

Analysis of Twitter Data Using Knitr

Matthew Scicluna

Monday, March 16, 2015

I used the twitterR R package to gather tweets with the word UofT in them. Here are a few such tweets.

```
## [1] "RT DaleBarbour The UofT needs honest UofTstrike Lets review theyre saying WeAreUofT httpcoWMnp"
```

```
## [1] "RT patitsel uoft just rejected overly modest proposal cupe httpcorQXnTESI So disappointed uni a"
```

```
## [1] "RT DiverseAlz Wouldnt great see conferences dementia RisingTide happening now UniversityofToron"
```

Here are the most frequent words found in those tweets, followed by how often each occurs.

##	uoft	risingtide	the	weareuoft	alzheimers	but
##	9	3	3	3	2	2
##	conference	dalebarbour	doe	education		
##	2	2	2	2		

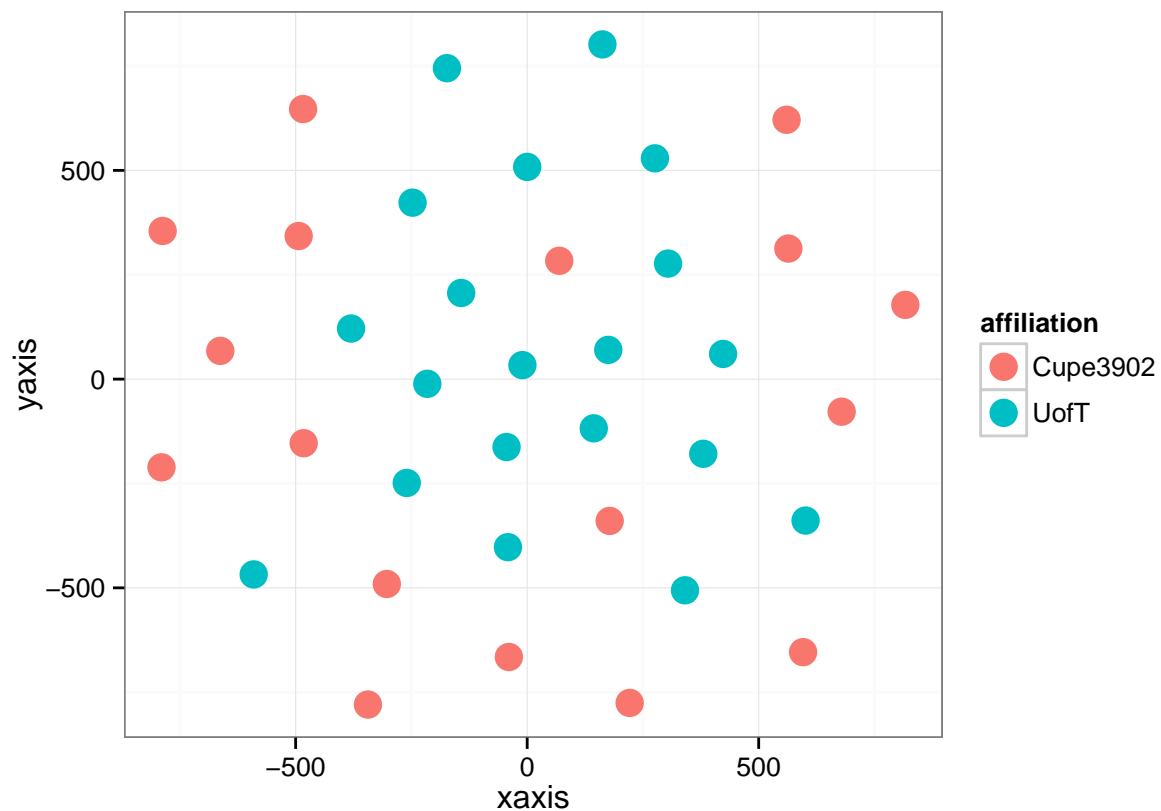
Now thats not very interesting! We should do some sort of comparative analysis. Lets analyze the tweets coming from the CUPE3092 and UofT twitter accounts and compare the word frequencies.

Here are the most common words used by UofT

##	congratulations	see	thats	fall
##	11	5	4	3
##	done	great		
##	2	2		

And here are the most common used by CUPE3902

##	weareuoft	cupe	strike	check	daily	members
##	10	6	5	4	3	3



Perhaps we should see what the data looks like projected onto 2 dimensions. We should also try to compute some sort of statistic at some point since this IS a statistics class after all!

```
##
## Paired t-test
##
## data: cupesums and uoftsums
## t = 2.95, df = 165, p-value = 0.00364
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1354650 0.6838121
## sample estimates:
## mean of the differences
##           0.4096386
```

Think about what kinds of analysis you can perform on this data or on data similar to it. You will need to be more rigorous than I am being if you want to do well on this assignment!