



COMP 150: Project Day 2 Week 14, Lecture 21



Agenda

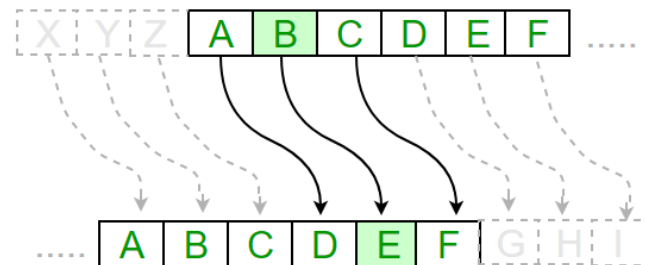
1. Attendance
2. The Shift Cypher
3. Frequency Analysis
4. The Word Cypher
5. Let's break it!

Project Timelines

Dates	Contents
Week 14: April 13, 15, 17	Code Project 13: Project Day 1 15: Project Day 2 17: Project Day 3, Select Report Topic
Week 15: April 20, 22, 24	Code Project 20: Project Day 4 22: Project Day 5 24: MT III -- Same format as MT II
Final: May 1, 4:15-6:16 PM	We are not going to meet on this day. Code Project Due Reports due

The Shift Cypher

- Encode each letter with a number
- Shift each letter by some number from 0 to 25
 - Ex, A shifted by 0 is A, A shifted by 1 is B, etc
- [Interactive wheel](#)



A	B	C	D	E	F	G	H	I	J	K	L	M
1	2	3	4	5	6	7	8	9	10	11	12	13

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
14	15	16	17	18	19	20	21	22	23	24	25	26

Frequency Analysis

- Since each letter always translates to another letter, we know:

Frequency of letter in plaintext == Frequency of letter in cyphertext

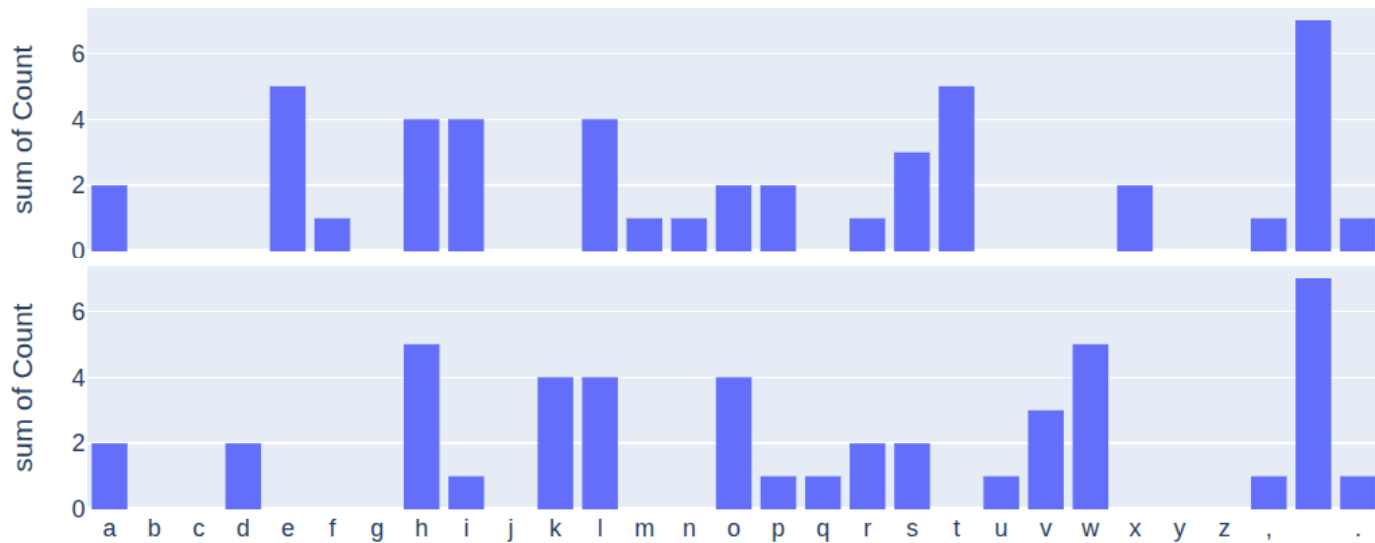
Example:

Key: 3

Plaintext: Hello, this is the plaintext for this example.

Cyphertext: Kloor, wklv lv wkh sodlqwhaw iru wklv hadpsoh.

Character Counts in Plaintext



Frequency Analysis

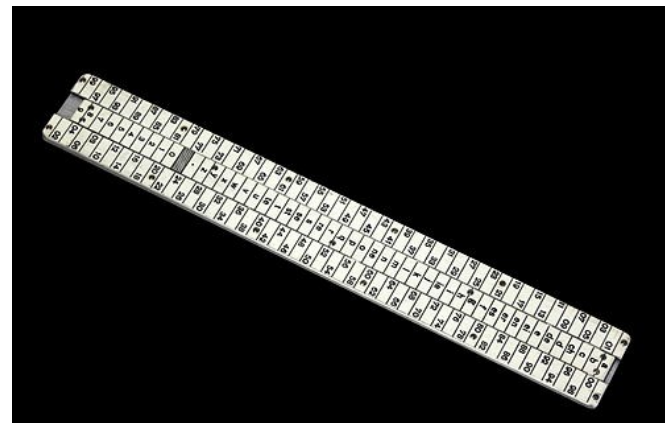
With sufficiently long cyphertext, frequency analysis is easy:

1. Take the cyphertext
2. Run frequency analysis on it
3. Find the *shifted* “E” and “T” - which will roughly be 13% and 9% of the letters
4. Count the distance between E’s shift and T’s shift to get the key
5. If wrong, go with the next best guess!

E	12.7 %	M	2.4 %
T	9.1 %	W	2.4 %
A	8.2 %	F	2.2 %
O	7.5 %	G	2.0 %
I	7.0 %	Y	2.0 %
N	6.7 %	P	1.9 %
S	6.3 %	B	1.5 %
H	6.1 %	V	1.0 %
R	6.0 %	K	0.8 %
L	4.0 %	J	0.2 %
D	4.3 %	X	0.2 %
C	2.8 %	Q	0.1 %
U	2.8 %	Z	0.1 %

Word Cypher

- Also known as the Vigenère cipher
 - Misattributed to the mathematician Blaise de Vigenere
 - Actually invented by Giovan Battista Bellaso in 1553
- Instead of encrypting using a single key, use a sequence of keys
- Much, much harder to break!
- Especially for single short messages.



Word Cypher

1. Convert letters to numeric representation
2. Convert key (sequence of values) to numeric representation
3. Shift each sequence of letters by the key, repeating the key as needed to cover the plaintext/cyphertext

Word Cypher

Plaintext: Hello, this is the plaintext for this example.

Key: BAT $\rightarrow$ 1, 0, 19

Ciphertext: Ieemo, miil js mie imabotxyt ypr miil fxtnp ef.

Plain H	Plain E	Plain L	Plain L	Plain O	Plain T	Plain H	Plain I	Plain S	Plain I	Plain S	Plain T	Plain H	Plain E	
Key B	Key A	Key T	Key B	Key A	Key T	Key B	Key A	Key T	Key B	Key A	Key T	Key B	Key A	
Cipher I	Cipher E	Cipher E	Cipher M	Cipher O	Cipher M	Cipher I	Cipher I	Cipher L	Cipher J	Cipher S	Cipher M	Cipher I	Cipher E	

Plain P	Plain L	Plain A	Plain I	Plain N	Plain T	Plain E	Plain X	Plain T	Plain F	Plain O	Plain R	Plain T	Plain H	Plain I	Plain S
Key T	Key B	Key A	Key T	Key B	Key A	Key T	Key B	Key A	Key T	Key B	Key A	Key T	Key B	Key A	Key T
Cipher I	Cipher M	Cipher A	Cipher B	Cipher O	Cipher T	Cipher X	Cipher Y	Cipher T	Cipher Y	Cipher P	Cipher R	Cipher M	Cipher I	Cipher I	Cipher L

Breaking the Word Cypher

Strategy is the same as for the shift cypher with two steps:

1. Figure out how long the key is
2. Chunk the cyphertext according to the key length, then treat each like an individual shift cypher

Breaking the Word Cypher

Strategy is the same as for the shift cypher with two steps:

1. Figure out how long the key is
2. Chunk the cyphertext according to the key length, then treat each like an individual shift cypher



Breaking the Word Cypher

Strategy is the same as for the shift cypher with two steps:

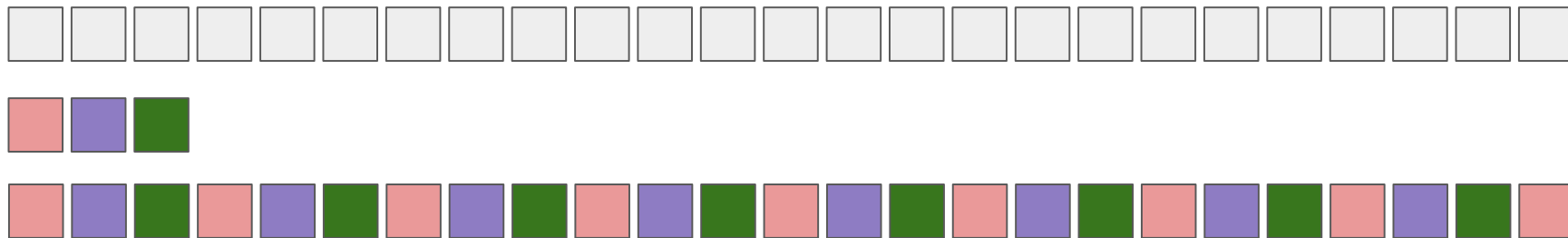
1. Figure out how long the key is
2. Chunk the cyphertext according to the key length, then treat each like an individual shift cypher



Breaking the Word Cypher

Strategy is the same as for the shift cypher with two steps:

1. Figure out how long the key is
2. Chunk the cyphertext according to the key length, then treat each like an individual shift cypher



Breaking the Word Cypher



Breaking the Word Cypher



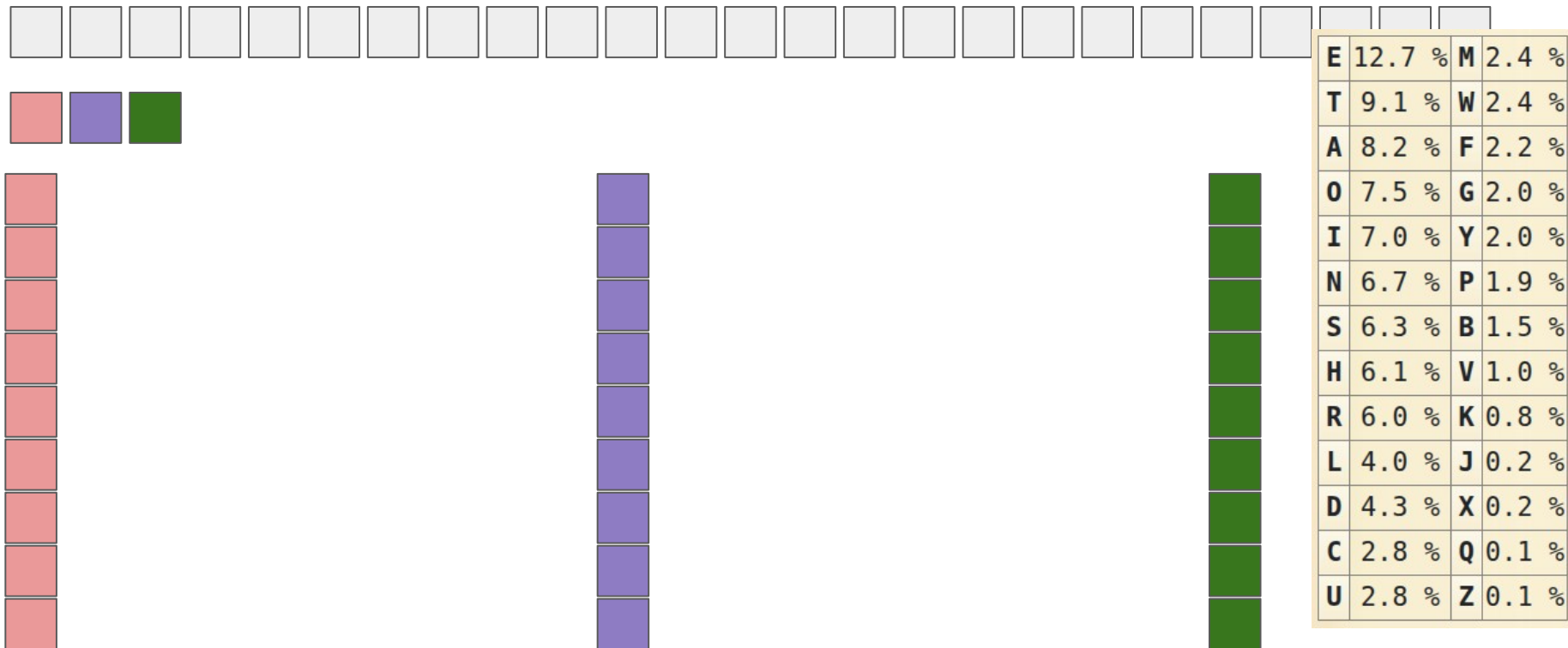
Breaking the Word Cypher

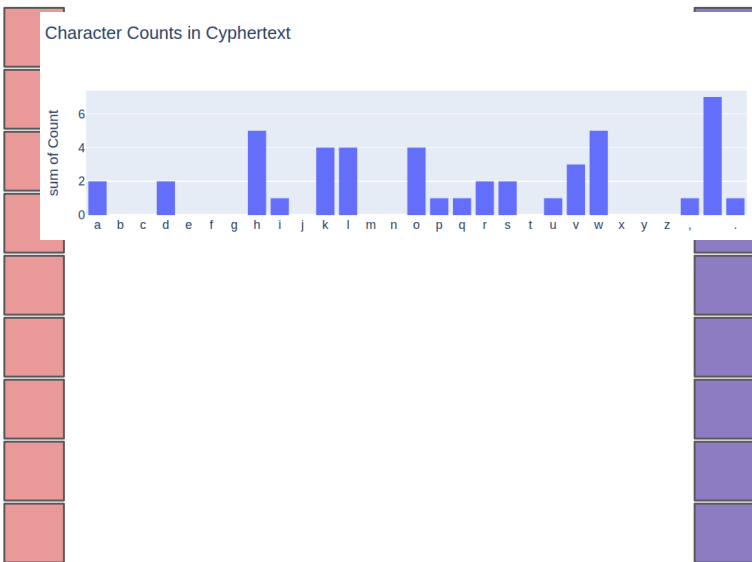


Breaking the Word Cypher



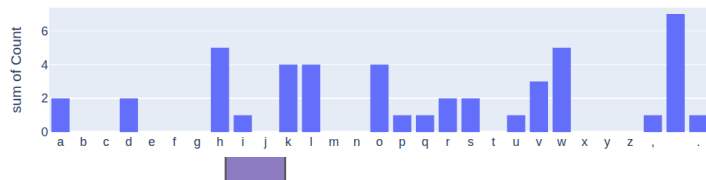
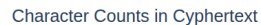
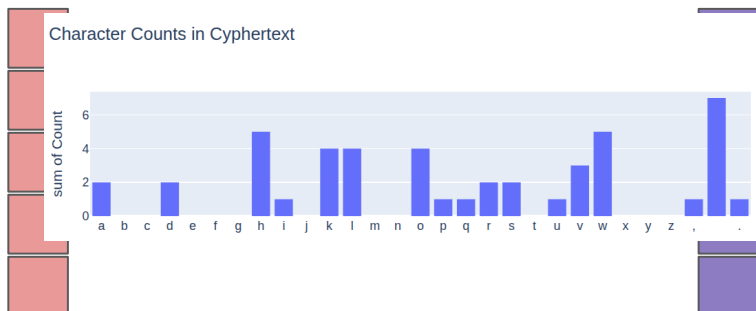
Breaking the Word Cypher



[illegible]

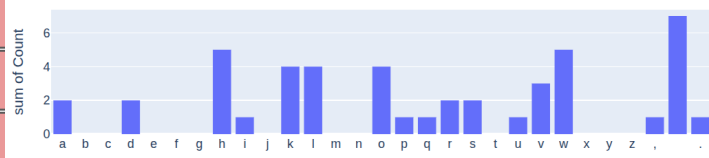
1
2
3
4
5
6
7
8
9

																							F	12	7	%	M
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	---	----	---	---	---



E	12.7 %	M	2.4 %
T	9.1 %	W	2.4 %
A	8.2 %	F	2.2 %
O	7.5 %	G	2.0 %
I	7.0 %	Y	2.0 %
N	6.7 %	P	1.9 %
S	6.3 %	B	1.5 %
H	6.1 %	V	1.0 %
R	6.0 %	K	0.8 %
L	4.0 %	J	0.2 %
D	4.3 %	X	0.2 %
C	2.8 %	Q	0.1 %
U	2.8 %	Z	0.1 %

																							F	12	7	%	M	1
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	---	----	---	---	---	---

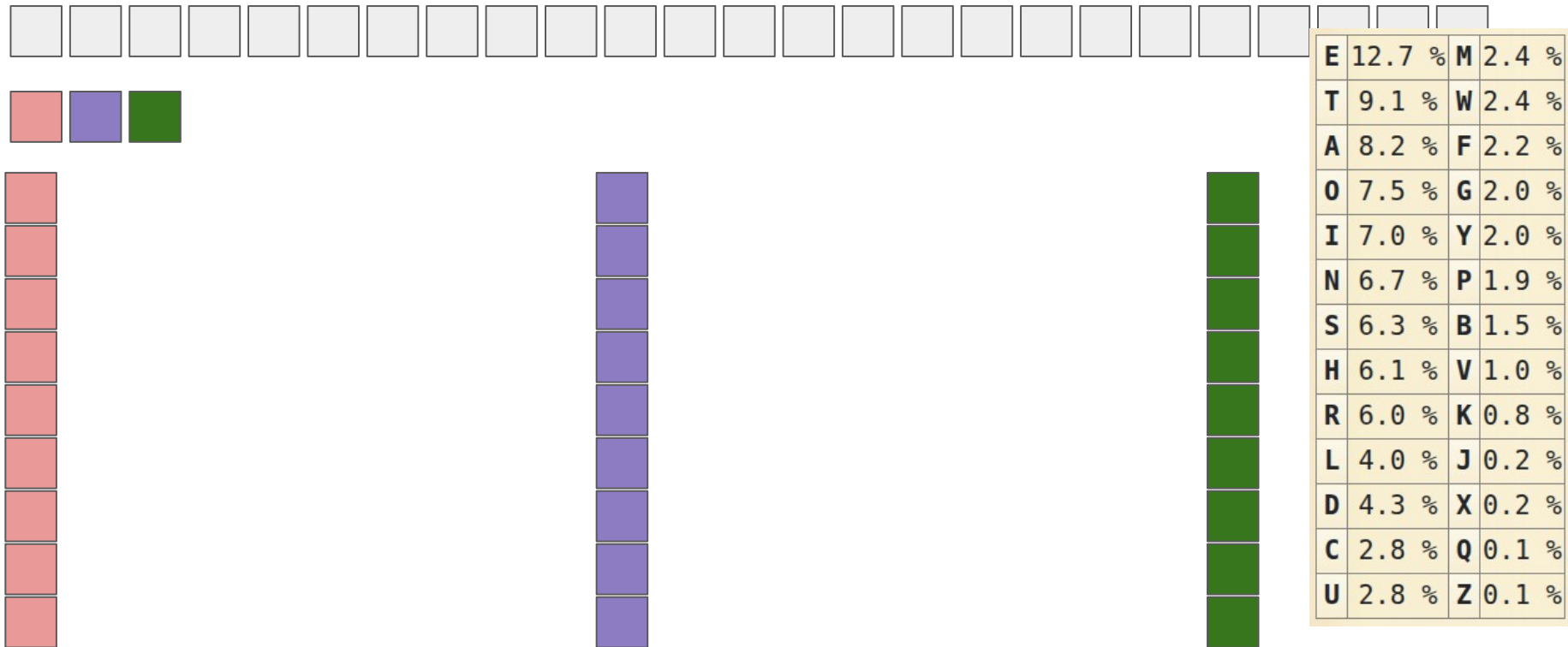


Character	sum of Count
a	2
b	0
c	0
d	2
e	0
f	0
g	0
h	5
i	1
j	0
k	4
l	4
m	0
n	0
o	4
p	1
q	1
r	2
s	2
t	0
u	1
v	3
w	5
x	0
y	0
z	0
,	1
.	1

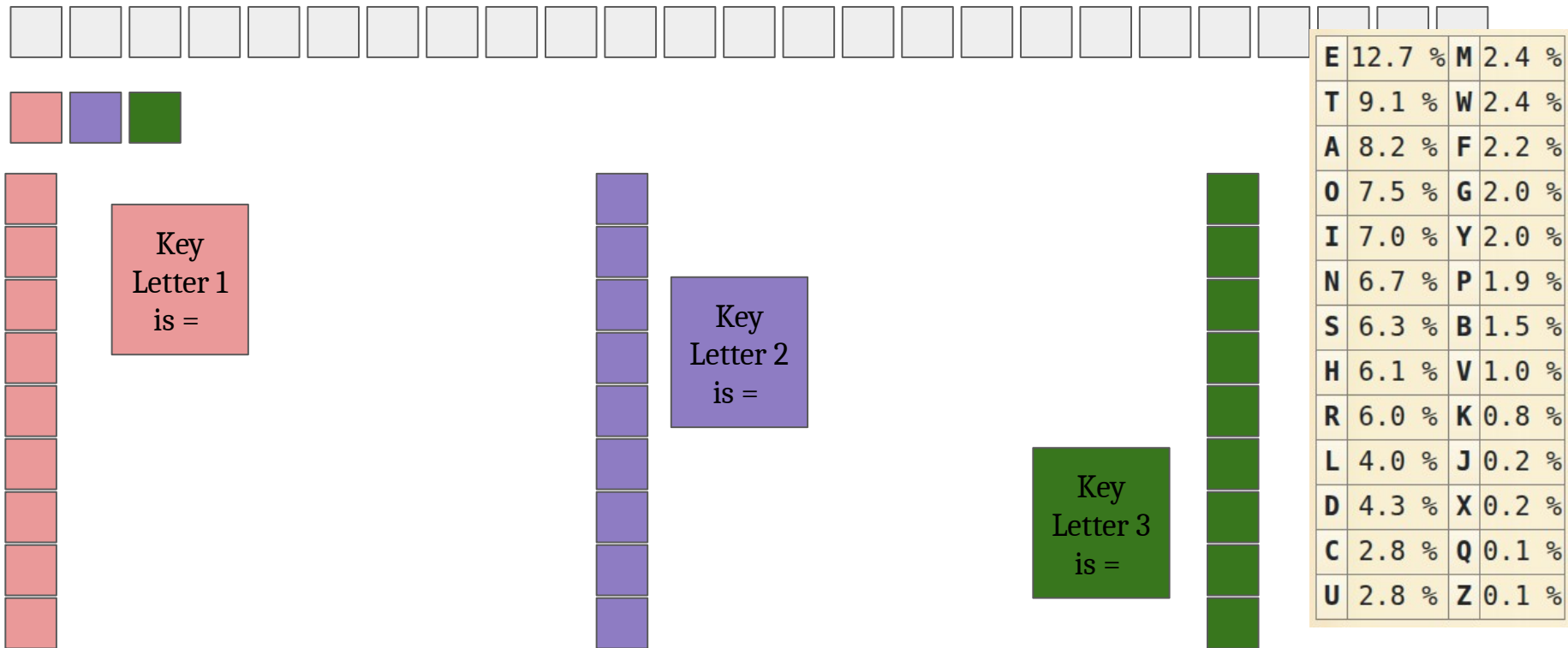
Character	sum of Count
a	2
b	0
c	0
d	2
e	0
f	0
g	0
h	5
i	1
j	0
k	4
l	4
m	0
n	0
o	4
p	1
q	1
r	2
s	2
t	0
u	1
v	3
w	5
x	0
y	0
z	0
,	1
.	1

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72
73	74	75	76	77	78
79	80	81	82	83	84
85	86	87	88	89	90
91	92	93	94	95	96
97	98	99	100	101	102
103	104	105	106	107	108
109	110	111	112	113	114
115	116	117	118	119	120
121	122	123	124	125	126
127	128	129	130	131	132
133	134	135	136	137	138
139	140	141	142	143	144
145	146	147	148	149	150
151	152	153	154	155	156
157	158	159	160	161	162
163	164	165	166	167	168
169	170	171	172	173	174
175	176	177	178	179	180
181	182	183	184	185	186
187	188	189	190	191	192
193	194	195	196	197	198
199	200	201	202	203	204
205	206	207	208	209	210
211	212	213	214	215	216
217	218	219	220	221	222
223	224	225	226	227	228
229	230	231	232	233	234
235	236	237	238	239	240
241	242	243	244	245	246
247	248	249	250	251	252
253	254	255	256	257	258
259	260	261	262	263	264
265	266	267	268	269	270
271	272	273	274	275	276
277	278	279	280	281	282
283	284	285	286	287	288
289	290	291	292	293	294
295	296	297	298	299	300
301	302	303	304	305	306
307	308	309	310	311	312
313	314	315	316	317	318
319	320	321	322	323	324
325	326	327	328	329	330
331	332	333	334	335	336
337	338	339	340	341	342
343	344	345	346	347	348
349	350	351	352	353	354
355	356	357	358	359	360
361	362	363	364	365	366
367	368	369	370	371	372
373	374	375	376	377	378
379	380	381	382	383	384
385	386	387	388	389	390
391	392	393	394	395	396
397	398	399	400	401	402
403	404	405	406	407	408
409	410	411	412	413	

Breaking the Word Cypher



Breaking the Word Cypher



Breaking the Word Cypher

- If we know the key length and the cyphertext is long enough, easy!
- What do we do if we **don't** know the key length?

More frequency analysis!

Breaking the Word Cypher

- Instead of looking for the most common letters (shifted E and T), look for repeated sequences (THE, ING, and so on)

the man and the woman retrieved the letter from the post office

- If we use the key “beads”:

bea dsb ead **sbe** adsbe adsbeadsb **ead** sbears bead **sbe** adsb eadsbe
the man and **the** woman retrieved **the** letter from **the** post office

- The key length will be a multiple of the distance between repeated sequences
- Here: the distance is 30, so we know the key length is a factor (2, 3, 5, 6, ...)

Your turn!

1. Download COMP150_project_day2.ipynb from **Piazza**
2. Upload it to Google Colab
3. Rename the file to: yourname_COMP150_project_day2.ipynb
4. Complete the code steps
5. Continue your document, yourname_COMP150_project
6. Complete analysis steps in this document under a section titled:
 - a. 1. The Word Cypher

The report and notebooks will all be due on **Fridays**.

An interactive word cypher webpage: <https://www.boxentriq.com/ciphers/vigenere-cipher>