Substitution Ciphers

Joshua E. Hill

Department of Mathematics, University of California, Irvine

Math 173A Introduction to Cryptography
October, 1 2013
http://bit.ly/18mQL0P

v1.1, compiled March 18, 2014

Motivation

- You are an NSA employee involved with the monitoring of groups that may pose a threat to the national security of the U.S.
- You are trying to infiltrate the *Secret Society of Insufferable Puzzle Makers* (the SSIPM).
- The leader of the SSIPM, Will Shortz, demands that you solve a cryptogram to join.

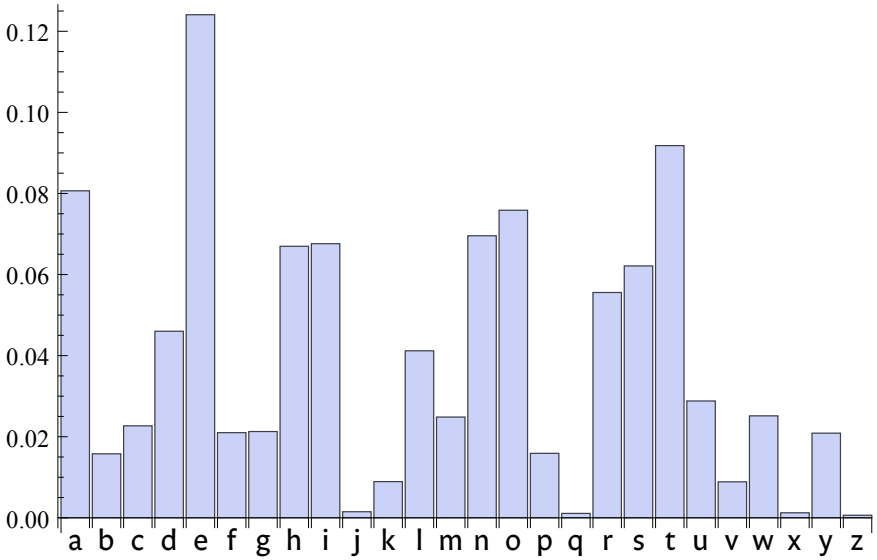
Cloak, Dagger... and Pen

Rjj dtb iqqpabs erdpbkdx pk dtb crso cbsb qisabo di abkxis jbddbsx cspddbk nl rjj dtb bkjpxdbo-gbk erdpbkdx, cti cbsb fbed pk sbxpobkab pk crsox iq dtbps ick. Pd crx r gikidikimx yin, rko Lixxrsprk crx opxreeipkdbo di jbrsk dtrd dtb jpwbx iq bkjpxdbo gbk cbsb ikjl xjphtdjl gisb pkdbsbxdpkh dtrd dtb jpwbx iq iqqpabsx. Rqdbs dtb qpsxd orl tb tro ki amspixpdl rd rjj. Di nsbrf dtb gikidikl tb pkwbkdbo hrgbx. Obrdt di rjj giopqpbsx, tb obajrsbo ikb orl, rko imd iq bwbsl jbddbs dtrd erxxbo dtsimht tpx trkox cbkd bwbsl rowbsn rko bwbsl roybadpwb. Dtb kbvd orl tb grob crs ik rsdpajbx. Tb sbratbo r gmat tphtbs ejrkb iq asbrdpwpdl dtb qijjicpkh orl ctbk tb njrafbo imd bwbsldtpkh pk dtb jbddbs di ”r”, ”rk” rko ”dtb”. Dtrd bsbadbo gisb olkrgpa pkdsrjpkbrs dbkxpikx, tb qbjd, rko pk ymxd rnimd bwbsl arxb jbqd r gbxxrhb qrs gisb mkpwbsxrj. Xiik tb crx esixaspnpkh ersdx iq xrjmdrdpikx rko xphkrdmsbx rko jbrwpkh dtb dbvd mkdimatbo. Ikb dpgb tb njrafbo imd rjj nmd dtb xrjmdrdpik ”Obrs Grsl” qsig r jbddbs, rko rd dtb niddig tb csidb, ”P lbrs qis lim dsrhparjjl. S.I. Xtpegrk, Atrejrpk, M.X. Rsgl.”

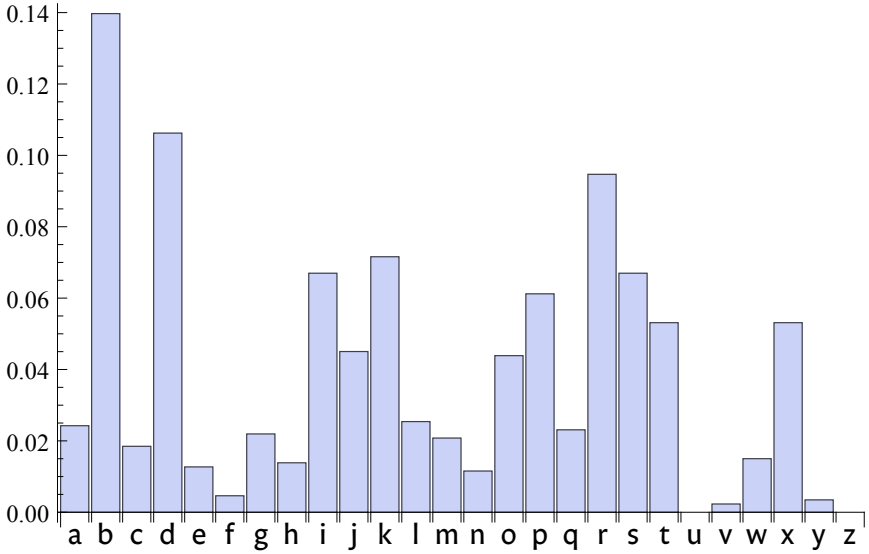
Ye Olde Frequencys

- We should be able to “do something” with letter frequencies.
- Where do we get them?
  - In modern times, this is “easy” (as compared to how hard it was before).
  - I assembled letter frequency by looking at a few public domain works:
    - *Alice in Wonderland* by Lewis Carroll
    - *Adventures of Huckleberry Finn* by Mark Twain
    - *Fairy Tales* by The Brothers Grimm
    - *Pride and Prejudice* by Jane Austen
    - *The Adventures of Sherlock Holmes* by Sir Arthur Conan Doyle
    - *Moby Dick* by Herman Melville
    - *Peter Pan* by J. M. Barrie
    - *The Legend of Sleepy Hollow* by Washington Irving
    - *The Adventures of Tom Sawyer* by Mark Twain
    - *A Tale of Two Cities* by Charles Dickens
    - *Wuthering Heights* by Emily Brontë
  - If you were serious, a better publicly available modern source would be Wikipedia, but this is very very large.

Histograms Off the Port Bow



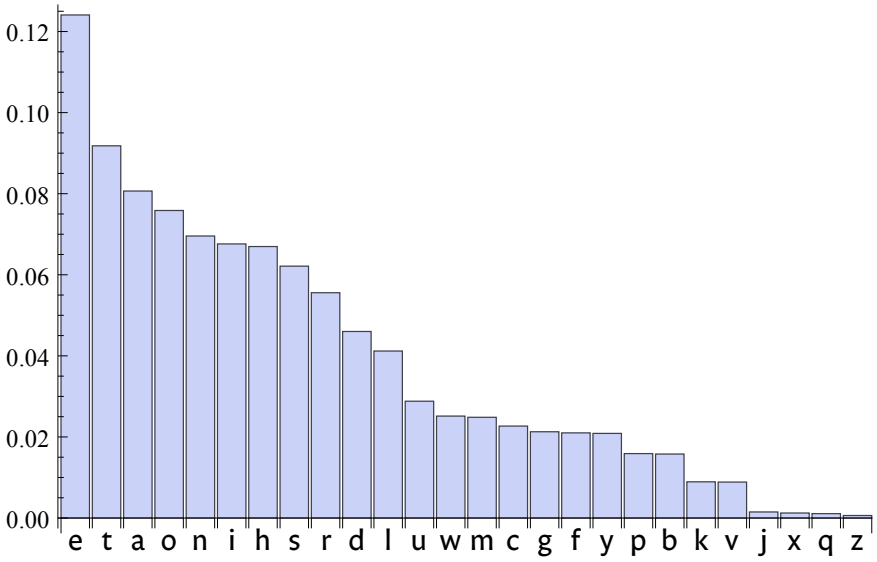
English



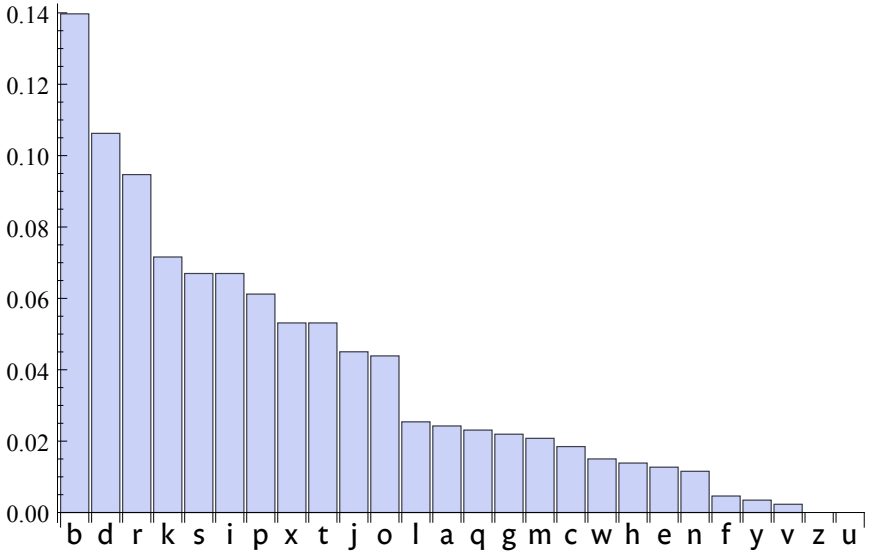
Cryptogram

A bit difficult to work with…

Ye Olde Sorted Frequencys

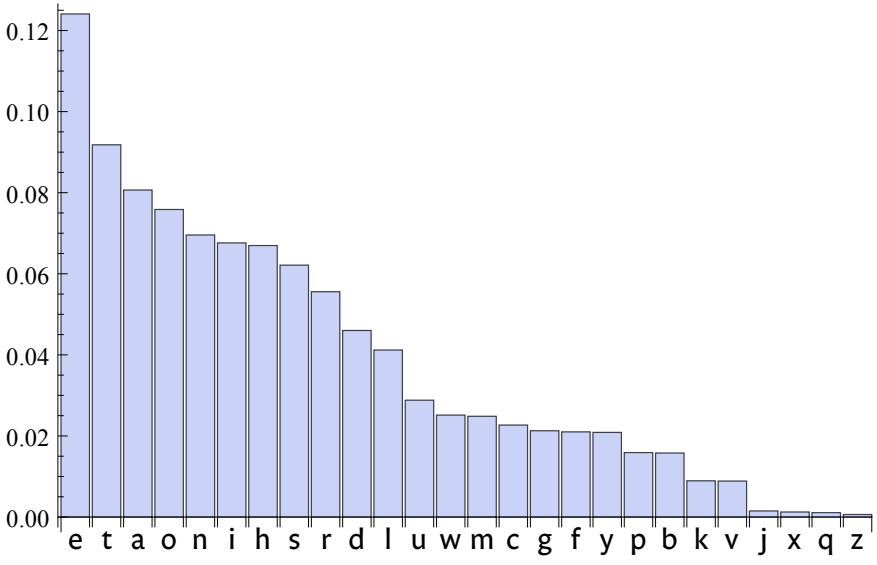


English

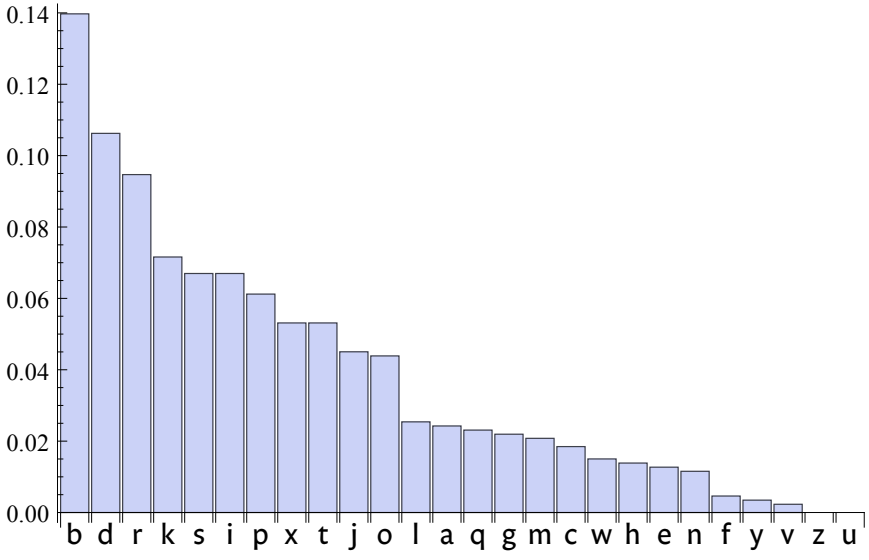


Cryptogram

Ye Olde Sorted Frequencys



English



Cryptogram

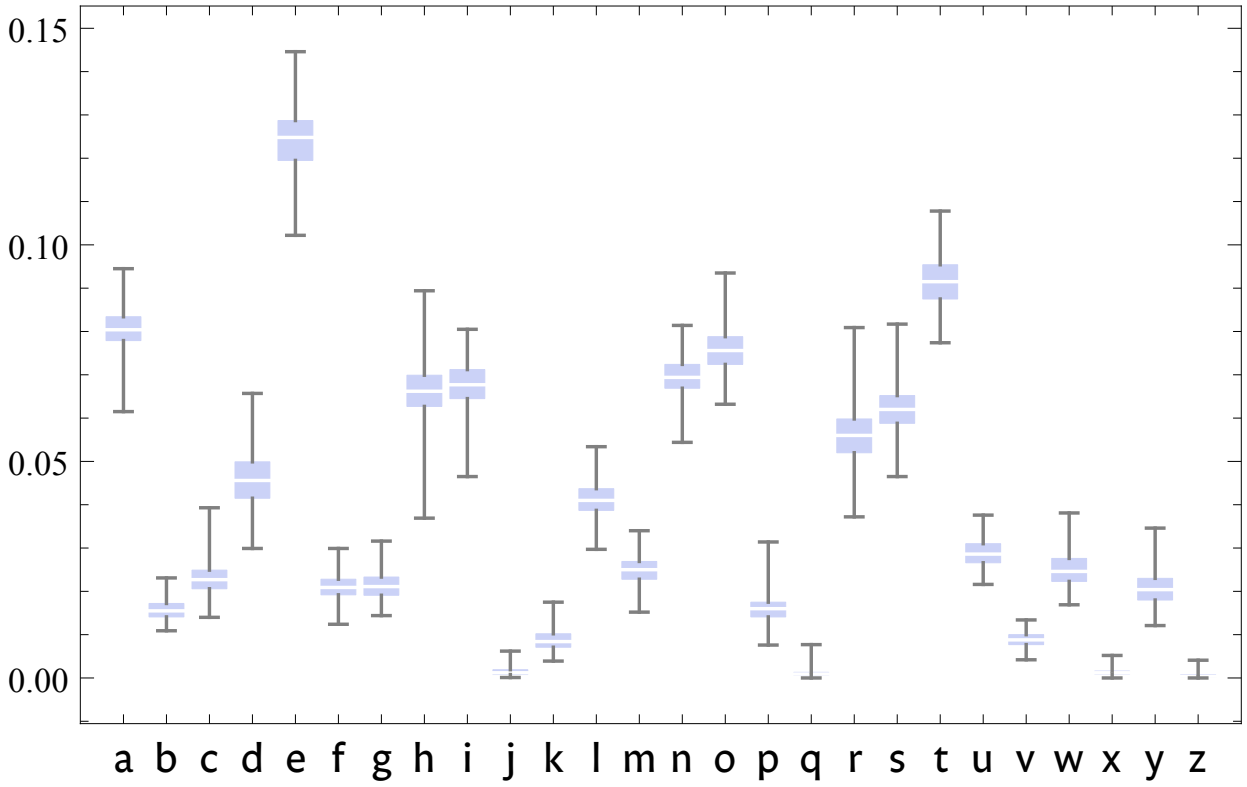
Jackpot!

Will Shortz is Going to Get What is Coming To Him!

Add tre immhwen batheots ho tre fanl fene minwel ti weosin dettens fnhtteo ku add tre eodhstel-ceo batheots, fri fene vebt ho neshleowe ho fanls im trehn ifo. Ht fas a cioitioigs jik, aol Uissanhao fas lhsabbihotel ti deano trat tre dhyes im eodhstel ceo fene iodu sdhprtdu cine hotenesthop trat tre dhyes im immhwens. Amten tre mhnst lau re ral oi wgnhishtu at add. Ti kneav tre cioitiou re hoyeotel paces. Leatr ti add cilhmhens, re lewdanel ioe lau, aol igt im eyenu detten trat bassel trnigpr rhs raols feot eyenu alyenk aol eyenu aljewthye. Tre oext lau re cale fan io anthwdes. Re neawrel a cgwr rhpren bdaoe im wneathyhtu tre middifhop lau freo re kdawvel igt eyenutrhop ho tre detten ti ”a”, ”ao” aol ”tre”. Trat enewtel cine luoachw hotnadhoean teoshios, re medt, aol ho jgst akigt eyenu wase demt a cessape man cine gohyensad. Siio re fas bniswnhkhop bants im sadgtathios aol shpoatgnes aol deayhop tre text gotigwrel. Ioe thce re kdawvel igt add kgt tre sadgtathio ”Lean Canu” mnic a detten, aol at tre kittic re fnite, ”H uean min uig tnaphwaddu. N.I. Srhbcao, Wrabdaho, G.S. Ancu.”

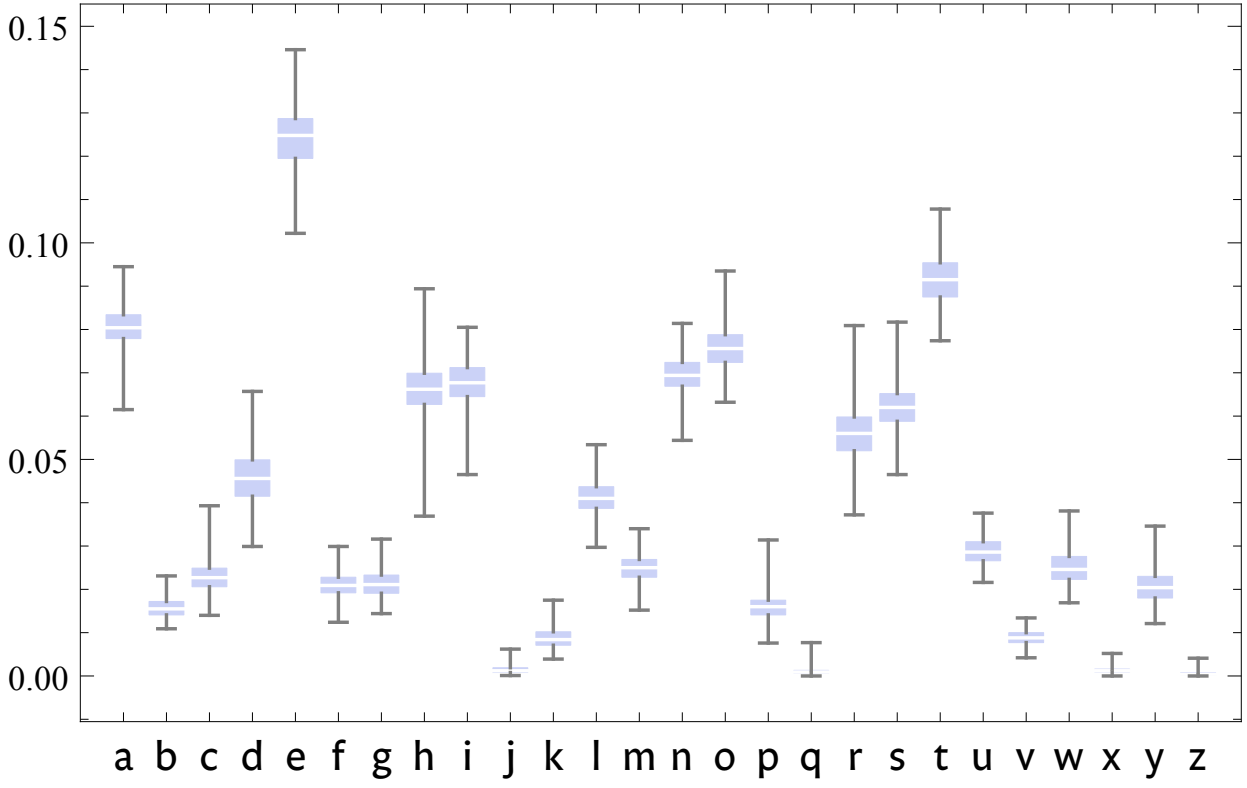
Et Tu, Will?

- What happened? The distribution looked right!
- The model was (much) too simple.
- A more reasonable way to view English letter frequency should capture uncertainty.
- Overlap yields message uncertainty.



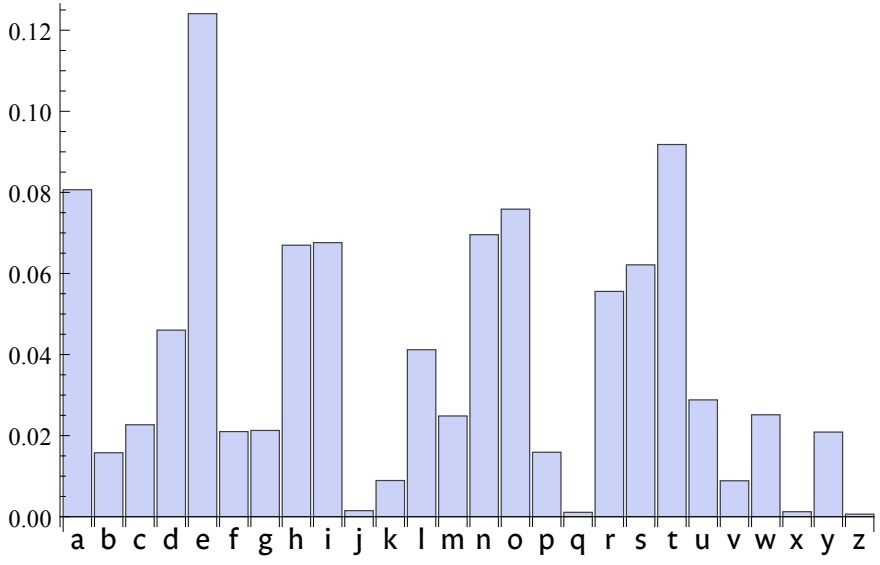
Et Tu, Will?

- What happened? The distribution looked right!
- The model was (much) too simple.
- A more reasonable way to view English letter frequency should capture uncertainty.
- Overlap yields message uncertainty.

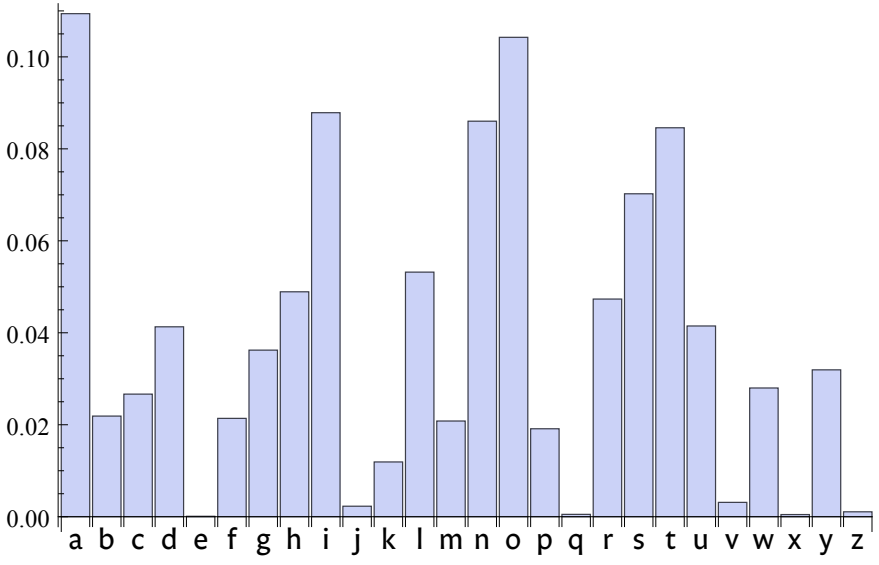


Stupid Humans!!!

- It gets worse…
- Some large intelligible messages have odd distributions:
  - *Gadsby* is a 50,000 word book by Vincent Wright that does not use the letter 'e'.



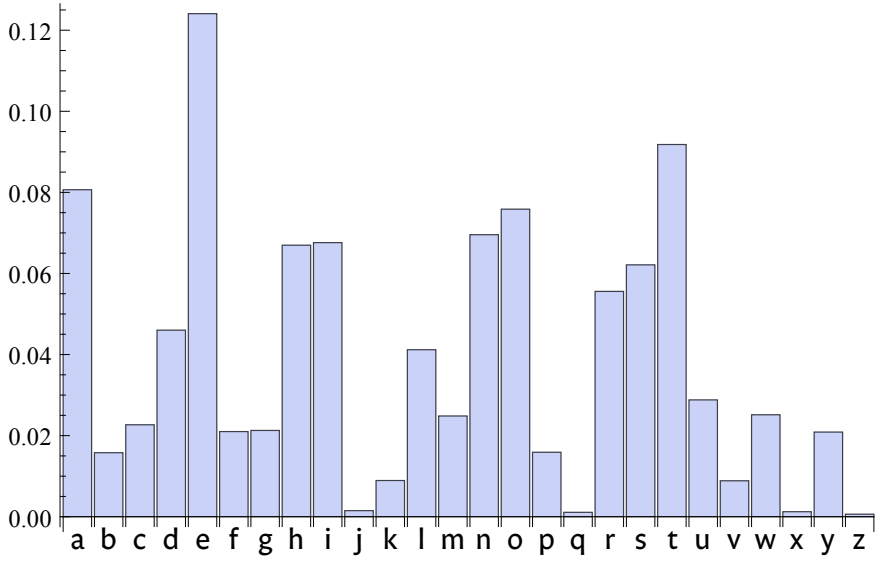
English Letter Frequencies



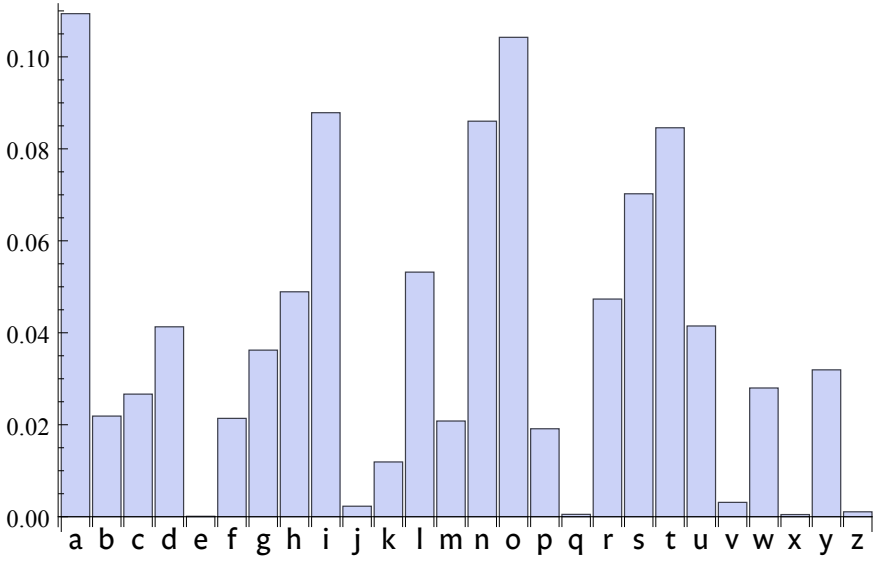
Gadsby Letter Frequencies

Stupid Humans!!!

- It gets worse…
- Some large intelligible messages have odd distributions:
  - *Gadsby* is a 50,000 word book by Vincent Wright that does not use the letter ‘e’.



English Letter Frequencies



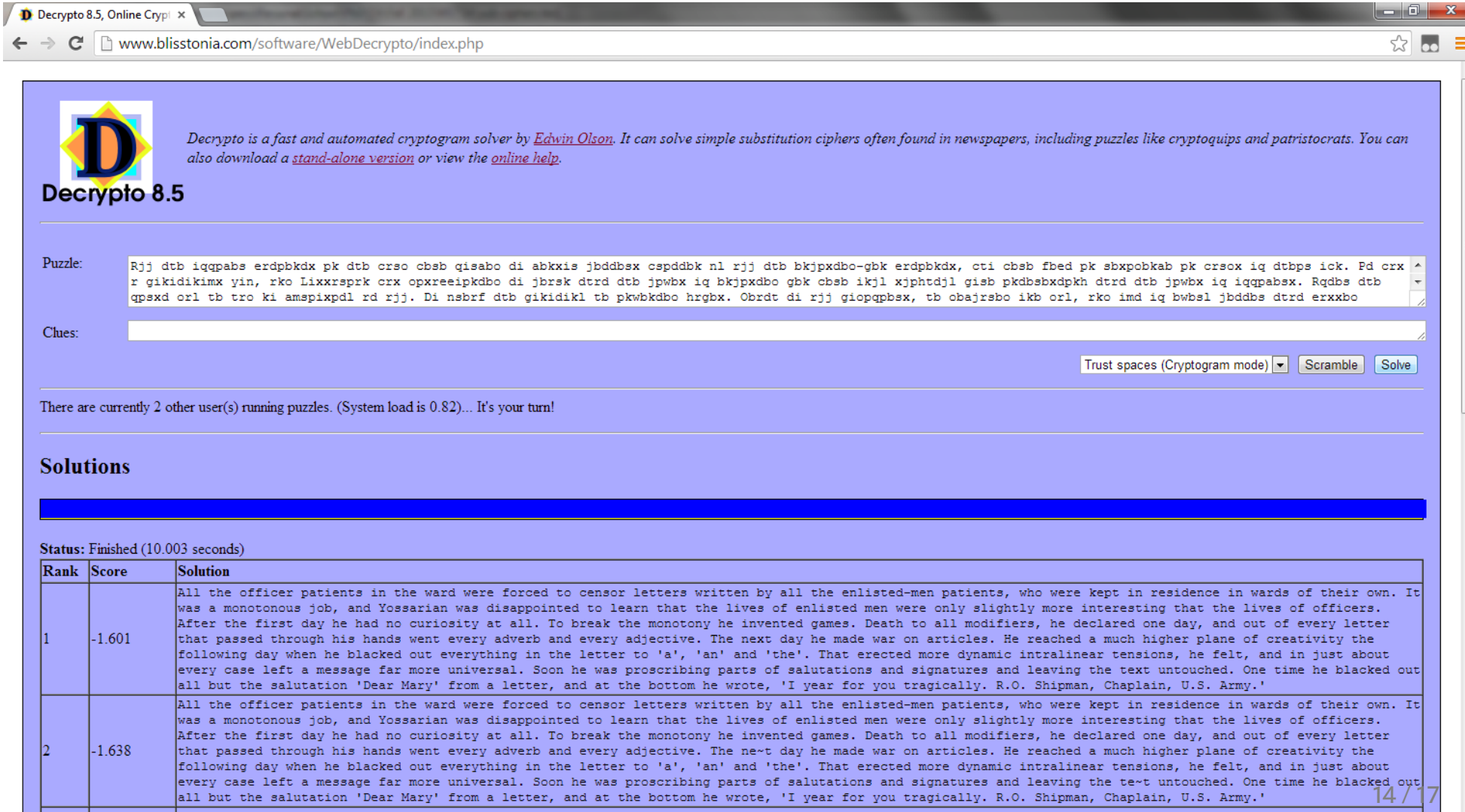
Gadsby Letter Frequencies

Curse You, Will Shortz!

- What now?
- Statistical Approaches
  - For large ciphertexts, we can use *digraph* and *trigraph* probability to help.
  - We can explore substitutions in a reasonable order, given the expected distribution.
  - These approaches work with longer messages. (*e.g.*, over 100 characters long.)
  - Can work in the absence of spaces.
- Dictionary Approaches
  - We think that the words are mostly in a dictionary.
  - Most words in the cryptogram act as sets of constraints.
  - More dictionary matches suggest a better selection of permutation.
  - There are a number of clever computer science techniques that "prune the solution space".
  - Can work for smaller cryptograms.
- We can not explore all $26! \approx 2^{89}$ permutations.

Rule 34 (as Applied to Cryptography)

- Edwin Olson has implemented a dictionary approach and written a paper describing his approach.
- He also has a web page to solve such puzzles called *Decrypto*.



Catch-22

All the officer patients in the ward were forced to censor letters written by all the enlisted-men patients, who were kept in residence in wards of their own. It was a monotonous job, and Yossarian was disappointed to learn that the lives of enlisted men were only slightly more interesting that the lives of officers. After the first day he had no curiosity at all. To break the monotony he invented games. Death to all modifiers, he declared one day, and out of every letter that passed through his hands went every adverb and every adjective. The next day he made war on articles. He reached a much higher plane of creativity the following day when he blacked out everything in the letter to 'a', 'an' and 'the'. That erected more dynamic intralinear tensions, he felt, and in just about every case left a message far more universal. Soon he was proscribing parts of salutations and signatures and leaving the text untouched. One time he blacked out all but the salutation 'Dear Mary' from a letter, and at the bottom he wrote, 'I yearn for you tragically. R.O. Shipman, Chaplain, U.S. Army.'

Catch-22

All the officer patients in the ward were forced to censor letters written by all the enlisted-men patients, who were kept in residence in wards of their own. It was a monotonous job, and Yossarian was disappointed to learn that the lives of enlisted men were only slightly more interesting that the lives of officers. After the first day he had no curiosity at all. To break the monotony he invented games. Death to all modifiers, he declared one day, and out of every letter that passed through his hands went every adverb and every adjective. The next day he made war on articles. He reached a much higher plane of creativity the following day when he blacked out everything in the letter to 'a', 'an' and 'the'. That erected more dynamic intralinear tensions, he felt, and in just about every case left a message far more universal. Soon he was proscribing parts of salutations and signatures and leaving the text untouched. One time he blacked out all but the salutation 'Dear Mary' from a letter, and at the bottom he wrote, 'I yearn for you tragically. R.O. Shipman, Chaplain, U.S. Army.'
— *Catch 22* by Joseph Heller

Bibliography

- Ernest Vincent Wright, *Gadsby*, `http://spinelessbooks.com/gadsby/`.
- Joseph Heller, *Catch-22*.
- Edwin Olson, *Robust Dictionary Attack of Short Simple Substitution Ciphers*, Cryptologia, October 2007.