# Assignment 1

Andrew Rosen

September 8, 2014

## Question 0

There are a couple of ways this could work. If we see substitution of Chinese characters for other Chinese characters, we are looking at a frequency table with an enormous $x$ axis. This just means we need a larger corpus to use brute force method.

On the other hand, Chinese is written down phonetically, it would be a frequency analysis on syllables, which is also a larger space.

## Question 1

The samples are too short to be analyzed significantly using $\Phi$ on single letters; they get close to identical values. Running $\Phi$ on dimers and trimers yields expected results on all but the first ciphertext, which gets close to 0 for $\Phi$ on the dimers and 0 on the trimers.

This means that no dimers or trimers reoccurred in sample 1, so it is most likely permuted and not a substitution.

## Question 2

No, there aren't enough letters to do a full mono-alphabetic cipher. See the next question.

## Question 3

We can discount other cipher systems because it's unlikely that an other cipher system would yield a ciphertext consisting only of `adfgvx`.

If we assume it's encoded with `adfgvx`, then it's almost certainly permuted, as $\Phi = 0.00539088589576$ on the corpus broken up into pairs of symbols, which looks just like random text. If it wasn't permuted, $\Phi$ would be much closer to 0.0385.

# Question 4

A $\Phi$ of 0.0363075549455 was found with 115 columns. This was significantly larger than the other $\Phi$ values. The resulting text:

vxvdaagvaffvxvdddddxavafvggvvdaagvvxaadfvggggvggvgfdxvdxgaaavg
gvdxxfxfddavggdgdgdfdxaaavxfafdxfvgvafvgddaavggadfafvdaaaxdxfadxd
xvgaadgaaavgvdfavgvggafgvaafvaafaggdfgvvddxvaaffvdxaafvdxvdaaavva
xfdfdxaxdxavxfdxddgvdgdxddggavaavaaadfgvvddxdfvaafvavggggvfxvgggv
xgvvddxvggvvdaagvafgvvxaafvdgdfxfafaxdxavdgdxvaafgaaaxffvgaafdxvg
gadxvxaafvafvgaaddaaavxfdxfvfvaavaaxaavggadxvafvgvaagvdxgvvdaavgv
gggvxaavgvagvvddxdfxvfvdxvagvggfgxvgvafgvvaggvxvggvggxfaaxaafvgdx
fvfvvxvddfdfggxvfvfxxvxffxafvgfdxfafgvgvxfdxvadxaxafxfdfggxvgvvdd
xdfvxggxvxfvafvaadffddxgvxvfgaavgvavaggfvggdgdxgvvdafvgfdddggavdf
ggxvavxfafaxafvgfdgaaavggvdfggxvvggggvfxvgggvxafvgfdggddgaggxvavf
vdxgvvdaagvafvxaafvafxfxfaavgvagvvddxdfvaafvavggvfdafaxdxdgdxfgaf
xfxffvgvvddxdffdaaaxdxdgdxgaxfxvdgfgfvggvggvvddxfvafvadxggddgvvdd
xvddxaavaaavgvafvgvavaavgfddxaafvafgvdgaadfaafgfgdxaaavgvvdggfvdx
gaxfxvdgfgfvggvggvvddxvddxaavaggddgvdxvggaxvavdxvadgdxddggavgvvdd
xgvafdgdxfadxafvgfdafvdaaaxdxfxvgggvxvgggvgdxgaxfxvdgfgggvggvvddx
vddxaavavdaaaxdxdgggavdxdxdddddxgagvxvfgggvgdgdfxfafaxdxavaavgvad
gaafxdxdgdxdddxdxxfdgggavdxaavgffafggxvfvgvggfdggfvgvavaaaffdvdgv
aavxaadfgvvddxvgaavgvagvvddxavdxaavgvavaggvxvdaagvvxaafvvxaavggvd
xvagvggfadxvaggvgdxvxafgvvdggxvgvddxvavgvvddxavxfggfvfvggddgvafdg
dxgvvdaavgaavxvdggxfdxfaggffggddfgafxfxffvvaggdxfvvgggvxdfggxvfxv
gggvxafgvggddgvdxvgaffvfvgggvvdggfvdxfvafdgfgxfdxggxfvaddaafvvdaf
ggvgdxvaavdxdgdxvaafdxfvaaavdxfvggdgdxgvafdgdxfvdgggavdxdxdddafg
aaagaafggxvfvgvvdaavgaaxfxfgvvddxvaaffvfgdxvgfvaaavdffvgvxvdddd

## Question 5

I decided to make a crib of the first 10 occurrences of the symbol `gg`, which occur at indices `[21, 23, 37, 73, 80, 109, 123, 127, 194, 208]`.

I searched each substring of the `3boat10` as the same length as the corpus, looking for an instance where characters at the listed indices were identical. The found substring was:

```
whatisufferinthatwaynotonguecantellfrommyearliestinfancyihaveb
eenamartyrtoitasaboythediseasehardlyeverleftmeforadaytheydidnotkn
owthenthatitwasmylivermedicalsciencewasinafarlessadvancedstatetha
nnowandtheyusedtoputitdowntolazinesswhyyouskulkinglittledevilyout
heywouldsaygetupanddosomethingforyourlivingcantyounotknowingofcou
rsethatiwasillandtheydidntgivemepillstheygavemeclumpsonthesideoft
heheadandstrangeasitmayappearthoseclumpsontheheadoftencuredmefort
hetimebeingihaveknownoneclumpontheheadhavemoreeffectuponmyliveran
dmakemefeelmoreanxioustogostraightawaythenandthereanddowhatwaswan
tedtobedonewithoutfurtherlossoftimethanawholeboxofpillsdoesnowyou
knowitoftenissothosesimpleoldfashionedremediesaresometimesmoreeff
icaciousthanallthedispensarystuffw
```