Austin Hunt
CS 6387 – Homework 2 (Crypto Homework) Part 1: Decode the Ciphertext
Sept 14, 2022

1. Who is the author of the work?
    i. Charles Dickens
2. Where did I get this text? (second two words, line #1)
    i. **Project Gutenberg,** a library of free eBooks
3. What is the key? I am looking for the mapping of cipher-text to plain-text characters. Specifically a python dictionary where the key is the letter in the ciphertext and the value is the value in the plaintext.

```
{
    "a": "o",
    "b": "n",
    "c": "w",
    "d": "d",
    "e": "t",
    "f": "l",
    "g": "b",
    "h": "i",
    "i": "f",
    "j": "u",
    "k": "c",
    "l": "z",
    "m": "q",
    "n": "p",
    "o": "k",
    "p": "r",
    "q": "a",
    "r": "y",
    "s": "h",
    "t": "g",
    "u": "m",
    "v": "v",
    "w": "s",
    "x": "x",
    "y": "j",
    "z": "e"
}
```

4. What technique(s) did you use to solve this problem?
    i. I first obtained the frequencies of each individual character. Then I disregarded this in favor of looking at word/pattern frequencies.
    ii. I obtained the frequencies of each N-length word for N between 1 and 6, considering very common words like "the", "and", "is", "are", "it", "that", "this", etc.

iii. Starting from word length 1, I found q to be most common, and replaced it with a; I found h to be the next most common, so replaced it with i. No other letters really made sense standing on their own.

iv. I then looked at longer patterns that were frequent; qbd was most common, esz was second most common; I replaced them respectively with "and" and "the" since "esz" was also found as the prefix to several high-frequency 4-letter words and several English words begin with "the" like "then", "their", etc.

v. I kept going in this direction using smaller cracked patterns to identify larger ones until the text was essentially obvious. After reaching the point where kHARfES DIkoENS was in the result, further replacements became increasingly easy.

My source code and documentation and produced files for frequency analysis are on Github here: https://github.com/austinjhunt/vanderbilt-cs6387-cryptography-homework/tree/main/decode-substitution-ciphertext
The source code has a lot of comments walking through my approach.