The monoalphabetic cipher

The monoalphabetic cipher replaces each letter with another according to a table, which forms the key of the cipher. Here is an example, derived from the word “university”:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
UNIVERSTYABCDFGHJKLMOPQWXZ
```

Monoalphabetics are easy to break as they do not hide the patterns which any text contains. For English text, these patterns include:

- *Common short words.* The most common three-letter words are THE and AND and the only possible one-letter words are I, A and O. Among two-letter words, some common ones are listed below - a key point to remember is that two-letter words generally contain one consonant and one vowel. The letter Y can function as a vowel if it is the second letter (as in BY, MY …).
  - starting with A: AM, AN, AS, AT
  - starting with O: OF, ON, OR, (OH)
  - starting with I: IN, IT, IS, IF
  - ending with O: TO, DO, GO, NO, SO
  - ending with S: IS, AS, US
  - ending with E: BE, HE, WE, ME
  - ending with N: IN, ON, AN
- *The relative frequencies of letters.* The most common letters in English are, with rough frequencies, E (12.5%), T (9%), A (8%), O (7.5%), I (7%), N (7%), S (6.5%), H (5.5%), R (6%), D (4%), L (4%) and U (3%).
- *Relative frequencies of letter pairs, triples and so on.* Among letter-pairs, TH (1.5%) and HE (1.2%) are the most common, followed by IN, ER, AN at just under 1% each. THE is the most common three-letter combination at 1.8%, AND and ING both follow at around 0.75%. As a rough rule, vowels tend to pair with lots of other letters, whereas consonants are less “sociable”.

A single good guess can instantly break a monoalphabetic cipher. For example, what might be the beginning of the following letter, which we know from the address is meant for a Mr. Smith?:

```
VEUK DK. LDYMT,

YM YL DX HCEULOKE MG ...
```

Task 1

1. Decrypt the following ciphertext with the cipher given above derived from “university”.

```
NKYLMGC OFYPEKLYMX YL KUFBEV UDGFS MTE MGH RYPE YF

MTE OB RGK KELEUKIT EWIECCEFIE.
```

2. Break the following monoalphabetic cipher. All text is capitalised but the spacing between words is correct. Some statistics on the ciphertext are given below.

```
QTH SIZUHOPZQX JA EOZPQJF VMP AJSICHC ZI HZNTQHHI

PHUHIQX-PZW MP SIZUHOPZQX RJFFHNH, EOZPQJF MIC NMZIHC

M OJXMF RTMOQHO ZI IZIHQHHI TSICOHC MIC IZIH. AOJG

ZQP AZOPQ CMX MP M RJFFHNH, ZQ MCGZQQHC EJQT VJGHI

MIC GHI, QTH AZOPQ SIZUHOPZQX ZI QTH RJSIQOX QJ CJ PJ

- QTJSNT QTHX VHOH MQ AZOPQ IJQ MFFJVHC QJ NMZI

CHNOHHP ZI GHCZRZIH.
```

- Length: 264 characters (not including spaces and punctuation).
- Most frequent letters: H (35, 13.3%), Q (27, 10.2%), Z, I (26, 9.8%), J (18, 6.8%), O, P, M (16, 6%), C (15, 5.6%).
- Most frequent letter pairs: ZI (9), IC, OP, QT, ZQ, SI, HC (6), HI,IH,IZ,HO,PZ,PQ (5).

3. One sometimes hears a suggestion that a monoalphabetic cipher can be made harder to break by replacing the 26 letters of the alphabet not with other letters, but with arbitrary shapes and symbols - after all, there are a lot more symbols than letters. What do you think of this suggestion?

⬡ *Bonus task: write a program in your favourite programming language that reads in a text file and computes letter, pair and triplet frequences. Obviously you only want to only process each character once as you read it, rather than reading the file many times to get the triplet counts.*

The Vigenère cipher

The following cipher has been used, in various forms, by the Confederate army in the US civil war, by the French monarchy in the 16th century and by the mathematician Charles Dodgson a.k.a. Lewis Carroll. It is sometimes also called a polyalphabetic cipher.

Begin by writing your message in capital letters, without spaces or punctuation. Pick a key word (which you share with your co-conspirators of course) and write it out repeatedly underneath the message:

```
MEETMEATMIDNIGHTBYTHEBRIDGE
BRISTOLBRISTOLBRISTOLBRISTO
```

Enciphering and deciphering is done with the alphabet square (in the appendix of this document). In the first column, we have message letter M and key letter B, so we look up the letter at the intersection of row M and column B, to find N which is our first ciphertext letter. Next, we have message letter E and key letter R, so we look up where the E-row meets the R-column to get the next ciphertext letter V. Our ciphertext reads:

```
NVMLFSLUDQVGWRIKJQMVPCIQVZS
```

Task 2

1. Explain how one decrypts a ciphertext with the key word and a Vigenère square.
2. Decrypt the following ciphertext with the keyword BRIDGE:
   `DCQIZSOUWZTWURBLUR`
3. A cipher disk (examples available in the lab or the appendix of this document) is a disk with two copies of the alphabet in concentric cricles. Explain how to encrypt and decrypt with the Vigenère cipher using a cipher disk.
4. A cipher slide (also in the appendix of this document) can also be used to encrypt and decrypt with Vigenère - explain how.

⬡ *Bonus task: Write a program in your favourite programming language that accepts a message and a key, and produces Vigenère-encrypted text. Can you write a version which also handles decryption?*

Note: Mathematically speaking, we can *encode* (not encrypt) the letters of the alphabet as A = 0, B = 1 … Z = 25. In this case, the Vigenère cipher is simply C = M + K (where C is a letter of ciphertext, M of the message and K of the key), added “modulo 26” (any sum that goes to 26 or over, you “wrap around” by subtracting 26 to get it back in range). Decryption is M = C - K (if you go beneath 0 you add 26 again). The same principle works of course if you have another “alphabet”, for example including punctuation and spaces.

Cryptanalysis of the Vigenère cipher

The Vigenére cipher is ‘polyalphabetic’ – it uses a different key for different letters, to disguise patterns that would be obvious for a monoalphabetic cipher. Yet it has a fatal flaw. Look what happens if we use the Vigenère cipher with a very short key, like the letter ‘D’:

```
KILLTHEEMPEROR
DDDDDDDDDDDDDD
NLOOWKHHPSHURU
```

Look familiar? The Vigenère cipher can be considered as a series of Caesar ciphers, one for each letter of the key. If we use only one letter for our key, it’s just a Caesar cipher. Now let’s try a slightly longer key, and a longer message.

```
FIRSTTHEYENCRYPTEDTHISWITHFOURLETTERS
HELPHELPHELPHELPHELPHELPHELPHELPHELPH
MMCHAXSTFIYRYCAILHEWPWHXALQDBVWTAXPGZ
```

If we break the ciphertext into rows every 4 letters, then every column is a Caesar cipher, a very weak monoalphabetic cipher.

```
HELP
MMCH
AXST
FIYR
YCAI
LHEW
PWHX
ALQD
BVWT
AXPG
Z
```

Taking each column as its own ‘message’, you can then tackle each column individually through frequency analysis. Each column can be treated independently like a weaker form of the monoalphabetic cipher you encountered above. Once you have a few letters of the key, analysis often becomes even easier.

Task 3

1. Try to decrypt the Vigenère-ciphered message below, which uses a three-letter word as its key. Characters which share a Caesar-cipher column have been highlighted in the same colour.

FIJCYHMR DCKJYRV YVGQI OGXZ Y KJYZW YRV QXSRIDW EAP EFB FJMYYFX EC

XZC FWCXDC JJMQ S EPSQW UYWW GR OFMUF ML UEK CRUJSKCH AR ASQ E

TCEMRMXSP KAEJYFSCYK YRV YX LFEL RMEC YFIRGUR LM RSRYJYPAQXK MJ

UMYJQI S EVWYX HPMRC MF Y WUGIFRMXGG HMMFR SX TMWU XZCVW UIJC XOM

VGSRV ZPSAO KNSLQ RWYV GLI WVXJCQARC GD XZC FSAO SLH S JSFE SFC RWYV

LFI GRLWP XZC WUYPWQ AWPI WVGWCHALKDW LSPH SLH YJSKQC OGXZ YPD RLW

YTHCEJYRUC SX ZYJLMKFIV ESDB XZC AWGKZR SX RLW GRKCGL UEK TIJW

VWKEJIETJI SLH LYOALK SJP LFMFEW ALXG ASFQMVCVSRMGL M UMYDB LSPHDW

FDYQW HYHGXWP JGP LAQ SHGRAMR JCWHCGLGRY GX

Tips

There are 156 characters in each cipher colum.
The most common red characters (the first cipher column) are, in order:
C (26), Y (20), R (13), Q (11), G (11), M (10), and L (9).

The most common black characters (the second cipher column) are:
X (18), R (17), M (14), I (12), E (11), S (11), and Y (9).

The most common blue characters (the final column) are:
W (20), S (17), J (15), L (10), K (10), D (9), G (9), A (9), and F (9).

Hint: The most common letter in plain English is ‘E’; a common three-letter word is ‘the’.

Kasiski analysis

The frequency analysis method used above relies on knowing the length of the key used in encryption. The real trick to breaking a Vigenère cipher is the method for figuring out the length of the key used by a cipher. This is known as **Kasiski analysis**, and it works by exploiting repetition in the underlying message. Take a look at this message:

cryptanalysis beats up **crypt**ography.

The sequence ‘crypt’ is repeated twice in the text, 20 characters apart (ignoring spaces, start to start). If we have a key that lines up with this gap (if the keylength divides 20), then look what happens:

CRYPTANALYSISBEATSUP**CRYPT**OGRAPHY

ABCDEABCDEABCDEABCDE**ABCDE**ABCDEAB

CSASXAOCOCSJUEIAUUXT**CSASX**OHTDTHZ

Because the same text was encoded with the same key material, the ciphertext shows these suspiciously similar repeated patterns. The gap between these patterns gives us some information on the key’s length. We usually find that the key length is a **divisor** of this gap, and is often the **greatest common divisor** of multiple such gaps in the ciphertext. Look at the repeted patterns in the ciphertext below.

```
ALPNALPCZIYIALTHALCDBKSHLGCTAGSPURPAZ
| 4 |   8   | 4 |
```

The distances between patterns (start-to-start) are 4 or 8 characters long (12 or 16 if you suspect one of the earlier ones is a coincidence). The **greatest common divisor** of 4 and 8 is 4, so this is a likely key length for this ciphertext. Other possibilities are 2 or 1. If the message was long enough, we could now perform frequency analysis on 4 cipher columns to unlock the content.

Task 4

1. Using Kasiski analysis, what key length might you suspect for the following ciphertext?

YCAMULRILETMYCASSCYCAMLYBEQYCAPUCQSCTELYWYPDQ

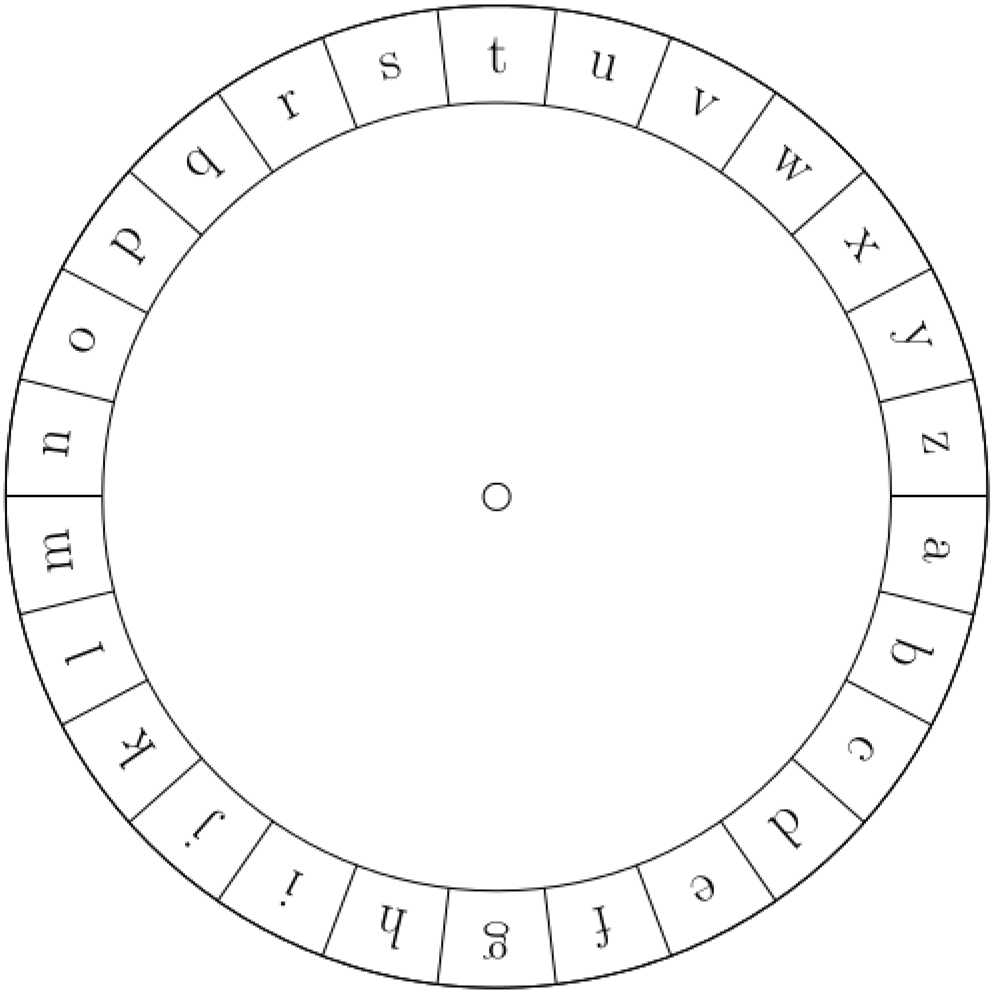
2. Kasiski analysis depends upon repetition of the key. How might you choose a key for Vigenère encryption so that this analysis is not possible?

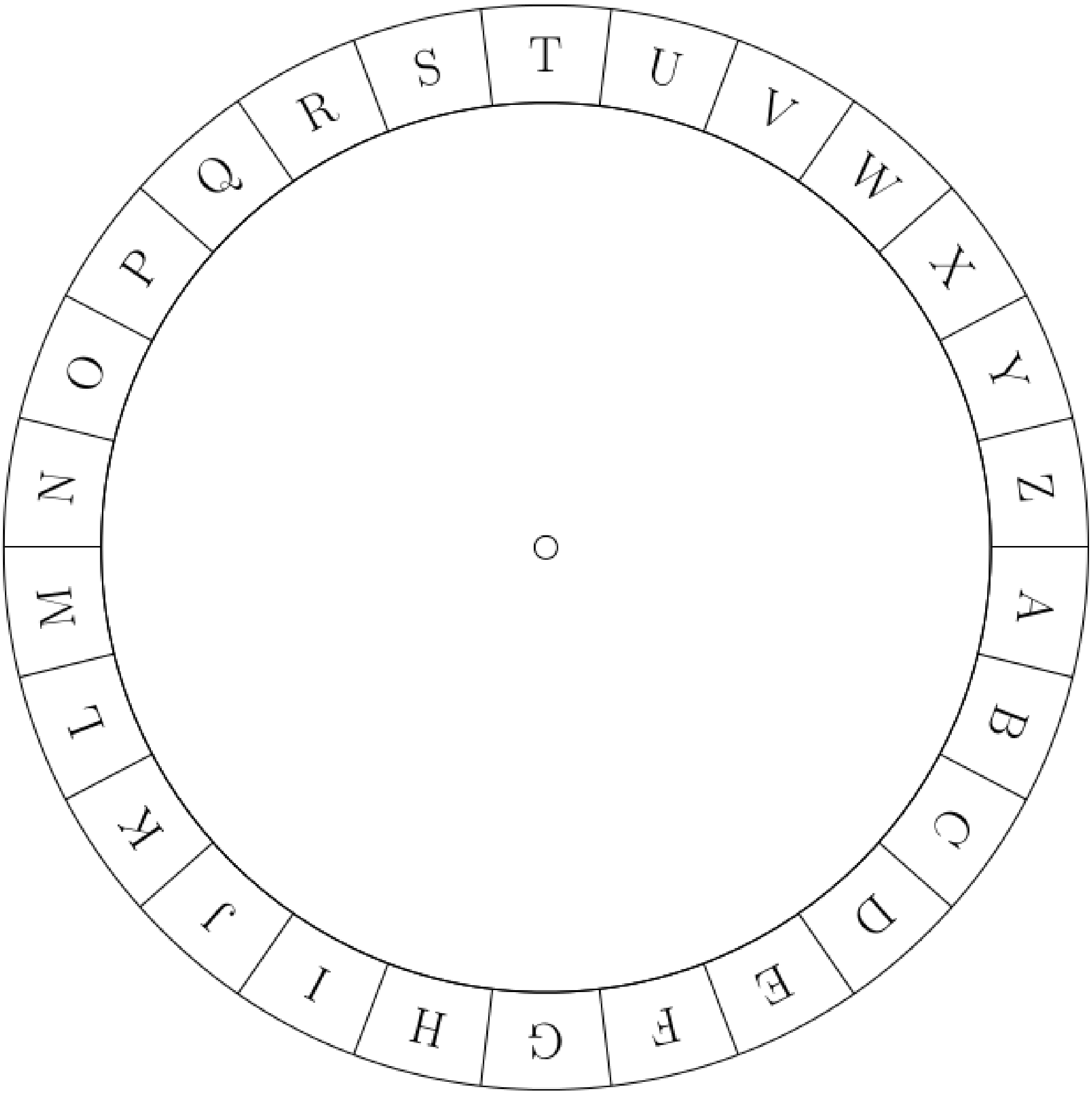
Appendix

The alphabet square for the Vigenère cipher:

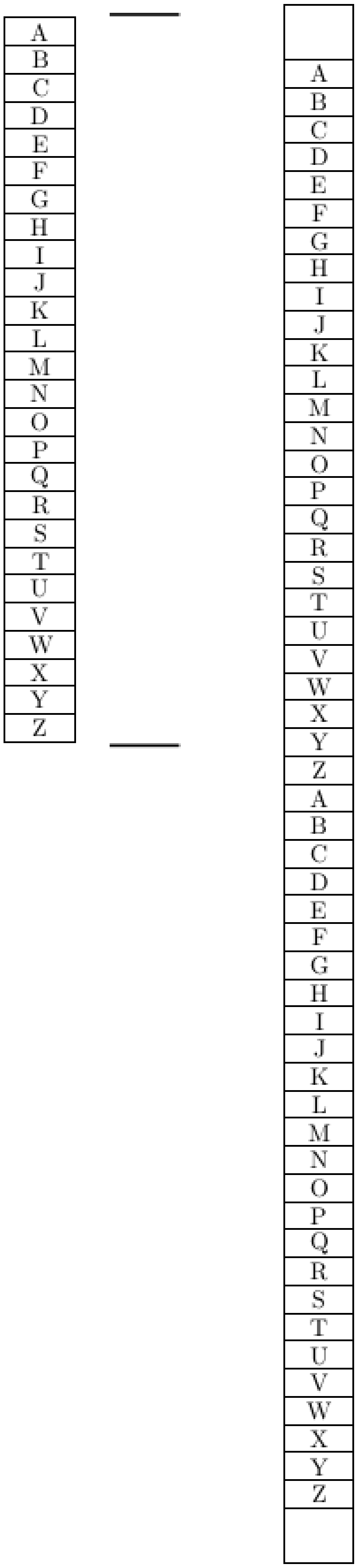
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V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

Cipher disk: cut out the inner ring and fasten it to the outer ring with a pin.




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

Cipher slide: cut out the right-hand (long) slide and make two slits at the thick black lines, then feed the long slide through so that 26 letters are visible.


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 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 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