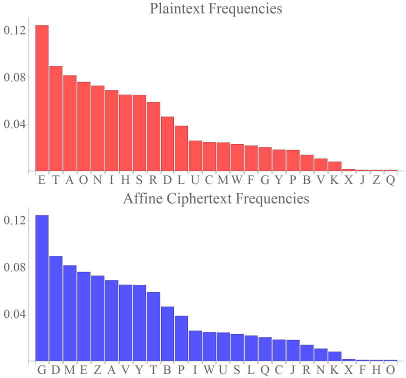
The Vigenère Cipher

Masking Letter Frequencies

CS 363 Computer Security

Letter Frequencies

The affine cipher (monoalphabetic substitution cipher) does not change the letter frequencies.



Letter Frequencies

Affine cipher: $E(x) = (a \cdot x + b) \pmod{26}$

For example, we can associate ciphertext `G` and `D` with plaintext `E` and `T`, respectively, and solve

$$\left.\begin{aligned} 4m + k &\equiv 6 \\ 19m + k &\equiv 3 \end{aligned}\right\} \quad \mathrm{mod}\ 26$$

to determine that $m = 5$ and $k = 12$.

Letter Frequency Masking

A good cipher masks the letter frequencies that exist in the plaintext.

How can we tell if frequencies are masked by a cipher?

One way is to look at the distribution of character frequencies, but we can also compute an important statistic instead.

William Friedman's Index of Coincidence (1925)

Let A_1 be the event that you get an a as the first chosen letter and A_2 be the event that you get an a as the second letter, etc... If n is the total number of characters in the text and there are n_1 a's, n_2 b's, etc..., then

$$\begin{aligned}
\text{Index of Coincidence} &= P(\text{two randomly chosen letters are the same}) \\
&= P\left[(A_1 \cap A_2) \cup (B_1 \cap B_2) \cup \ldots \cup (Z_1 \cap Z_2)\right] \\
&= P(A_1 \cap A_2) + P(B_1 \cap B_2) + \ldots + P(Z_1 \cap Z_2) \\
&= \left(\frac{n_1}{n}\right)\left(\frac{n_1 - 1}{n - 1}\right) + \left(\frac{n_2}{n}\right)\left(\frac{n_2 - 1}{n - 1}\right) + \ldots \left(\frac{n_{26}}{n}\right)\left(\frac{n_{26} - 1}{n - 1}\right) \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{26} n_i(n_i - 1) \\
&\approx \frac{1}{n^2} \sum_{i=1}^{26} n_i^2.
\end{aligned}$$

Example

``The quick brown fox jumps over the lazy dog.'' is a short sentence of 36 characters that famously uses each letter of the alphabet at least once.

Only 6 letters are used more than once: e (three times), h (twice), o (four times), r (twice), t (twice), and u (twice).

$$\begin{aligned}
\text{IoC} &= \frac{1}{n(n-1)} \sum_{i=1}^{26} n_i(n_i - 1) \\
&= \frac{1}{35(34)} \left[3(2) + 2(1) + 4(3) + 2(1) + 2(1) + 2(1)\right] \\
&= \frac{26}{1190} \\
&= \frac{13}{595} \\
&\approx 0.0218
\end{aligned}$$

More Examples

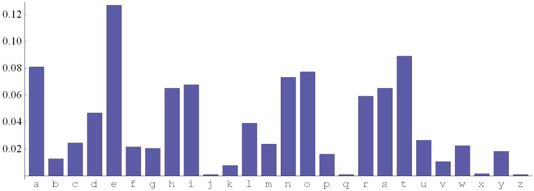
What are more typical values of the IoC for 26-letter English?

Text	Number of Characters	IoC
“The Gold Bug”	58,270	0.066
2006 State of the Union Address	25,940	0.066
“Julius Caesar”	86,699	0.064
USA Patriot Act	286,260	0.070

What should we expect the IoC to be for 26-letter English?

Letter Frequencies (English)

Letter	Relative Frequency
a	0.082
b	0.014
c	0.025
d	0.046
e	0.124
f	0.022
g	0.020
h	0.065
i	0.069
j	0.001
k	0.008
l	0.039
m	0.024
n	0.073
o	0.076
p	0.018
q	0.001
r	0.059
s	0.065
t	0.089
u	0.026
v	0.011
w	0.023
x	0.002
y	0.018
z	0.001



Theoretical Index of Coincidence

For a sufficiently long text, all letter pair events (like A_1 and A_2) should be almost independent, so

$$\begin{aligned}\text{IoC} &= P(A_1 \cap A_2) + P(B_1 \cap B_2) + \ldots + P(Z_1 \cap Z_2) \\ &= P(A_1)P(A_2|A_1) + P(B_1)P(B_2|B_1) + \ldots + P(Z_1)P(Z_2|Z_1) \\ &\approx P(A_1)^2 + P(B_1)^2 + \ldots + P(Z_1)^2 \\ &\approx (0.082)^2 + (0.014)^2 + \ldots + (0.001)^2 \\ &\approx 0.0658\end{aligned}$$

using the letter frequencies on the previous slide. In other words, in long English texts, there is about a 6.6% chance that two randomly selected letters are the same.

Other Indices of Coincidence

Language	IoC
Arabic	0.0759
Danish	0.0707
Finnish	0.0738
French	0.0746
German	0.0767
Greek	0.0692
Hebrew	0.0768
Italian	0.0733
Japanese	0.0772
Malay	0.0853
Norwegian	0.0694
Portuguese	0.0745
Russian	0.0561
Serbo Croatian	0.0644
Spanish	0.0766
Swedish	0.0645

The Index of Coincidence for Ciphertext

A necessary condition for a good cipher is that it masks all of the letter frequencies, so let's assume that every letter in the ciphertext is equally likely, with probability $1/26$. Then

$$\begin{aligned}\text{IoC} &\approx P(A_1)^2 + P(B_1)^2 + \ldots + P(Z_1)^2 \\ &= \left(\frac{1}{26}\right)^2 + \left(\frac{1}{26}\right)^2 + \ldots + \left(\frac{1}{26}\right)^2 \\ &= \frac{26}{26^2} \\ &= \frac{1}{26} \\ &\approx 0.0385.\end{aligned}$$

So, the better a cipher masks letter frequencies, the closer the IoC of the ciphertext is to 0.0385.

Homophonic Cipher

The simplest approach to disguising the frequencies of a letter is to use multiple symbols for each letter. The number of symbols should be proportionate to the frequency of the letter.

Letter	Alternate Characters
a	> 8 & ¶ ? Å ± r̊
b	μ
c	; § ©
d	! #) $\int$
e	7 ∪ * □ $\oint$ % ® ∘ Q Æ ⊕ ⊔ ◁
f	2 @
⋮	⋮

What are the advantages and disadvantages of this polyalphabetic substitution cipher?

Example

a	15, 33, 37, 55, 57, 72, 91, 96
b	24
c	03, 39, 67
d	04, 43, 61, 88
e	08, 12, 20, 46, 47, 59, 64, 79, 81, 85, 90, 94, 97
f	40, 48
g	29, 53
h	05, 16, 30, 42, 69, 99
i	14, 45, 50, 60, 73, 82, 93
j	11
k	77
l	01, 26, 71, 98
m	34, 87
n	06, 17, 22, 31, 49, 58
o	02, 10, 41, 51, 66, 75, 83
p	13, 18
q	36
r	21, 25, 65, 68, 92, 95
s	00, 28, 52, 63, 74, 78
t	07, 19, 23, 35, 38, 54, 70, 84, 89
u	09, 32
v	44
w	56, 80
x	86
y	62, 76
z	27

`Remember the Alamo!` can be encrypted in 795,971,764,224 ways:

68903459872447687099795501728766

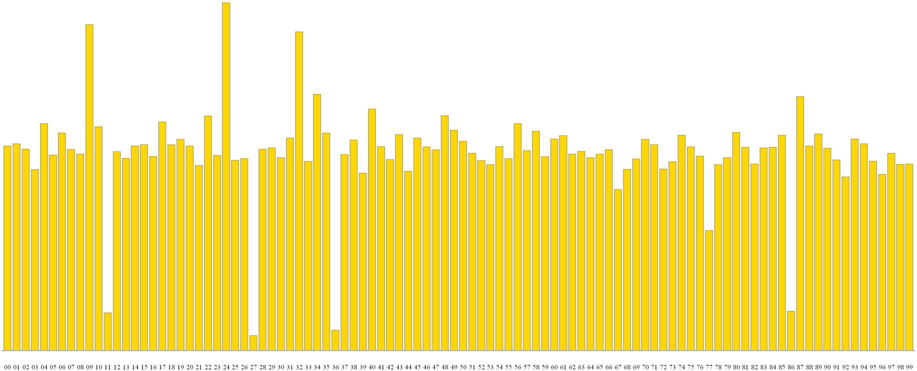
or

25793447872479920730945571723475

or

92853490342497213816943371558783 etc. . .

A Homophonic Cipher Disguises Letter Frequencies



Numerical code frequencies $\{00, 01, 02, \ldots, 99\}$ for a homophonic encryption of Edgar Allan Poe's "The Gold Bug" using the table on the previous slide. Note how flat the distribution is.

The Vigenère Cipher

	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	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
b	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
c	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	B
d	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
e	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
f	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
g	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
h	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
i	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
j	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
k	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
l	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
m	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
n	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
o	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
p	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
r	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
s	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
t	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
u	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
v	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
w	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
x	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
z	Z	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

plain	i	c	o	u	l	d	c	r	u	s	h	y	o	u	l	i	k	e	a	w	o	r	m
key	h	a	m	i	s	h	h	a	m	i	s	h	h	a	m	i	s	h	h	a	m	i	s
cipher	P	C	A	C	...																		

The plaintext letter corresponds to the column and the key letter corresponds to the row.

A Mathematical Version of the Vigenère Cipher

Encryption and decryption are much easier using modular arithmetic than using the Vigenère square.

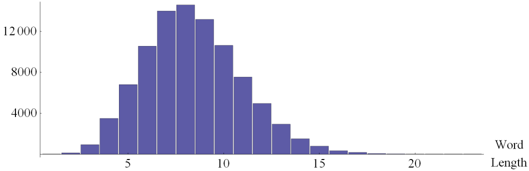
plain	I	C	O	U	L	D	C	R	U	S	H	Y	O	U	L	I	K	E	A	W	O	R	M
	8	2	14	20	11	3	2	17	20	18	7	24	14	20	11	8	10	4	0	22	14	17	12
key	H	A	M	I	S	H	H	A	M	I	S	H	H	A	M	I	S	H	H	A	M	I	S
	7	0	12	8	18	7	7	0	12	8	18	7	7	0	12	8	18	7	7	0	12	8	18
cipher	P	C	A	C	D	K	J	R	G	A	Z	F	V	U	X	Q	C	L	H	W	A	Z	E
	15	2	0	2	3	10	9	17	6	0	25	5	21	20	23	16	2	11	7	22	0	25	4

Note that the Vigenère cipher is really just a combination of shift ciphers. If an adversary knew that the length of the keyword is 7, then he/she would only have to break 7 shift ciphers.

Keyspace for the Vigenère Cipher (English Keywords)

How long can the keywords be if we only use English words? According to *Mathematica*, there are only seven words with more than 20 letters, so it seems reasonable to restrict our attention for the time being to words up to length 20.

If we insist on actual English words for keywords, then there are $92,518 \approx 2^{16.5}$ words in *Mathematica*'s dictionary. That sounds like a big number, but it's not for a computer.



Keyspace for the Vigenère Cipher (Non-English Keywords)

If we accept any string of characters up to and including 20 letters, then there are

$$26+26^2+26^3+\ldots+26^{20} = 20,725,274,851,017,785,518,433,805,270 \approx 2^{94}$$

possible keywords. That looks like a big number (20 octillion plus change), and it is – even for a modern computer. So exhaustion is out of the question in this case.

Strength of the Vigenère Cipher

Blaise de Vigenère is credited with inventing his namesake cipher in 1585 and it was considered to be unbreakable for centuries. It was broken by Charles Babbage in 1854 and by Friedrich Kasiski in 1863. Nevertheless, the Confederate Army used it (often incorrectly) during the American Civil War.

From Scientic American, Supplement LXXXIII on January 27, 1917, referring to the Vigenère cipher:

> *"The Vigenère cipher method used for the preparation and reading of code messages is simple in the extreme and at the same time impossible of translation unless the key is known. The ease with which the key may be changed is another point in favor of the adoption of this code by those desiring to transmit important messages without the slightest danger of their messages being read by political or business rivals etc."*

Periodicity of the Vigenère Cipher

Plain:	howmuchwoodwouldawoodchuckchuckifawoodchuckcouldchuckwood
Key:	twisttwisttwisttwisttwisttwisttwisttwisttwisttwisttwisttw
Cipher:	akeenvdeghwswmewweghwypmvdypmvdensphkluanysuhnhluanysohhz

Sometimes strings of plaintext characters (*e.g.* `wood`, `chuck`, `dchuck`) are encrypted the same way and sometimes they aren't. Why?

Polygraphs in Cipher

Here is a message encrypted with the Vigenère cipher using a key we don't know.

```
OIGGBWAGNUZSBYHTWSQPAMMVXIGGOUFWNLEQAIGVQNRDANTDWNTXBWACCCZTWNMCNQZPCCACLIZRNCHTMCZARVQGCSMC
MXQSRWMINXFDCBQEAIBDBCFXXHFWJNMAUGQCJLQRAYMINXQFDUXCXQITJLQTWAMVNXUCJADTJNOXECXLJLFTBNUCPQTT
CBQGCBMIWUFXXHAGJHKCJNUDWMARXHOTRPQSJHPHXXQSRWMINXOPWFACPYZSDLQLNUDTVYFDWUSGNUFQJNFANZUTUXAU
CBMIFUDLNBMKNWABNNASNXURJNQPYIDIRIZDONTPCZUTUXMHJZUCJFDTBNUCPJXPLYRDANTDBYIWXBQGNAMKNNTTRLXX
EYEIQUFIQUFCJNUDWGUVQNXXEYUIRMMACISTCBQGOCFIRHSPWXBGXJQGCBMIFYEWXOXSMIFWRMNJCCZPUUDVNLETWMQL
NWMCWIFSNXURJNQLNWMCWIFRXHETLLMINQQRJHZDCBMAUIIIQCEVAIGCMNTTKLMKNGQCUCHXWAMCMXQPMQTDBNDJPAXT
MBQGNBMKNWACBYOGJNQSRNRPAUNDEYAJAJADAJALNLFDJXPDAXQIAUOICBQLXLXSFCXAUCFIUYZDCYZDAFACPLQBNGNT
AQTPCQQHJSTTAYNJCCFRJHZTEYDUXLSTCQTPCNTTHXUSQYDTRNUHOIDJBNTTUCHXWADPCBQGCINTMYPXLUFTMBQGNNAI
QYGCOCZXBBQSFIDZFBURQNTTHQTDOIGVQNTTAYTPEYFWDMRPAMACXVXNJXHPWWQSRNUHAUFWNLRDAOEIXVQWNLQSNXUR
JNQSCIFWNADTJNFPBEDTVUUCRHSQNZAGNOEIQUFUAIYIQYETQIZDAYPSNUPLNNMZNCZRAYMHNXPTEIFXXHFDCBMILUGH
NZAGFBURQNTTHAMKNNTTUUEIOOXAVYMHDLQDOXQKXNUDWNTPCQQWNLQWRATAHLQHXFHTCBMICBQHNXQPMMTPUFZDCBMK
NXUTMCZKJCZIQUFIQCECJNUDWOZSNLSDMMTPUFTPEYMCNQNXANTDOZDTNXABJHPIQUFVXPQGWGQCCIRIQYBTXJXTKSFW
NJQDYFQUXLFWNJQDYFQHQUXAWIFENLUHQZDDVNTTNUDIQ
```

The Kasiski Test

If a string of characters appears in a polyalphabetic ciphertext message, it is possible that the distance between the occurrences is a multiple of the length of the keyword.

For the cipher on the previous slide,

Polygraph	Starting Positions	Differences
SNXURJNQ	296	$172 = 2^2 \cdot 43$
	468	$356 = 2^2 \cdot 89$
	824	
XQSRWMIN	94	$132 = 2^2 \cdot 3 \cdot 11$
	226	
WNJQDYFQ	1104	$12 = 2^2 \cdot 3$
	1116	

The common factors are 2 and 4, so the keyword is probably 4 letters long.

Vigenère Cipher Assuming a 4-Letter Keyword

```
OIGGBWAGNUZSBYHTWSQPAMMVXIGGOUFWNLEQAIGVQNRDANTDWNTXBWACCCZTWNMCNQZPCCACLIZRNCHTMCZARVQGCSMC
MXQSRWMINXFDCBQEAIBDBCFXXHFWJNMAUGQCJLQRAYMINXQFDUXCXQITJLQTWAMVNXUCJADTJNOXECXLJLFTBNUCPQTT
CBQGCBMIWUFXXHAGJHKCJNUDWMARXHOTRPQSJHPHXXQSRWMINXOPWFACPYZSDLQLNUDTVYFDWUSGNUFQJNFANZUTUXAU
CBMIFUDLNBMKNWABNNASNXURJNQPYIDIRIZDONTPCZUTUXMHJZUCJFDTBNUCPJXPLYRDANTDBYIWXBQGNAMKNNTTRLXX
EYEIQUFIQUFCJNUDWGUVQNXXEYUIRMMACISTCBQGOCFIRHSPWXBGXJQGCBMIFYEWXOXSMIFWRMNJCCZPUUDVNLETWMQL
NWMCWIFSNXURJNQLNWMCWIFRXHETLLMINQQRJHZDCBMAUIIIQCEVAIGCMNTTKLMKNGQCUCHXWAMCMXQPMQTDBNDJPAXT
MBQGNBMKNWACBYOGJNQSRNRPAUNDEYAJAJADAJALNLFDJXPDAXQIAUOICBQLXLXSFCXAUCFIUYZDCYZDAFACPLQBNGNT
AQTPCQQHJSTTAYNJCCFRJHZTEYDUXLSTCQTPCNTTHXUSQYDTRNUHOIDJBNTTUCHXWADPCBQGCINTMYPXLUFTMBQGNNAI
QYGCOCZXBBQSFIDZFBURQNTTHQTDOIGVQNTTAYTPEYFWDMRPAMACXVXNJXHPWWQSRNUHAUFWNLRDAOEIXVQWNLQSNXUR
JNQSCIFWNADTJNFPBEDTVUUCRHSQNZAGNOEIQUFUAIYIQYETQIZDAYPSNUPLNNMZNCZRAYMHNXPTEIFXXHFDCBMILUGH
NZAGFBURQNTTHAMKNNTTUUEIOOXAVYMHDLQDOXQKXNUDWNTPCQQWNLQWRATAHLQHXFHTCBMICBQHNXQPMMTPUFZDCBMK
NXUTMCZKJCZIQUFIQCECJNUDWOZSNLSDMMTPUFTPEYMCNQNXANTDOZDTNXABJHPIQUFVXPQGWGQCCIRIQYBTXJXTKSFW
NJQDYFQUXLFWNJQDYFQHQUXAWIFENLUHQZDDVNTTNUDI
```

Let’s extract the cipher by color.

Extracted Subsequences of Vigenère Ciphertext

```
OBNBWAXONAQAWBCWNCLNMRCMRNCABXJUJANDXJWNJJEJBPCCWXJJWXRJXRNWPDNVWNJNUCFNNN
NJYROCUJJBPLABXNNREQQJWQERCCORWXCFXMRCUNWNWNJNWXLNJCUQAMKNUWMMBPMNNBJRAEAA
NJAACXFUUCAPNACJACJEXCCHQROBUWCCMLMNQOBFFQHOQAEDAXJWRANAXNNJCNJBVRNNQAQQAN
NNANEXCLNFQHNUOVDOXWCNRHXCCNMUCNMJQQJWNMUENAONJQXWCQXKNYXNYQWNQVN
```

IoC ≈ 0.077

```
IWUYSMIULINNNWCNQCICCVSXWXBICHNGLYXUQLAXANCLNQBBUHHNMHPHXWXFYLUYUUNZXBUBWN
XNIINZXZFNJYNYBANLYUUNGNYMIBCHXJBYOIMCULMWIXNWIHLQHBICINLGCAXQNABBWYNNUYJJ
LXXUBLCCYYFLGQQSYCHYLQNXYNINCABIYUBNYCBIBNQINYYMMVXWNULOVLXNIANEUHZOUIYIYU
NCYXIHBUZBNANUOYLXNNQLALFBBXMFBXCCUCNOLMFYQNZXHUPGIYJSJFLJFUILZNU
```

IoC ≈ 0.067

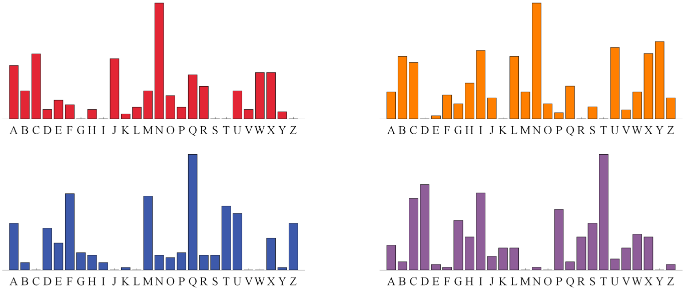
```
GAZHQMGFEGRTTAZMZAZHZQMQMFQBFFMQQMQXIQMUDOXFUTQMFAKUAOQPQMOAZQDFSFFUAMDMAA
UQDZTUMUDUXRTIQMTXEFFUUXUMSQFSBQMEXFNZDEQMFUQMFEMQZMIEGTMQHMQTDXQMAOQRNAAA
FPQOQXXFZZAQNTQTNFZDSTTUDUDTHDQNPFQAGZQDUTTGTTFRAXHQUFREQQUQFDFDUSAEFYEZPP
MZMPFFMGAUTMTEXMQQUTQQTQHMQQTZMUZZFEUZSTTMNTDAPFQQRBXFQQFQQXFUDTD
```

IoC ≈ 0.080

```
GGSTPVGWQVDDXCTCPCRTAGCSIDEDXWACRIFCTTVCTXLTCTGIXGCDRTSHSIPCSLTDGQATUILKBS
RPIDPTHCTCPDDWGKTXIICDVXIATGIPGGIWSWJPVTLCSRLCRTIRDAIVCTKCXCPDJTGKCGSPDJDL
DDIILSAIDDCBTPHTJRTUTPTSTHJTXPGTXTGICXSZRTDVTPWPCNPSHWDIWSRSWTPTCQGIUITDSL
ZRHTXDIHGRTKTIAHDKDPWWAHTIHPPDKTKIICDSDPPCXDTBIVGCITTWDUWDHAEHDTI
```

IoC ≈ 0.072

Can We Find the Key?



The frequency charts suggest that e encrypts to:

		Putative
Image	Shift	Keyword
of e	Size	Letter
N	9	j
N	9	j
Q	12	m
T	15	p

Decrypting with `jjmp`

```
fzursnorelndspvenjeardagozurflthecsbrzugheforehonehisnonttneneanehnatton
czncetvedtnlimertjandoedinateototseprzposttioythaeallxenacecrpateoequlln
ohweaceenrageoinarreaecivtlwacteseinghhetsertsatnltioyorayynaeiondocoyce
igedaydsooedinateocanwongpnducewelremptonlgreltbaetleqielooftsatwlrwesav
enomeeodeoicaeeapzrtiznofehatqieloasaqinawreseingalacpforehospwhosererav
eeheiclivpsthltthltnaeionxighelivpitidaltzgetserftttiyganoproaertsatwpsh
oflddzthidbuttnallrgecsendewenannztdeoicaeewenannztcoyseccatehecaynotsal
lzwthtsgrzundehebcavexenltvinrandoeadhhoserugrledseresavenonspcraeediefa
rlbovpouraooraowectoaodoroetrlcttsewocldwtllltttlpnotpnorwongcemexberhha
thesajherpbutttcaynevprfocgethhateheyoidhpreieisfzruseheltvinrratsertzbe
dpdicltedsereeothpunftnissedwzrkwsicheheyhhofzugheherphavpthudfardonomly
aovannedieisrlthecforfstomehecedeoicaeedtztherreaetasvremliniygbeqorefst
hltfrzmthpsehznorpddeldweeaketncrpaseodevztioytotsatcluseqorwsicheheyrav
eehellstffllmpasuceofoevoeionehathehecehirhlycesowvetsattseseoeaddhalwno
tsaveoiedtnvatnthltthtsnaeionfndecgoddhalwhavpanehbirehofqreeoomaydthltg
ogernxentzfthppeoalebjtheaeopwefoctheaeopweshlllnztpecishqromeheelrth
```

This doesn't look quite right. What should we do?

Decrypting with `jump`

```
fourscoreandsevenyearsagoourfathersbroughtforthonthiscontinentanewnation
conceivedinlibertyanddedicatedtothepropositionthatallmenarecreatedequaln
owweareengagedinagreatcivilwartestingwhetherthatnationoranynationsoconce
ivedandsodedicatedcanlongendurewearemetonagreatbattlefieldofthatwarwehav
ecometodedicateaportionofthatfieldasafinalrestingplaceforthosewhoheregav
etheirlivesthatthatnationmightliveitisaltogetherfittingandproperthatwesh
oulddothisbutinalargersensewecannotdedicatewecannotconsecratewecannothal
lowthisgroundthebravemenlivinganddeadwhostruggledherehaveconsecrateditfa
raboveourpoorpowertoaddordetracttheworldwilllittlenotenorlongrememberwha
twesayherebutitcanneverforgetwhattheydidhereitisforusthelivingrathertobe
dedicatedheretotheunfinishedworkwhichtheywhofoughtherehavethusfarsonobly
advanceditisratherforustobeherededicatedtothegreattaskremainingbeforeust
hatfromthesehonoreddeadwetakeincreaseddevotiontothatcauseforwhichtheygav
ethelastfullmeasureofdevotionthatweherehighlyresolvethatthesedeadshallno
thavediedinvainthatthisnationundergodshallhaveanewbirthoffreedomandthatg
overnmentofthepeoplebythepeopleforthepeopleshallnotperishfromtheearth
```

Using the IoC to Find the Keyword Length

Let's measure the IoC of subsequences of the ciphertext.

Keyword	Subsequence									
Length	1	2	3	4	5	6	7	8	9	10
1	0.046									
2	0.058	0.049								
3	0.044	0.048	0.047							
4	0.077	0.067	0.081	0.072						
5	0.043	0.051	0.044	0.046	0.044					
6	0.055	0.055	0.060	0.047	0.058	0.047				
7	0.045	0.045	0.048	0.044	0.052	0.043	0.045			
8	0.074	0.067	0.080	0.065	0.077	0.062	0.080	0.076		
9	0.041	0.050	0.044	0.043	0.042	0.061	0.048	0.048	0.044	
10	0.054	0.051	0.054	0.052	0.057	0.049	0.061	0.047	0.062	0.045

Keywords of length 4 and 8 seem to give subsequences with IoCs compatible with English.

The Fibonacci Sequence

Let $a_0 = 0$, $a_1 = 1$, and

$$a_n = a_{n-1} + a_{n-2}, \qquad n \geq 2.$$

This generates the sequence

$$\{a_n\}_{n=0}^{\infty} = \{0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots\},$$

which is known as the Fibonacci sequence after the Italian mathematician Leonardo of Pisa (*a.k.a.* Leonardo Fibonacci - circa 1170-1250) who introduced Arabic numerals into the Latin West in his text *Liber Abaci*.

A Modular Fibonacci Sequence

Let $a_0 = 0$, $a_1 = 1$, and

$$a_n \equiv a_{n-1} + a_{n-2} \mod 26, \qquad n \geq 2.$$

This generates the sequence

$$\{a_n\}_{n=0}^{\infty} = \{0, 1, 1, 2, 3, 5, 8, 13, 21, 8, 3, 11, \ldots\}$$

that can be used as an expanded key in a Vigenère cipher. Note that the secret key now includes the initial keyword (0, 1 in this case) and the form of the recursion.

Periodicity of the Expanded Key

The sequence defined by $a_0 = 0$, $a_1 = 1$, and $a_n \equiv a_{n-1} + a_{n-2} \mod 26$, $n \geq 2$,

$$
\begin{aligned}
\{a_n\}_{n=0}^{\infty} = \{&0, 1, 1, 2, 3, 5, 8, 13, 21, 8, 3, 11, 14, 25, 13, 12, 25, 11, 10, 21, 5, 0, 5, 5, 10, 15, 25, 14,\\
&13, 1, 14, 15, 3, 18, 21, 13, 8, 21, 3, 24, 1, 25, 0, 25, 25, 24, 23, 21, 18, 13, 5, 18, 23, 15, 12, 1,\\
&13, 14, 1, 15, 16, 5, 21, 0, 21, 21, 16, 11, 1, 12, 13, 25, 12, 11, 23, 8, 5, 13, 18, 5, 23, 2, 25, 1,\\
&0, 1, 1, 2, 3, 5, 8, 13, 21, 8, 3, 11, 14, 25, 13, 12, 25, 11, 10, 21, 5, 0, 5, 5, 10, 15, 25, 14,\\
&13, 1, 14, 15, 3, 18, 21, 13, 8, 21, 3, 24, 1, 25, 0, 25, 25, 24, 23, 21, 18, 13, 5, 18, 23, 15, 12, 1,\\
&13, 14, 1, 15, 16, 5, 21, 0, 21, 21, 16, 11, 1, 12, 13, 25, 12, 11, 23, 8, 5, 13, 18, 5, 23, 2, 25, 1,\\
&0, 1, 1, 2, 3, 5, 8, 13, 21, 8, 3, 11, 14, 25, 13, 12, 25, 11, 10, 21, 5, 0, 5, 5, 10, 15, 25, 14,\\
&13, 1, 14, 15, 3, 18, 21, 13, 8, 21, 3, 24, 1, 25, 0, 25, 25, 24, 23, 21, 18, 13, 5, 18, 23, 15, 12, 1,\\
&\ 13, 14, 1, 15, 16, 5, 21, 0, 21, 21, 16, 11, 1, 12, 13, 25, 12, 11, 23, 8, 5, 13, 18, 5, 23, 2, 25, 1, \ldots\},
\end{aligned}
$$

is periodic with period 84. That is, $a_{n+84} = a_n$. Thus we have extended a “keyword” of length 2 to a “keyword” of length 84, which is a substantial improvement.

Some recursions are better than others. For example, $a_n \equiv a_{n-1} + a_{n-3} \mod 26$ has period twice as long; $a_{n+168} = a_n$. Finding the best recursions is beyond the scope of this course.

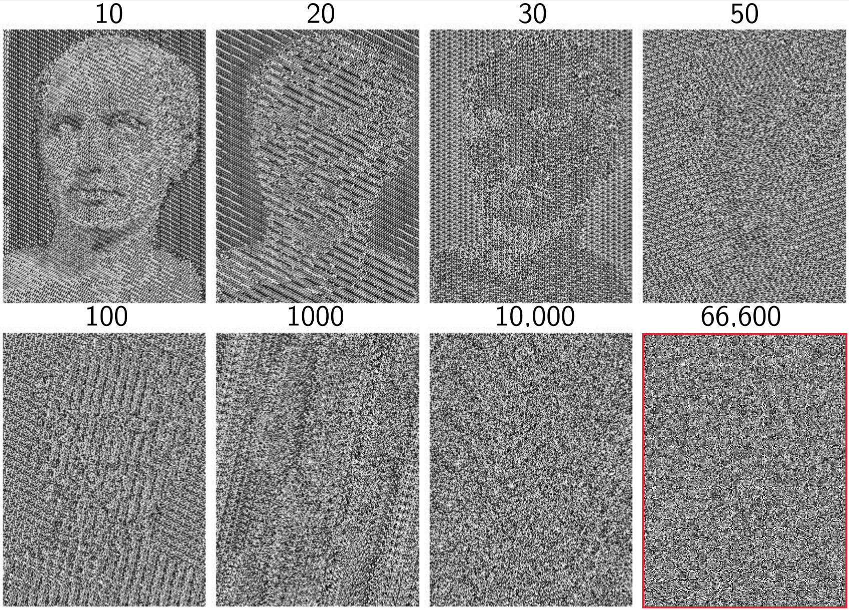
Approximating a One-Time Pad

If the key is a sequence of *random* characters that is as long as the plaintext, then the Vigenère cipher is unbreakable[1] if the key is only used *once*. In this case, we refer to it as a one-time pad.

One-time pads are impractical because of the huge key sizes typically needed. However, using key expansion via recursion for the Vigenère cipher approximates a one-time pad.

[1] The perfect security of the one-time pad was proven by Claude Shannon in 1949.

Vigenère Ciphers with Varying Key Lengths



10

20

30

50

100

1000

10,000

66,600

Vigenère Cipher with Key Expansion

On the left is an encryption of Julius Caesar's image using a Vigenère cipher with a random choice of only 8 integers (0-255) expanded using the recursion

$$a_n = a_{n-1} + 2a_{n-2} + 3a_{n-3} + 4a_{n-4} + 5a_{n-5} + 6a_{n-6} + 7a_{n-7} + 8a_{n-8} \mod 256.$$

On the right is an OTP encryption of the same image. There is no noticeable difference since the period of the recursion $> 66,600$.