

How to Learn to Code: Homework 1

Due Tues 9AM

Kimiko Suzuki and Brandon Price

June 22, 2017

1. Read in the *virus.txt* file and compute the percentages of each nucleotide and the GC content of its sequence.
2. Compare these calculations to known viruses from **Auewarakul et. al.** paper and determine which of the known viruses our unknown virus is most similar based on nucleotide composition (you can do this visually rather than computing it).
3. Output all five percentages and the name of the known virus into an *output.txt* file.

A sample *output.txt* file would look like the following (N.B. these are not the answers):

```
Percent A: 0.25
Percent T: 0.25
Percent C: 0.25
Percent G: 0.25
Percent GC: 0.5
Based on nucleotide composition, our virus is most similar to: 'Ebola'
```

Please submit your script and your *output.txt* files to Brandon and me by email or on the Google Classroom.