1   1.1: Elementary Ciphers

Problem 2. *The following ciphertext was generated by a Vigenère cipher with a repeating key. All spaces and punctuation marks were removed from the plaintext, and the resulting ciphertext was broken into six-letter blocks.*

```
NRUATW   YAHJSE   DIODII   TLWCIJ   DIOPRA   DPANTO   EOOPEG
TNCWAS   DOBYAP   FRALLW   HSQNHW   DTDPIJ   GENDEO   BUWCEH
LWKQGN   LVEEYZ   ZEOYOP   XAGPIP   DEHQOX   GIKFSE   YTDPOX
DENGEZ   AHAYOI   PNWZNA   SAOEOH   ZOGQON   AAPEEN   YSWYDB
TNZEHA   SIZOEJ   ZRZPRX   FTPSEN   PIOLNE   XPKCTW   YTZTFB
PRAYCA   MEPHEA   YTDPSA   EWKAUN   DUEESE   YCNJPP   LNWWYO
TSKYEG   YOSDTD   LTPSED   TDZPNK   CDACWW   DCKYSP   CUYEEZ
MYDFMW   YIJEEH   WICPNY   PWDPRA   LSPSEK   CDACOB   YAPFRA
LPLLRA   YTHJCK   XEOQRK   XAOZUN   NEKFTO   TDAZFK   FROPLR
PSWYDE   DMKCEI   JSPPRE   ZUO
```

(a) *Look for strings of three or more letters that are repeated in the ciphertext. From the separations of different instances of the same string, try to infer the length of the key.*

(b) *Using frequency analysis or any other means, try to find the key and the plaintext.*

After briefly scanning the text, we'll use the trigram TDP as it appears three times in the 399-letter ciphertext.[1]

```
NRUATW   YAHJSE   DIODII   TLWCIJ   DIOPRA   DPANTO   EOOPEG
TNCWAS   DOBYAP   FRALLW   HSQNHW   DTDPIJ   GENDEO   BUWCEH
LWKQGN   LVEEYZ   ZEOYOP   XAGPIP   DEHQOX   GIKFSE   YTDPOX
DENGEZ   AHAYOI   PNWZNA   SAOEOH   ZOGQON   AAPEEN   YSWYDB
TNZEHA   SIZOEJ   ZRZPRX   FTPSEN   PIOLNE   XPKCTW   YTZTFB
PRAYCA   MEPHEA   YTDPSA   EWKAUN   DUEESE   YCNJPP   LNWWYO
TSKYEG   YOSDTD   LTPSED   TDZPNK   CDACWW   DCKYSP   CUYEEZ
MYDFMW   YIJEEH   WICPNY   PWDPRA   LSPSEK   CDACOB   YAPFRA
LPLLRA   YTHJCK   XEOQRK   XAOZUN   NEKFTO   TDAZFK   FROPLR
PSWYDE   DMKCEI   JSPPRE   ZUO
```

The original ciphertext with all `TDP`s shown.

The first appearance starts at the 67th letter, the second starts at 121st letter, and the final instance of the trigram appears at the 223rd letter. The distance between the first two

---

[1]Its important to note that there are five trigrams that appear three times: `PRA`, `RAL`, `DTD`, `PSE`, and our choice `TDP`. I quickly eliminated `DTD`, as the greatest common denominator between instances was 1. While `PRA` and `RAL` have a gcd of 3 and 9, respectively, the six-letter partitioning presented by the textbook could be a hint that the gcd would also be 6, which was the case for `TDP` and `PSE`. This would also cover a key of period 3, meaning three of the five possible keys would be included in the solution. In the end, this process was *not* brief.

instances is 54, and the distance between the last two instances is 102. The factors of 54 are $\{1, 2, 3, 6, 9, 18, 27, 54\}$ while the factors of 102 are $\{1, 2, 3, 6, 17, 34, 51, 102\}$. Consequently, $\gcd(54, 102) = 6$, so the periodic key is potentially 6 characters long. We'll take every sixth letter of the ciphertext and count their frequency.[2]

The first column of frequency-distribution pairs represents the first character in the hexagram, the second column the second letter, and so on. We start with the original trigram,

Letter	# 1	% 1	# 2	% 2	# 3	% 3	# 4	% 4	# 5	% 5	# 6	% 6
A	2	3.0	6	9.0	7	10.4	2	3.0	2	3.0	**9**	**13.6**
B	1	1.5	0	0.0	1	1.5	0	0.0	0	0.0	3	4.5
C	3	4.5	2	3.0	2	3.0	6	9.1	2	3.0	0	0.0
D	**10**	**14.9**	4	6.0	5	7.5	3	4.5	2	3.0	2	3.0
E	2	3.0	**7**	**10.4**	2	3.0	7	10.6	**14**	**21.2**	6	9.1
F	3	4.5	0	0.0	0	0.0	4	6.1	2	3.0	0	0.0
G	2	3.0	0	0.0	2	3.0	1	1.5	1	1.5	2	3.0
H	1	1.5	1	1.5	3	4.5	1	1.5	2	3.0	3	4.5
I	0	0.0	**7**	**10.4**	0	0.0	0	0.0	4	6.1	3	4.5
J	1	1.5	0	0.0	1	1.5	3	4.5	0	0.0	3	4.5
K	0	0.0	0	0.0	**8**	**11.9**	0	0.0	0	0.0	5	7.6
L	6	9.0	1	1.5	1	1.5	3	4.5	2	3.0	0	0.0
M	2	3.0	1	1.5	0	0.0	0	0.0	1	1.5	0	0.0
N	2	3.0	4	6.0	3	4.5	2	3.0	4	6.1	6	9.1
O	0	0.0	4	6.0	**10**	**14.9**	1	1.5	7	10.6	4	6.1
P	5	7.5	3	4.5	7	10.4	**12**	**18.2**	1	1.5	5	7.6
Q	0	0.0	0	0.0	1	1.5	4	6.1	0	0.0	0	0.0
R	0	0.0	5	7.5	0	0.0	0	0.0	7	10.6	1	1.5
S	2	3.0	6	9.0	1	1.5	3	4.5	5	7.6	1	1.5
T	6	9.0	**7**	**10.4**	0	0.0	1	1.5	5	7.6	0	0.0
U	0	0.0	4	6.0	1	1.5	0	0.0	2	3.0	0	0.0
V	0	0.0	1	1.5	0	0.0	0	0.0	0	0.0	0	0.0
W	1	1.5	3	4.5	6	9.0	2	3.0	1	1.5	6	9.1
X	4	6.0	0	0.0	0	0.0	0	0.0	0	0.0	3	4.5
Y	**10**	**14.9**	1	1.5	1	1.5	**8**	**12.1**	2	3.0	1	1.5
Z	4	6.0	0	0.0	5	7.5	3	4.5	0	0.0	3	4.5

Counts and frequencies of ciphertext character by their hexagram position.

TDP. Since our process depends on this trigram representing a common trigram in English, we should aim for one when we attempt difference shifts in each position of the hexagram in addition to using frequency analysis as an aid. The frequency-distribution pair table above has the most common characters bolded.

Starting with the most common English trigram, THE, we can see not only that our common cipher trigram starts with a T as well, but that it also could feasibly not be a shifted position via the reasonable frequency of E, I, and T. The procedure continues with

---

[2]This was done using Python; script found here.

guessing a shift of 4 in the third position, as that would correspond to a mapping $D \mapsto H$, our second target letter. Finally, after seeing that our ciphertext's P is the most common letter in the fourth position, we can reasonably guess TDP $\mapsto$ THE. We obtain the following partially deciphered text. After verifying by inspection that the shifted characters could lead

```
NRYPTW   YALYSE   DISSII   TLARIJ   DISERA   DPECTO   EOSEEG
TNGLAS   DOFNAP   FREALW   HSUCHW   DTHEIJ   GERSEO   BUAREH
LWOFGN   LVITYZ   ZESNOP   XAKEIP   DELFOX   GIOUSE   YTHEOX
DERVEZ   AHENOI   PNAONA   SASTOH   ZOKFON   AATTEN   YSANDB
TNDTHA   SIDDEJ   ZRDERX   FTTHEN   PISANE   XPORTW   YTDIFB
PRENCA   METWEA   YTHESA   EWOPUN   DUITSE   YCRYPP   LNALYO
TSONEG   YOWSTD   LTTHED   TDDENK   CDERWW   DCONSP   CUCTEZ
MYHUMW   YINTEH   WIGENY   PWHERA   LSTHEK   CDEROB   YATURA
LPPARA   YTLYCK   XESFRK   XASOUN   NEOUTO   TDEOFK   FRSELR
PSANDE   DMOREI   JSTERE   ZUS
```

Partially deciphered ciphertext using shifts of $\{0, \mathbf{0}, \mathbf{4}, -\mathbf{11}, 0, 0\}$

to a valid English message, we complete our procedure with frequency analysis. Positions five and six each have a letter significantly more frequent than the rest, which we'll assume is E in our plaintext. Finally, we can complete our deciphering attempt by checking if we can choose a reasonably frequent letter to complete our first word. In this case, we have NRYPTA. . . which most likely is CRYPTA. . . , especially taking context into consideration. We have our final set of deciphered hexagrams!

```
CRYPTA   NALYSI   SISSIM   ILARIN   SISERE   SPECTS   TOSEEK
INGLAW   SOFNAT   UREALA   WSUCHA   STHEIN   VERSES   QUAREL
AWOFGR   AVITYD   OESNOT   MAKEIT   SELFOB   VIOUSI   NTHEOB
SERVED   PHENOM   ENAONE   HASTOL   OOKFOR   PATTER   NSANDF
INDTHE   HIDDEN   ORDERB   UTTHER   EISANI   MPORTA   NTDIFF
ERENCE   BETWEE   NTHESE   TWOPUR   SUITSI   NCRYPT   ANALYS
ISONEK   NOWSTH   ATTHEH   IDDENO   RDERWA   SCONST   RUCTED
BYHUMA   NINTEL   LIGENC   EWHERE   ASTHEO   RDEROF   NATURE
APPARE   NTLYCO   MESFRO   MASOUR   CEOUTS   IDEOFO   URSELV
ESANDI   SMOREM   YSTERI   OUS
```

Our deciphered hexagrams!

The key, $\{15, 0, 4\}$, corresponds to PAE and is notably only 3 letters long.

> Cryptanalysis is similar in respects to seeking laws of nature. A law such as the inverse square law of gravity does not make itself obvious in the observed phenomena. One has to look for patterns and find the hidden order. But there is an important difference between these two pursuits. In cryptanalysis, one knows that the hidden order was constructed by human intelligence,whereas the

order of nature apparently comes from a source outside of ourselves and is more mysterious.