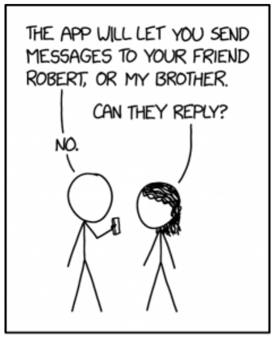
# Day 11

1. probability paradox
2. break Caesar/Vigenere
3. index of coincidence
4. inequality identity




MY NEW SECURE TEXTING APP ONLY ALLOWS PEOPLE NAMED ALICE TO SEND MESSAGES TO PEOPLE NAMED BOB.

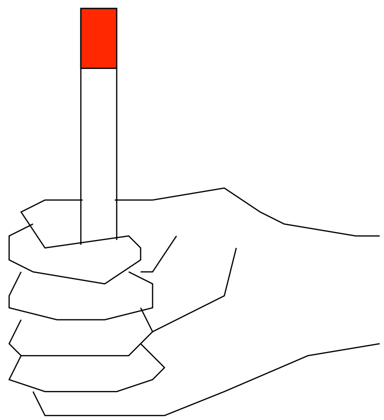
Take three sticks which have their ends colored and place them in a bag. The first stick has two red ends, the second has two black ends and the third stick has a red and a black end.

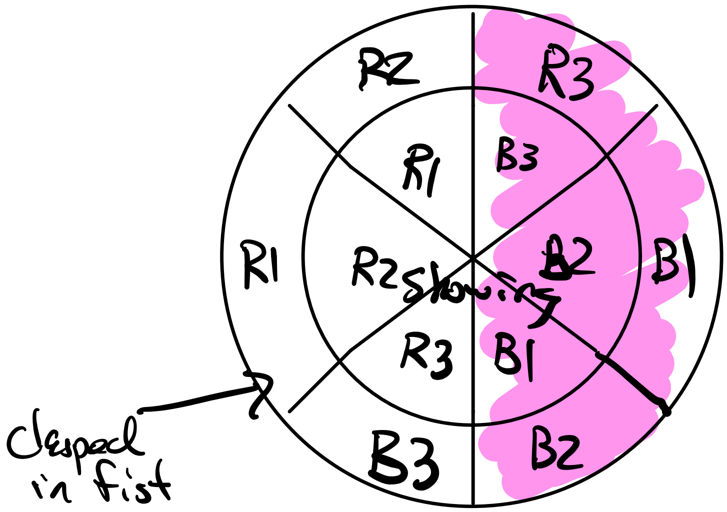
Now, reach into this bag (no peeking) and grasp one of the sticks by an end so that the other end is showing and pull the stick out. Say that a red end is showing.

What color is most likely clasped in your fist?

Is the answer?

1 A) red
2 B) black
7 C) red/black are equally likely
0 D) don't know/care






$$P(\text{clasped}=R \mid \text{showing}=R) = 2/3$$

---

Below this point I'm writing what I had on the whiteboard + one example of the Caesar shift (not all!)

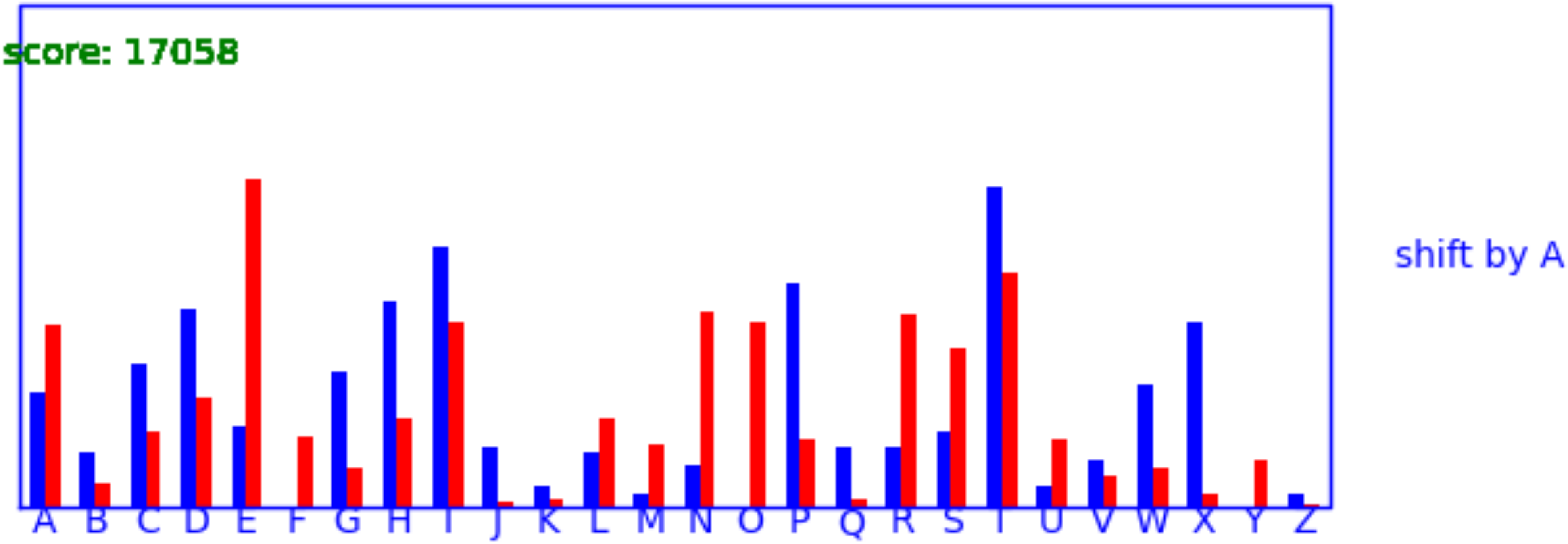
Let $p^{eng} = \langle p_A^{eng}, p_B^{eng}, \ldots, p_Z^{eng} \rangle$ be the vector of English stats.

Let $q^{ctext} = \langle q_A, q_B, \ldots, q_Z \rangle$ be counts of #'s of letters in the cyphertext.

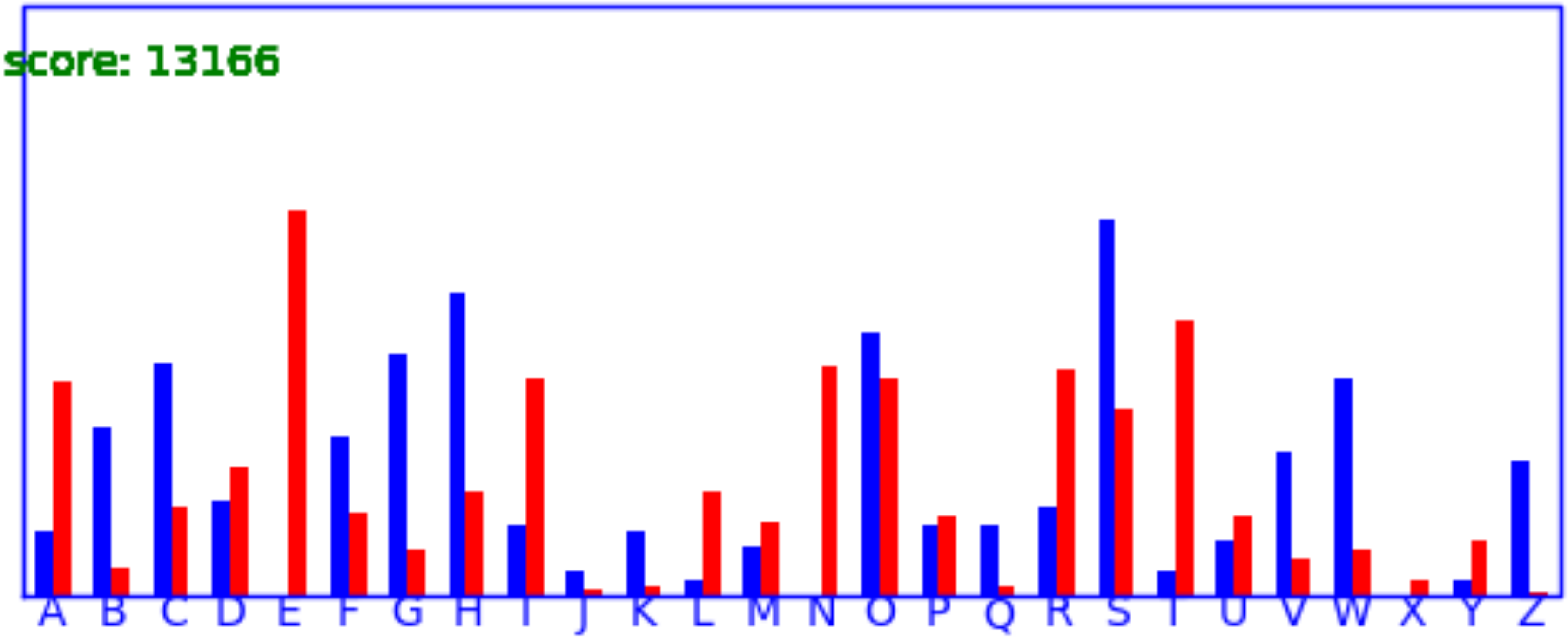
$$p^{eng} \cdot q^{ctext} = \sum_{\alpha=A}^{Z} p_\alpha^{eng} \cdot q_\alpha = |p^{eng}| \cdot |q^{ctext}| \cdot \cos\theta$$

if English $\approx \sum_{\alpha=A}^{Z} (p_\alpha^{eng})^2 N \approx 0.027 \cdot N$

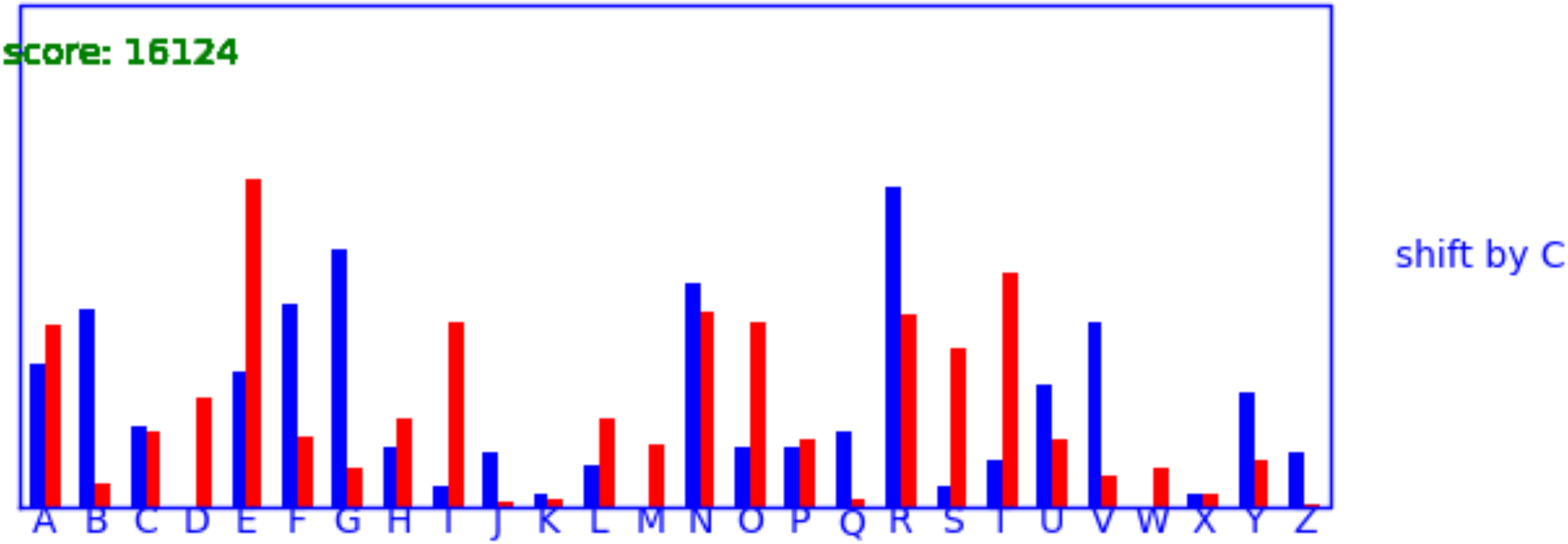
if not.... it should be smaller by factor of "$\cos\theta$"

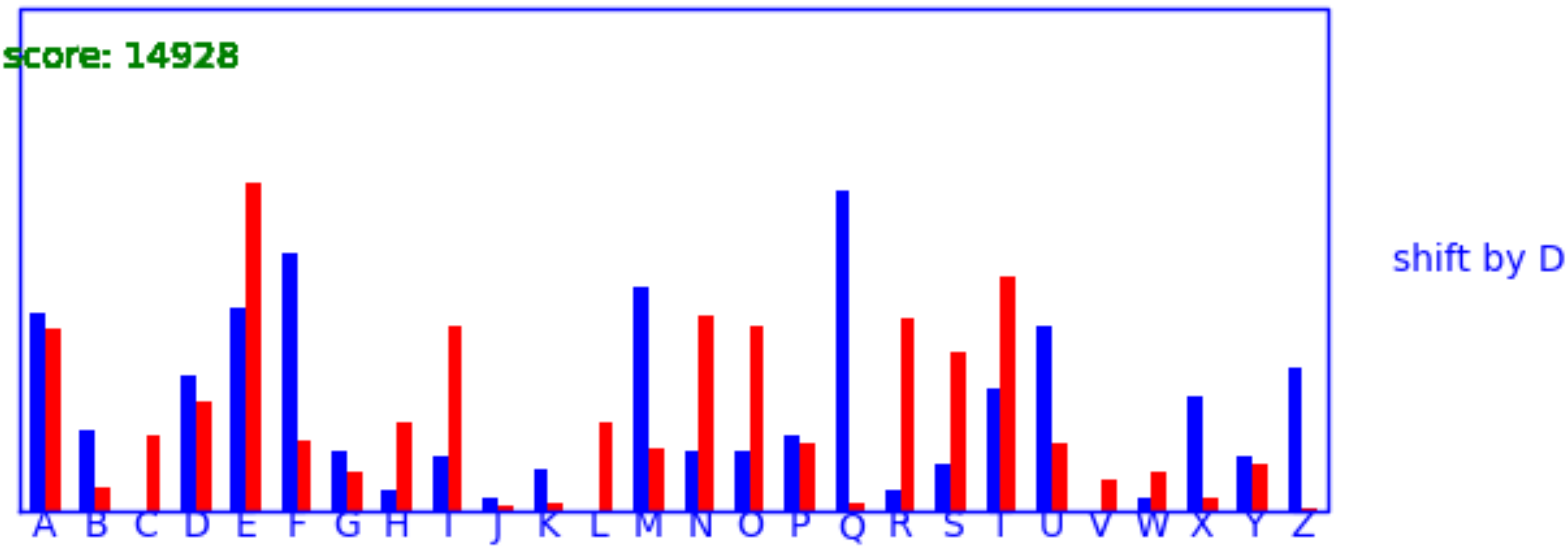


score: 17058
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by A

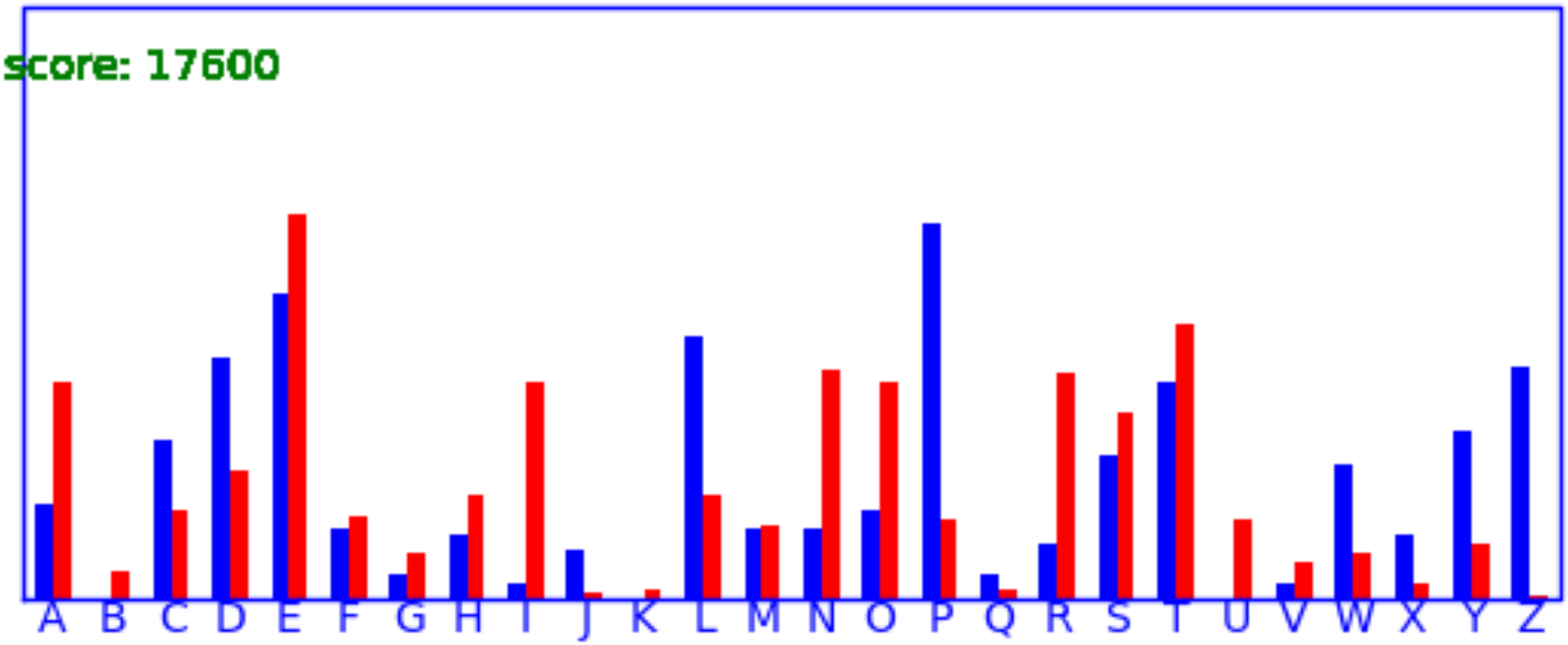



shift by B

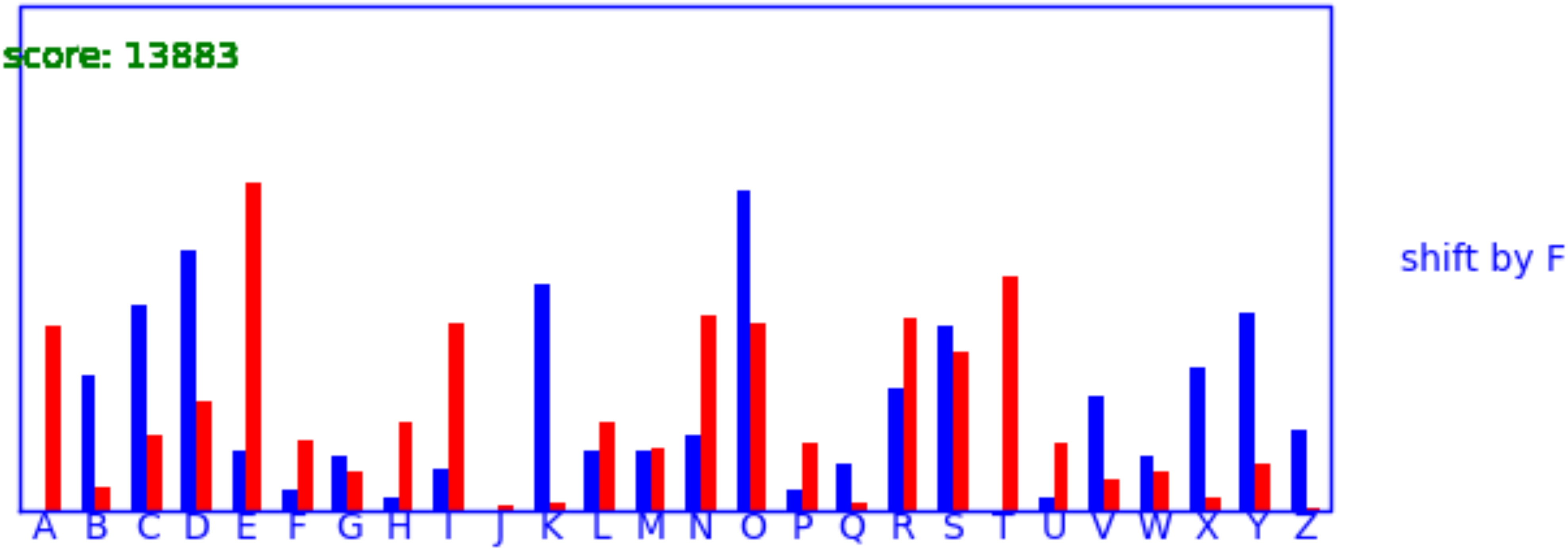

score: 16124
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by C

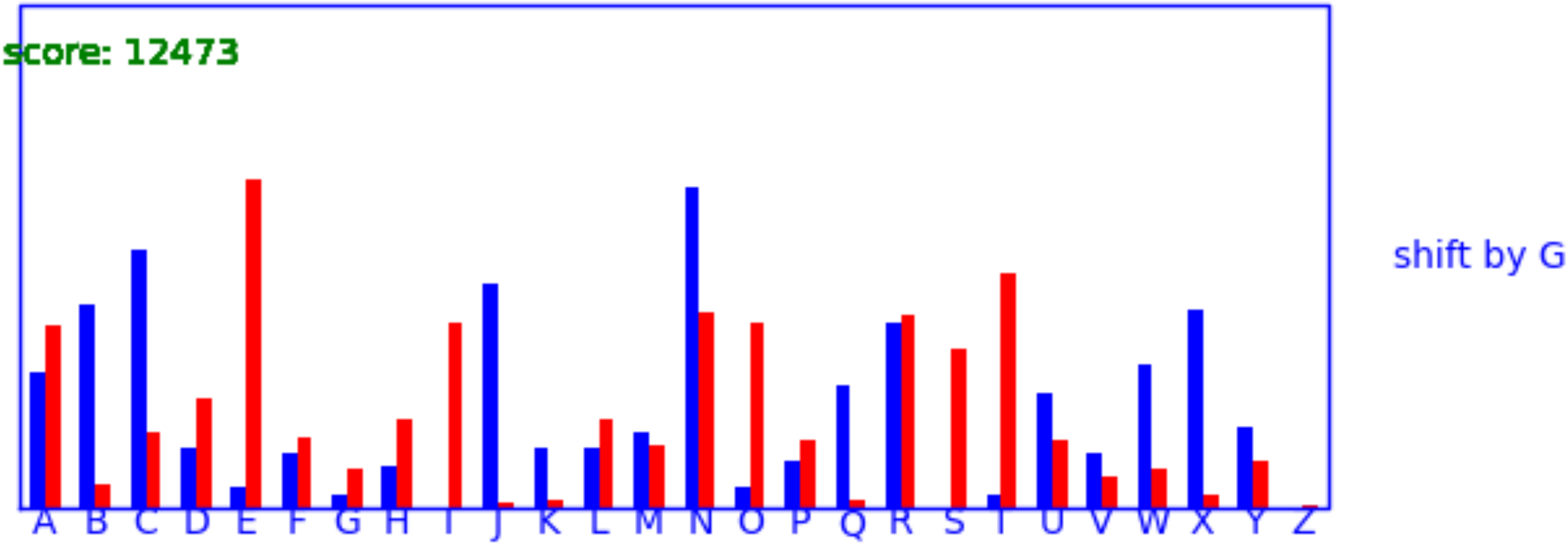


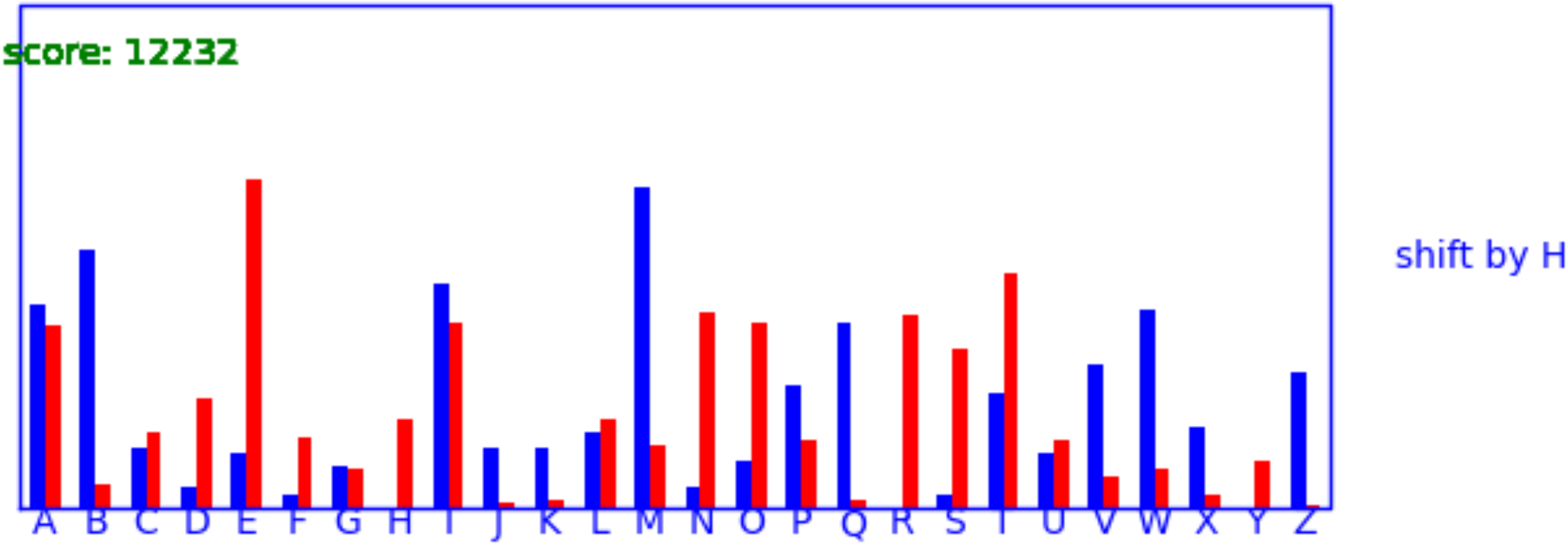
score: 14928
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by D

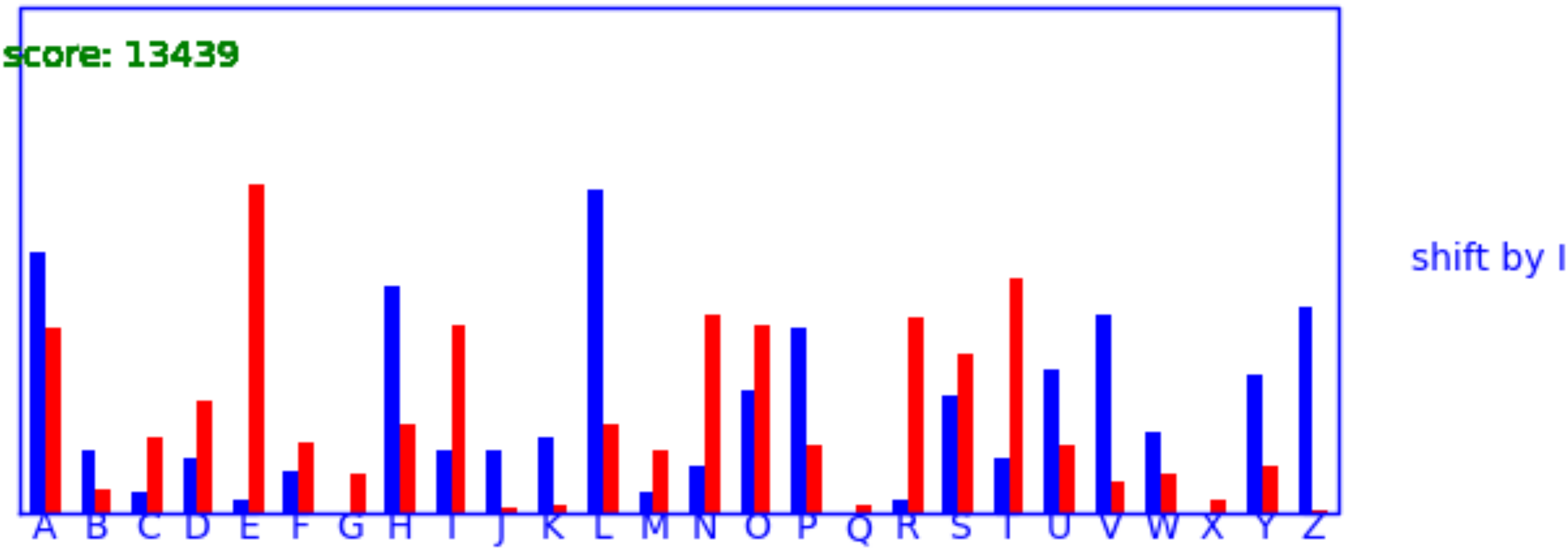


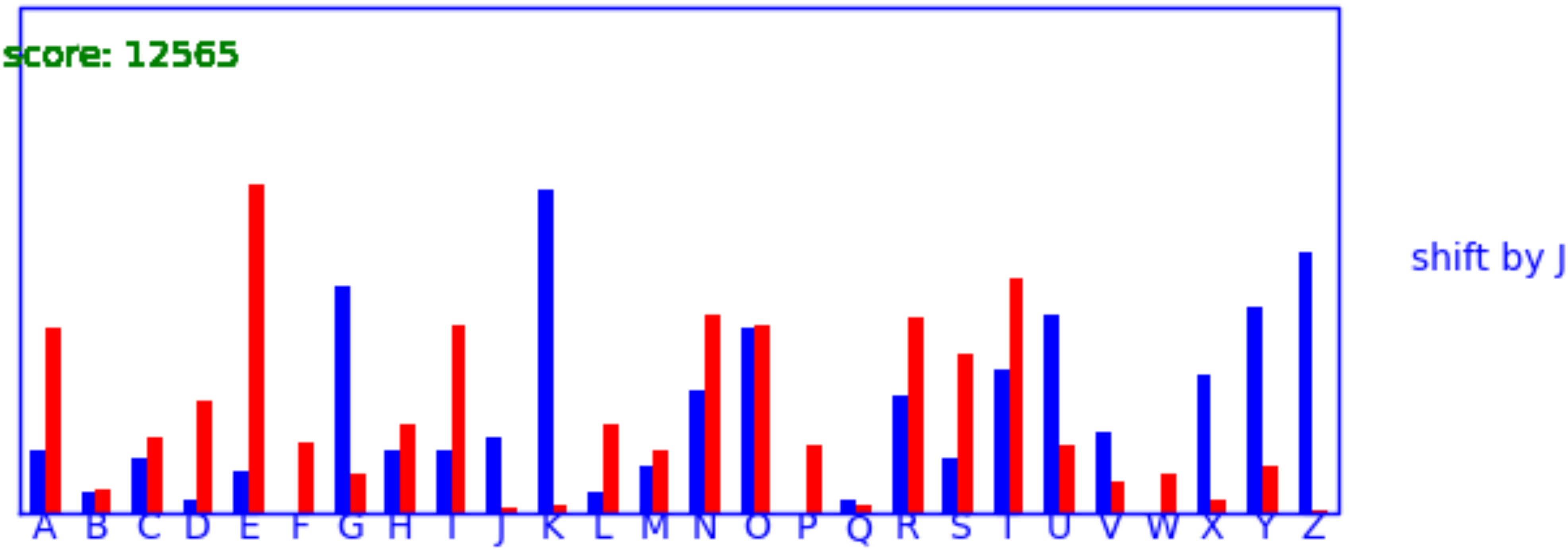

shift by E

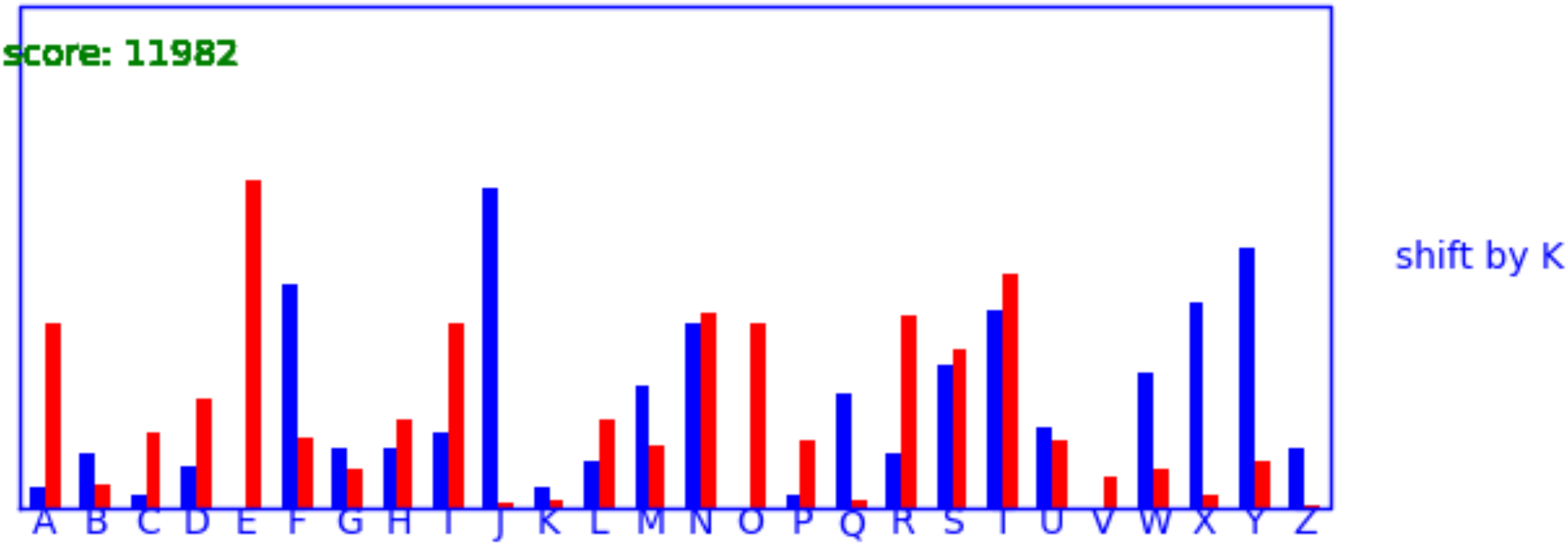

score: 13883
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by F


score: 12473
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
shift by G

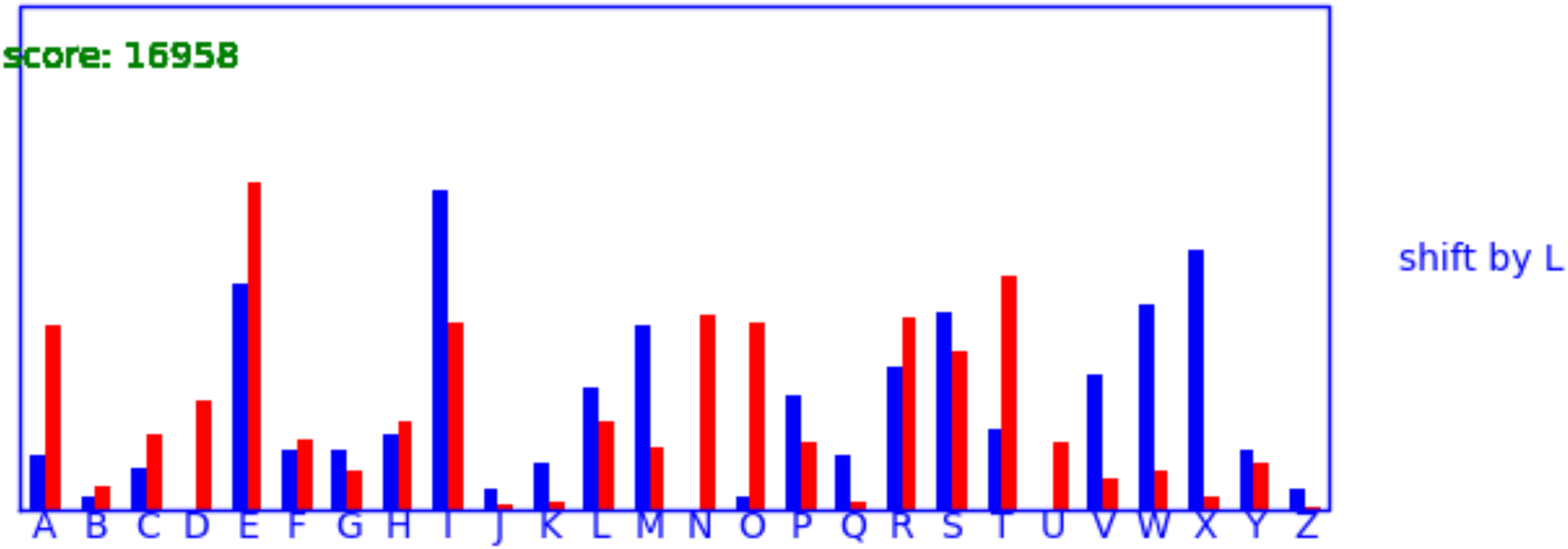

score: 12232
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by H

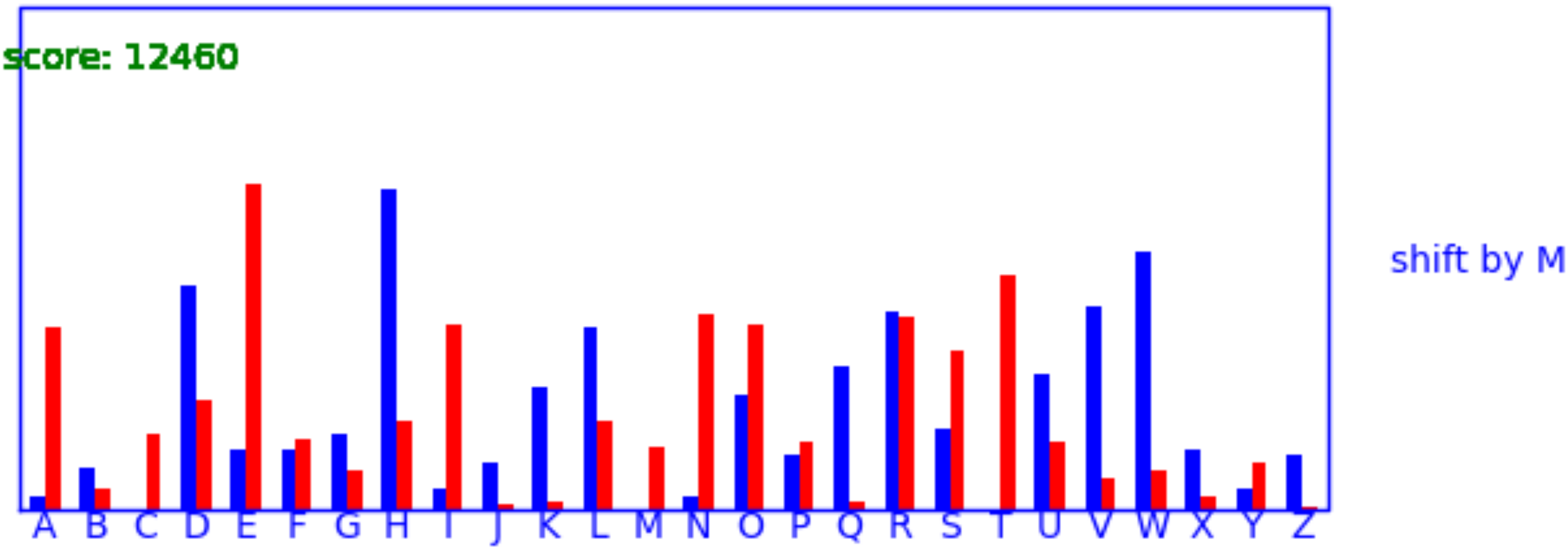

score: 13439
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
shift by I

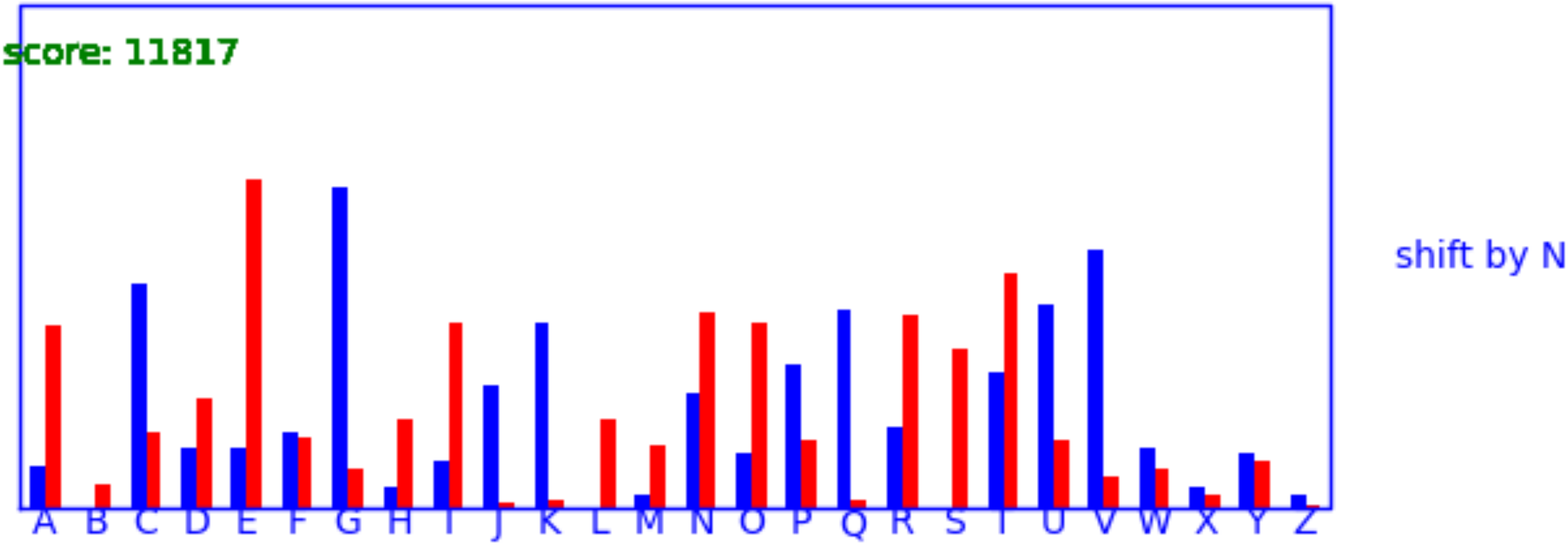

score: 12565
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by J

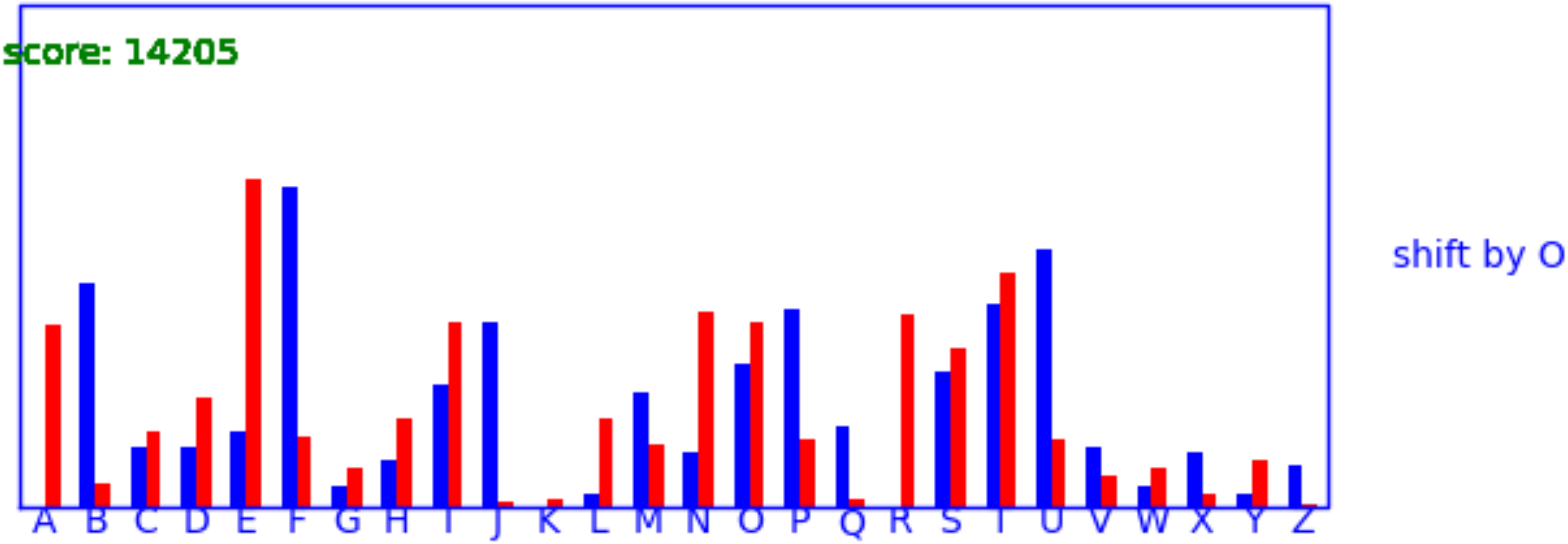


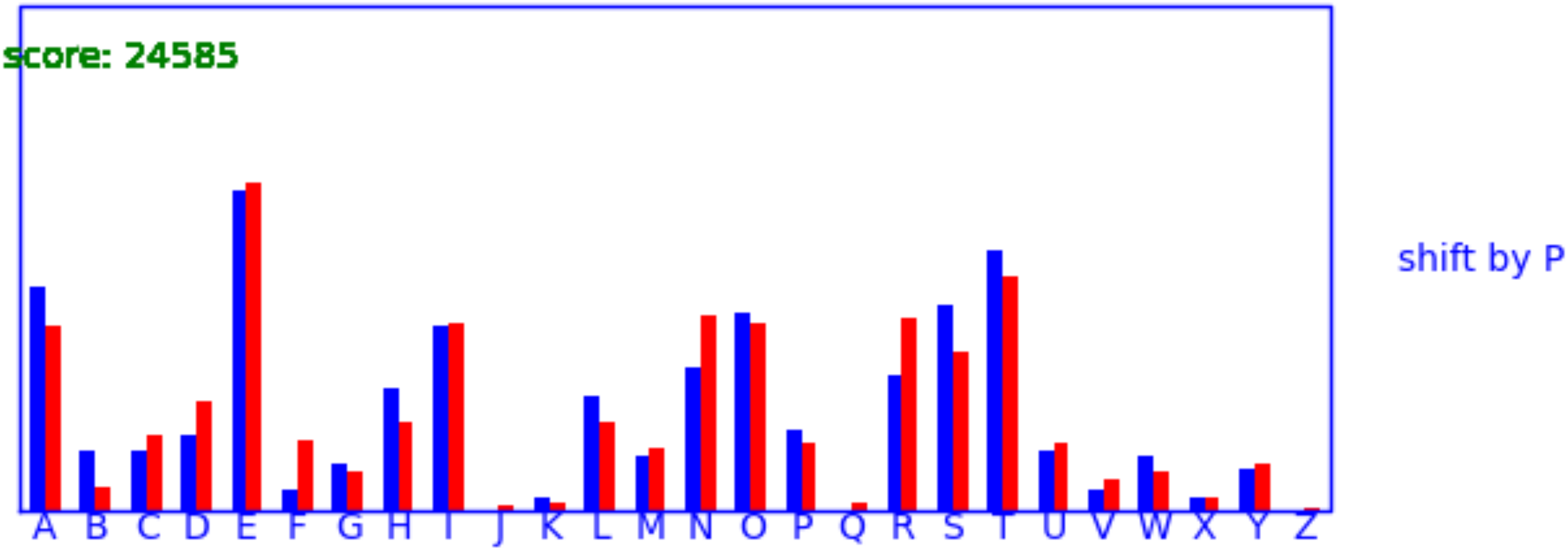
score: 11982
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
shift by K

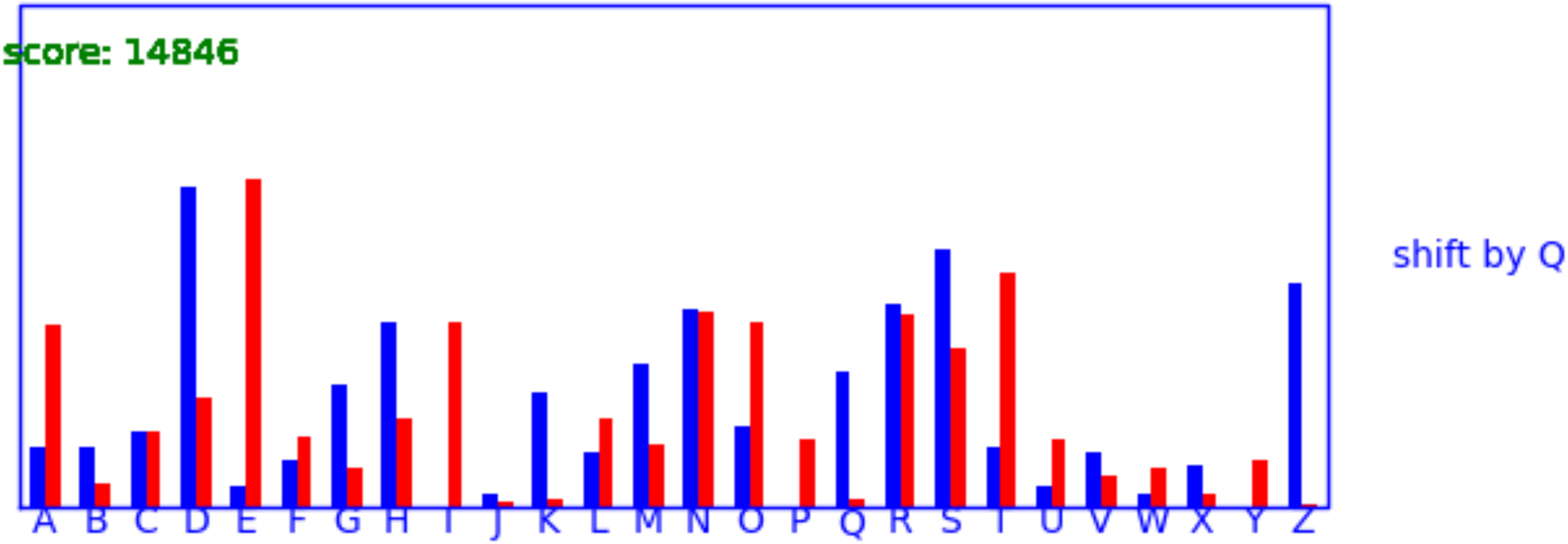

score: 16958
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by L

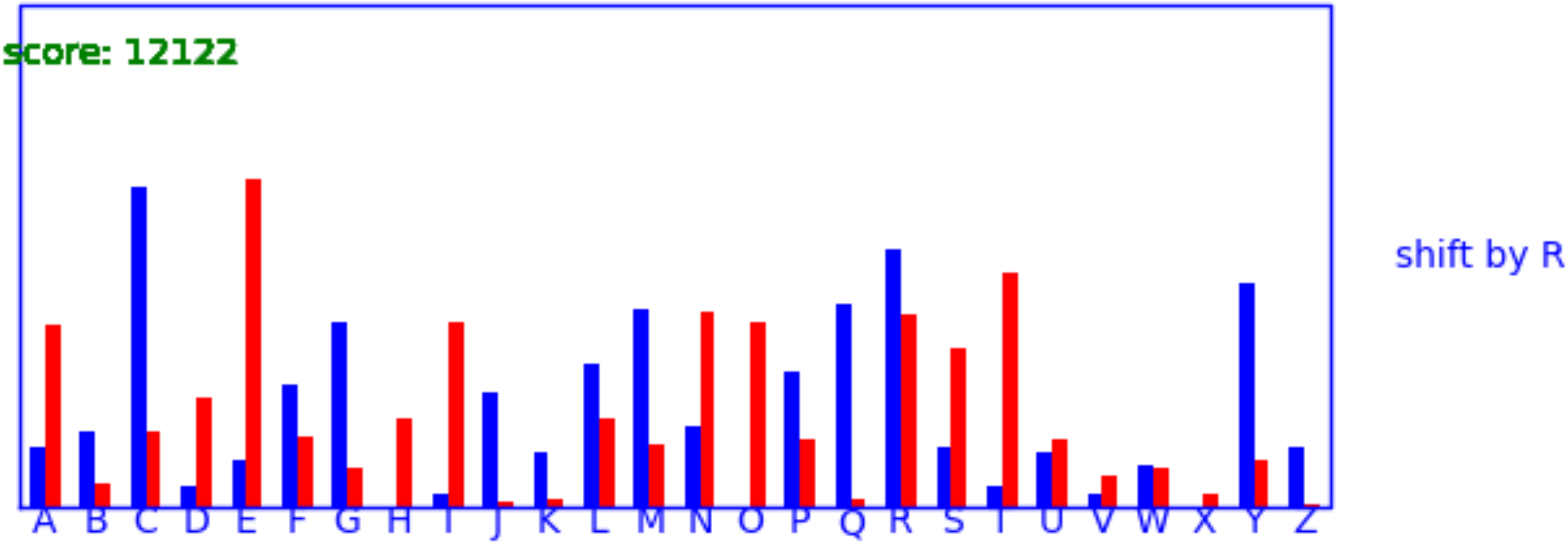

score: 12460
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
shift by M

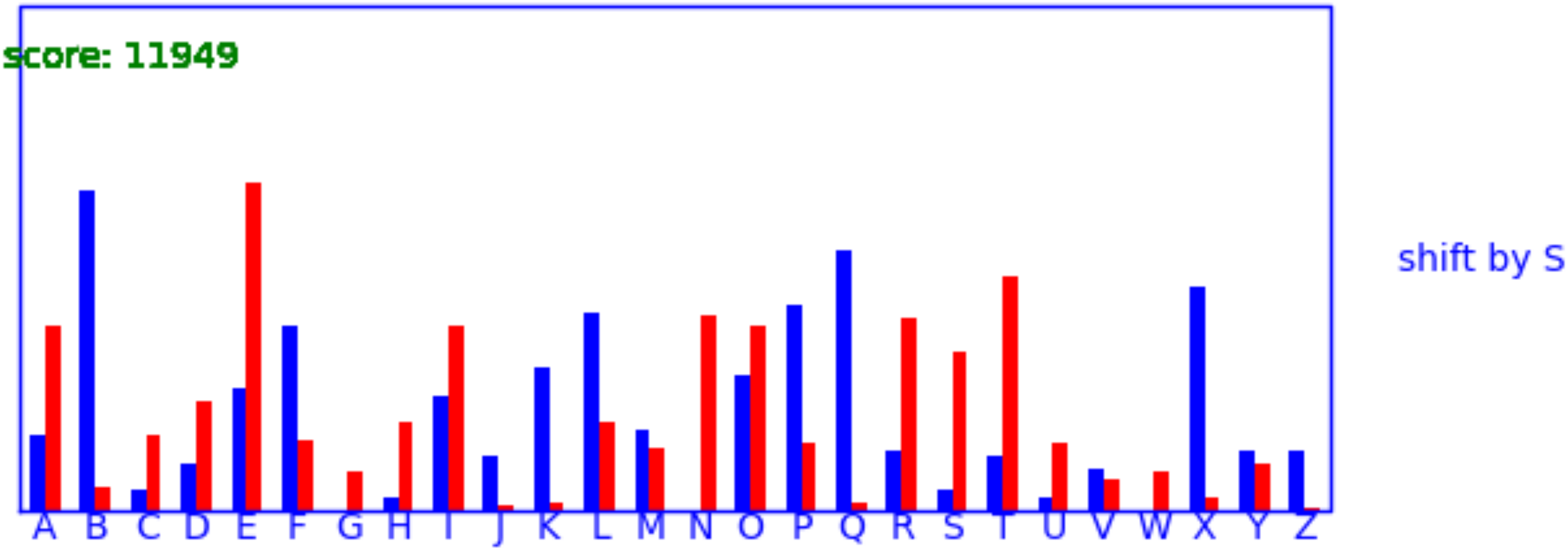

score: 11817
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by N

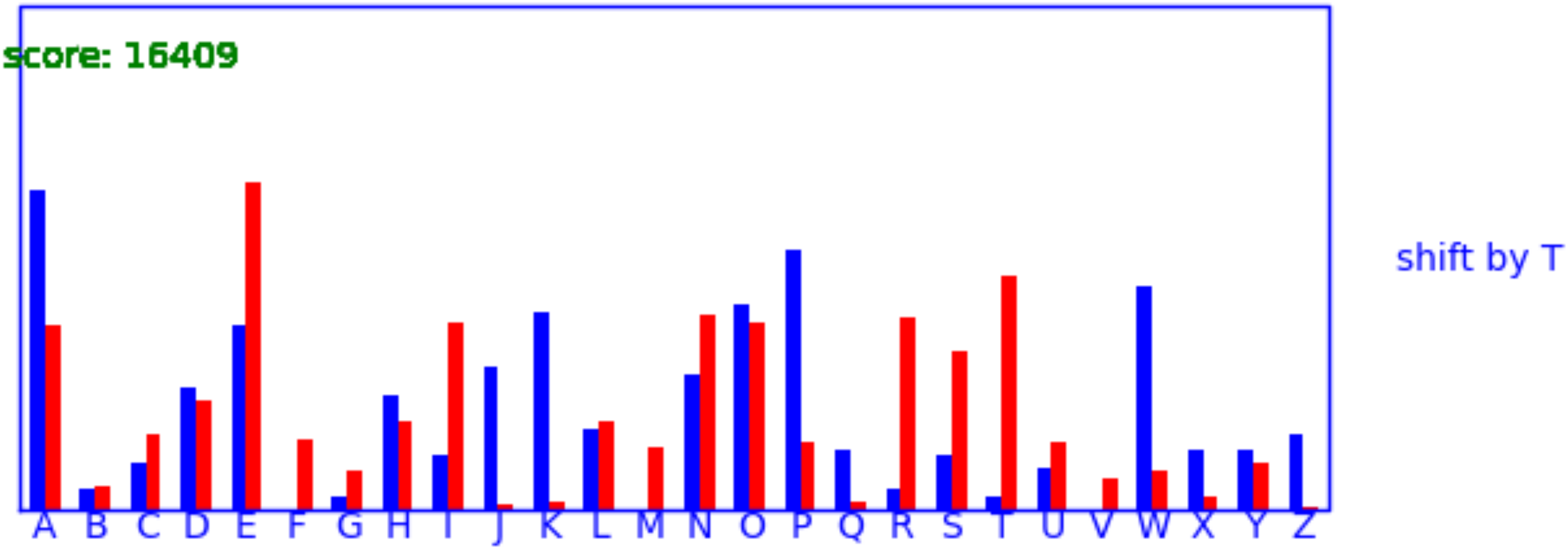

score: 14205
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
shift by O

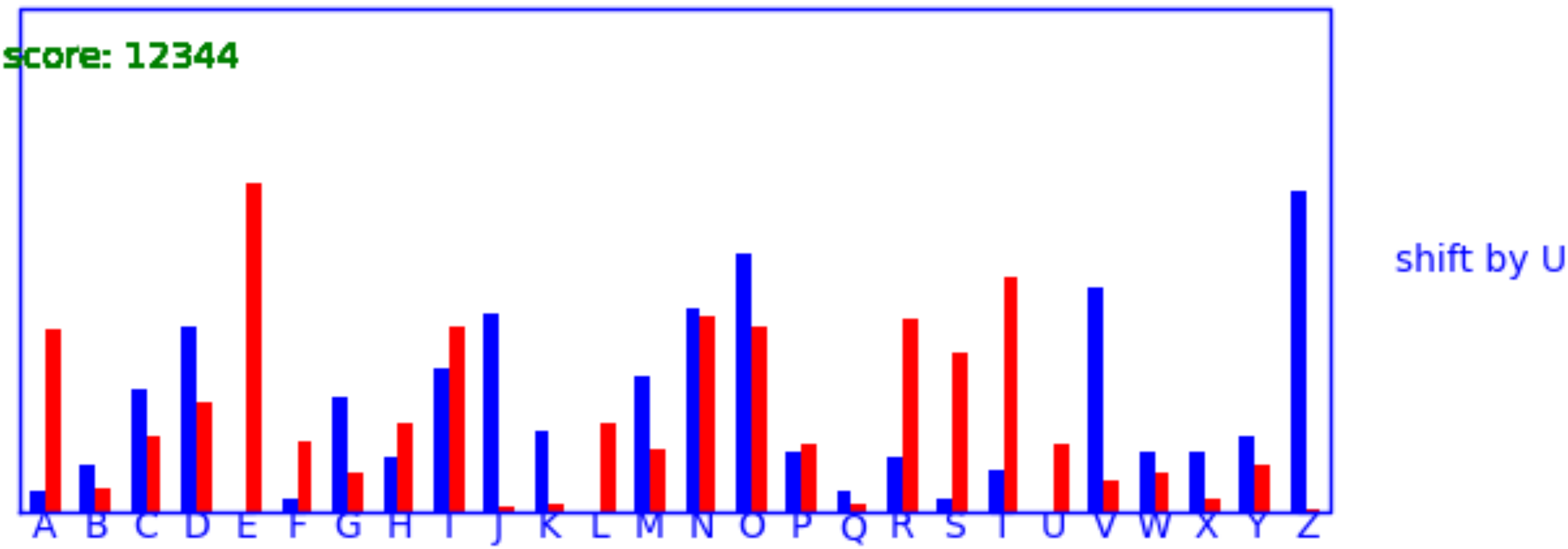

score: 24585
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by P

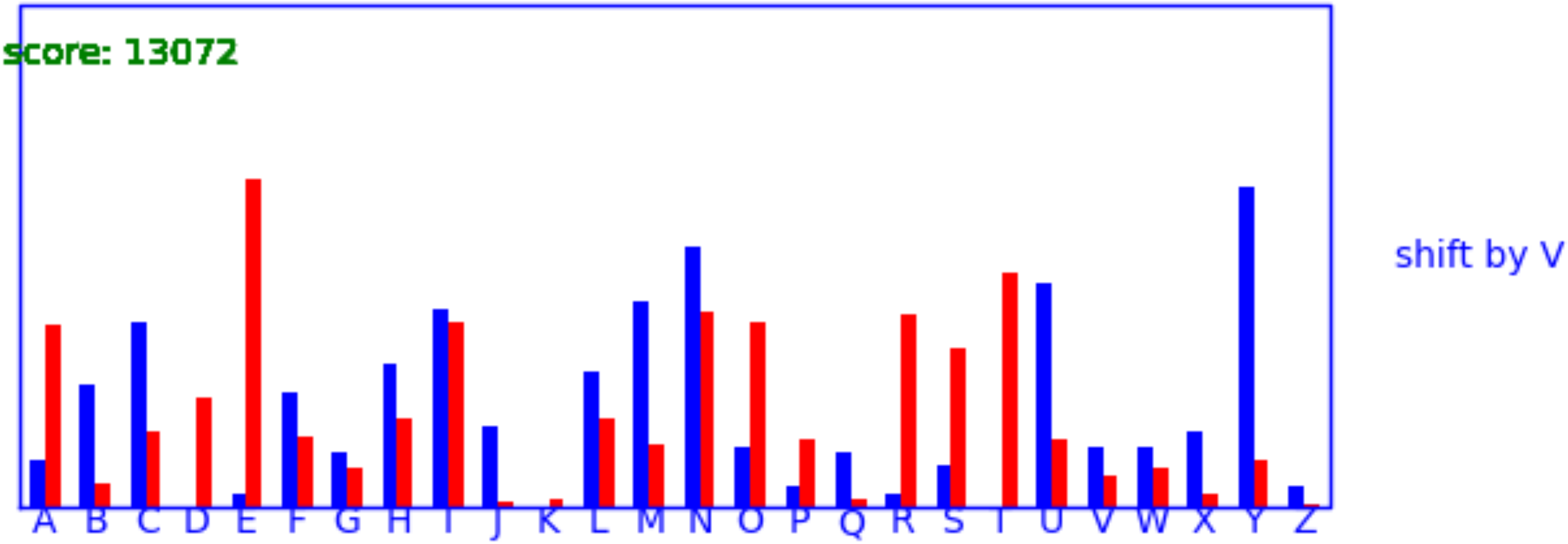

score: 14846
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by Q

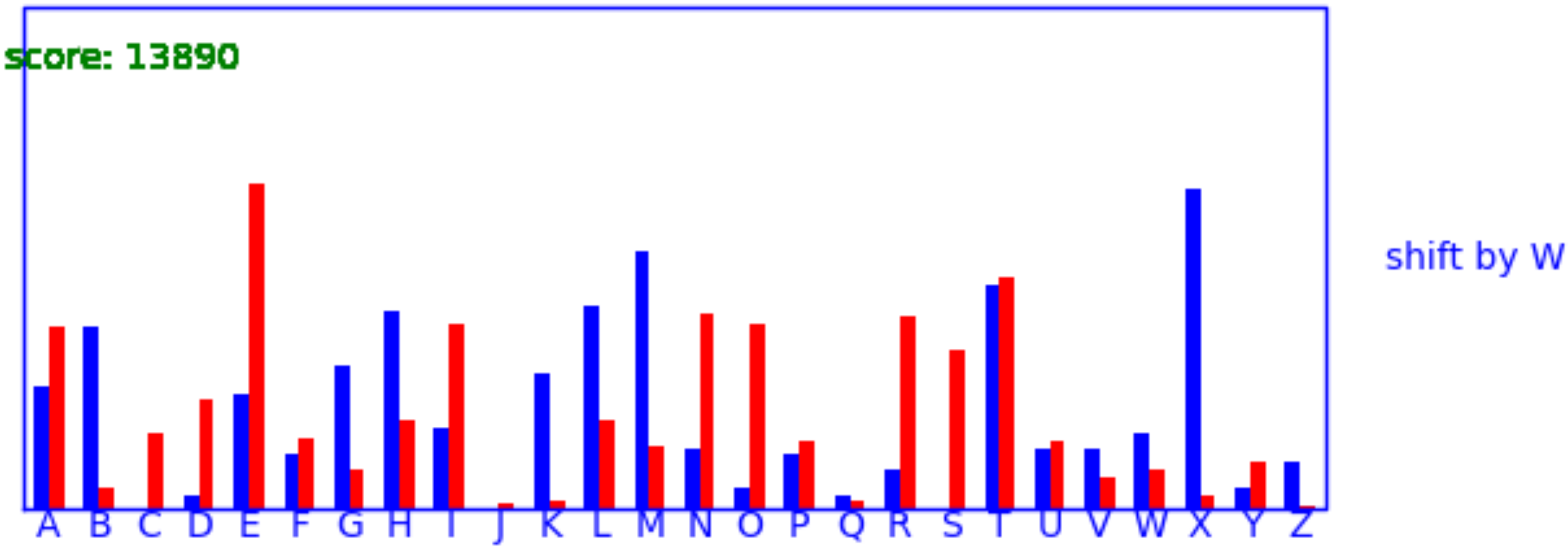

score: 12122
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by R

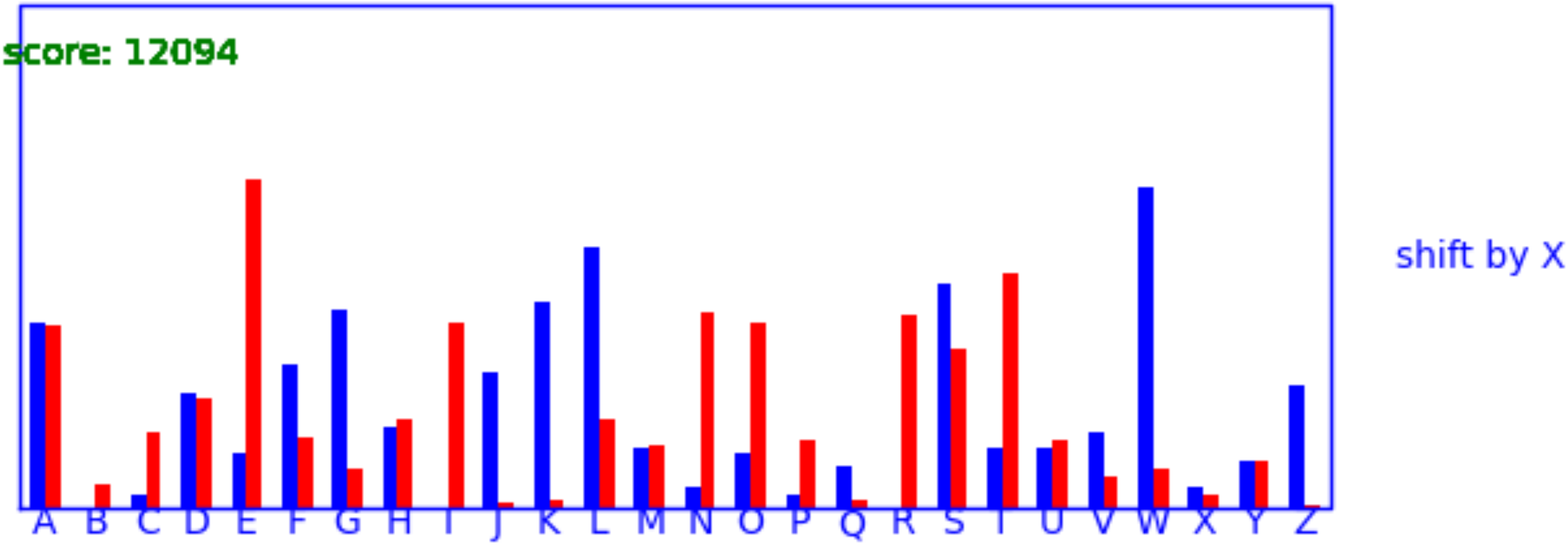

score: 11949
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by S

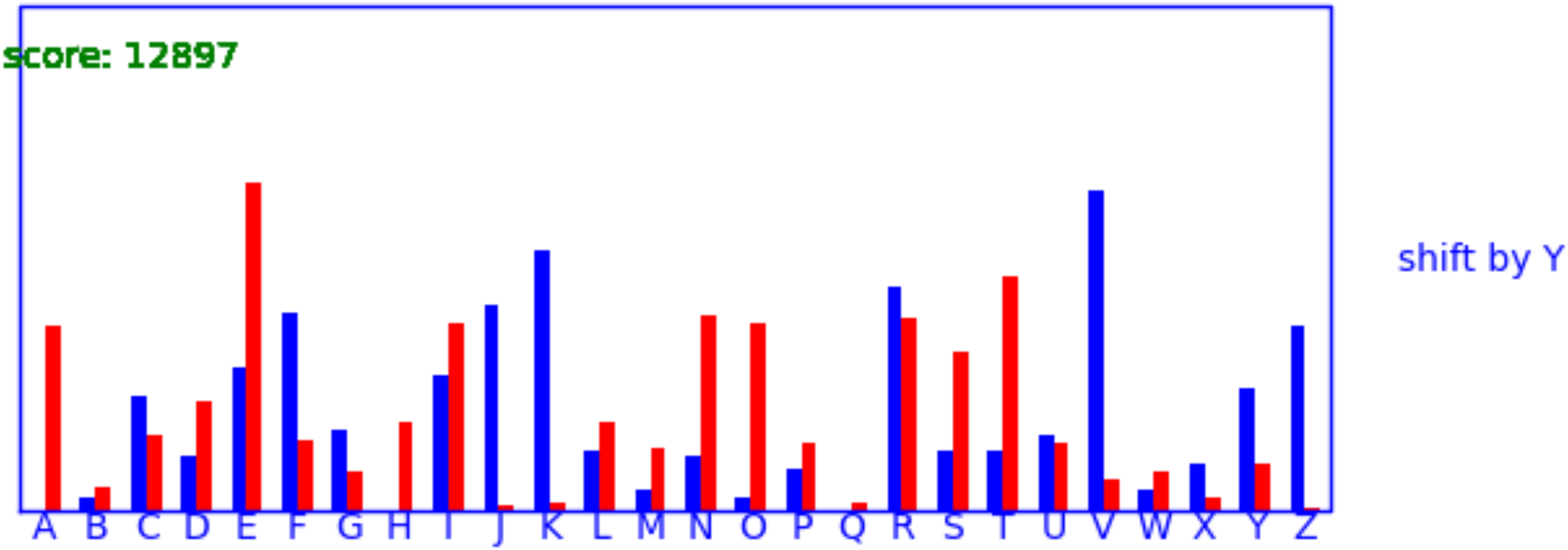

score: 16409
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by T

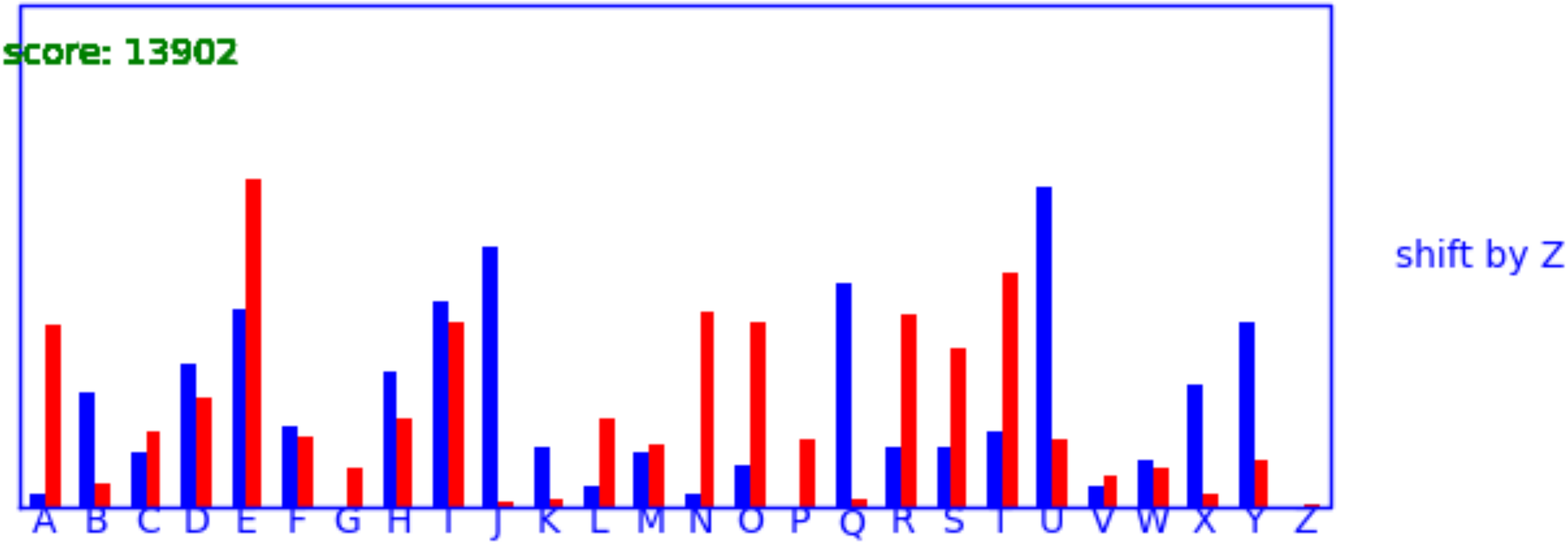

score: 12344
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
shift by U


score: 13072
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
shift by V


score: 13890
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
shift by W


score: 12094
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
shift by X


score: 12897
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
shift by Y


score: 13902
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
shift by Z

# Index of coincidence

$$I_c = \frac{\#\text{ of pairs of equal letters in } C}{\text{total }\#\text{ of pairs of letters in } C}$$

$\ldots E \ldots\ldots\ldots\ldots R \ldots$

pair of letters
not equal

$\ldots E \ldots\ldots\ldots\ldots E \ldots$

pair of equal letters

$$I_c = \frac{\sum_{\alpha=A}^{Z} N_\alpha(N_\alpha - 1)/2}{N(N-1)/2}$$

where $N$ = length of $C$
and $N_\alpha$ = # of $\alpha$ in $C$

Note $I_c$ is the same if you apply a Caesar OR Monoalphabetic substitution
for English expect $I_c \approx 0.065$

Say that my cypertext is grouped into $p$ blocks each with the same monoalphabetic substitution: Vigenere

$N$ = total letters in cypertext

$M$ = total letters in each block

$N = Mp$ $\quad$ $p$ = period

$$I_c = \frac{\sum_{i=1}^{p} \sum_{\alpha=A}^{Z} M_\alpha^{(i)}(M_\alpha^{(i)}-1)}{N(N-1)}$$

The index of coincidence is defined as

$$I_c = \frac{\text{number of pairs of equal letters in ciphertext}}{\text{the total number of pairs of letters}}$$

That is if we set

- $N_\alpha =$ the number of occurrences of the letter $\alpha$ in the cyphertext
- $$D_c = \sum_{\alpha=A}^{Z} \binom{N_\alpha}{2}$$

  $D_c$ represents the number of pairs of equal letters in the cyphertext.
- then $I_c = \frac{D_c}{\binom{N}{2}}$
- where $N =$ the number of letters in the cyphertext

The index of coincidence is invariant under monoalphabetic cyphers and we estimate under this condition that $N_\alpha = N * p_{\sigma(\alpha)}$ for some permutation of the alphabet $\sigma$ and so

$$\begin{aligned}
I_c &= \frac{\sum_{\alpha=A}^{Z}(N_\alpha^2 - N_\alpha)}{N(N-1)} \\
&\approx \frac{N^2(\sum_{\alpha=A}^{Z} p_\alpha^2) - N}{N(N-1)} \\
&= \frac{N(.065) - 1}{N-1} \\
&\approx .065
\end{aligned}$$

If the cyphertext was obtained from a polyalphabetic cipher then the index of coincidence can also be used to estimate the period of the cipher.
Let $p$ be the period of the cyphertext and place the letters of the cyphertext into groups of $p$ so that the letters in the $i^{th}$ position of the groups are all encrypted with the same key.

- Let $M_\alpha^{(i)}$ equal the number of occurrences of the letter $\alpha$ that appears in the $i^{th}$ positions in the groups.
- If there are $M$ groups of $p$, then $\sum_{\alpha=A}^{Z} M_\alpha^{(i)} = M$
- We also have $N = Mp$
- Also we can estimate that $M_\alpha^{(i)} \approx Mp_{\sigma(\alpha)}$ (again for some permutation for the alphabet $\sigma$)