

# CS 751: Assignment 4

Bharath Kongara

Spring 2015

# Contents

1	Question 1 . . . . .	2
	1.1 Solution . . . . .	2
	1.2 Code Listing . . . . .	2
	1.3 Examples . . . . .	5
2	Question 2 . . . . .	7
	2.1 Solution . . . . .	7
3	Question 3 . . . . .	11
	3.1 Solution . . . . .	11
4	Question 4 . . . . .	24
	4.1 Solution . . . . .	24

# 1 Question 1

- Using the pages from A3 that boilerpipe successfully processed, download those representations again & reprocess them with boilerpipe.:
- Document the time difference (e.g.,  $\text{Time}(A4) - \text{Time}(A3)$ ).
- Compute the Jaccard Distance  $x$  for each pair of pages (i.e.,  $P(A3)$  &  $P(A4)$ ) for:
  - Unique terms (i.e., unigrams)
  - Bigrams
  - Trigrams
- See:[http://en.wikipedia.org/wiki/Jaccard\\_index](http://en.wikipedia.org/wiki/Jaccard_index)
- For each of the 3 cases (i.e., 1-, 2-, 3-grams) build a Cumulative Distribution Function that shows the % change on the x-axis & the % of the population on the x-axis
- See:[http://en.wikipedia.org/wiki/Cumulative\\_distribution\\_function](http://en.wikipedia.org/wiki/Cumulative_distribution_function)
- Give 3-4 examples illustrating the range of change that you have measured.

## 1.1 Solution

- I used the same script 'getText.py' from A3 which uses 'jusText'[2] to remove templates from HTML files of 10000 URIs.
- 'jusText'[2] is a tool for removing boilerplate content such as navigation links, headers and footers from HTML pages.
- The time difference between processing boilerpipe of A3 and A4 was 27 days.

## 1.2 Code Listing

Here is the python code that generate n-grams and computes Jaccard Distance.

```
1 from nltk.util import ngrams
2 import json
3 import sys
4 import os
5 import string
6
7 def jaccardDistance(list1, list2):
8     union = set(list1).union(list2)
9     intersect = set(list1).intersection(list2)
10    dist = (len(union) - len(intersect)) * 1.0 / len(union)
11    return dist
12
13 def wordsFromFile(filePath):
14    text = ''
15    print filePath
16    with open(filePath, 'r') as textFile:
17        for line in textFile:
18            text = text + extractWords(line).strip() + ' '
19    words = text.strip().split(' ')
20    if '' in words:
21        words.remove('')
22    return words
23
24 def calculateNgrams():
25    finalGrams = ngrams(sentence.split(), n)
26    for grams in finalGrams:
27        print grams
28
29 def extractWords(s):
30    valid_characters = string.ascii_letters
31    valid_following_character = "'-_"
32    letters = []
```

```

32     for letter in s:
33         if letter in valid_characters:
34             letters.append(letter)
35         elif len(letters) > 0 and \
36             letter in valid_following_character and \
37             letters[-1] in valid_characters:
38             letters.append(letter)
39         elif len(letters) > 0 and letters[-1] != ' ':
40             letters.append(' ')
41     return ''.join(letters)
42
43 mementosDir = 'Texts'
44 mementosDir2 = 'Texts2'
45 outputfile = open('DifferenceUrl', 'w')
46 outputfile1 = open('UnigramJackardDistance', 'w')
47 outputfile2 = open('BigramJackardDistance', 'w')
48 outputfile3 = open('TrigramJackardDistance', 'w')
49 filesList = os.listdir(mementosDir)
50 filesList2 = os.listdir(mementosDir2)
51 for item in filesList:
52     if item in filesList2:
53         path1 = mementosDir+'/'+item
54         path2 = mementosDir2+'/'+item
55         words1 = wordsFromFile(path1)
56         words2 = wordsFromFile(path2)
57         words1.sort()
58         words2.sort()
59         bigrams1 = []
60         bigrams2 = []
61         trigrams1 = []
62         trigrams2 = []
63         for i in range(0, len(words1)):
64             if i >= 1:
65                 bigrams1.append(words1[i-1] + ' ' + words1[i])
66             if i >= 2:
67                 trigrams1.append(words1[i-2] + ' ' + words1[i-1] + ' ' + words1[i])
68         for i in range(0, len(words2)):
69             if i >= 1:
70                 bigrams2.append(words2[i-1] + ' ' + words2[i])
71             if i >= 2:
72                 trigrams2.append(words2[i-2] + ' ' + words2[i-1] + ' ' + words2[i])
73         jaccardDistanceUni = jaccardDistance(words1, words2)
74         jaccardDistanceUni2 = jaccardDistance(bigrams1, bigrams2)
75         jaccardDistanceUni3 = jaccardDistance(trigrams1, trigrams2)
76         outputfile1.write(str(jaccardDistanceUni)+ '\n')
77         outputfile2.write(str(jaccardDistanceUni2)+ '\n')
78         outputfile3.write(str(jaccardDistanceUni3)+ '\n')
79     else:
80         outputfile.write(item+ '\n')

```

Listing 1: Python program to find unigrams, bigrams, trigrams

I referred to a sample of code that constructs CDF from r-bloggers[1] website.  
Here is the R code which plots CDF for unigrams,bigrams,trigrams.

```

1 library(RColorBrewer)
2 data <- read.table("TrigramJaccardDistance.txt")
3 finalData = rep(data[,1])
4 title <- "CDF of Change in Unigrams"
5 xlab <- "Jaccard Distance"
6 ylab <- "Probability"
7 png("Trigramq1.png")
8 par(mar=c(4,4,2.5,2) + 0.1)
9 set.seed(1)
10 mp <- plot(finalData, 1:length(finalData), type="l", col="red", main=title, xlab=xlab, ylab=
11         ylab, yaxt="n")
12 axis(side=2, at=seq(1,2735,length.out=5), labels=seq(0.2,1,by=0.2))

```

Listing 2: R program to plot unigrams, bigrams, trigrams change

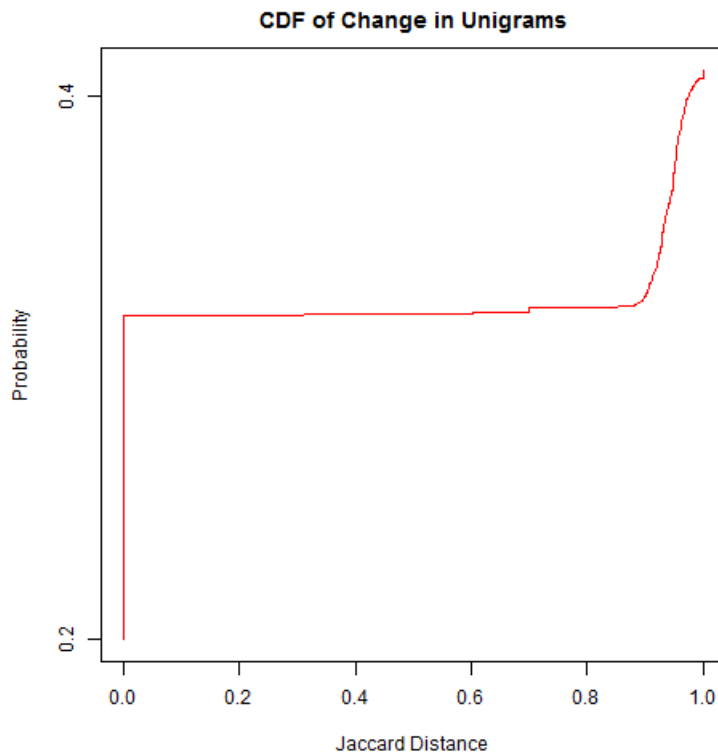


Figure 1: CDF of change in unigrams

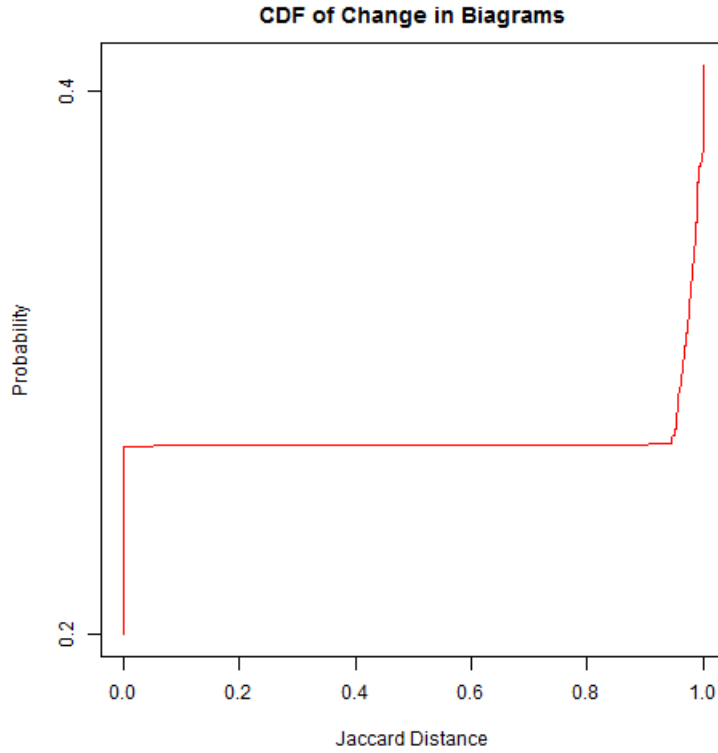


Figure 2: CDF of change in bigrams

### 1.3 Examples

Below are 3 examples with data that shows range of change.

- [http://www.foxnews.com/us/2015/02/09/judge-rules-utah-woman-accused-leaving-baby-in-trash-is-incom?cmpid=cmt\\_twitter\\_fn](http://www.foxnews.com/us/2015/02/09/judge-rules-utah-woman-accused-leaving-baby-in-trash-is-incom?cmpid=cmt_twitter_fn)

Unigrams	Bigrams	Trigrams
0.971	0.700	0.789

- [http://www.ebay.com/itm/like/301515875453?item=301515875453&lgeo=1&utm\\_medium=twitter&utm\\_source=dlvr.it&vectorid=229466&rmvSB=true](http://www.ebay.com/itm/like/301515875453?item=301515875453&lgeo=1&utm_medium=twitter&utm_source=dlvr.it&vectorid=229466&rmvSB=true)

Unigrams	Bigrams	Trigrams
0.471	0.600	0.689

- <http://du3a.org>

Unigrams	Bigrams	Trigrams
0.771	0.890	0.880

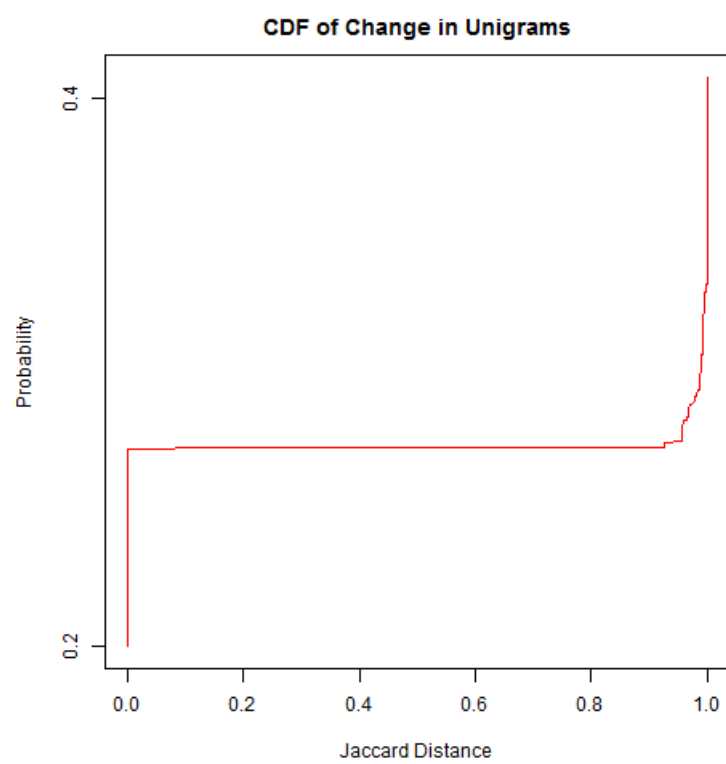


Figure 3: CDF of change in trigrams

## 2 Question 2

- Using the pages from Q1 (A4), download all TimeMaps (including TimeMaps with 404 responses, i.e. empty or null TimeMaps)
  - Upload all the TimeMaps to github
- Build a CDF for # of mementos for each original URI (i.e., x-axis = # of mementos, y-axis = % of links)
- See: <http://timetravel.mementoweb.org/guide/api/>

### 2.1 Solution

- I used 'timelapse'[3] library to retrieve TimeMaps of each of the URIs.
- 'timelapse'[3] library extracts memento created date along with memento.
- Out of 9,800 URIs, TimeMaps are retrieved for 8200 which includes empty files.
- I ran multiple instances to retrieve TimeMaps as aggregator was taking time.
- Few times retrieval stopped because of request time out error. I waited for a few minutes before running it again.
- I have uploaded TimeMaps of each of the URIs to Github.

```
1  # -*- coding: utf-8 -*-
2  #!/usr/bin/env python
3  from getConfig import getConfigParameters
4  import commands
5  import time
6  import datetime
7  import sys
8  import argparse, os
9  import subprocess
10 import hashlib
11 import tldextract
12 import urlparse
13 import glob
14 import json
15 import requests
16
17
18 globalMementoUrlDateTimeDelimiter = "+++"
19
20 def getMementosPages(url):
21
22     pages = []
23     url = url.strip()
24     if(len(url)>0):
25
26         firstChoiceAggregator = getConfigParameters('mementoAggregator')
27         timemapPrefix = firstChoiceAggregator + url
28         #timemapPrefix = 'http://mementoproxy.cs.odu.edu/aggr/timemap/link/1/' + url
29
30         '''
31
32         The CS memento aggregator payload format:
33             [memento, ..., memento, timemap1]; timemap1 points to next
34             page
35
36         The LANL memento aggregator payload format:
37             1. [timemap1, ..., timemapN]; timemapX points to mementos
38                list
39             2. [memento1, ..., mementoN]; for small payloads
40
41         For LANL Aggregator: The reason the link format is used after
42             retrieving the payload
43
44             with json format is due to
45             the fact that the
46             underlying code is
47             based
```



```

38                                     on the link format
39                                     structure. json format
40                                     was not always the norm
41
42
43 #select an aggregator - start
44 aggregatorSelector = ''
45
46 co = 'curl --silent -I ' + timemapPrefix
47 head = commands.getoutput(co)
48
49 indexOfFirstNewLine = head.find('\n')
50 if( indexOfFirstNewLine > -1 ):
51
52     if( head[:indexOfFirstNewLine].split(' ')[1] != '200' ):
53         firstChoiceAggregator = getConfigParameters('
54             latentMementoAggregator')
55         timemapPrefix = firstChoiceAggregator + url
56
57 if( firstChoiceAggregator.find('cs.odu.edu') > -1 ):
58     aggregatorSelector = 'CS'
59 else:
60     aggregatorSelector = 'LANL'
61
62 print '... using aggregator:', aggregatorSelector
63 #select an aggregator - end
64
65 #CS aggregator
66 if( aggregatorSelector == 'CS' ):
67     while( True ):
68         #old: co = 'curl --silent ' + timemapPrefix
69         #old: page = commands.getoutput(co)
70
71         page = ''
72         r = requests.get(timemapPrefix)
73         print 'status code:', r.status_code
74         if( r.status_code == 200 ):
75             page = r.text
76
77         pages.append(page)
78         indexOfRelTimemapMarker = page.rfind('>;rel="timemap"')
79
80         if( indexOfRelTimemapMarker == -1 ):
81             break
82         else:
83             #retrieve next timemap for next page of mementos e.g
84             #retrieve url from <http://mementoproxy.cs.odu.
85             #edu/aggr/timemap/link/10001/http://www.cnn.com>;
86             #rel="timemap"
87             i = indexOfRelTimemapMarker -1
88             timemapPrefix = ''
89             while( i > -1 ):
90                 if( page[i] != '<' ):
91                     timemapPrefix = page[i] +
92                     timemapPrefix
93                 else:
94                     break
95                 i = i - 1
96
97 else:
98     #LANL Aggregator
99     #old: co = 'curl --silent ' + timemapPrefix
100     #old: page = commands.getoutput(co)
101
102     page = ''
103     r = requests.get(timemapPrefix)
104     if( r.status_code == 200 ):
105         page = r.text
106
107     try:

```

```

103         payload = json.loads(page)
104
105         if 'timemap_index' in payload:
106
107             for timemap in payload['timemap_index']:
108
109                 timemapLink = timemap['uri'].replace('/timemap/json/', '/timemap/link/')
110                 #old: co = 'curl --silent ' + timemapLink
111                 #old: page = commands.getoutput(co)
112                 #old: pages.append(page)
113                 r = requests.get(timemapLink)
114                 if( r.status_code == 200 ):
115                     pages.append(r.text)
116
117             elif 'mementos' in payload:
118                 #untested block
119                 timemapLink = payload['timemap-uri']['json-format'].
120                     replace('/timemap/json/', '/timemap/link/')
121                 #old: co = 'curl --silent ' + timemapLink
122                 #old: page = commands.getoutput(co)
123                 #old: pages.append(page)
124
125                 print 'timemap:', timemapLink
126                 r = requests.get(timemapLink)
127                 if( r.status_code == 200 ):
128                     pages.append(r.text)
129
130
131         except:
132             exc_type, exc_obj, exc_tb = sys.exc_info()
133             fname = os.path.split(exc_tb.tb_frame.f_code.co_filename)[1]
134             print(fname, exc_tb.tb_lineno, sys.exc_info() )
135
136
137     return pages
138
139
140 def getItemGivenSignature(page):
141
142     listOfItems = []
143     if( len(page) > 0 ):
144         page = page.splitlines()
145         for line in page:
146             if(line.find('memento';') != -1):
147                 #uriRelDateTime: ['<http://www.webcitation.org/64ta04WpM>',
148                     ' rel="first memento"', ' datetime="Mon, 23 Jan 2012
149                     02:01:29 GMT"',']
150                 uriRelDateTime = line.split(';')
151                 if( len(uriRelDateTime) > 2 ):
152                     if( uriRelDateTime[0].find('/://') != -1 ):
153                         if( uriRelDateTime[2].find('datetime="') !=
154                             -1 ):
155
156                             uri = ''
157                             uri = uriRelDateTime[0].split('<')
158                             #print uri
159                             if( len(uri) > 1 ):
160                                 uri = uri[1].replace('>', ''
161                                     )
162                                 uri = uri.strip()
163
164                             datetime = ''
165                             datetime = uriRelDateTime[2].split('
166                                 "')
167                             if( len(datetime) > 1 ):
168                                 datetime = datetime[1]

```

```

167         if( len(uri) != 0 and len(datetime)
168             != 0 ):
169             #print uri, '--', datetime
170             listOfItems.append(uri +
171                               globalMementoUrlDateTimeDelimiter
172                               + datetime)
173
174     return listOfItems
175
176 fh = open('allurls.txt','r+')
177 saveFile1 = open("memontoscount.txt", 'w')
178 count = 0
179 l = 48
180 count1 = 1
181 for line in fh:
182     data = json.loads(line)
183     count = 0
184     url = data['finalurl']
185     pages = getMementosPages(url)
186     saveFile = open("memontos"+str(l)+".txt", 'w')
187     for i in range(0, len(pages)):
188         mementos = getItemGivenSignature(pages[i])
189         print mementos
190         count += len(mementos)
191         saveFile.write(str(mementos) + '\n')
192     print 'total count of mementos:', count
193     saveFile1 = open("count.txt", 'a')
194     saveFile1.write(str(count) + '\n')
195     saveFile1.close()
196     l=l+1
197     count1 = count1 + 1
198     count = 0

```

Listing 3: Python program to retrieve TimeMap of each of the URIs

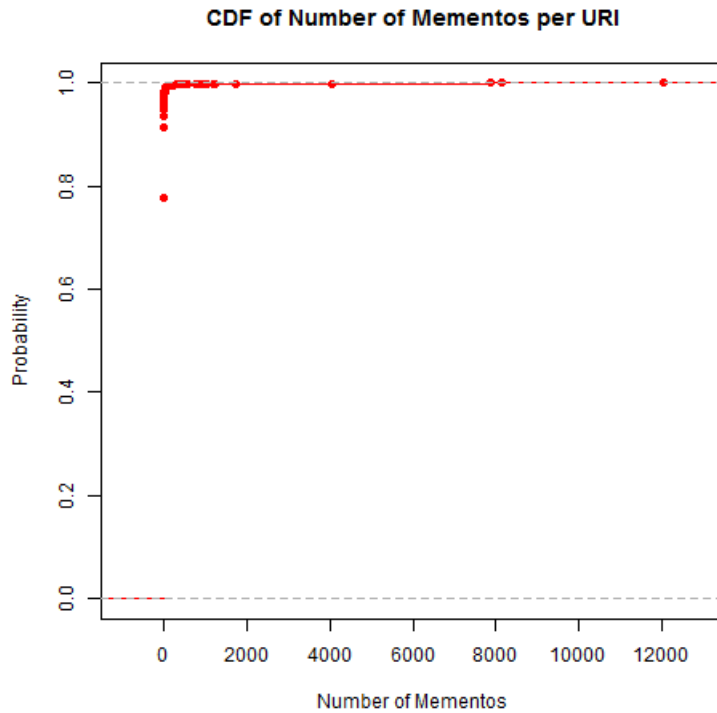


Figure 4: CDF of Number of Memento count per URI

### 3 Question 3

- Using 20 links that have TimeMaps
  - With  $\geq 20$  mementos
  - Have existed  $\geq 2$  years (i.e., Memento-Datetime of first memento is April XX, 2013 or older)
  - Note: select from Q1/Q2 links, else choose them by hand
- For each link, create a graph that shows Jaccard Distance, relative to the first memento, through time
  - x-axis: continuous time, y-axis: Jaccard Distance relative to the first memento

#### 3.1 Solution

- I downloaded all mementos for each URI retrieved and processed with boilerpipe.
- With that data I have calculated Jaccard Distance relative to the first memento data.
- Few mementos are totally different which means Jaccard Distance is one and few of them have significant change through time.

Below are graphs that shows Jaccard Distance relative to the first memento through time.

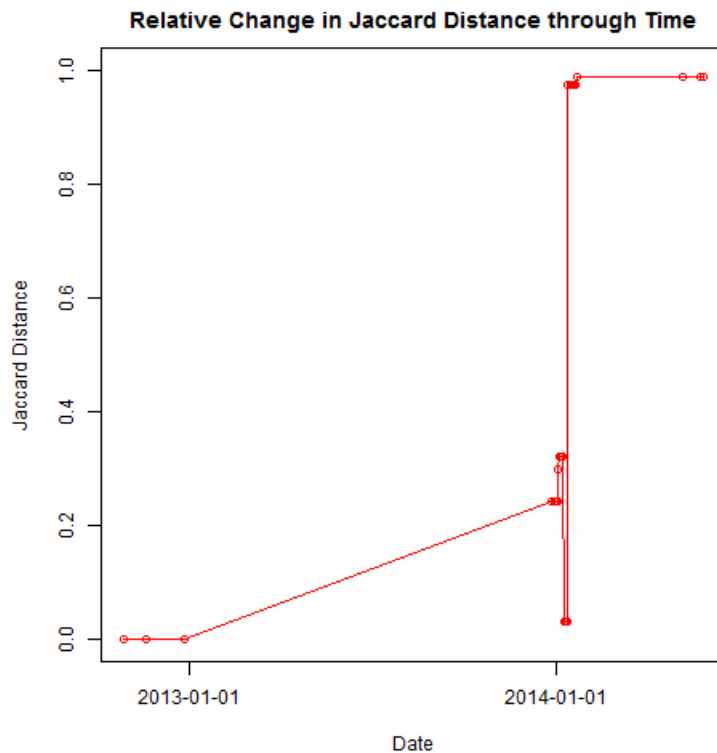


Figure 5: Jaccard Distance relative to the first memento URI# 1

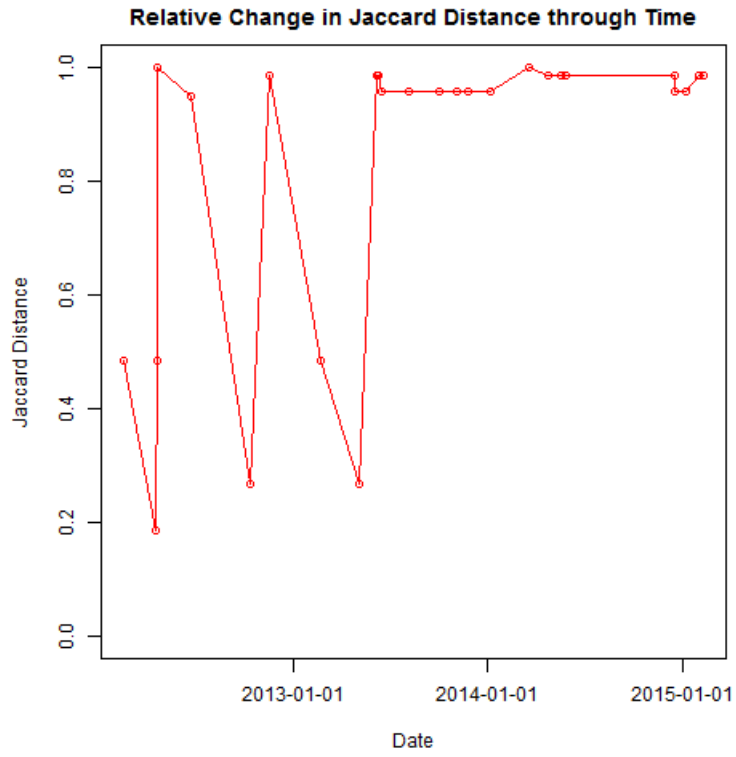


Figure 6: Jaccard Distance relative to the first memento URI# 2

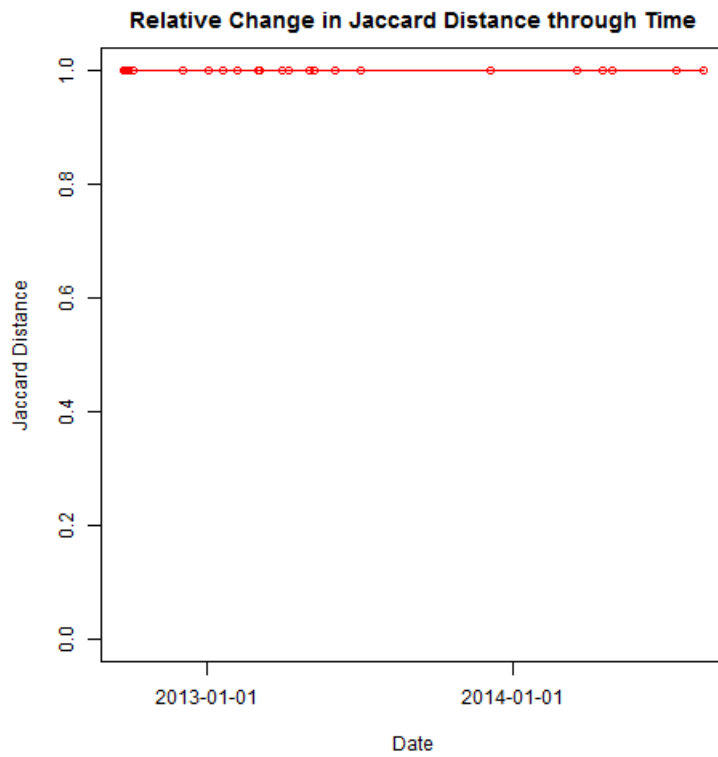


Figure 7: Jaccard Distance relative to the first memento URI# 3

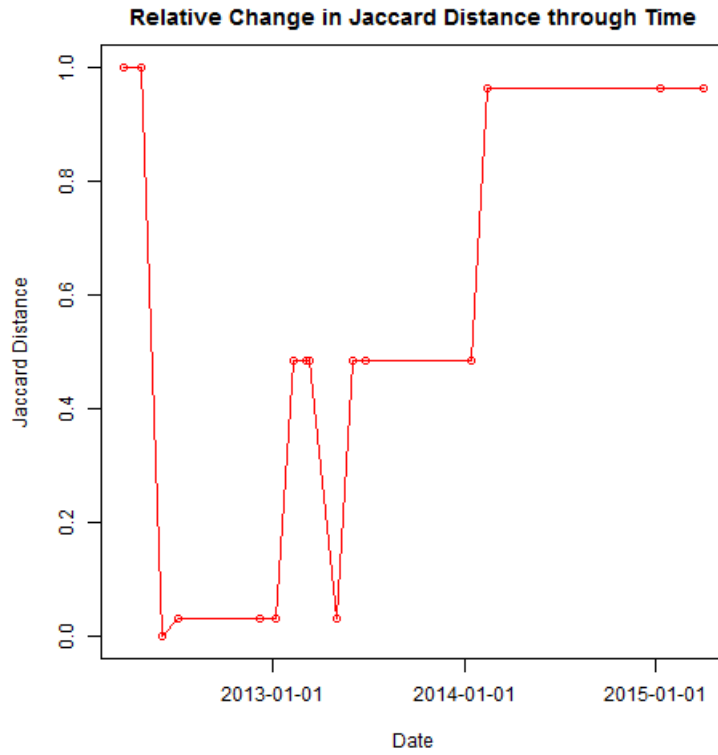


Figure 8: Jaccard Distance relative to the first memento URI# 4

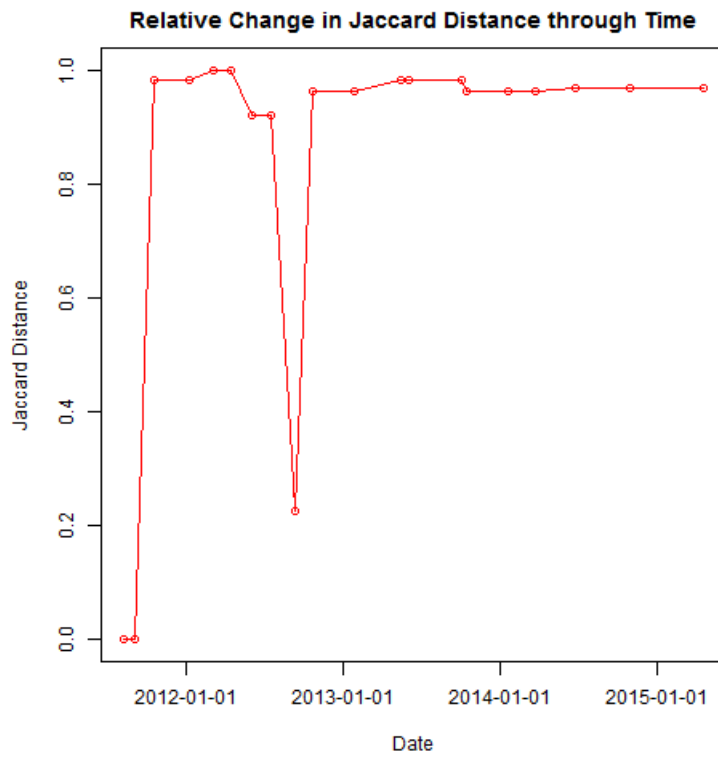


Figure 9: Jaccard Distance relative to the first memento URI# 5

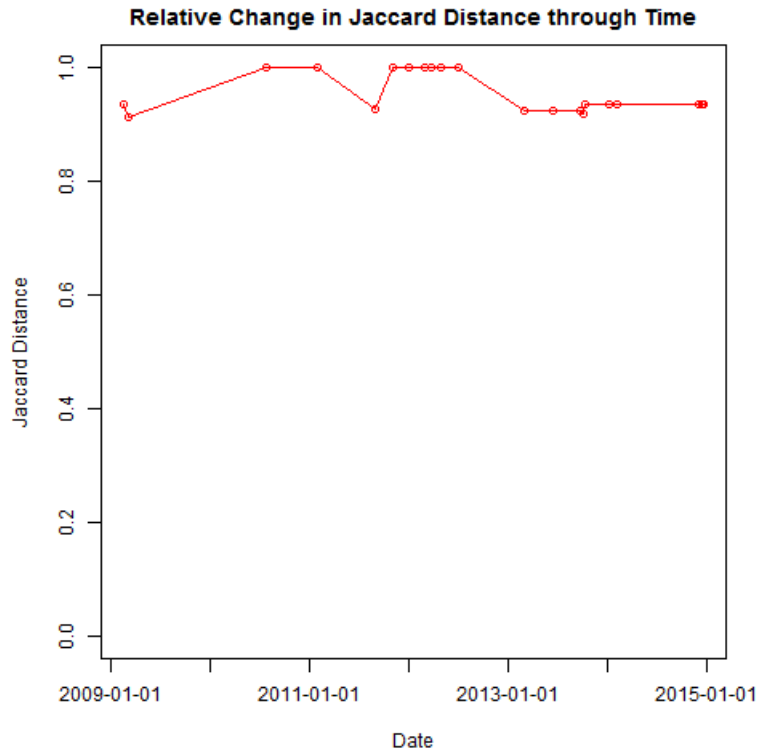


Figure 10: Jaccard Distance relative to the first memento URI# 5

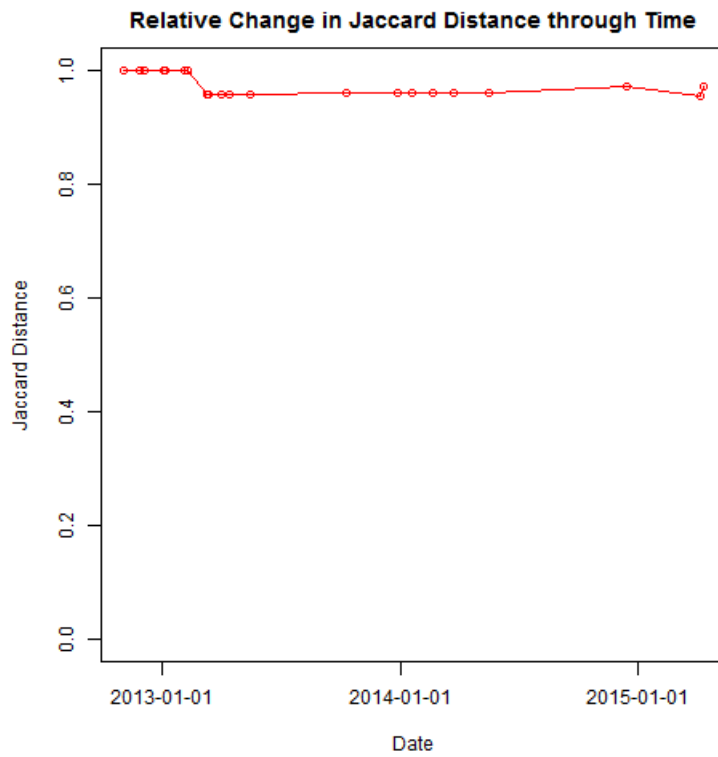


Figure 11: Jaccard Distance relative to the first memento URI# 7

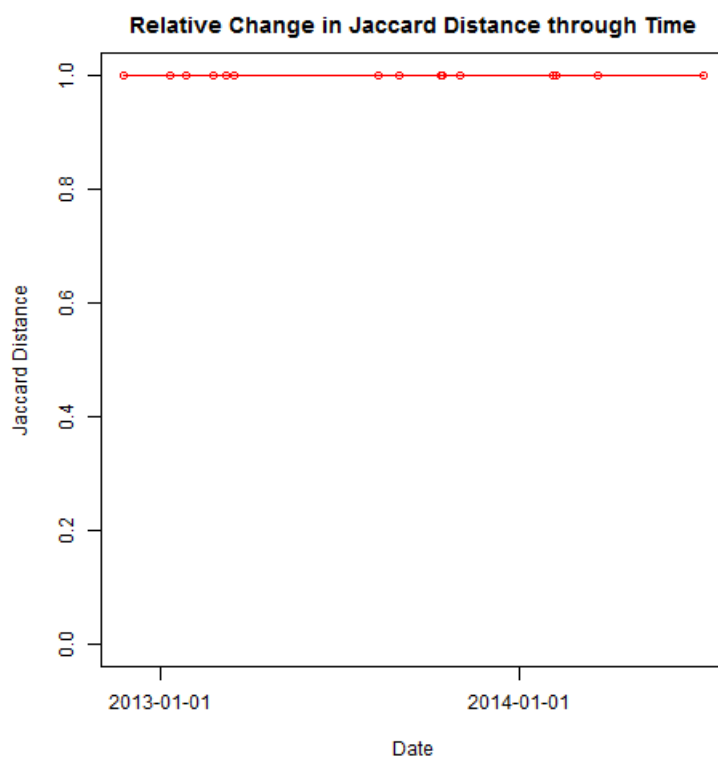


Figure 12: Jaccard Distance relative to the first memento URI# 8



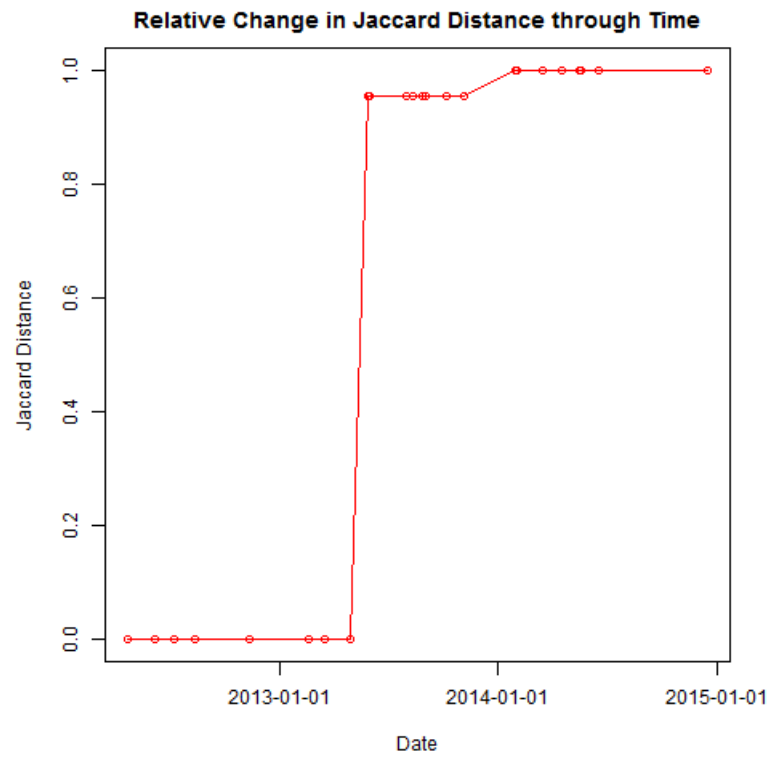


Figure 13: Jaccard Distance relative to the first memento URI# 9

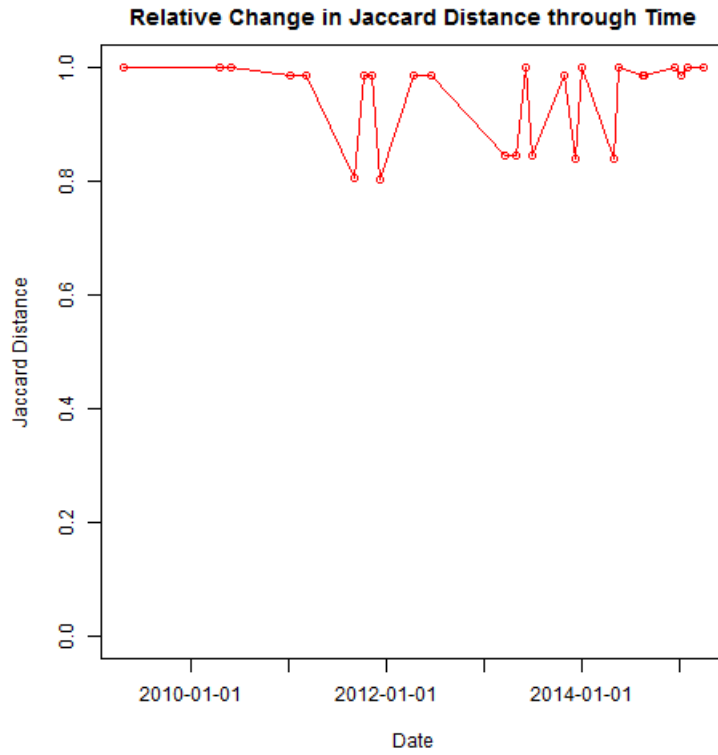


Figure 14: Jaccard Distance relative to the first memento URI# 10

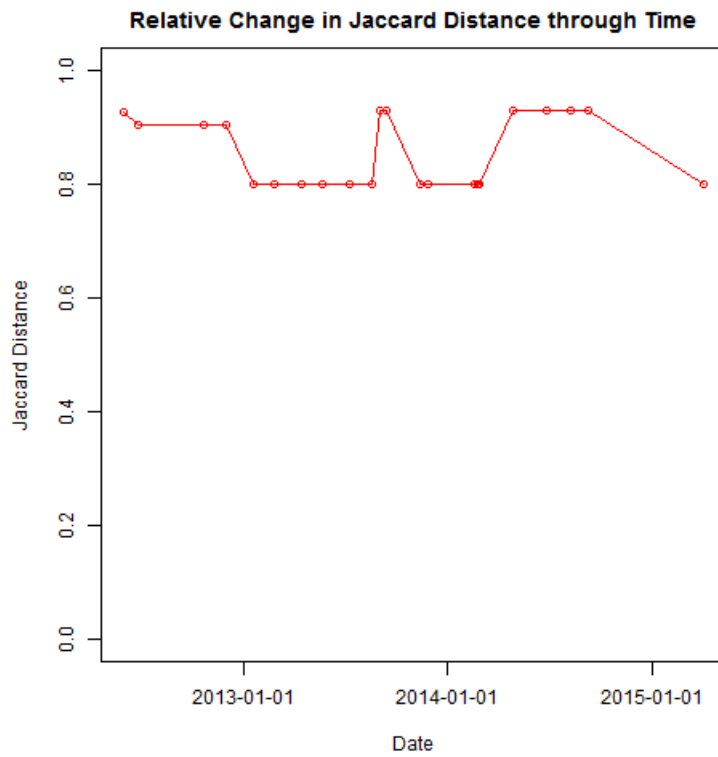


Figure 15: Jaccard Distance relative to the first memento URI# 11

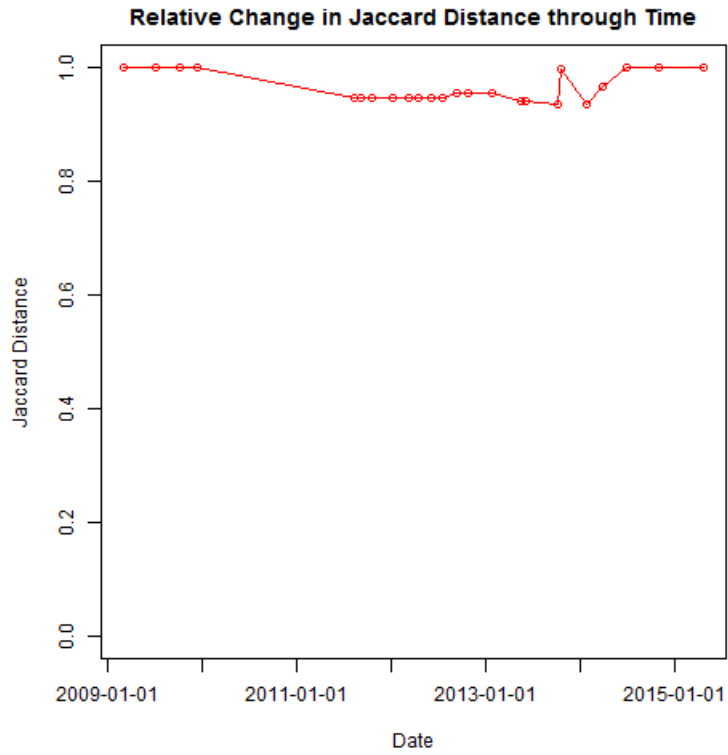


Figure 16: Jaccard Distance relative to the first memento URI# 12

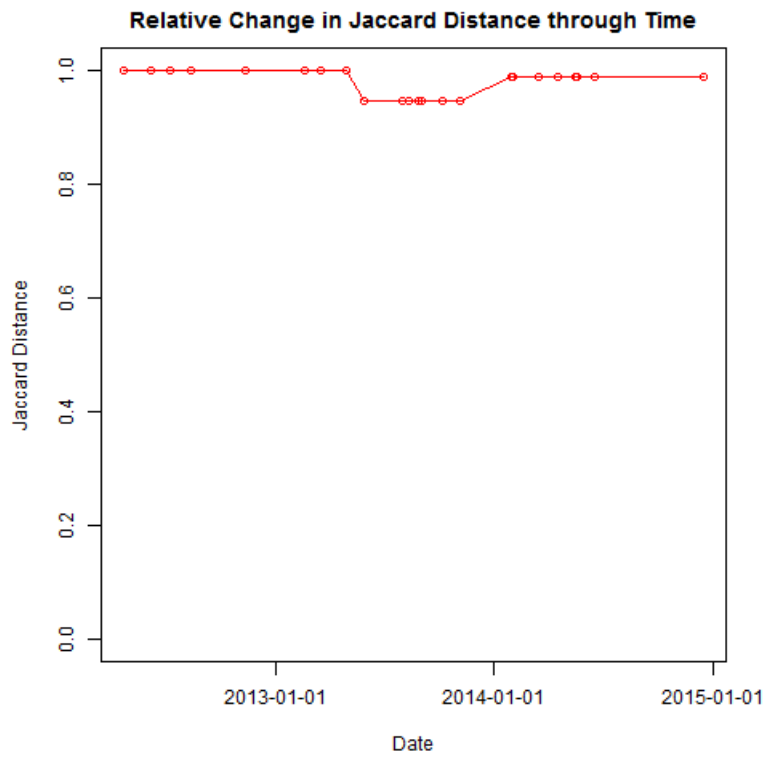


Figure 17: Jaccard Distance relative to the first memento URI# 13

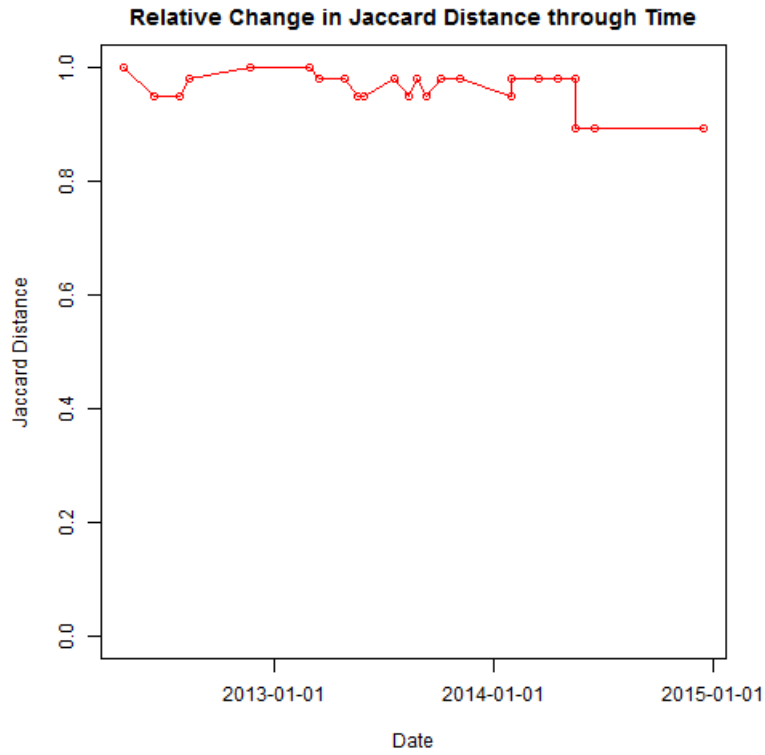


Figure 18: Jaccard Distance relative to the first memento URI# 14

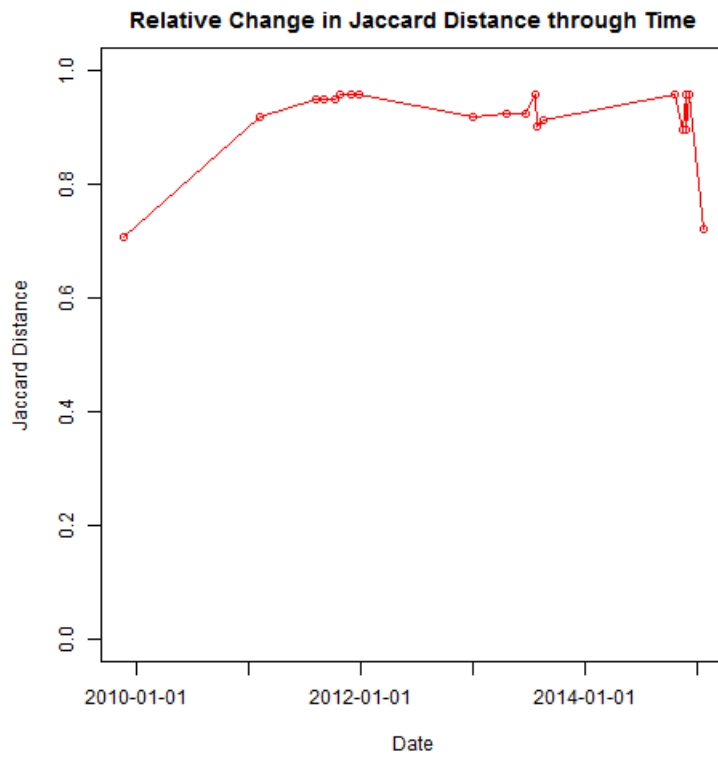


Figure 19: Jaccard Distance relative to the first memento URI# 15

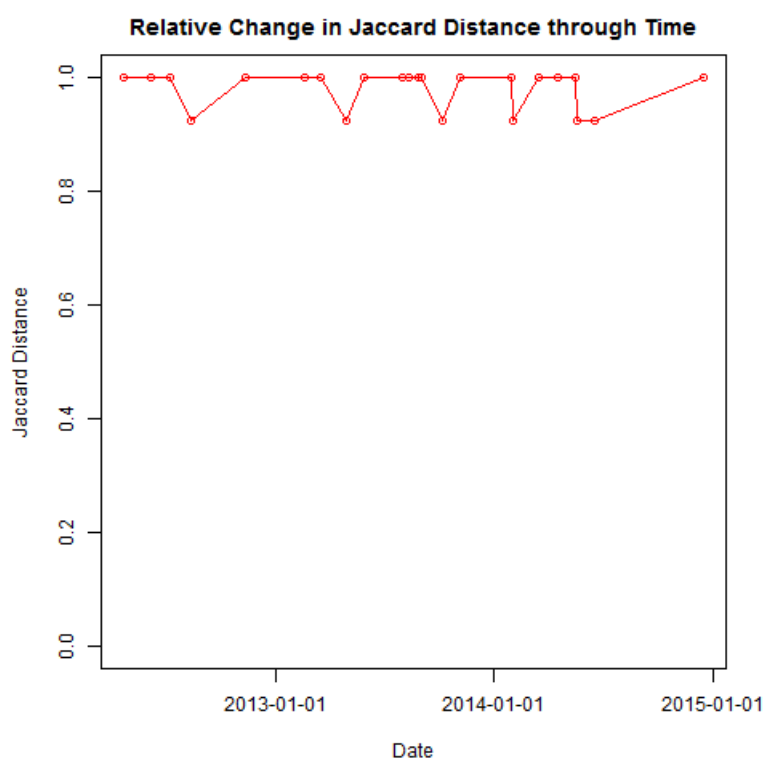


Figure 20: Jaccard Distance relative to the first memento URI# 16

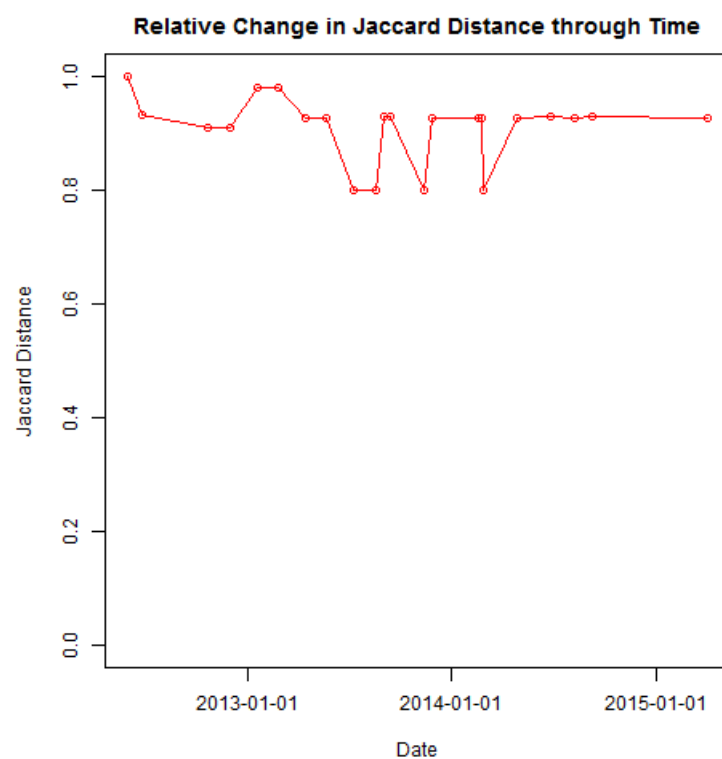


Figure 21: Jaccard Distance relative to the first memento URI# 17

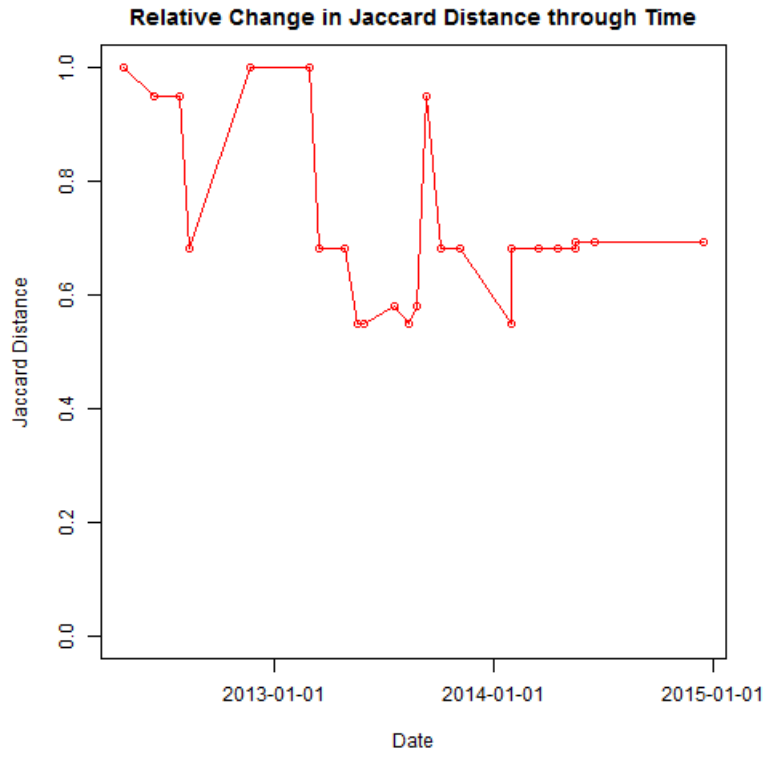


Figure 22: Jaccard Distance relative to the first memento URI# 18

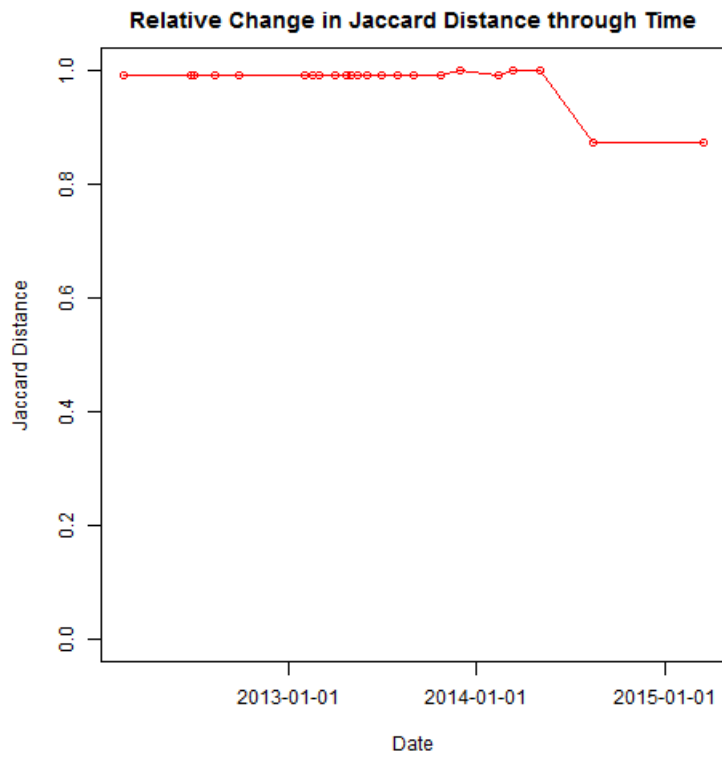


Figure 23: Jaccard Distance relative to the first memento URI# 19

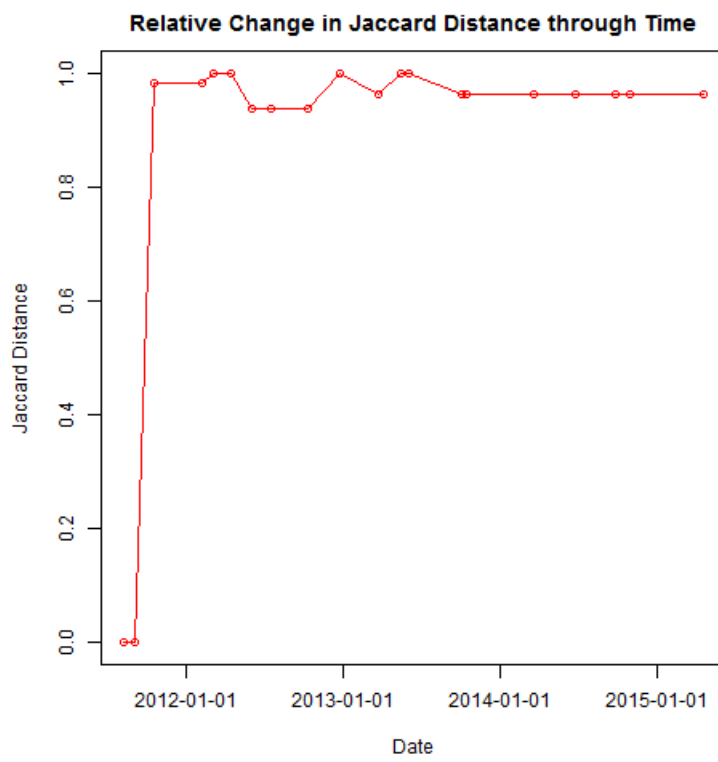


Figure 24: Jaccard Distance relative to the first memento URI# 20



## 4 Question 4

- Choose a news-related event
- Use twarc.py to collect 1000 tweets, every day for 5 different days
  - See: <https://github.com/edsu/twarc>
- For each day:
  - Create a wall
  - Build a tag/word cloud for each day
  - Create a map using GeoJSON & Github
    - \* <https://help.github.com/articles/mapping-geojson-files-on-github/>
- Discuss in detail lessons learned, experiences, etc.

### 4.1 Solution

Below are screen shots of five day tweets wall and word cloud.

- Installed twarc[4] to retrieve 1000 tweets.
- I used twarc to build tweets wall, word cloud and to retrieve geojson.
- "World Earth day" was my news trending topic and I retrieved 1000 tweets daily.
- I found few interesting things with word cloud and GeoJSON.
- GeoJSON showed exact geographical locations where this topic is being discussed each day.
- Word cloud clubbed all words and displayed the words in different sizes based on the word frequency in the tweets. Each day earth was the word most frequently used.
- I have uploaded GeoJSON files for each of the day to Github. GeoJSON can be visualized at <https://github.com/bharath559/cs851-s15/tree/master/Assignment4/Q4/geojson>

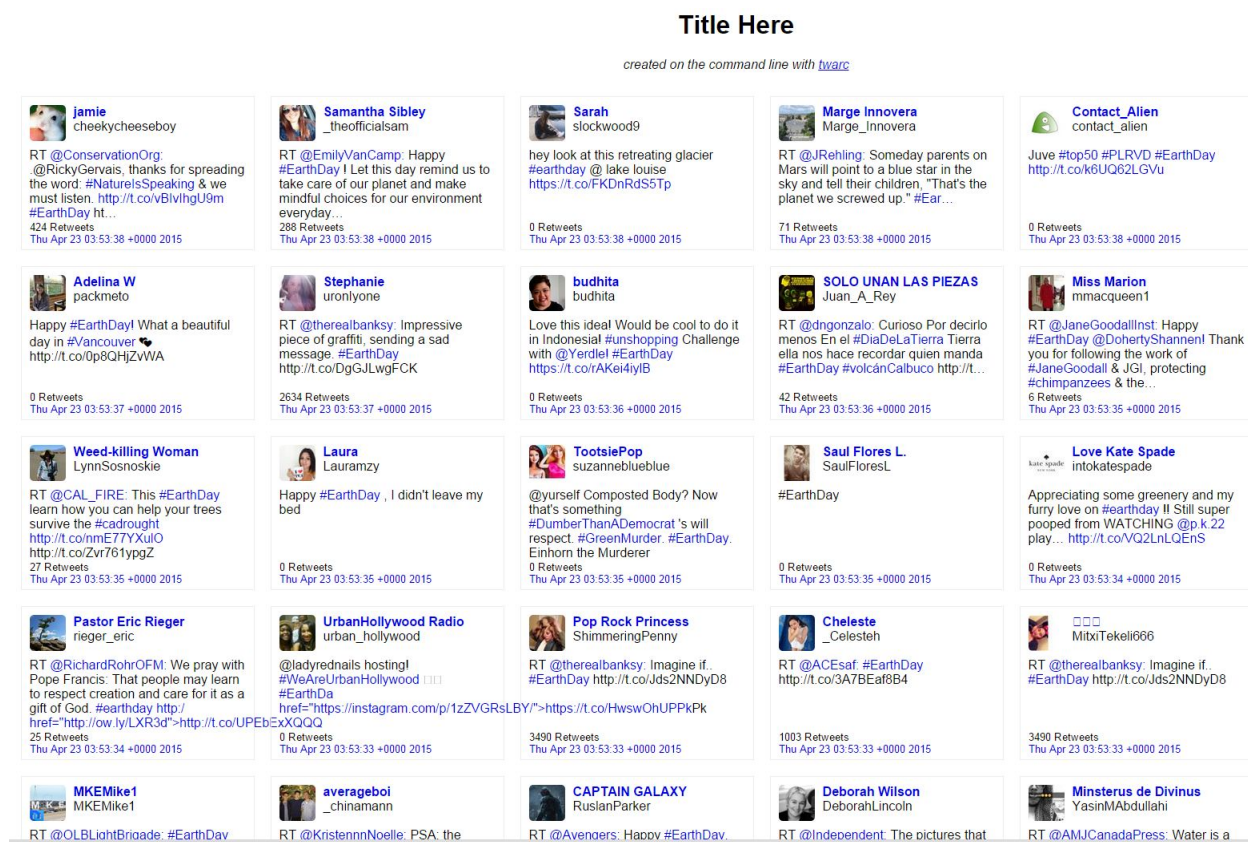


Figure 25: Tweets Wall Day1

## Title Here

created on the command line with [tvarc](#)


























 <b>Estelle cartidge</b> Stelc81 RT @loveyogabears: Happy Earth Day to all Earthlings ♥ blessings from @loveyogabears @shaktistudios @bikramliverpool @Stelc81 #earthday ht... 2 Retweets Thu Apr 23 22:45:36 +0000 2015	 <b>Sumicel USA</b> sumicelusa #Samsung patents a lightweight #fitness tracking band (+photos) - <a href="http://t.co/g16thFDV0J">http://t.co/g16thFDV0J</a> #News #EarthDay 0 Retweets Thu Apr 23 22:45:34 +0000 2015	 <b>Jessica Orwig</b> orwigwrites RT @mashable: We know #EarthDay can be depressing. So here are 4 reasons to be optimistic about the planet <a href="http://t.co/7mMK1SqYEB">http://t.co/7mMK1SqYEB</a> <a href="http://t.co/7mMK1SqYEB">http://t.co/7mMK1SqYEB</a> ... 65 Retweets Thu Apr 23 22:45:34 +0000 2015	 <b>Bri</b> BriThibodeaux RT @NigelBarker: It's #EarthDay & the perfect time for us to remember to look after & respect our beautiful planet #Roma #hugatree x <a href="http://t.co/7mMK1SqYEB">http://t.co/7mMK1SqYEB</a> ... 61 Retweets Thu Apr 23 22:45:32 +0000 2015	 <b>Consulado Ecu Queens</b> CEcuQueens RT @ApoyaAIEcuador: Our voices are stronger together than apart. Help our brothers and sisters in the Amazons #WeDeserveJustice #EarthDay h... 12 Retweets Thu Apr 23 22:45:31 +0000 2015
 <b>Irina Burkova</b> IrinaBurkova RT @Jasmine_Soul: Happy #EarthDay 🌱🌱 <a href="http://t.co/OvaWf2QrmE">http://t.co/OvaWf2QrmE</a> 4 Retweets Thu Apr 23 22:45:26 +0000 2015	 <b>Heather</b> ourcirtetime Happy #EarthDay! We made a bird feeder during the craft segment on today's <a href="http://t.co/AvDOQj87BL">http://t.co/AvDOQj87BL</a> video activity. <a href="http://t.co/AvDOQj87BL">http://t.co/AvDOQj87BL</a> 0 Retweets Thu Apr 23 22:45:26 +0000 2015	 <b>Hello Enterprise</b> HelloEnterpris RT @GoodsForGoodCo: BETTER LIFE - what-EVER! <a href="http://t.co/lcBhTLE2pQ">http://t.co/lcBhTLE2pQ</a> #BetterLife #EcoFriendly #homecare #EarthDay #Online #shopping <a href="http://t.co/7mMK1SqYEB">http://t.co/7mMK1SqYEB</a> ... 3 Retweets Thu Apr 23 22:45:23 +0000 2015	 <b>bsbafflesbrains</b> bsbafflesbrains RT @elliosch: @firedoglake nails it. Thankfuly @motherjones didn't realize their grave mistake earlier. <a href="http://t.co/TKQvW8yWVn">http://t.co/TKQvW8yWVn</a> <a href="http://t.co/TKQvW8yWVn">http://t.co/TKQvW8yWVn</a> ... 3 Retweets Thu Apr 23 22:45:23 +0000 2015	 <b>oops_im_amazing</b> DestinyFugate #EarthDay 0 Retweets Thu Apr 23 22:45:21 +0000 2015
 <b>Brian David Stockman</b> Carlisleboy RT @WAGNERGIRLE: #APictureIsWorthOneThousandWords 🌱 Obama's empty #earthday speech <a href="http://t.co/nlNk5oXZUsUs">http://t.co/nlNk5oXZUsUs</a> #tcot #pcra #ccot <a href="http://t.co/nlNk5oXZUsUs">http://t.co/nlNk5oXZUsUs</a> ... 10 Retweets Thu Apr 23 22:45:21 +0000 2015	 <b>Sean Nelson</b> opinionatedwino #earthday <a href="http://t.co/stfprs3jyB">http://t.co/stfprs3jyB</a> 0 Retweets Thu Apr 23 22:45:20 +0000 2015	 <b>r.e.e.n.a</b> reenahuffman13 RT @JoeyGraceffa: HAPPY #EarthDay in the spirit I've been playing @TreeStorygame try it out its so cute! #HaveFunDoGood <a href="http://t.co/FjGdzY0j2y">http://t.co/FjGdzY0j2y</a> 373 Retweets Thu Apr 23 22:45:18 +0000 2015	 <b>Lachlan Milt</b> MiltPozbk RT @4BillLewis: The most expensive #Starbucks drinks ever <a href="http://t.co/q8KUXVsp9S">http://t.co/q8KUXVsp9S</a> via @nypost #socialmedia #reteeet #coffee #StarbucksProble... 133 Retweets Thu Apr 23 22:45:18 +0000 2015	 <b>DavidPMiranda™</b> DavidPMiranda Happy Earth Day! So excited I got up before Dawn. Fortunately Dawn sleeps until noon. #EarthDay 0 Retweets Thu Apr 23 22:45:17 +0000 2015
 <b>Abbey Farkas</b> AbFark RT @radicalbytes: This #EarthDay remember personal recycling & water conservation initiatives are designed to shift blame from corporations... 237 Retweets Thu Apr 23 22:45:15 +0000 2015	 <b>sal</b> levett RT @WAGNERGIRLE: #APictureIsWorthOneThousandWords 🌱 Obama's empty #earthday speech <a href="http://t.co/nlNk5oXZUsUs">http://t.co/nlNk5oXZUsUs</a> #tcot #pcra #ccot <a href="http://t.co/nlNk5oXZUsUs">http://t.co/nlNk5oXZUsUs</a> ... 10 Retweets Thu Apr 23 22:45:15 +0000 2015	 <b>Jonathan Cohee</b> joncohee RT @PaigeToccoTerra: Kudos to my daughter Emmalyn, and all who dedicate their careers to preserving wildlife habitats #EarthDay 2 Retweets Thu Apr 23 22:45:15 +0000 2015	 <b>papadukeschaney</b> KevinDChaneySr RT @mashable: We know #EarthDay can be depressing. So here are 4 reasons to be optimistic about the planet <a href="http://t.co/7mMK1SqYEB">http://t.co/7mMK1SqYEB</a> <a href="http://t.co/7mMK1SqYEB">http://t.co/7mMK1SqYEB</a> ... 65 Retweets Thu Apr 23 22:45:15 +0000 2015	 <b>Jaime Silva</b> Jaime0 RT @FortuneMagazine: See Apple's glossy #EarthDay video <a href="http://t.co/1eXXRKjPcl">http://t.co/1eXXRKjPcl</a> <a href="http://t.co/1eXXRKjPcl">http://t.co/1eXXRKjPcl</a> <a href="http://t.co/1eXXRKjPcl">http://t.co/1eXXRKjPcl</a> ... 16 Retweets Thu Apr 23 22:45:14 +0000 2015
 <b>Yordano</b> YordanoNapa RT @NASA: Ah. #Earth. there's 0 Retweets Thu Apr 23 22:45:14 +0000 2015	 <b>SOLFARMERS</b> SOLFARMERS Thursdav dinner idea 4 #EarthDay by 0 Retweets Thu Apr 23 22:45:14 +0000 2015	 <b>Abbey Farkas</b> AbFark RT @radicalbytes: This #EarthDay 0 Retweets Thu Apr 23 22:45:14 +0000 2015	 <b>Keith Tharp</b> KeithTharp Our Earth Day haul at Hodason 0 Retweets Thu Apr 23 22:45:14 +0000 2015	 <b>Giovanna Quina Klapp</b> gioquina #chile #volcanes esta es 1 fumarola 0 Retweets Thu Apr 23 22:45:14 +0000 2015

Figure 26: Tweets Wall Day2

## Title Here

created on the command line with [twarc](#)








 <b>GreeneaseApp</b> GreeneaseApp Our fabulous intern, @sevincsey, blending in some smoothies at GWU's #EarthDay <a href="http://t.co/0w17ULrReB">http://t.co/0w17ULrReB</a> 0 Retweets Fri Apr 24 13:40:21 +0000 2015	 <b>Little Critterz</b> LittleCritterz3 #Sloths are an integral part of tropical rain forest ecosystems #EarthDay #nature #environment #Earthdayeveryday 0 Retweets Fri Apr 24 13:40:21 +0000 2015	 <b>Anjana</b> AnjanaKandhan RT @Avengers: Happy #EarthDay, from Earth's Mightiest! <a href="http://t.co/IRBeo0kbT">http://t.co/IRBeo0kbT</a> 1879 Retweets Fri Apr 24 13:40:21 +0000 2015	 <b>CJ Ronneberg</b> braintwat RT @deborahhallsted: As if on cue, snow strikes America as Pres. Obama talks global warming for #EarthDay [photos] <a href="http://t.co/bK4HYA4HL5">http://t.co/bK4HYA4HL5</a> 1 Retweets Fri Apr 24 13:40:16 +0000 2015	 <b>Connected Universe</b> ConnectdUnivrse Nature is a portal to deeper understanding ~ Malcom Carter, director of #ConnectedUniverse #EarthDay <a href="http://t.co/W5YNI7i9of">http://t.co/W5YNI7i9of</a> 0 Retweets Fri Apr 24 13:40:15 +0000 2015
 <b>OEM Off-Highway</b> OEMOffHighway It was a busy social media week for us! Check out @intematparis coverage with #OEMatINTERMAT and #EarthDay coverage with #GreenOEM 0 Retweets Fri Apr 24 13:40:14 +0000 2015	 <b>sevenbowie</b> sevenbowie RT @ConservationOrg: @RickyGervais, thanks for spreading the word: #NatureIsSpeaking & we must listen. <a href="http://t.co/vB1vIhgU9m">http://t.co/vB1vIhgU9m</a> #EarthDay ht... 732 Retweets Fri Apr 24 13:40:12 +0000 2015	 <b>Danny Ivatt</b> Daniellvatt According to #earthday crap where I am, I need to eat less meat and be more vegan. Neither of which is conducive to being swole. #nothanks 0 Retweets Fri Apr 24 13:40:10 +0000 2015	 <b>Ashley</b> Ashleyash27 RT @Martha_x15: @LisaVanderpump >Important reminder on #EarthDay. <a href="http://t.co/CiObnwtkm2&lt;...loll">http://t.co/CiObnwtkm2&lt;...loll</a> 10 Retweets Fri Apr 24 13:40:09 +0000 2015	 <b>Chris</b> cmdiehl29 RT @VictoriaJustice: How did you celebrate #EarthDay? Join me & @DoSomething by recycling clothes at @hmusal <a href="http://t.co/dFL8Ob0bAv">http://t.co/dFL8Ob0bAv</a> 382 Retweets Fri Apr 24 13:40:07 +0000 2015
 <b>gidgiegoomie</b> gidgiegoomie RT @Martha_x15: @LisaVanderpump >Important reminder on #EarthDay. <a href="http://t.co/CiObnwtkm2&lt;...loll">http://t.co/CiObnwtkm2&lt;...loll</a> 10 Retweets Fri Apr 24 13:40:04 +0000 2015	 <b>Dolores Briseno</b> mofodolo RT @Martha_x15: @LisaVanderpump >Important reminder on #EarthDay. <a href="http://t.co/CiObnwtkm2&lt;...loll">http://t.co/CiObnwtkm2&lt;...loll</a> 10 Retweets Fri Apr 24 13:40:04 +0000 2015	 <b>giveaway 100K coins</b> coins_100k RT @SlashedBeauty: It's #EarthDay! Discover #ecofriendly #skincare brand, @derma_e & win their best products: <a href="http://t.co/tkQ1TyvA3J">http://t.co/tkQ1TyvA3J</a> <a href="http://t.co/tkQ1TyvA3J">http://t.co/tkQ1TyvA3J</a> 19 Retweets Fri Apr 24 13:40:04 +0000 2015	 <b>PinkGreen Creative</b> WeArePinkGreen #FeatureFontFriday is here, with #earthday as the theme. @miradax goes green with Grafolita. #typography #design #... <a href="http://t.co/6rHY8MnjAA">http://t.co/6rHY8MnjAA</a> 0 Retweets Fri Apr 24 13:39:57 +0000 2015	 <b>Agility CSR</b> AgilityCSR Good #EarthDay memories indeed. <a href="https://t.co/YNFdUpE1y3">https://t.co/YNFdUpE1y3</a> 0 Retweets Fri Apr 24 13:39:55 +0000 2015
 <b>Team Lupold Thoman</b> LupoldThoman Green Team!!! Thank you for all your hard work! Wonderful video about recycling! @NH_Elementary #earthday <a href="http://t.co/3obgMhVwPY">http://t.co/3obgMhVwPY</a> 0 Retweets Fri Apr 24 13:39:51 +0000 2015	 <b>Ashwini Kaikini</b> ashkaikini RT @Nriyandhi: The earth & I have a lot in common. We're made mostly of water & every year we both get hotter! ~ @TheEllenShow   @ashkai... 1 Retweets Fri Apr 24 13:39:48 +0000 2015	 <b>Abbi</b> Abi_Frann RT @Martha_x15: @LisaVanderpump >Important reminder on #EarthDay. <a href="http://t.co/CiObnwtkm2&lt;...loll">http://t.co/CiObnwtkm2&lt;...loll</a> 10 Retweets Fri Apr 24 13:39:46 +0000 2015	 <b>1victoria bailey</b> gabjamom1 RT @Martha_x15: @LisaVanderpump >Important reminder on #EarthDay. <a href="http://t.co/CiObnwtkm2&lt;...loll">http://t.co/CiObnwtkm2&lt;...loll</a> 10 Retweets Fri Apr 24 13:39:46 +0000 2015	 <b>gareth lee-bell</b> gingerkermi RT @LuxeTiffany: Happy #EarthDay <a href="https://t.co/Wmectr8IVb">https://t.co/Wmectr8IVb</a> (Victoria Falls, Zambia) 3 Retweets Fri Apr 24 13:39:41 +0000 2015
 <b>Mauri Solages</b> MauriSolages RT @ColleenNGRID: Gearing up! 0 Retweets Fri Apr 24 13:39:41 +0000 2015	 <b>Rita</b> Rin2379 RT @findmeaifit: #Competition 0 Retweets Fri Apr 24 13:39:41 +0000 2015	 <b>EZstorit.com</b> ezstorit RT @PublicStorage: Inspired by 0 Retweets Fri Apr 24 13:39:41 +0000 2015	 <b>Claire Basiloña</b> ClaireBasilona Calling on Young People to Lead Us 0 Retweets Fri Apr 24 13:39:41 +0000 2015	 <b>Burke County Georgia</b> BurkeCountyGA Master Gardeners Plant Sale today at 0 Retweets Fri Apr 24 13:39:41 +0000 2015

Figure 27: Tweets Wall Day3



## Title Here

created on the command line with [twarc](#)


























 <b>aaron k.</b> snidesnark RT @EricWolfson: Best. #ClimateChange. Tweet. EVER. (via @ScottWesterfeld) #ClimateAction #EarthDay #EarthDay2015 #EarthDayEveryDay http://... 1101 Retweets Sat Apr 25 15:14:48 +0000 2015	 <b>Boulder Book Store</b> boulderbooks RT @techtionary: Today #Boulder @alfalfasmarket B'Earthday #EarthDay Bash 11-4 https://t.co/yTlmVgHJm w/ @AsherBrewing @GrowingGardensB 2 Retweets Sat Apr 25 15:14:44 +0000 2015	 <b>CHANDRA PRAKASH</b> dv891798 RT @BrightHouseNow: April 25, Earthday Birthday, Central Florida Fairgrounds Join @1011wjrr & Bright House Networks! Lineup of bands: http://t.co/8gMoi5PpQ0 2 Retweets Sat Apr 25 15:14:44 +0000 2015	 <b>Nita Pujiastuti</b> nee_ta15 RT @tupperwareID: Sebelum mulai menanam #mangrove, jangan lupa berfoto bersama :) #EarthDay http://t.co/8HW81UZAHa 3 Retweets Sat Apr 25 15:14:39 +0000 2015	 <b>CHANDRA PRAKASH</b> dv891798 RT @BrightHouseNow: TODAY- Earthday Birthday, Central Florida Fairgrounds Join @1011wjrr & Bright House Networks! Lineup of bands: http://t.co/8gMoi5PpQ0 1 Retweets Sat Apr 25 15:14:30 +0000 2015
 <b>पंडित जी Pandit ji</b> dheerajrazdan @AgentSaffron #Nepal #NepalEarthquake God have mercy on the Nepalese population #IndiaWithNepal #EarthDay #earthquake 1 Retweets Sat Apr 25 15:14:29 +0000 2015	 <b>Nita Pujiastuti</b> nee_ta15 RT @tupperwareID: Ayo kita tanam 10.000 #mangrove sebagai bentuk cinta kita pada lingkungan bersama WWF & Tupperware! #EarthDay http://t.co/http://t.co/YwMD39TbwD 3 Retweets Sat Apr 25 15:14:29 +0000 2015	 <b>haltonmastergardener</b> HaltonMGs RT @henhere: All set up at Georgetown Market Place for #EarthDay Fair, till 4pm. @ImagineHH http://t.co/1DcdT4er05 1 Retweets Sat Apr 25 15:14:14 +0000 2015	 <b>TheEcoVeganEvents</b> TomasinaBarr RT @missrainbird: This #EarthDay, watch Cowspracy online for just \$1 and learn the truth about who's killing the planet! #documentary http://t.co/Dp28uvT96U 1 Retweets Sat Apr 25 15:14:08 +0000 2015	 <b>Robert Rossachaj</b> RRossachaj RT @the_sca: SCA CEO Jaime Matyas greets 400+ volunteers at today's #EarthDay service project w/ @AmericanExpress in #DC. http://t.co/0yhnkn... 2 Retweets Sat Apr 25 15:13:58 +0000 2015
 <b>Oak Hills Sports</b> OakHillsSports RT @kenmeibers: Giving back to the community. #earthday http://t.co/LpwwQ74Gyx 3 Retweets Sat Apr 25 15:13:57 +0000 2015	 <b>Ahmed Numen</b> AhmedNumen Sad to hear the terrible news of massive earthquake hit in Nepal. My prayers and thoughts are with the people of Nepal #EarthDay #Earthquake 0 Retweets Sat Apr 25 15:13:53 +0000 2015	 <b>Primped and Improper</b> primpdNimproper Pace Team Cleanup! #earthday #oakville #joshuacreek #realtor #realtorlife #realestate #century21... https://t.co/dg4De0l2Ox 0 Retweets Sat Apr 25 15:13:42 +0000 2015	 <b>money honey</b> g_money_honey29 RT @Martha_x15: @LisaVanderpump -Important reminder on #EarthDay. http://t.co/CiObnwtkm2<...lol 58 Retweets Sat Apr 25 15:13:40 +0000 2015	 <b>Crystal Coast</b> CrystalCoast_NC The #CrystalCoast #EarthDay Celebration is going on @ #FortMacon. Enjoy FREE kids activities, demos & more, 10am-2pm. http://t.co/J9vXiOnKsC 0 Retweets Sat Apr 25 15:13:38 +0000 2015
 <b>Yenny T</b> UNSWstudentleif RT @ThePatriot143: CLIMATE ACTIVISTS IN MELTDOWN AS 'SCEPTIC' GETS AUSTRALIAN GOVERNMENT FUNDING http://t.co/mv4xisaWuG #EarthDay 1246 Retweets Sat Apr 25 15:13:36 +0000 2015	 <b>Estuk</b> T_F_L RT @reason: News flash: #NorthKorea wins #EarthDay for record 45th time in a row! http://t.co/JALvCO3WT HT @Neoavataara http://t.co/MUynUG0... 1246 Retweets Sat Apr 25 15:13:24 +0000 2015	 <b>Angelo W</b> Waglio_O RT @AlexineAudiard: #EarthDay http://t.co/gmoF0DireT 36 Retweets Sat Apr 25 15:13:24 +0000 2015	 <b>Goran Kolasinac</b> onikapova RT @MariaSharapova: Everyday is #EarthDay #HugATree http://t.co/SS2svT2AmA 424 Retweets Sat Apr 25 15:13:23 +0000 2015	 <b>Ferguson-Florisant</b> FergFlorSchools Great crowd @ our Little Creek Nature Area celebrating #EarthDay big thanks to @missioncontinue for building seating http://t.co/hPUtDsyAGJ 0 Retweets Sat Apr 25 15:13:20 +0000 2015
 <b>Kim Rees</b> kimba420 RT @WeeLittleHawk: @Stantec 	 <b>li'l gal with aGlock</b> lilglockgal Al Gore fact #EarthDay 	 <b>mercedes.m</b> LuisaM56_ RT @AlexineAudiard: #EarthDay 	 <b>Warsame Abdi</b> warsameeabdi RT @Stanford: Stanford's new 	 <b>Barrett Henry Team</b> bhenryteam Pace Team Cleanup! #oakville 

Figure 28: Tweets Wall Day4

## Title Here

created on the command line with [twarc](#)

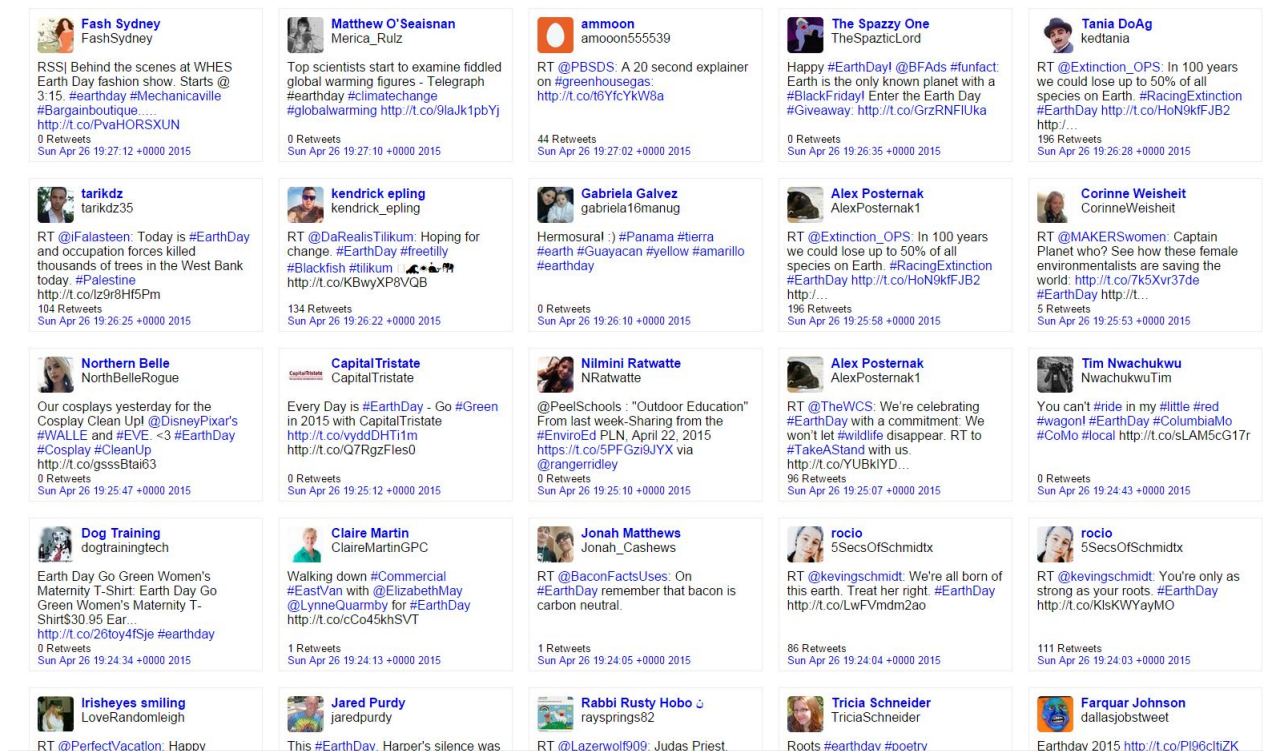


Figure 29: Tweets Wall Day5







Figure 31: Tweets Word Cloud Day2













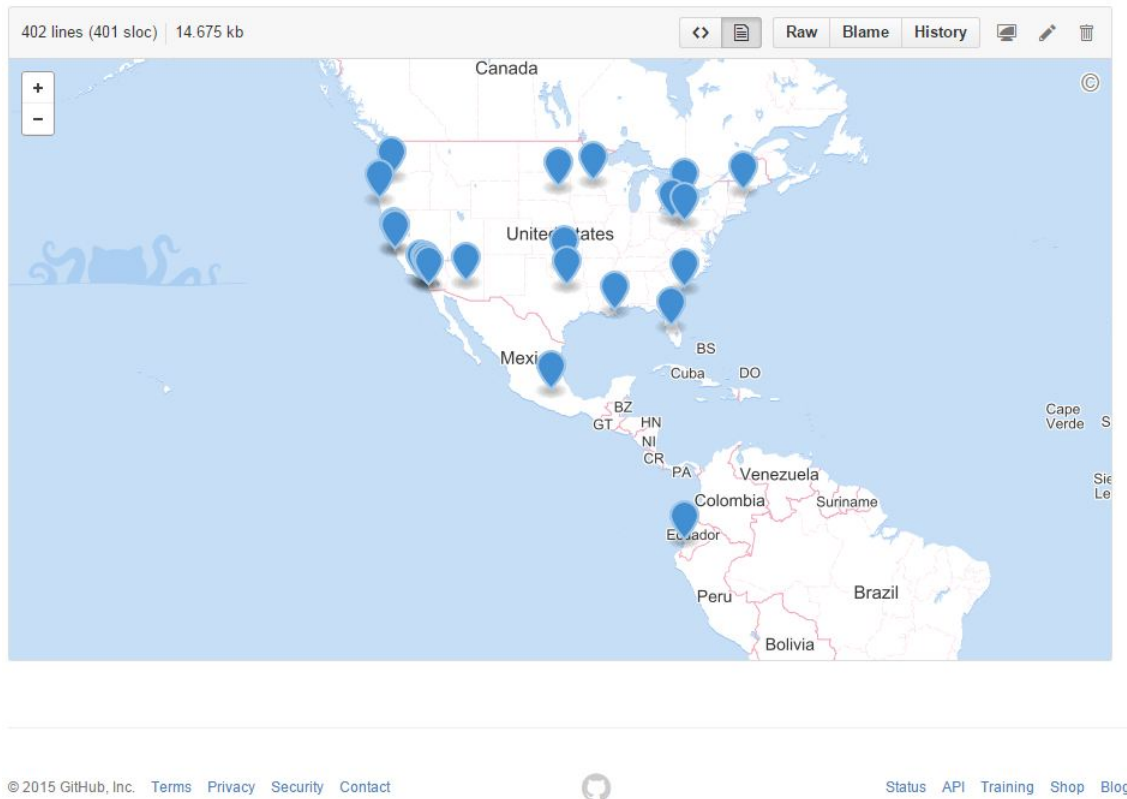


Figure 35: Map for Day1

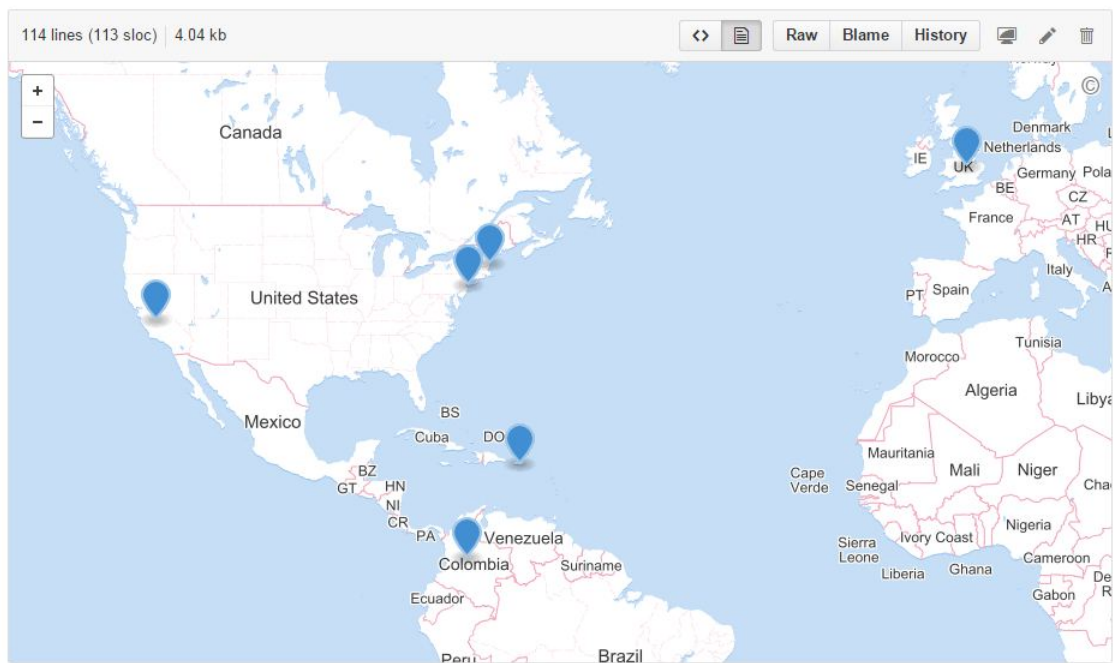


Figure 36: Map for Day2

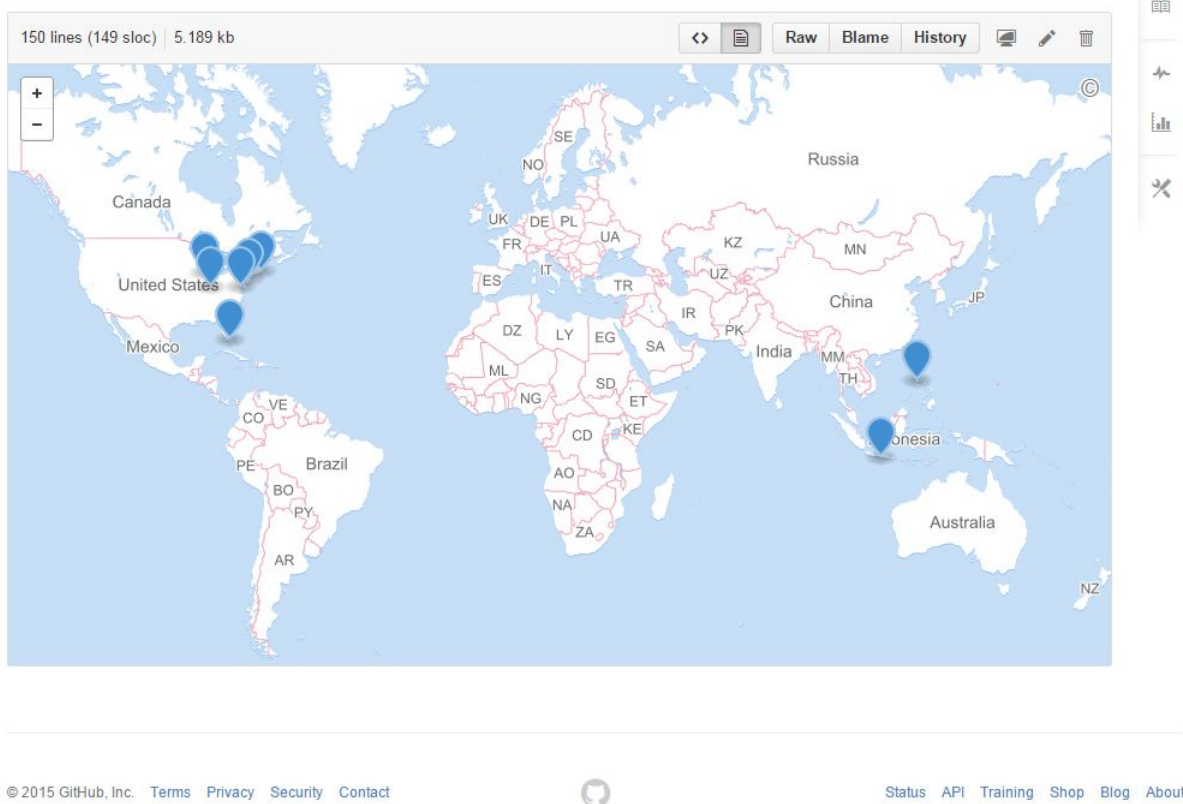


Figure 37: Map for Day3

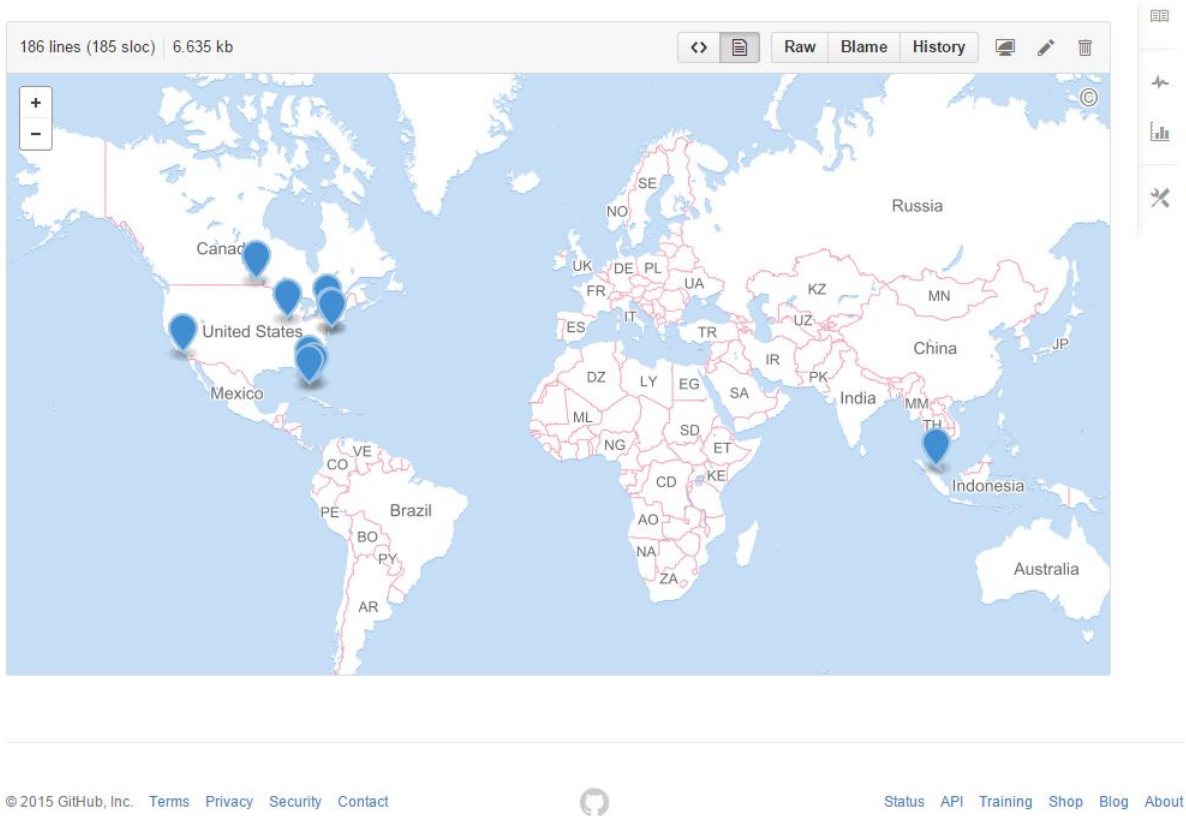


Figure 38: Map for Day4

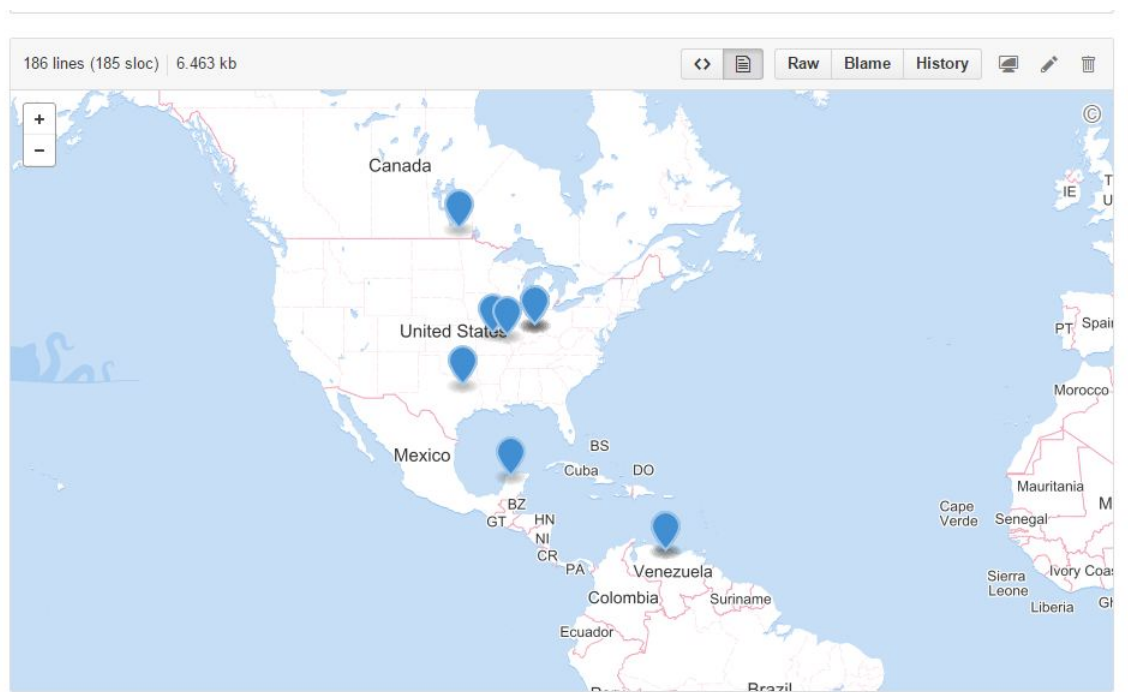


Figure 39: Map for Day5

# Bibliography

- [1] Cumulative. <http://www.r-bloggers.com>.
- [2] jusText. <https://pypi.python.org/pypi/jusText/2.1.0>.
- [3] timelapse. <https://github.com/burg/timelapse/>.
- [4] twarc. <https://github.com/edsu/twarc>.
- [5] Wordcount. <https://github.com/mlafeldt/google-python-class/blob/master/basic/solution/wordcount.py>.  
Accessed: 2015-04-03.