COMP 150:
Project Introduction
Week 14, Lecture20

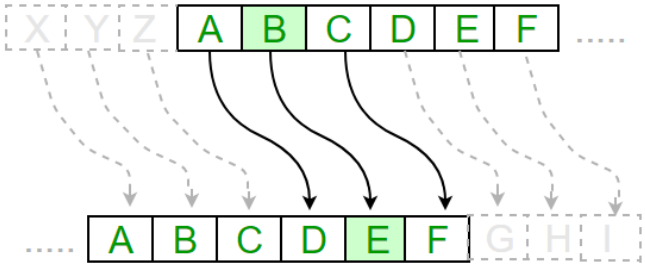
Agenda

1. Attendance
2. Timeline
3. The Shift Cypher
4. Frequency Analysis
5. Your Turn!

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



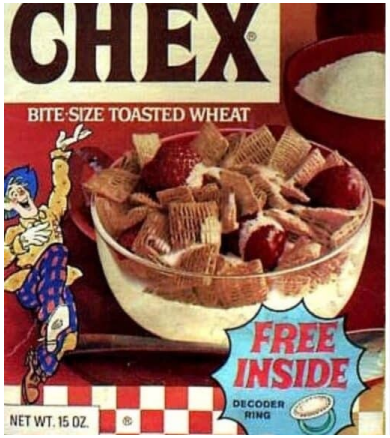
- Encode each letter with a number
- Shift each letter by some number from 0 to 25
  - Ex, A shifted by 0 is A, A shifted by 1 is B, etc
- Interactive wheel

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

The Shift Cypher

- Historic
  - Also called “Caesar Cypher”
- Quick and easy
  - We can do this in about 15 lines of code
- Predictably fragile
  - We can reliably break this in about 15 seconds
  - How?



Breaking The Shift Cypher

- Brute force
  - There are only so many shift options
  - Try them all and have a human verify the results!
- Crib dragging
  - “Crib”: known piece of plaintext
  - British slang (thanks, Bletchley Park!)
  - Ex: very likely in English that a three-letter-word is “the”
- Known Plaintext Attack (KPA)
  - Since there’s a 1:1 relationship between plaintext and cyphertext, KPA breaks shift cyphers
  - Crib dragging is a guesswork/minor form of KPA
- Frequency analysis

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.

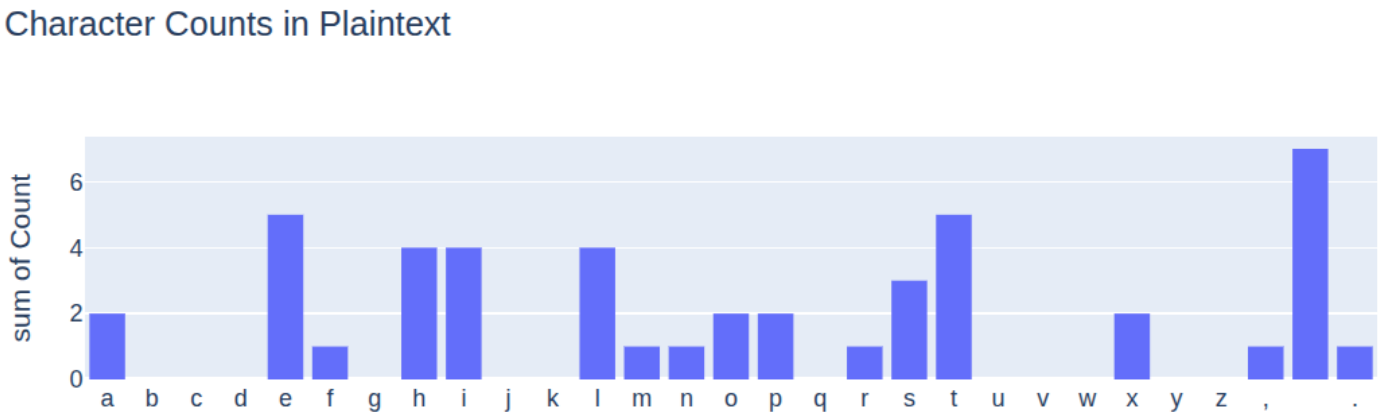
Cyphertext: Khoor, wklv lv wkh sodlqwhaw iru wklv hadpsoh.

Frequency Analysis

Key: 3

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

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

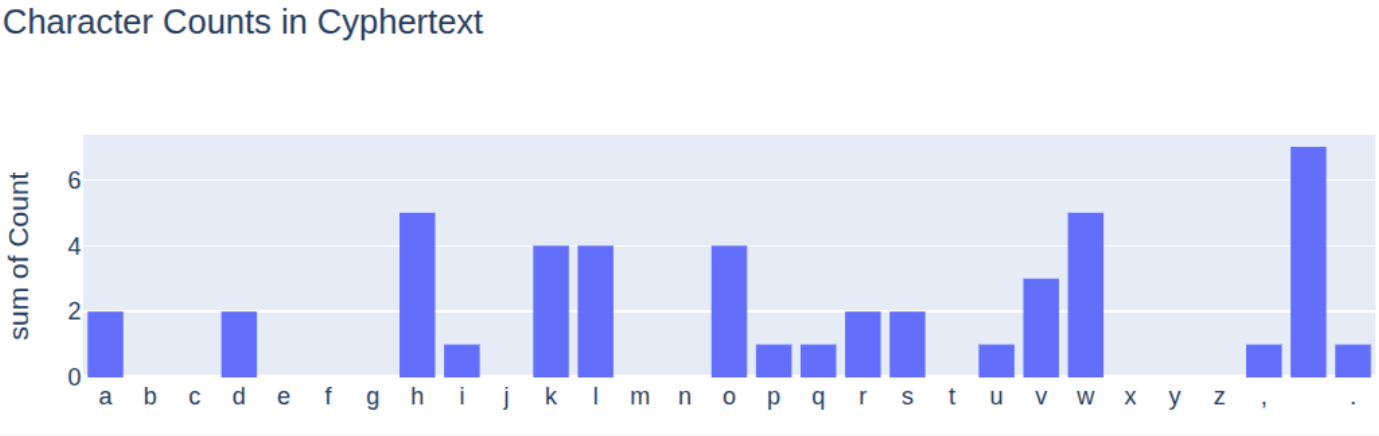


Frequency Analysis

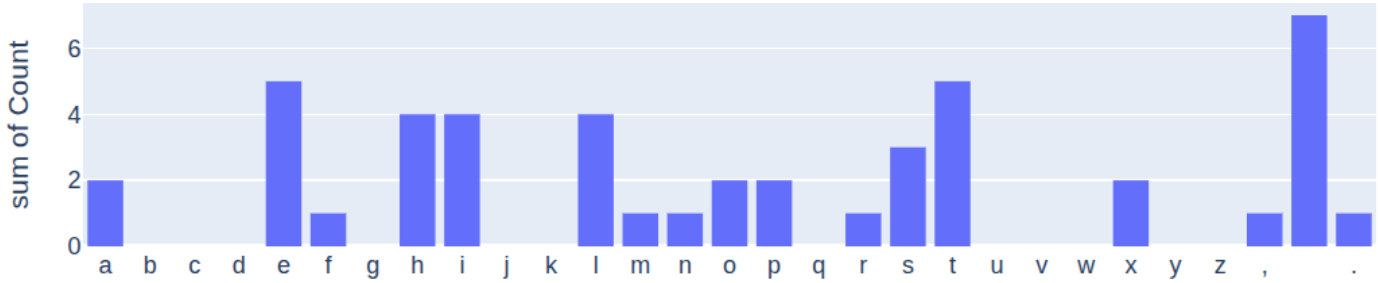
Key: 3

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

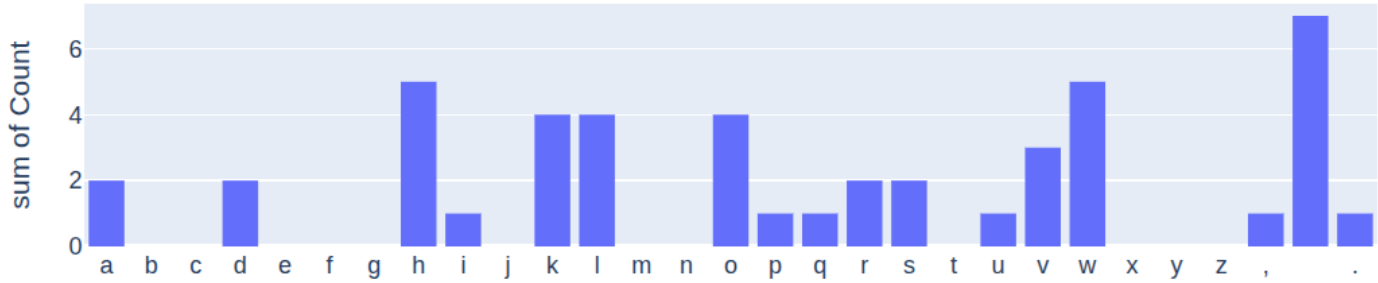
Cyphertext: Khoor, wklv lv wkh sodlqwhaw iru wklv hadpsoh.

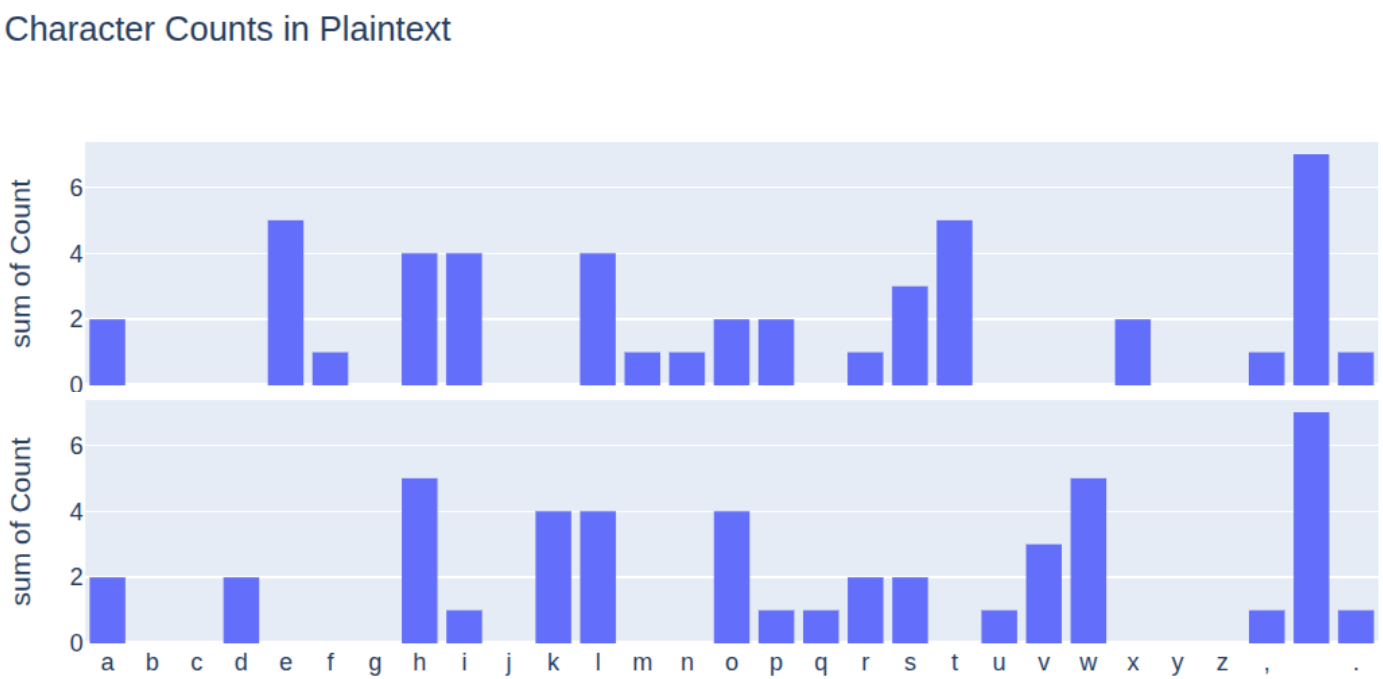


Character Counts in Plaintext



Character Counts in Cyphertext




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 %

Your turn!

1. Download COMP150_project_day1.ipynb from Sakai
2. Upload it to Google Colab
3. Rename the file to: yourname_COMP150_project_day1.ipynb
4. Complete the code steps
5. Begin a document, yourname_COMP150_project
6. Complete analysis steps in this document under a section titled:
    a. 1. The Shift Cypher

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