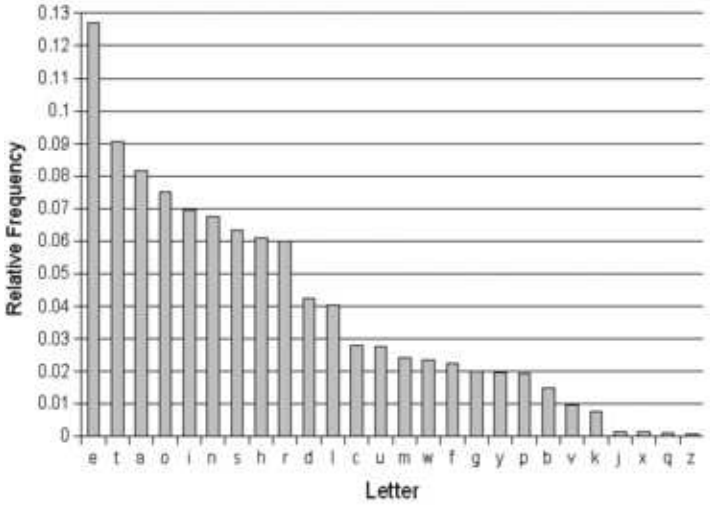
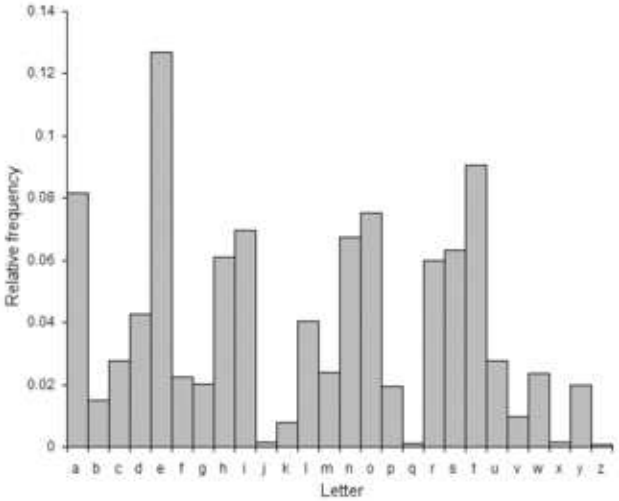
Letter and Digraph Frequency in English

Letter	Frequency (%)	Letter	Frequency (%)
a	8.167	e	12.702
b	1.492	t	9.056
c	2.782	a	8.167
d	4.253	o	7.507
e	12.702	i	6.966
f	2.228	n	6.749
g	2.015	s	6.327
h	6.094	h	6.094
i	6.966	r	5.987
j	0.153	d	4.253
k	0.772	l	4.025
l	4.025	c	2.782
m	2.406	u	2.758
n	6.749	m	2.406
o	7.507	w	2.360
p	1.929	f	2.228
q	0.095	g	2.015
r	5.987	y	1.974
s	6.327	p	1.929
t	9.056	b	1.492
u	2.758	v	0.978
v	0.978	k	0.772
w	2.360	j	0.153
x	0.150	x	0.150
y	1.974	q	0.095
z	0.074	z	0.074

Digraph	Frequency
th	1.52
he	1.28
in	0.94
er	0.94
an	0.82
re	0.68
nd	0.63
at	0.59
on	0.57
nt	0.56
ha	0.56
es	0.56
st	0.55
en	0.55
ed	0.53
to	0.52
it	0.50
ou	0.50
ea	0.47
hi	0.46
is	0.46
or	0.43
ti	0.34
as	0.33
te	0.27
et	0.19





Most common double letters: ll, ee, ss, oo, ff, tt

Most common trigraphs: the and tha ent ion tio for nde has nce

Most common starting letters: t o a w b c d s f m

Half of all words end in: e t d s

Example cryptanalysis of a Monoalphabetic Cipher:

FQJCBRWJWJVNJAXBNKHJWHXCQNAWJVNFXDUMBVNUUJBBFNNC

Frequency Count:

```
j-7   frequency = 0.14583333333333334
n-7   frequency = 0.14583333333333334
b-5   frequency = 0.10416666666666667
w-4   frequency = 0.08333333333333333
c-3   frequency = 0.0625
f-3   frequency = 0.0625
u-3   frequency = 0.0625
v-3   frequency = 0.0625
x-3   frequency = 0.0625
a-2   frequency = 0.041666666666666664
h-2   frequency = 0.041666666666666664
q-2   frequency = 0.041666666666666664
d-1   frequency = 0.020833333333333332
k-1   frequency = 0.020833333333333332
m-1   frequency = 0.020833333333333332
r-1   frequency = 0.020833333333333332
```

Either J or N is plaintext E. From digram frequencies, most common digram is EE. So, choose N as E.

```
Cipher: FQJCBRWJWJVNJAXBNKHJWHXCQNAWJVNFXDUMBVNUUJBBFNNC
Plain:  FQJCBRWJWJVeJAXBeKHJWHXCQeAWJVeFXDUMBVeUUJBBFeeC
```

Next, consider BB. From the frequency of B and the presence of digram BB, we choose B as S.

```
Cipher:  FQJCBRWJWJVeJAXBeKHJWHXCQeAWJVeFXDUMBVeUUJBBFeeC
Plain:   FQJCsRWJWJVeJAXseKHJWHXCQeAWJVeFXDUMsVeUUJssFeeC
```

Next, consider UU. From the frequency of U and the presence of digram UU we chose U as L.

```
Cipher:  FQJCsRWJWJVeJAXseKHJWHXCQeAWJVeFXDUMsVeUUJssFeeC
Plain:   FQJCsRWJWJVeJAXseKHJWHXCQeAWJVeFXDlMsVellJssFeeC
```

Consider the ellJss. J is high frequency letter. So it has to map to T, A, O, I, etc (in decreasing order of probabilities). Putting J as T, elltss does not make a proper word. Putting J as A we get ellass which seems to be proper ell as s. So, putting J as A, we get

```
Cipher:  FQJCsRWJWJVeJAXseKHJWHXCQeAWJVeFXDlMsVellJssFeeC
Plain:   FQaCsRWaWaVeaAXseKHaWHXCQeAWaVeFXDlMsVellassFeeC
```

Now attack the last word. sFeeC, this most likely is sweet (other words do not fit in the context of smell). So, replacing F with w and C with t,

```
Cipher:  FQaCsRWaWaVeaAXseKHaWHXCQeAWaVeFXDlMsVellassFeeC
Plain:   wQatsRWaWaVeaAXseKHaWHXtQeAWaVewXDlMsVellassweet
```

The trigram tQe is most likely the. So replacing Q with t, we have

```
Cipher:  wQatsRWaWaVeaAXseKHaWHXtQeAWaVewXDlMsVellassweet
Plain:   whatsRWaWaVeaAXseKHaWHXtheAWaVewXDlMsVellassweet
```

The first word is whats. The second word that follows must be ‘in’. So replacing R with I and W with n, we get

```
Cipher:  whatsRWaWaVeaAXseKHaWHXtheAWaVewXDlMsVellassweet
Plain:   whatsinanaVeaAXseKHanHXtheAnaVewXDlMsVellassweet
```

naVe must be name. So replacing V with m, we get,

```
Cipher:  whatsinanaVeaAXseKHanHXtheAnaVewXDlMsVellassweet
Plain:   whatsinanameaAXseKHanHXtheAnamewXDlMsmellassweet
```

Now splitting it in to words,

```
Whats in a name AXseKHanHXtheA name wXDlM smell as sweet
```

Little googling (or from memory) we can recognize this as quote from Shakespear’s Romeo and Juliet.

So we get the plaintext:

```
Whats in a name? Rose by any other name would smell as sweet.
```