COMP 150:
Project Day 2
Week 14, Lecture21

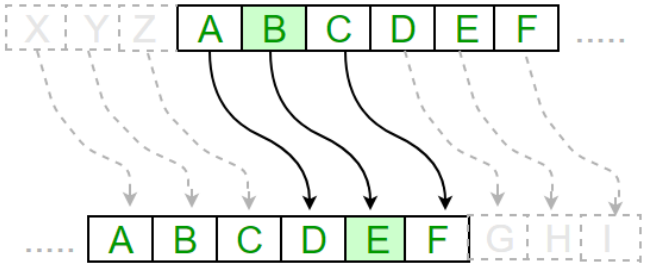
Agenda

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:

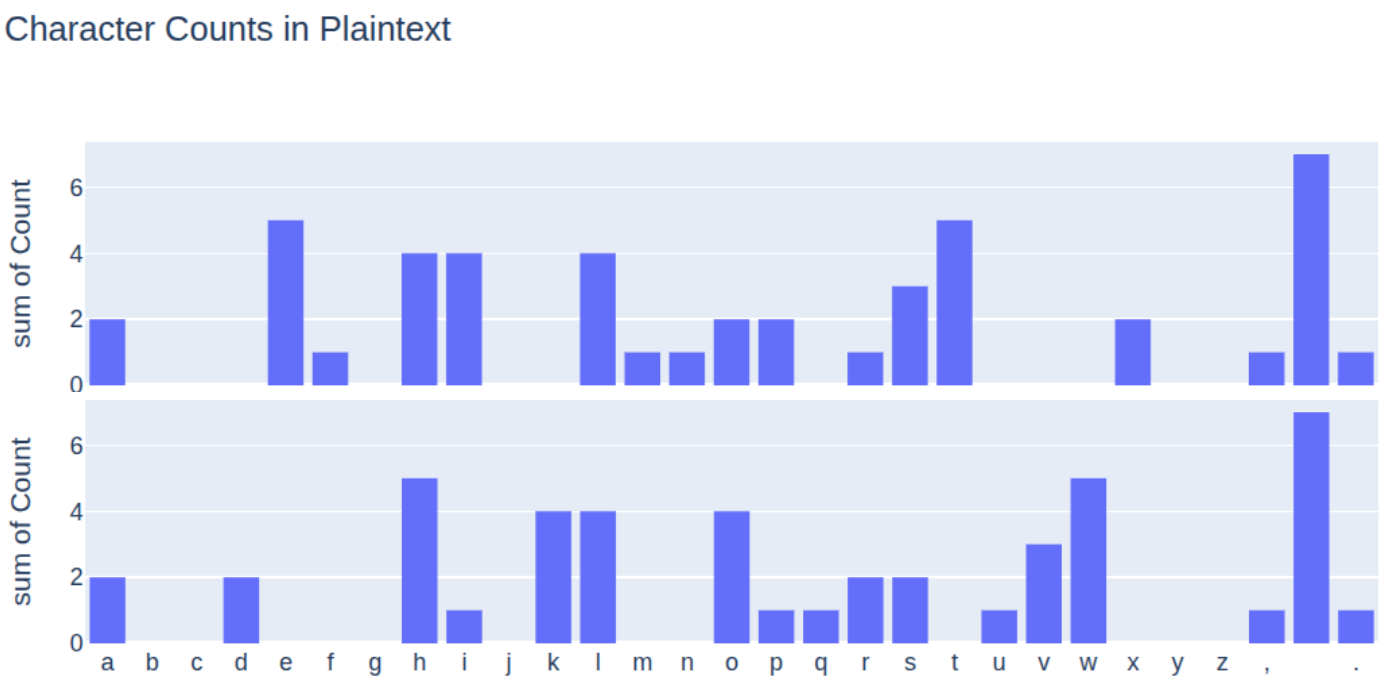
  Frequency of letter in plaintext == Frequency of letter in cyphertext

Example:

Key: 3

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

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


Character Counts in Plaintext
sum of Count
0
2
4
6
sum of Count
0
2
4
6
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 Analysis

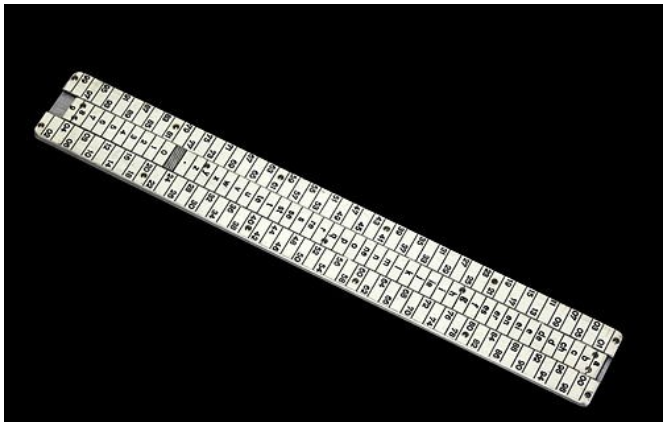
With sufficiently long cyphertext, frequency analysis is easy:

1. Take the cypehrtext
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

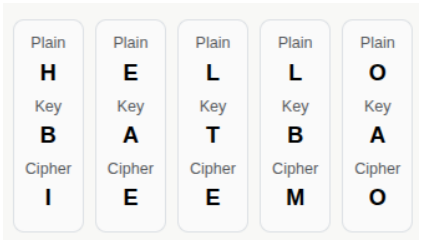
1. Convert letters to numeric representation
2. Convert key (sequence of values) to numeric representation
3. Shift each sequence of letters by the key, repeating the key as needed to cover the plaintext/cyphertext

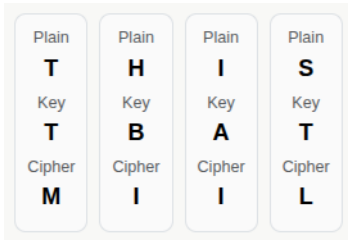
Word Cypher

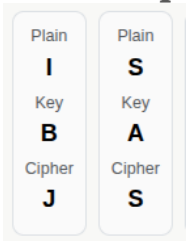
Plaintext: Hello, this is the plaintext for this example.

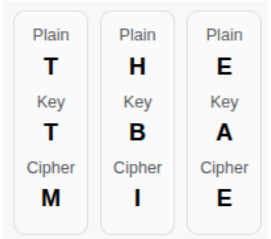
Key: BAT → 1, 0, 19

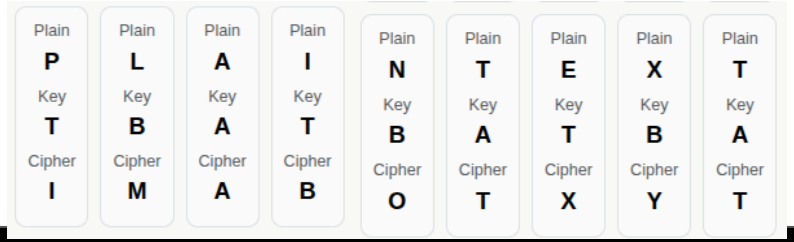
Ciphertext: Ieemo, miil js mie imabotxyt ypr miil fxtnpef.

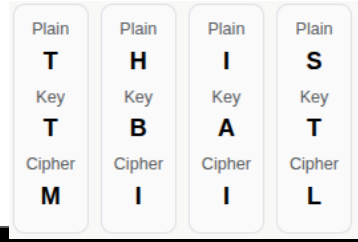












Breaking the Word Cypher

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

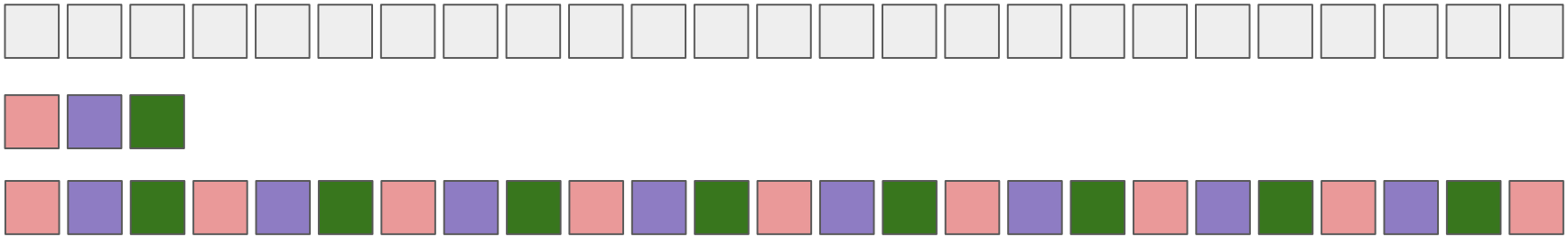
2. Chunk the cyphertext according to the key length, then treat each like an individual shift cypher

Breaking the Word Cypher

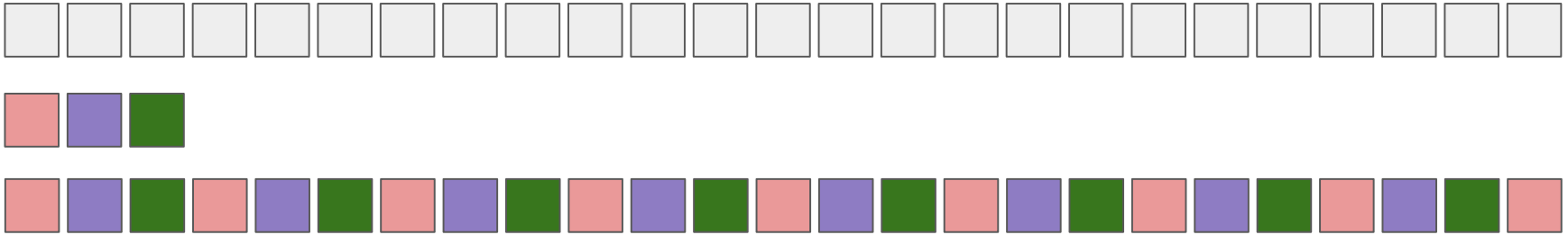
Strategy is the same as for the shift cypher with two steps:

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

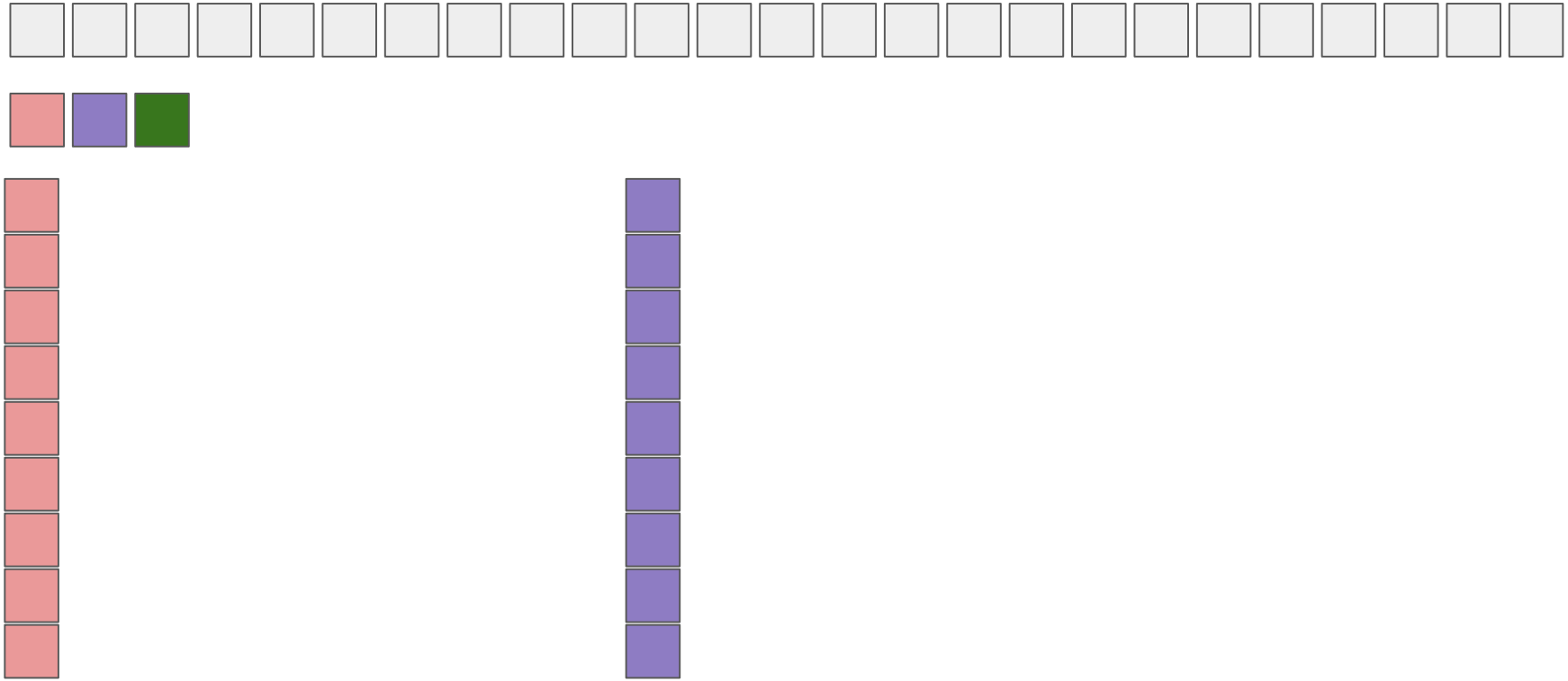


Breaking the Word Cypher

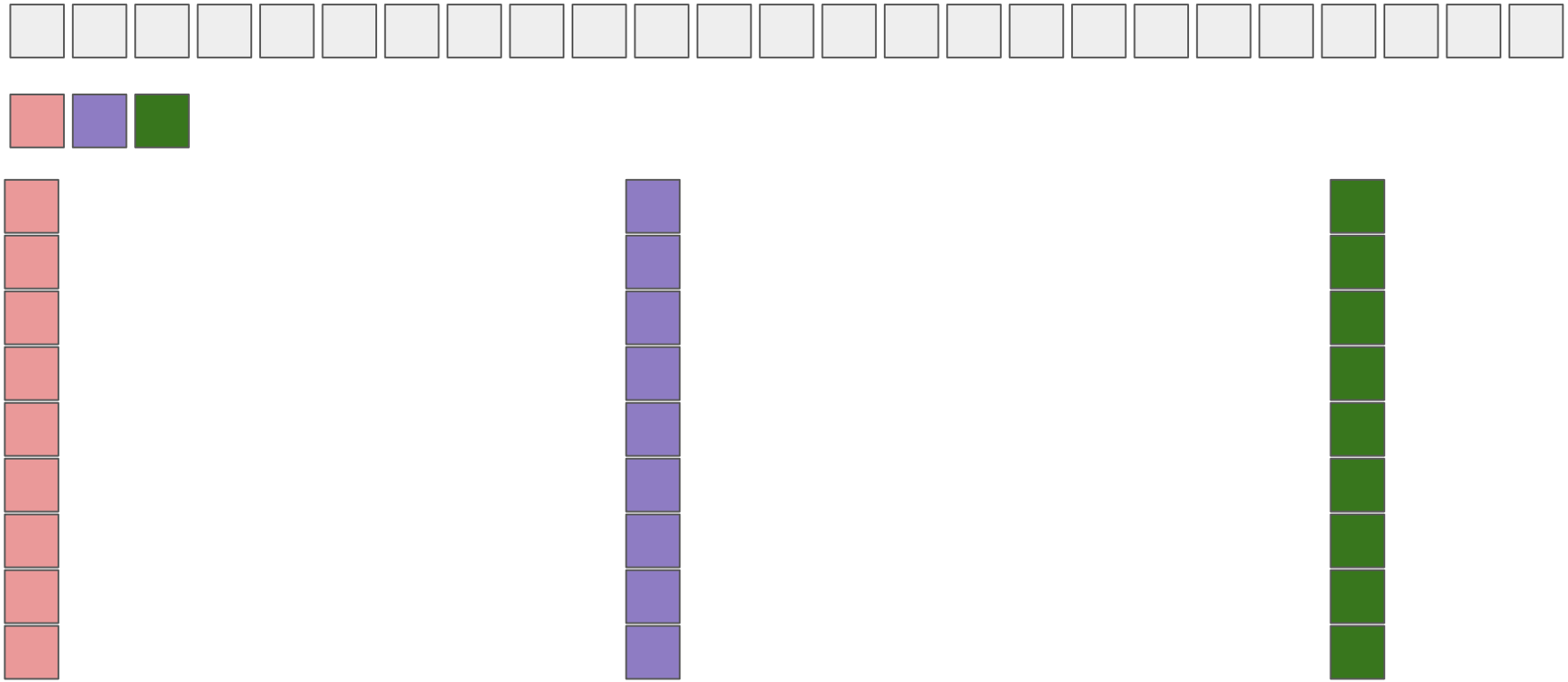


Breaking the Word Cypher

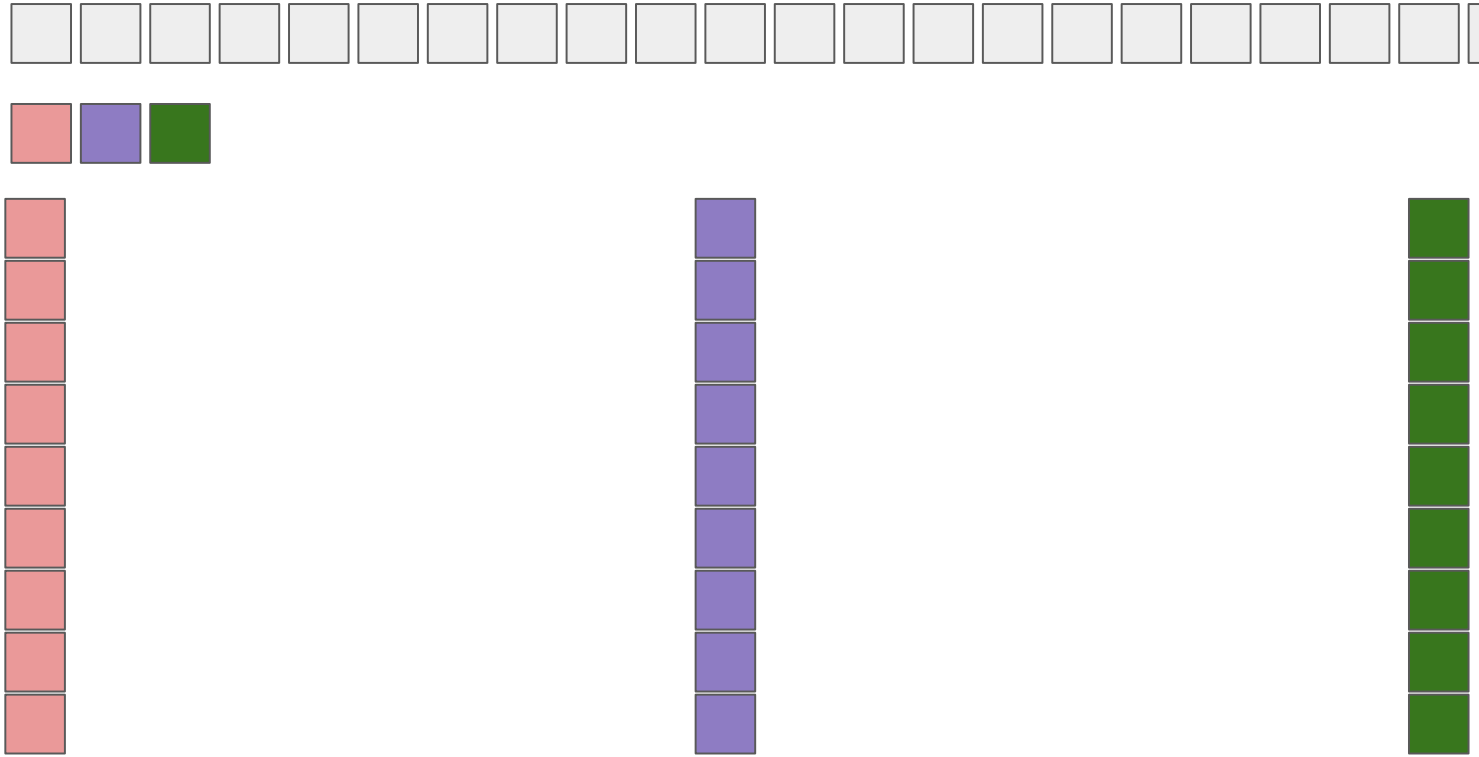
Breaking the Word Cypher



Breaking the Word Cypher

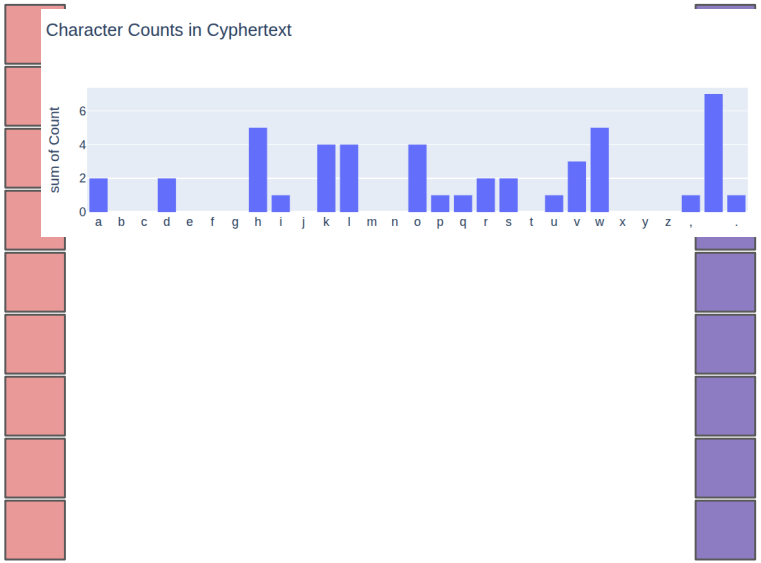


Breaking the Word Cypher



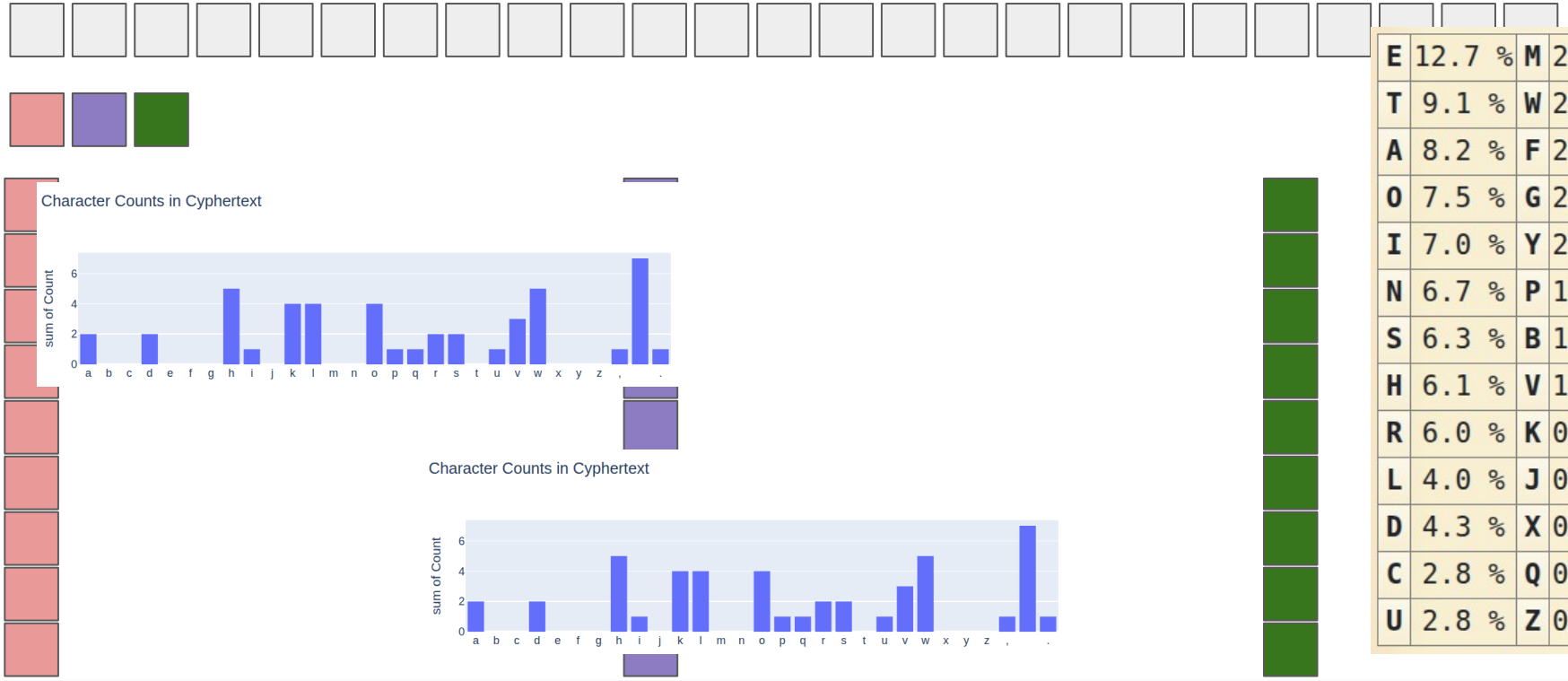
E	12.7 %	M	2.4 %
T	9.1 %	W	2.4 %
A	8.2 %	F	2.2 %
O	7.5 %	G	2.0 %
I	7.0 %	Y	2.0 %
N	6.7 %	P	1.9 %
S	6.3 %	B	1.5 %
H	6.1 %	V	1.0 %
R	6.0 %	K	0.8 %
L	4.0 %	J	0.2 %
D	4.3 %	X	0.2 %
C	2.8 %	Q	0.1 %
U	2.8 %	Z	0.1 %

Breaking the Word Cypher



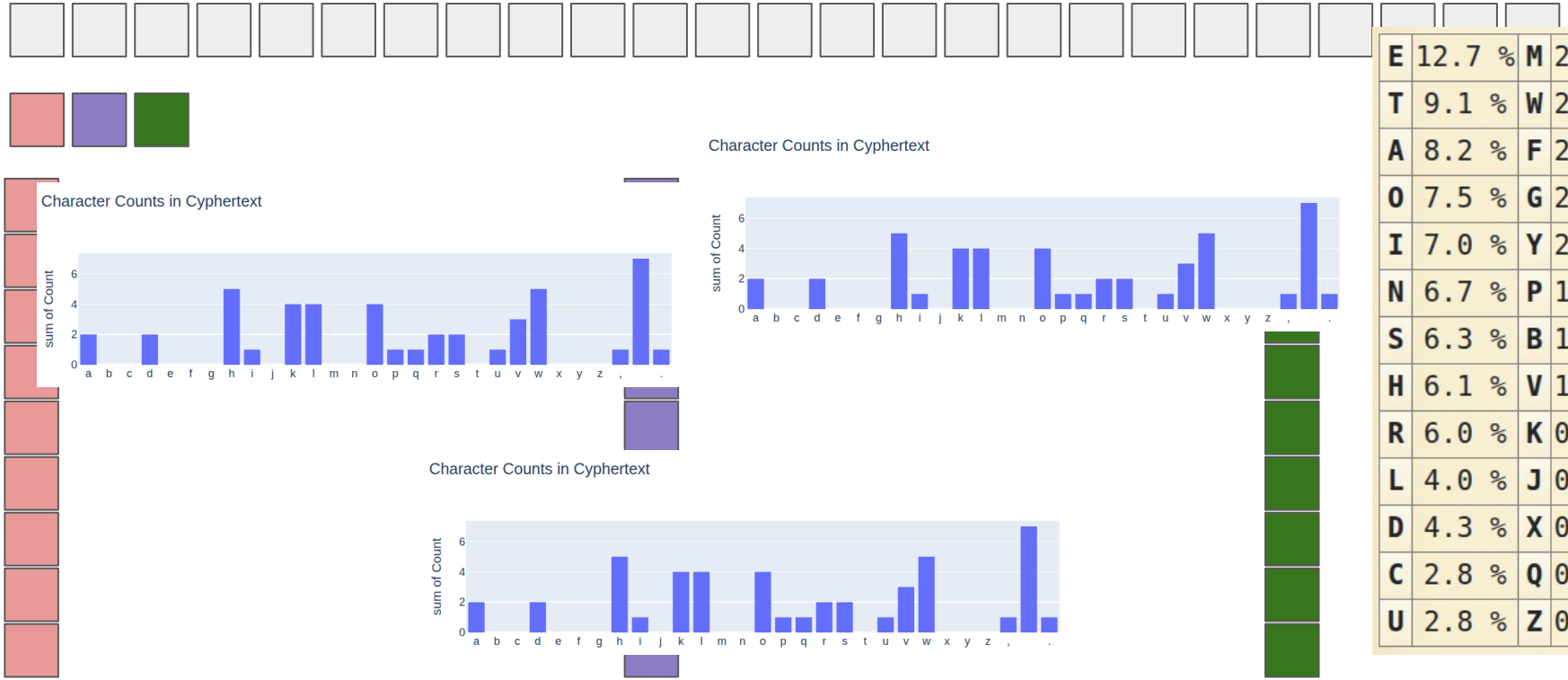
E	12.7 %	M	2.4 %
T	9.1 %	W	2.4 %
A	8.2 %	F	2.2 %
O	7.5 %	G	2.0 %
I	7.0 %	Y	2.0 %
N	6.7 %	P	1.9 %
S	6.3 %	B	1.5 %
H	6.1 %	V	1.0 %
R	6.0 %	K	0.8 %
L	4.0 %	J	0.2 %
D	4.3 %	X	0.2 %
C	2.8 %	Q	0.1 %
U	2.8 %	Z	0.1 %

Breaking the Word Cypher



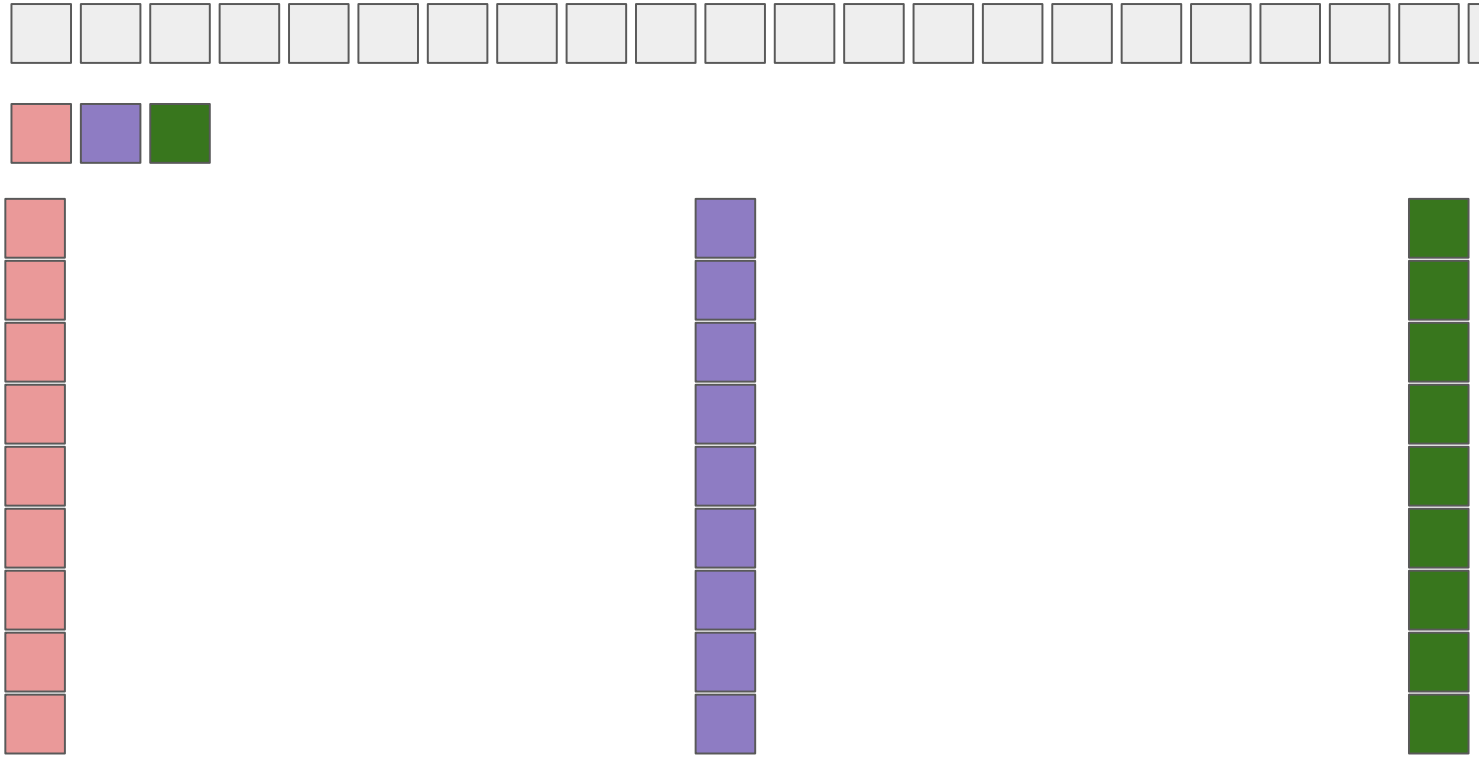
E	12.7 %	M	2.4 %
T	9.1 %	W	2.4 %
A	8.2 %	F	2.2 %
O	7.5 %	G	2.0 %
I	7.0 %	Y	2.0 %
N	6.7 %	P	1.9 %
S	6.3 %	B	1.5 %
H	6.1 %	V	1.0 %
R	6.0 %	K	0.8 %
L	4.0 %	J	0.2 %
D	4.3 %	X	0.2 %
C	2.8 %	Q	0.1 %
U	2.8 %	Z	0.1 %

Breaking the Word Cypher



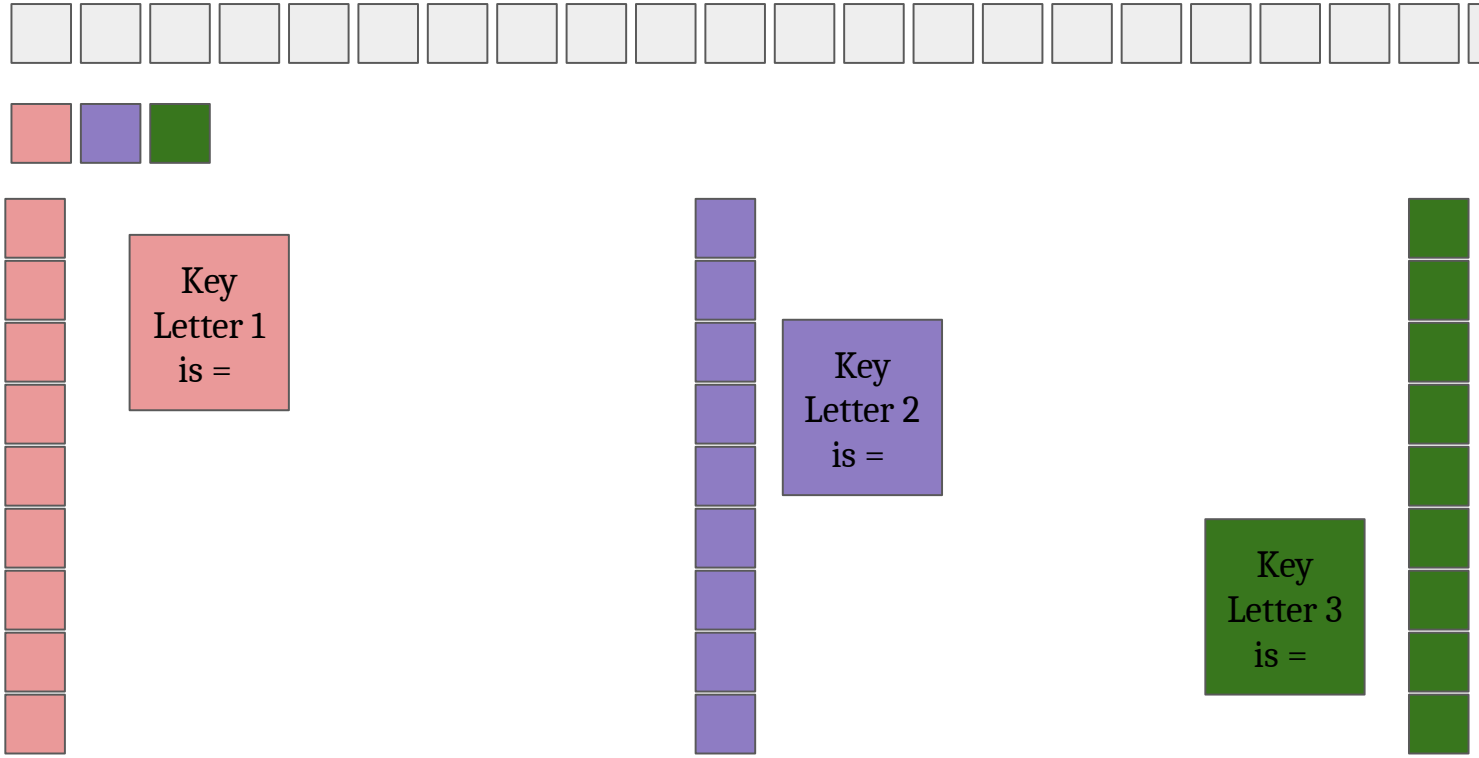
E	12.7 %	M	2.4 %
T	9.1 %	W	2.4 %
A	8.2 %	F	2.2 %
O	7.5 %	G	2.0 %
I	7.0 %	Y	2.0 %
N	6.7 %	P	1.9 %
S	6.3 %	B	1.5 %
H	6.1 %	V	1.0 %
R	6.0 %	K	0.8 %
L	4.0 %	J	0.2 %
D	4.3 %	X	0.2 %
C	2.8 %	Q	0.1 %
U	2.8 %	Z	0.1 %

Breaking the Word Cypher



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 %

Breaking the Word Cypher



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 %

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 sbeads 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