Anamitra Bhattacharyya
Predict 420-DL, Section 55
Extra Credit Assignment (May 29, 2016)

1) Python Code

```
import re
import collections

#The re package opens the text file and reads all alphanumeric words/characters
#and makes them all lowercase
words = re.findall('\w+', open('babble-words.txt').read().lower())
print(words)#prints a list of all words in 'babble-words.txt'
print collections.Counter(words)#prints a sorted list of words with their occurrence freq
```

2) Sample output

Sorted word frequency list:

```
In [1134]: print collections.Counter(words)
Counter({'sed': 17, 'ut': 15, 'in': 11, 'nulla': 10, 'nunc': 8, 'nec': 8, 'amet': 8, 'sit': 8, 'turpis': 8, 'vel': 7, 'et':
7, 'eget': 7, 'ac': 7, 'at': 7, 'vitae': 7, 'sagittis': 6, 'eu': 6, 'ultrices': 6, 'purus': 6, 'felis': 6, 'etiam': 5,
'non': 5, 'bibendum': 5, 'odio': 5, 'ligula': 5, 'sem': 5, 'aliquam': 5, 'eleifend': 5, 'justo': 5, 'enim': 5, 'quis': 5,
'donec': 5, 'duis': 5, 'id': 5, 'a': 5, 'scelerisque': 5, 'blandit': 4, 'eros': 4, 'metus': 4, 'tincidunt': 4, 'neque': 4,
'consequat': 4, 'ipsum': 4, 'nisi': 4, 'nisl': 4, 'porta': 4, 'auctor': 4, 'molestie': 4, 'risus': 4, 'euismod': 4, 'ante':
4, 'pretium': 4, 'feugiat': 4, 'luctus': 4, 'tempus': 4, 'quam': 4, 'ultricies': 4, 'ornare': 4, 'erat': 4, 'mauris': 4,
'lorem': 4, 'tristique': 4, 'dolor': 4, 'laoreet': 3, 'elit': 3, 'accumsan': 3, 'consectetur': 3, 'curabitur': 3, 'mattis':
3, 'viverra': 3, 'velit': 3, 'sollicitudin': 3, 'ex': 3, 'hendrerit': 3, 'orci': 3, 'magna': 3, 'phasellus': 3, 'proin': 3,
'varius': 3, 'morbi': 3, 'venenatis': 3, 'tellus': 3, 'fringilla': 3, 'lobortis': 3, 'efficitur': 3, 'suspendisse': 3,
'sapien': 3, 'dapibus': 3, 'tortor': 3, 'sodales': 3, 'lectus': 3, 'urna': 3, 'nullam': 3, 'massa': 3, 'dignissim': 3,
'est': 3, 'cursus': 3, 'porttitor': 3, 'pellentesque': 3, 'fermentum': 3, 'lacus': 2, 'integer': 2, 'faucibus': 2, 'augue':
2, 'vivamus': 2, 'iaculis': 2, 'quisque': 2, 'leo': 2, 'imperdiet': 2, 'vestibulum': 2, 'libero': 2, 'tempor': 2,
'placerat': 2, 'pharetra': 2, 'interdum': 2, 'vehicula': 2, 'posuere': 2, 'vulputate': 2, 'maximus': 2, 'dictum': 2,
'semper': 2, 'arcu': 2, 'commodo': 2, 'gravida': 2, 'finibus': 2, 'habitasse': 1, 'condimentum': 1, 'hac': 1, 'facilisi': 1,
'congue': 1, 'nibh': 1, 'malesuada': 1, 'lacinia': 1, 'diam': 1, 'mollis': 1, 'suscipit': 1, 'elementum': 1, 'maecenas': 1,
'aenean': 1, 'senectus': 1, 'fusce': 1, 'facilisis': 1, 'dui': 1, 'rhoncus': 1, 'ullamcorper': 1, 'egestas': 1, 'volutpat':
1, 'fames': 1, 'habitant': 1, 'platea': 1, 'netus': 1, 'convallis': 1, 'nam': 1, 'adipiscing': 1, 'aliquet': 1, 'dictumst':
1, 'potenti': 1})
```

Summary (parsed from output above)

1. sed 17
2. ut 15
3. in 11
4. nulla 10
5. nunc 8
6. nec 8
7. amet 8
8. sit 8
9. turpis 8
10. vel 7