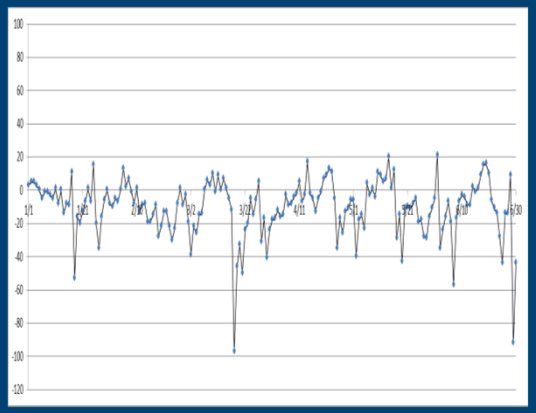
Michigan SAS Conference
June 7, 2017

Lag Models with Social Response Outcomes

David J. Corliss, PhD

Director, Peace-Work

OUTLINE

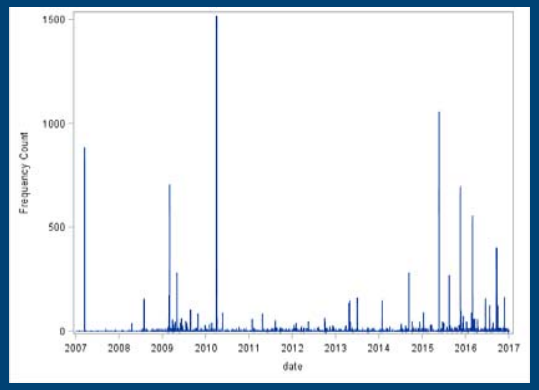


Lag Models

Social Media Data

Linking Solar Storms & Tweets

Hate Tweets and Violence



Google Trends

Summary

Lag Models – An Example

Lead / Lag	(None)		-3 Weeks		-2 Weeks		-1 Week		Even		+1 Week		+2 Weeks	
Correlation	0.399		0.718		0.747		0.526		0.399		0.193		-0.355	
Week	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
				1.00										
				1.12		1.00								
				0.89		1.12		1.00						
Week 1	0.22%	1.00	0.22%	1.06	0.22%	0.89	0.22%	1.12	0.22%	1.00	0.22%		0.22%	
Week 2	0.526	1.12	0.80%	1.12	0.80%	1.06	0.80%	0.89	0.526	1.12	0.80%	1.00	0.80%	
Week 3	0.56%	0.89	0.56%	1.05	0.56%	1.12	0.56%	1.06	0.56%	0.89	0.56%	1.12	0.56%	1.00
Week 4	-0.50%	1.06	-0.50%	0.69	-0.50%	1.05	-0.50%	1.12	-0.50%	1.06	-0.50%	0.89	-0.50%	1.12
Week 5	-0.18%	1.12	-0.18%	0.74	-0.18%	0.69	-0.18%	1.05	-0.18%	1.12	-0.18%	1.06	-0.18%	0.89
Week 6	0.15%	1.05	0.15%	0.53	0.15%	0.74	0.15%	0.69	0.15%	1.05	0.15%	1.12	0.15%	1.06
Week 7	-1.01%	0.69	-1.01%	0.60	-1.01%	0.53	-1.01%	0.74	-1.01%	0.69	-1.01%	1.05	-1.01%	1.12
Week 8	-1.74%	0.74	-1.74%	0.52	-1.74%	0.60	-1.74%	0.53	-1.74%	0.74	-1.74%	0.69	-1.74%	1.05
Week 9	-1.24%	0.53	-1.24%	0.81	-1.24%	0.52	-1.24%	0.60	-1.24%	0.53	-1.24%	0.74	-1.24%	0.69
Week 10	-0.18%	0.60	-0.18%	0.79	-0.18%	0.81	-0.18%	0.52	-0.18%	0.60	-0.18%	0.53	-0.18%	0.74
Week 11	0.40%	0.52	0.40%	0.97	0.40%	0.79	0.40%	0.81	0.40%	0.52	0.40%	0.60	0.40%	0.53
Week 12	0.10%	0.81	0.10%		0.10%	0.97	0.10%	0.79	0.10%	0.81	0.10%	0.52	0.10%	0.60
Week 13	0.05%	0.79	0.05%		0.05%		0.05%	0.97	0.05%	0.79	0.05%	0.81	0.05%	0.52
Week 14	-0.20%	0.97	-0.20%		-0.20%		-0.20%		-0.20%	0.97	-0.20%	0.79	-0.20%	0.81
												0.97		0.79
														0.97

Creating Lagged Variables

```
%macro lag(var1,var2,lag_n);
data lag_data;
   set source_data;

   x1_lag0 = &var1;
   x2_lag0 = &var2;

   %do i = 1 %to &lag_n;
      %let j = %eval(&i - 1);
      x1_lag&i = lag(x1_lag&j);
      x2_lag&i = lag(x2_lag&j);
   %end;
run;
%mend;

%lag(gdp,unemployment,3);
```

A Model With Lagged Variables

```
proc reg data=ht_data;
   model count = count_tminus1 forecl_change_rate
                 PCI rtw_state Poverty_pct_tminus2 ;
run;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-64.24238	23.89955	-2.69	0.0120
count_tminus1	1	2.33827	0.46725	5.00	<.0001
forecl_change_rate	1	-15.60110	4.52329	-3.45	0.0018
PCI	1	0.00150	0.00047598	3.16	0.0038
RTW_State	1	14.38219	5.94202	2.42	0.0222
Poverty_pct_tminus2	1	1.25805	0.55883	2.25	0.0324

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	21446	4289.28049	22.61	<.0001
Error	28	5311.18600	189.68521		
Corrected Total	33	26758			

Root MSE	13.77263	R-Square	0.8015
Dependent Mean	50.70471	Adj R-Sq	0.7661
Coeff Var	27.16242		

Social Media Data

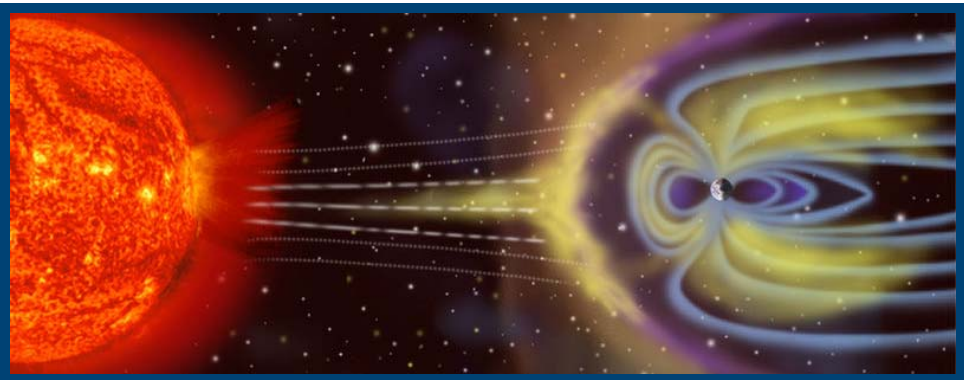
Different Social Media channels have unique characteristics

Twitter: Short duration and highly reactive

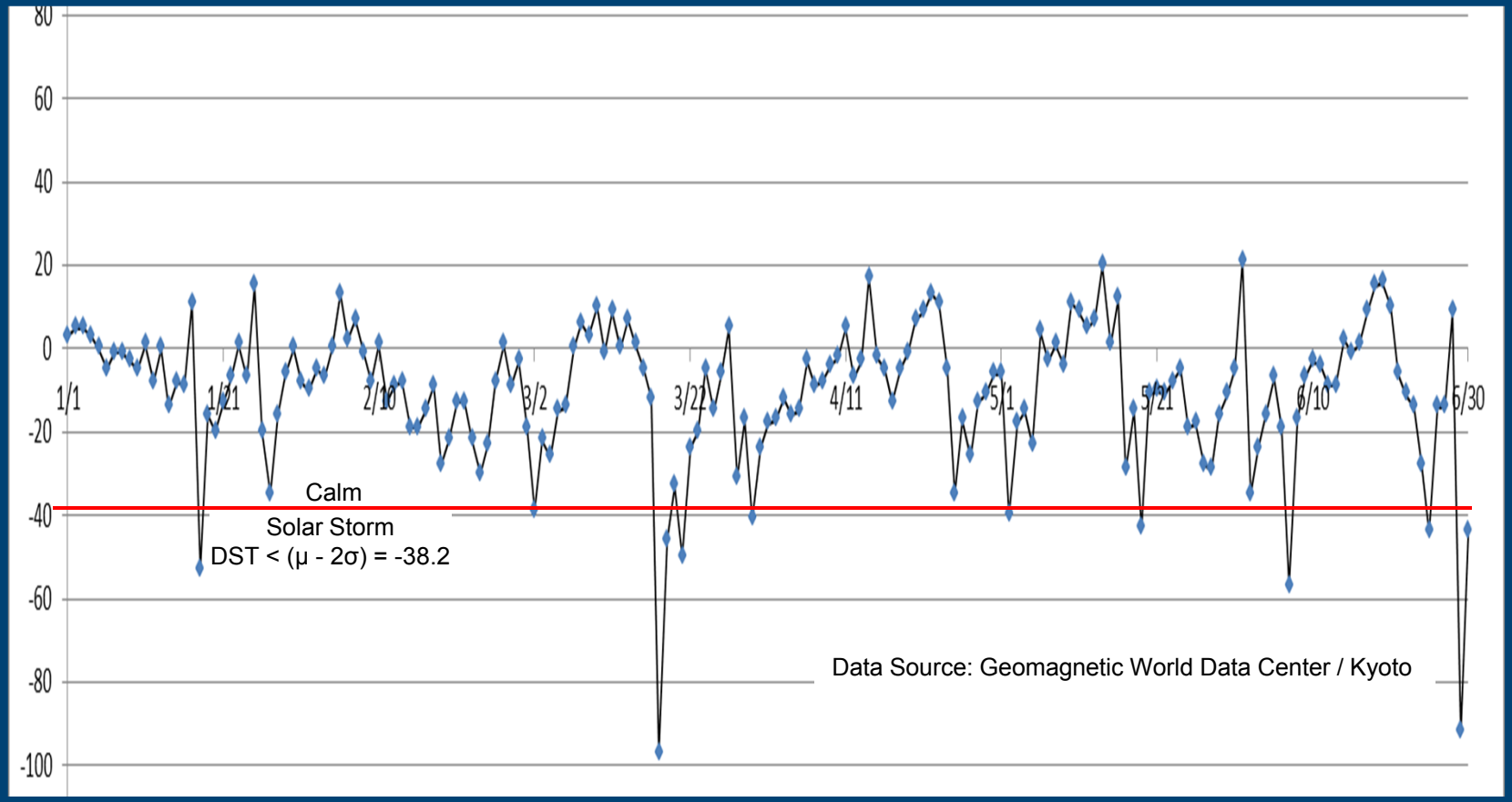
Facebook: medium term, extended conversation threads

Google Trends: May be more predictive. The Social Media equivalent of Durable Goods.

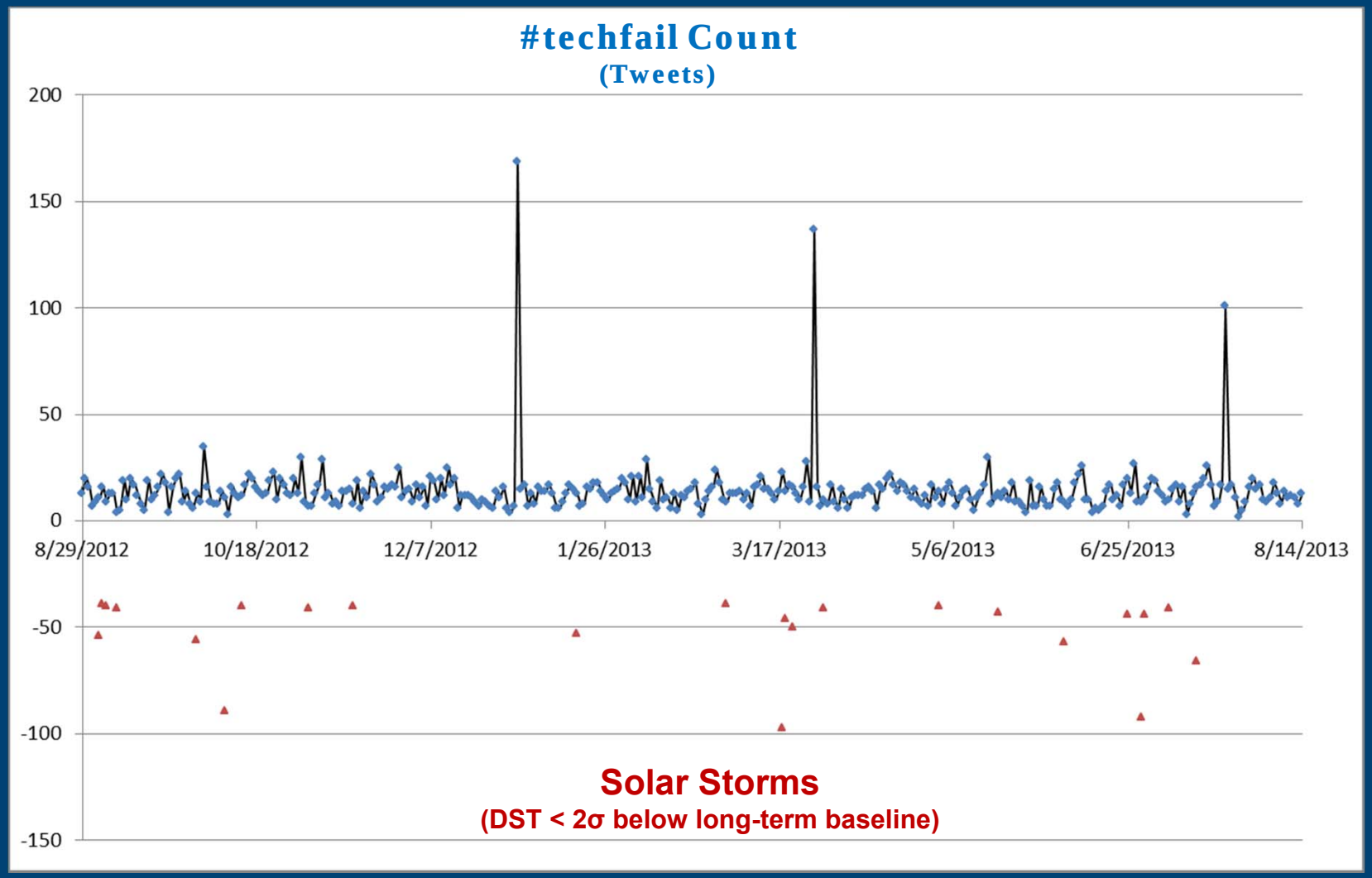
YouTube – the #2 Social Media channel - and #3 Instagram: –no effective means to mine images at this time



Solar Eruptions and Twitter Data



#techfail: Tweets & Astrophysics



#techfail: Tweets & Astrophysics

Time Window: 2 to 4 days following a storm

16.0% of all days follow a solar storm

Top #techfail days: 57.9% follow a solar storm

Burst of #techfail tweets 3.6 times more likely

Statistically significant at > 99% confidence

Mining Twitter Feeds: A Process

1. **Register with Twitter as a developer to get access**
2. **API to deploy search terms and receive tweets**
3. **Parse plain text tweet stream into a SAS dataset**
4. **Exploratory Data Analysis to find best search terms**
5. **High-volume search using results from #4**

1. Register With Twitter

Secure | https://dev.twitter.com/oauth/overview

Developers Products Documentation Community Build My apps Join

Twitter Developer Documentation

Docs / OAuth / Overview

Products & Services

- Best practices
- API overview
- Twitter for Websites
- Twitter Kit
- Cards
- OAuth
  - Overview
    - Using OAuth
    - OAuth FAQ
    - Application Permission Model

Overview

In order to make authorized calls to Twitter's APIs, an application must first obtain an OAuth **access token** on behalf of a Twitter user (or, issue Application-only authenticated requests, when user context is not required). The way to obtain such tokens will depend on the use case.

If you...	Use...
Want to offer a "Sign in with Twitter" button on your website...	Sign in with Twitter
Want to read or post Twitter data on behalf of visitors to your website...	3-legged OAuth
Have a mobile, desktop, or embedded app which can't access a browser...	PIN-based OAuth
Just want to access the API from your own account...	Tokens from dev.twitter.com
Offer an API where clients send you data on behalf of Twitter users...	OAuth Echo

https://dev.twitter.com/oauth/overview

2. Twitter API: Obtain Tokens

1. **consumerKey: Twitter User ID**
2. **consumerSecret: Twitter login password**
3. **Bearer Token: Used in SAS code to access tweets**

2. Twitter API: SAS Code

```
/* Code from Isabel Litton and Rebecca Ottesen, California
Polytechnic State University. Paper: "%GrabTweet: A SAS®
Macro to Read JSON Formatted Tweets", WUSS 2013. Modified
by D. Corliss 2017 to parameterize TYPE */

%MACRO grab_tweet_100(search_term, type, target_amount);

/* Location of Token File */
filename auth "C:\dcorliss\SAS\Twitter\token.txt";

/* Location of Output File */
filename twtOut "C:\dcorliss\SAS\Twitter\Tweets.txt";

/* Set search parameters: COUNT = and TYPE = Popular,
Recent, or Mixed */

%IF &target_amount < 100 %THEN %LET
   num_tweet = %NRSTR(&count=)&target_amount;
%ELSE %LET num_tweet = %NRSTR(&count=100);
%LET type = %NRSTR(&result_type=&type);
```

3. Parse Plain Text Tweet Stream

Raw Twitter Text Output:

Tweets.txt - Notepad

File Edit Format View Help


```
{"statuses":[{"created_at":"Thu Jun 08 08:52:55 +0000 2017","id":872738196777426944,"id_str":"872738196777426944","text":"@lsarsour You hate America &amp
ars","url":null,"entities":{"description":{"urls":[]}},"protected":false,"followers_count":2013,"friends_count":2511,"listed_count":26,"created_at":"Wed
idebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,"has_extended_profile":false,"default_profile":true,"defaul
"recent"},"source":"\u003ca href=\"http:\/\/twitter.com\/#!\/download\/ipad\" rel=\"nofollow\"\u003eTwitter for iPad\u003c\/a\u003e","in_reply_to_status_
ges\/themes\/theme1\/bg.png","profile_background_image_url_https":"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_background_tile":fal
guess what...America is on to you and your sharia law promotion.  Your Alinsky tactics are transparent.","truncated":false,"entities":{"hashtags":[],"sym
one":null,"geo_enabled":true,"verified":false,"statuses_count":1015,"lang":"en","contributors_enabled":false,"is_translator":false,"is_translation_enable
s":null,"translator_type":"none"},"geo":null,"coordinates":null,"place":null,"contributors":null,"is_quote_status":false,"retweet_count":1,"favorite_coun
_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":1216957579,"id_str":"1216957579","name":"Karen D. Scio
g.com\/profile_background_images\/799702831\/608c64c78f543cfda598bc9311553460.png","profile_background_image_url_https":"https:\/\/pbs.twimg.com\/profile
at":"Wed Jun 07 14:39:45 +0000 2017","id":872463091677540353,"id_str":"872463091677540353","text":"Agree our schools are run by commies &amp; traitors wh
n":"Libertarian Vice Presidential nominee, Best Selling Author, Youtube: https:\/\/t.co\/Cxln91HRMg WAR Now Radio: https:\/\/t.co\/WmcCUVjLrr","url":"htt
r":false,"is_translation_enabled":false,"profile_background_color":"222222","profile_background_image_url":"http:\/\/pbs.twimg.com\/profile_background_im
"place":null,"contributors":null,"is_quote_status":false,"retweet_count":36,"favorite_count":41,"favorited":false,"retweeted":false,"possibly_sensitive":
_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":3290719405,"id_str":"3290719405","name":"Maine4Trumpism","screen_name":"Maine4Tru
s\/theme1\/bg.png","profile_background_image_url_https":"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_background_tile":false,"profil
k white men are most dangerous demographic, subscribe to primitive totalitaria\u2026 https:\/\/t.co\/r1CHYXsHsb","truncated":true,"entities":{"hashtags":
http:\/\/larrylockman.com","display_url":"larrylockman.com","indices":[0,23]}]},"description":{"urls":[]}},"protected":false,"followers_count":199,"frien
ink_color":"DD2E44","profile_sidebar_border_color":"000000","profile_sidebar_fill_color":"000000","profile_text_color":"000000","profile_use_background_i
2","indices":[95,107]},{"screen_name":"mainegop","name":"Maine GOP","id":22251316,"id_str":"22251316","indices":[120,129]},{"screen_name":"gehrig38","nam
te of Maine \ud83c\udf32 (DM US WITH TIPS) #mepolitics #mepol Recently Featured on @TuckerCarlson & @infowars","url":"https:\/\/t.co\/iuS68vcV1X","entiti
/7fqZZBA1_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/872561048095784961\/7fqZZBA1_normal.jpg","profile_banner_url":"
nt":0,"favorited":false,"retweeted":false,"lang":"en"},{"created_at":"Thu Jun 08 02:42:52 +0000 2017","id":872645070616489984,"id_str":"87264507061648998
coggin County GOP. Put Maine Citizens First! #mepolitics Award-Winner! https:\/\/t.co\/CItRSg8f8N","url":"https:\/\/t.co\/PjLJESlnmp","entities":{"url":{
file_background_tile":false,"profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/779527995501506560\/tqs6TWcE_normal.jpg","profile_image_url_http
,"truncated":true,"entities":{"hashtags":[],"symbols":[],"user_mentions":[],"urls":[{"url":"https:\/\/t.co\/r1CHYXsHsb","expanded_url":"https:\/\/twitter
ected":false,"followers_count":199,"friends_count":95,"listed_count":9,"created_at":"Wed Feb 19 03:09:53 +0000 2014","favourites_count":51,"utc_offset":-
color":"000000","profile_use_background_image":false,"has_extended_profile":false,"default_profile":false,"default_profile_image":false,"following":null,
[120,129]},{"screen_name":"gehrig38","name":"Curt Schilling","id":14945261,"id_str":"14945261","indices":[130,139]}],"urls":[{"url":"https:\/\/t.co\/NvM4
url":"https:\/\/t.co\/iuS68vcV1X","entities":{"url":{"urls":[{"url":"https:\/\/t.co\/iuS68vcV1X","expanded_url":"http:\/\/mainefirstmedia.wordpress.com",
qZZBA1_normal.jpg","profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/816402981851660288\/1496869729","profile_link_color":"ABB8C2","profile
9091027914753,"id_str":"872639091027914753","text":"RT @WayneRoot: Agree our schools are run by commies &amp; traitors who hate America &amp; Christianit
```

3. Parse Plain Text Tweet Stream

```
/* Code from Isabel Litton and Rebecca Ottesen, California
Polytechnic State University. Paper: "%GrabTweet: A SAS®
Macro to Read JSON Formatted Tweets", WUSS 2013 */

   INPUT
      @'"created_at":' created_at1
      @'"id":' tweet_id
      @'"text":' text
      @'"user":{"id":' user_id
      @'"name":' name
      @'"screen_name":' screen_name
      @'"location":' location
      @'"created_at":' created_at2
      @'"lang":' lang
      @'"contributors":null,' retweeted @@
      ;
```

4. Exploring Search Terms

```
**** Mining Twitter for Hate Speech ****;

%grab_tweet_100(Sharia America hate, recent, 100);
```

	created_at1	tweet_id	text	user_id	name	screen_name	location
1	Thu Jun 08 08:52:55 +0000 2017	872738196777426944	@lsarsour You hate America & support Sharia here. Your time will come.	1223623118	Michael Erwin	michaeloferwins	California, USA
2	Thu Jun 08 03:52:00 +0000 2017	872662466714624001	RT @LMARocks: @lsarsour Not a hate crime, and guess what...America is on to you and your sharia law promotion. Your Alinsky tactics are tr\	4223111532	PorkChop Express	China_is_here	Wing Kong Exchange
3	Mon Feb 23 03:30:48 +0000 2009	872660605043449856	RT @WayneRoot: Agree our schools are run by commies & traitors who hate America & Christianity? What exactly do you call this?\nhttp	1216957579	Karen D. Scioscia	KDScioscia	
4	Wed May 06 21:43:22 +0000 2009	872645595068067840	RT @lockman4mehouse: Many \"new Mainers\" hate America	3290719405	Maine4Trumpism	Maine4Trump	#ME02 Moscow, Maine
5	Wed Feb 19 03:09:53 +0000 2014	872598556640780291	#NewRelease \n\nMaine's Refugee Community: The Best And The Brightest \n\nhttps:\/\/t.co\/NvM4koRXXR\n\n #mepolitics @mainegop @	816402981851660	Maine First Media	MaineFirstMedia	Waterville, ME
6	Thu Jun 08 02:42:52 +0000 2017	872645070616489984	RT @lockman4mehouse: Many \"new Mainers\" hate America	779527044287848	Androscoggin GOP	AndroGOP	Maine, USA
7	Wed Feb 19 03:09:53 +0000 2014	872598556640780291	#NewRelease \n\nMaine's Refugee Community: The Best And The Brightest \n\nhttps:\/\/t.co\/NvM4koRXXR\n\n #mepolitics @mainegop @	816402981851660	Maine First Media	MaineFirstMedia	Waterville, ME
8	Thu Jun 08 02:19:06 +0000 2017	872639091027914753	RT @WayneRoot: Agree our schools are run by commies & traitors who hate America & Christianity? What exactly do you call this?\nhttp	2937854619	Sammie Snickers	Sammie_Snickers	
9	Wed May 06 21:43:22 +0000 2009	872610362209837056	RT @lockman4mehouse: Many \"new Mainers\" hate America	816402981851660	Maine First Media	MaineFirstMedia	Waterville, ME
10	Wed Feb 19 03:09:53 +0000 2014	872598556640780291	#NewRelease \n\nMaine's Refugee Community: The Best And The Brightest \n\nhttps:\/\/t.co\/NvM4koRXXR\n\n #mepolitics @mainegop @	816402981851660	Maine First Media	MaineFirstMedia	Waterville, ME
11	Thu Jun 08 00:24:21 +0000 2017	872610213991501828	Many \"new Mainers\" hate America	2351033034	Larry Lockman	lockman4mehouse	
12	Wed Jun 07 23:38:02 +0000 2017	872598556640780291	#NewRelease \n\nMaine's Refugee Community: The Best And The Brightest \n\nhttps:\/\/t.co\/NvM4koRXXR\n\n #mepolitics @mainegop @	816402981851660	Maine First Media	MaineFirstMedia	Waterville, ME

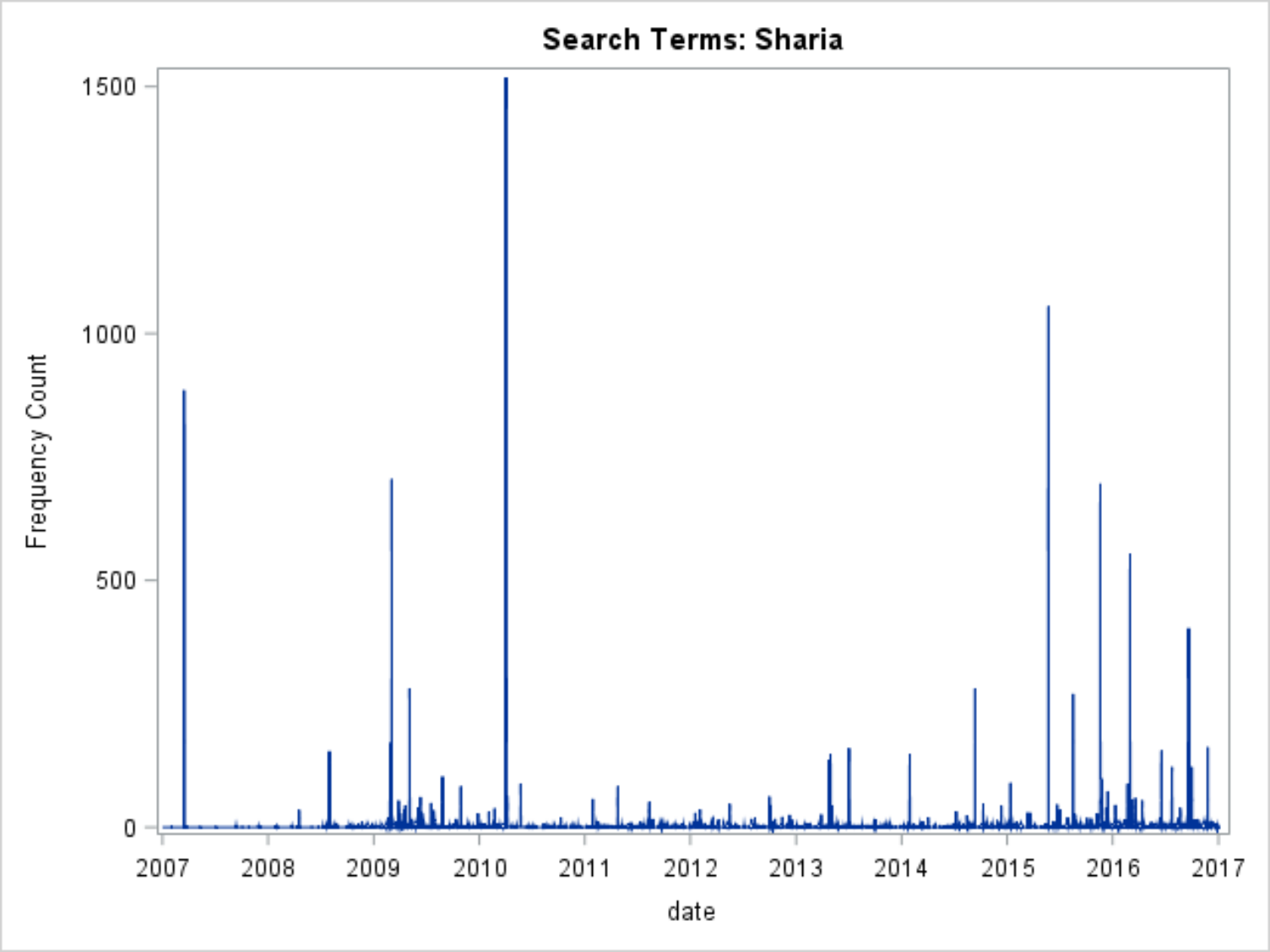
5. High-Volume Search

1. **Litton-Ottesen macro to access Twitter-imposed limit of 100 tweets at a time**

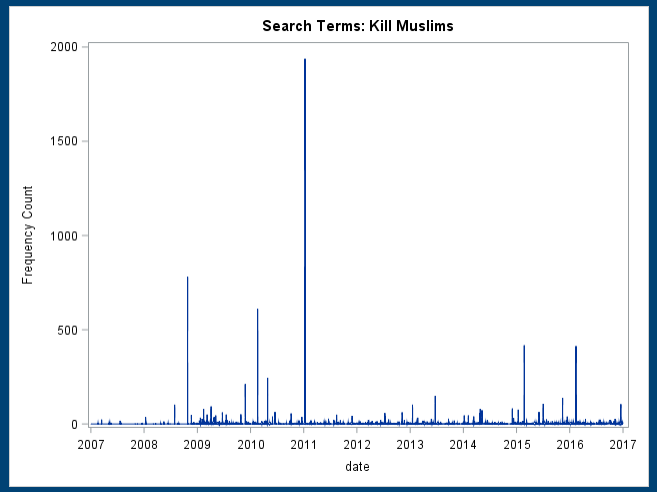
2. **Macro wrapper to repeat until specified count, going backwards from end of previous iteration**

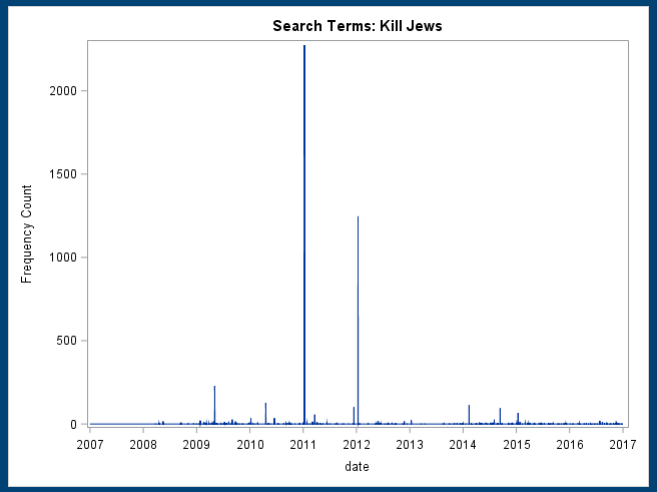
3. **Data transformation required for time series dataset**

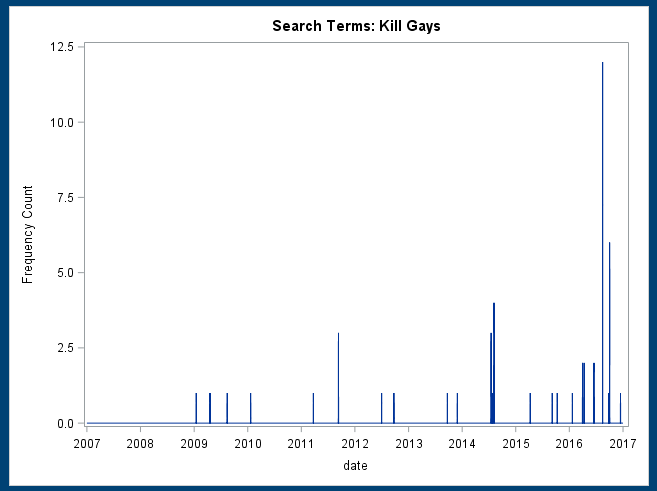
Analytic Results

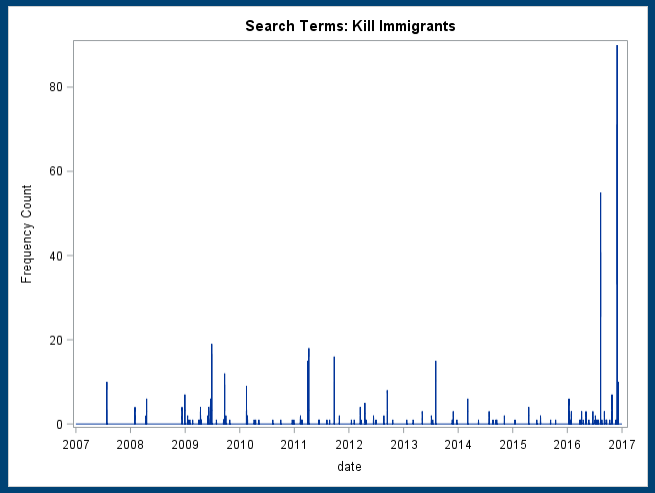


Analytic Results

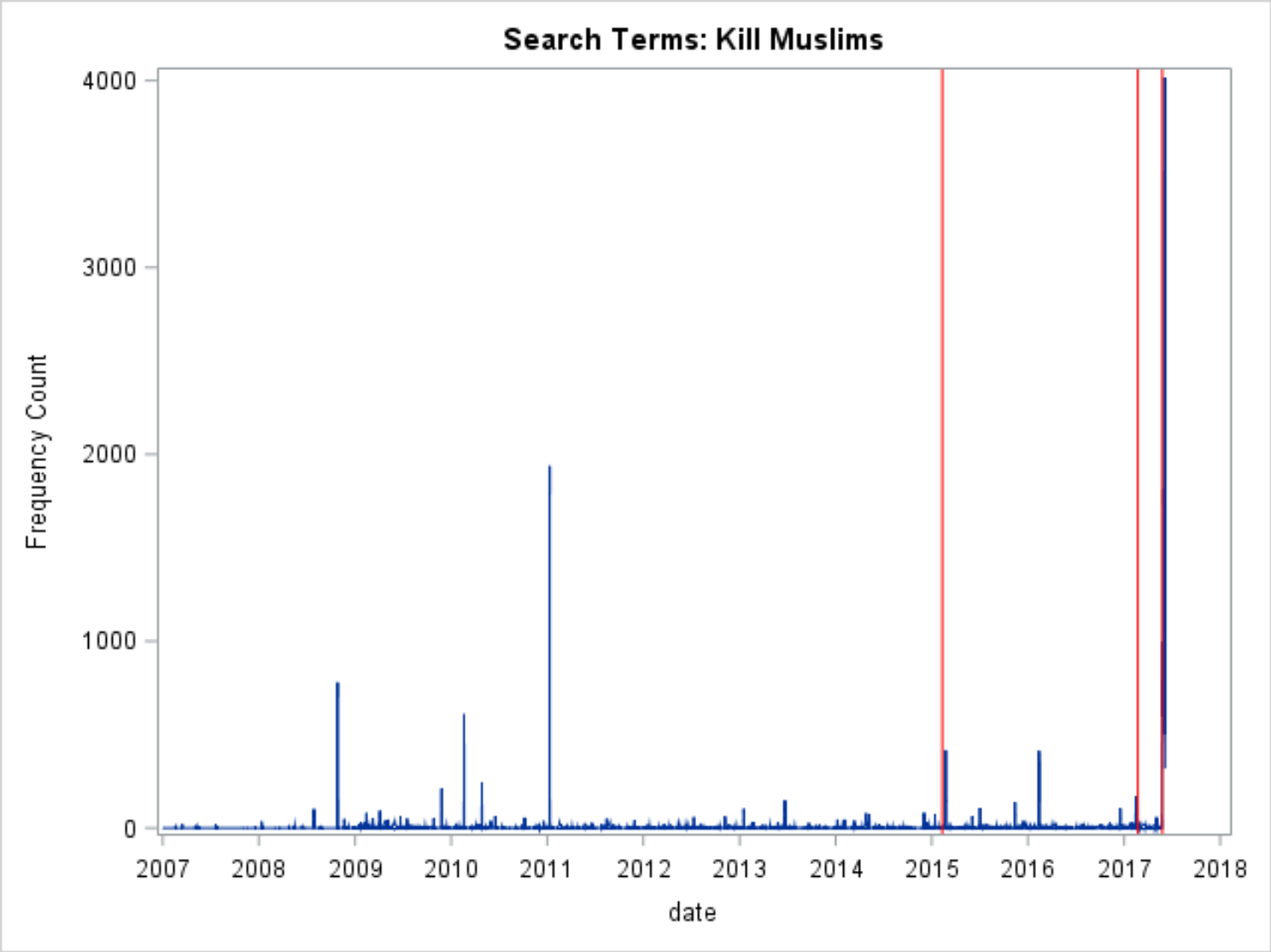




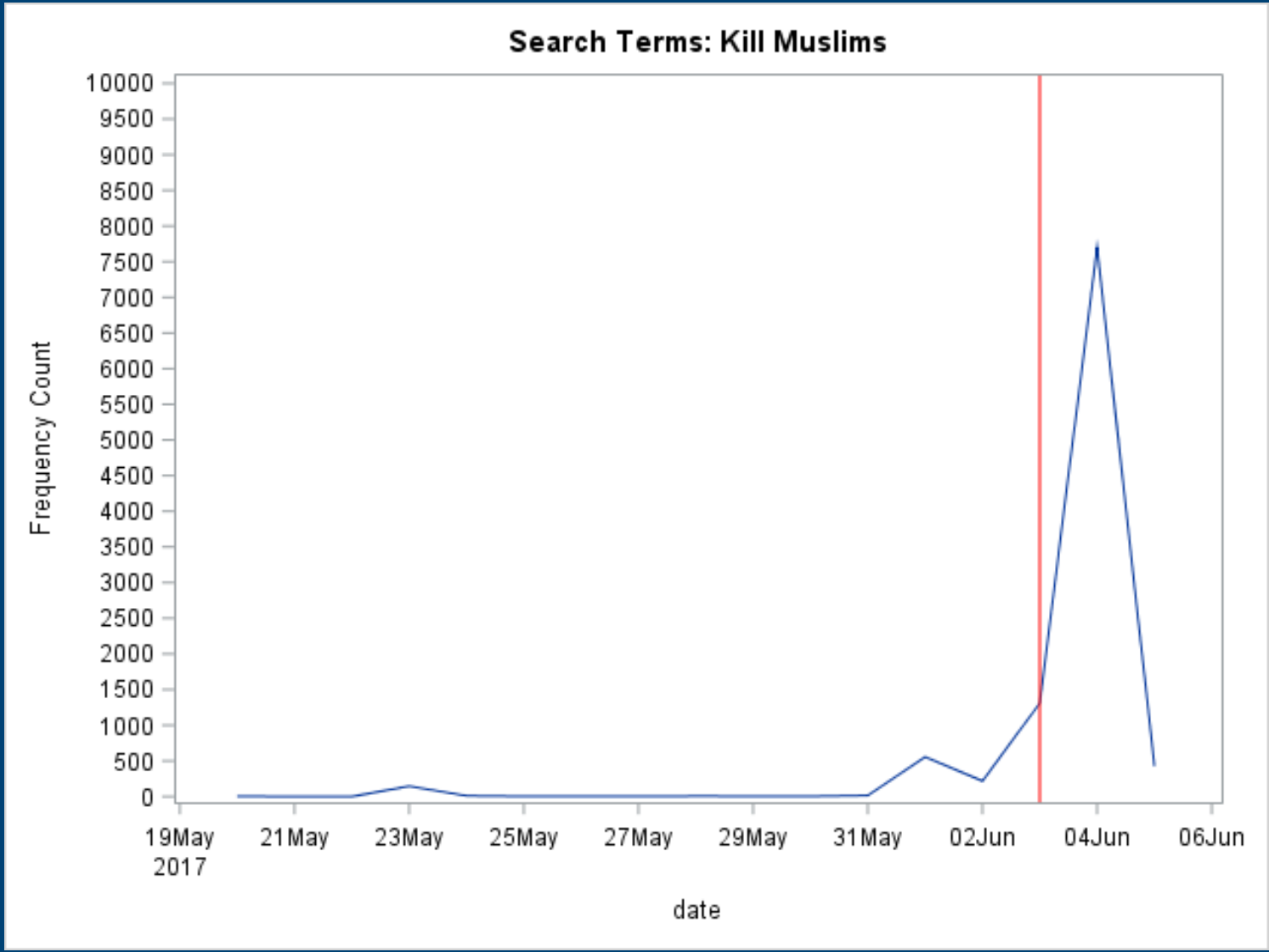




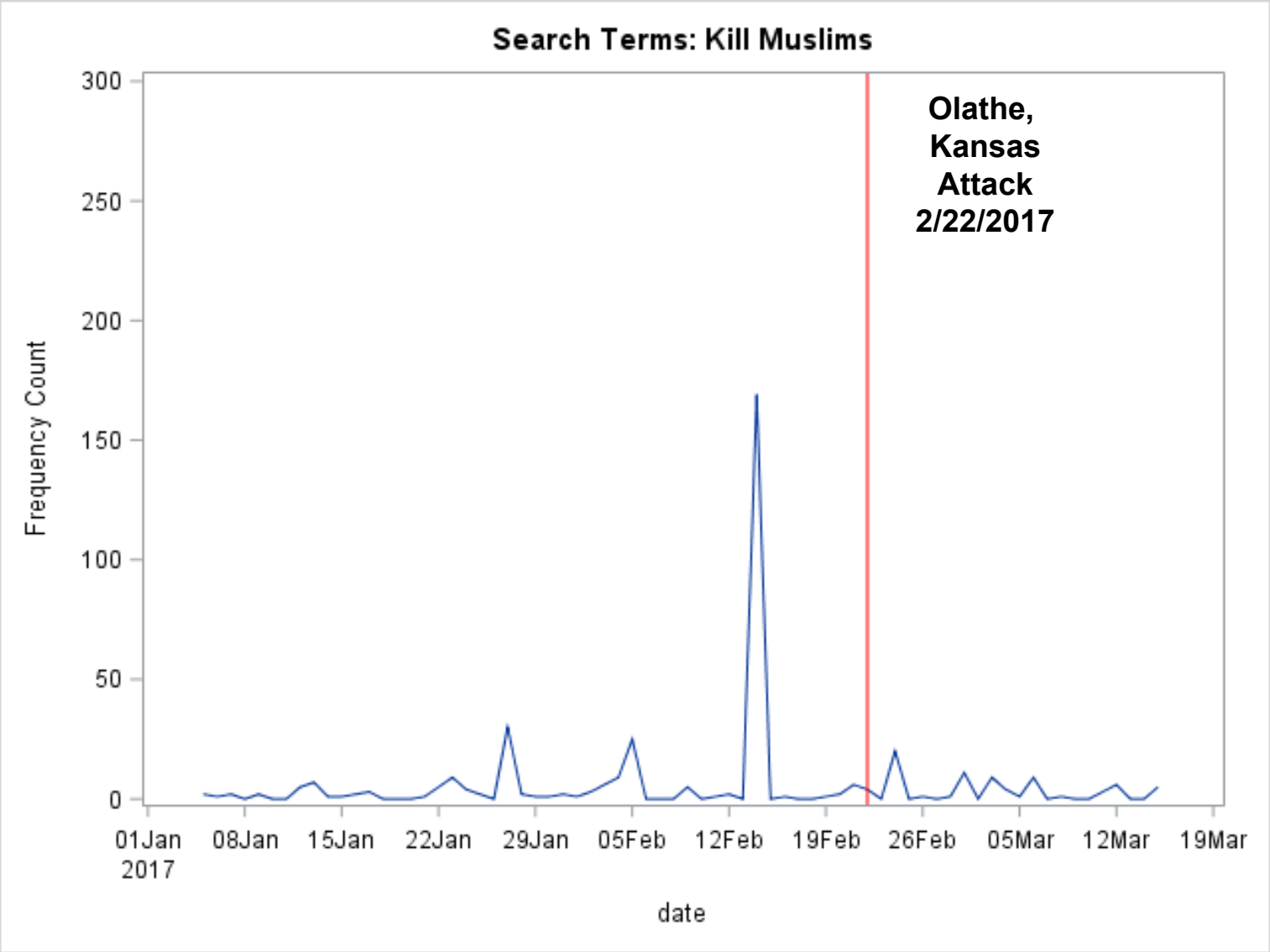
Analytic Results



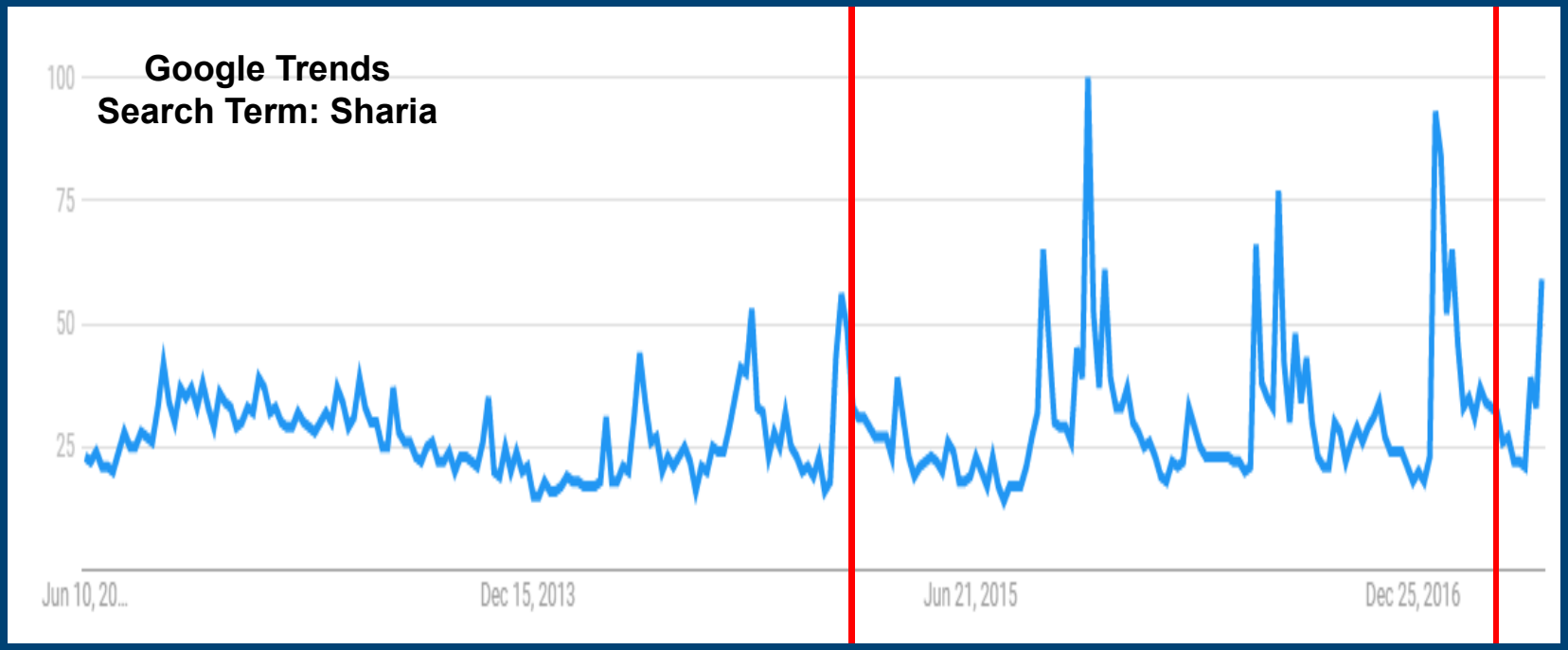
Analytic Results



Analytic Results



Google Trends: A Leading Indicator?



Summary

Social media data can be matched to event data for Time Series Analysis, including lead / lag modeling

Different social channels have different characteristics, including lead or lag times

Analytics methods optimized for big data methods may be necessary

Twitter data can be mined and analyzed at high volume

Analysis of Hate Crimes with social media data shows promise but better crime data is needed.

Questions

David J. Corliss

davidjcorliss@peace-work.org

www.peace-work.org