

## Lab 2: Lists

Due date: Thursday January 23, 11:59pm

1. The main task in this lab is to separate text into lists and do some statistics on these lists. The data for this lab needs to be downloaded from

[https://www.gutenberg.org/ebooks/search/?sort\\_order=downloads](https://www.gutenberg.org/ebooks/search/?sort_order=downloads)

pick a few books you like and report (you will need to download the textfile (plain text UTF-8))

- The average sentence length.
  - The average word length.
  - How many 1-letter words, 2-letter words, ... 20-letter words are present.
  - What are the 100 most common words [use a dictionary] (20 points extra credit)
2. Send the python script and the output of your experiment to Philip (pb12c@my.fsu.edu) by email before the deadline. Your output should contain the book titles and the statistics.

```
#!/usr/bin/env python
# (c) yourname

#
def readbook(filename):
    '''
    Reads a textfile into a single string, ignoring line breaks.
    The function has a single argument: the infile name
    and it returns a soingle string containing the whole file
    '''
    with open(filename, 'r') as myfile:
        data=myfile.read().replace('\n', '').replace('\r','')
    return data

if __name__ == '__main__':
    data = readbook('testbook.txt')
# more stuff to follow
```