

data·visualization

at Twitter

data at Twitter

“Tweets”

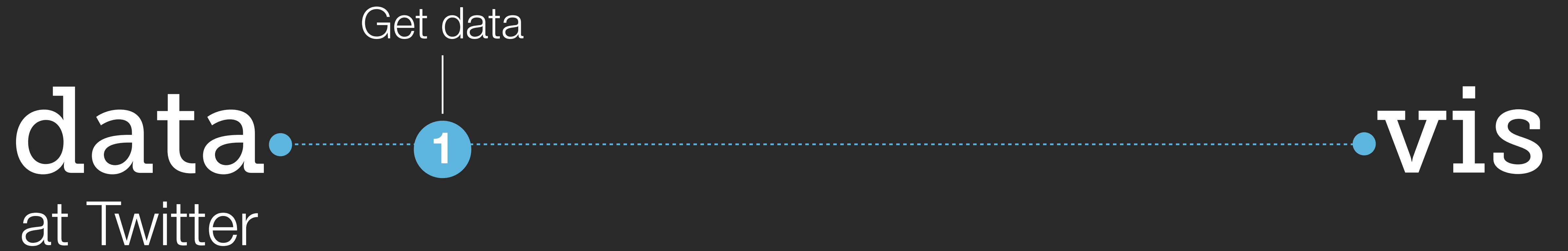
#events

TV Shows New Year
Earthquake Protest
Oscars Super Bowl
World Cup Election
Breaking news ...

#curiosity

Sleep pattern
Human behavior
Language ...

What could we learn from the Tweets?



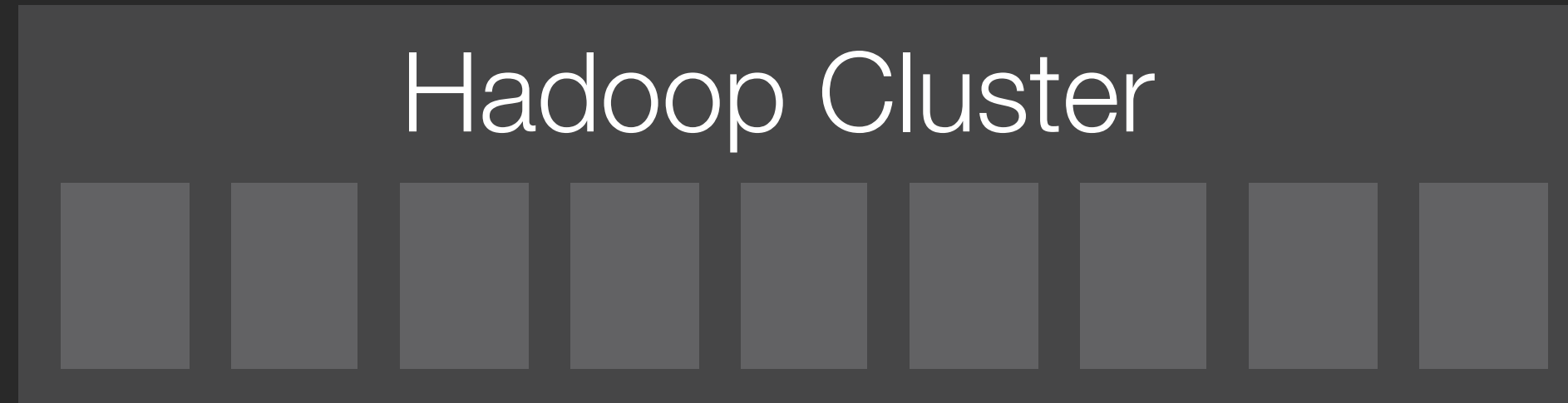
“Tweets”

Challenges

- Too much data
 - Want only relevant Tweets
 - hashtag: #BRA
 - keywords: “goal”
 - Need to aggregate & reduce size
 - Long processing time (hours)

Workflow

Data Storage



Tool

Pig / Scalding (slow)



SQL

Workflow

Data Storage

Hadoop Cluster

Vertica

Tool

Pig / Scalding (slow)

SQL

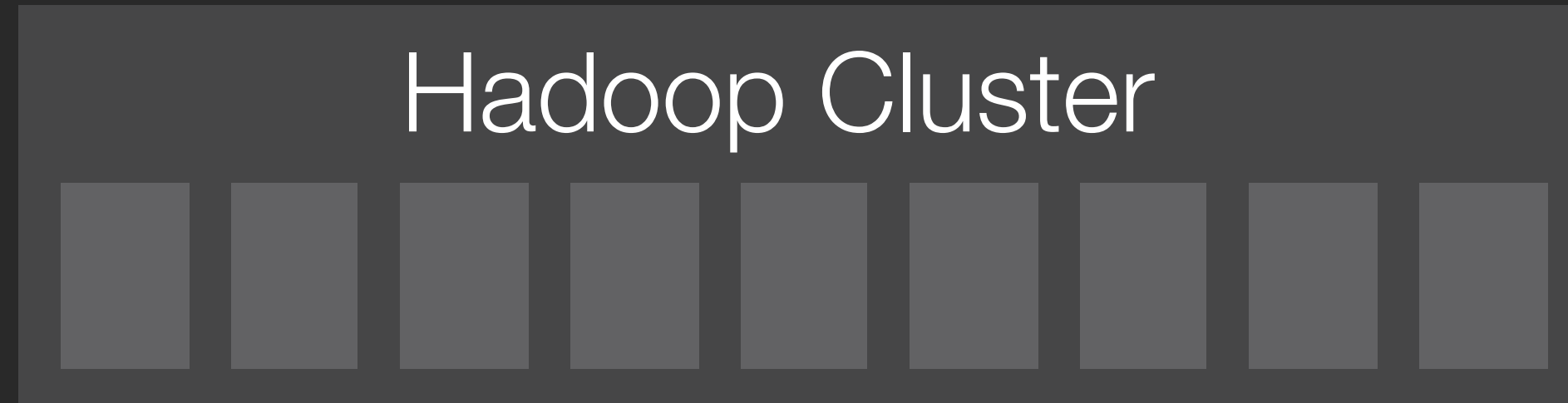
```
-- PIG

tweets = load '/tables/statuses/$DATE' using TwadoopLoader();

filtered_tweets = filter tweets by twtext matches '.*#BRA.*';
filtered_tweets = foreach filtered_tweets
  generate '#BRA' as name:chararray, extract_minute(created_at) as minute;
counted_hashflag = foreach (group filtered_tweets by minute)
  generate FLATTEN(group) as (name), COUNT(filtered_tweets) as count;
```

Workflow

Data Storage

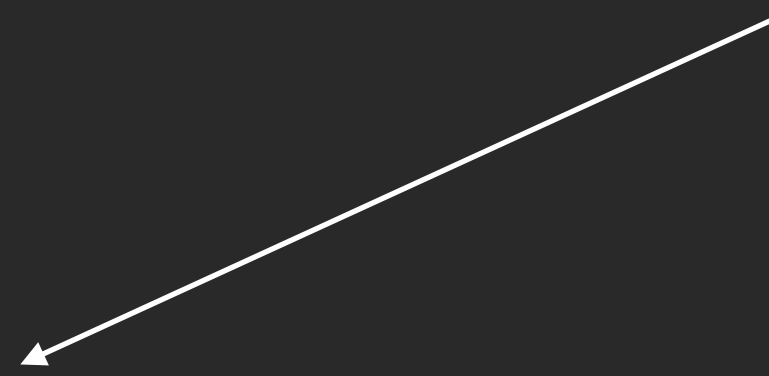


Tool

Pig / Scalding (slow)

SQL

Your laptop



Workflow

Data Storage

Hadoop Cluster

Vertica

Tool

Pig / Scalding (slow)

SQL

Your laptop

Smaller dataset

Tool

node.js / python / excel (fast)

Final dataset

data.
at Twitter

“Tweets”

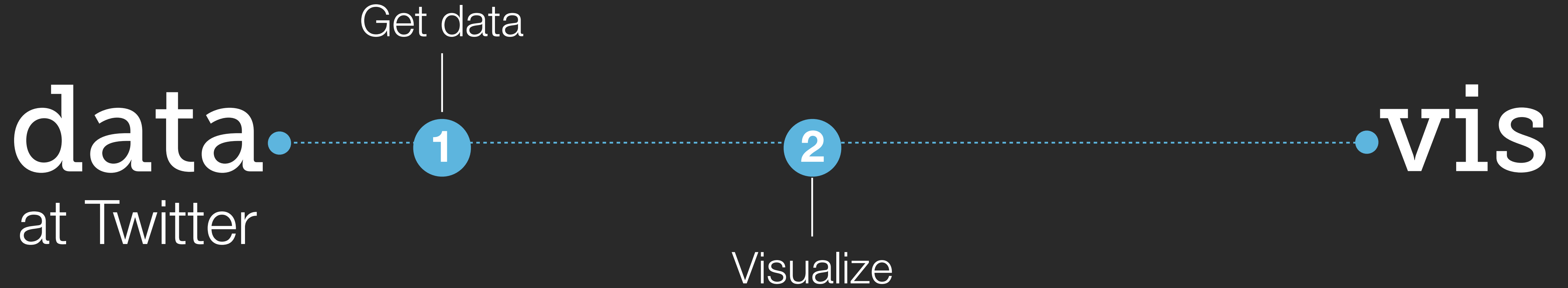
Get data

1

2

Visualize

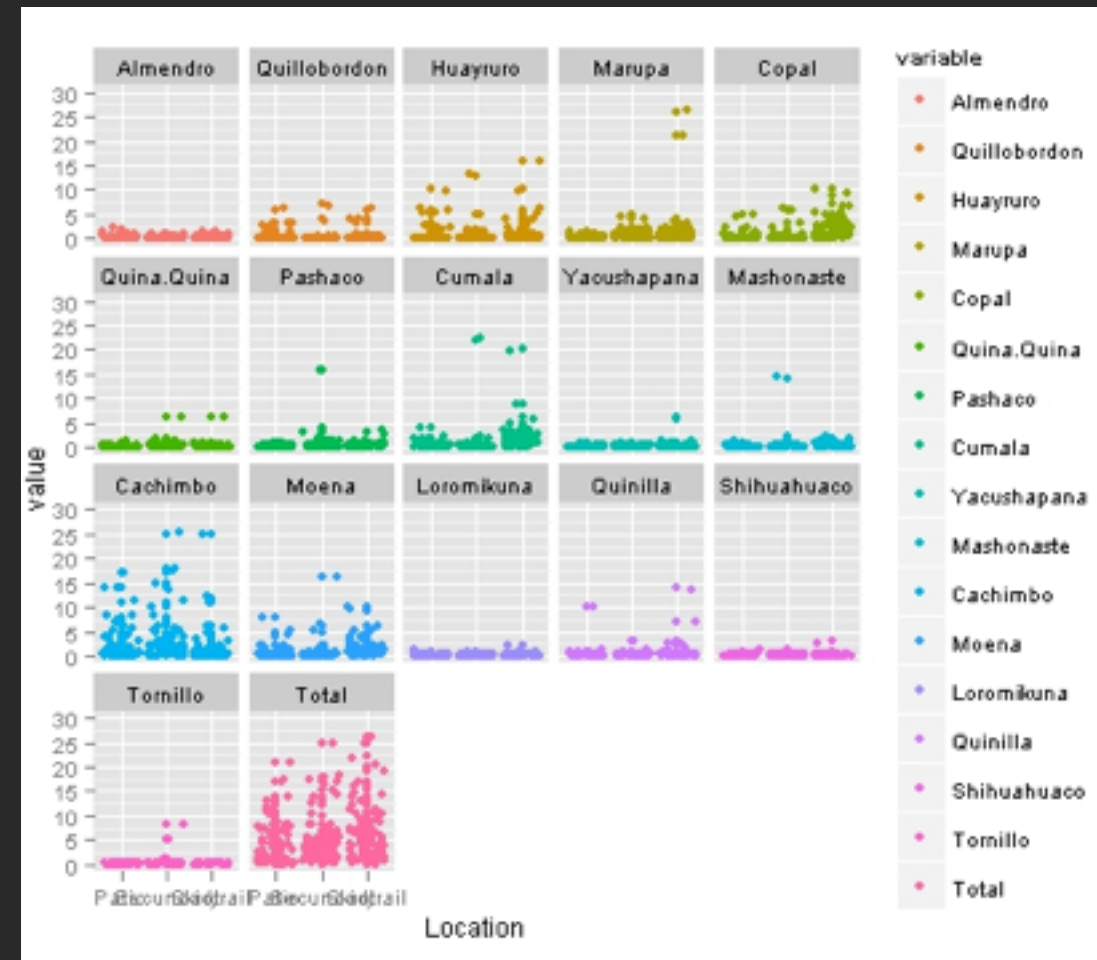
vis



Visualize

- Peek into data
- Check data & test ideas
- Decide how to visualize
 - Guided by data type
- Choose tools
- Start building

R



d3

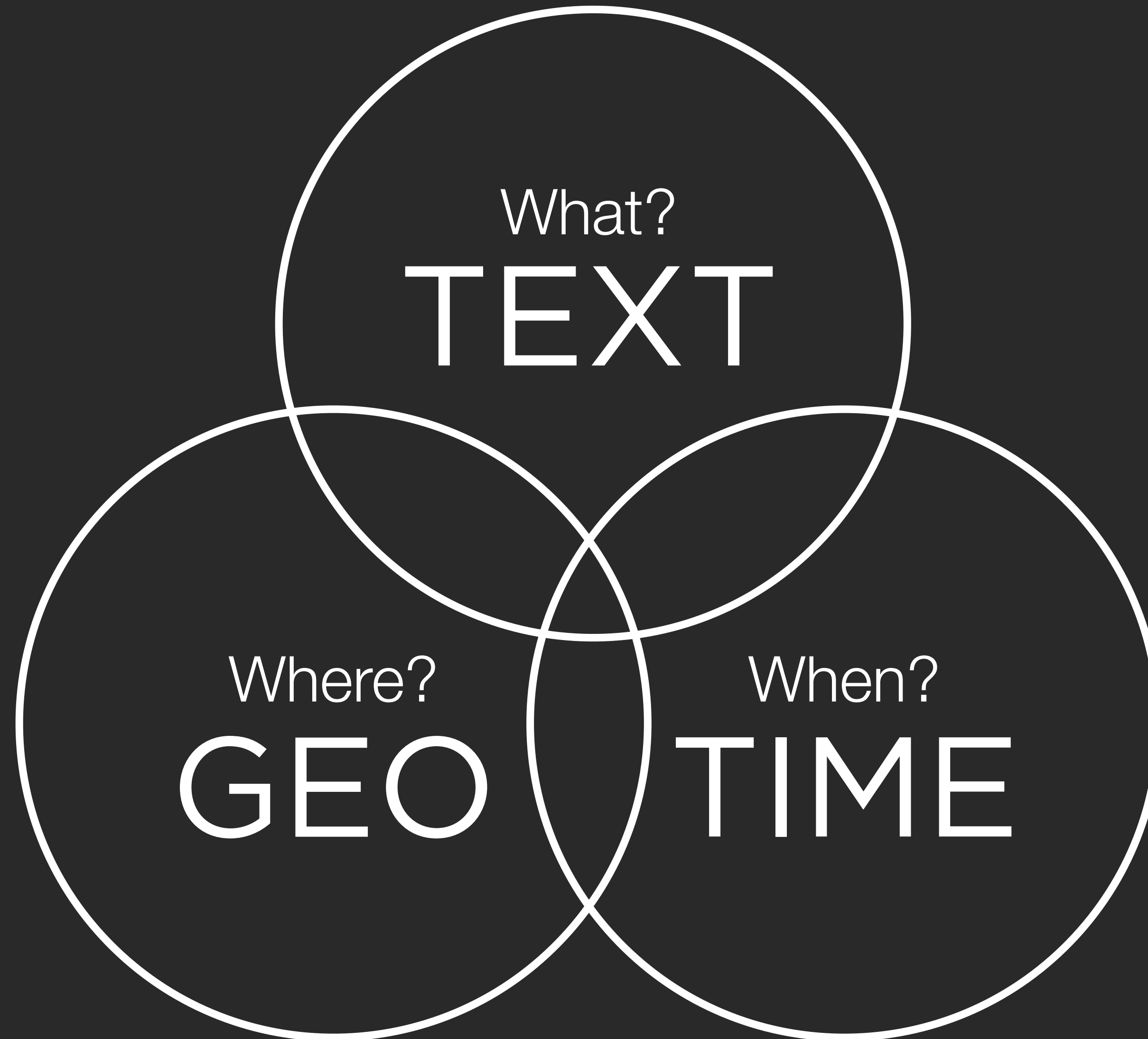


Tableau



Yeoman

Visualize Data



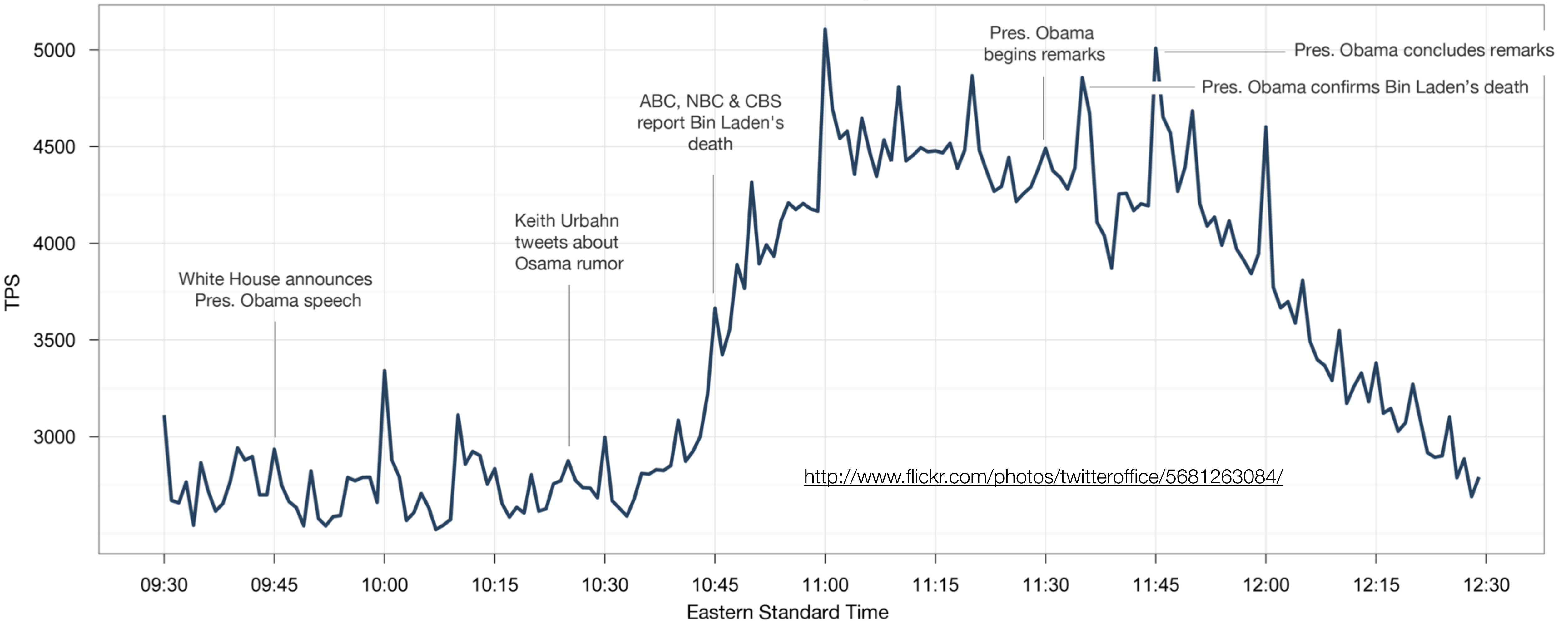
Visualize Data



Time

Tweets/second + Annotation

Tweets Per Second: Evening of May 1, 2011



Visualize Data



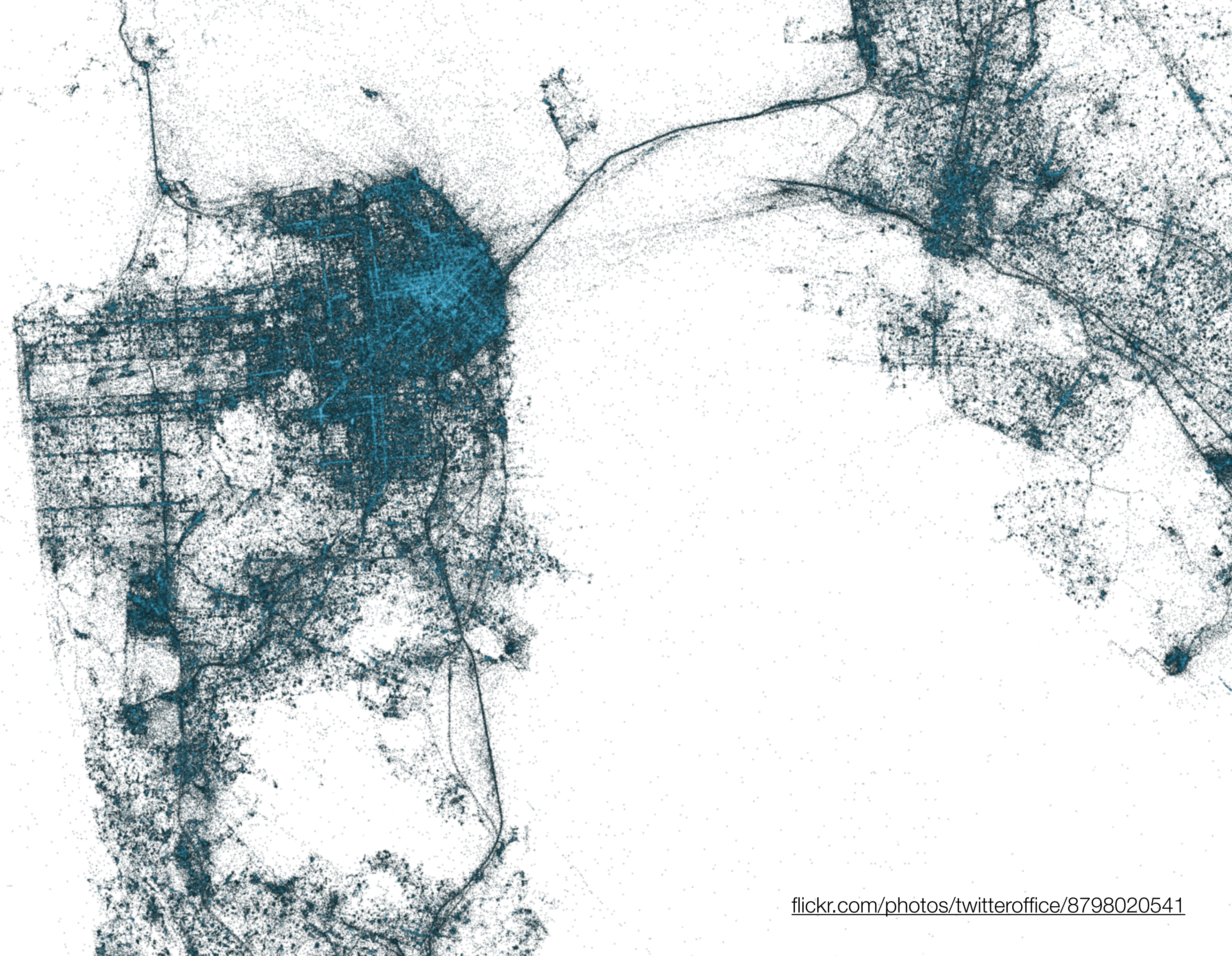
Geo

San Francisco

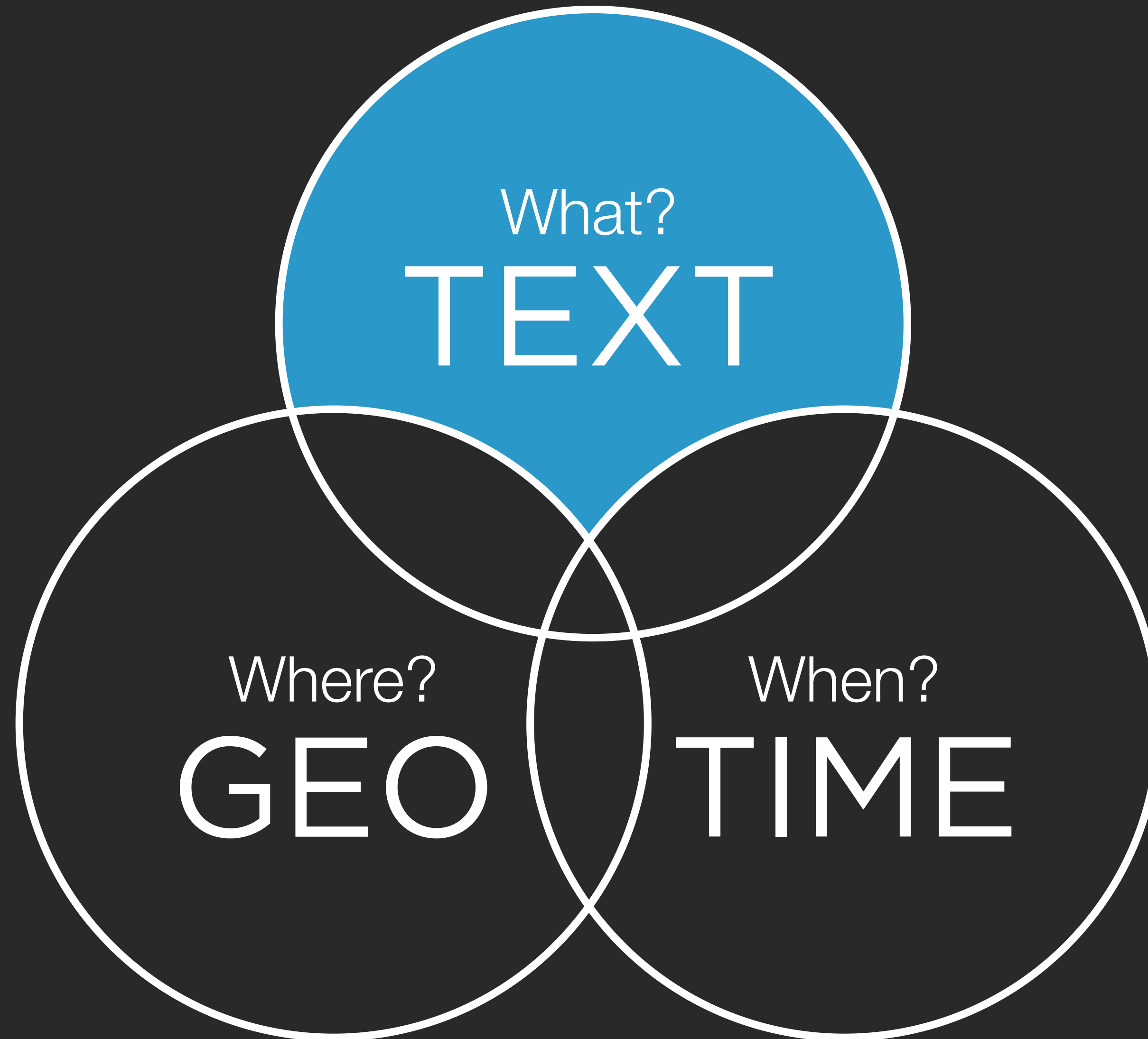
Low density



High density



Visualize Data

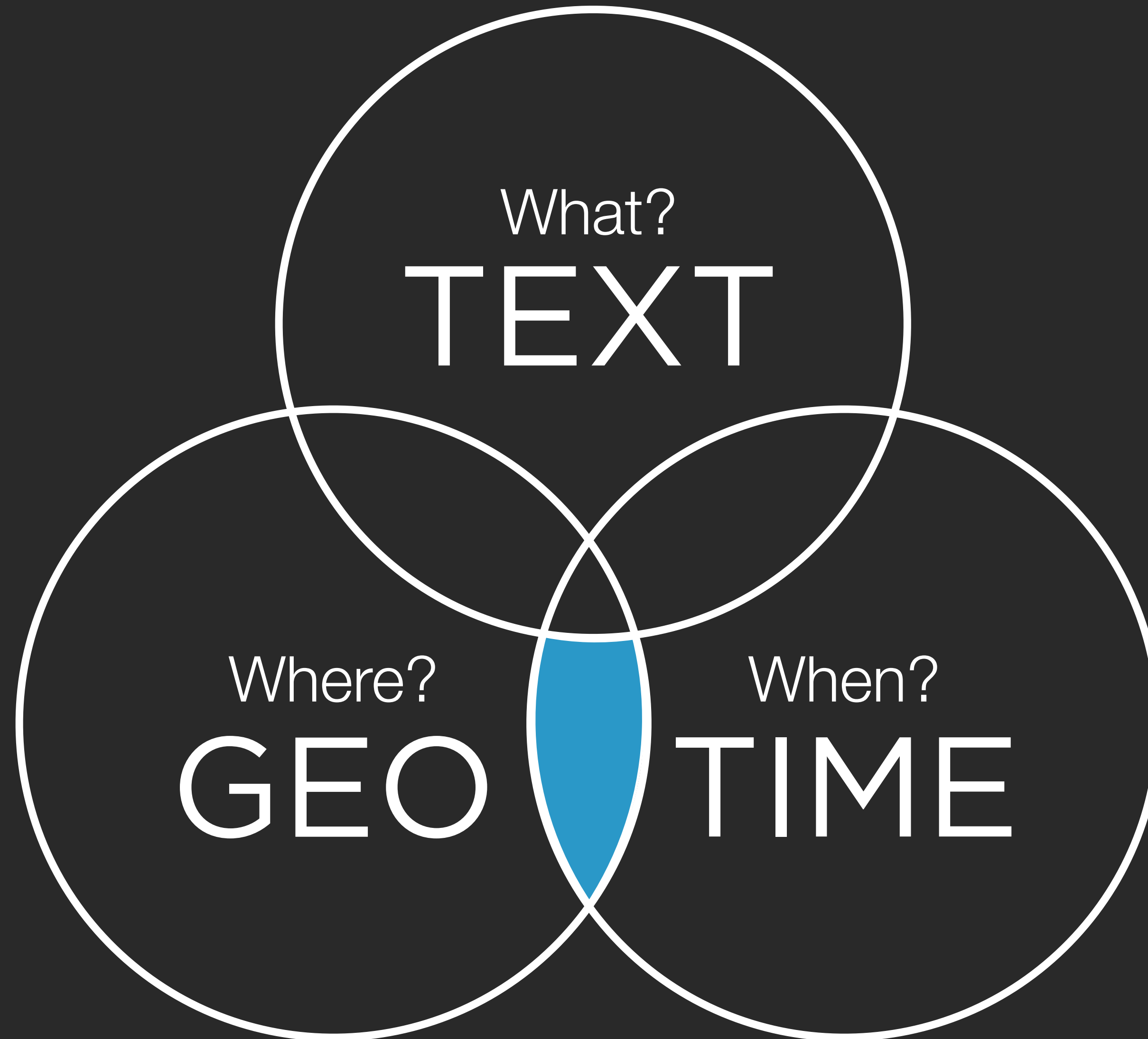


Text

WordTree [Wattenberg & Viégas 2008]

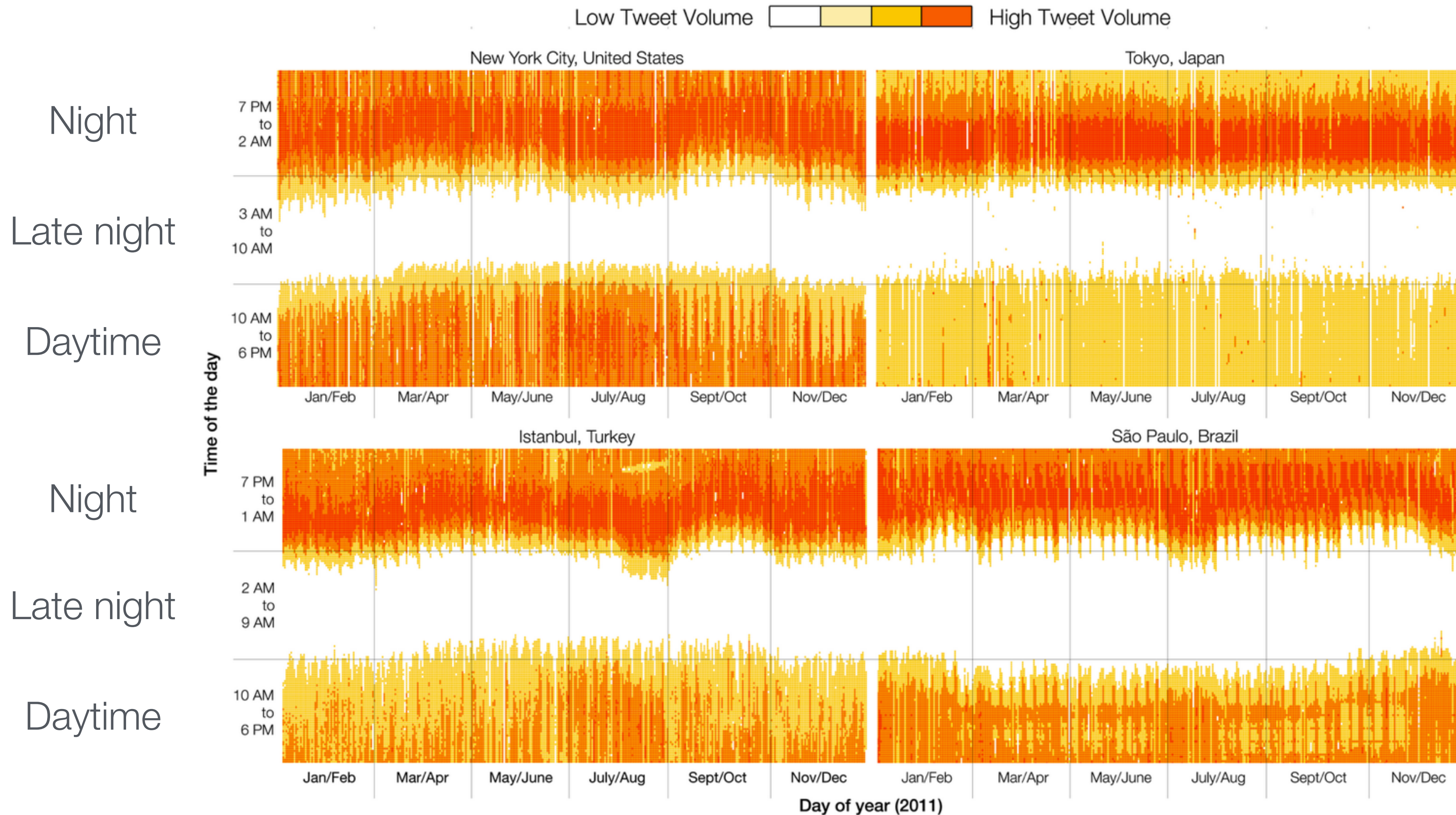


Visualize Data

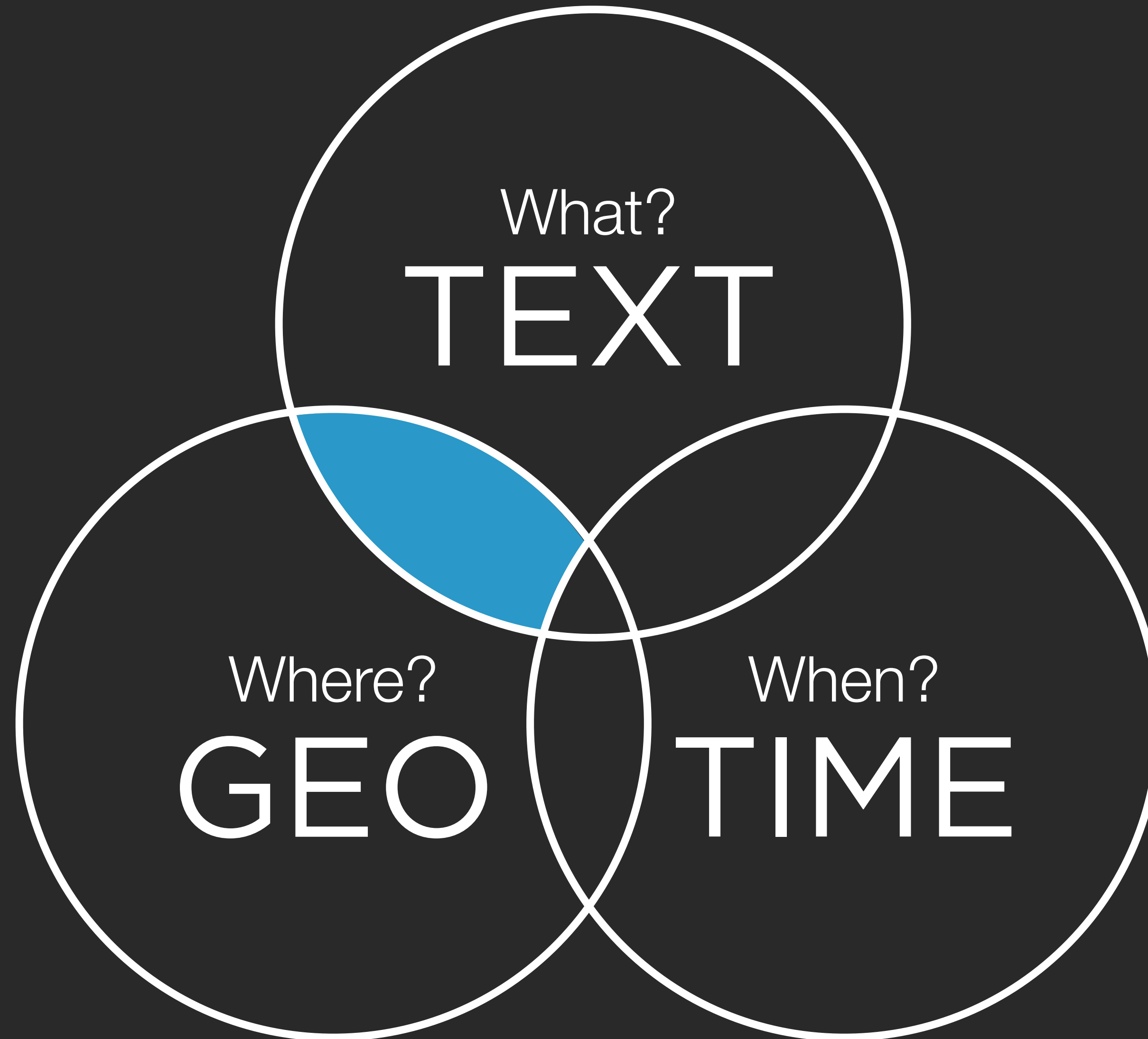


Time + Geo

Tweet pattern [Rios & Lin 2012]



Visualize Data



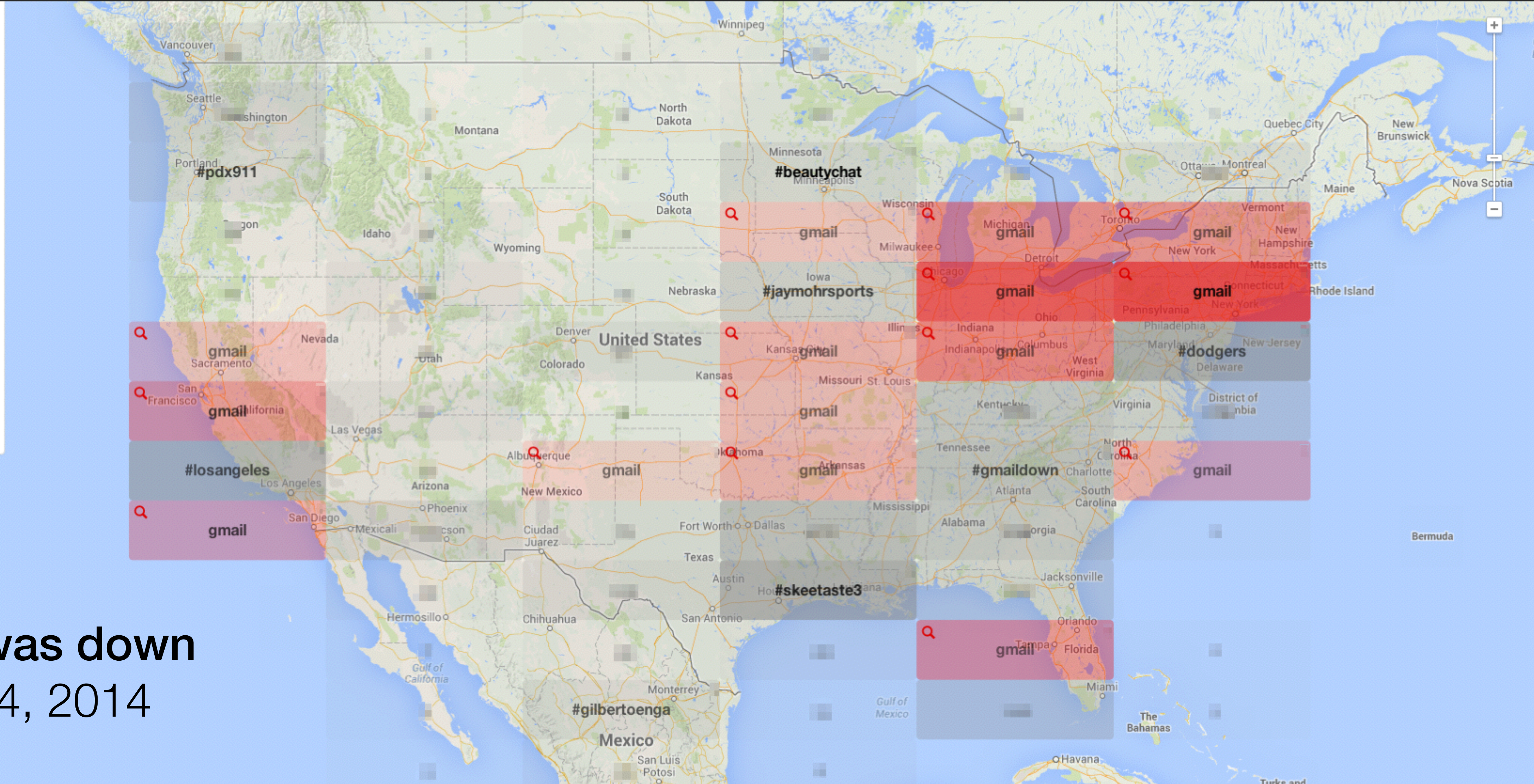
Geo + Text

Real-time Tweet map

Search

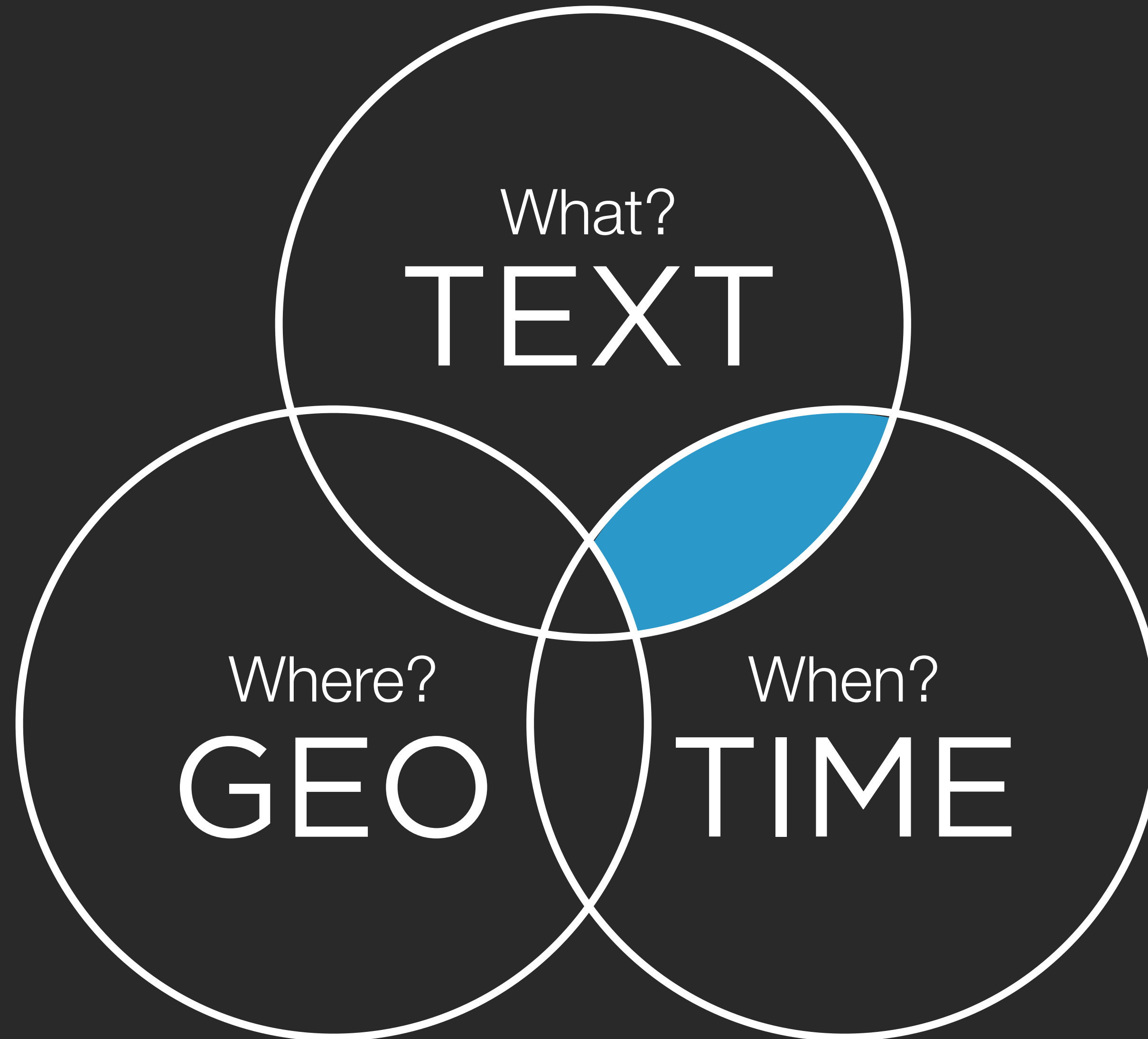
gmail clear

@trebor - visual-insights@twitter.com - help



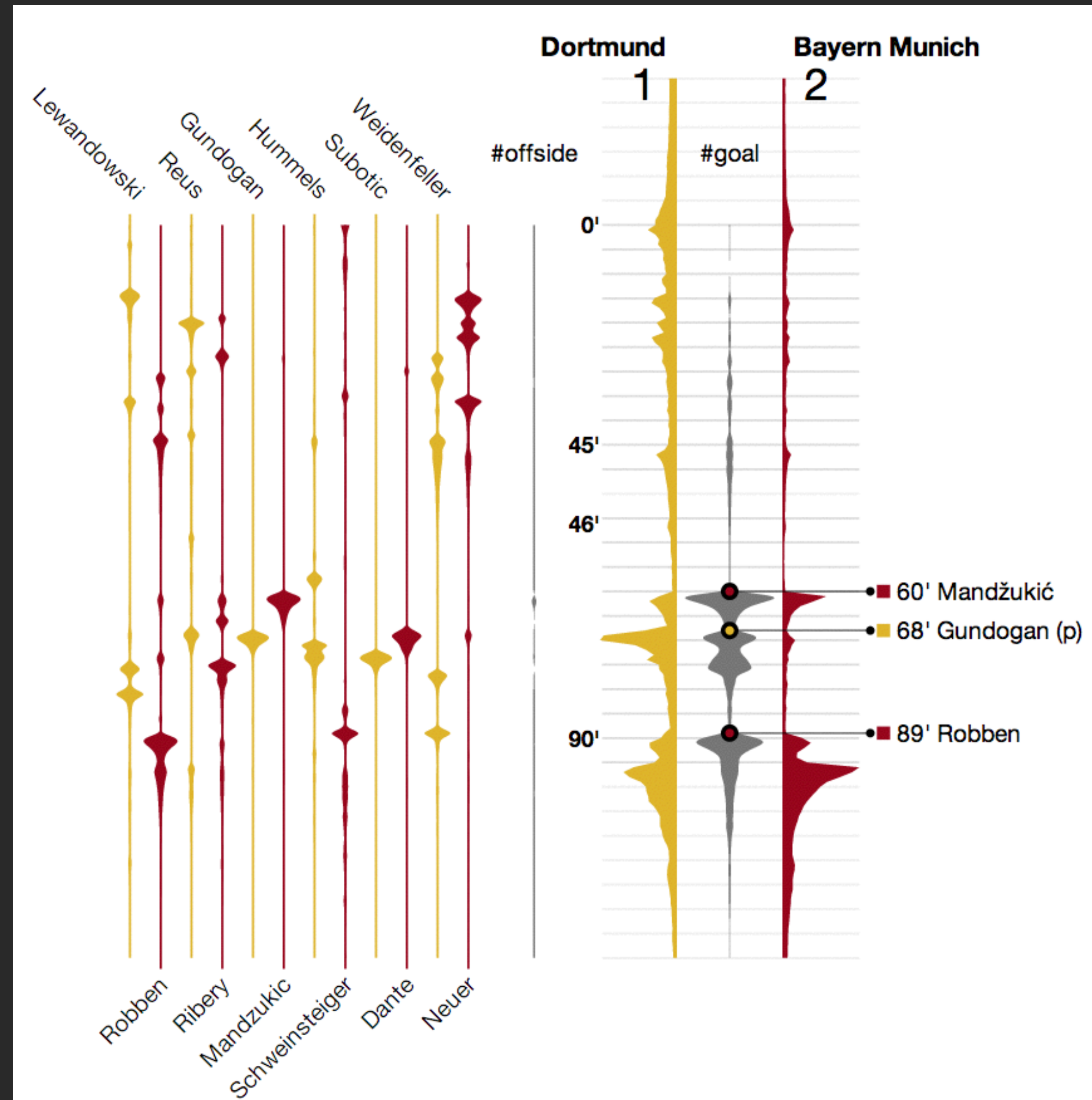
Gmail was down
Jan 24, 2014

Visualize Data



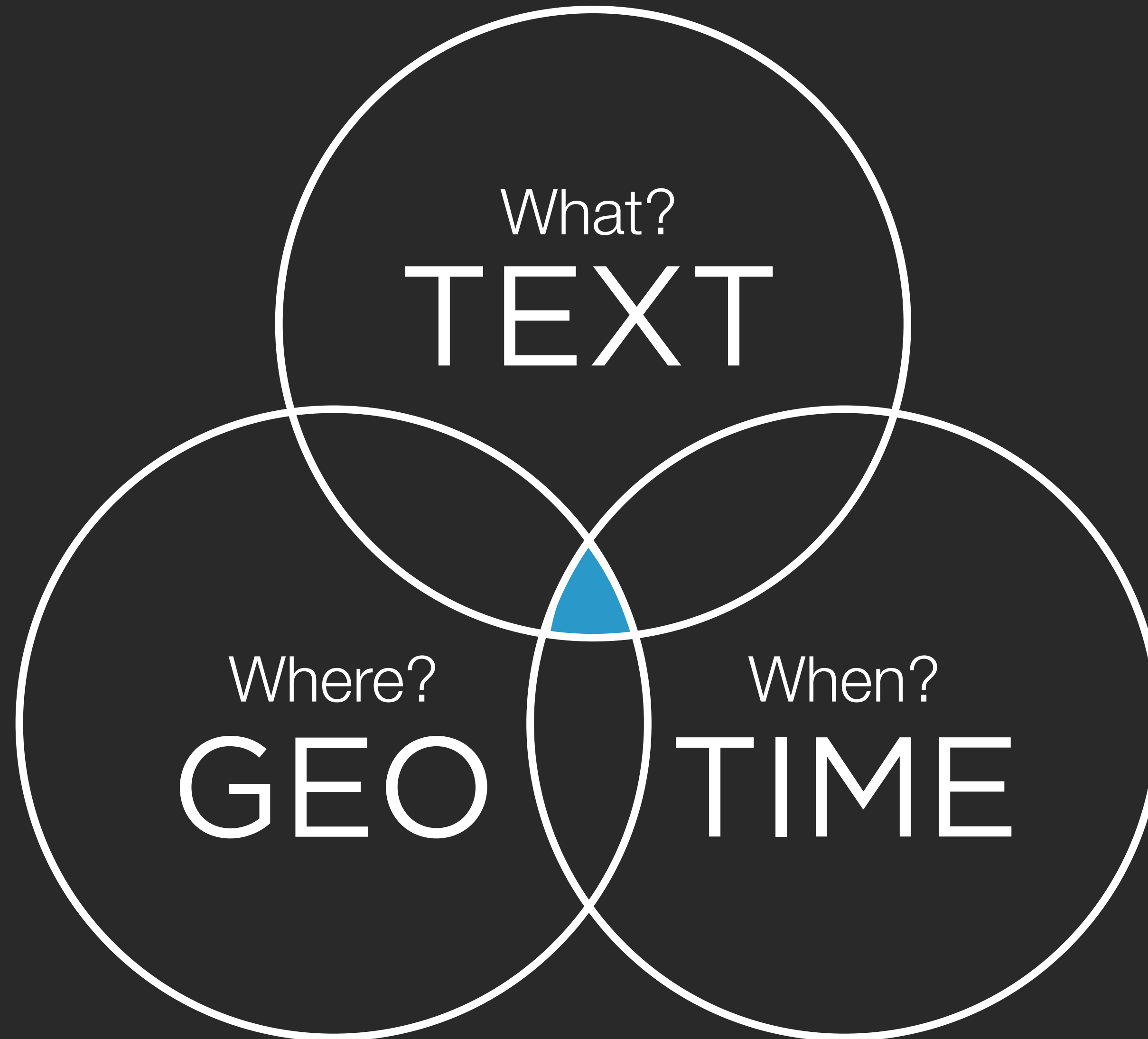
Time + Text

UEFA Champions League

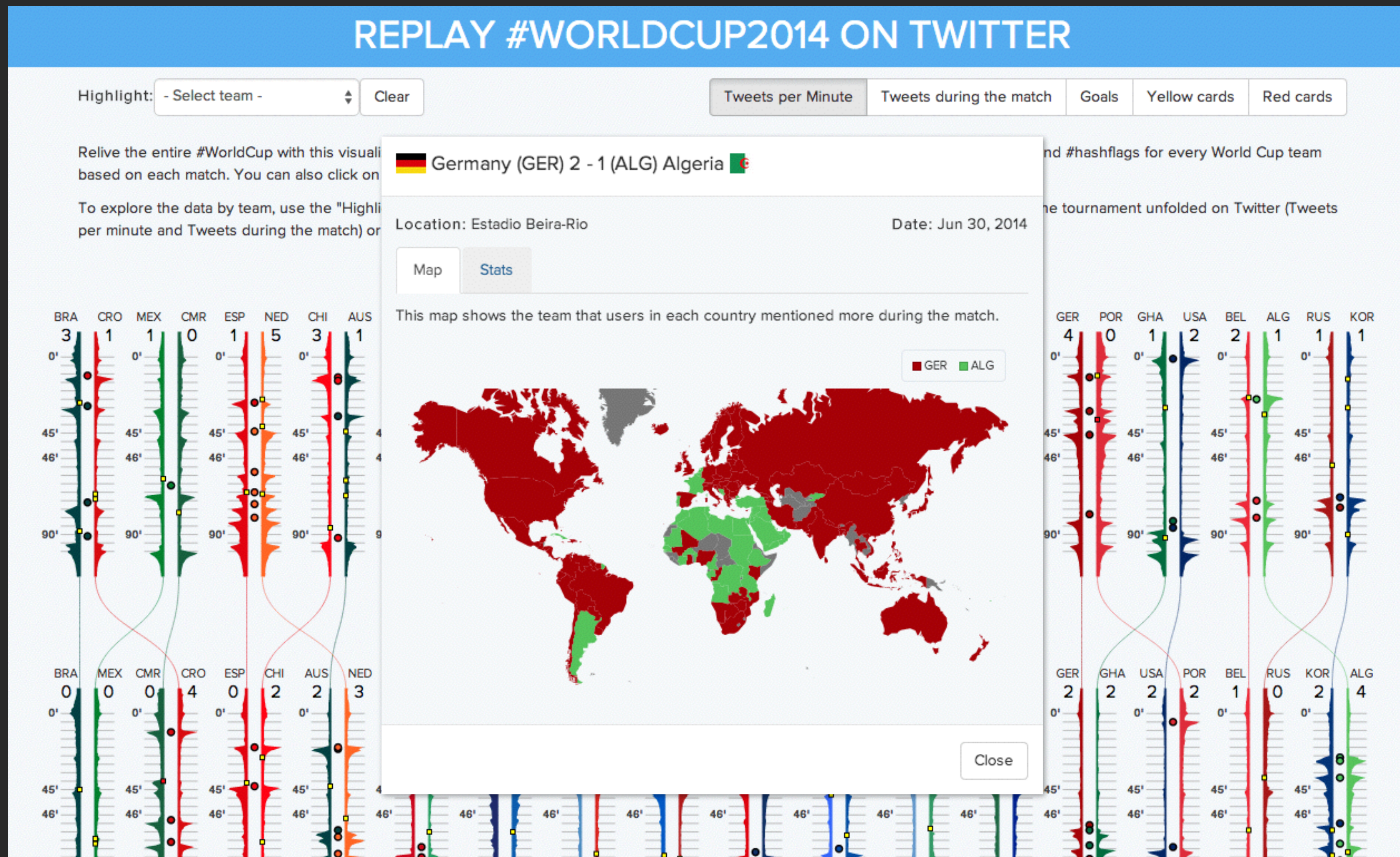


+ players

Visualize Data



Time + Text + Geo World Cup 2014



data. at Twitter

“Tweets”

- users
- followers graph
- logs
- etc.

big data => small data

Get data

1

2

self, peer, external

Evaluate

3

vis

Visualize

What? Where? When?

Who? ...

(with deadline)

@kristw / <https://interactive.twitter.com>

+ visualizations by @philogb, @miguelrios & @trebor