Decrypting a Text (1964 letters) Encrypted by Substitution

Here is a 1964-character text encrypted by substitution (non-letters removed). We know that the plain text was English language text. We decrypt it by frequency analysis.

```
RIRPKQJHIRBBTCLSWQKUXBVERIVEHUZGCIGIVQJWKGHWJCSGUVLUGHVBCSCLKVXVQBPJRLHURBTKTKUH
EGRKJRLWHWVLKIGHWQJHCLUGHQLCUHWKURUHKRUUGHHLWVXRKCNTHRBECKKCVLUGHKURBWQKUBHUQBLI
RPKQJHKJVZHWXBVEEPGUVXHHUPHBKHIVLWRKCUBHUQBLHWRUREERBDCLSUGHXCBKUKQIIHKKXQJIVJJH
IUCVLVXIVEHURBTRLWCLUHBKUHJJRBERUHBCRJRLWWBRZCLSIGHHBKXBVELRKRKICHLUCKUKOVHMHJJC
LSRRLWKHMHBRJWVYHLIVJJHRSQHKSRUGHBHWRLNCVQKJTTHKUHBWRTRURECJCURBTUHKUCLSBRLSHCLQ
URGZRUIGCLSRKUGHPBVFHUGHKCYHVXRJRBSHPJRLUPVUGQBUJHWCLUVUGHRUEVKPGHBHXRKUHBUGRLRL
TERLERWHVFOHIUGRWWVLHFHXVBHRKIVEHUKRBHJRBSHFRJJKVXWCBUTCIHJHXUVMHBXBVEUGHIBHRUCV
LVXUGHKVJRBKTKUHEFCJJCVLTHRBKRSVKICHLUCKUKHNPHIUUGRUUGHKREPJHKZCJJBHMHRJIJQHKRFV
QUGVZUGHPJRLHUKZHBHXVBEHWRLWPVKKCFJTUGHVBCSCLKVXJCXHEBMHJJCLSRKLHBMVQKWCKPVKCUCV
LZRKLVUKVJHJTFHIRQKHVXUGHGCSGKPHHWIQJECLRUCVLVXUGHECKKCVLXVBZGCIGGHCKPBVSBREEHER
LRSHBXVBJVIDGHHWERBUCLKPRIHKTKUHEKUGHIVEPRLTUGRUWHKCSLHWRLWIVLKUBQIUHWUGHPBVFHRL
WCUKEVUGHBKGCPKCNUHHLEVLUGKRSVGHZRUIGHWCLGVBBVBRKUGHSHLHKCKBHUQBLPBVFHIRBBTCLSKV
JRBZCLWPRBUCIJHKFQCJUFTUGHKREHUHREQLWHBGCKWCBHIUCVLRLWFRKHWVLKCECJRBUHIGLVJVSTKJ
REEHWCLUVUGHSBVQLWRUEPGHMHBTFVWTCKIVLXCWHLUCLZGRUZHGRMHWHKCSLHWRLWFQCJURLWUHKUHW
GHKRCWFQUHMHBTFVWTRJZRTKZVBBCHKRFVQUZGRUZHGRMHLVUUGVQSGUVXKURBWQKUZRKJRQLIGHWCLX
HFBQRBTCUHLIVQLUHBHWUGHZCJWIVEHUUZVTHRBKRSVRLWGVPHXQJJTSRUGHBHWWQKUXBVECUKURCJQK
CLSRPRWWJHKGRPHWJCDHRUHLLCKBRIAQHURHBVSHJRIVJVQBJHKKSRKHVQKOHJJTUGRUCKUGHJCSGUHK
UERUHBCRJVLHRBUGRLWCKDLVZLUVKICHLUCKUKRKXBVYHLKEVDHZRKQKHWCLUGHRUUHEPUUVUBRPUGHE
RUHBCRJWBERUUSHLSHXBVECEPHBCRJIVJJHSHJVLWVLRJHRWCLSHNPHBUVLHNUBRUHBBHKUBCRJWQKUZ
CJJFHVLHVXUGHXCBKUFBCUCKGKICHLUCKUKUVBHIHCMHWQKUKREPJHKUVRLRJTKHUGCKCKUGHXCBKUUC
EHKCLIHUGHRPVJJVECKKCVLKUGRUKREPJHKVXBVIDGRMHFHHLBHUQBLHWXBVEKPRIHUVHRBUGGHKRCWU
GCKUGVQKRLWUGVXRSBREEHVXWQKUXBVEZCJWZCJJPBVFRFJTUHJJQKEVBHRFVQUUGHXVBERUCVLVXUGH
KVJRBKTKUHEUGRLUGHPRKUTHRBKVXUHJHKIVPHVFKHBMRUCVLKVXUGHKHVFOHIUKCUCKRSBHRUUCEHUV
FHCLUVWQKUERLTFHJCHMHUGRUIVEHUKWHJCMHBHWEVKUVXUGHZRUHBVLHRBUGRLWPVKKCFJTUGHVBSRL
CIIVEPVQLWKUGRUZHBHMCURJUVUGHHMVJQUCVLVXJCXH
```

1) It is very likely that e → H, h → G, t → U.

Look at the frequencies of common plaintext and ciphertext letters and digrams, all expressed in occurrences per 10000 characters of text. Plaintext frequencies are estimated using about 3.2 million characters of what is (hopefully) fairly typical English language text. (Recall lower case indicates plaintext, upper case indicates cibertext).

Letter	Frequency (per 10,000)
e	1237
t	921
a	821
o	767
n	705
i	676
h	645
s	617
r	550
d	479
l	393
u	291

Letter	Frequency (per 10,000)
H	1253
U	1013
R	845
K	804
V	743
C	642
B	616
L	580
J	484
G	479
W	397
E	316
I	280
Q	260
X	229
S	204

Digram	Frequency (per 10,000)
th	330
he	302
an	181
in	179
er	169
nd	146
re	133
ed	126
es	115
ou	115
to	115
ha	114
en	111
ea	110
st	109
nt	106
on	106
at	104

Digram	Frequency (per 10,000)
UG	290
GH	250
KU	219
HK	163
RU	158
UH	153
HB	148
RL	143
HW	138
VL	133
CL	133
RB	133
UC	107
LW	107
HU	102
HR	102
VX	102
BH	102
CK	102

Merely from the single-letter frequencies, it appears very likely that **e → H**. Otherwise the plain text character λ for which λ → **H** would have *freq*(λ) ≤ 921, versus *freq*(**H**) = 1253. So we assume **e → H**. In the unlikely event this is wrong, we will discover our error, and backtrack.

The highest-frequency plaintext digram **th** (*freq* = 330) is very likely mapped to one of the highest-frequency ciphertext digrams: **UG** (*freq* = 290), **GH** (*freq* = 250), or **KU** (*freq* = 219. But **GH** is impossible because **e → H**, and **KU** is unlikely because *freq*(**U**) = 1013 is way higher than *freq*(**h**) = 645. So we assume **th → UG**. In particular **t → U** and **h → G**.

Comparing frequencies of characters we have decrypted so far, we have

plaintext	**e**	**t**	**h**	**ee**	**tt**	**hh**	**et**	**te**	**eh**	**he**	**th**	**ht**
frequency	1237	921	645	48	56	6	83	75	33	302	330	32

ciphertext	**H**	**U**	**G**	**HH**	**UU**	**GG**	**HU**	**UH**	**HG**	**GH**	**UG**	**GU**
frequency	1253	1013	479	41	56	10	102	153	15	250	290	20

Agreement is quite good (better than we might expect) except in two entries:

i) The frequency **G** is rather low, compared to **h**.

ii) The frequency of **UH** is surprisingly high, given that of **te**.

However, even in these cases the deviations aren't as large as those we used above to reject alternatives (e.g., to reject **t** → **H** or **h** → **U** or **th** → **RU**).

2) It is very likely that n → L, and {a,o,i} → {R,V,C} in some order.

In our plaintext data, the third and fourth most common digrams are **an** (*freq*=181) and **in** (*freq*=179). In addition, **en** (*freq*=111) and **on** (*freq*=106) are very common. On the other hand, **tn** (*freq* = 7) and **hn** (*freq* = 1) are quite rare.

We don't yet know how to encrypt **a**, **i**, **o**, or **n**, but they are among the six most frequent plaintext letters, so we expect them to encrypt to high-frequency ciphertext letters. It is probably safe to assume that **a** and **o** encrypt to a letter **L** or above, in the list of ciphertext letters sorted by frequency, and **n** and **i** to a letter **G** or above. Excluding **H**, **U**, and **G**, this means each of **a**, **i**, **o**, **n** must encrypt to ones of **R**, **K**, **V**, **C**, **B**, **L**, **J**.

Consider the following table of digram frequencies below. The first letter of the digram is given at left, the second letter at top.

		Alternatives for the letter that **n** *encrypts to*						
What should occur in the column for the letter that n encrypt to?		**R**	**K**	**V**	**C**	**B**	**L**	**J**
Close to *freq*(**en**) = 111.	**H**	102	163	61	25	148	76	71
Close to *freq*(**tn**) = 7.	**U**	66	71	82	107	31	0	5
Close to *freq*(**hn**) = 1	**G**	92	15	31	41	0	5	0
The three highest frequency entries should be a reasonable match for the 181, 179 and 106 (the frequencies of **an**, **in** and **on**.)	**R**	---	87	0	15	133	143	66
	K	87	---	51	92	10	10	25
	V	10	36	---	0	66	133	56
	C	25	102	66	---	25	133	56
	B	46	61	97	46	---	20	10
	L	36	46	46	15	5	---	0
	J	76	5	36	56	5	0	---

The only good matches in the table above occur in the column for L, and in the bottom part of the table they occur in the rows for R, V, and C. So it is very likely that **n** → **L**, and **{a, o, i}** → **{R, V, C}** in some order.

3) It is likely that a → R, o → V, i → C.

We already know each of **a**, **o**, and **i** encrypts to one of **R**, **V**, or **C**. Consider the following table of single-letter and digram frequencies. Each box contains a letter or digram and its frequency.

a 821	**aa** 1	**ae** 0	**ea** 110	**at** 104	**ta** 59	**ah** 5	**ha** 114	**an** 181	**na** 40
o 767	**oo** 36	**oe** 5	**eo** 33	**ot** 57	**to** 115	**oh** 11	**ho** 49	**on** 106	**no** 60
i 676	**ii** 1	**ie** 23	**ei** 41	**it** 93	**ti** 76	**ih** 6	**hi** 97	**in** 179	**ni** 33

R 854	**RR** 5	**RH** 5	**HR** 102	**RU** 158	**UR** 66	**RG** 5	**GR** 92	**RL** 143	**LR** 36
V 743	**VV** 0	**VH** 10	**HV** 61	**VU** 41	**VR** 82	**VG** 5	**GV** 31	**VL** 133	**LV** 46
C 642	**CC** 0	**CH** 31	**HC** 25	**CU** 51	**CR** 107	**CG** 0	**GC** 41	**CL** 133	**LC** 15

Consider the row for **a**. It matches the row for **R** fairly reasonably in all positions, and very well in most positions. It doesn't match the row for **V** as well due to major variations in the columns for **at** and **ha**. And it doesn't match the row for **C** well at all; note especially the columns for **a**, **ea**, and **ha**.

So we conclude it is likely that **a** → **R**.

We can try to match the row for **o** with the rows for **V** or **C**, but there is no clear winner.

To separate **i** and **o**, we can make use of the fact that **ou** is a very common diagram (*freq*=**115**), while **iu** is quite rare (*freq*=1). The plaintext letter **u** (*freq* = 291) should encrypt to a ciphertext letter of somewhat comparable frequency – probably a letter between **S** (*freq* = 204) and **W** (*freq* = 397) in the list of ciphertext letters, sorted by frequency. These letters are **S**, **X**, **Q**, **I**, **E**, **W**.

Consider this table of digram frequencies. The first letter of the digram is taken from the column at left, the second from the row at top.

	S	**X**	**Q**	**I**	**E**	**W**
V	15	<u>102</u>	<u>71</u>	10	<u>82</u>	15
C	36	10	0	25	20	15

Only **VX**, **VQ**, and **VE** are reasonable candidates for **ou** encrypted. In any case, **o** → **V**. This leaves us with **i** → **C**.

4) It is likely that s → K .

Consider what letter encrypts to **K** (*freq* = 804). All plaintext letters with frequencies higher than **s** (*freq*=617) have been accounted for. Although the frequencies of **s** and **K** are a bit further apart than we would expect for **s→K**, any other letter that could encrypt to **K** would have a frequency of 550 or less — a substantially worse discrepancy. So we assume **s→K**.

At this point, we have deduced the probable encryptions of the eight most common plaintext letter. Here we decrypt the text, using a dot to indicate a letter we do not yet know how to decrypt.

```
a.a.s..e.a...in...st..o.a.o.et.hi.h.o...she..i.htontheo.i.inso.o....aneta..s.ste
.has.an.e.ons.he...einthe.nite.statesattheen.o.asi..ea..issionthesta...st.et..n.
a.s..es.o.e...o...hto.eet.e.se.on.asit.et..ne.ata..a..in.the.i.sts...ess....o..e
.tiono..o.eta..an.inte.ste..a..ate.ia.an...a.in..hee.s..o.nasas.ientists.oe.e..i
n.aan.se.e.a..o.en.o..ea..es.athe.e.an.io.s...este..a.ata.i.ita..testin..an.ein.
tah.at.hin.asthe..o.ethesi.eo.a.a..e..ant.oth..t.e.intotheat.os.he.e.aste.thanan
..an.a.eo..e.tha..one.e.o.eas.o.etsa.e.a..e.a..so..i.t.i.e.e.to.e...o.the..eatio
no.theso.a.s.ste..i..ion.ea.sa.os.ientistse..e.tthatthesa...es.i...e.ea....esa.o
.tho.the..anets.e.e.o..e.an..ossi...theo.i.inso..i.e...e..in.asne..o.s.is.ositio
n.asnotso.e...e.a.seo.thehi.hs.ee.....inationo.the.ission.o..hi.hheis..o..a..e.a
na.e..o..o..hee..a.tins.a.es.ste.sthe.o..an.that.esi.ne.an..onst...te.the..o.ean
.its.othe.shi.si.teen.onthsa.ohe.at.he.inho..o.asthe.enesis.et..n..o.e.a...in.so
.a..in..a.ti..es..i.t..thesa.etea..n.e.his.i.e.tionan..ase.onsi.i.a.te.hno.o..s.
a..e.intothe..o.n.at..he.e...o..is.on.i.entin.hat.eha.e.esi.ne.an...i.tan.teste.
hesai...te.e...o..a..a.s.o..iesa.o.t.hat.eha.enottho..hto.sta...st.as.a.n.he.in.
e...a..iten.o.nte.e.the.i...o.ett.o.ea.sa.oan.ho.e......athe.e...st..o.itstai..s
in.a.a...esha.e..i.eatennis.a...etae.o.e.a.o.o...ess.aseo.s.e...thatisthe.i.htes
t.ate.ia.onea.than.is.no.ntos.ientistsas..o.ens.o.e.as.se.intheatte..ttot.a.the.
ate.ia....att.en.e..o.i..e.ia..o..e.e.on.ona.ea.in.e..e.tone.t.ate..est.ia...st.
i...eoneo.the.i.st..itishs.ientiststo.e.ei.e..stsa...estoana..sethisisthe.i.stti
.esin.ethea.o..o.issionsthatsa...eso..o..ha.e.een.et..ne...o.s.a.etoea.thhesai.t
histho.san.tho.a..a..eo...st..o..i...i....o.a...te...s.o.ea.o.tthe.o..ationo.the
so.a.s.ste.thanthe.ast.ea.so.te.es.o.eo.se..ationso.theseo..e.tsitisa..eatti.eto
.einto..st.an..e.ie.ethat.o.ets.e.i.e.e..osto.the.ate.onea.than..ossi...theo..an
i..o..o.n.sthat.e.e.ita.tothee.o..tiono..i.e
```

At this point, we could finish the decoding by inspection of the text above. It is easier if we discover the encodings of a few more letters by frequency analysis.

5) It is likely that d → W, r → B, and l → J.

Here is a table of useful letter and digram frequencies for **r**, **d**, and **l**, followed by the corresponding frequencies of ciphertext letters/digrams to which they might reasonably encrypt (based on character frequencies).

λ	Plaintext estimated frequency of λ and digrams involving λ																	
	λ	λλ	λe	eλ	λt	tλ	λh	hλ	λa	aλ	λn	nλ	λo	oλ	λi	iλ	λs	sλ
r	550	14	133	169	42	28	12	8	50	75	16	3	55	84	50	27	37	4
d	479	13	57	126	56	7	25	1	48	52	16	146	41	18	50	33	35	7
l	393	56	64	55	15	17	4	2	40	57	2	9	41	26	47	37	11	13

Φ	Cibertext frequency of Φ and digrams involving Φ																	
	Φ	ΦΦ	ΦH	HΦ	ΦU	UΦ	ΦG	GΦ	ΦR	RΦ	ΦL	LΦ	ΦV	VΦ	ΦC	CΦ	ΦK	KΦ
B	616	25	102	148	51	31	5	0	46	133	20	5	97	66	46	25	61	10
J	484	76	87	71	15	5	0	0	76	66	0	0	36	56	56	56	5	25
W	397	20	36	138	25	5	10	0	41	20	0	107	31	15	61	15	20	15
E	316	25	61	46	10	20	5	0	61	51	5	10	31	82	41	20	10	15
I	280	10	31	66	31	15	41	5	25	31	0	25	87	10	20	25	0	41
Q	260	0	15	5	31	20	0	5	0	10	31	5	0	71	10	0	82	20

d might reasonably encrypt to **B**, **J**, **W**, or possibly **E**. But only **LW** (*freq*=107) comes anywhere close to matching the high frequency of **nd** (*freq*=146). Noting also that the entire row for **W** matches that for **d** fairly well, we assume **d→W**.

r might reasonably encrypt to **B** or **J**. **BH** and **HB** provides a significantly better match for the high-frequency digrams **re** and **er**, than do **JH** and **HJ**. In fact, in 14 columns of the table, row **r** more closely matches row **B** than row **J**, in two columns it matches row **J** more closely, and in two columns there is a tie. (In many rows, the differences are not significant by themselves.). So we assume **r → B**.

Suppose **l** does not encrypt to **J**? The next-highest frequency plaintext letter available is **u**. It seems unlikely that **u** (*freq*=291) would encrypt to **J** (*freq*=484).

We have deduced the probable encryptions of the eleven most common plaintext letters. Here is our text with these eleven letters decrypted.

```
aIaPsQleIarrTinSdQstXroEaIoEetZhiIhIoQldshedliShtontheoriSinsoXoQrPlanetarTsTste
EhaslandedonsIhedQleintheQnitedstatesattheendoXasiNTearEissionthestardQstretQrnI
aPsQlesloZedXroEEPhtoXeetPerseIondasitretQrnedataEEarDinStheXirstsQIIessXQlIolle
ItionoXIoEetarTandinterstellarEaterialanddraZinSIheersXroEnasasIientistsOoeMelli
nSaandseMeraldoYenIolleaSQesSatheredanNioQslTTesterdaTataEilitarTtestinSranSeinQ
tahZatIhinSasthePrFroFethesiYeoXalarSePlantPothQrtledintotheatEosPhereXasterthanan
TEanEadeoFOeIthaddoneFeXoreasIoEetsarelarSeFallsoXdirtTiIeleXtoMerXroEtheIreatio
noXthesolarsTsteEFillionTearsaSosIientistseNPeItthatthesaEPlesZillreMealIlQesaFo
QthoZthePlanetsZereXorEedandPossiFlTtheoriSinsoXliXeErMellinSasnerMoQsdisPositio
nZasnotsolelTFeIaQseoXthehiShsPeedIQlEinationoXtheEissionXorZhiIhheisProSraEEeEa
naSerXorloIDheedEartinsPaIesTsteEstheIoEPanTthatdesiSnedandIonstrQItedtheProFean
ditsEothershiPsiNteenEonthsaSoheZatIhedinhorrorastheSenesisretQrnProFeIarrTinSso
larZindPartiIlesFQiltFTthesaEeteaEQnderhisdireItionandFasedonsiEilarteIhnoloSTsl
aEEedintotheSroQndatEPheMerTFodTisIonXidentinZhatZehaMedesiSnedandFQiltandtested
hesaidFQteMerTFodTalZaTsZorriesaFoQtZhatZehaMenotthoQShtoXstardQstZaslaQnIhedinX
eFrQarTitenIoQnteredtheZildIoEettZoTearsaSoandhoPeXQllTSatheredd QstXroEitstailQs
inSaPaddleshaPedliDeatennisraIAQetaeroSelaIoloQrlessSaseoQsOellTthatistheliShtes
tEaterialonearthandisDnoZntosIientistsasXroYensEoDeZasQsedintheatteEPttotraPtheE
aterialdrEattSenSeXroEiEPerialIolleSelondonaleadinSeNPertoneNtraterrestrialdQstZ
illFeoneoXtheXirstFritishsIientiststoreIeiMedQstsaEPlestoanalTsethisistheXirstti
EesinIetheaPolloEissionsthatsaEPlesoXroIDhaMeFeenretQrnedXroEsPaIetoearthhesaidt
histhoQsandthoXaSraEEeoXdQstXroEZildZillProFaFlTtellQsEoreaFoQttheXorEationoXthe
solarsTsteEthanthePastTearsoXtelesIoPeoF
```

Consider the second line. I have inserted a little space at some likely word breaks.

Ehas landedons Ihed Qle in the Qnited states at the endo Xasi NTear Eission the stard Qst ret Qrn

u → Q

Ehas landedons Ihedule in the united states at the endo Xasi NTear Eission the stardust return

c → I

f → X

m → E

mhas landed on schedule in the united states at the end of asi NTear mission the stardust return

Now let us look at the 11th and 12th lines, with the characters above decoded.

k → D

y → T

g → S

b → F

naSerforlocDheedmartinsPacesTstemsthecomPanTthatdesiSnedandconstructedthePro Fean
ditsmothershiPsiNteenmonthsaSoheZatchedinhorrorastheSenesisreturnProFecarrTinSso

p → P

x → N

w → Z

nager for lockheed martin space systems the company that designed and constructed the probe a
n
dits mother ship sixteen months ago he watched in horror as the genesis return probe carrying s
o

So far we have decrypted 22 of the 26 characters; these lines will yield three more.

ymanmadeobOecthaddonebeforeascometsarelargeballsofdirtyiceleftoMerfromthecreatio

j → O

v → M

z → Y

tahwatchingastheprobethesiYeofalargeplantpothurtledintotheatmospherefasterthanan

The final plain text character **q** must encrypt to the final ciphertext character **A**.
So the encryption key is the permutation

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

and our plaintext was

```
acapsulecarryingdustfromacometwhichcouldshedlightontheoriginsofourplanetarysyste
mhaslandedonscheduleintheunitedstatesattheendofasixyearmissionthestardustreturnc
apsuleslowedfrommphtofeetpersecondasitreturnedatammarkingthefirstsuccessfulcolle
ctionofcometaryandinterstellarmaterialanddrawingcheersfromnasascientistsjoevelli
ngaandseveraldozencolleaguesgatheredanxiouslyyesterdayatamilitarytestingrangeinu
tahwatchingastheprobethesizeofalargeplantpothurtledintotheatmospherefasterthanan
ymanmadeobjecthaddonebeforeascometsarelargeballsofdirtyiceleftoverfromthecreatio
nofthesolarsystembillionyearsagoscientistsexpectthatthesampleswillrevealcluesabo
uthowtheplanetswereformedandpossiblytheoriginsoflifemrvellingasnervousdispositio
nwasnotsolelybecauseofthehighspeedculminationofthemissionforwhichheisprogrammema
nagerforlockheedmartinspacesystemsthecompanythatdesignedandconstructedtheprobean
ditsmothershipsixteenmonthsagohewatchedinhorrorasthegenesisreturnprobecarryingso
larwindparticlesbuiltbythesameteamunderhisdirectionandbasedonsimilartechnologysl
ammedintothegroundatmpheverybodyisconfidentinwhatwehavedesignedandbuiltandtested
hesaidbuteverybodyalwaysworriesaboutwhatwehavenotthoughtofstardustwaslaunchedinf
ebruaryitencounteredthewildcomettwoyearsagoandhopefullygathereddustfromitstailus
ingapaddleshapedlikeatennisracquetaerogelacolourlessgaseousjellythatisthelightes
tmaterialonearthandisknowntoscientistsasfrozensmokewasusedintheattempttotrapthem
aterialdrmattgengefromimperialcollegelondonaleadingexpertonextraterrestrialdustw
illbeoneofthefirstbritishscientiststoreceivedustsamplestoanalysethisisthefirstti
mesincetheapollomissionsthatsamplesofrockhavebeenreturnedfromspacetoearthhesaidt
histhousandthofagrammeofdustfromwildwillprobablytellusmoreabouttheformationofthe
solarsystemthanthepastyearsoftelescopeob
```

It is interesting to compare the frequencies and ranks of the ciphertext letters (and of the actual plaintext letters) with those predicted from our “typical” English language text data.

Letter		Frequency of Letter		Rank of Letter	
typical plain text	cipher text	typical plain text	cipher text	typical plain text	cipher text
e	**H**	1237	1253	1	1
t	**U**	921	1013	2	2
a	**R**	821	845	3	3
o	**V**	767	743	4	5
n	**L**	705	580	5	8
i	**C**	676	642	6	6
h	**G**	645	479	7	10
s	**K**	617	804	8	4
r	**B**	550	616	9	7
d	**W**	479	397	10	11
l	**J**	393	484	11	9
u	**Q**	291	260	12	14
w	**Z**	254	137	13	20
m	**E**	254	316	14	12
c	**I**	230	280	15	13
f	**X**	225	229	16	15
g	**S**	208	204	17	16
y	**T**	195	178	18	18
p	**P**	163	199	19	17
b	**F**	150	153	20	19
k	**D**	87	31	21	22
v	**M**	87	87	22	21
j	**O**	18	20	23	24
x	**N**	13	31	24	22
q	**A**	9	5	25	26
z	**Y**	6	15	26	25