Cryptography

Jason Baldridge
UT Austin
Language and Computers

Many slides used from Chris Brew's *Codes and Code Breaking* course at OSU, and much material taken from Simon Singh's The Code Book: http://www.simonsingh.net/The_Code_Book.html

Decode these

```
gur urqtrubt va gur pntr pheyrq vagb n onyy
```

Decode these

```
gur urqtrubt va gur pntr pheyrq vagb n onyy
```

```
the hedgehog in the cage curled into a ball
```

```
Unix command to encode:
> echo "the hedgehog in the cage curled into a ball" | tr 'a-z' 'n-za-m'
```

Decode these

gur urqtrubt va gur pntr pheyrq vagb n onyy

the hedgehog in the cage curled into a ball

```
Unix command to encode:
> echo "the hedgehog in the cage curled into a ball" | tr 'a-z' 'n-za-m'
```

o homen alto esta dirigindo um carro grande na minha estrada

Decode these

gur urqtrubt va gur pntr pheyrq vagb n onyy

the hedgehog in the cage curled into a ball

```
Unix command to encode:
> echo "the hedgehog in the cage curled into a ball" | tr 'a-z' 'n-za-m'
```

o homen alto esta dirigindo um carro grande na minha estrada

the tall man is driving a big car on my street

Decode these

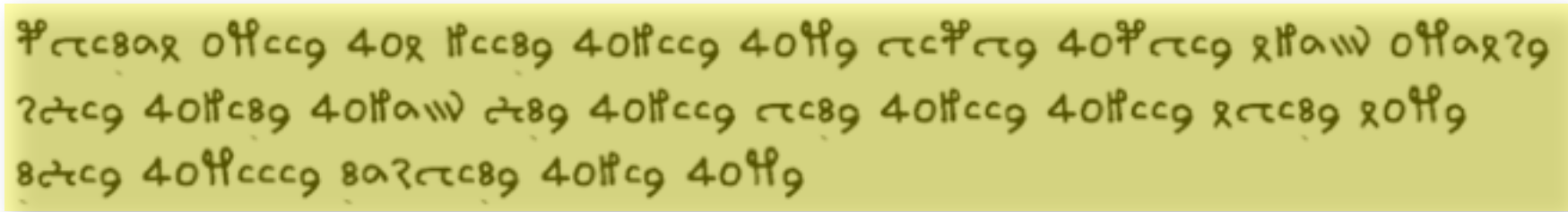
gur urqtrubt va gur pntr pheyrq vagb n onyy

the hedgehog in the cage curled into a ball

```
Unix command to encode:
> echo "the hedgehog in the cage curled into a ball" | tr 'a-z' 'n-za-m'
```

o homen alto esta dirigindo um carro grande na minha estrada

the tall man is driving a big car on my street



Decode these

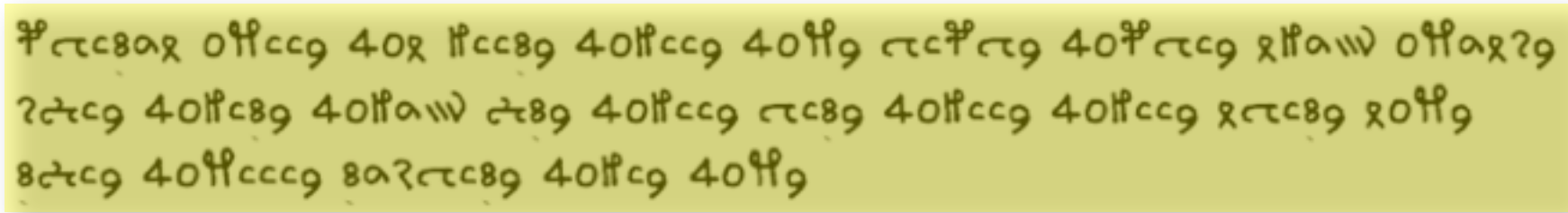
gur urqtrubt va gur pntr pheyrq vagb n onyy

the hedgehog in the cage curled into a ball

```
Unix command to encode:
> echo "the hedgehog in the cage curled into a ball" | tr 'a-z' 'n-za-m'
```

o homen alto esta dirigindo um carro grande na minha estrada

the tall man is driving a big car on my street



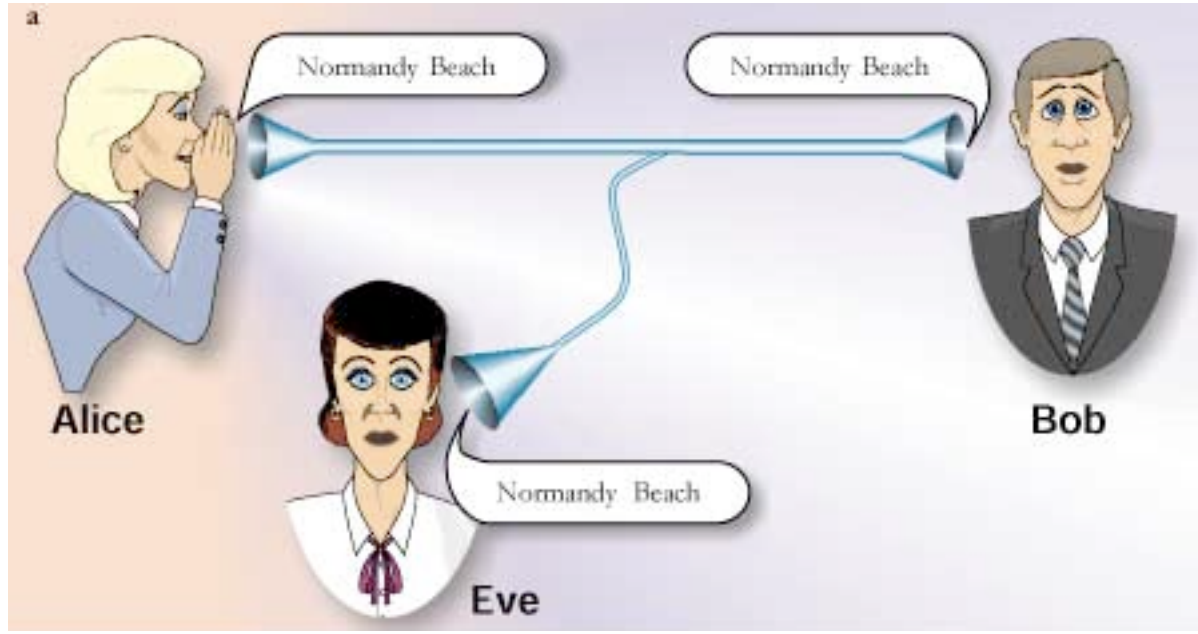
The Voynich manuscript script - nobody knows!

Getting the message across

- If you want to get a message to someone, what can you do to prevent eavesdropping?
- This problem, the solutions to it, and the ways of breaking through the solutions have shaped history.
- They have also helped us crack forgotten writing systems such as Egyptian hieroglyphics and Linear B.
- The sophistication of codes and code-breaking has evolved greatly over the last several thousand years.
- We'll start simple and get a glimpse of how things work today.

Alice, Bob, and Eve

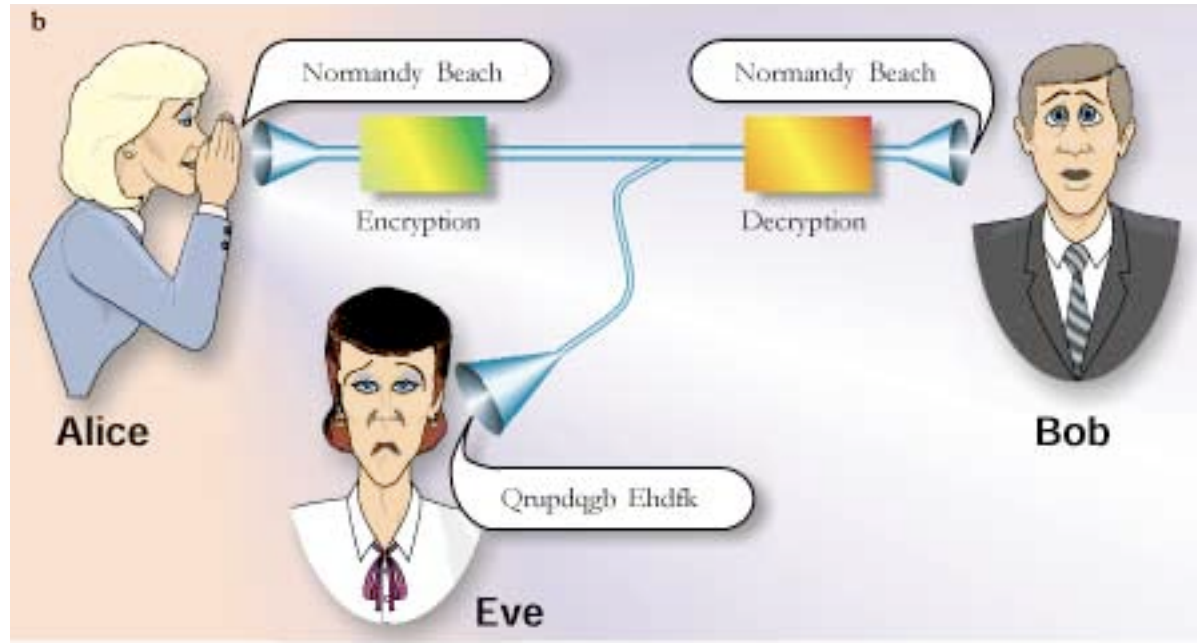
- Alice wants to send a message to Bob, and Eve is trying to eavesdrop.
- If Alice doesn't do anything, Eve will hear what Bob hears.
- However, if she encrypts the message and Bob knows how to decrypt it, Eve is out of luck.



From: www.physicstoday.org/pt/vol-53/iss-11/p22.htm

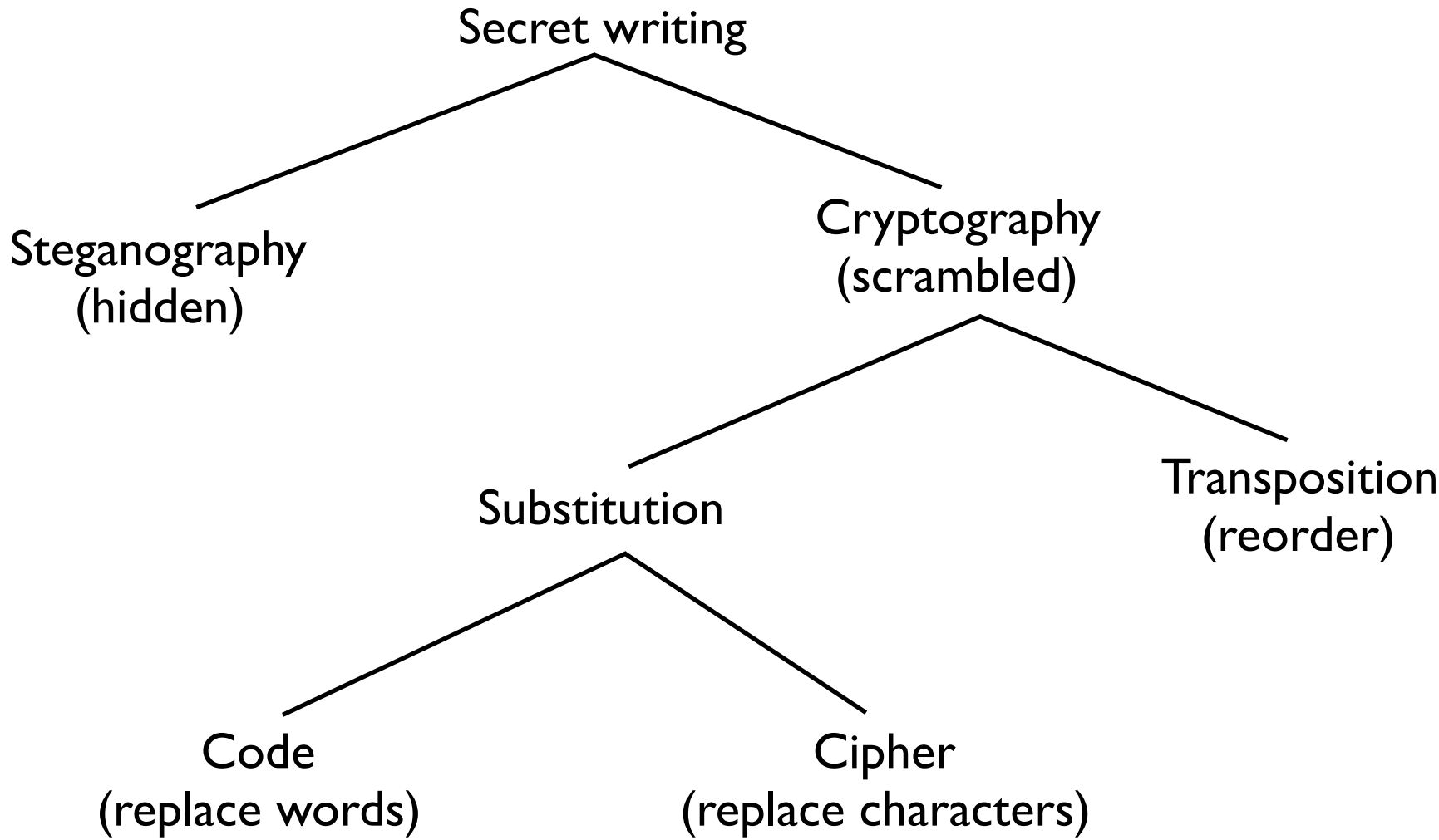
Alice, Bob, and Eve

- Alice wants to send a message to Bob, and Eve is trying to eavesdrop.
- If Alice doesn't do anything, Eve will hear what Bob hears.
- However, if she encrypts the message and Bob knows how to decrypt it, Eve is out of luck.



From: www.physicstoday.org/pt/vol-53/iss-11/p22.htm

The branches of secret writing



Steganography: hidden writing

- Eavesdropping was avoiding early on by simply hiding the message.
  - put the message in a false heel
  - Histaiaeus (494 BC): shave messenger's head, write the message, let the hair grow and then the messenger could travel unhindered
  - invisible ink
- **Steganography** is derived from *steganos* "covered" and *graphein* "to write"
- Provides some security, but if the message is detected, the contents are immediately known to the interceptors.
- Modern steganography is very advanced, with messages being embedded in text, images, and video.

Image steganography examples

This avatar contains the message "Boss said that we should blow up the bridge at midnight." encrypted with mozaiq using "växjö" as password.
http://en.wikipedia.org/wiki/Steganography

According to the FBI, this image contains a map of the Burlington, Vermont airport.
http://www.wired.com/dangerroom/2010/06/alleged-spies-hid-secret-messages-on-public-websites/

Cryptography: scrambling the message

- Encrypted messages can be seen by others, but their contents are hidden because the text itself has been transformed by some algorithm. The recipient must know how to reverse that algorithm.
- Ways of encrypting messages:
  - **transposition**: reordering the letters
  - **substitution**: replace words or letters with other words, letters, or symbols

Transposition

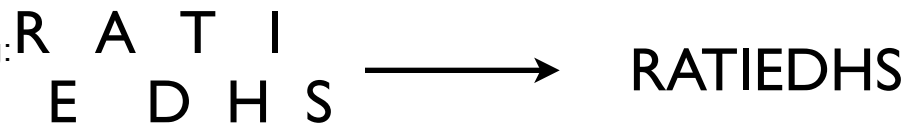
- A simple way to scramble a message is **transposition**: reorder the symbols.
  - Example: READ THIS
    - random:
    - alternating:
    - insertion (more effective when spoken, as with Ubbi Dubbi):
- **Scytales** were a way of doing alternating transposition easily. The message is encoded on a strip of leather on a cylinder, and then the decoder uses a cylinder of the same diameter to reveal the message.

Transposition

- A simple way to scramble a message is **transposition**: reorder the symbols.
  - Example: READ THIS
    - random: EDRA HSTI
    - alternating:
    - insertion (more effective when spoken, as with Ubbi Dubbi):
- **Scytales** were a way of doing alternating transposition easily. The message is encoded on a strip of leather on a cylinder, and then the decoder uses a cylinder of the same diameter to reveal the message.

Transposition

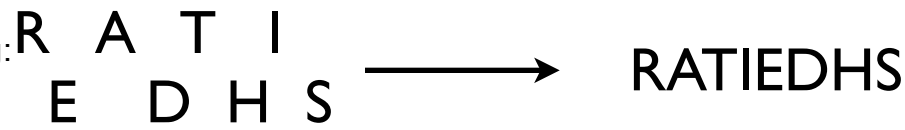
- A simple way to scramble a message is **transposition**: reorder the symbols.
  - Example: READ THIS
    - random: EDRA HSTI
    - alternating:



    - insertion (more effective when spoken, as with Ubbi Dubbi):
- **Scytales** were a way of doing alternating transposition easily. The message is encoded on a strip of leather on a cylinder, and then the decoder uses a cylinder of the same diameter to reveal the message.

Transposition

- A simple way to scramble a message is **transposition**: reorder the symbols.
  - Example: READ THIS
    - random: EDRA HSTI
    - alternating:



    - insertion (more effective when spoken, as with Ubbi Dubbi): RUBEAD THUBIS
- **Scytales** were a way of doing alternating transposition easily. The message is encoded on a strip of leather on a cylinder, and then the decoder uses a cylinder of the same diameter to reveal the message.

Transposition

- A simple way to scramble a message is **transposition**: reorder the symbols.
  - Example: READ THIS
    - random: EDRA HSTI
    - alternating:
      R A T I
      E D H S
      → RATIEDHS
    - insertion (more effective when spoken, as with Ubbi Dubbi): RUBEAD THUBIS

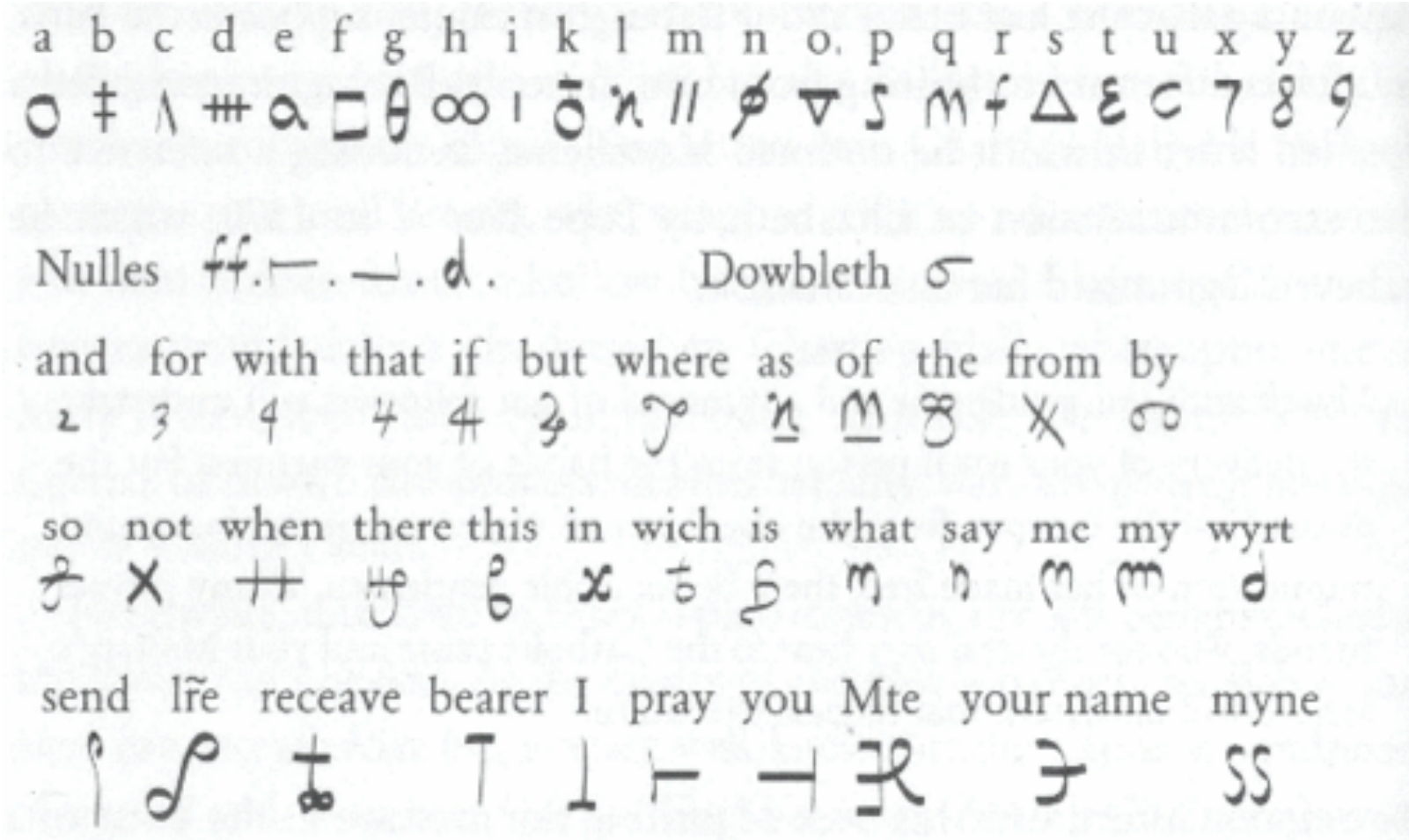
- **Scytales** were a way of doing alternating transposition easily. The message is encoded on a strip of leather on a cylinder, and then the decoder uses a cylinder of the same diameter to reveal the message.

Substitution

- With transposition, all the original characters of the underlying message are still available -- with enough time the message can be decoded easily.
- Substitution involves replacing the letters or words systematically:
  - **code**: replace words
  - **cipher**: replace letters
- The cipher of Mary Queen of Scots used both a cipher and coded words, and provides a dramatic example of the importance of using a strong encryption method.

The Cipher of Mary Queen of Scots

- A simple substitution cipher with codes for frequent words



From: http://www.simonsingh.com/The_Black_Chamber/maryqueen.html

The Babington Plot

- Mary was imprisoned by Queen Elizabeth in 1567. After 18 years, she was contacted by Anthony Babington, who was plotting to free her and assassinate Queen Elizabeth.
- Their correspondence was encrypted using the cipher shown previously, and it was delivered by Gilbert Gifford.
- Unbeknownst to Mary and Babington, Gifford was a double agent, working for Sir Francis Walsingham, Principal Secretary to Queen Elizabeth and also her spymaster.

Weak encryption is worse than no encryption

- Walsingham was aware of recent advances in **cryptanalysis**, including **frequency analysis**. His cipher secretary, Thomas Phelippes, easily cracked the cipher and decode the messages.
- These messages were the key evidence that she was a knowing participant in the plot. With that evidence, Walsingham had Mary arrested and put on trial. The judges recommended the death penalty and she was executed on February 8, 1587.
- Moral of the story: don't use weak encryption!!!!

Weak encryption is worse than no encryption

- Walsingham was aware of recent advances in **cryptanalysis**, including **frequency analysis**. His cipher secretary, Thomas Phelippes, easily cracked the cipher and decode the messages.
- These messages were the key evidence that she was a knowing participant in the plot. With that evidence, Walsingham had Mary arrested and put on trial. The judges recommended the death penalty and she was executed on February 8, 1587.
- Moral of the story: don't use weak encryption!!!!

Back to simple substitution ciphers

- **Caeser shift ciphers**: just shift the alphabet
  - e.g., shift-3:

```
a b c d e .... w x y z
D E F G H .... Z A B C
```

- **plain text**: the original message
- **cipher text**: the encoded message

```
read this
UHDG QKLV
```

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Alice

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Alice

Bob

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

`readthis`

Alice

Bob

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Algorithm: Caesar shift
Key: Shift-3

`readthis`

Alice

Bob

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Algorithm: Caesar shift
Key: Shift-3

`readthis`

Alice

Algorithm: Caesar shift
Key: Shift-3

Bob

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Algorithm: Caesar shift
Key: Shift-3

`readthis`

Alice

Algorithm: Caesar shift
Key: Shift-3

Bob

Eve

Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption

Algorithm: Caesar shift
Key: Shift-3

`readthis` ⟶ `UHDGWKLV`

Algorithm: Caesar shift
Key: Shift-3

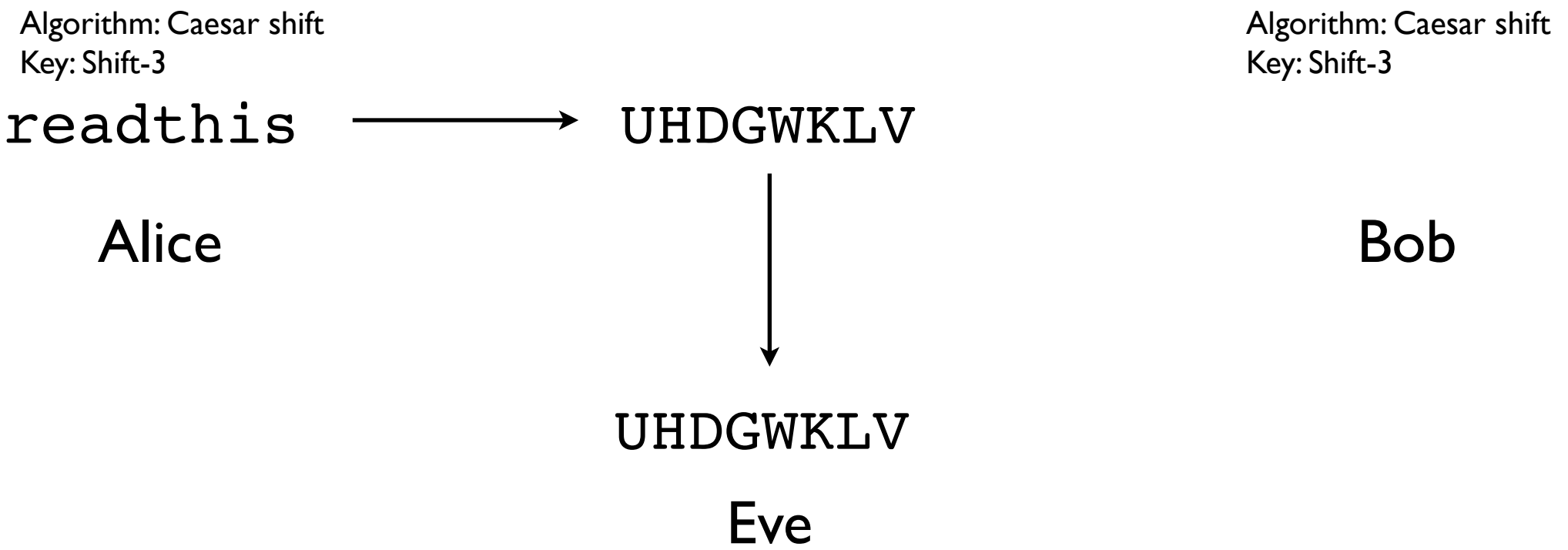
Alice

Bob

Eve

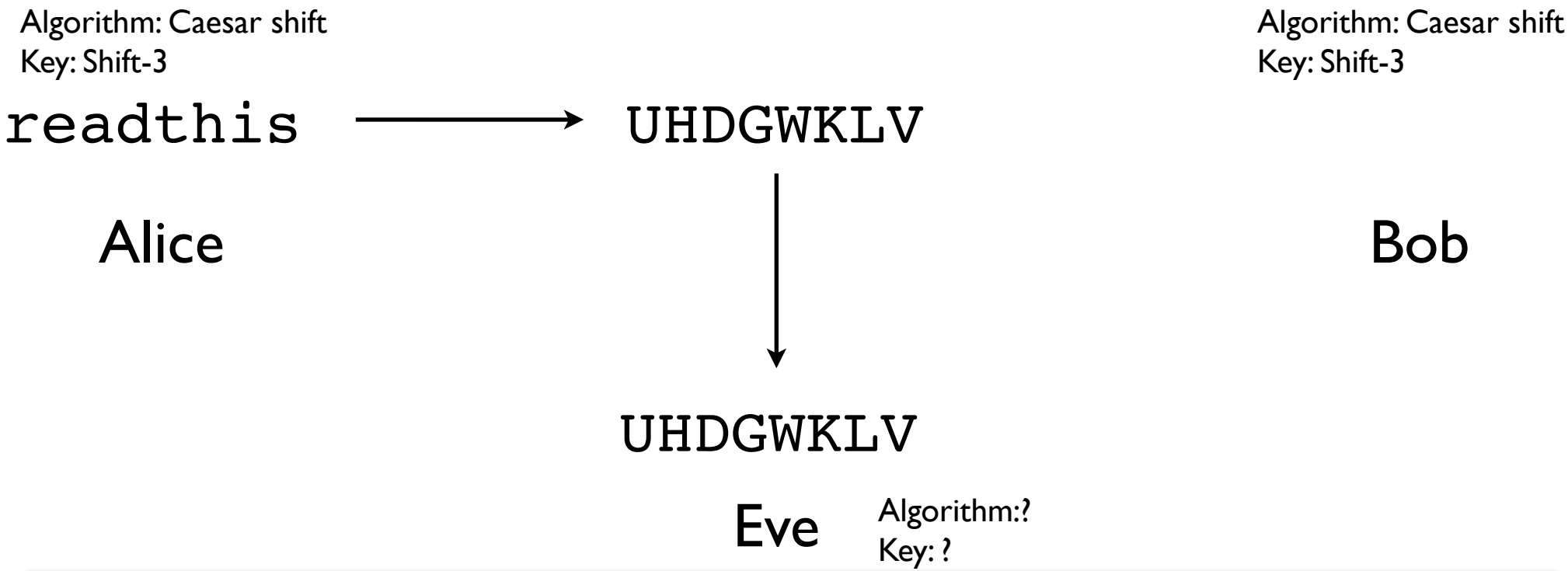
Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption



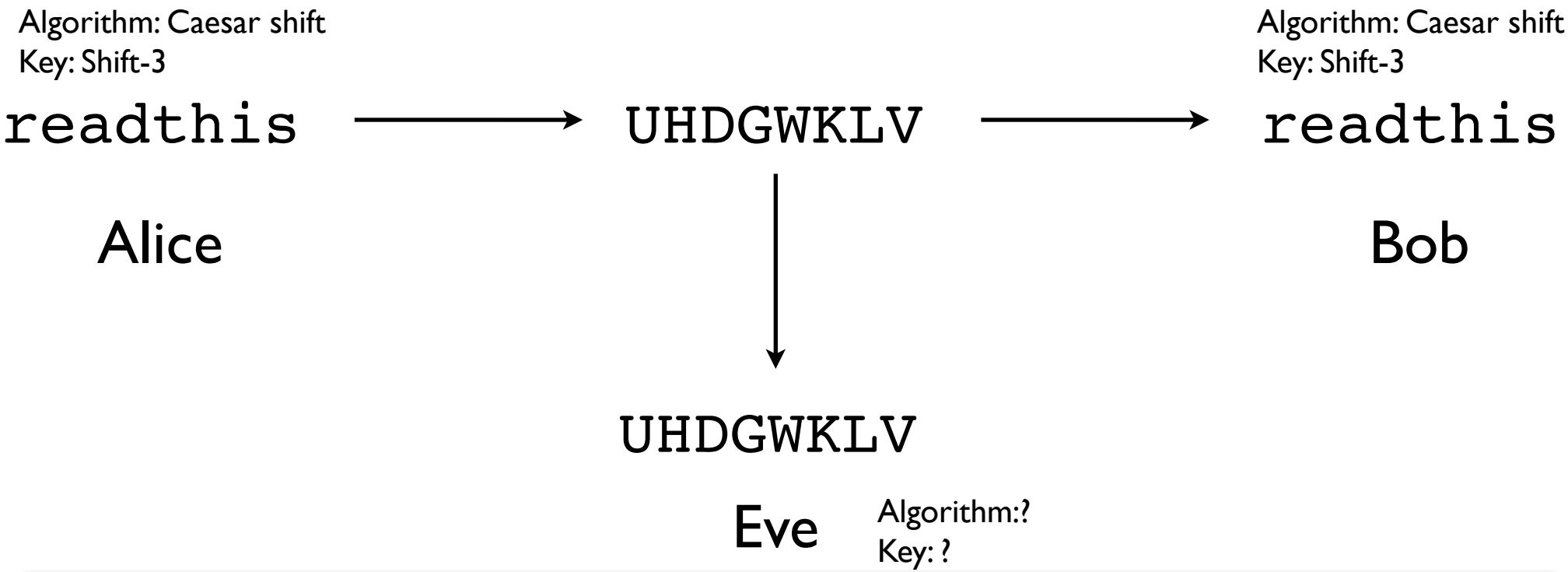
Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption



Basic components of encryption

- **algorithm**: the encryption method that precisely defines how to produce cipher text
- **key**: details for the particular encryption



Basic components of encryption

- The most important aspect of encryption is for the secret to be the key, not the algorithm.
  - “the enemy knows the system”
  - the more keys the better
- How many keys are there for Caesar shift?
- **Brute-force attack**: try all combinations (all possible keys)
- So, this is pretty easy to do for Caesar shift. (Try the message on the course syllabus.)

Basic components of encryption

- The most important aspect of encryption is for the secret to be the key, not the algorithm.
  - "the enemy knows the system"
  - the more keys the better
- How many keys are there for Caesar shift? 25
- **Brute-force attack**: try all combinations (all possible keys)
- So, this is pretty easy to do for Caesar shift. (Try the message on the course syllabus.)

General substitution

- Caesar shift maintains the order of the original alphabet, thereby limiting the number of keys and leaving messages open to brute-force attacks.
- General substitution: any letter can substitute for any letter.

```
Plain alphabet:  abcdefghijklmnopqrstuvwxyz
Cipher alphabet: JLPAWIQBCTRZYDSKEGFXHUONVM
```

- This allows 400,000,000,000,000,000,000,000,000 keys. A brute force attack checking one per per second would take roughly a billion times the lifetime of the universe to decipher a message.

```
Plain text:  et tu, brute?
Cipher text: WX XH, LGHXW?
```

Key phrases

- General substitution allows many more keys: but how can you easily remember the key in order to transmit it to the receiver?
- By using **keywords** or **key phrases**, it becomes easy to remember the key while still keeping a large number of possible keys. How to do it:
  - Choose a phrase, like JULIUS CAESAR
  - Remove spaces and duplicate letters: JULISCAER
  - Use this as the beginning of the cipher alphabet, and use the rest of the letters in order, starting where the key phrase ends.

Key phrases

- With key phrase JULIUS CAESAR:

```
Plain alphabet:  abcdefghijklmnopqrstuvwxyz
Cipher alphabet: JULISCAERTVWXYZBDFGHKMNOPQ
```

- Advantages:
  - key phrase is easily committed to memory: no need to write it down on paper that could be intercepted
  - not as many keys as general case, but still too many for brute force
- What is the major problem with this encryption method?

Many following slides from Chris Brew (OSU)

Decode this…

ZM VOWVI HRHGVI XZNV GL ERHRG SVI BLFMTVI HRHGVI RM GSV XLFMGIB. GSV VOWVI DZH NZIIRVW GL Z GIZWVHNZM RM GLDM, GSV BLFMTVI GL Z KVZHZMG RM GSV EROOZTV. ZH GSV HRHGVIH HZG LEVI GSVRI GVZ GZOPRMT, GSV VOWVI YVTZM GL YLZHG LU GSV ZWEZMGZTVH LU GLDM ORUV: HZBRMT SLD XLNULIGZYOB GSVB OREVW GSVIV, SLD DVOO GSVB WIVHHVW, DSZG URMV XOLGSVH SVI XSROWIVM DLIV, DSZG TLLW GSRMTH GSVB ZGV ZMW WIZMP, ZMW SLD HSV DVMG GL GSV GSVZGIV, KILNVMZWVH, ZMW VMGVIGZRMNVMGH.ZM VOWVI HRHGVI XZNV GL ERHRG SVI BLFMTVI HRHGVI RM GSV XLFMGIB. GSV VOWVI DZH NZIIRVW GL Z GIZWVHNZM RM GLDM, GSV BLFMTVI GL Z KVZHZMG RM GSV EROOZTV. ZH GSV HRHGVIH HZG LEVI GSVRI GVZ GZOPRMT, GSV VOWVI YVTZM GL YLZHG LU GSV ZWEZMGZTVH LU GLDM ORUV: HZBRMT SLD XLNULIGZYOB GSVB OREVW GSVIV, SLD DVOO GSVB WIVHHVW, DSZG URMV XOLGSVH SVI XSROWIVM DLIV, DSZG TLLW GSRMTH GSVB ZGV ZMW WIZMP, ZMWSLD HSV DVMG GL GSV GSVZGIV, KILNVMZWVH, ZMW VMGVIGZRMNVMGH.

Intuitively...

Intuitively...

- What clues do we have?

Intuitively...

- What clues do we have?
- How shall we work with them?

Intuitively...

- What clues do we have?
- How shall we work with them?
- What are we assuming?

A systematic approach

- Make a table of the characters used

Character set

'B', 'D', 'E', 'F', 'G', 'H',

'I', 'K', 'L', 'M', 'N', 'O',

'P', 'R', 'S', 'T', 'U', 'V',

'W', 'X', 'Y', 'Z'

Character set

- No ‘A’,’C’,’J’,’Q’

'B', 'D', 'E', 'F', 'G', 'H',
'I', 'K', 'L', 'M', 'N', 'O',
'P', 'R', 'S', 'T', 'U', 'V',
'W', 'X', 'Y', 'Z'

Character set

- No ‘A’,’C’,’J’,’Q’
- Why not?

'B', 'D', 'E', 'F', 'G', 'H',
'I', 'K', 'L', 'M', 'N', 'O',
'P', 'R', 'S', 'T', 'U', 'V',
'W', 'X', 'Y', 'Z'

What to expect.

- Make the same table for known English text
- Same number of characters from lead sport article in Sunday's Columbus Dispatch

In a city synonymous with hope against all odds, the Ohio State men's basketball team stared down another sticky situation in the Alamodome to defeat Memphis and advance to the NCAA Final Four. Madness is on the march -- to Atlanta.

"Three years ago, we had a vision for this program. It just became reality," OSU coach Thad Matta said as chants of O-H-I-O filled the arena after the Buckeyes' 92-76 win against Memphis. OSU now heads to Saturday's national semifinals.

The reality didn't come easy.

The No. 1 Buckeyes seldom take the simple route to success, as proved in the past two games when they needed late and big comebacks against Xavier and Tennessee.

Yesterday's win against the second-seeded Tigers in the South Regional final was no different, despite the 16-point margin of victory.

Ohio State (34-3) needed its four freshmen to play like seniors, and needed one of those kids, 7-foot center Greg Oden, to help wipe away a five-point deficit with 12:39 to play.

Known English

'A', 'B', 'C', 'D', 'E', 'F',

'G', 'H', 'I', 'J', 'K', 'L', 'M',

'N', 'O', 'P', 'R', 'S', 'T',

'U', 'V', 'W', 'X', 'Y'

Known English

- No 'Q','Z'

'A', 'B', 'C', 'D', 'E', 'F',
'G', 'H', 'I', 'J', 'K', 'L', 'M',
'N', 'O', 'P', 'R', 'S', 'T',
'U', 'V', 'W', 'X', 'Y'

Known English

- No ’Q’,’Z’
- Why not?

'A', 'B', 'C', 'D', 'E', 'F',
'G', 'H', 'I', 'J', 'K', 'L', 'M',
'N', 'O', 'P', 'R', 'S', 'T',
'U', 'V', 'W', 'X', 'Y'

Known English

- No 'Q','Z'
- Why not?
- Would this be same for other texts?

'A', 'B', 'C', 'D', 'E', 'F',
'G', 'H', 'I', 'J', 'K', 'L', 'M',
'N', 'O', 'P', 'R', 'S', 'T',
'U', 'V', 'W', 'X', 'Y'

A systematic approach

- Make a table of the characters used
- Keep track of frequencies
- We'll return to this in a second…

Intuitive approach

- How did you do it?

Systematizing intuition

Systematizing intuition

- Word spotting

Systematizing intuition

- Word spotting
- Start with short, common words

Word spotting…

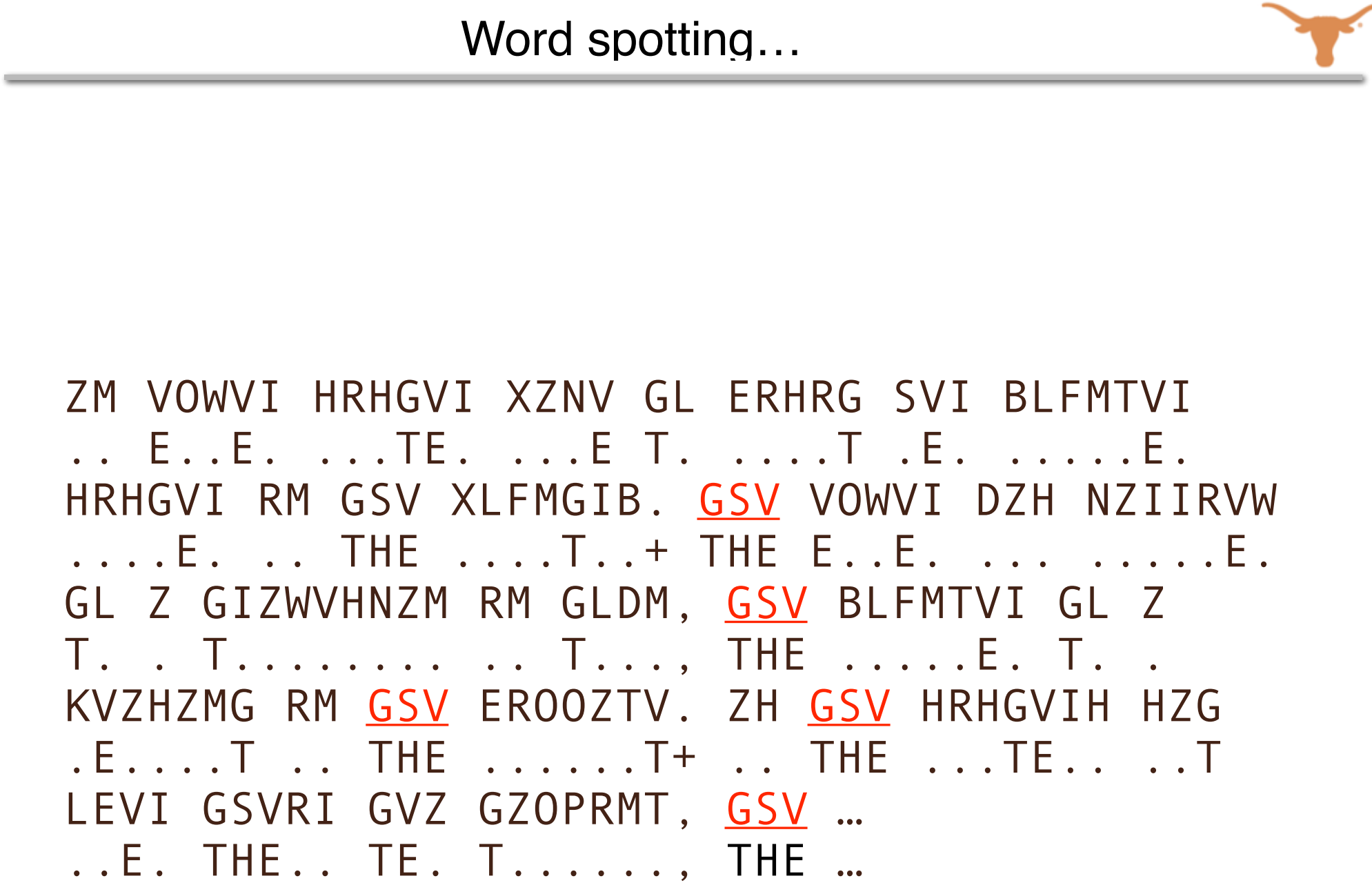
ZM VOWVI HRHGVI XZNV GL ERHRG SVI BLFMTVI

HRHGVI RM GSV XLFMGIB. GSV VOWVI DZH NZIIRVW
?THE?
GLZ GIZWVHNZM RM GLDM, GSV BLFMTVI GL Z

KVZHZMG RM GSV EROOZTV. ZH GSV HRHGVIH HZG

LEVI GSVRI GVZ GZOPRMT, GSV …

Word spotting…



Word spotting…

```
ZM VOWVI HRHGVI XZNV GL ERHRG SVI BLFMTVI
.. E..E. ...TE. ...E T. ....T .E. .....E.
HRHGVI RM GSV XLFMGIB. GSV VOWVI DZH NZIIRVW
....E. .. THE ....T..+ THE E..E. ... .....E.
GL Z GIZWVHNZM RM GLDM, GSV BLFMTVI GL Z
T. . T..... .. T..., THE .....E. T. .
KVZHZMG RM GSV EROOZTV. ZH GSV HRHGVIH HZG
.E....T .. THE ......T+ .. THE ...TE.. ..T
LEVI GSVRI GVZ GZOPRMT, GSV …
..E. THE.. TE. T....., THE …
```

Word spotting…

ZM VOWVI HRHGVI XZNV GL ERHRG SVI BLFMTVI
A. E..E. ...TE. ...E TO ....T .E. .O...E.
HRHGVI RM GSV XLFMGIB. GSV VOWVI DZH NZIIRVW
....E. .. THE .O..T..+ THE E..E. .A. .A...E.
GL Z GIZWVHNZM RM GLDM, GSV BLFMTVI GL Z
TO A T.A..... .. TO.., THE .....E. TO A
KVZHZMG RM GSV EROOZTV. ZH GSV HRHGVIH HZG
.E....T .. THE ......T+ A. THE ...TE.. ..T
LEVI GSVRI GVZ GZOPRMT, GSV …
..E. THE.. TEA TA....., THE …

How are we doing?

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

Cut to the chase?

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

Does it work…

```
ZM VOWVI HRHGVI XZNV GL ERHRG SVI BLFMTVI
AN ELDER SISTER CAME TO VISIT HER YOUNGER
HRHGVI RM GSV XLFMGIB. GSV VOWVI DZH NZIIRVW
SISTER .. THE .O..T..+ THE E..E. .A. .A...E.
GL Z GIZWVHNZM RM GLDM, GSV BLFMTVI GL Z
TO A T.A..... .. TO.., THE .....E. TO A
KVZHZMG RM GSV EROOZTV. ZH GSV HRHGVIH HZG
.E....T .. THE ......T+ A. THE ...TE.. ..T
LEVI GSVRI GVZ GZOPRMT, GSV …
..E. THE.. TEA TA....., THE …
```

Word spotting

Word spotting

- Focused on short common words

Word spotting

- Focused on short common words
- Spotted a few words

Word spotting

- Focused on short common words
- Spotted a few words
- Guessed it was a reversed alphabet.

Word spotting

- Focused on short common words
- Spotted a few words
- Guessed it was a reversed alphabet.
- Checked it.

Word spotting

- Focused on short common words
- Spotted a few words
- Guessed it was a reversed alphabet.
- Checked it.
- Why do we know this is the answer?

Why we think it is right.

- It looks like English
- The encoding we found makes sense

Character set

'B', 'D', 'E', 'F', 'G', 'H',

'I', 'K', 'L', 'M', 'N', 'O',

'P', 'R', 'S', 'T', 'U', 'V',

'W', 'X', 'Y', 'Z'

Character set

- No ‘A’,’C’,’J’,’Q’

'B', 'D', 'E', 'F', 'G', 'H',
'I', 'K', 'L', 'M', 'N', 'O',
'P', 'R', 'S', 'T', 'U', 'V',
'W', 'X', 'Y', 'Z'

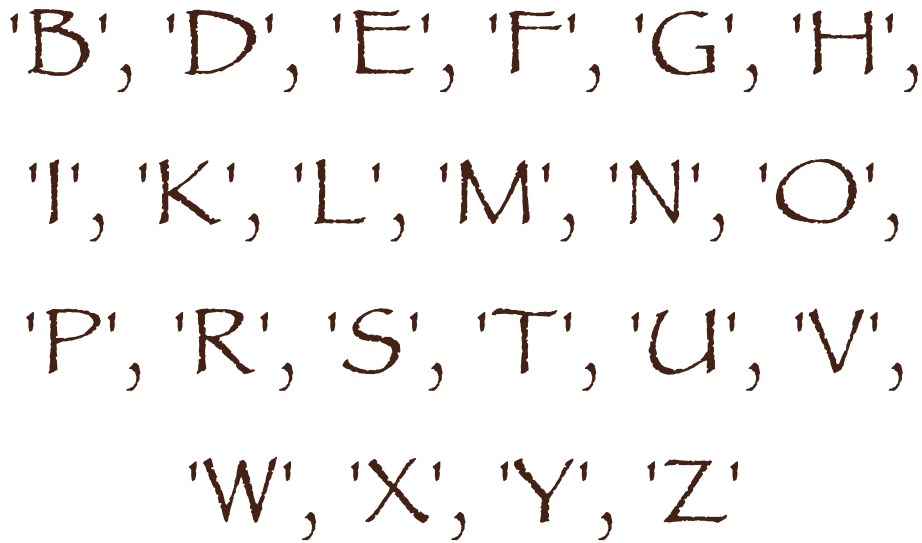
Character set

- No ‘A’,’C’,’J’,’Q’
- Why not?

'B', 'D', 'E', 'F', 'G', 'H',
'I', 'K', 'L', 'M', 'N', 'O',
'P', 'R', 'S', 'T', 'U', 'V',
'W', 'X', 'Y', 'Z'

Character set

- No ‘A’,’C’,’J’,’Q’
- Why not?
  - No ‘Z’,’X’,’Q’,’J’ in plaintext.



Character set

- No ‘A’,’C’,’J’,’Q’
- Why not?
  - No ‘Z’,’X’,’Q’,’J’ in plaintext.
  - Makes sense

'B', 'D', 'E', 'F', 'G', 'H',
'I', 'K', 'L', 'M', 'N', 'O',
'P', 'R', 'S', 'T', 'U', 'V',
'W', 'X', 'Y', 'Z'

Exercise

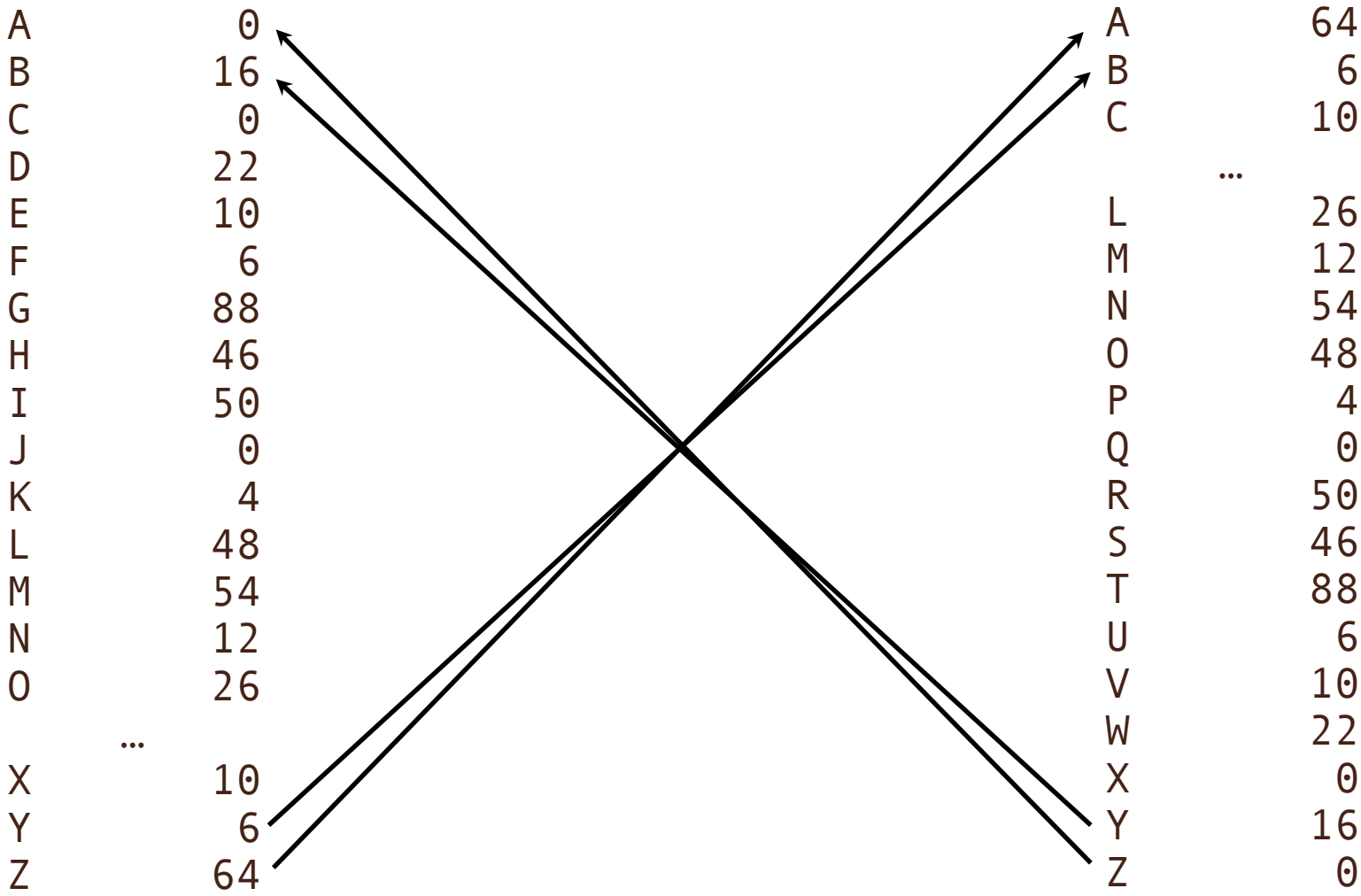
- Not the reversed alphabet, but similar.
- Use word spotting as just shown.
- See the last page of these slides for the answer.

```
"E QYSBJ ZYT KFMZGI AO QMO YH BEHI HYV OYSVU," UMEJ UFI. "QI AMO
BERI VYSGFBO, LST MT BIMUT QI MVI HVII HVYA MZPEITO. OYS BERI EZ
LITTIV UTOBI TFMZ QI JY, LST TFYSGF OYS YHTIZ IMVZ AYVI TFMZ OYS
ZIIJ, OYS MVI RIVO BECIBO TY BYUI MBB OYS FMRI. OYS CZYQ TFI XVYRIVL,
'BYUU MZJ GMEZ MVI LVYTFIVU TQMEZ.' ET YHTIZ FMXXIZU TFMT XIYXBI QFY
MVI QIMBTFO YZI JMO MVI LIGGEZG TFIEV LVIMJ TFI ZIPT. YSV QMO EU
UMHIV. TFYSGF M XIMUMZT'U BEHI EU ZYT M HMT YZI, ET EU M BYZG YZI.
QI UFMBB ZIRIV GVYQ VEKF, LST QI UFMBB MBQMOU FMRI IZYSGF TY IMT."
```

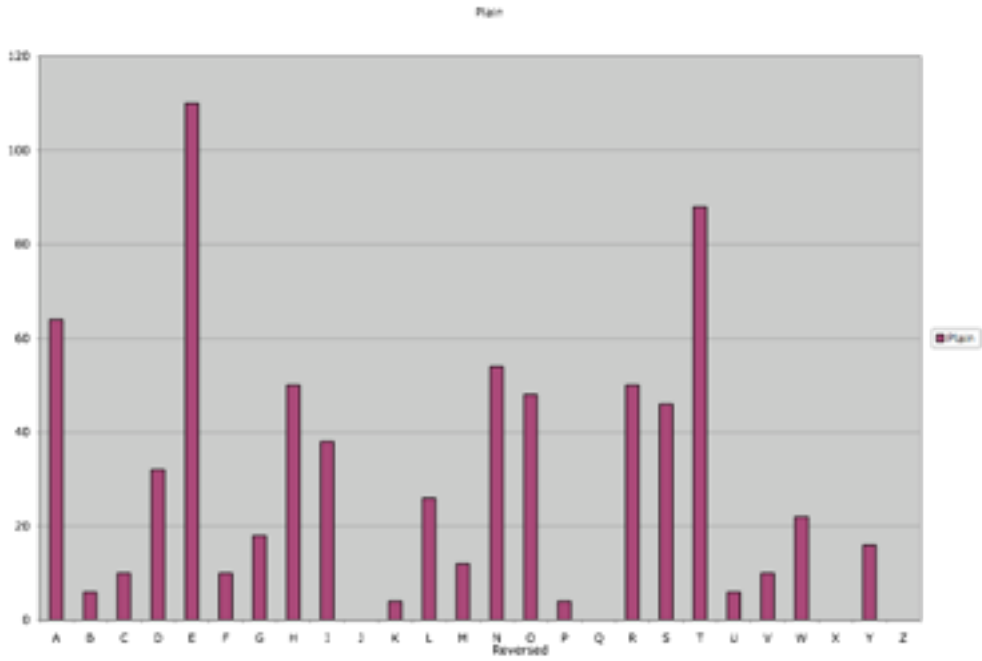
Character frequency distributions

- Not just present/absent but count
- We know which letters will probably be common
- By counting the frequency of each character in the cipher text, we can compare the relative frequency of cipher text characters to the frequency of plain text characters (using existing unencrypted text).
- A table of frequencies for all characters is a **frequency distribution**.

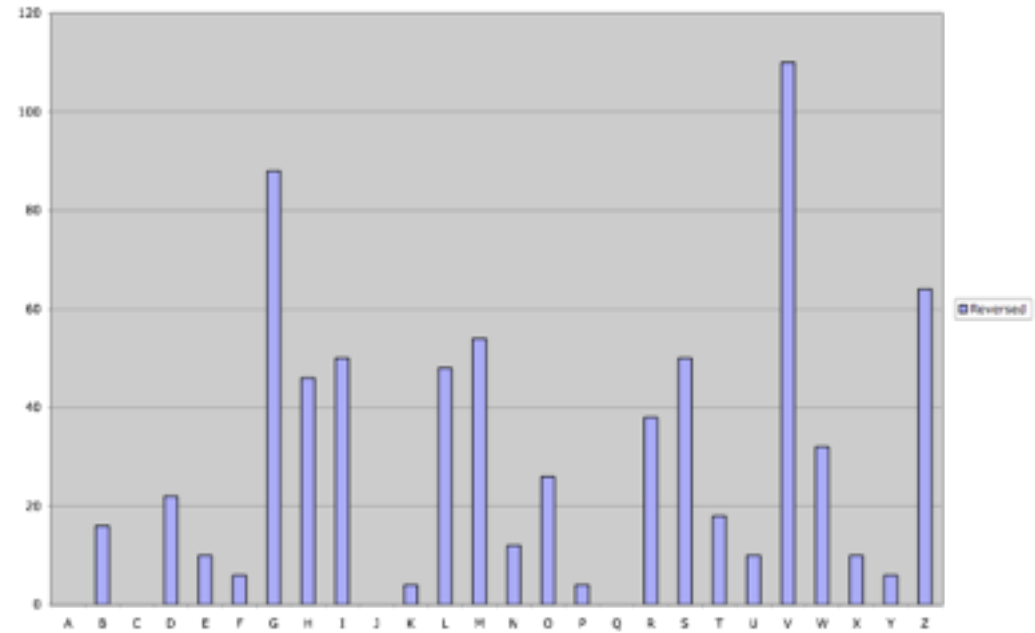
Frequencies



Histogram of frequencies



Plain

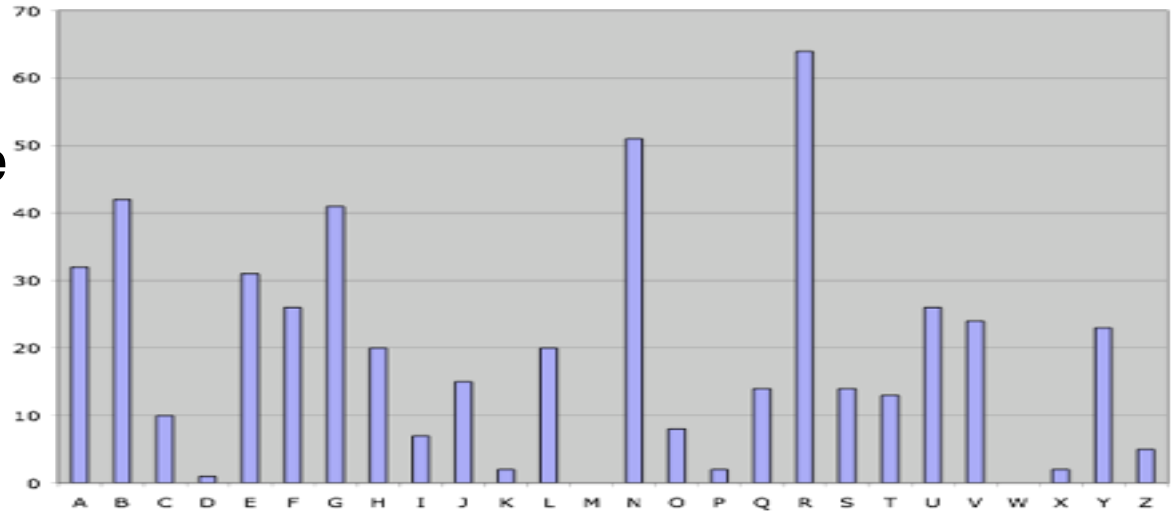


Reversed

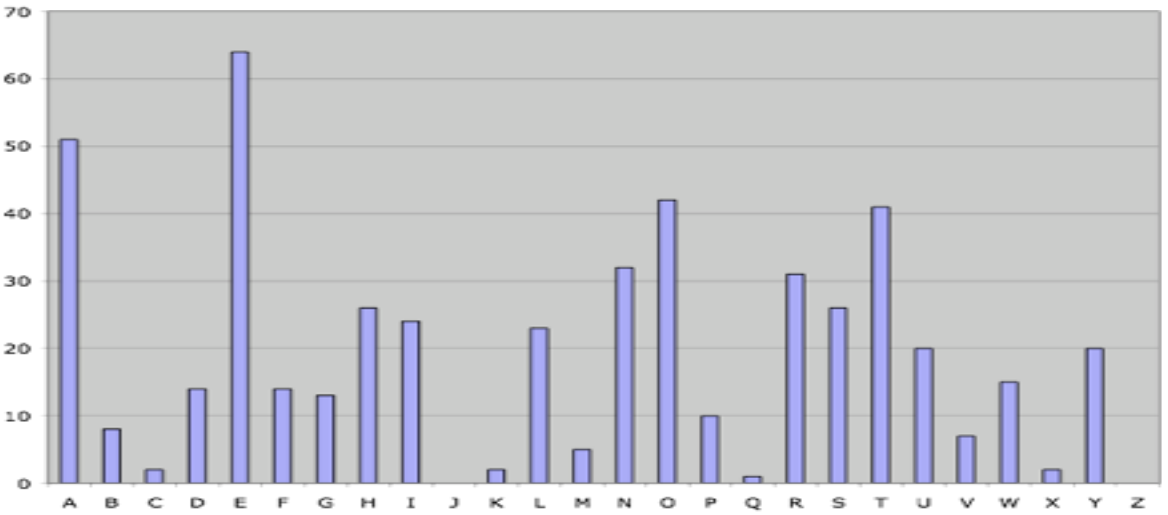
Frequencies

- The frequency pattern for the reversed alphabet exactly mirrors that of the plain text
- A Caesar shift will just show a shift in such frequency.
- What does a cipher letter “N” encode given the cipher and plain text frequency distributions on the right?

Cipher text

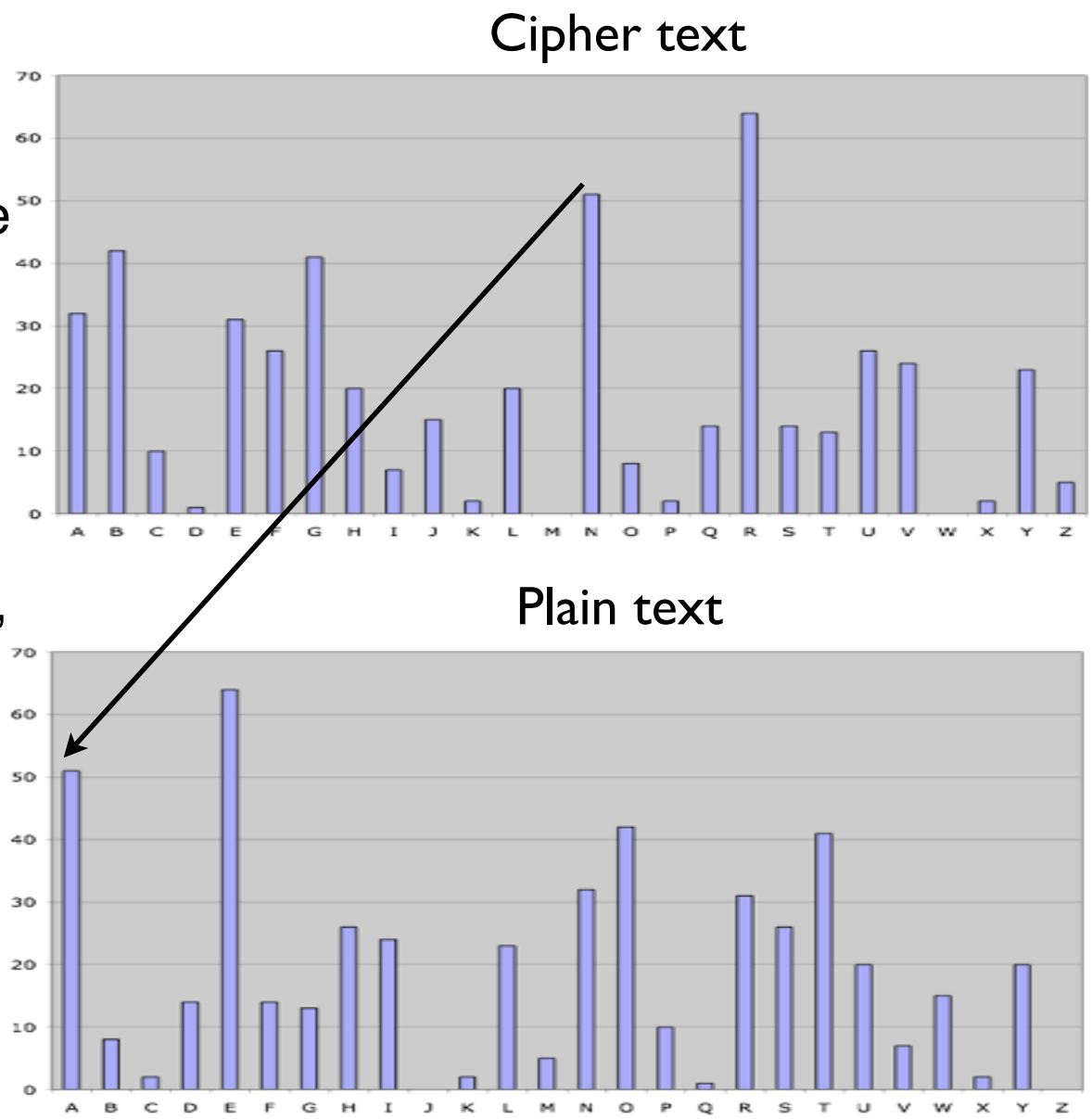


Plain text

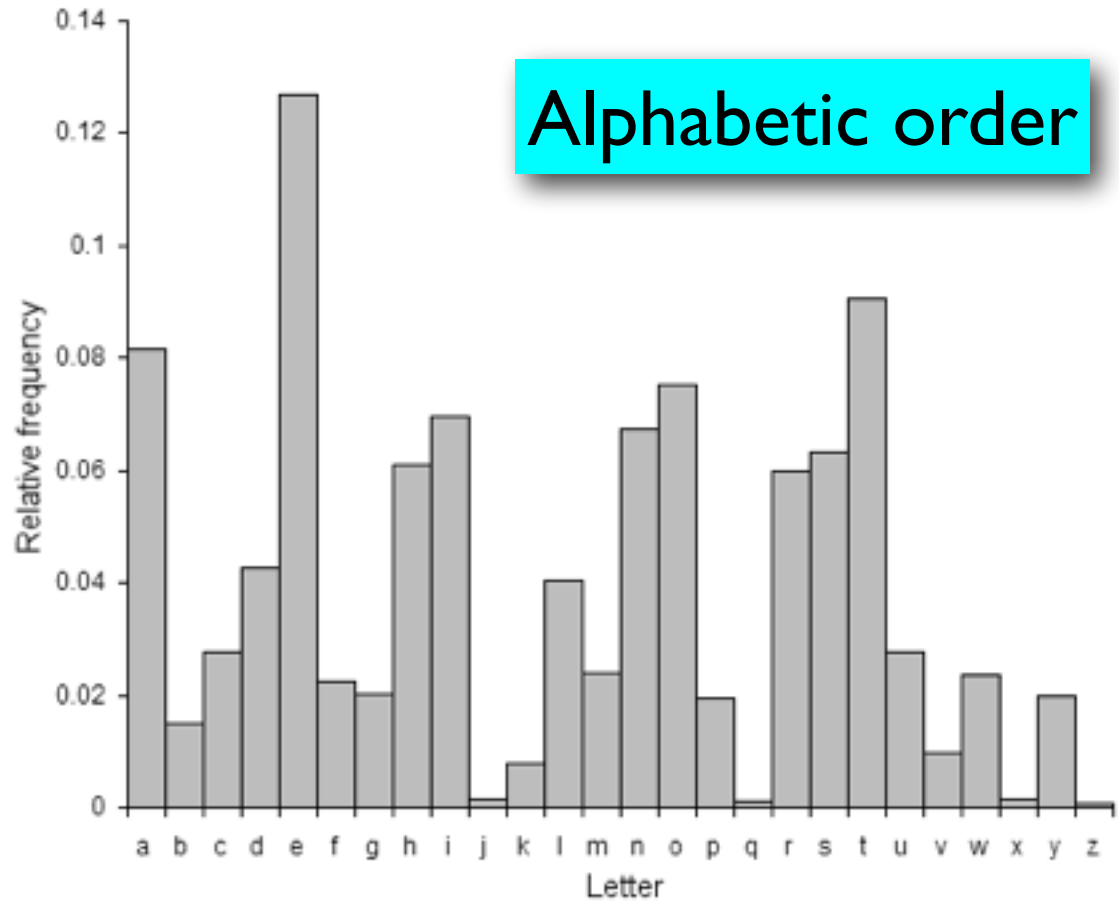


Frequencies

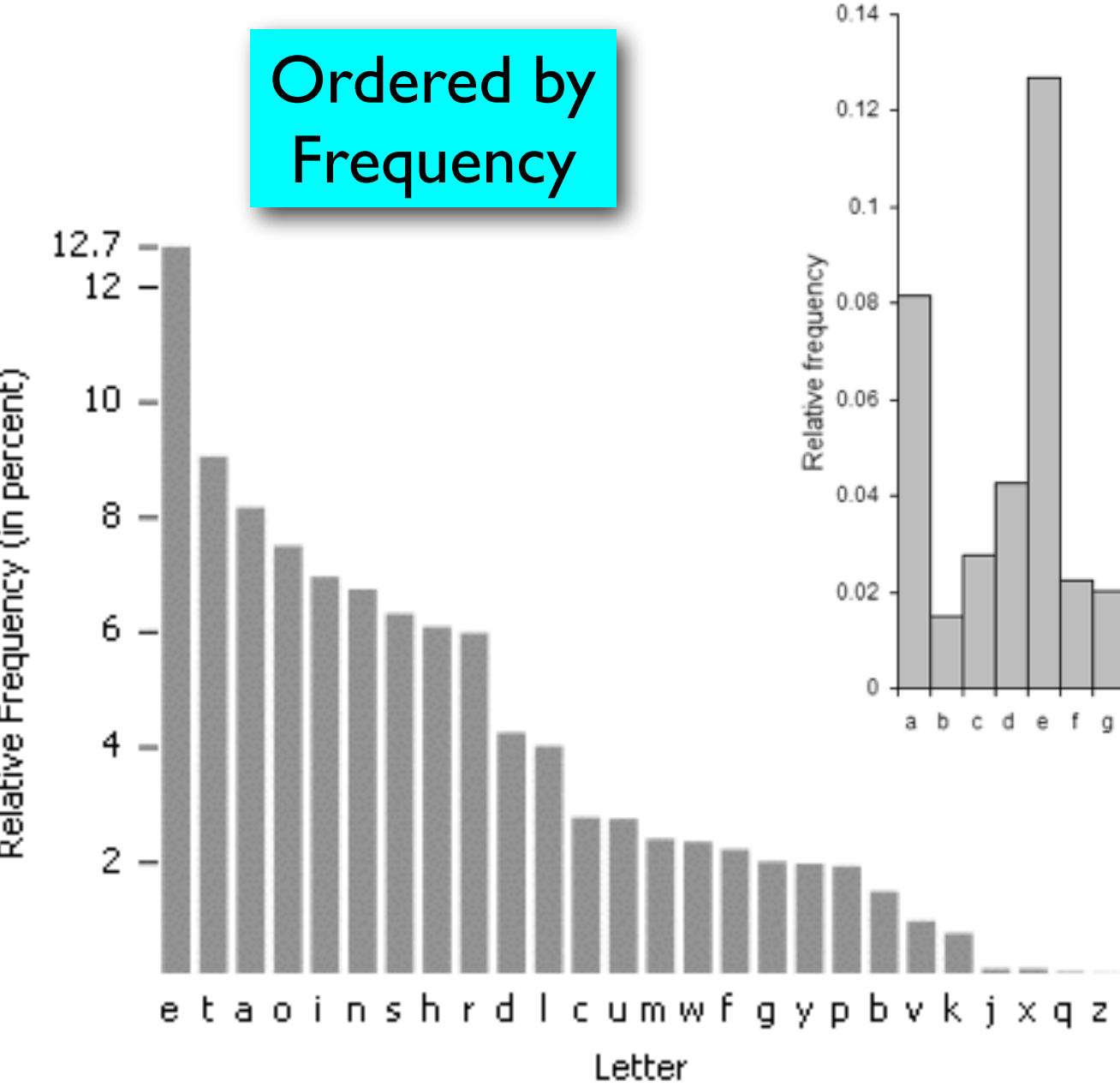
- The frequency pattern for the reversed alphabet exactly mirrors that of the plain text
- A Caesar shift will just show a shift in such frequency.
- What does a cipher letter “N” encode given the cipher and plain text frequency distributions on the right?

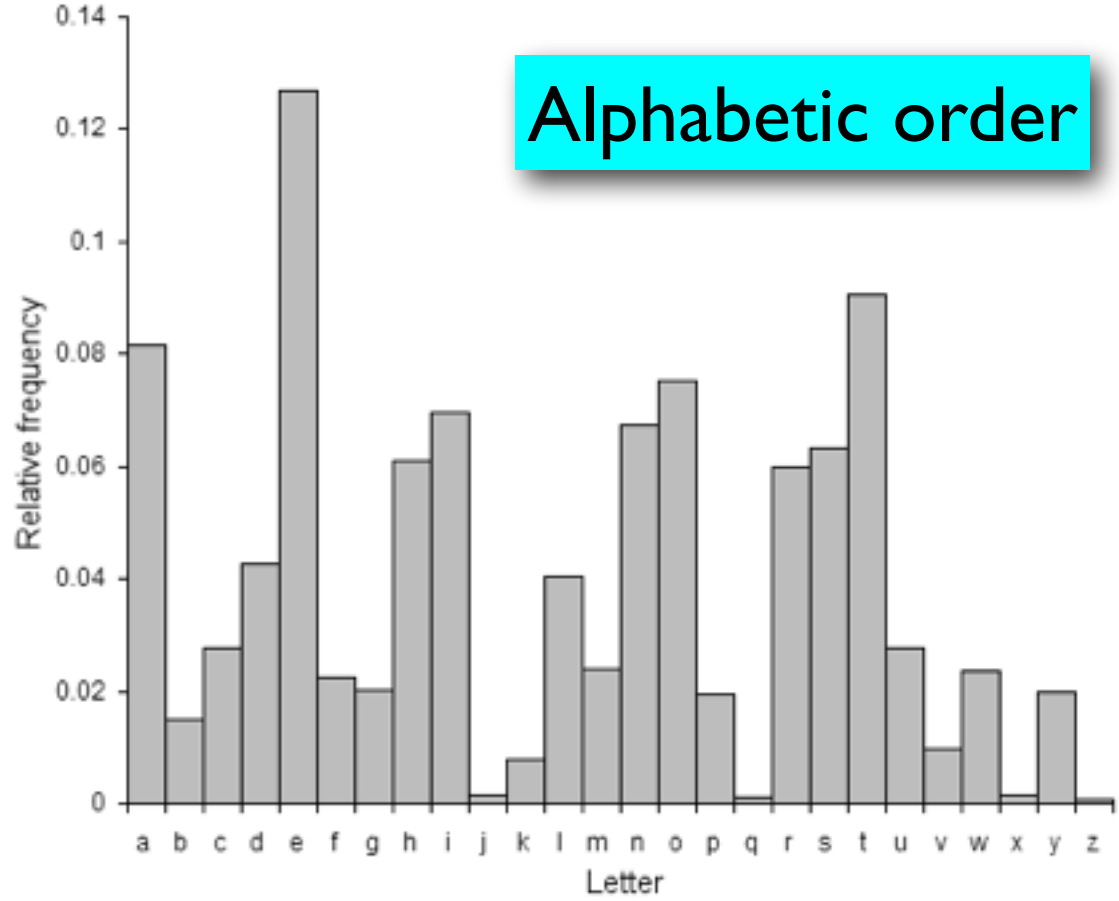


Relative frequency of English characters (from wikipedia)



Relative frequency of English characters (from wikipedia)





Big assumption

- Each time a letter appears in the plaintext it will map to the *same* letter in the ciphertext.
- Technically, this makes the ciphers we have considered so far **monoalphabetic**.
  - The problem with a monoalphabetic cipher is that it is easy to decode with word spotting and frequency analysis because each character has only one way to be encoded.
- Let's have a look at **polyalphabetic** ciphers, which provide an extra level of protection.

The Vigenere square

- The Vigenere square, published in 1586 by Blaise de Vigenere, allows all 25 Caesar shift keys to be used for the same encryption.
- The important thing is that each plain text character will be encoded in multiple ways.
- The encoding is determined by the Vignere square plus a keyphrase, such as KING or WHITE.

	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

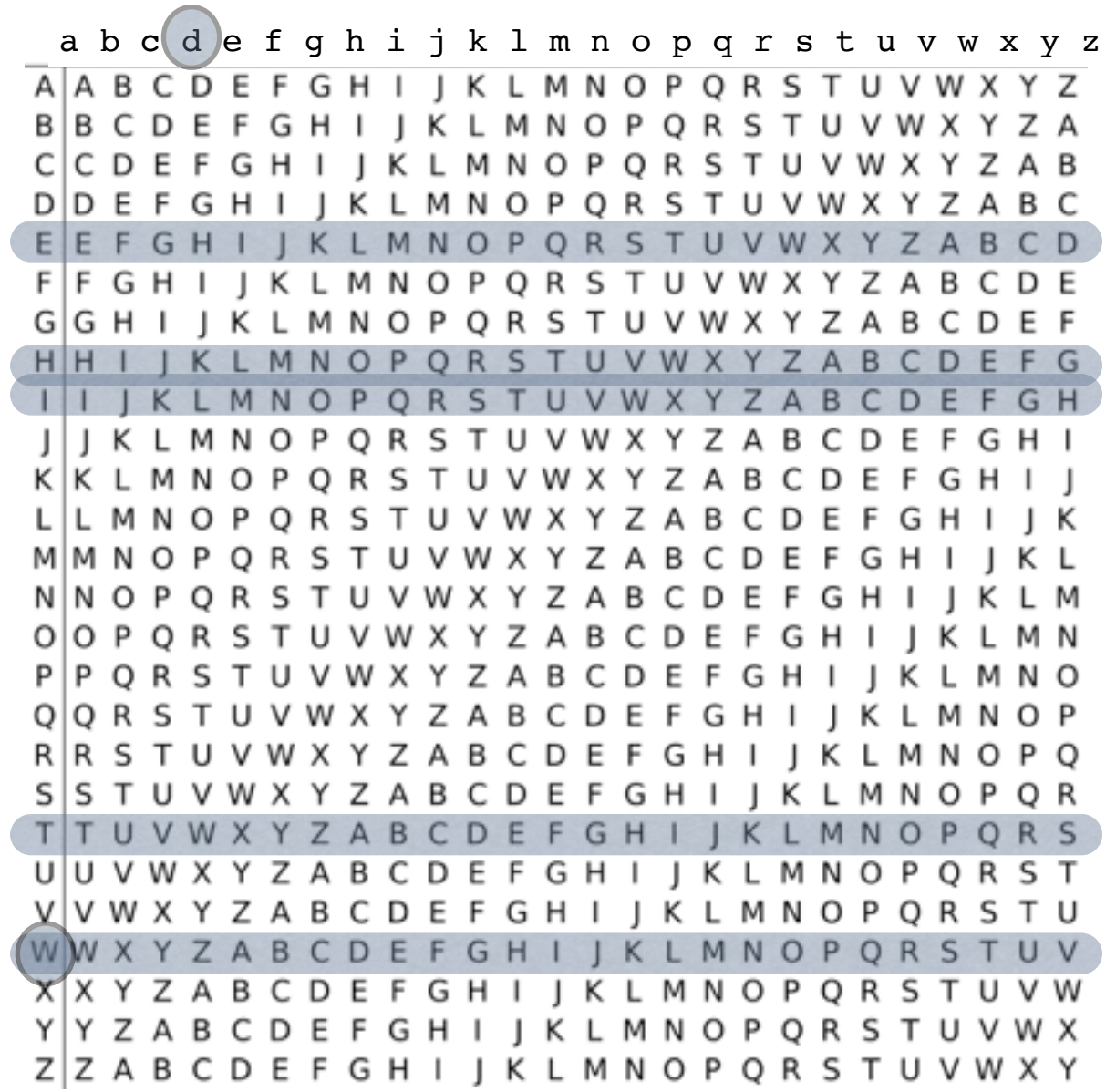
The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.

	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

The Vigenere Square

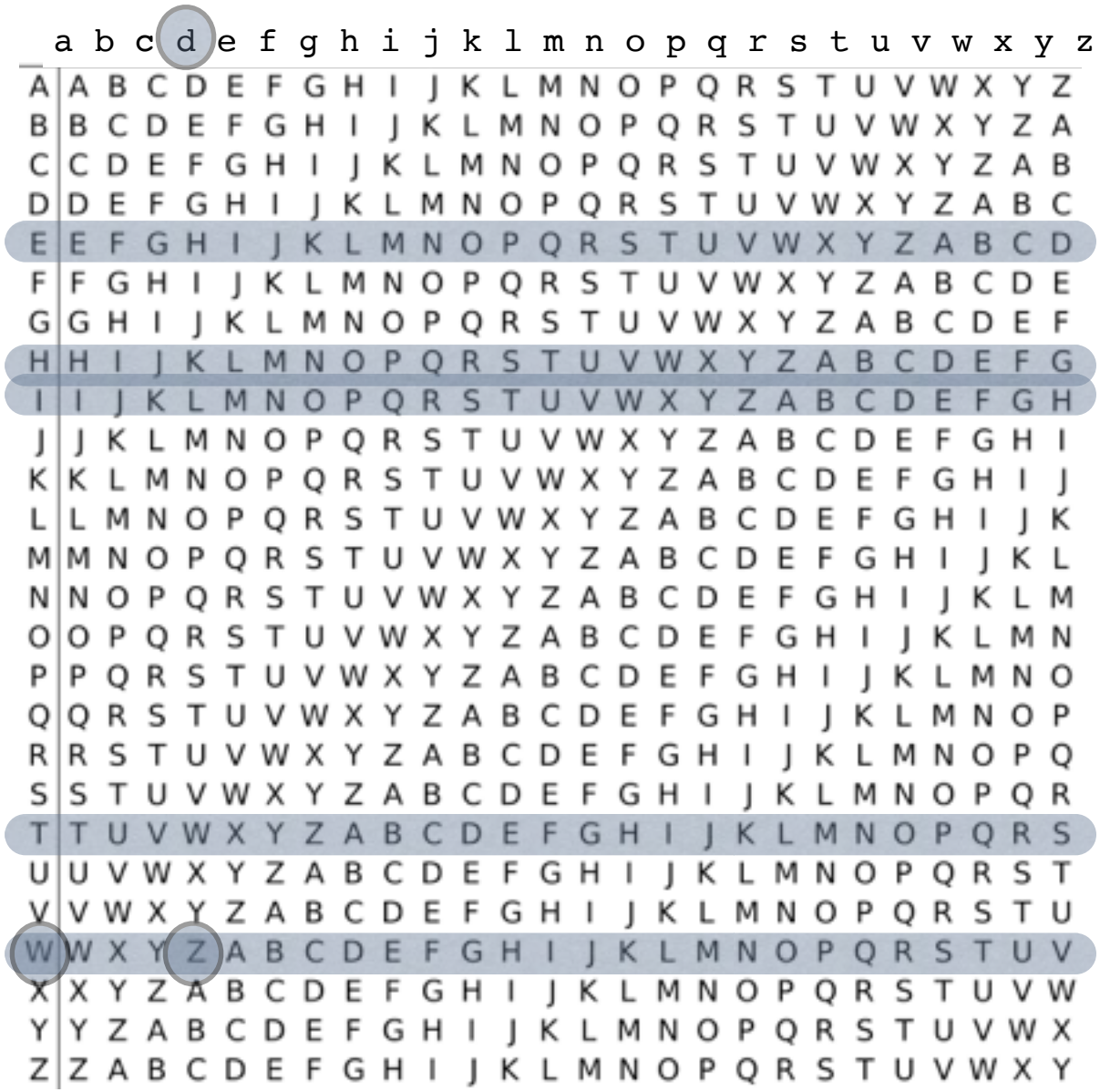
- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.



	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

The Vigenere Square

- The square defines mappings from plain text characters (column headings) to cipher text (in the square) using a key phrase letter (row headings).
- For example, if the key phrase is WHITE, the highlighted rows will be used for encryption.
- To encode a ‘d’ with a W, we look down the ‘d’ column to the W row.



	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

Vigenere example

- To encode with Vigenere, the key phrase is repeated above the plain text, and the corresponding row of the square for each key phrase character is used to encode each plain text character.

- To encode the message “divert troops to east” with the keyword WHITE:

```
Key phrase:
Plain text: diverttroopstoeast
Cipher:
```

- Note that the same letter is encoded in many different ways. For example, “t” becomes P, A and, B in the above message.

Vigenere example

- To encode with Vigenere, the key phrase is repeated above the plain text, and the corresponding row of the square for each key phrase character is used to encode each plain text character.

- To encode the message “divert troops to east” with the keyword WHITE:

```
Key phrase: WHITEWHITEWHITEWHI
Plain text: diverttroopstoeast
Cipher:
```

- Note that the same letter is encoded in many different ways. For example, “t” becomes P, A and, B in the above message.

Vigenere example

- To encode with Vigenere, the key phrase is repeated above the plain text, and the corresponding row of the square for each key phrase character is used to encode each plain text character.

- To encode the message “divert troops to east” with the keyword WHITE:

```
Key phrase: WHITEWHITEWHITEWHI
Plain text: diverttroopstoeast
Cipher:     Z
```

- Note that the same letter is encoded in many different ways. For example, “t” becomes P, A and, B in the above message.

Vigenere example

- To encode with Vigenere, the key phrase is repeated above the plain text, and the corresponding row of the square for each key phrase character is used to encode each plain text character.

- To encode the message “divert troops to east” with the keyword WHITE:

```
Key phrase: WHITEWHITEWHITEWHI
Plain text: diverttroopstoeast
Cipher:     ZP
```

- Note that the same letter is encoded in many different ways. For example, “t” becomes P, A and, B in the above message.

Vigenere example

- To encode with Vigenere, the key phrase is repeated above the plain text, and the corresponding row of the square for each key phrase character is used to encode each plain text character.

- To encode the message “divert troops to east” with the keyword WHITE:

```
Key phrase: WHITEWHITEWHITEWHI
Plain text: diverttroopstoeast
Cipher:     ZPDXVPAZHSLZBHIWZB
```

- Note that the same letter is encoded in many different ways. For example, “t” becomes P, A and, B in the above message.

Vigenere's weakness

- Because it was not susceptible to word spotting and frequency analysis, the Vigenere method became known as *Le Chiffre Indechiffrable*, "The Undecipherable Cipher". However, the use of a repeating key phrase was its weakness. Charles Babbage discovered how to crack such ciphers in the mid 1800's.
- Basic idea:
  - for a key phrase w/ N letters, each letter can only be encoded N ways.
  - look for common repeating sequences to find the length of the key phrase
  - use frequency analysis for everything Nth character
- Example:

```
Key phrase: KINGKINGKINGKINGKINGKING
Plain text: thesunandthemaninthemoon
Cipher:     DPRYEVNTNBUKWIAOXBUKWWBT
```

Vigenere's weakness

- Because it was not susceptible to word spotting and frequency analysis, the Vigenere method became known as *Le Chiffre Indechiffrable*, "The Undecipherable Cipher". However, the use of a repeating key phrase was its weakness. Charles Babbage discovered how to crack such ciphers in the mid 1800's.
- Basic idea:
  - for a key phrase w/ N letters, each letter can only be encoded N ways.
  - look for common repeating sequences to find the length of the key phrase
  - use frequency analysis for everything Nth character
- Example:

```
Key phrase: KINGKINGKINGKINGKINGKINGKING
Plain text: thesunandthemaninthemoon
Cipher:     DPRYEVNTNBUKWIAOXBUKWWBT
```

Vigenere's weakness

- Because it was not susceptible to word spotting and frequency analysis, the Vigenere method became known as *Le Chiffre Indechiffrable*, "The Undecipherable Cipher". However, the use of a repeating key phrase was its weakness. Charles Babbage discovered how to crack such ciphers in the mid 1800's.
- Basic idea:
  - for a key phrase w/ N letters, each letter can only be encoded N ways.
  - look for common repeating sequences to find the length of the key phrase
  - use frequency analysis for everything Nth character
- Example:

```
Key phrase:  KINGKINGKINGKINGKINGKINGKING
Plain text:  thesunandthemaninthemoon
Cipher:      DPRYEVNTNBUKWIAOXBUKWWBT
```

8 chrs = 2 x length("KING")

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: ????????????????????
Plain text: ????????????????????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: ????????????????????
Plain text: the???the?????the????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN???BSJ?????YPT????
Plain text: the???the?????the????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN???BSJ?????YPT????
Plain text: the???the?????the????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

CAN, CANteen,
CANada, CANny

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN???BSJ?????YPT????
Plain text: the???the?????the????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

CAN, CANteen, CANada, CANny

??? ... Doesn’t look like English...

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN???BSJ?????YPT????
Plain text: the???the?????the????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

CAN, CANteen, CANada, CANny

??? ... Doesn’t look like English...

apocalYPTic, crYPT, egYPT

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN???BSJ?????YPT????
Plain text: the???the?????the????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN?????APOCALYPTIC??
Plain text: the?????nqcbeothexg??
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CAN?????????EGYPT????
Plain text: the?????????atthe????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CANADA?????EGYPT????
Plain text: themee?????atthe????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CANADA?????EGYPT????
Plain text: themeeting??atthe????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CANADABRAZ??EGYPT????
Plain text: themeeting??atthe????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CANADABRAZILEGYPT????
Plain text: themeetingisatthe????
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

What about a non-repeating key phrase?

- One could use a poem or a book, or the names of all the presidents as a key phrase. This would be much more impervious to this style of decipherment.
- But, we can play a variant of the word spotting game even in this case! Assume that some common word, like “the” is in various parts of the plain text, and see if an interesting key phrase word would have produced

```
Key phrase: CANADABRAZILEGYPTCUBA
Plain text: themeetingisatthedock
Cipher:     VHRMHEUZNFQDEZRWXFIDK
```

Mechanization of polyalphabetic ciphers

Mechanization of polyalphabetic ciphers

Confederate Cipher Disk

Enigma Machine

Substitution in the digital age

- To a computer, letters are just binary numbers (e.g., ASCII)
- Encryption then becomes a question of manipulating numbers.
  - "HELLO" = `1001000 1000101 1001100 1001100 1001111 (Decimal: 18,391,344,324)`
  - "DAVID" = `1000100 1000001 1010110 1001001 1000100 (Decimal: 19,473,311,311)`
- Operation: bitwise XOR (0 XOR 0 = 0, 0 XOR 1=1, 1 XOR 0=1, 1 XOR 1=0)

Substitution in the digital age

- To a computer, letters are just binary numbers (e.g., ASCII)
- Encryption then becomes a question of manipulating numbers.
  - "HELLO" = `1001000 1000101 1001100 1001100 1001111 (Decimal: 18,391,344,324)`
  - "DAVID" = `1000100 1000001 1010110 1001001 1000100 (Decimal: 19,473,311,311)`
- Operation: bitwise XOR (0 XOR 0 = 0, 0 XOR 1=1, 1 XOR 0=1, 1 XOR 1=0)

```
Key phrase:  10001001000001101011010010011000100
Plain text:  10010001000101100110010011001001111
Cipher text: 00011000000100001101000001010001011
```

Substitution in the digital age

- To a computer, letters are just binary numbers (e.g., ASCII)
- Encryption then becomes a question of manipulating numbers.
  - "HELLO" = `1001000 1000101 1001100 1001100 1001111 (Decimal: 18,391,344,324)`
  - "DAVID" = `1000100 1000001 1010110 1001001 1000100 (Decimal: 19,473,311,311)`
- Operation: bitwise XOR (0 XOR 0 = 0, 0 XOR 1=1, 1 XOR 0=1, 1 XOR 1=0)

```
Key phrase:   10001001000001101011010010011000100
Plain text:   10010001000101100110010011001001111
Cipher text:  00011000000100001101000001010001011
```

"DAVID"

"HELLO"

3,230,040,715
(No simple character string)

Exploits

- Encrypted messages have actual content underlying them, so educated guesses about the keys and the content could often be exploited:
  - frequency
  - repetition
    - many words are more common and will be repeated
    - many messages will start with the same pattern, e.g., a date or location
  - meaning: both keys and message have semantic patterns

The Navajo code talkers

- During WWII, the American military used Navajos as radio operators who could speak in a code (i.e., the Navajo language) to transmit messages.
- A message in English would be given to a Navajo radio operator, who would speak a Navajo translation into the radio. Another Navajo radio operator would hear it on the other side, and translate it back into English easily.
- Code talkers had been used in WWI, so Hitler had sent anthropologists to study native American languages before the outbreak of WWII, but could not cover all the languages and dialects that existed: the Navajo was one of the tribes that had not been studied.

The Navajo code talkers

- Code talkers were amazingly effective for several reasons.
  - the Japanese and German militaries had no expertise in Navajo. It belongs to the Na-Dene family of languages, which has no link to Asian or European languages
  - in trials, American cryptanalysts couldn't even transcribe it, much less crack it, calling Navajo "a weird succession of guttural, nasal, tongue-twisting sounds"
  - encoding and decoding was extremely fast, so Navajo soldiers were extremely useful in battle groups that couldn't wait for decipherment with more complex techniques for hiding English messages.

The Navajo code talkers

- Code talkers were amazingly effective for several reasons.
  - the Japanese and German militaries had no expertise in Navajo. It belongs to the Na-Dene family of languages, which has no link to Asian or European languages
  - in trials, American cryptanalysts couldn't even transcribe it, much less crack it, calling Navajo "a weird succession of guttural, nasal, tongue-twisting sounds"
  - encoding and decoding was extremely fast, so Navajo soldiers were extremely useful in battle groups that couldn't wait for decipherment with more complex techniques for hiding English messages.

Deciphering lost languages and ancient scripts

- Many writing systems have been developed over the ages, and some were forgotten.
  - Ancient Egyptian hieroglyphs 
  - Linear B 
- And mysterious manuscripts have come to light, such as the Voynich manuscript.
  - unknown script, unknown language
  - fake or real?

Egyptian hieroglyphs

- It was originally thought that the hieroglyphic writing system was completely logographic: each character represents a concept.
- In 1652, the Jesuit scholar Athanasius Kircher published a dictionary of hieroglyphs based on the logographic assumption. This assumption persisted for another century and a half.
- in 1799, the Rosetta stone was discovered: it contained a single text in three different writing systems: Greek, demotic, and hieroglyphic. This is known as a **parallel text**, which is important in current machine translation techniques.
- The fact that the Greek portion could be read easily was the key: it provided the "plain text" for discovering the hieroglyphic system (the "cipher text")

The Rosetta stone (196 BC)

Hieroglyphic

Demotic

Greek

Deciphering hieroglyphics

- In 1814, Thomas Young focused on the **cartouche**: a set of hieroglyphs surround by a loop. The Rosetta stone had the cartouche of Pharaoh Ptolemy, who was mentioned in the Greek text several times.

p l
t o m i i s **Ptolmiis**

- Young determined a number of sound correspondences correctly for hieroglyphs found in cartouches. Unfortunately, he didn't follow this through because of the Kircher's argument that hieroglyphs were logographic.

- Jean-Francois Champollion took the next step in 1822, and applied Young's approach to other cartouches.

Champollion's next step

- Deciphered the cartouche of Cleopatra using another bilingual text.
- Based on his ideas about the sound values of glyphs, he decoded his first "mystery" cartouche (no bilingual) text: *alksentrs*, i.e., Alexandros (Alexander the Great)
- He then got his first hieroglyphs from before the Graeco-Roman period, and "deciphered" the cartouche of Ramses.
- To do this, he made an educated guess that the Coptic language was the language of ancient Egyptian writing.

The decoding of Ramses's cartouche

- Champollion knew that  was "s", so he had ?-?-s-s
- Thought the  could be the sun, which was "ra" in Coptic, so ra-?-s-s.
- Observed that vowels were often left out, and only one Pharaonic name fit: Ramses, so  was "m".
- Egyptian scribes had used **the rebus principle**: long words are broken into their phonetic components, which are then represented as logographs:
  - E.g., "belief" can be rewritten as "bee-leaf", and then as  
- Egyptian hieroglyphs is a mixture of such logographs and phonetic symbols.

Wrapping up hieroglyphs

- The fact that the sun - ‘ra’ connection was established made the underlying language of ancient hieroglyphics known: Coptic. As we know from our previous discussion of decryption, knowing the language the cipher text is written in is a huge clue to deciphering it!
- After this breakthrough, Champollion went on to break the rest of the system and published his work in 1824: for the first time in 14 centuries, it was possible to read the history of the pharaohs as written by their scribes.

The Voynich manuscript

- Slides from Kevin Knight, full talk available at:

http://www.isi.edu/natural-language/people/voynich.pdf

- Note that VMS means “Voynich Manuscript”.

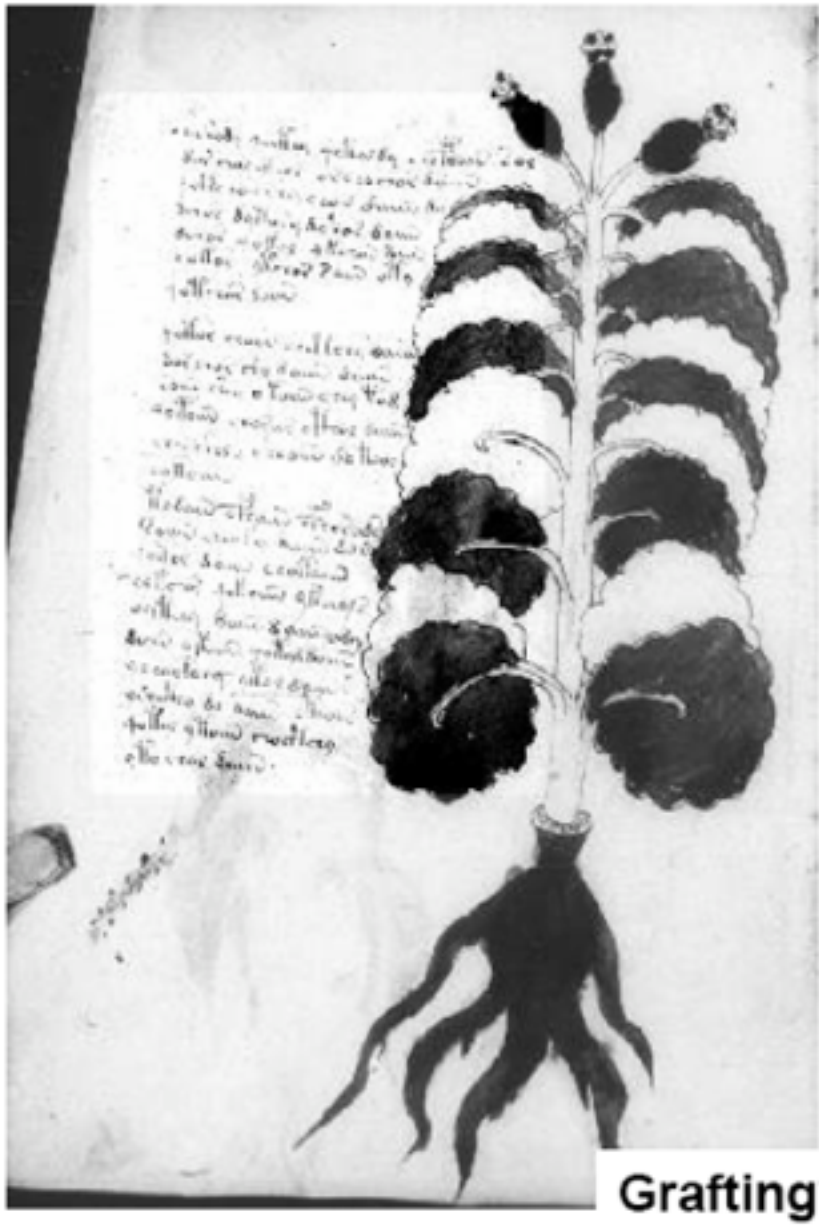
What is it?

- Medieval illustrated manuscript
- Approx. 235 pages on vellum material
- Color drawings of plants, nymphs, stars, etc.
- Approx. 38,000 words written in an unknown script
- Undeciphered!!! Meaning is unknown
- Currently owned by Yale University

Apparent Sections of VMS

- Herbal (11,938 words)
- Astrological (2594 words)
- Biological (6915 words)
- Cosmological (679 words)
- Pharmacological (5111 words)
- Pure Text (10,682 words)

The Pictures: Herbal



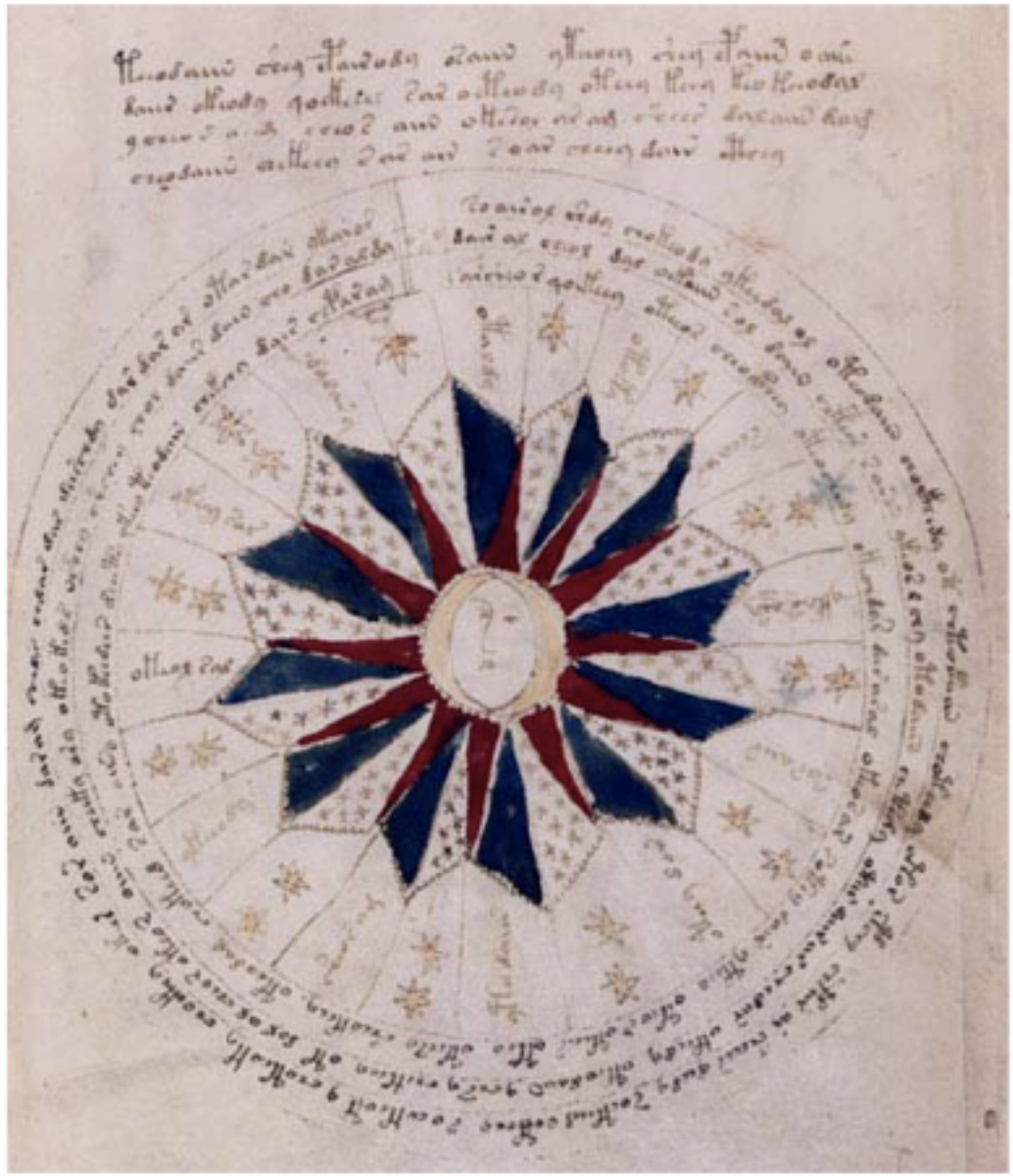
Grafting?

Sunflower?

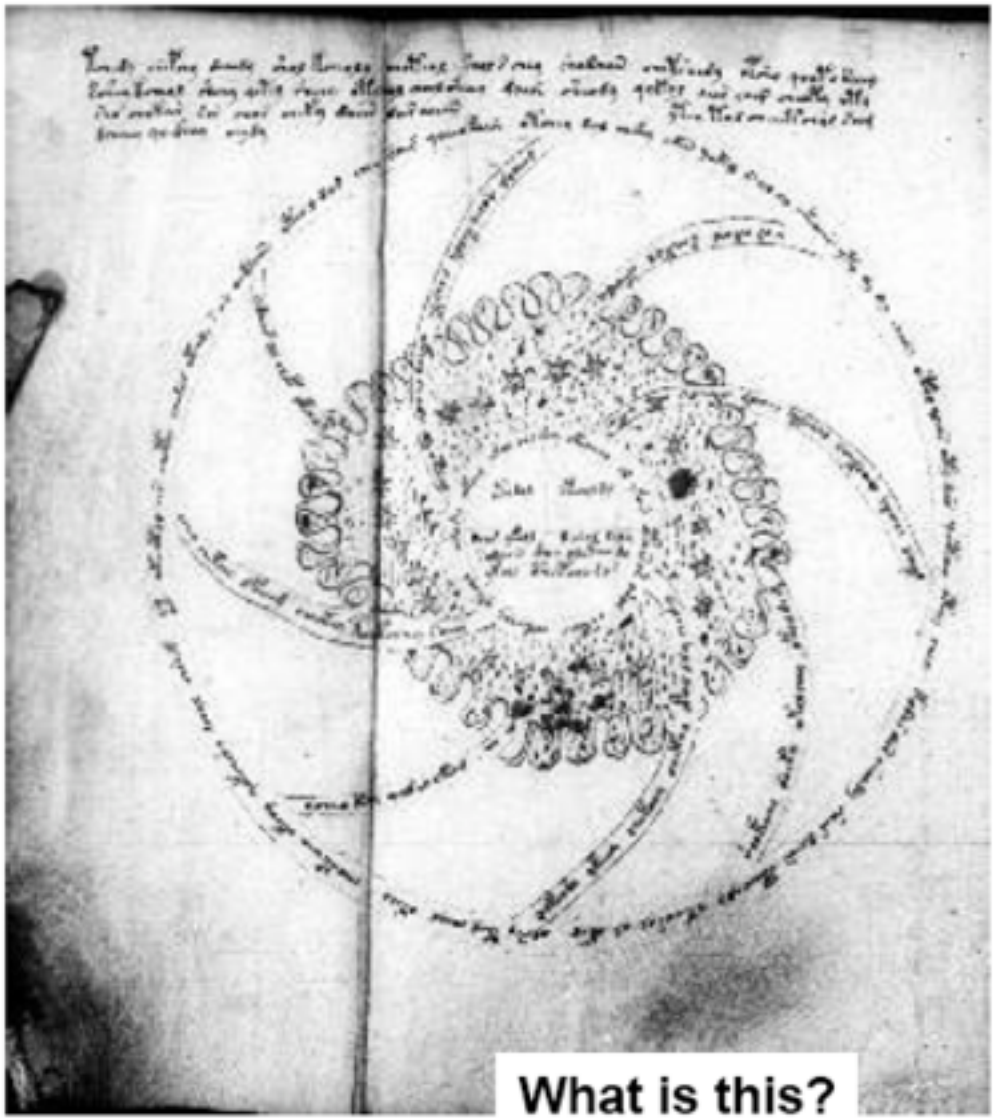
The Pictures: Herbal

- Strange vs ridiculous vs possible
- Many stems grafted onto roots
- Sunflower? Would date VMS as post-1492
- Dana Scott: 21 identifications (5 with confidence)

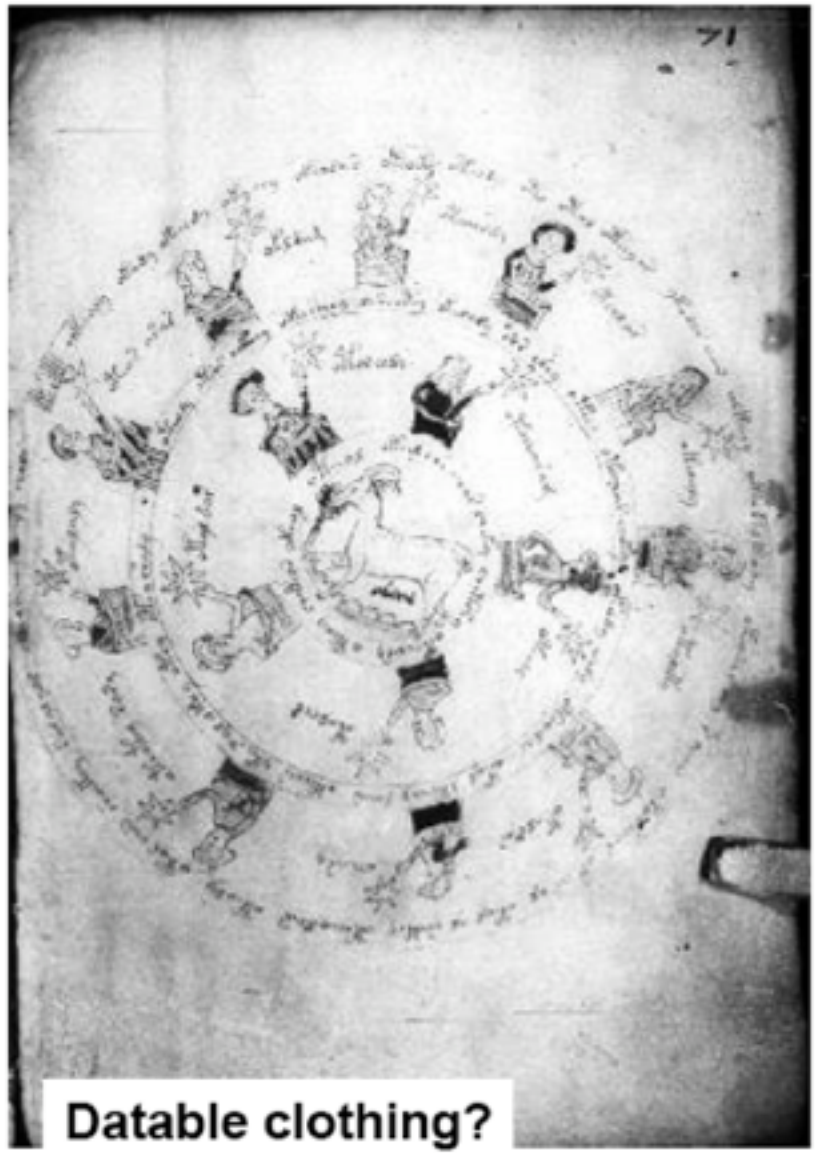
The Pictures: Astrological



The Pictures: Astrological



What is this?



Datable clothing?

The Pictures: Biological

Small nudes in baths

Interconnecting tubes of liquids

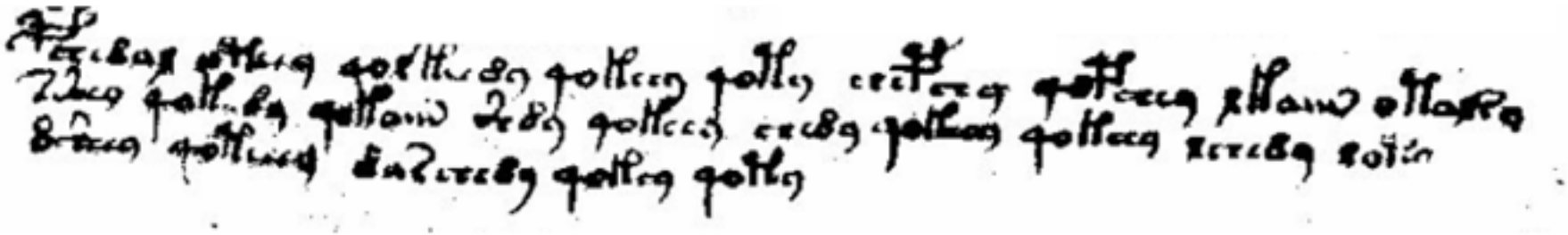
The Pictures: Pharmacological

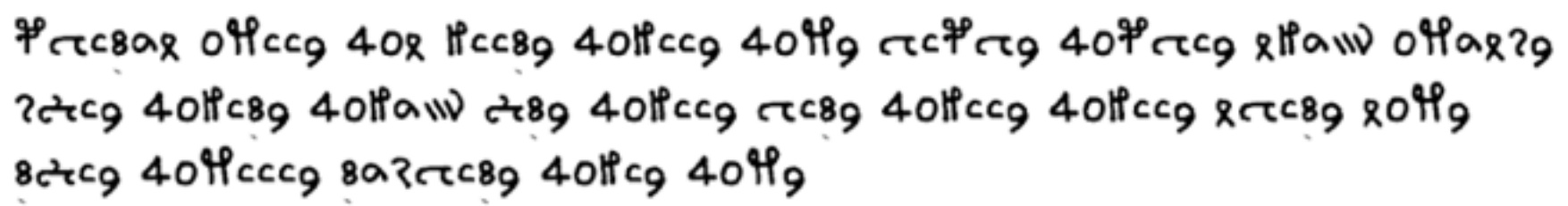
medicine jar?

The Text

- Approx. 38,000 words, unknown script
- Writing style similar to 15th century Florentine “humanist” hand
- Between 23 and 40 distinct characters
- No corrections, likely to have been copied
- Writing was done after illustrations

Transcription



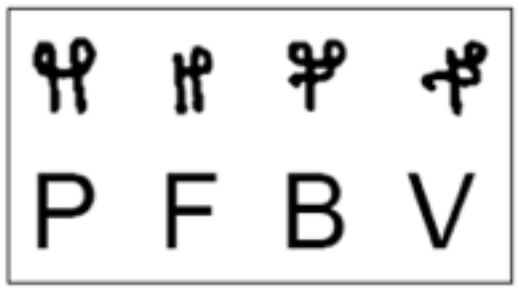


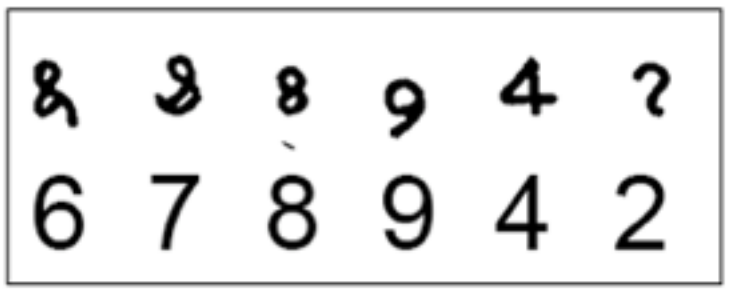
BSC8AE OPCC9 4OE FCC89 4OFCC9 4OP9 SCBS9 4OBSC9 EFAM OPAE29
2ZC9 4OFC89 4OFAM Z89 4OFCC9 SC89 4OFCC9 4OFCC9 ESC89 EOP9
8ZC9 4OPCCC9 8ARSC89 4OFC9 4OP9

last paragraph, f103r

Alphabet: currier/D'Imperio

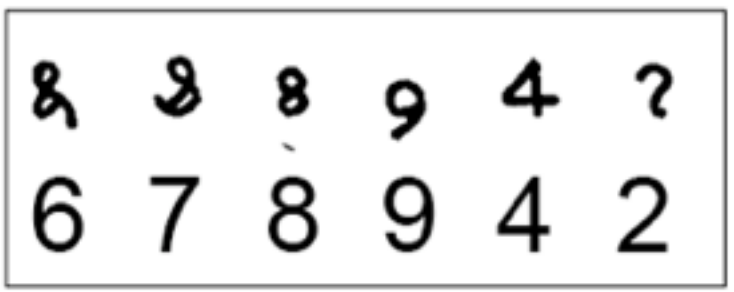
Transcription





Alphabet: currier/D'Imperio

Transcription



← Maybe this is really

IR IIR IIIR

There are several transcription schemes to choose from.

Letter Frequencies

count	letter	
25468	O	
20227	C	
17655	9	
14281	A	
12973	8	
11008	S	
10471	E	
10026	F	
6716	R	
5994	P	
5423	4	
4501	Z	
4076	M	

count	letter	
2886	2	
1752	N	
1413	B	
1046	J	
950	Q	
908	X	
591	T	
524	*	
431	V	
316	I	
217	W	
157	D	
156	3	

count	letter	
148	U	
96	6	
74	Y	
52	K	
31	G	
17	L	
14	H	
2	1	
1	5	
1	0	

Total
63k running characters

Most Frequent Words

count	word	
863	8AM	
537	OE	
501	SC89	
469	AM	
426	ZC89	
396	SOE	
363	OR	
350	AR	
344	SC9	
318	8AR	
308	4OFCC9	
305	4OFCC89	
283	ZC9	
279	4OFAN	
272	4OFC89	
270	89	
262	4OFAM	
260	AE	
253	8AE	
243	2	
219	SOR	

count	word	
212	OFAM	
211	8AN	
191	4OFAE	
186	ZOE	
177	OFCC9	
174	SCC9	
172	SCOE	
155	S9	
155	OPC89	
154	OPAM	
152	4OFAR	
151	9	
151	4OE	
150	S89	
147	4OF9	
144	ZCC9	
144	OFAN	
144	2AM	
143	OPAE	
141	OPAR	
140	SX9	

count	word	
140	OPCC9	
138	OFAE	
130	ZO	
129	OFAR	
119	ESC89	
118	OFC89	

etc

Totals:

8116 distinct words
38k running words

Word Length Distributions

Voynich

Length	Distribution
1	0.02
2	0.10
3	**0.22**
4	**0.23**
5	**0.21**
6	0.12
7	0.05
8	0.01
9	0.003
10	0.001
11	0.0001
12	0.00007
13	0.00002
35	0.00002

English

Length	Distribution
1	0.03
2	**0.15**
3	**0.16**
4	**0.15**
5	**0.11**
6	**0.09**
7	**0.11**
8	0.08
9	0.05
10	0.03
11	0.01
12	0.006
13	0.002

Counts on vocabulary, not running text

Features of the Text

- 115 (out of 8116) words appear doubled at least once

  … 4olfcc89 4olfcc89 …

- 8 words appear tripled at least once

  … 4olfc89 4olfc89 4olfc89 …

  … ctor ctor ctor …

  … ctcor ctcor ctcor …

  … olfaw olfaw olfaw …

  … or or or …

  … 9lfaw 9lfaw 9lfaw …

  … 8aw 8aw 8aw …

  … 4olfcc89 4olfcc89 4olfcc89 …

Kevin Knight

Some Experiments I Did

- Is VMS a phonetic writing system for some known language?
- Is VMS a sort of substitution cipher?

- It's been proposed that VMS is written in a form of vowel-less Ukranian …

Automatic deciphering

- Writing systems can be seen as substitution ciphers for spoken languages.
  - Speech=plaintext: `D IY S AY F ER M EH N T IH Z`
  - Writing=ciphertext: `decipherment is ...`

- So, we'd like to find the most probable sequence of sounds p (for plaintext) for a given writing sample c (ciphertext)
  - This means we want to find $\text{argmax}_p P(p|c)$

Automatic deciphering

- The noisy channel model again!

$$P(p|c) = \frac{P(c|p) \times P(p)}{P(c)} \propto P(c|p) \times P(p)$$

- So, we can solve:

$$\text{argmax}_p \, P(p|c) = \text{argmax}_p \, P(c|p) \times P(p)$$

Automatic deciphering

- The noisy channel model again!

$$P(p|c) = \frac{P(c|p) \times P(p)}{P(c)} \propto P(c|p) \times P(p)$$

- So, we can solve:

$$\text{argmax}_p \, P(p|c) = \text{argmax}_p \, P(c|p) \times P(p)$$

Substitution Model (like the error model in spelling correction)

Automatic deciphering

- The noisy channel model again!

$$P(p|c) = \frac{P(c|p) \times P(p)}{P(c)} \propto P(c|p) \times P(p)$$

- So, we can solve:

$$\text{argmax}_p\ P(p|c) = \text{argmax}_p\ P(c|p) \times P(p)$$

Substitution Model (like the error model in spelling correction)

Language Model

Automatic deciphering

- The noisy channel model again!

$$P(p|c) = \frac{P(c|p) \times P(p)}{P(c)} \propto P(c|p) \times P(p)$$

- So, we can solve:

$$\text{argmax}_p\ P(p|c) = \text{argmax}_p\ P(c|p) \times P(p)$$

Substitution Model (like the error model in spelling correction)

Language Model

We know how to build this for a given language.

Automatic deciphering

- The noisy channel model again!

$$P(p|c) = \frac{P(c|p) \times P(p)}{P(c)} \propto P(c|p) \times P(p)$$

- So, we can solve:

$$\text{argmax}_p\ P(p|c) = \text{argmax}_p\ P(c|p) \times P(p)$$

Substitution Model (like the error model in spelling correction)

But where do we get this?!

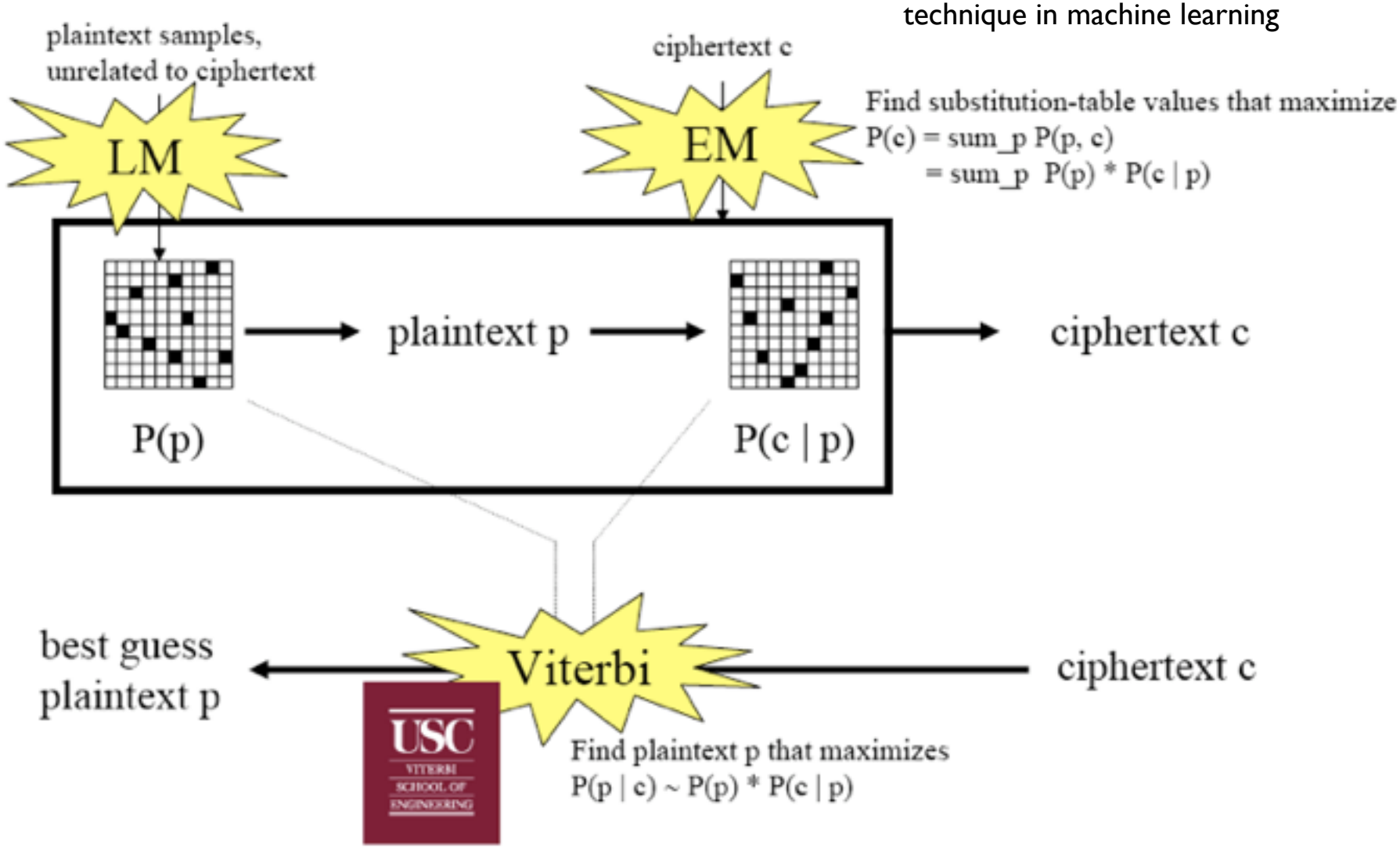
Language Model

We know how to build this for a given language.

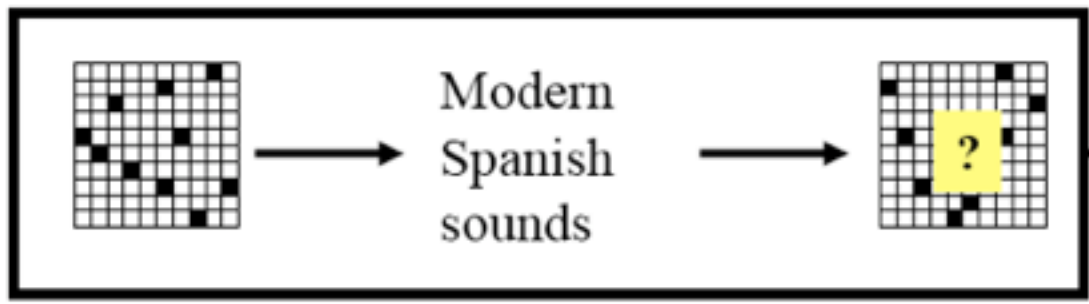
Automatic Decipherment Using EM

[Knight, Nair, Rathod, Yamada, 2006]

Expectation-Maximization: a very important technique in machine learning



Phonetic Decipherment



ciphertext (6980 letters)

```
primera parte
del ingenioso
hidalgo don ...
```

(Don Quixote)

Decoder maximize P(p) * P(c | p)^3	805 errors / 6980
Smooth P(p) with lambdas	684
Use per-symbol lambdas	621
Final Trigram P(p)	492 (7%)

Automatic decipherment pronounces 93% of written letters correctly

Unknown Source Language

- Suppose source language is unknown?

```
ceze ceg qy ataf uqyt qa dwg q y zapu …
VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR …
```

- Decode against all spoken languages:
  - Pre-collect phonetic models for 300 languages
  - Decipher against each
  - See which decoding run yields highest probability

UN Declaration of Human Rights

300+ words in many of world's languages, UTF-8 encoding

No one shall be arbitrarily deprived of his property
Niemand se eiendom sal arbitrêr afgeneem word nie
Asnjeri nuk duhet të privohet arbitrarisht nga pasuria e tij
لا يجوز تجريد أحد من ملكه تعسفا
Janiw khitisa utaps oraqeps inaki aparkaspati
Arrazoirik gabe ez zaio inori bere jabegoa kenduko
Den ebet ne vo tennet e berc'hentiezh digantañ diouzh c'hoant
Никой не трябва да бъде произволно лишен от своята собственост
Ningú no serà privat arbitràriament de la seva propietat
任何人的财产不得任意剥夺。
Di a so prupiità ùn ni pò essa privu nimu di modu tirannicu
Nitko ne smije samovoljno biti lišen svoje imovine
Nikdo nesmí být svévolně zbaven svého majetku
Ingen må vilkårligt berøves sin ejendom
Niemand mag willekeurig van zijn eigendom worden beroofd
Nul ne peut être arbitrairement privé de sa propriété
Nimmen mei samar fan syn eigendom berôve wurde
Ninguín será privado arbitrariamente da súa propiedade
Niemand darf willkürlich seines Eigentums beraubt werden
Κανείς δεν μπορεί να στερηθεί αυθαίρετα την ιδιοκτησία του
Avavégui ndojepe'a va'erâi oimeháicha reinte imbáe teéva
Ba wanda za a kwace wa dukiyarsa ba tare da cikakken dalili ba
Senkit sem lehet tulajdonától önkényesen megfosztani
Engan má eftir geðþótta svipta eign sinni
Tak seorang pun boleh dirampas hartanya dengan semena-mena
Necuno essera private arbitrarimente de su proprietate
Ní féidir a mhaoin a bhaint go forlámhach de dhuine ar bith
Al neniu estu arbitre forprenita lia proprieto
Kelleltki ei tohi tema vara meelevaldselt ära võtta
Eingin skal hissini vera fyri ongartøku
Me kua ni dua e kovei vua na nona iyau
Keltään älköön mielivaltaisesti riistettäkö hänen omaisuuttaan

Unknown Source Language

- Input:

  `cevzren cnegr qry vatravbfb uvqnytb qba dhvwbgr qr yn znapun …`

- Languages with best Prob after deciphering?

Probability

Unknown Source Language

- Input:

  `cevzren cnegr qry vatravbfb uvqnytb qba dhvwbgr qr yn znapun …`

- Top 5 languages with best Prob after deciphering:

-5.29120	spanish
-5.43346	galician
-5.44087	portuguese
-5.48023	kurdish
-5.49751	romanian

- Best-path decoding assuming plaintext is Spanish:

  `primera parte del ingenioso hidalgo don quijote de la mancha …`

- Simultaneous decipherment and language ID

Voynich manuscript

- Input:

  VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR ZOE89 …

- Languages with best Prob after deciphering?

Voynich manuscript

- Input:

  `VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR ZOE89 …`

- Top 10 languages with best Prob after deciphering:

-1.03444	romanian	-1.03546	occitan
-1.03490	zhuang	-1.03568	croatian
-1.03494	polish	-1.03575	chinese
-1.03498	kurdish	-1.03587	albanian
-1.03516	siswati	-1.03594	lingala

- Best-path decoding assuming plaintext is Latin:

  `quiss squm is ONUM pom quss hates s qum hatis …`

Summing up the Voynich manuscript

- Frequency analysis of characters and words provides evidence that it is a real text. (Though, actually, there are ways of mimicking even this.)
- But, even if it isn't a hoax, we don't know the language in which the Voynich manuscript is written, which makes it much harder to get anywhere with decoding it.
- Modern computational linguistics techniques that can be used for deciphering might allow us to detect what the source *language* actually is (though not necessarily the source *text*).

Appendix: The key and answer to the cipher text...

- Reverse the alphabet and then shift:

```
Plain alphabet:  abcdefghijklmnopqrstuvwxyz
Cipher alphabet: MLKJIHGFEDCBAZYXWVUTSRQPON
```

- Here's the unix command:

```
tr 'MLKJIHGFEDCBAZYXWVUTSRQPON' 'a-z'
```

- And the decoded text (from Tolstoy):

```
i would not change my way of life for yours," said she. "we may
live roughly, but at least we are free from anxiety. you live in
better style than we do, but though you often earn more than you
need, you are very likely to lose all you have. you know the proverb,
'loss and gain are brothers twain.' it often happens that people who
are wealthy one day are begging their bread the next. our way is
safer. though a peasant's life is not a fat one, it is a long one.
we shall never grow rich, but we shall always have enough to eat.
```