

Visualizing Twitter Trends

By Kevin Wong

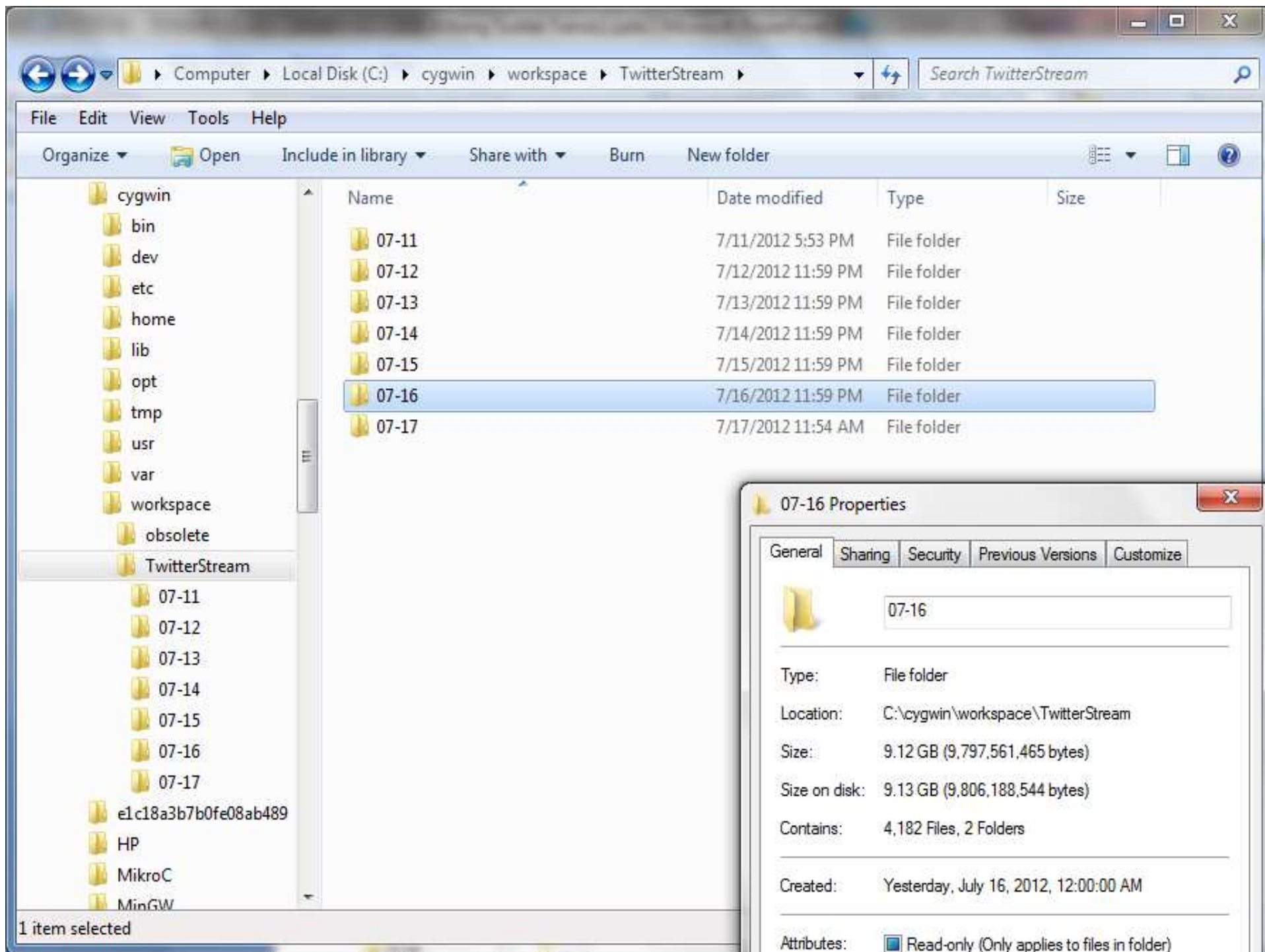
Mentor: James Abello & Yifan Hu

Project Description

- What are people talking about?
- Are there emerging trends?
- Tweets are public
- Visualize tweets over time & space
- **How does slang usage differ...**
 - Between regions?
 - Over time?

Collecting Twitter Data

```
{
  "in_reply_to_status_id_str": null, "place": null, "favorited": false, "text": "(\\u00d8\\u00b5\\u02d8\\u01aa)", "in_reply_to_screen_name": null, "in_reply_to_user_id_str": null, "retweet_count": 0, "in_reply_to_status_id": null, "source": "\\u00Ca href=\\\"http:\\/\\/twitterous.com\\\" rel=\\\"nofollow\\\"\\u003E\\u003C\\u003C\\u003E\", \"coordinates\": null, \"geo\": null, \"retweeted\": false, \"truncated\": false, \"created_at\": \"Tue Jul 17 06:53:01 +0000 2012\", \"id_str\": \"225120907964850176\", \"in_reply_to_user_id\": null, \"contributors\": null, \"user\": {
    \"profile_sidebar_fill_color\": \"#ffffff\", \"followers_count\": 615, \"profile_image_url_https\": \"https:\\/\\/si0.twimg.com/profile_images\\/2398710539\\/5a9cm9v279mas9h3d9y1_normal.jpeg\", \"profile_image_url\": \"http:\\/\\/a0.twimg.com/profile_images\\/2398710539\\/5a9cm9v279mas9h3d9y1_normal.jpeg\", \"listed_count\": 1, \"contributors_enabled\": false, \"geo_enabled\": false, \"follow_request_sent\": null, \"following\": null, \"profile_sidebar_border_color\": \"#CODEED\", \"screen_name\": \"riscayunita\", \"default_profile_e\": false, \"description\": \"Unperfect | Unbeautiful | Friendly | Single keep woles and stay high | 27 juni 1997 (\\u0254 \\u02c6\\u0437\\u02c6)\\u02c6\\u03b5\\u02c6 c)\", \"is_translator\": false, \"show_all_inline_media\": false, \"profile_use_background_image\": true, \"profile_text_color\": \"#333333\", \"profile_background_image_url_https\": \"https:\\/\\/si0.twimg.com/profile_background_images\\/570127226\\/17mg8aavsx07uhx9b0we.jpeg\", \"profile_background_image_url\": \"http:\\/\\/a0.twimg.com/profile_bac kground_images\\/570127226\\/17mg8aavsx07uhx9b0we.jpeg\", \"created_at\": \"Sun Mar 20 04:29:39 +0000 2011\", \"default_profile_image\": false, \"time_zone\": \"Central Time (US & Canada)\", \"profile_link_color\": \"#0084B4\", \"protected\": false, \"url\": \"http:\\/\\/www.facebook.com\\/riscayunita\", \"id_str\": \"269127316\", \"favourites_count\": 58, \"friends_co unt\": 321, \"name\": \"RISCA YUNITA\", \"statuses_count\": 29464, \"verified\": false, \"profile_background_color\": \"#CODEED\", \"id\": 269127316, \"lang\": \"en\", \"notifications\": null, \"profile_background_tile\": true, \"utc_offset\": -21600, \"location\": \"Balaraja, Tangerang\", \"id\": 225120907964850176, \"entities\": {\"hashtags\": [], \"user_mentions\": [], \"urls\": []}}
  }
  \"in_reply_to_status_id_str\": null, \"place\": null, \"favorited\": false, \"text\": \"Apa itu cium cium ini masih siang\", \"in_reply_to_screen_name\": null, \"in_reply_to_user_id_str\": null, \"retweet_count\": 0, \"in_reply_to_status_id\": null, \"source\": \"\\u003Ca href=\\\"http:\\/\\/twitter.com/download/iphone\\\" rel=\\\"nofollow\\\"\\u003ETwitter for iPhone\\u003C\\u003C\\u003E\", \"coordinates\": null, \"geo\": null, \"retweeted\": false, \"truncated\": false, \"created_at\": \"Tue Jul 17 06:53:01 +0000 2012\", \"id_str\": \"225120907985825794\", \"in_reply_to_user_id\": null, \"contributors\": null, \"user\": {
    \"profile_sidebar_fill_color\": \"#DDEEF6\", \"followers_count\": 139, \"profile_image_url_https\": \"https:\\/\\/si0.twimg.com/profile_images\\/2399640344\\/jwqh26252syOuli373z5_normal.jpeg\", \"profile_image_url\": \"http:\\/\\/a0.twimg.com/profile_images\\/2399640344\\/jwqh26252syOuli373z5_normal.jpeg\", \"listed_count\": 0, \"contributors_enabled\": false, \"geo_enabled\": false, \"follow_request_sent\": null, \"following\": null, \"profile_sidebar_border_color\": \"#CODEED\", \"screen_name\": \"ASLizzyFR\", \"default_profile\": false, \"description\": \"VRP from @Friends RP | lead rapper kece-nya AfterSchool B | 92lines | gue sama schee member geng kapak B | Bulliers dan Bullied ~~~\", \"is_translator\": false, \"show_all_inline_media\": false, \"profile_use_background_image\": true, \"profile_text_color\": \"#333333\", \"profile_background_image_url_https\": \"https:\\/\\/si0.twimg.com/profile_background_images\\/601649520\\/igap2vbwlvsdvvcyb3b.png\", \"profile_background_image_url\": \"http:\\/\\/a0.twimg.com/profile_bac kground_images\\/601649520\\/igap2vbwlvsdvvcyb3b.png\", \"created_at\": \"Sat Jun 16 03:34:12 +0000 2012\", \"default_profile_image\": false, \"time_zone\": \"Bangkok\", \"profile_link_color\": \"#0084B4\", \"protected\": false, \"url\": \"http:\\/\\/lizzykecetiadatara.jangandibantah.co m\", \"id_str\": \"609652946\", \"favourites_count\": 4, \"friends_count\": 139, \"name\": \"Park Rizzy ule2606\\u0F61\", \"statuses_count\": 4645, \"verified\": false, \"profile_background_color\": \"#CODEED\", \"id\": 609652946, \"lang\": \"en\", \"notifications\": null, \"profile_background_t ile\": true, \"utc_offset\": 25200, \"location\": \"After School's Dorm\", \"id\": 225120907985825794, \"entities\": {\"hashtags\": [], \"user_mentions\": [], \"urls\": []}}
  }
  \"place\": null, \"in_reply_to_screen_name\": null, \"text\": \"#MePasaSoloAmi quedarme encerrada 1 hora y media sola en un ascensor en el piso 14.\", \"in_reply_to_user_id_str\": null, \"retweet_count\": 0, \"in_reply_to_status_id\": null, \"source\": \"web\", \"retweeted\": false, \"truncated\": false, \"id_str\": \"22512090797744 1280\", \"coordinates\": null, \"geo\": null, \"in_reply_to_user_id\": null, \"contributors\": null, \"created_at\": \"Tue Jul 17 06:53:01 +0000 2012\", \"user\": {
    \"listed_count\": 0, \"contributors_enabled\": false, \"Geo_enabled\": true, \"follow_request_sent\": null, \"notifications\": null, \"profile_sidebar_border_color\": \"#eeeeee\", \"loc ation\": \"\", \"default_profile\": false, \"following\": null, \"is_translator\": false, \"show_all_inline_media\": false, \"profile_use_background_image\": true, \"description\": \"Amo los animales, soy obsesiva con cuidar el medio ambiente. Soy muy positiva sociable y caprichosa siempre consigo lo que quiero, no bajo los brazos f\\u00e9cilmente\", \"profile_text_color\": \"#333333\", \"screen_name\": \"gandisolll\", \"profile_background_image_url_https\": \"https:\\/\\/si0.twimg.com/profile_background_im ages\\/597022180\\/tpcd3ameoozrofexq75.jpeg\", \"profile_background_image_url\": \"http:\\/\\/a0.twimg.com/profile_background_images\\/597022180\\/tpcd3ameoozrofexq75 .jpeg\", \"default_profile_image\": false, \"time_zone\": \"Santiago\", \"profile_link_color\": \"#1F9EC7\", \"id_str\": \"229546176\", \"favourites_count\": 127, \"friends_count\": 147, \"prot ected\": false, \"statuses_count\": 31365, \"verified\": false, \"created_at\": \"Wed Dec 22 17:47:25 +0000 2010\", \"profile_background_color\": \"#FCF3F3\", \"name\": \"Tndira\", \"profile_background_tile\": true, \"url\": null, \"id\": 229546176, \"lang\": \"es\", \"utc_offset\": -14400, \"profile_s
```



Tweets

- Live stream
 - prgm runs all day
- 6MB (~3600 tweets)/min
- 9GB/day
- 30% with location
- 7% with useful locations
- .7% with geotag
- Tweets in JSON

The JSON Format

- JavaScript Object Notation
- “a lightweight data-interchange format”
- Relatively clean text
- Consists of:
 - Name/value pairs (object/struct)
 - Ordered lists (arrays)

Some JSON

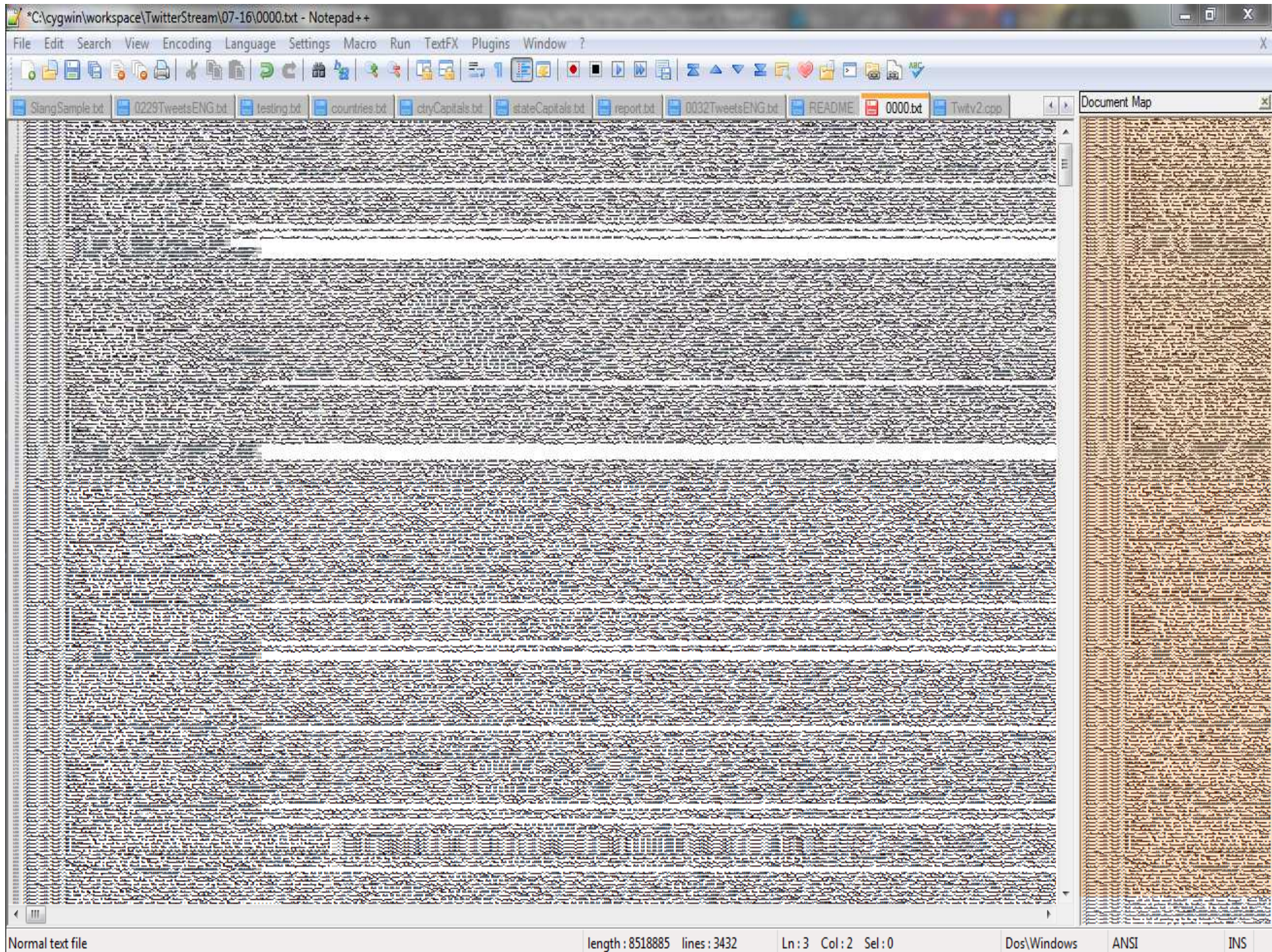
```
Kevin@Kevin-PC /cygdrive/c/users/kevin/downloads/cjsonfiles/cjson
$ ./test
{
  "name": "Jack (\\"Bee\\") Nimble",
  "format": {
    "type": "rect",
    "width": 1920,
    "height": 1080,
    "interlace": false,
    "frame rate": 24
  }
}
```




```

1 {
2   "entities":
3   {
4     "urls": [],
5     "media": [{"type": "photo", "indices": [15, 35], "display_url": "pic.twitter.com/XnaKz5Kj", "media_url_https": "https://p.twimg.com/Ax2Ta7qCQAK7930.jpg", "id_str": "21549043787850096", "expanded_url": "http://twitter.com/ChelseaAllHood/status/21549043787850096/"},
6     "user_mentions": [],
7     "hashtags": [{"text": "webcamtoy", "indices": [4, 14]}]
8   },
9   "text": "not #webcamtoy http://t.co/XnaKz5Kj",
10  "place": null,
11  "in_reply_to_status_id": null,
12  "in_reply_to_user_id": null,
13  "truncated": false,
14  "possibly_sensitive_editable": true,
15  "id_str": "21549043787850096",
16  "coordinates": null,
17  "geo": null,
18  "retweeted": false,
19  "source": "\u003Cp href='\"http://neave.com/webcam/\"' rel='\"nofollow\"'\u003EWebcam Toy\u003C/a\u003E",
20  "possibly_sensitive": false,
21  "in_reply_to_screen_name": null,
22  "created_at": "Wed Jun 20 17:04:59 +0000 2012",
23  "in_reply_to_status_id_str": null,
24  "contributors": null,
25  "user":
26  {
27    "default_profile": false,
28    "lang": "en",
29    "location": "",
30    "is_translator": false,
31    "contributors_enabled": false,
32    "geo_enabled": true,
33    "time_zone": "Quito",
34    "profile_sidebar_border_color": "FF0000",
35    "follow_request_sent": null,
36    "following": null,
37    "profile_use_background_image": true,
38    "description": "you need me, man i don't need you +",
39    "id_str": "342420274",
40    "default_profile_image": false,
41    "show_all_inline_media": false,
42    "verified": false,
43    "profile_text_color": "000000",
44    "profile_background_image_url_https": "https://a0.twimg.com/profile_background_images/557535011/b.gif",
45    "profile_background_image_url": "http://a0.twimg.com/profile_background_images/557535011/b.gif",
46    "favourites_count": 355,
47    "friends_count": 43,
48    "profile_link_color": "0084B4",
49    "created_at": "Tue Jul 26 00:14:47 +0000 2011",
50    "protected": false,
51    "statuses_count": 612,
52    "profile_background_color": "002233",
53    "followers_count": 94,
54    "url": null,
55    "profile_image_url_https": "https://a0.twimg.com/profile_images/2302423545/O4wed5Em_normal",
56    "profile_image_url": "http://a0.twimg.com/profile_images/2302423545/O4wed5Em_normal",
57    "screen_name": "ChelseaAllHood",
58    "name": "Chas Finster",
59    "listed_count": 0,
60    "notifications": null,
61    "profile_background_tile": true,
62    "id": "342420274",
63    "utc_offset": -18000,
64    "profile_sidebar_fill_color": "F7B1AD"
65  },
66  "retweet_count": 0,
67  "favorite": false,
68  "id": "21549043787850096",
69  "in_reply_to_user_id_str": null
70 }
71

```



Justin Bieber ✓

@justinbieber

#BELIEVE is on ITUNES and in STORES WORLDWIDE! - SO MUCH LOVE FOR THE FANS...you are always there for me and I will always be there for you. MUCH LOVE. thanks
All Around The World - <http://www.youtube.com/justinbieber>

Following



17,245 TWEETS

123,022 FOLLOWING

25,091,606 FOLLOWERS

Tweet to Justin Bieber

@justinbieber

Tweets

Following

Followers

Favorites

Lists

Recent images



Similar to Justin Bieber



Christian Beadles ✓ @LittlecBea...
Follow



Rvan Butler ✓ @itsRvanButler

Tweets All / No replies



Justin Bieber @justinbieber

6h

going to stay in the TOP 3 on BILLBOARD this week again! thank u. Proud of OUR ALBUM! #BIEBERBLAST #BELIEVE - itunes.apple.com/us/album/belie...
🎵 View album



Justin Bieber @justinbieber

6h

but now...IM TIRED> but i love u Australia. gonna meet with some special kids right now.
Expand



Justin Bieber @justinbieber

7h

that was incredible. thanks to all the fans that came out to @sunriseon7 - had to keep singing for you guys. too much fun.
Expand



Justin Bieber @justinbieber

9h

Grumpy justin haha youtube.com/watch?v=8sv2Iz...
📺 View video



Justin Bieber @justinbieber

10h

good morning AUSTRALIA!! let's do this! @sunriseon7 see u soon
Expand

Tweet Metadata

- Picture URL's, hashtags, user mentions
- Tweet info
 - Time of post
 - text
 - Retweet
- User info
 - Account creation date
 - language
 - Location
 - Follower count

Identifying Trends

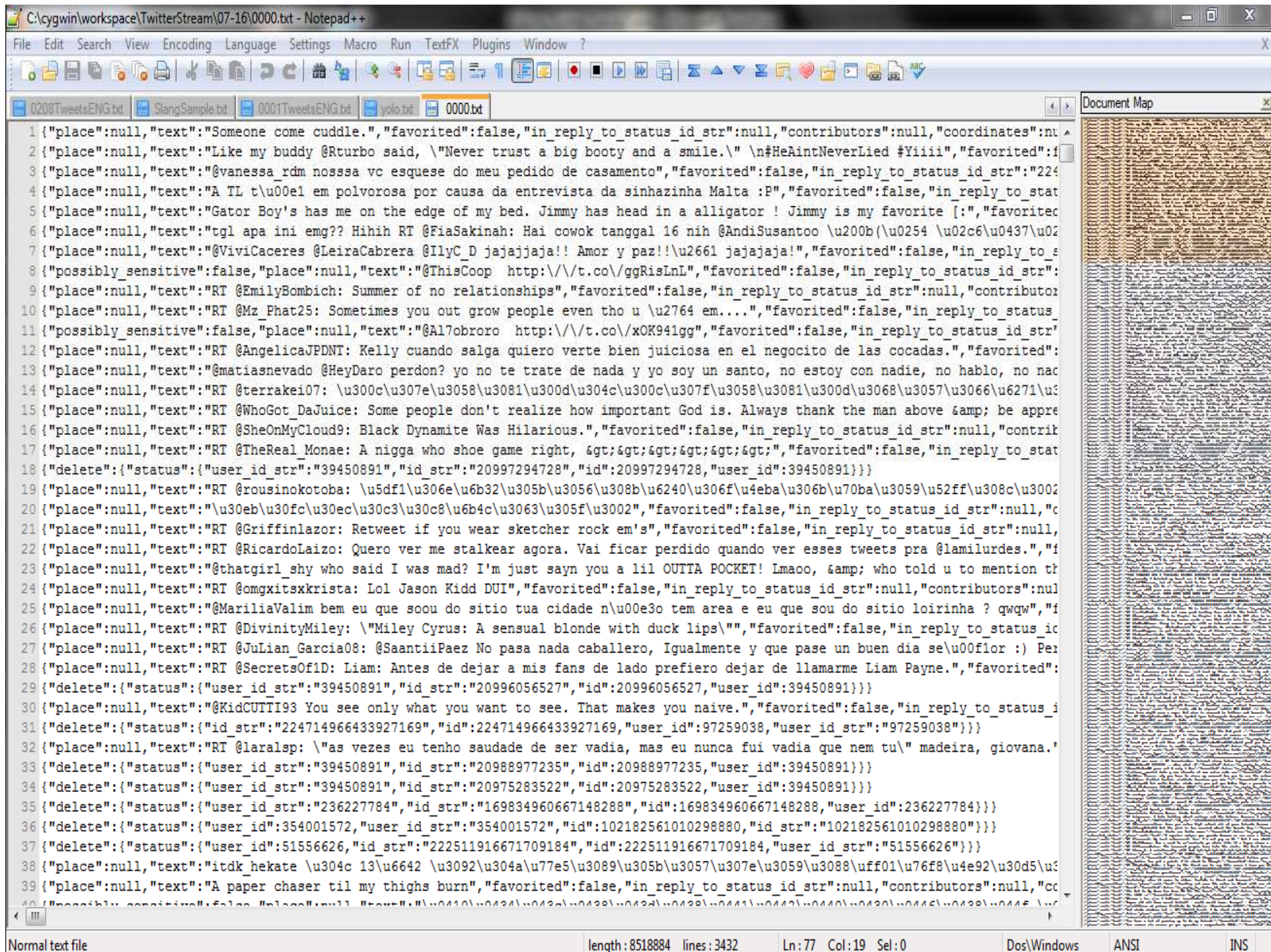
- Have sense of time
- Have sense of location
- Now what's interesting?
 - Frequency Normalization
 - “Slang Velocity”
- Focus: Slang

Identifying Trends

- Frequency Normalization
 - Compares a word's expected vs. actual usage
 - Expected = last week's frequency / 7 days
 - Actual = today's frequency / mins since 12:00
 - Score = actual / expected
 - Interest = $\max(\text{score}, 1/\text{score})$
- Slang Velocity
 - Compute a word's "instantaneous" usage
 - Change in frequency / increment time
 - Positive velocity over extended time = trending

Approach

- Collect Raw tweets using libcurl
- Parse with cJSON library
- Remove non-english tweets
- Remove tweets without location info(~30%)
- Remove tweets that do not contain slang (100+ word slang dictionary)
- Discard useless locations
- July 16:
 - ~5000000 tweets total
 - 1 499 014 with location
 - 103 358 identifiable locations
- Visualization



C:\Users\Kevin\Desktop\EclipseWorkspace\VelocityFinder\output 07-16\plothour0.txt - Notepad++

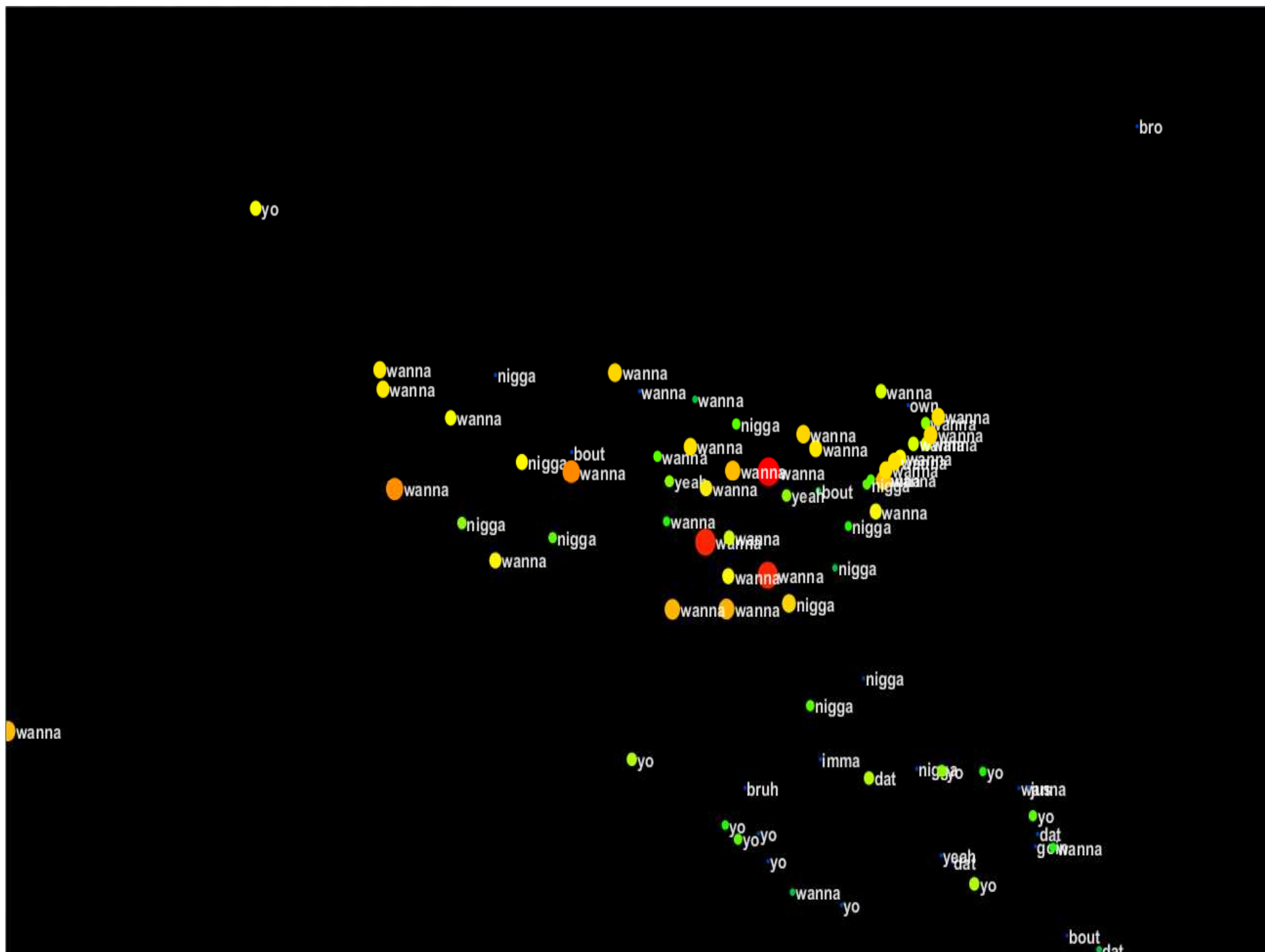
File Edit Search View Encoding Language Settings Macro Run TextFX Plugins Window ?

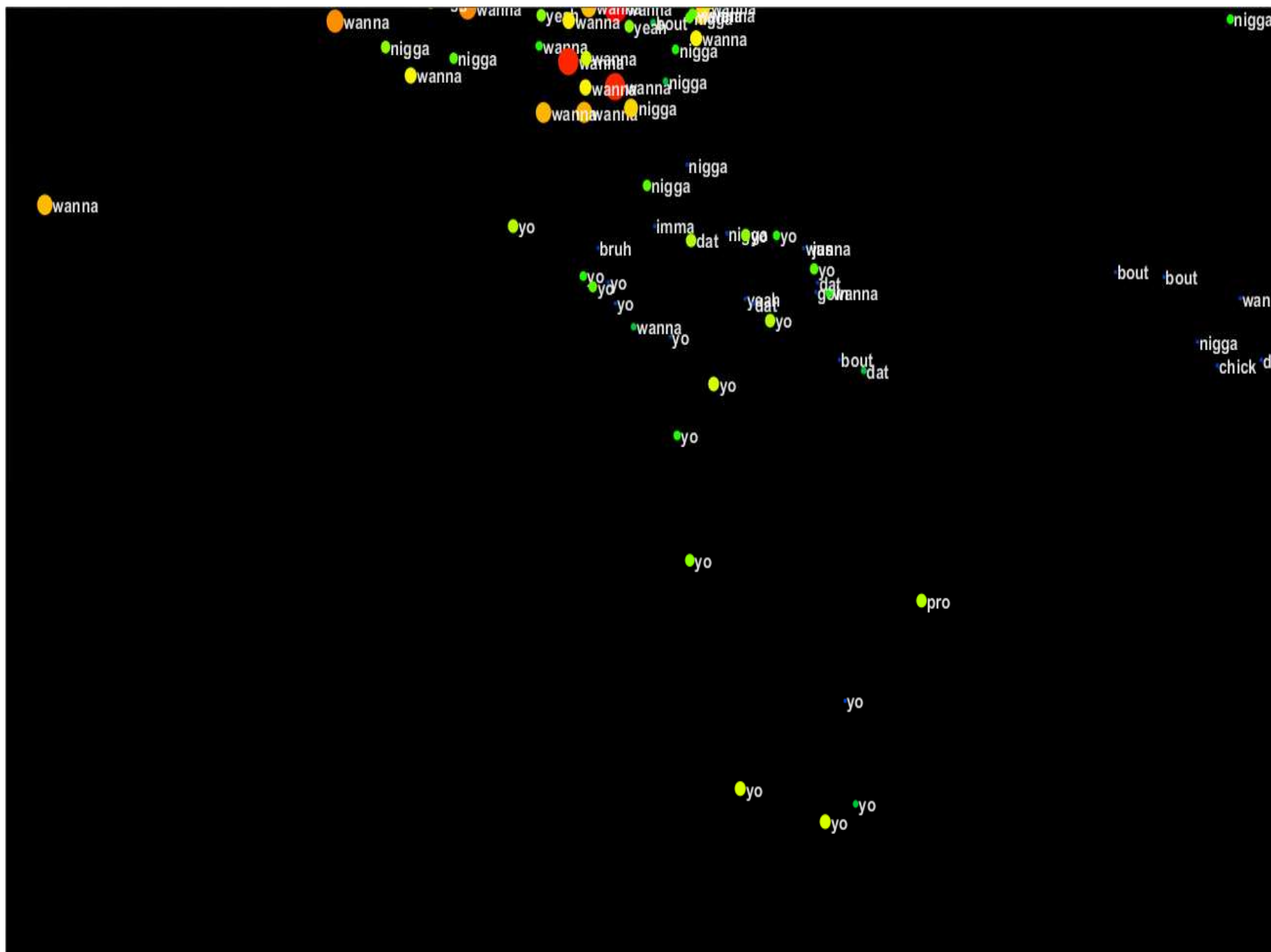
0208TweetsENG.txt SlangSample.txt 0001TweetsENG.txt yolo.txt 0000.txt 0000TweetsENG.txt plothour0.txt

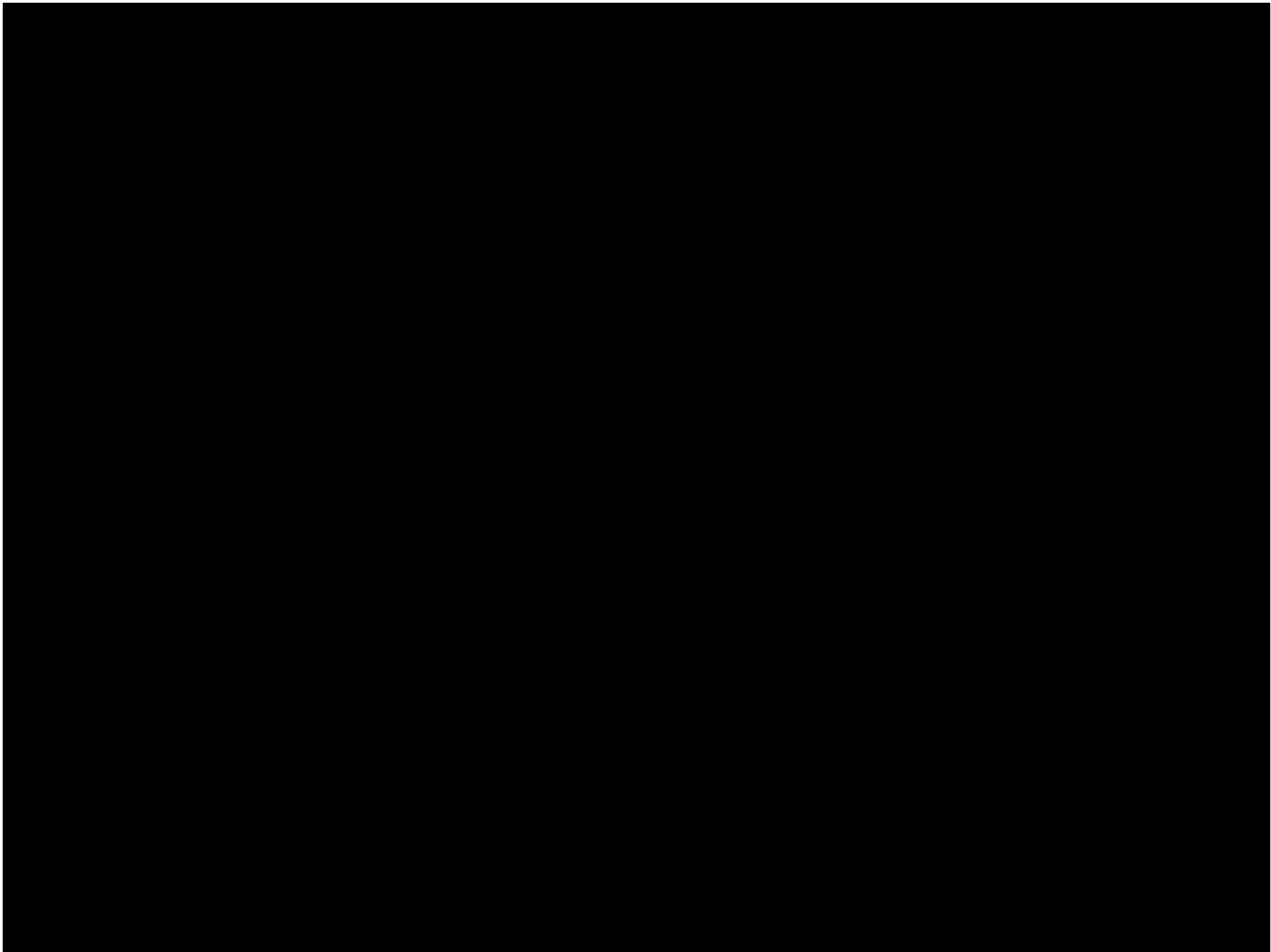
Document Map

```
1 (afghanistan,34.4666666666667,69.1833333333334,wanna,0)
2 (african,4.38333333333334,18.5833333333333,yall,0)
3 (alabama,32.3833333333333,-86.3666666666666,wanna,63)
4 (alaska,58.3666666666667,-134.5833333333334,yo,12)
5 (albania,41.3,19.8166666666666,bs,0)
6 (angola,-8.83333333333334,13.25,yo,0)
7 (antigua,17.3333333333333,-61.8,jus,0)
8 (antilles,12.0833333333334,-69.0,dat,0)
9 (argentina,-36.5,-60.0,yo,3)
10 (arizona,33.4333333333333,-112.0166666666667,wanna,3)
11 (arkansas,34.7333333333334,-92.2333333333333,wanna,78)
12 (armenia,40.1666666666666,44.5166666666666,yeah,0)
13 (aruba,12.5333333333333,-70.0333333333333,yeah,1)
14 (australia,-35.25,149.133333333333,yeah,7)
15 (austria,48.2,16.3666666666667,wanna,0)
16 (azerbaijan,40.4833333333334,49.9333333333333,cuz,0)
17 (bahamas,25.0833333333333,-77.3333333333333,nigga,1)
18 (bahrain,26.1666666666666,50.5,bs,0)
19 (bangladesh,23.7166666666666,90.4333333333334,wit,0)
20 (barbados,13.0833333333334,-59.5,wanna,1)
21 (belarus,53.8666666666667,27.5,yeah,0)
22 (belgium,50.85,4.35,dat,0)
23 (belize,17.3,-88.5,bruh,0)
24 (benin,6.38333333333334,2.7,bro,0)
25 (bosnia,43.8666666666667,18.4333333333334,chill,0)
26 (botswana,-24.75,25.95,dey,0)
27 (brazil,-15.7833333333333,-47.9166666666666,pro,3)
28 (brunei,4.8666666666667,115.0,nah,0)
29 (bulgaria,42.75,23.3333333333333,wanna,0)
30 (caledonia,-22.2833333333335,166.5,wanna,0)
31 (california,38.5166666666666,-121.5,wanna,25)
32 (cambodia,11.55,104.916666666667,yeah,1)
33 (cameroon,3.83333333333335,11.5833333333334,bout,0)
34 (canada,45.45,-75.7,wanna,5)
35 (cayman,19.3333333333333,-81.4,imma,0)
36 (chad,12.1666666666666,14.9833333333333,yo,0)
37 (chile,-33.4,-70.6666666666667,yo,6)
38 (china,39.9166666666666,116.333333333333,wanna,0)
39 (colombia,4.5666666666666,-74.0,yo,4)
40 (colorado,39.75,-104.866666666666,wanna,27)
41 (congo,-4.3333333333333,15.25,bruh,0)
```

Normal text file length : 9682 lines : 213 Ln: 79 Col: 1 Sel: 0 UNIX ANSI INS







Questions & Comments

References

- Project Proposal:
 - <http://dimacs.rutgers.edu/REU/2012/proposed.html>
- cJSON Library:
 - [Json.org](http://json.org)
- cURL library:
 - <http://curl.haxx.se/>
- Twitter stream:
 - <https://stream.twitter.com/1/statuses/sample.json>