Question #2 on 2018 Sample 5 is a Patristocrat with a K1 alphabet and a simple clue:

2) **[300 Points]** Solve this Patristocrat which is a quote by Albert Camus which is encoded using a K1 key and starts with the word **IN**.

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
YUTVJ CGFUV GCGLC GPCZR UJKFF UY

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement																										

Here's one approach to solving it.

1. Since we are given that **CG** corresponds to **IN**, we can go through and make that substitution globally as well as put it in our replacement below

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
IN IN I IN N
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
IN NIN I N I

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement			I				N																			

2. Next we see that the frequency of **U** is so much higher than everything else, we are going to guess that it is the letter **E**.

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
IN E E IN E I IN E NE E
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
E IN E NIN I N I E

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement			I				N														E					

3. We are looking for high frequency letters to match **T** — the **W** is extremely suspicious. Given that we see **WX** occur multiple times in the code, and thinking how it starts with **IN**, the logical conclusion is that this might start as **IN THE**. Filling that in gives us:

```
CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE  E TH    IN TE I  IN      E   NE TH  TTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
 E    IN E  NIN I N I   E
```

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement			I				N														E		T	H		

4. A quick scan of the text so far shows a **TH?T THE** in the text which is just screaming to be **THAT** which means that **V** must be **A**

```
CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE  E TH    IN TE I  INA     EA  NE TH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
 E A  INlmEA NIN I N I   E  lmlm E
```

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement			I				N														E	A	T	H		

5. We can now take advantage of the fact that this is a K1 alphabet. The letters for **UVWX** are not in ascending order and as such must be part of the key word. Taking a quick gander at the **C-G** span where we have **I-N** and doing a quick count of **IJKLMN** we can see that it is part of the continuous alphabet and **F** must be either **M** or **L**. We also know that **YZAB** must be some of the letters **BCDFG** depending on where the keyword ends.

```
CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE  E TH  df IN TEbcIdf INA     EAbc NE TH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
bcE A  INlmEA NIN I N IcdE  lmlm Ebc
```

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	df	fg	I	jk	kl	lm	N														E	A	T	H	bc	cd

6. A quick look at the end of the end of the phrase tells us that there is no way that **Y** could be either **B** or **C** since we don't know of any reasonable words that end in **EB** or **EC**, particularly that have a **LL** or **MM** before the **E**. As such we must be able to assume that Y must match some other letter that is part of the keyphrase. This compresses our **BCDFG** sequence and tells us that two of the letters must appear in the keyphrase. Thinking of words that end in **LLE?** or **MME?** tells us that **Y** must be either a **D S** or **R** (It might have been **N** but that is already used).

```
C GWXU  SUHWX QA TCG  WUY C A  CGVRR NR UVY  GUSWX  VWWXU
INTHE   E  TH   cdf  IN  TEdsrI cdf  INA      EAdsr  NE  TH  AT THE
YUTVJ CGFUV GC GLC  GPC Z R  UJKFF  UY
dsrE A  INlmEA NI N  I N  I bcd   E   lmlm  Edsr
```

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	cdf	dfg	I	jk	kl	lm	N														E	A	T	H	dsr	bcd

7. Looking at the sequence **INTE**dsr**I**cdf**IN** doesn't give us any clues, but the **EA**dsr**NE?** Quickly eliminates **D** and **S** since there are almost no words that end **EADNE?** Or **EASNE?** and which lets us assume the **Y** must be **R**.

```
CGWXU  SUHWX QA TCG  WUYC A  CGVRR NRUVY  GUSWX  VWWXU
INTHE   E  TH   cdf  IN  TERI cdf  IN       EAR NE  TH  AT THE
YUTVJ CGFUV GC GLC  GPCZ R  UJKFF  UY
RE A  INlmEA NI N  I N  I bcd   E   lmlm  ER
```

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	cdf	dfg	I	jk	kl	lm	N														E	A	T	H	R	bcd

8. Now that we have the sequence **EARNE?** it is apparent that the **S** must stand for **D** since we don't know of any other letters that could end that word. We can also drop **D** from the running for **Z**, **A** and **B**. Additionally, this tells us that the keyword must span from **S** to **Y** at least. Given that **R** has such a high frequency, we know that it couldn't possibly be the letter **Z** so we can assume the keyword at least spans from **R** to **Y**

```
CGWXU  SUHWX QATCG  WUYCA CGVRR NRUVY  GUSWX  VWWXU
INTHE  DE  TH   cf  IN  TERIcf  INA      EAR NEDTH  AT THE
YUTVJ CGFUV GCGLC  GPCZR UJKFF  UY
RE A  INlmEA NIN  I N  I bc   E   lmlm  ER
```

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	cf	fg	I	jk	kl	lm	N												D		E	A	T	H	R	bc

9. Reading the beginning of the phrase, there is only one word which starts **DE?TH** (it can't be **DEATH** since **A** is already known) so **H** must stand for **P**. We also see that the letter O has been skipped in the alphabet which tells us the **T R** or even **P** or **Q** must stand for **O** since **O** must be part of the keyphrase. We know it can't be **T** since it would make **REOA** at the start of the second line since that would give us either **THERE OA** or **THE REOA**. Likewise it can't be **R** since it would give us **OEARNED** near the end of the first line. This means that **Q** must be part of the key phrase (and possibly more). As we read the start testing for **O** for **Q**, we see **IN THE DEPTH O**cf which implies that **Q** could be **O** and **A** would be **F** (making **B** be **G** even though it isn't used).

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE DEPTH OF IN TERIF IN EAR NEDTH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
RE A INlm**EA NIN I N I**bc **E** lmlm **ER**

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	F	G	I	jk	kl	lm	N	P									O		D		E	A	T	H	R	bc

10. This also gives us another couple of words splits as there is only one word left which could start **?INTER** – **WINTER** – which means that **T** must be **W** and we can see that the part of the code words from **T** to **Y** is **WEATHER** (you drop the duplicate **E**.

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE DEPTH OFWIN TERIF INA EAR NEDTH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
REWA INlm**EA NIN I N I**bc **E** lmlm **ER**

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	F	G	I	jk	kl	lm	N	P									O		D	W	E	A	T	H	R	bc

11. Looking at the **RR** that is next we go thought he list of common double letters to realize it must be **LL** which means that **N** must be a **Y** for the word **FINALLY** giving us a few good keys including knowing that **F** must not be **L** but must be **M**.

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE DEPTH OFWIN TERIF INALL YLEAR NEDTH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
REWA INMEA NIN I N Ibc**L E MM ER**

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	F	G	I	J	K	M	N	P						Y			O	L	D	W	E	A	T	H	R	bc

12. Remembering that this is a K1 alphabet we know that either **B** or **C** must be part of the keyword (since they both can't be used for **Z**). A little thought looking at **P** says that it must be **C** since **BOLD WEATHER** doesn't make sense and **COLD WEATHER** goes well with the **WINTER** word in the phrase. So we map **P** to **B**, **Z** to **B** and then fill in the rest of the alphabet since **O** has to be **Z** (either it isn't used in the key phrase and would have to go after **Y** or it is used in the keyphrase and it is the only spot left over). We can also fill in **IJKLM** to be **QSUVX** since **R T** and **W** are used in the keyphrase.

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE DEPTH OFWIN TERIF INALL YLEAR NEDTH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
REWAS INMEA NINVI NCIBL ESUMM ER

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	F	G	I	J	K	M	N	P	Q	S	U	V	X	Y	Z	C	O	L	D	W	E	A	T	H	R	B

13. As we fill them in and put in the word breaks to read it, it is worth noticing that the **THAT THE** was actually **THAT THERE**. Sometimes the line breaks can cause you to think that a word ends when in fact it continued on to the next line. For purposes of scoring on the test it isn't necessary to separate out the words, but it can be helpful in solving.

CGWXU SUHWX QATCG WUYCA CGVRR NRUVY GUSWX VWWXU
INTHE DEPTH OFWIN TERIF INALL YLEAR NEDTH ATTHE
YUTVJ CGFUV GCGLC GPCZR UJKFF UY
REWAS INMEA NINVI NCIBL ESUMM ER

K1	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
Frequency	2		8			3	8	1		2	1	1		1		1	1	4	2	2	10	5	6	4	4	1
Replacement	F	G	I	J	K	M	N	P	Q	S	U	V	X	Y	Z	C	O	L	D	W	E	A	T	H	R	B