The Vigenère Cipher

Math for America · Classical Cryptography Worksheet

Name: _____ **Keyword:** _____

How it works. Number the letters $\mathrm{A} = 0,\ \mathrm{B} = 1,\ \ldots,\ \mathrm{Z} = 25$. Write the keyword under the message, repeating it until every letter has a key letter beneath it. Each ciphertext letter is $C = (P + K) \bmod 26$; to decrypt, compute $P = (C - K) \bmod 26$. On the cipher wheel: turn the inner disk so its **A** sits under the key letter on the outer disk, then find the plaintext letter on the *outer* ring and read the ciphertext letter directly *inside* it.

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Part 1 — Encrypt. Write your message in the top row (one letter per box, no spaces), the repeated keyword in the middle row, and the ciphertext in the bottom row.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plaintext																				
Key																				
Ciphertext																				

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plaintext																				
Key																				
Ciphertext																				

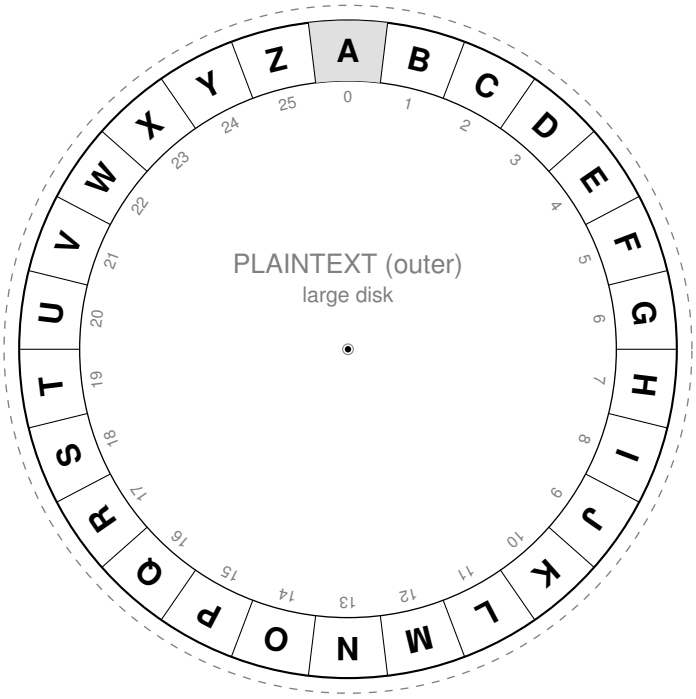
Part 2 — Decrypt. Trade ciphertext and keyword with a partner. Write the ciphertext in the bottom row, the repeated keyword in the middle, and recover the plaintext on top.

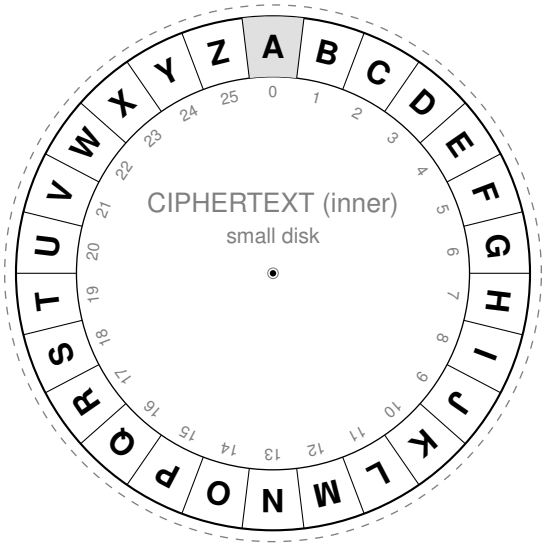
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plaintext																				
Key																				
Ciphertext																				

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Plaintext																				
Key																				
Ciphertext																				

Vigenère Cipher Wheel

Cut out both disks along the dashed lines. Place the small disk on top of the large one and push a brass fastener (or a pin) through the center dots.





Using the wheel. For each letter of the message: (1) find the current key letter on the *outer* disk and rotate the inner disk so its **A** lines up with it; (2) find the plaintext letter on the outer disk; (3) the letter directly inside it on the inner disk is the ciphertext. To decrypt, keep the same alignment, find the ciphertext letter on the *inner* disk, and read the plaintext outside it. Check with numbers: the inner disk's **A** sits at key position K, so the inner letter at plaintext P is $P + K \pmod{26}$.

Arithmetic Mod 26

Math for America · Classical Cryptography Worksheet

Name: _____

The idea. Working "mod 26" means we only care about the remainder after dividing by 26, so every answer is one of the numbers $0, 1, \ldots, 25$. Think of a clock with 26 hours: if a sum goes past 25, wrap around by subtracting 26; if a difference goes below 0, wrap around by adding 26. For example $22 + 7 = 29 \equiv 3 \pmod{26}$ and $4 - 9 = -5 \equiv 21 \pmod{26}$.

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Part 1 — Addition. Reduce each sum mod 26.

$7 + 12 \equiv$ _____ $\quad 20 + 9 \equiv$ _____ $\quad 25 + 2 \equiv$ _____

$13 + 13 \equiv$ _____ $\quad 18 + 22 \equiv$ _____ $\quad 3 + 25 \equiv$ _____

$24 + 24 \equiv$ _____ $\quad 9 + 21 \equiv$ _____ $\quad 16 + 17 \equiv$ _____

Part 2 — Subtraction. Reduce each difference mod 26 (answers must be between 0 and 25).

$5 - 3 \equiv$ _____ $\quad 2 - 9 \equiv$ _____ $\quad 0 - 1 \equiv$ _____

$14 - 20 \equiv$ _____ $\quad 7 - 7 \equiv$ _____ $\quad 11 - 25 \equiv$ _____

$3 - 5 \equiv$ _____ $\quad 22 - 25 \equiv$ _____ $\quad 1 - 23 \equiv$ _____

Part 3 — Letters. Convert each letter to its number using the strip above, add or subtract mod 26, and convert back to a letter. (This is exactly one step of the Vigenère cipher: plaintext + key = ciphertext.)

$\mathrm{H} + \mathrm{E} =$ _____ $\quad \mathrm{T} + \mathrm{K} =$ _____ $\quad \mathrm{W} + \mathrm{S} =$ _____

$\mathrm{Y} + \mathrm{C} =$ _____ $\quad \mathrm{Q} + \mathrm{L} =$ _____ $\quad \mathrm{Z} + \mathrm{Z} =$ _____

$\mathrm{K} - \mathrm{C} =$ _____ $\quad \mathrm{B} - \mathrm{H} =$ _____ $\quad \mathrm{A} - \mathrm{Z} =$ _____

$\mathrm{E} - \mathrm{R} =$ _____ $\quad \mathrm{X} - \mathrm{X} =$ _____ $\quad \mathrm{D} - \mathrm{W} =$ _____

Part 4 — Think about it.

1. Adding A (that is, 0) to a letter leaves it unchanged. Which key letter shifts a letter exactly halfway around the alphabet, so that applying it twice brings you back where you started?

2. Subtracting K mod 26 gives the same answer as adding some other number K'. Find K' in terms of K, and use it to explain why decrypting a Vigenère message is just encrypting with a different keyword. What is the "decryption keyword" for the keyword CAB?

3. Explain why, if you know the plaintext letter and the ciphertext letter, you can recover the key letter.

How Many Keys?

Math for America · Classical Cryptography Worksheet

Name: _____

Use the *Powers of 2 and 10* ladder on the next page. Rough answers are fine: we care about the *exponent*, not the last digit. Assume a fast computer can test $10^9 \approx 2^{30}$ keys per second.

Part 1 — Simple substitution cipher. A key is any rearrangement of the 26 letters.

1. Count the keys. There are 26 choices for what A becomes, then ___ for B, then ___ for C, … so the number of keys is $26 \times 25 \times \cdots \times 1 = 26!$. Written out, $26! = 403{,}291{,}461{,}126{,}605{,}635{,}584{,}000{,}000$. That is about $4 \times 10^{\text{___}}$, which on the ladder is roughly $2^{\text{___}}$.

2. At 10^9 keys per second, how many seconds does it take to try every key? About $10^{\text{___}}$ seconds. Using 1 year $\approx 3 \times 10^7$ seconds, that is about $10^{\text{___}}$ years. Compare this with the age of the universe (1.4×10^{10} years).

3. Yet a substitution cipher can be broken by hand in a few minutes (see the *Frequency Analysis* page). What does this tell you about the relationship between the *number* of keys and the *security* of a cipher?

Part 2 — Vigenère cipher. A key is a word of m letters, so there are 26^m keys.

4. Fill in the table. (Hint: each letter of the key is worth $\log_2 26 \approx 4.7$ bits, so $26^m \approx 2^{4.7m}$.)

Key length m	Number of keys 26^m	$\approx 10^{?}$	$\approx 2^{?}$	Time to try all keys at 10^9/sec
3				
5				
10				

5. Modern ciphers such as AES use 2^{128} keys. How long would a Vigenère keyword have to be to have that many keys? $m \approx$ _____ letters

6. Suppose the keyword is chosen from an English dictionary of about $100{,}000$ words rather than at random. Now how many keys are there really, regardless of the keyword's length? About $2^{\text{___}}$. How long does the 10^9/sec computer take to try them all? _____

Powers of 2 and 10: The Ladder

Reference sheet

The one rule to remember: $2^{10} = 1{,}024 \approx 10^3$. So every 10 powers of 2 is about 3 powers of 10: $2^n \approx 10^{0.3\,n}$ and $10^k \approx 2^{3.3\,k}$. (Exactly: $\log_{10} 2 = 0.301$ and $\log_2 10 = 3.32$.)

Power of 2	≈ Power of 10	Decimal	Something about that big
2^{0}	10^{0}	1	
2^{10}	10^{3}	1,024 ≈ one thousand	words in a short essay
2^{17}	10^{5}	131,072 ≈ one hundred thousand	words in an English dictionary
2^{20}	10^{6}	1,048,576 ≈ one million	seconds in 12 days
2^{25}	3×10^{7}	33,554,432	**seconds in one year**
2^{30}	10^{9}	≈ one billion	operations per second, one fast computer
2^{33}	10^{10}	≈ 8.6 billion	people on Earth
2^{40}	10^{12}	≈ one trillion	
2^{47}	10^{14}	$\approx 1.4 \times 10^{14}$	Vigenère keys of length 10
2^{50}	10^{15}	≈ one quadrillion	
2^{56}	10^{17}	$\approx 7 \times 10^{16}$	DES keys (1977); brute-forced in 1998
2^{59}	4×10^{17}		**seconds since the Big Bang**
2^{60}	10^{18}	≈ one quintillion	
2^{63}	10^{19}	$\approx 7.5 \times 10^{18}$	grains of sand on Earth
2^{70}	10^{21}	≈ one sextillion	guesses per second, worldwide Bitcoin mining
2^{80}	10^{24}	≈ one septillion	“infeasible” for most attackers today
2^{88}	4×10^{26}	$26!$	simple substitution cipher keys
2^{100}	10^{30}		
2^{128}	3×10^{38}		AES-128 keys
2^{166}	10^{50}		atoms in the Earth
2^{256}	10^{77}		AES-256 keys
2^{266}	10^{80}		atoms in the observable universe

Powers of 10, with names

10^k	Name	Prefix	$\approx 2^?$
10^{3}	thousand	kilo	2^{10}
10^{6}	million	mega	2^{20}
10^{9}	billion	giga	2^{30}
10^{12}	trillion	tera	2^{40}
10^{15}	quadrillion	peta	2^{50}
10^{18}	quintillion	exa	2^{60}
10^{21}	sextillion	zetta	2^{70}
10^{24}	septillion	yotta	2^{80}
10^{30}	nonillion		2^{100}
10^{38}			2^{126}
10^{77}			2^{256}

Time, at one billion (2^{30}) guesses per second

Guesses	Time
2^{30}	1 second
2^{40}	17 minutes
2^{47}	1.6 days
2^{50}	13 days
2^{55}	1 year
2^{60}	36 years
2^{64}	585 years
2^{70}	37,000 years
2^{80}	38 million years
2^{88}	13 billion years (age of the universe)
2^{128}	10^{22} years

Doubling the speed of the computer, or using a billion computers (2^{30}), only moves you a few rungs down the ladder. Adding one bit to the key moves you one rung up.

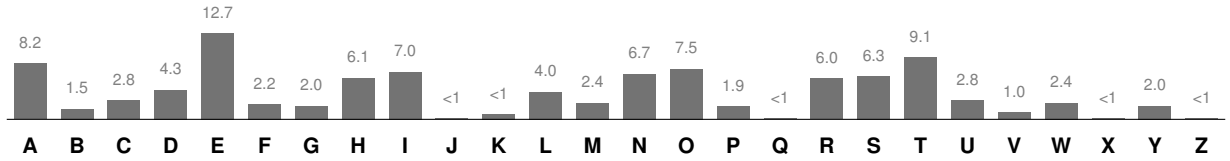
Frequency Analysis

Math for America · Classical Cryptography Worksheet

Name: _____

The idea. In ordinary English some letters are far more common than others. A Caesar shift (or any substitution) changes *which* symbol is written, but not *how often* it appears. So the most common ciphertext letter is probably E in disguise.

Letter frequencies in English text (percent)



Exercise. The message below was encrypted with a Caesar shift (every letter moved the same number of places).

```
TLLAT LHAAO LVSKI YPKNL HAZLC LUAOP ZLCLU PUNHU KIYPU NAOLS LAALY DPAOA OLZLH S
```

1. Tally each ciphertext letter in the grid (tally marks in the tall row, totals in the bottom row). There are 66 letters.

	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
tally																										
count																										

2. The most common ciphertext letter is ____. If it stands for E, the shift is ____. (Ciphertext letter $-$ E, mod 26.)

3. Check your guess: with that shift, T would become ____ and A would become ____. Are those letters also common in the ciphertext?

4. Decrypt the message (subtract the shift from each letter, or use the wheel):

   _____

   _____

5. **Extension.** Why does a Vigenère cipher with a 3-letter keyword *not* fall to this attack directly? What if you first split the ciphertext into three streams: letters 1, 4, 7, . . . ; letters 2, 5, 8, . . . ; and letters 3, 6, 9, . . . ?

The Index of Coincidence

Math for America · Classical Cryptography Worksheet

Name: _____

The idea. Pick two letters from a text at random. The *index of coincidence* (IC) is the probability that they are the same letter. If $n_{\mathrm{A}}, n_{\mathrm{B}}, \ldots, n_{\mathrm{Z}}$ are the letter counts in a text of N letters,

$$\mathrm{IC} \;=\; \frac{n_{\mathrm{A}}(n_{\mathrm{A}}-1) + n_{\mathrm{B}}(n_{\mathrm{B}}-1) + \cdots + n_{\mathrm{Z}}(n_{\mathrm{Z}}-1)}{N(N-1)}.$$

For English text, $\mathrm{IC} \approx 0.066$. For letters chosen uniformly at random, $\mathrm{IC} = 1/26 \approx 0.038$. A substitution or Caesar cipher only renames letters, so it leaves the IC unchanged. A Vigenère cipher mixes several alphabets, which flattens the counts and pushes the IC down toward 0.038. That is how you can tell them apart *without decrypting anything*.

Exercise 1. Copy your counts from the *Frequency Analysis* page into the row n, compute $n(n-1)$ for each letter (it is 0 whenever n is 0 or 1), and add them up.

	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
n																										
$n(n-1)$																										

Sum of $n(n-1) =$ _____, $N = 66$, $N(N-1) =$ _____, $\mathrm{IC} =$ _____. Is this closer to 0.066 or to 0.038?

Exercise 2. This 86-letter message was encrypted with a Vigenère cipher. The counts are done for you.

```
ELQHE PEEIE PMEPH CMQRO PRQOG SXTSG LPXSF ESMYG CIQCA ELQBB CXTGV OIATG SIDWI PVZSN CXTSB
WHEHB YIYWY W
```

	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
n	2	4	5	1	9	1	5	4	6	0	0	3	3	1	3	6	5	2	7	4	0	2	4	4	4	1
$n(n-1)$																										

Sum of $n(n-1) =$ _____, $N(N-1) = 86 \times 85 = 7310$, $\mathrm{IC} =$ _____. Which cipher does the IC point to, and why?

Exercise 3 (challenge). A keyword of length m mixes m alphabets, so roughly $1/m$ of letter pairs come from the same alphabet (IC ≈ 0.066) and the rest from different ones (IC ≈ 0.038). Setting $\mathrm{IC} \approx \frac{1}{m}(0.066) + \left(1 - \frac{1}{m}\right)(0.038)$ and solving gives the estimate

$$m \;\approx\; \frac{0.028}{\mathrm{IC} - 0.038}.$$

Use your IC from Exercise 2 to estimate the keyword length: $m \approx$ _____. (The true keyword has 5 letters; on a text this short the estimate is rough.) How would you finish breaking the cipher?

Answer Keys

Arithmetic Mod 26

Part 1 (addition): $19,\ 3,\ 1;\quad 0,\ 14,\ 2;\quad 22,\ 4,\ 7.$

Part 2 (subtraction): $2,\ 19,\ 25;\quad 20,\ 0,\ 12;\quad 24,\ 23,\ 4.$

Part 3 (letters): L, D, O; A, B, Y; I, U, B; N, A, H.

Part 4:

1. N ($= 13$): $13 + 13 = 26 \equiv 0$, so shifting by N twice is the identity.
2. $K' = 26 - K$ (with $K' = 0$ when $K = 0$), since $-K \equiv 26 - K \pmod{26}$. Decrypting with keyword $k_1 k_2 \ldots k_m$ is encrypting with keyword $(26 - k_1)(26 - k_2) \ldots (26 - k_m)$. For $\mathrm{CAB} = (2, 0, 1)$ the decryption keyword is $(24, 0, 25) = \mathrm{YAZ}$.
3. $K = C - P \pmod{26}$. This is why a Vigenère cipher falls to a known-plaintext attack: a few matched letters reveal the keyword, and its length, immediately.

How Many Keys?

1. 25 for B, 24 for C. $26! \approx 4 \times 10^{26} \approx 2^{88}$.
2. $4 \times 10^{26} / 10^9 = 4 \times 10^{17}$ seconds $\approx 1.3 \times 10^{10}$ years: about the age of the universe.
3. A huge key space is *necessary* for security (otherwise brute force works) but not *sufficient*: the attacker need not try keys one at a time if the cipher leaks structure.
4. $26^3 = 17{,}576 \approx 10^{4.2} \approx 2^{14}$, instant. $26^5 \approx 1.2 \times 10^7 \approx 2^{23.5}$, about 0.01 seconds. $26^{10} \approx 1.4 \times 10^{14} \approx 2^{47}$, about 1.4×10^5 seconds ≈ 1.6 days.
5. $128 / 4.7 \approx 27.2$, so 28 letters.
6. $10^5 \approx 2^{17}$ keys; about 0.0001 seconds.

Frequency Analysis

Counts: A 10, C 2, D 1, H 4, I 2, K 3, L 14, N 3, O 5, P 5, S 3, T 2, U 5, V 1, Y 3, Z 3. L is the most common (14 of 66), so the shift is $\mathrm{L} - \mathrm{E} = 11 - 4 = 7$. Under shift 7, T $\to$ A (10 occurrences) and A $\to$ H (4), both common. Plaintext: MEET ME AT THE OLD BRIDGE AT SEVEN THIS EVENING AND BRING THE LETTER WITH THE SEAL. Extension: with three alphabets the counts blur together; each of the three streams is a plain Caesar cipher and can be solved separately.

Index of Coincidence

Exercise 1. $\sum n(n-1) = 380$, $N(N-1) = 4290$, $\mathrm{IC} = 0.089$. Clearly on the "English" side (short texts run high, but nowhere near 0.038): a single-alphabet cipher.
Exercise 2. $\sum n(n-1) = 330$, $\mathrm{IC} = 330/7310 = 0.045$, much closer to 0.038: a polyalphabetic cipher.
Exercise 3. $m \approx 0.028/(0.045 - 0.038) \approx 3.9$ (true length 5; keyword LEMON). To finish: split the ciphertext into m streams and run frequency analysis on each; the plaintext begins THE TREASURE IS BURIED BENEATH THE TALLEST OAK TREE.