## Homework 8 Ryan Kingery

STAT 5014 Dr. Settlage

## 1 Problem 2

Since everyone loves a word cloud, why not do that? The following Python code loads in the file survey\_data.txt as a string doc. Some minor text processing is then done to make the word cloud more representative. After this, a word cloud object is generated for doc using the Python wordcloud package, and displayed using the matplotlib.pyplot.imshow function. The code is shown below.

```
import matplotlib.pyplot as plt
import textmining as tm
from wordcloud import WordCloud
DIR = '~/STAT_5014/08_text_mining_Rnotebooks_bash_sed_awk/survey_data.txt'
def process_doc(doc):
   doc = doc.lower()
   doc = doc.replace('intermediate', 'int')
   doc = doc.replace('intermed','int')
   doc = doc.replace('int', 'intermediate')
   doc = doc.replace('beginner','beg')
   doc = doc.replace('beg', 'beginner')
   doc = doc.replace('master', 'ms')
   doc = doc.replace('obj-c','objC')
   doc = doc.replace('c++','cpp')
   doc = doc.replace('/','')
   doc = doc.replace('-',' ')
   for stopword in tm.stopwords:
       doc = doc.replace(' '+stopword+' ',' ')
   return doc
doc = open(DIR).read()
doc = process_doc(doc)
```

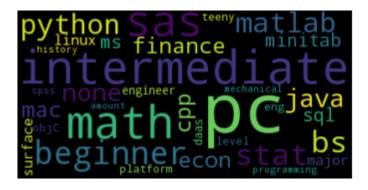


Figure 1: Word cloud of the survey text.

```
wordcloud = WordCloud().generate(doc)
plt.imshow(wordcloud,interpolation='bilinear')
plt.axis('off')
plt.show()
```

The word cloud is shown in figure 1. We can easily see from this the following: A plurality of students were math majors in college. There is a fairly even split between people with beginner and intermediate R experience. Most students are using PCs. Finally, there is a wide mix of other languages students have used before.

## 2 Problem 3

I'll have to take a zero for this problem. I spent way too much time trying to make the word cloud in the last problem look good (over 5 hours). I've got a Bayes homework I have to finish that's due today as well, and stuff for the Data Analytics project to do. Now seems to be that time of year when Scotland likes to torture his students. C'est la vie.