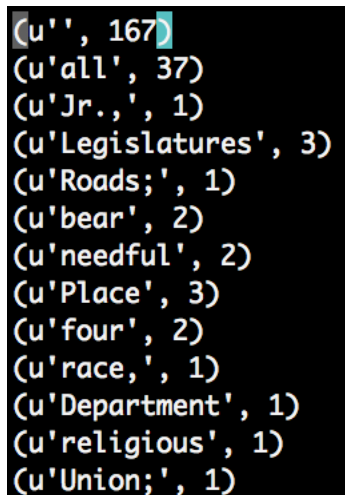


Karl Roth
Lab 12
GEOG479

In this lab we explored the use of spark in ROGER. First we loaded the constitution data into a variable from hdfs. We counted the number of elements to be 865. And the first line was:

u'We the People of the United States, in Order to form a more perfect Union.'

17 lines contained the word United States. Then we counted the number of words in the text file. The result is shown in Fig1 below.



```
(u'', 167)
(u'all', 37)
(u'Jr.', 1)
(u'Legislatures', 3)
(u'Roads;', 1)
(u'bear', 2)
(u'needful', 2)
(u'Place', 3)
(u'four', 2)
(u'race,', 1)
(u'Department', 1)
(u'religious', 1)
(u'Union;', 1)
```

Fig1. The first few word count examples.

Then we ran the New York taxi density map code. This code performs a map reduce on the taxi data. It creates an index with flatmap and then reduces by key, adding together the taxi and passenger data. It then maps to a key value pair, and sorts the data. Then the data is written to an ascii file.