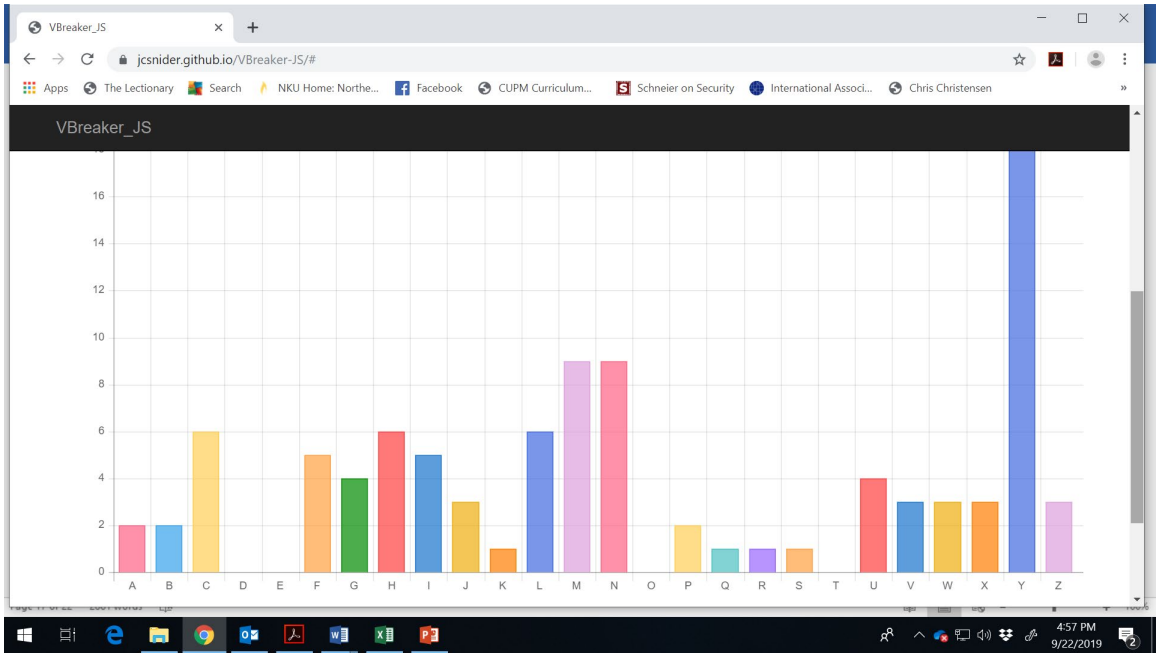
It appears that `M` might correspond to `e`; that would make `I` correspond to `a`. The second letter of the keyword would be `i`. `v i _ _ _` (Guess?)

Alphabet number three



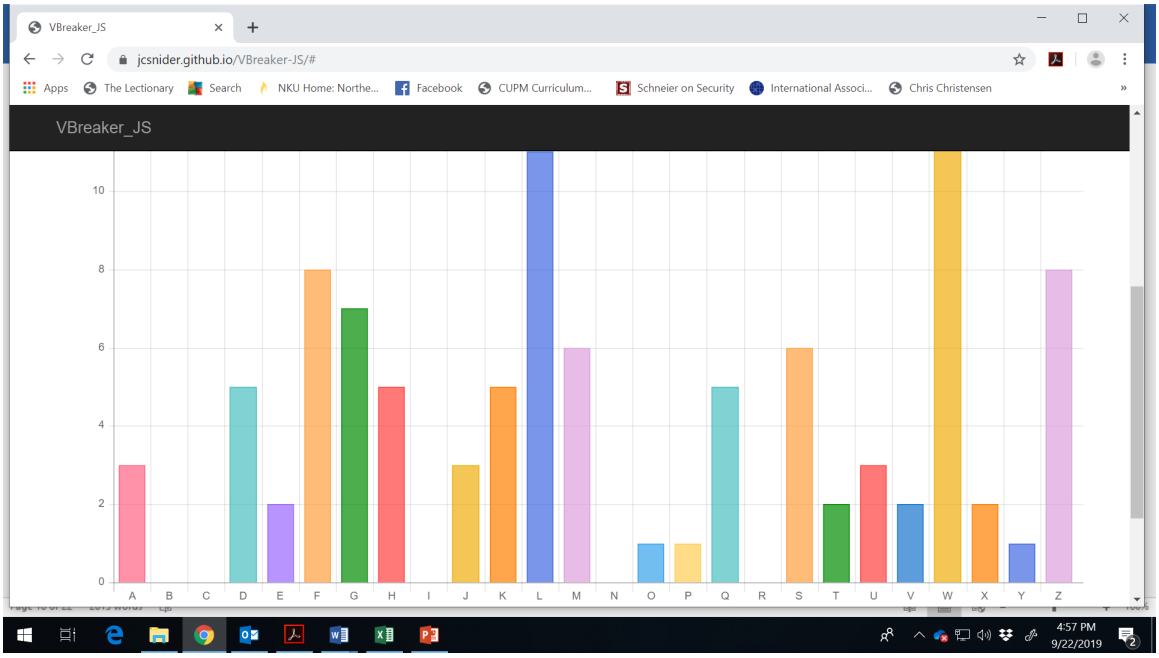
It appears that `V` might correspond to `e`; that would make `R` correspond to `a`. The third letter of the keyword would be `r. v i r _ _` (Guess?)

Alphabet number four



It appears that `Y` might correspond to `e`; that would make `U` correspond to `a`. The fourth letter of the keyword would be `u`. `v i r u _` (Not many possibilities.)

Alphabet number five



It appears that `W` might correspond to `e`; that would make `S` correspond to `a`. The fifth letter of the keyword would be `s`. The keyword would be `v i r u s`.

Now we can decipher the message:

```
ITISOF GREATH ELPTOT HECRYP TANALO STTOFI NDOUTH OWTANY ALPHAB ETSARE
INUSEF ORTHIS THEREA REVARI OUSTEC HNIQUE SAVAIL ABLEON EOFTHE MDEPEN
DSONTH EFREQU ENTREP ETITIO NOFLET TERGRO UPSINT EXTTHE LETTER STHEAR
EVERYC OMMONI NENGLI SHINAS UFFICI ENTLYL ONGTEX TTHERE ISAVER YGOODC
HANCET HASCOR RESPON DINGRE PEATED LETTER GROUPS MAYBEF OUNDIN THECIP
HERTEX TANDTH EYWILL SOMETI MESBEL OCATED ATINTE RVALSO FSOMEM ULTIPL
EOFTHE NUMBER OFALPH ABETSI NUSETH ESECAN BEFOUN DBYSCA NNINGT HECIPH
ERTEXT ANDFRO MTHEIR SPACIN GINTEL LIGENT GUESSE SMADEA STOTHE NUMBER
OFALPH ABETS
```

Of course, for Kasiski's method to work, it is necessary for the keyword to be repeated. In fact, what we depended upon was that we had a very long message and relatively short keyword so that the keyword was repeated many times and that when we stripped off the various Caesar cipher alphabets each alphabet contained enough letters to enable us to spot the shift.