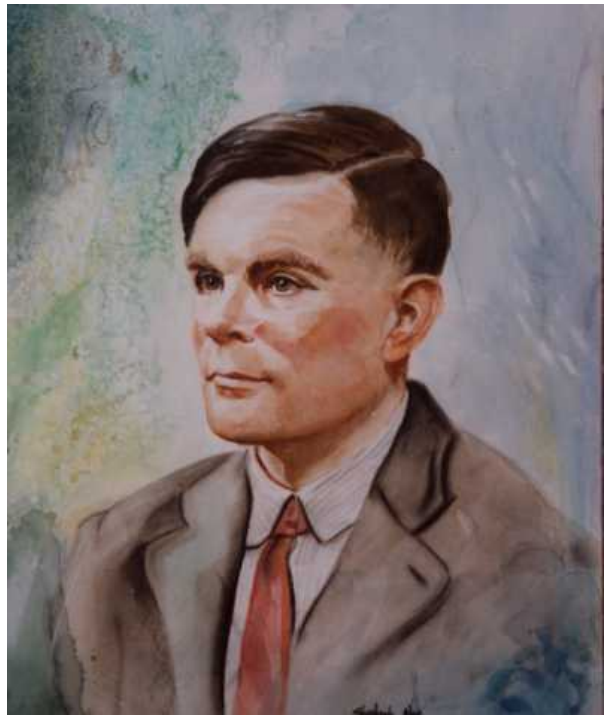


Code Makers and Code Breakers From Julius Caesar To Alan Turing And Beyond III

An Introduction To Some Mathematics and Linguistics Used to Crack Secret Ciphers

Monoalphabetic Substitution

Need some way to scramble the alphabet

Keyword Method

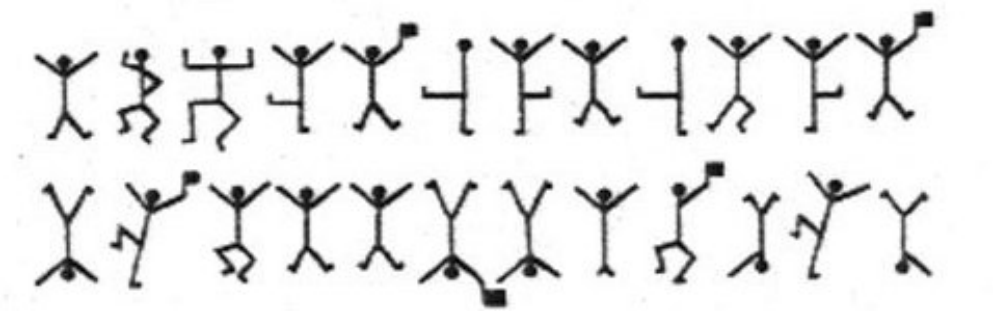
Keyword: **VERMONT**

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

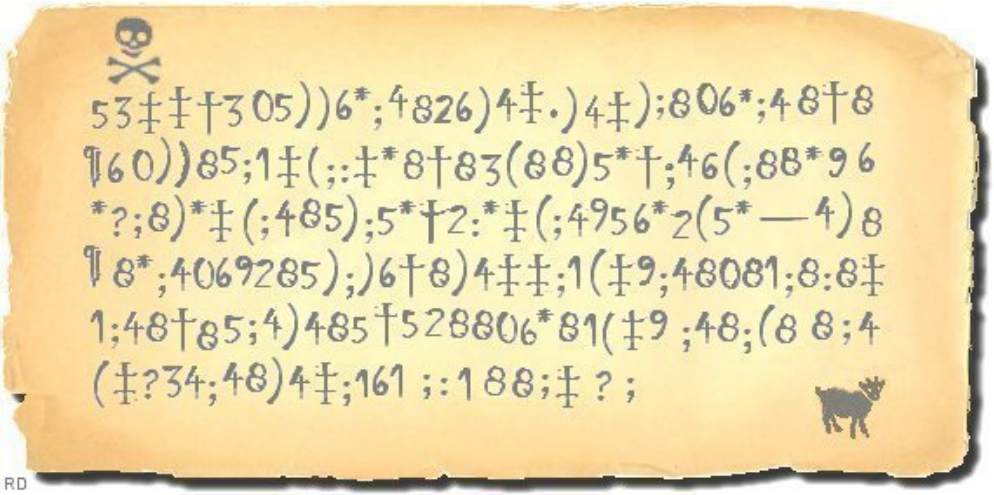
ABCDEFGHIJKLMNOPQRSTUVWXYZ

VAIUEBJWRCKXMDLYOFPZNGQTHS

Two Famous Monoalphabetic Ciphers



Sherlock Holmes, "The Adventure of the Dancing Men" (1903)



Edgar Allen Poe, "The Gold Bug" (1843)

How Does One Begin To Crack a Cipher?

Single Most Important Tool:

Frequency Count of Characters

The percentage of occurrence of any particular letter, its **Characteristic Frequency**, will not vary widely from one plaintext message to the next.

HAPPY FAMILIES ARE ALL ALIKE; EVERY UNHAPPY FAMILY IS UNHAPPY IN ITS OWN WAY. EVERYTHING WAS IN CONFUSION IN THE OBLONSKYS' HOUSE. THE WIFE HAD DISCOVERED THAT THE HUSBAND WAS CARRYING ON AN INTRIGUE WITH A FRENCH GIRL, WHO HAD BEEN A GOVERNESS IN THEIR FAMILY, AND SHE HAD ANNOUNCED TO HER HUSBAND THAT SHE COULD NOT GO ON LIVING IN THE SAME HOUSE WITH HIM. THIS POSITION OF AFFAIRS HAD NOW LASTED THREE DAYS, AND NOT ONLY THE HUSBAND AND WIFE THEMSELVES, BUT ALL THE MEMBERS OF THEIR FAMILY AND HOUSEHOLD, WERE PAINFULLY CONSCIOUS OF IT. EVERY PERSON IN THE HOUSE FELT THAT THERE WAS SO SENSE IN THEIR LIVING TOGETHER, AND THAT THE STRAY PEOPLE BROUGHT TOGETHER BY CHANCE IN ANY INN HAD MORE IN COMMON WITH ONE ANOTHER THAN THEY, THE MEMBERS OF THE FAMILY AND HOUSEHOLD OF THE OBLONSKYS. THE WIFE DID NOT LEAVE HER OWN ROOM, THE HUSBAND HAD NOT BEEN AT HOME FOR THREE DAYS. THE CHILDREN RAN WILD ALL OVER THE HOUSE; THE ENGLISH GOVERNESS QUARRELED WITH THE HOUSEKEEPER, AND WROTE TO A FRIEND ASKING HER TO LOOK OUT FOR A NEW SITUATION FOR HER; THE MAN-COOK HAD WALKED OF THE DAY BEFORE JUST AT DINNER-TIME; THE KITCHEN-MAID, AND THE COACHMAN HAD GIVEN WARNING. THREE DAYS AFTER THE QUARREL, PRINCE STEPAN ARKADYEVITCH OBLONSKY–STIVA, AS HE WAS CALLED IN THE FASHIONABLE WORLD–WOKE UP AT HIS USUAL HOUR, THAT IS, AT EIGHT O'CLOCK IN THE MORNING, NOT IN HIS WIFE'S BEDROOM, BUT ON THE LEATHER-COVERED SOFA IN HIS STUDY. HE TURNED OVER HIS STOUT, WELL-CARED-FOR PERSON ON THE SPRINGY SOFA, AS THOUGH HE WOULD SINK INTO A LONG SLEEP AGAIN; HE VIGOROUSLY EMBRACED THE PILLOW ON THE OTHER SIDE AND BURIED HIS FACE IN IT; BUT ALL AT ONCE HE JUMPED UP, SAT UP ON THE SOFA, AND OPENED HIS EYES. "YES, YES

Plaintext Example
Opening Chapter of *Anna Karenina*

A	331	E	515
B	57	T	380
C	83	H	332
D	200	A	331
E	515	I	324
F	107	N	308
G	94	O	307
H	332	S	281
I	324	R	219
J	2	D	200
K	27	L	163
L	163	U	117
M	88	F	107
N	308	W	100
O	307	G	94
P	75	M	88
Q	4	Y	85
R	219	C	83
S	281	P	75
T	380	B	57
U	117	V	45
V	45	K	27
W	100	Q	4
X	4	X	4
Y	85	J	2
Z	0	Z	0

Characters Counted: 4248

Characteristic Frequencies

Why Do They Exist?

Why don't all letters appear equally often?

How Can We Exploit Them?

Redundancy Example

Pl*t Su*m*ry f*r St*r W*rs: Epis*de III- R*v*g* f te Sith

Thr** y**rs *ft*r th* *n*t**l b*ttl* of th* Clon* W*rs, th* long *nd t*r*ng confl*ct b*tw**n th* S*p*r*t*sts *nd th* R*publ*c *s n**r*ng *ts *nd. How*v*r, Ob*-W*n K*nob*, now * J*d* M*st*r, *nd *n*k*n Skyw*lk*r, now * J*d* Kn*ght , l**rn * shock*ng d*v*lopm*nt. Corusc*nt *s und*r *tt*ck! Th*y hurry b*ck to Corusc*nt to count*r th* *tt*ck l*d by G*n*r*l Gr**vous, th* l**d*r of th* S*p*r*t*st dro*d *rm**s.

Plot Summary for Star Wars: Episode III- Revenge of the Sith

Three years after the initial battle of the Clone Wars, the long and tiring conflict between the Separatists and the Republic is nearing its end. However, Obi-Wan Kenobi, now a Jedi Master, and Anakin Skywalker, now a Jedi Knight , learn a shocking development. Coruscant is under attack! They hurry back to Coruscant to counter the attack led by General Grievous, the leader of the Separatist droid armies.

Redundancy Example II

The New York Times

Th* f*d*r*l gov*rnm*nt on Mo*d*y *nno*nc*d * n*w s*t of mon*tor*ng gu*d*l*n*s for p*opl* *rr*v*ng from W*st *fr*c* th*t st*pp*d sh*rt of th* to*gh m**sur*s *nst*tut*d *n N*w Y*rk *nd N*w J*rs*y l*st w**k, *n *ffort to br*ng un*form*ty to * m*ssy p*tchw*rk of r*spons*s by st*t*s.

Th* n*w pol*cy, wh*ch f*d*r*l h**lth off*c**ls s**d w*s *n *ffort to str*k* * b*l*nce b*tw**n s*f*ty *nd c*v*l l*b*rt**s, would r*qu*r* r*turn*ng h**th c*r* work*rs, or p*opl* who h*d b**n n**r *bol p*t* o subm*t to *n *np*rson ch*ckup *nd * phon* c*ll from * loc*l publ*c h**lth *uthor*ty.

Redundancy Example II

The New York Times

The federal government on Monday announced a new set of monitoring guidelines for people arriving from West Africa that stopped short of the tough measures instituted in New York and New Jersey last week, an effort to bring uniformity to a messy patchwork of responses by states.

The new policy, which federal health officials said was an effort to strike a balance between safety and civil liberties, would require returning heath care workers, or people who had been near Ebola patients, to submit to an in-person checkup and a phone call from a local public health authority.

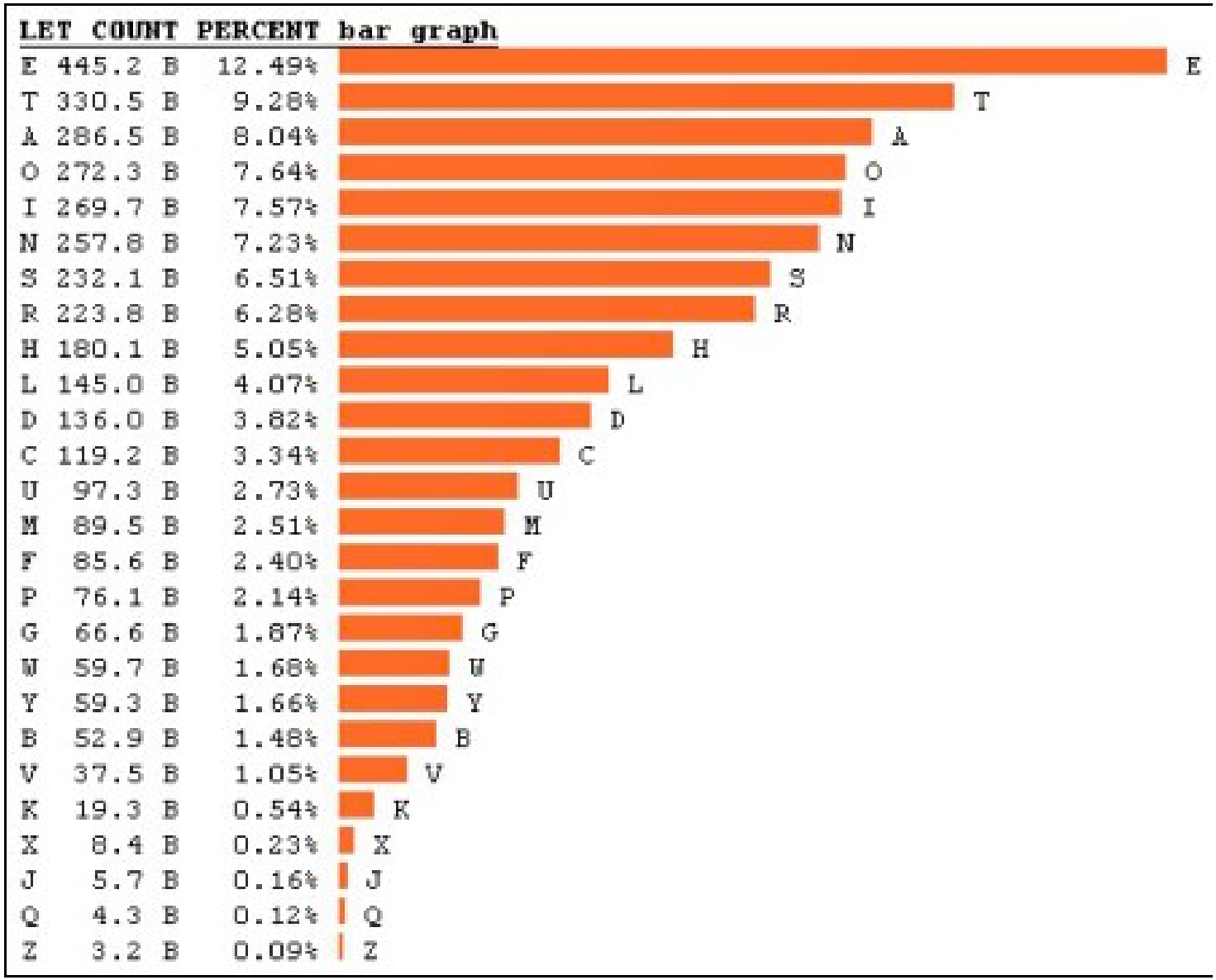
ENGLISH LANGUAGE DATA

E	1231	L	403	B	162
T	959	D	365	G	161
A	805	C	320	V	93
O	794	U	310	K	52
N	719	P	229	Q	20
I	718	F	228	X	20
S	659	M	225	J	10
R	603	W	203	Z	9
H	514	Y	188		

SAMPLE CIPHER MESSAGE

H	Y	U	D	P	Q	G	V	O	R	X	Z	S
23	19	17	16	16	16	13	13	12	12	12	12	10
F	E	K	L	M	N	J	T	B	I	A	C	W
10	9	9	9	9	9	8	8	4	4	3	3	0

English Frequency and Sequence Data
3,563,505,777,820 letters



Solving a Cryptogram Using Frequency Information

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

is part of a longer ciphertext which has the following frequency distribution of letters:

A	89		N	367		S	523		Z	112
B	229		O	89		V	406		W	111
C	147		P	389		P	389		X	105
D	331		Q	58		N	367		A	89
E	7		R	32		I	339		O	89
F	72		S	523		D	331		F	72
G	158		T	131		H	308		Q	58
H	308		U	31		J	306		R	32
I	339		V	406		B	229		U	31
J	306		W	111		L	172		M	8
K	5		X	105		G	158		E	7
L	172		Y	4		C	147		K	5
M	8		Z	112		T	131		Y	4

E	The Highest Frequency
T A O I N S H R	High Frequency Group
D L	Medium Frequency
C UM W F G Y P B	Low Frequency
V K J X Q Z	Rare

Standard Written English: Characteristic Frequencies and Sequence Data

Order and Frequency of Single Letters:

E 12311	L 403	B 162
T 959	D 365	G 161
A 805	C 320	V 93
O 794	U 310	K 52
N 719	P 229	Q 20
I 718	F 228	X 20
S 659	M 225	J 10
R 603	W 203	Z 9
H 514	Y 188	

Group Percentages

AEIOU	38.58%
LNRST	33.43%
JKQXZ	1.11%
ETAON	45.08%
ETAONISRH	70%

Order and Frequency of Leading Digrams

TH 315	TO 111	SA 75	MA 56
HE 251	NT 110	HI 72	TA 56
AN 172	ED 107	LE 72	CE 55
IN 169	IS 106	SO 71	IC 55
ER 154	AR 101	AS 67	LL 55
RE 148	OU 96	NO 65	NA 54
ES 145	TE 94	NE 64	RO 54
ON 145	OF 94	EC 64	OT 53
EA 131	IT 88	IO 63	TT 53
TI 128	HA 84	RT 63	VE 53
AT 125	SE 84	CO 59	NS 51
ST 121	ET 80	BE 58	UR 49
EN 120	AL 77	DI 57	ME 49
OR 113	NG 75	RA 57	LY 47

List of Common Reversals

ER	RE	ON	NO	TE	ET	ST	TS
ES	SE	IN	NI	OR	RO	IS	SI
AN	NA	EN	NE	TO	OT	ED	DE
TI	IT	AT	TA	AR	RA	OF	FO

Order of Leading Trigrams in 10,000 letters of text:

THE	THA	ION	FOR	HAS	EDT	OFT	MEN
AND	ENT	TIO	NDE	NCE	TIS	STH	

Order of Letters as *Initial* Letters of Words:
T A S O I C W P B F H M R E N L G U Y V J K Q X Z

Order of Letters as *Final* Letters of Words:
E S D N T R Y O F L A G H M W K C P I X U B V J Z Q

Solving a Cryptogram Using Frequency Information

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

is part of a longer ciphertext which has the following frequency distribution of letters:

A	89		N	367		S	523		Z	112
B	229		O	89		V	406		W	111
C	147		P	389		P	389		X	105
D	331		Q	58		N	367		A	89
E	7		R	32		I	339		O	89
F	72		S	523		D	331		F	72
G	158		T	131		H	308		Q	58
H	308		U	31		J	306		R	32
I	339		V	406		B	229		U	31
J	306		W	111		L	172		M	8
K	5		X	105		G	158		E	7
L	172		Y	4		C	147		K	5
M	8		Z	112		T	131		Y	4

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

(replace S by e)

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

(replace S by e)

PBe FHIIeD NZ PBe IeF BVCWJBHDe WDHCVDQ BVJ V ONNL GBVIGe NZ AeGNCHIO PBe IeKP WDeJHLeIP NZ PBe XIHPeL JPVPeJ.

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

(replace S by e)

PBe FHIIeD NZ PBe IeF BVCWJBHDe WDHCVDQ BVJ V ONNL GBVIGe NZ AeGNCHIO PBe IeKP WDeJHLeIP NZ PBe XIHPeL JPVPeJ.

(replace P by t)

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

(replace S by e)

PBe FHIIeD NZ PBe IeF BVCWJBHDe WDHCVDQ BVJ V ONNL GBVIGe NZ AeGNCHIO PBe IeKP WDeJHLeIP NZ PBe XIHPeL JPVPeJ.

(replace P by t)

tBe FHIIeD NZ tBe IeF BVCWJBHDe WDHCVDQ BVJ V ONNL GBVIGNZ AeGNCHIO tBe IeKt WDeJHLeIt NZ tBe XIHteL JtVteJ

PBS FHIISD NZ PBS ISF BVCWJBHDS WDHCVDQ BVJ V ONNL GBVIGS NZ ASGNCHIO PBS ISKP WDSJHLSIP NZ PBS XIHPSL JPVPSJ.

(replace S by e)

PBe FHIIeD NZ PBe IeF BVCWJBHDe WDHCVDQ BVJ V ONNL GBVIGe NZ AeGNCHIO PBe IeKP WDeJHLeIP NZ PBe XIHPeL JPVPeJ.

(replace P by t)

tBe FHIIeD NZ tBe IeF BVCWJBHDe WDHCVDQ BVJ V ONNL GBVIGNZ AeGNCHIO tBe IeKt WDeJHLeIt NZ tBe XIHteL JtVteJ

(replace B by h)

the FHIIeD NZ the IeF hVCWJhHDe WDHCVDQ hVJ V ONNL GhVIGNZ AeGNCHIO the IeKt WDeJHLeIt NZ the XIHteL JtVteJ.

the FHIIeD NZ the IeF hVCWJhHDe WDHCVDQ hVJ V ONNL GhVIGNZ AeGNCHIO the IeKt WDeJHLeIt NZ the XIHteL JtVteJ.

(replace J by s)

the FHIIeD NZ the IeF hVCWJhHDe WDHCVDQ hVJ V ONNL GhVIGNZ AeGNCHIO the IeKt WDeJHLeIt NZ the XIHteL JtVteJ.

(replace J by s)

the FHIIeD NZ the IeF hVCWshHDe WDHCVDQ hVs V ONNL GhVIGNZ AeGNCHIO the IeKt WDesHLeIt NZ the XIHteL stVtes.

the FHIIeD NZ the IeF hVCWJhHDe WDHCVDQ hVJ V ONNL GhVIGNZ AeGNCHIO the IeKt WDeJHLeIt NZ the XIHteL JtVteJ.

(replace J by s)

the FHIIeD NZ the IeF hVCWshHDe WDHCVDQ hVs V ONNL GhVIGNZ AeGNCHIO the IeKt WDesHLeIt NZ the XIHteL stVtes.

(replace V by a)

the FHIIeD NZ the IeF hVCWJhHDe WDHCVDQ hVJ V ONNL GhVIGNZ AeGNCHIO the IeKt WDeJHLeIt NZ the XIHteL JtVteJ.

(replace J by s)

the FHIIeD NZ the IeF hVCWshHDe WDHCVDQ hVs V ONNL GhVIGNZ AeGNCHIO the IeKt WDesHLeIt NZ the XIHteL stVtes.

(replace V by a)

the FHIIeD NZ the IeF haCWshHDe WDHCaDQ has a ONNL GhaIGNZ AeGNCHIO the IeKt WDesHLeIt NZ the XIHteL states.

replace N with o

the FHIIeD oZ the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe oZ AeGoCHIO the IeKt WDesHLeIt oZ the XIHteL states.

replace N with o

the FHIIeD oZ the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe oZ AeGoCHIO the IeKt WDesHLeIt oZ the XIHteL states.

(replace Z with f)

replace N with o

the FHIIeD oZ the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe oZ AeGoCHIO the IeKt WDesHLeIt oZ the XIHteL states.

(replace Z with f)

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe of AeGoCHIO the IeKt WDesHLeIt of the XIHteL states

replace N with o

the FHIIeD oZ the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe oZ AeGoCHIO the IeKt WDesHLeIt oZ the XIHteL states.

(replace Z with f)

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe of AeGoCHIO the IeKt WDesHLeIt of the XIHteL states

replace N with o

the FHIIeD oZ the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe oZ AeGoCHIO the IeKt WDesHLeIt oZ the XIHteL states.

(replace Z with f)

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL GhaIGe of AeGoCHIO the IeKt WDesHLeIt of the XIHteL states

(replace G with c)
the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL chaIce of AecoCHIO the IeKt WDesHLeIt of the XIHteL states

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL chaIce of AecoCHIO the IeKt WDesHLeIt of the XIHteL states

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL chaIce of AecoCHIO the IeKt WDesHLeIt of the XIHteL states

(replace I with n)

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL chaIce of AecoCHIO the IeKt WDesHLeIt of the XIHteL states

(replace I with n)
the FHnneD of the neF haCWshHDe WDHCaDQ has a OooL chance of AecoCHnO the neKt WDesHLent of the XnHteL states

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL chaIce of AecoCHIO the IeKt WDesHLeIt of the XIHteL states

(replace I with n)
the FHnneD of the neF haCWshHDe WDHCaDQ has a OooL chance of AecoCHnO the neKt WDesHLent of the XnHteL states

(replace F with w, H with i, D with r)

the FHIIeD of the IeF haCWshHDe WDHCaDQ has a OooL chaIce of AecoCHIO the IeKt WDesHLeIt of the XIHteL states

(replace I with n)
the FHnneD of the neF haCWshHDe WDHCaDQ has a OooL chance of AecoCHnO the neKt WDesHLent of the XnHteL states

(replace F with w, H with i, D with r)

the winner of the new haCWshire WrdCarQ has a OooL chance of AecoCinO the neKt WresiLent of the XniteL states

How can we get rid of the tell tale fingerprints of the characteristic frequencies?