[CSCI 55500 Cryptography] HW1
David Xu

All attached code was written in Python3 (3.6.1) using packages preinstalled by Anaconda (4.4.0).

Cipher 1

a. Vigenere cipher with key:

DARKKNIGHT

Plaintext:

"the batman phenomena gives me great satisfaction to realize that i have millions of fans in the world. that i have created a hero that influenced so many people in the world and it gives me a wonderful spiritual satisfaction to know that and it really touches my heart, i feel so gratified that i created a superhero who can reach all elements of our society, and that's quite fulfilling" -bob kane

batman is a superhero co-created by artist bob kane and writer bill finger and published by dc comics. the character made his first appearance in detective comics #27 (may, 1939). batman is the secret identity of bruce wayne. witnessing the murder of his parents as a child leads him to train himself to physical and intellectual perfection and don a bat-themed costume in order to fight crime. batman operates in gotham city, assisted by various supporting characters including his sidekick robin and his butler alfred pennyworth, and fights an assortment of villains influenced by the characters' roots in film and pulp magazines. unlike most superheroes, he does not possess any superpowers; he makes use of intellect, detective skills, science and technology, wealth, physical prowess, and intimidation in his war on crime.

batman became a popular character soon after his introduction, and eventually gained his own title, batman. as the decades wore on, differing takes on the character emerged. the late 1960s batman television series utilized a camp aesthetic associated with the character for years after the show ended. various creators worked to return the character to his dark roots, culminating in the 1986 miniseries batman: the dark knight returns, by writer-artist frank miller. that and the success of director tim burton's 1989 batman motion picture helped reignite popular interest in the character. a cultural icon, batman has been licensed and adapted into a variety of media, from radio to television and film, and appears on a variety of merchandise sold all over the world. the batman goes by numerous nicknames, such as the dark knight, the caped crusader, or simply the bat.

batman's history has undergone various revisions, both minor and major. few elements of the character's history have remained constant. scholars william uricchio and roberta e. pearson noted in the early 1990s, "unlike some fictional characters, the batman has no primary urtext set in a specific period, but has rather existed in a plethora of equally valid texts constantly appearing over more than five decades."

the central fixed event in the batman stories is the character's origin story. as a little boy, bruce wayne is shocked to see his parents, the physician dr. thomas wayne and his wife martha wayne, being murdered by a mugger in front of his very eyes. this drives him to fight crime in gotham city as the batman.

pearson and uricchio also noted beyond the origin story and such events as the introduction of robin, "until recently, the fixed and occurring and hence, canonized, events have been few in number," a situation altered by an increased effort by later batman editors such as dennis o'neil to ensure consistency and continuity between stories.

batman was involved in a battle with dr. simon hurt and the "black glove," a criminal organization dedicated to corrupting virtue, as they attempted to destroy batman and everything for which he stands.
using a mixture of physical and psychological attacks, the black glove tests batman's resolve, forcing him to temporarily adopt the crazed persona of the "batman of zur-en-arrh." he is then led to arkham asylum to face the joker. seemingly defeated, batman is buried alive by the black glove, a group that includes bruce wayne's girlfriend, jezebel jet, who has betrayed him. with the assistance of robin and nightwing he turns the tables on his foes. in a final confrontation with dr. hurt, batman is caught in a helicopter crash.

the storyline concludes with both batman's fate and the true identity of dr. hurt still in the air. hurt himself repeatedly claimed to be thomas wayne throughout the story, while batman apparently believed him to be mangrove pierce, a crazed actor.

b. The vigenere cipher was the second that I decrypted. I used the framework from the substitution cipher to initially determine the letter frequencies and found a normal distribution, indicating that cipher 1 was most likely the vigenere cipher.

Letter	Count	Frequency
K	203	0.06077844311377246
V	182	0.05449101796407185
R	171	0.05119760479041916
O	163	0.04880239520958084
B	152	0.045508982035928146
I	144	0.04311377245508982
A	144	0.04311377245508982
U	141	0.04221556886227545
E	141	0.04221556886227545
L	140	0.041916167664670656
Y	136	0.0407185628742515
N	136	0.0407185628742515
H	132	0.03952095808383234
D	130	0.038922155688622756
G	128	0.03832335329341317
Z	127	0.03802395209580838
X	122	0.03652694610778443
S	121	0.03622754491017964
M	116	0.03473053892215569
T	113	0.033832335329341316
W	104	0.031137724550898204
C	92	0.027544910179640718
P	86	0.025748502994011976
F	80	0.023952095808383235
Q	72	0.02155688622754491
J	64	0.019161676646706587

I implemented an incidence of coincidence function to determine the key length $m = 10$.

m	char	Incidence of coincidence
1	0	0.04059051707904945
2	0	0.0427219856273074
2	1	0.04460270591232155
3	0	0.04094583194207191
3	1	0.040772233964849684
3	2	0.039915776273859616
4	0	0.04245035109636842
4	1	0.045345280661698184
4	2	0.0426542598256724
4	3	0.04390355978690102
5	0	0.05242438660909066
5	1	0.0491520706712512
5	2	0.05384283905951216
5	3	0.046535115675695085
5	4	0.04964134699117507
6	0	0.04257132893326272
6	1	0.04540640378182194
6	2	0.043275254123451685
6	3	0.04501246399648683
6	4	0.04211549679175578
6	5	0.042718257826171495
7	0	0.03988491530924625
7	1	0.04059862938886246
7	2	0.03954160280464387
7	3	0.040633863608336415
7	4	0.04020224441978049
7	5	0.04051935239504607
7	6	0.041012631467681415
8	0	0.04269503057840809
8	1	0.04575860842426537
8	2	0.043394949112480354

8	3	0.04315399355157023
8	4	0.04142455266555986
8	5	0.045863309352517985
8	6	0.04232383324109943
8	7	0.04363816638996495
9	0	0.042010839637133
9	1	0.040984920230203246
9	2	0.038216653310992936
9	3	0.04005245137320609
9	4	0.040722663364172795
9	5	0.04013987032854957
9	6	0.04081008231951628
9	7	0.04073723319006338
9	8	0.041917389087200406
10	**0**	**0.07178435920950892**
10	**1**	**0.06624588660516804**
10	**2**	**0.06592220963478448**
10	**3**	**0.06263149376921832**
10	**4**	**0.06606606606606606**
10	**5**	**0.07216198234162306**
10	**6**	**0.06054557551563539**
10	**7**	**0.0650410889931848**
10	**8**	**0.061228893564222904**
10	**9**	**0.06022189854525184**

For each k_i in $k_1k_2k_3 \ldots k_{10}$, I iteratively shifted through 26 keys and sorted each key by their incidence of coincidence. I found multiple possible shifts for each of the positions that had high incidence of coincidences. For example, the sorted IOCs for each position were as follows,

0 [('W', 0.0959), ('S', 0.0901), **('D', 0.0823),** ('L', 0.0809), ('K', 0.0806), ('Z', 0.0782), ('V', 0.0774), ('F', 0.0729), ('J', 0.0672), ('E', 0.0658), ('A', 0.0633), ('U', 0.0624), ('R', 0.0596), ('T', 0.0558), ('G', 0.0534), ('H', 0.0519), ('M', 0.0516), ('O', 0.0499), ('Q', 0.0487), ('Y', 0.0466), ('C', 0.0459), ('X', 0.0459), ('N', 0.0458), ('I', 0.0451), ('P', 0.0444), ('B', 0.033)]

1 [('P', 0.0915), **('A', 0.0839),** ('H', 0.0836), ('T', 0.0821), ('G', 0.0803), ('I', 0.0756), ('W', 0.0754), ('C', 0.071), ('S', 0.067), ('B', 0.0652), ('O', 0.0651), ('R', 0.0573), ('X', 0.0572), ('N', 0.0562), ('D', 0.0561), ('J', 0.0539), ('F', 0.052), ('E', 0.0497), ('Q', 0.0497), ('V', 0.0492), ('K', 0.0481), ('Z', 0.0477), ('L', 0.0475), ('M', 0.0473), ('U', 0.0462), ('Y', 0.0361)]

2 [('K', 0.0923), **('R', 0.0922),** ('Y', 0.0852), ('G', 0.0812), ('N', 0.0789), ('X', 0.0782), ('Z', 0.0756), ('T', 0.068), ('J', 0.0677), ('O', 0.0656), ('S', 0.0649), ('V', 0.0568), ('F', 0.0565), ('Q', 0.0558), ('H', 0.054), ('E', 0.0528), ('I', 0.0525), ('U', 0.0513), ('M', 0.0512), ('W', 0.0512), ('L', 0.05), ('C', 0.0473), ('A', 0.0466), ('D', 0.0424), ('B', 0.0387), ('P', 0.038)]

3 [**('K', 0.0876),** ('Z', 0.0792), ('R', 0.079), ('Q', 0.0784), ('S', 0.0768), ('G', 0.0758), ('D', 0.0756), ('C', 0.0722), ('H', 0.0712), ('M', 0.069), ('L', 0.0652), ('B', 0.0645), ('N', 0.0595), ('A', 0.0579), ('J', 0.0575), ('Y', 0.0541), ('X', 0.0532), ('P', 0.052), ('F', 0.0501), ('O', 0.0489), ('T', 0.0482), ('E', 0.048), ('W', 0.0459), ('V', 0.0437), ('U', 0.0429), ('I', 0.0385)]

4 [('D', 0.0926), **('K', 0.0909),** ('Z', 0.0845), ('S', 0.0835), ('R', 0.0833), ('G', 0.0787), ('C', 0.0739), ('M', 0.068), ('Q', 0.0652), ('L', 0.0646), ('B', 0.0629), ('H', 0.0606), ('Y', 0.0571), ('N', 0.0547), ('J', 0.0543), ('A', 0.0536), ('V', 0.053), ('O', 0.0516), ('F', 0.0506), ('U', 0.0486), ('T', 0.0473), ('X', 0.0455), ('P', 0.045), ('E', 0.0442), ('W', 0.0431), ('I', 0.0377)]

5 [('G', 0.0932), **('N', 0.0867),** ('C', 0.0836), ('U', 0.0834), ('J', 0.0827), ('P', 0.079), ('T', 0.0776), ('V', 0.0744), ('F', 0.07), ('O', 0.0643), ('Q', 0.0616), ('K', 0.0597), ('B', 0.0591), ('I', 0.0554), ('A', 0.0529), ('R', 0.0528), ('E', 0.0519), ('M', 0.0508), ('S', 0.0507), ('Y', 0.0474), ('H', 0.0469), ('D', 0.0456), ('X', 0.0444), ('W', 0.0438), ('Z', 0.0431), ('L', 0.0341)]

```
6 [('X', 0.089), ('I', 0.0864), ('B', 0.0793), ('P', 0.0781), ('Q', 0.0769), ('E',
0.0765), ('A', 0.074), ('J', 0.0718), ('O', 0.0713), ('K', 0.0681), ('Z', 0.0637),
('F', 0.0618), ('W', 0.0615), ('Y', 0.0552), ('L', 0.055), ('H', 0.0529), ('T',
0.0522), ('M', 0.0516), ('D', 0.0505), ('N', 0.0497), ('R', 0.0494), ('S', 0.0477),
('C', 0.0471), ('V', 0.0454), ('U', 0.0412), ('G', 0.0386)]

7 [('V', 0.0885), ('C', 0.087), ('Z', 0.0848), ('N', 0.0819), ('G', 0.0813), ('O',
0.0736), ('M', 0.0699), ('I', 0.0697), ('D', 0.0688), ('Y', 0.0686), ('U', 0.0655),
('J', 0.0638), ('H', 0.0612), ('B', 0.0576), ('X', 0.057), ('W', 0.0561), ('K',
0.0507), ('T', 0.0499), ('F', 0.0487), ('Q', 0.0485), ('R', 0.0479), ('A', 0.0463),
('L', 0.0451), ('P', 0.0448), ('S', 0.0423), ('E', 0.0355)]

8 [('W', 0.0857), ('O', 0.0811), ('A', 0.0785), ('D', 0.0781), ('J', 0.0754), ('N',
0.075), ('P', 0.0739), ('H', 0.0722), ('Z', 0.0721), ('I', 0.0637), ('Y', 0.0636),
('E', 0.0616), ('X', 0.0596), ('V', 0.0586), ('C', 0.057), ('U', 0.0566), ('K',
0.0534), ('Q', 0.0532), ('B', 0.0521), ('G', 0.051), ('M', 0.0495), ('T', 0.0483),
('L', 0.0479), ('R', 0.0453), ('S', 0.0444), ('F', 0.0371)]

9 [('P', 0.0812), ('T', 0.0805), ('I', 0.0804), ('M', 0.0797), ('Z', 0.0775), ('A',
0.0769), ('V', 0.0743), ('B', 0.0699), ('L', 0.0686), ('W', 0.0674), ('Q', 0.0643),
('H', 0.0603), ('X', 0.0603), ('U', 0.0602), ('K', 0.0582), ('G', 0.0544), ('E',
0.0531), ('O', 0.0528), ('Y', 0.0523), ('S', 0.0507), ('N', 0.0488), ('J', 0.0486),
('C', 0.048), ('D', 0.0474), ('F', 0.0427), ('R', 0.0365)]
```

I guessed that the first word in the ciphertext "WHY" mapped to the plaintext word "the", indicating a left shift of $(3, 0, 17) = (D, A, R)$ for the key. Additional substitutions were used to generate the remaining k_i to give the key to the cipher. After making additional choices for keys, I realized that the plaintext passage was in reference to Batman and guessed the remainder of the key.

c. The major difficulty of this cipher was after determining the key length m and calculating the incidence of coincidences for each of the 26 possible shifts of the key. I excepted fewer possible shifts for each position with IOC ≥ 0.065 and initially used the maximum IOC shift as the key. However, the resulting plaintext was incoherent. Therefore, I used a manual strategy of assigning keys for each position.

Cipher 2

a. Permutation cipher with key:

Decrypt

x	1	2	3	4	5	6	7	8	9
$\pi^{-1}(x)$	4	5	1	9	7	2	8	6	3

Plaintext:

```
lordeddardnedstarkistheheadofthestarkfamilywhosemembersareinvolvedinmostoftheseriessintertwinedplo
tlinesheandhiswifecatelyntullyhavefivechildrentheeldestrobbthedaintysansathetomboyaryatheadventuro
usbranandtheyoungestrickonnedshostageandwardtheongreyjoyusedtolivewiththestarksrobbswifeistheheale
rtalisamaegyrandaryahasbefriendedtheblacksmithsapprenticegendrynedsbastardsonjonsnowandhisfriendsa
mwelltarlyserveinthenightswatchunderlordcommanderjeormormontheredhairedwildlingygritteisjonsnowsr
omanticinterestnedsoldfriendkingrobertbaratheonsharesalovelessmarriagewithqueencerseilannisterwhoh
astakenhertwinthekingslayerserjaimelannisterashersecretlovershesloathesheryoungerbrotherthecleverdw
arftyrionlannisterwhoisattendedbyhismistressshaeandthesellswordbronncerseisfatheristhefabulouslywe
althylordtywinlannisterandhersonjoffreyisguardedbythescarfacedwarriorsandorthehoundcleganethekings
```

smallcouncilofadvisorsincludesthecraftymasterofofcoinlordpetyrlittlefingerbaelishandtheeunuchmaste
rofwhispererslordvarysrobertsbrotherstannisbaratheonisadvisedbytheforeignpriestessmelisandreandthe
formersmugglerserdavosseaworththewealthytyrellfamilyisrepresentedatcourtbytheambitiousmargaerytyre
llacrossthenarrowseasiblingsviserysanddaenerystargaryentheexiledchildrenofthekingoverthrownbyrober
tbaratheonareontherunfortheirlivestryingtowinbackthethronedaeneryshasbeenmarriedtokhaldrogothelead
erofthenomadicdothrakiandisguardedbytheexileknightserjorahmormontttheshowscostumesareinspiredbymany
culturessuchasjapaneseandpersiandothrakioutfitsresemblethatofthebedouinsonewasmadeoutoffishskinsto
resembledragonscalesandthewildlingswearfursideinandskinsideoutliketheinuitwildlingbonearmorismadeo
fmoldstakenofrealbonesandassembledwithstringandlatexresemblingcatgutwhileextraswhoportraywildlings
andthenightswatchwearhatsaswouldbenormalinacoldclimatemainactorsusuallydonotsoviewerscanidentifyth
echaractersbjorksalexandermcqueenhighnecklinedressesinspireddormersunusualfunnelneckoutfitandprost
itutecostumesaredesignedtobequicklyremovedallclothingwhetherforwildlingsorforwomenattheroyalcourti
sagedfortwoweekstoimproverealismonthehighdefinitiontelevisionabouttwodozenwigsareusedforactorssuch
asheadeydormervanhoutenandclarkemadeofhumanhairanduptotwofeetinlengththeycostupto7000usdeachandare
washedandstyledlikerealhairapplyingthewigsisalengthyprocessclarkeforexamplerequiresabouttwohoursto
styleherbrunettehairwithaplatinumblondewigandbraidsotheractorssuchasgleesonandturnerreceivefrequen
thaircoloringforcharacterssuchasclarkeandherdothrakihairwigsandcostumesareprocessedsotheyappearasi
ftheyhavenotbeenwashedforweeks

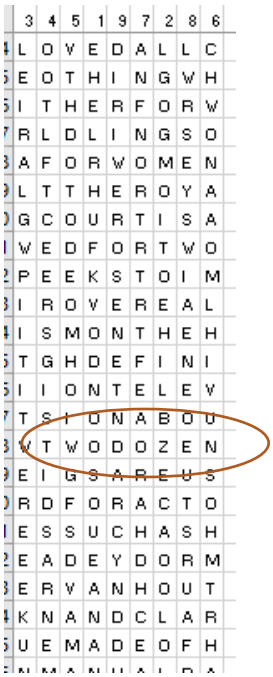
b. I initially calculated letter frequency and incidence of coincidence for the permutation cipher:

single	letter	freq
E	326	0.12665112665112666
R	218	0.08469308469308469
A	211	0.08197358197358197
S	202	0.07847707847707848
T	197	0.07653457653457653
N	176	0.06837606837606838
O	166	0.0644910644910645
I	163	0.06332556332556333
H	140	0.05439005439005439
D	126	0.04895104895104895
L	113	0.043900543900543904
U	63	0.024475524475524476
C	62	0.024087024087024088
W	55	0.021367521367521368
Y	55	0.021367521367521368
M	55	0.021367521367521368
G	50	0.019425019425019424
F	49	0.019036519036519036
B	48	0.018648018648018648
K	30	0.011655011655011656
V	24	0.009324009324009324
P	24	0.009324009324009324
J	9	0.0034965034965034965
X	6	0.002331002331002331
Q	5	0.0019425019425019425
Z	1	0.0003885003885003885

m	pos	IOC
1	0	0.06568027127685115
2	0	0.06740693210366616
2	1	0.063675394935115
3	0	0.0681539386323517
3	1	0.06421272232240727
3	2	0.06447655805882177
4	0	0.0648503231166021
4	1	0.06591288892323445
4	2	0.0709001322655194
4	3	0.06194192913862686
5	0	0.06614785992217899
5	1	0.06393411658040875
5	2	0.06641985569113369
5	3	0.065770088020853
5	4	0.06667880249694708
6	0	0.07598631897697318
6	1	0.06400453129425092

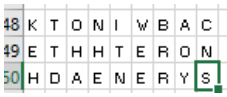
6	2	0.06349258218417098
6	3	0.06094372916802823
6	4	0.06425505958216239
6	5	0.06446201773304577
7	0	0.0666686411562611
7	1	0.07143703352683331
7	2	0.06714251865892666
7	3	0.06141156261106504
7	4	0.0666686411562611
7	5	0.064605946903709
7	6	0.06094310686261372
8	0	0.06855517501596331
8	1	0.06791664247982818
8	2	0.07165109034267912
8	3	0.06071863934521391
8	4	0.06186025812194036
8	5	0.06393065149668156
8	6	0.0691588785046729
8	7	0.06466121495327103
9	0	0.0685069316648264
9	1	0.06124401913875598
9	2	0.06440927493559072
9	3	0.06377131640289535
9	4	0.06453195926880137
9	5	0.06374677953625323
9	6	0.07000368052999632
9	7	0.06597963440068703
9	8	0.06475279106858055
10	0	0.07127560100141767
10	1	0.06180436159623563
10	2	0.0650318221578741
10	3	0.06276958344644527
10	4	0.07125486381322957
10	5	0.06064567120622568
10	6	0.06490150778210117
10	7	0.06861016536964981
10	8	0.06940053501945526
10	9	0.06319917315175097

This indicated that cipher 2 was the permutation cipher, as the IOC at each position in m was roughly the same. With the absence of punctuations and spacing, and an unknown m length, I first tried to brute force $8 \leq m \leq 10$ since the number of keys is $m!$ at $7! = 5040$ is still manageable by hand. Seeing as m must be $8 \leq m \leq 10$, I initially tried $m = 9$, as the number of characters in the ciphertext was divisible by 9 without a remainder. From the letter frequencies, I noticed that the letter z occurs once. I then split the letters into columns and plotted in Excel. I initially guessed that the word with 'z' was 'dozen':



Looking at the partially deciphered text, I saw partial text that unscrambled to references in Game of Thrones:

 -> bastard, jon snow

 -> daenerys

Rearranging the permutation led to the deciphered text with the above key.

c. The difficulty in this cipher was determining plaintext when there were no indicators of where sentences or words began or ended. I initially tried to find the text around the only number in the ciphertext, 7000, but the letters surrounding it were not helpful in determining the context of the number. I then switched to letter frequency analysis, as there should be few words that contain the letter x, q, or z. Since there was only one instance of the letter z, guessing common words that contained one z led to solving the cipher.

Cipher 3

a. Substitution cipher with key:

`ABCDEFGHIJKLMNOPQRSTUVWXYZ | abcdefghijklmnopqrstuvwxyz`

`kxvmcnophqrszyijadlegwbuft | QWERTYUIOPASDFGHJKLZXCVBNM`

Plaintext:

the common house martin is a migratory passerine bird of the swallow family which breeds in europe, north africa and temperate asia, and winters in sub-saharan africa and tropical asia. it feeds on insects which are caught in flight. it has a blue head and upperparts, white rump and pure white underparts, and is found in both open country and near human habitation. it is a noisy species, especially at its breeding colonies. it is similar in appearance to the two other martin species of the delichon genus, which are endemic to eastern and southern asia. its scientific name (delichon urbicum) and common name both relate to its use of man-made structures. it builds a closed cup nest from mud pellets under eaves or similar locations on buildings usually in colonies, but sometimes fouling below nests can be a problem. it is hunted by the eurasian hobby, and like other birds is affected by internal parasites and external fleas and mites, although its large range and population mean that it is not threatened globally. it's proximity to man has led to some cultural and literary references, including a description in macbeth.

the common house martin was originally a cliff and cave nester, and some cliff-nesting colonies still exist, with the nests built below an overhanging rock. it now largely uses human structures such as bridges and houses. unlike the barn swallow, it uses the outside of inhabited buildings, rather than the inside of barns or stables. the nests are built at the junction of a vertical surface and an overhang, such as on house eaves, so that they may be strengthened by attachment to both planes.

breeding birds return to europe between april and may, and nest building starts between late march in north africa and mid-june in lapland. the nest is a neat closed convex cup fixed below a suitable ledge, with a narrow opening at the top. it is constructed by both sexes with mud pellets collected in their beaks, and lined with grasses, hair or other soft materials. the mud, added in successive layers, is collected from ponds, streams or puddles. house sparrows frequently attempt to take over the nest during construction, with the house martins rebuilding elsewhere if they are successful. the entrance at the top of the cup is so small once it is complete that sparrows cannot take over the nest.

the common house martin has been regularly recorded as hybridising with the barn swallow, this being one of the most common passerine interspecific crosses. the frequency of this hybrid has led to suggestions that delichon is not sufficiently separated genetically from hirundo to be considered a separate genus.

b. The repetition at the start of each paragraph in the encoded message suggested that it was a substitution cipher. I implemented a sliding window on the unstripped ciphertext using regular expression to calculate the most frequent single letter, 2-letter, and 3-letter frequencies among the words:

Letter	Count	Frequency
Single letter frequency		
T	255	0.11849442379182157
Z	195	0.09061338289962825
Q	180	0.08364312267657993
L	176	0.08178438661710037
O	169	0.07853159851301116
F	168	0.07806691449814127
G	134	0.062267657992565055
K	129	0.05994423791821561
I	100	0.046468401486988845
S	95	0.04414498141263941
R	83	0.03856877323420074
E	80	0.03717472118959108
X	79	0.03671003717472119
D	56	0.026022304832713755
W	51	0.023698884758364312
H	44	0.020446096654275093
Y	39	0.01812267657992565
U	35	0.016263940520446097
V	30	0.013940520446096654
N	28	0.013011152416356878
C	10	0.004646840148698885
A	6	0.002788104089219330 7
B	6	0.002788104089219330 7

```
P       2       0.0009293680297397769
J       2       0.0009293680297397769
```

2-gram frequency

```
('Z', 'I') 56 0.026022304832713755
('O', 'F') 41 0.019052044609665426
('Q', 'F') 41 0.019052044609665426
('T', 'L') 38 0.017657992565055763
('I', 'T') 36 0.016728624535315983
('Q', 'K') 31 0.014405204460966542
('O', 'Z') 31 0.014405204460966542
('F', 'R') 29 0.013475836431226766
('G', 'F') 28 0.013011152416356878
('T', 'K') 28 0.013011152416356878
('Z', 'T') 27 0.01254646840148699
('K', 'T') 24 0.011152416356877323
('Q', 'Z') 23 0.010687732342007435
('T', 'R') 23 0.010687732342007435
('L', 'Z') 23 0.010687732342007435
('L', 'T') 21 0.009758364312267658
('F', 'T') 21 0.009758364312267658
('T', 'F') 19 0.008828996282527882
('Z', 'O') 18 0.008364312267657992
('O', 'L') 18 0.008364312267657992
```

3-gram frequency

```
('Z', 'I', 'T') 34 0.015799256505576207
('Q', 'F', 'R') 23 0.010687732342007435
('O', 'F', 'U') 14 0.006505576208178439
('F', 'T', 'L') 11 0.005111524163568773
('T', 'L', 'Z') 11 0.005111524163568773
('X', 'L', 'T') 10 0.004646840148698885
('Q', 'S', 'S') 9 0.004182156133828996
('I', 'G', 'X') 8 0.0037174721189591076
('Q', 'K', 'Z') 8 0.0037174721189591076
('Z', 'T', 'K') 8 0.0037174721189591076
('G', 'X', 'L') 7 0.0032527881040892194
('Q', 'Z', 'T') 7 0.0032527881040892194
('H', 'Q', 'K') 7 0.0032527881040892194
('G', 'Z', 'I') 7 0.0032527881040892194
('Z', 'O', 'G') 7 0.0032527881040892194
('O', 'G', 'F') 7 0.0032527881040892194
('R', 'O', 'F') 7 0.0032527881040892194
('W', 'X', 'O') 7 0.0032527881040892194
('X', 'O', 'S') 7 0.0032527881040892194
('Z', 'T', 'R') 7 0.0032527881040892194
```

I found overrepresentation of the ZIT and QFR tuples and T as the most frequently observed single letter at ~12%. Substitution of ZIT and QFR to ‘the’ and ‘and’, respectively, and additional substitution of common letters resulted in the above cipher key. For example, matching OL to the word ‘is’.

```
dmap = {}
dmap['Z'] = 't'
dmap['I'] = 'h'
dmap['T'] = 'e'
dmap['Q'] = 'a'
dmap['F'] = 'n'
dmap['R'] = 'd'
dmap['O'] = 'i'
dmap['L'] = 's'
```

And comparing with the decrypted plaintext with the undecrypted letters in both the plaintext and ciphertext. For example, after the above substitutions:

```
        unmapped CRYPTKEYS ['A', 'B', 'C', 'D', 'E', 'G', 'H', 'J', 'K', 'M', 'N', 'P', 'S', 'U',
'V', 'W', 'X', 'Y']
        unmapped plaintext ['b', 'c', 'f', 'g', 'j', 'k', 'l', 'm', 'o', 'p', 'q', 'r', 'u', 'v',
'w', 'x', 'y', 'z']
```

```
the EGDDGn hGXse DaKtin is a DiUKatGKN HasseKine WiKd GY the sVaSSGV YaDiSN VhiEh WKeeds in eXKGHe,
nGKth aYKiEa and teDHeKate asia, and VinteKs in sXW-sahaKan aYKiEa and tKGHiEaS asia. it Yeeds Gn
inseEts VhiEh aKe EaXUht in YSiUht. it has a WSXe head and XHHeKHaKts, Vhite KXDH and HXKe Vhite
XndeKHaKts, and is YGXnd in WGth GHen EGXntKN and neaK hXDan haWitatiGn. it is a nGisN sHeEies,
esHeEiaSSN at its WKeedinU EGSGnies. it is siDiSaK in aHHeaKanEe tG the tVG GtheK DaKtin sHeEies GY
the deSiEhGn UenXs, VhiEh aKe endeDiE tG easteKn and sGXtheKn asia. its sEientiYiE naDe (deSiEhGn
XKWiEXD) and EGDDGn naDe WGth KeSate tG its Xse GY Dan-Dade stKXEtXKes. it WXiSds a ESGsed EXH nest
YKGD DXd HeSSets XndeK eaCes GK siDiSaK SGEatiGns Gn WXiSdinUs XsXaSSN in EGSGnies, WXt sGDetiDes
YGXSinU WeSGV nests Ean We a HKGWSeD. it is hXnted WN the eXKasian hGWWN, and SiAe GtheK WiKds is
aYYeEted WN inteKnaS HaKasites and eBteKnaS YSeas and Dites, aSthGXUh its SaKUe KanUe and HGHXSatiGn
Dean that it is nGt thKeatened USGWaSSN. it's HKGBiDitN tG Dan has Sed tG sGDe EXStXKaS and SiteKaKN
KeYeKenEes, inESXdinU a desEKiHtiGn in DaEWeth. … the EGDDGn hGXse DaKtin has Ween KeUXSaKSN KeEGKded
as hNWKidisinU Vith the WaKn sVaSSGV, this WeinU Gne GY the DGst EGDDGn HasseKine inteKsHeEiYiE
EKGsses. the YKeJXenEN GY this hNWKid has Sed tG sXUUestiGns that deSiEhGn is nGt sXYYiEientSN
seHaKated UenetiEaSSN YKGD hiKXndG tG We EGnsideKed a seHaKate UenXs.
```

c. Since the substitution cipher was the first one I decrypted, the most difficult part was implementing the n-grams using a sliding window using the text with punctuations. Once I solved that problem, guessing which letter in the ciphertext matched which letter in the plaintext was achieved iteratively.

Cipher 4

a. LFSR4 cipher with unknown key

b. After determining the cipher used on the other ciphertexts, this was assumed to be the LFSR4 cipher. Assuming that $m = 4$, the goal was to find a portion of plaintext of length $2m = 8$ to decipher the text. I initially guessed three portions of the ciphertext and possible plaintext from the ciphertext, for example:

```
QJ ZVO 70'I. XVI 70'A SEZI
in the 70's. the 70's were
by the 70’s. the 70’s were

ZXA  70'U FCJ BDQ 60'M
the  70's and the 60's
the  70’s but the 60’s
the  70’s but not 60’s
the  70’s and not 60’s
the  70’s not the 60’s

49% YB TRK CENNV'G FWLGFQVEAB. 49% YB DBU WIBVH'G NQFIBGNSEX
49% of the world's population. 49% of the world's population

WN 1066. AADHWGO JDC QYVMASBUV WRDWDMH IZWFGJT EF 1066. EMHPQEU JBG UABWMKDIB
in 1066. william the conqueror invaded England in 1066. william the conqueror
```

Using an string of 8 characters with which I had the highest confidence of being correct:

```
ZVO 70'I. XVI 70'A
the 70's. the 70's
```

and found $Z_1 \dots Z_m \dots Z_{2m}$:

$$Y = (X - Z) (\text{mod } 26)$$

```
y [ Z,  V,  O,  I,  X,  V, I,  A]
Y [25, 21, 14,  8, 23, 21, 8,  0]

x [ t,  h,  e,  s,  t,  h, e,  s]
X [19,  7,  4, 18, 19,  7, 4, 18]

Z [ 6, 14, 10, 16,  4, 14, 4,  8]
```

I used this to find the set of coefficients:

$$(z_5, z_6, z_7, z_8) = (c_0, c_1, c_2, c_3) \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \\ z_2 & z_3 & z_4 & z_5 \\ z_3 & z_4 & z_5 & z_6 \\ z_4 & z_5 & z_6 & z_7 \end{pmatrix}$$

$$= (6,14,10,16) = (c_0, c_1, c_2, c_3) \begin{pmatrix} 6 & 14 & 10 & 16 \\ 14 & 10 & 16 & 4 \\ 10 & 16 & 4 & 14 \\ 16 & 4 & 14 & 4 \end{pmatrix}$$

However, I found that the square Z matrix did not have a modular inverse. For substrings of text that were longer than the period 2^{m-1}, I implemented a sliding window to test all possible substrings of length $2m$. I found that even for these lengths of text, I was unable to find the coefficients. This was true for additional the remaining substrings I had found, which included lengths of text that were greater than the period 2^{m-1}.

c. The most difficult part in deciphering LFSR is finding plaintext that have lengths of $2m$ to generate a period. Another difficulty I had was calculating the inverse matrix from the square Z matrix, as I was unable to find a string of plaintext that resulted in an invertible matrix.

d. The number of keys for a brute force for the LFSR cipher is 26^{2m}, since there are 26^m possible values for both the coefficients $c_0 c_1 c_2 \dots c_m$ and $z_1 z_2 z_3 \dots z_m$ values of the initial vector key. All other z_{m+i} can then be derived from the initial coefficients and key vector by linear combination.

Cipher 5

a. Hill cipher with key:

$$K = \begin{bmatrix} 4 & 3 \\ 7 & 8 \end{bmatrix}, K^{-1} = \begin{bmatrix} 22 & 21 \\ 23 & 24 \end{bmatrix}$$

Plaintext:

```
she sells seashells by the seashore. the shells she sells are surely seashells.
so if she sells shells on the seashore, i'm sure she sells seashore shells.
```

b. After we were given that cipher 5 was a Hill cipher with $m = 2$ in class, I guessed the plaintext for the cipher as the tongue twister "she sells seashells by the seashore". This was then used to setup the 2×2 matrices for both the plaintext and ciphertext and calculate K:

$$\begin{bmatrix} R & G \\ M & A \end{bmatrix} = \begin{bmatrix} s & h \\ e & s \end{bmatrix} K$$
$$\begin{bmatrix} 17 & 6 \\ 12 & 0 \end{bmatrix} = \begin{bmatrix} 18 & 7 \\ 4 & 18 \end{bmatrix} K$$
$$\begin{bmatrix} 18 & 7 \\ 4 & 18 \end{bmatrix}^{-1} = No\ modular\ inverse$$

Since the first substring did not have a modular inverse, I tried the second one:

$$\begin{bmatrix} P & W \\ O & V \end{bmatrix} = \begin{bmatrix} e & l \\ l & s \end{bmatrix} K$$
$$\begin{bmatrix} 15 & 22 \\ 14 & 21 \end{bmatrix} = \begin{bmatrix} 4 & 11 \\ 11 & 18 \end{bmatrix} K$$
$$X^{-1} = \begin{bmatrix} 4 & 11 \\ 11 & 18 \end{bmatrix}^{-1} = \begin{bmatrix} 6 & 5 \\ 5 & 10 \end{bmatrix}$$
$$K = \begin{bmatrix} 6 & 5 \\ 5 & 10 \end{bmatrix} \begin{bmatrix} 15 & 22 \\ 14 & 21 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 7 & 8 \end{bmatrix}$$
$$K^{-1} = \begin{bmatrix} 4 & 3 \\ 7 & 8 \end{bmatrix}^{-1} = \begin{bmatrix} 22 & 21 \\ 23 & 24 \end{bmatrix}$$

The values of K and K^{-1} were confirmed for additional substrings:

```
x ('h', 'e', 'l', 'l')
y ('E', 'B', 'R', 'R')
x inverse Matrix([[9, 18], [17, 1]])
k Matrix([[4, 3], [7, 8]])
k inverse Matrix([[22, 21], [23, 24]])

x ('n', 't', 'h', 'e')
y ('D', 'J', 'E', 'B')
x inverse Matrix([[16, 15], [11, 13]])
k Matrix([[4, 3], [7, 8]])
k inverse Matrix([[22, 21], [23, 24]])
```

This was also applied back to the instance were the modular inverses could not be determined:

$$\begin{bmatrix} R & G \\ M & A \end{bmatrix} = \begin{bmatrix} s & h \\ e & s \end{bmatrix} K$$
$$\begin{bmatrix} 17 & 6 \\ 12 & 0 \end{bmatrix} = \begin{bmatrix} 18 & 7 \\ 4 & 18 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 7 & 8 \end{bmatrix}$$

c. The difficulty of this cipher lies in guessing the plaintext and determining m, the size of the matrix. Without knowing either, there are 26^{m^2} possible keys, minus those that do not have modular inverses. Another difficulty was after determining the plaintext, many of the X matrices could not be invertible. Of the 31 substrings of length $2m$, only 8 had modular inverses.