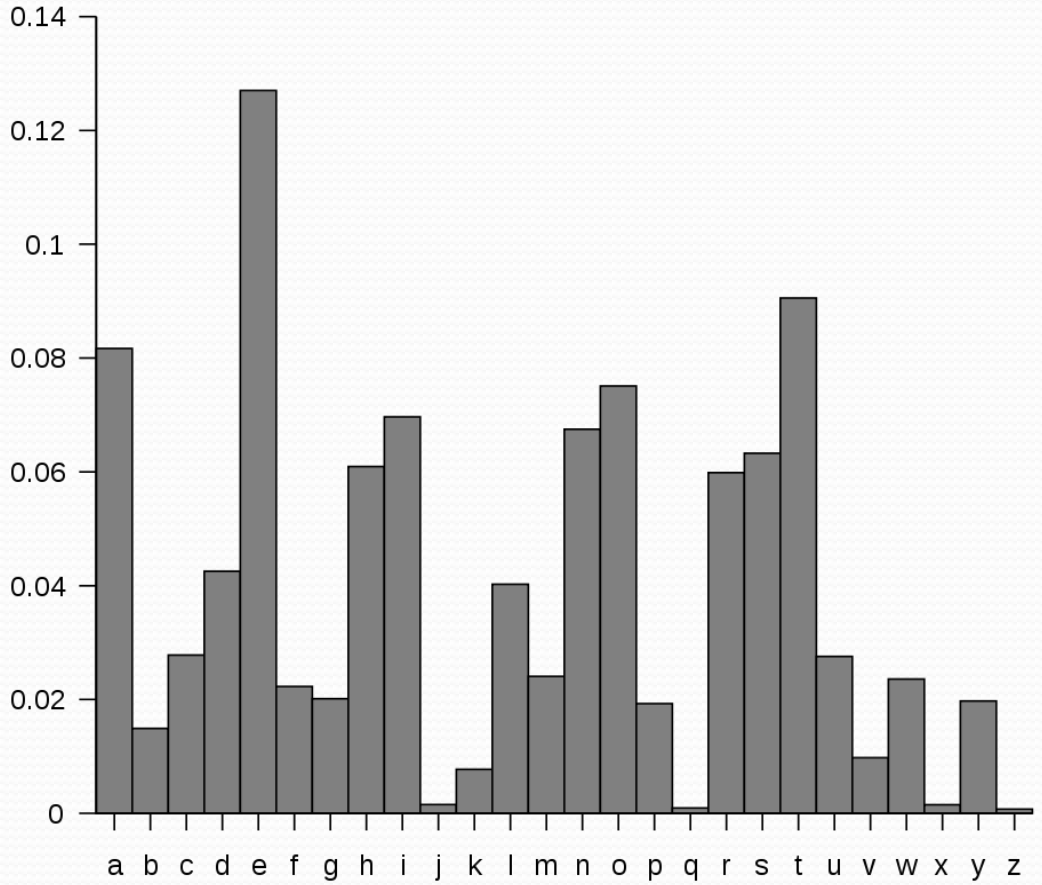
Cryptanalysis

Frequency Analysis

- Based on the fact that certain letters and combinations of letters are more common than others.
- E, T, A, O are most common letters
- Z, Q, X, J are rarest
- Naïve frequency analysis: match the most frequent letters in the ciphertext with the most common letters overall

English Letter Frequency Chart



Encrypted Text

vk ejl qlbr 1628, ejl lbrf sw nluskxjvrl nqvkd, bwelr sgr bgejsr jbn xlruln jvp 20 qlbrx, jl erbulffln bdbvk vkes wrbkzl tvej b xsk sw xvr dlrubx zfvwesk

Decrypted with Frequency Analysis

in the pear 1628, the earl ow devonshire dpiny, awter our author had served him 20 pears, he travelled ayain into wrance fith a son ow sir yervas cliwton

(using a text with 945,826 characters to analyze)

What do you think this text should be?

What ideas do you have for improving the decryption process?

Original Text

In the year 1628, the earl of Devonshire dying, after our author had served him 20 years, he travelled again into France with a son of Sir Gervas Clifton

Improvements

- Look at pairs of letters
  - The most common pair likely decodes to “th”, for example, especially if found in lots of three-letter words
- Look at words
  - A word decoded to “rtate” should most likely be “state”, so r can be replaced with s in the decoding.
- Use context
  - “in the pear 1628” doesn’t make sense, but it’s very common for the word “year” to be followed by a four-digit number

Modern Cryptography

- A simple substitution cipher can be broken in fractions of a second by any modern computer
- Modern cryptography uses the idea of computational intractability – problems which take unreasonable amounts of time to solve
- Many algorithms are based on large prime numbers
  - Multiplying two large primes: 15 microseconds
  - Recovering the original two factors: 200,000 years