

# Quantitative Forecasting of Political Conflict

Philip A. Schrodt  
Department of Political Science  
Pennsylvania State University  
[schrodt@psu.edu](mailto:schrodt@psu.edu)  
<http://eventdata.psu.edu>

Presentation at the State Dept INR Workshop on  
Data-Driven Forecasting and Quantitative Analysis  
Arlington VA  
23 April 2013

# The Debate



## ► ARGUMENT

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

### Why the World Can't Have a Nate Silver

The quants are riding high after Team Data crushed Team Gut in the U.S. election forecasts. But predicting the Electoral College vote is child's play next to some of these hard targets.

BY JAY ULFELDER | NOVEMBER 8, 2012

# Vs.



## ► ARGUMENT

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

### Predicting the Future Is Easier Than It Looks

Nate Silver was just the beginning. Some of the same statistical techniques used by America's forecaster-in-chief are about to revolutionize world politics.

BY MICHAEL D. WARD, NILS METTERNICH | NOVEMBER 16, 2012

# Main points

- Event data analysis, originally developed under DARPA funding 1965-1980, is a well-understood technique for collecting systematic information on political interactions over time
- Contemporary automated coding methods allow data to be collected in a transparent and reproducible manner in real time at a very low marginal cost
- Statistical models provide 70%-80% accuracy in predicting violence in protracted conflicts in out-of-sample tests at policy-relevant forecast leads
- Exponential increases in the availability of information on political events has produced a major change in the viability and utility of these methods

# Drivers of change in social science research in the 21<sup>st</sup> century

- Big data
  - Effective use of high performance computing
  - Wider range of analytical methods
  - Wider range of data
- Decentralized collaborative environments
  - Open source / open access
  - Two weeks after GDELT was released, one could download multiple tutorials, dozens of visualizations, and hundreds of lines of ready-to-use R code
  - Increased interest in policy and private-sector applications of cutting-edge techniques

# Science

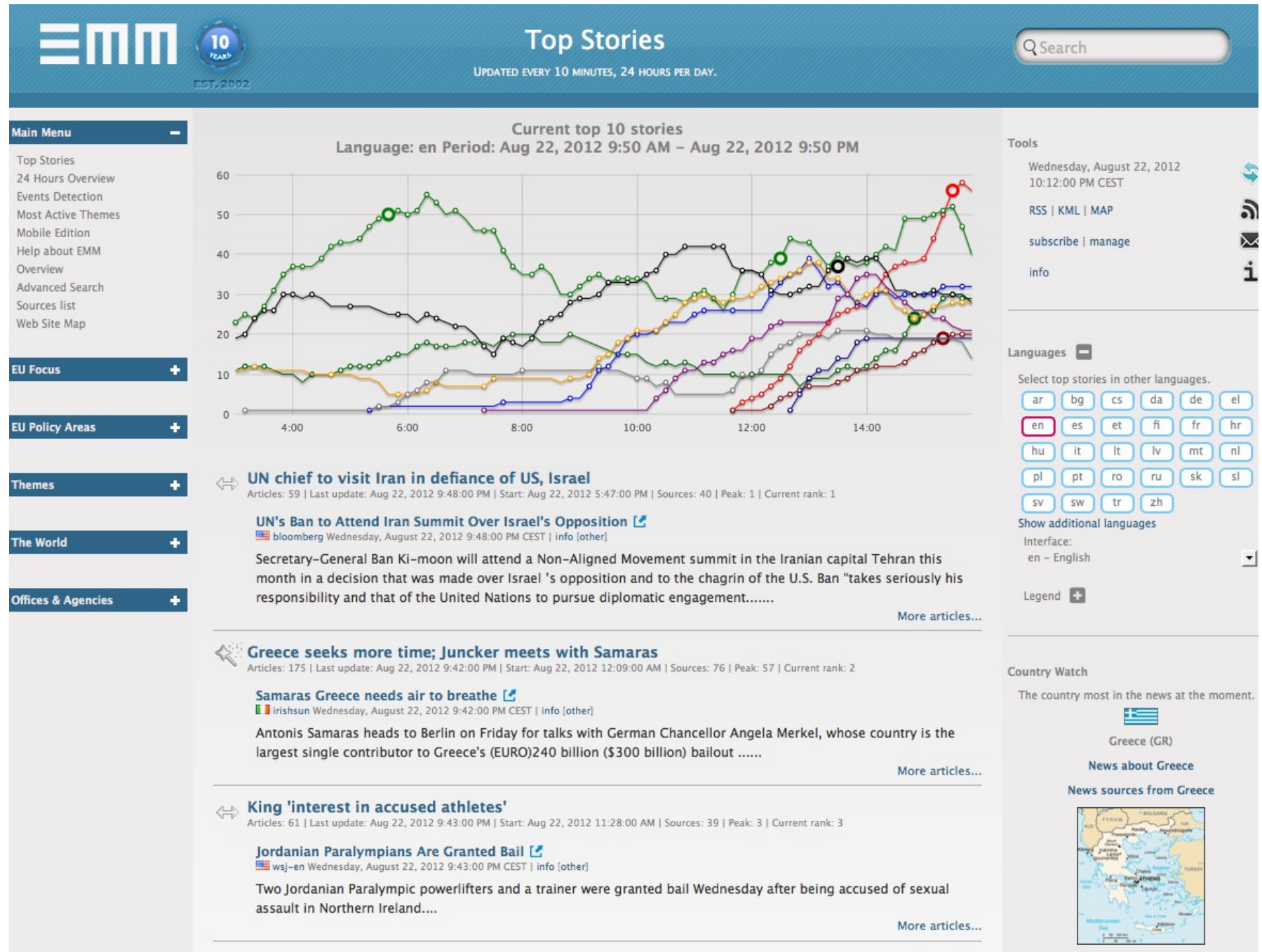
11 February 2011 | \$10

example knowledge

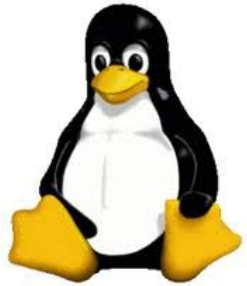
scientific  
information  
research  
data  
climate  
time  
science  
scientists  
brain  
analysis  
visualization  
many  
new  
access  
researchers  
community  
funding  
including  
first  
online  
algorithms  
publications  
policy  
understanding  
complex  
systems  
interdisciplinary  
approaches  
must  
important  
use  
cell  
future  
work  
model  
need  
large  
sharing  
may  
systems  
public  
tools  
2009  
health  
project  
digital  
used  
quality  
communication  
changes  
methods  
one  
human  
become  
2010  
et  
make  
zone  
AAAS



# European Media Monitor



# Open Source Software



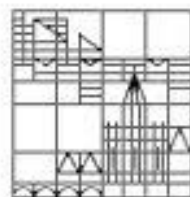
# Political Data Sources



Minorities at Risk



Universität  
Konstanz





**SpatialKey** North-Central Hotspot Save

Xycentalaf 0 Xychad 0 Xysudan 0

Map 1

**Chad Timeline : Event Date**

Count

| Year | Count |
|------|-------|
| 1997 | 10    |
| 1998 | 10    |
| 1999 | 10    |
| 2000 | 10    |
| 2001 | 10    |
| 2002 | 10    |
| 2003 | 10    |
| 2004 | 10    |
| 2005 | 10    |
| 2006 | 120   |
| 2007 | 100   |
| 2008 | 80    |
| 2009 | 20    |
| 2010 | 10    |
| 2011 | 10    |

**CAR Timeline : Event Date**

Count

| Year | Count |
|------|-------|
| 1997 | 40    |
| 1998 | 20    |
| 1999 | 10    |
| 2000 | 10    |
| 2001 | 40    |
| 2002 | 50    |
| 2003 | 60    |
| 2004 | 30    |
| 2005 | 20    |
| 2006 | 50    |
| 2007 | 100   |
| 2008 | 80    |
| 2009 | 80    |
| 2010 | 120   |
| 2011 | 10    |

**Sudan Timeline : Event Date**

Count

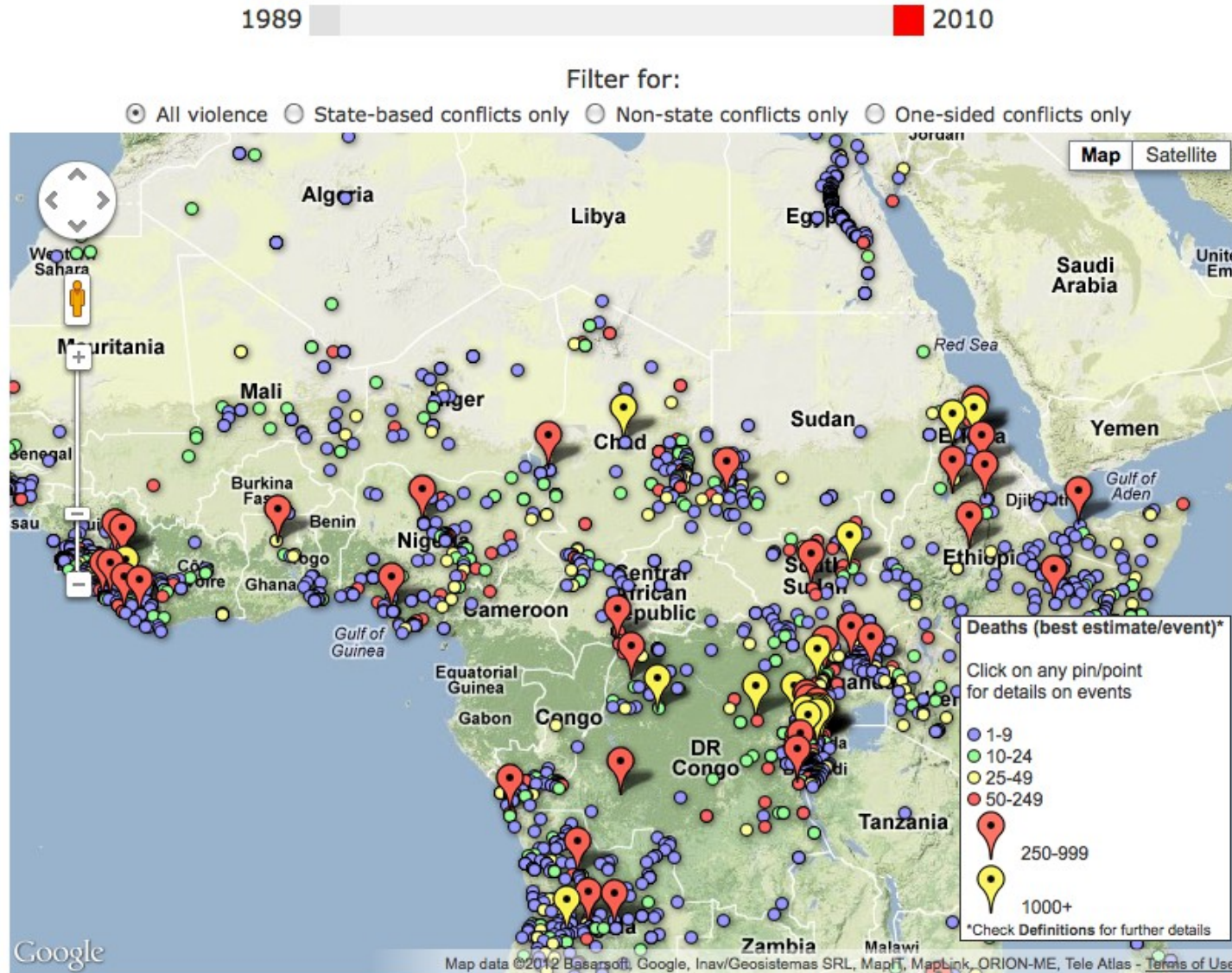
| Year | Count |
|------|-------|
| 1996 | 200   |
| 1997 | 200   |
| 1998 | 200   |
| 1999 | 200   |
| 2000 | 200   |
| 2001 | 200   |
| 2002 | 280   |
| 2003 | 80    |
| 2004 | 240   |
| 2005 | 320   |
| 2006 | 120   |
| 2007 | 100   |
| 2008 | 280   |
| 2009 | 240   |
| 2010 | 200   |
| 2011 | 280   |



# Uppsala Georeferenced Event Dataset

## Welcome to the UCDP GED - Uppsala Conflict Data Program's Georeferenced Event Dataset

The interactive map below covers all the locations geocoded as part of GED. **Drag and move the red and grey thumbs of the slider below to filter for particular intervals, years, or even particular months or days.** Use the filters to select only one of the three UCDP types of violence. Clicking on each dot or pin will display information regarding events taking place in those respective locations. [Download data here.](#)



The GED is the product of two and a half years of work at the **Department of Peace and Conflict Research, Uppsala University**. The UCDP GED contains conflict data disaggregated spatially and temporally down to the level of the individual incidents of violence. For more details please see the [About UCDP](#) link above.

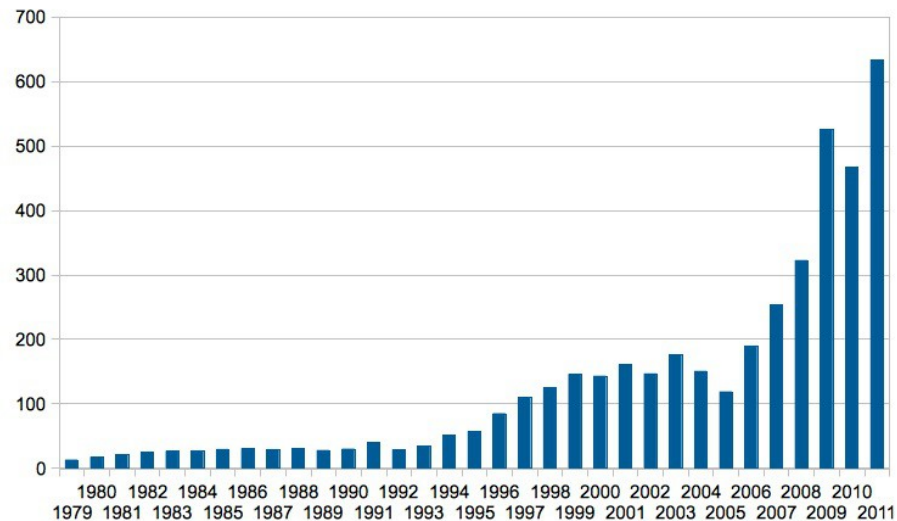
# ICEWS Phase 1 Event Data

- 30-gigabytes of text from Lexis-Nexis
- 25 sources
- 8-million stories
- 26-million sentences
  - Only first four sentences coded in each story
- 3-million events
- Generally two orders of magnitude greater than any prior event coding effort



# GDELT Event Data

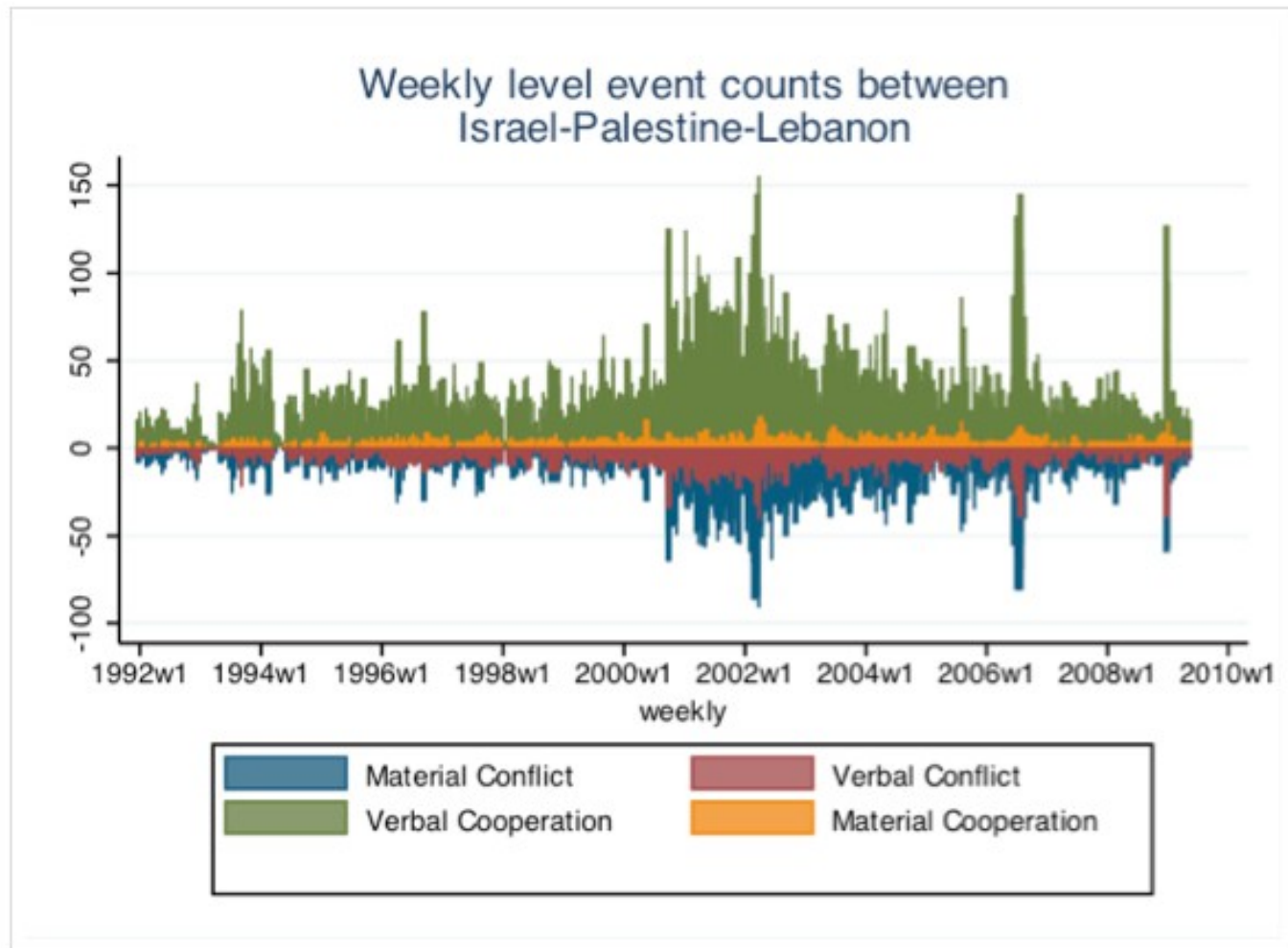
- Open source collaboration of Univ of Illinois (Kalev Leetaru) and Penn State
- Global coding, Jan-1979 to Jun-2012
- 200-million events based on open news sources
- CAMEO event, actor and sub-state agent coding
  - 15,000 verb phrase dictionary
  - 40,000+ political actors and agents
- Geolocated to city level
- Current intake of 20,000 to 100,000 stories per day
- Planned daily updates, backfit to 1800



- <http://eventdata.psu.edu/data.dir/GDELT.html>

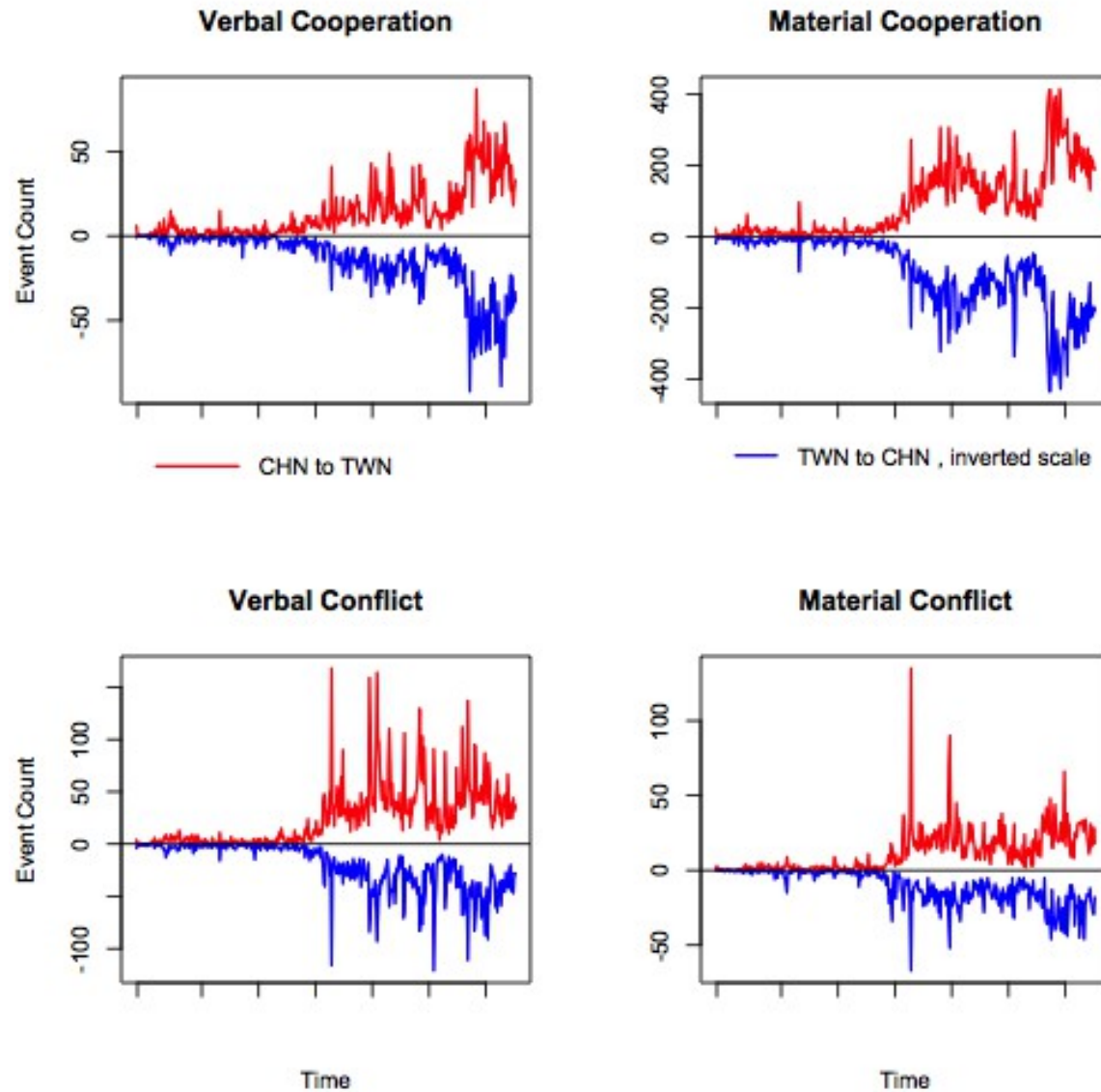
# Israel-Palestine 1992-2010

Visualization by Jay Yonamine

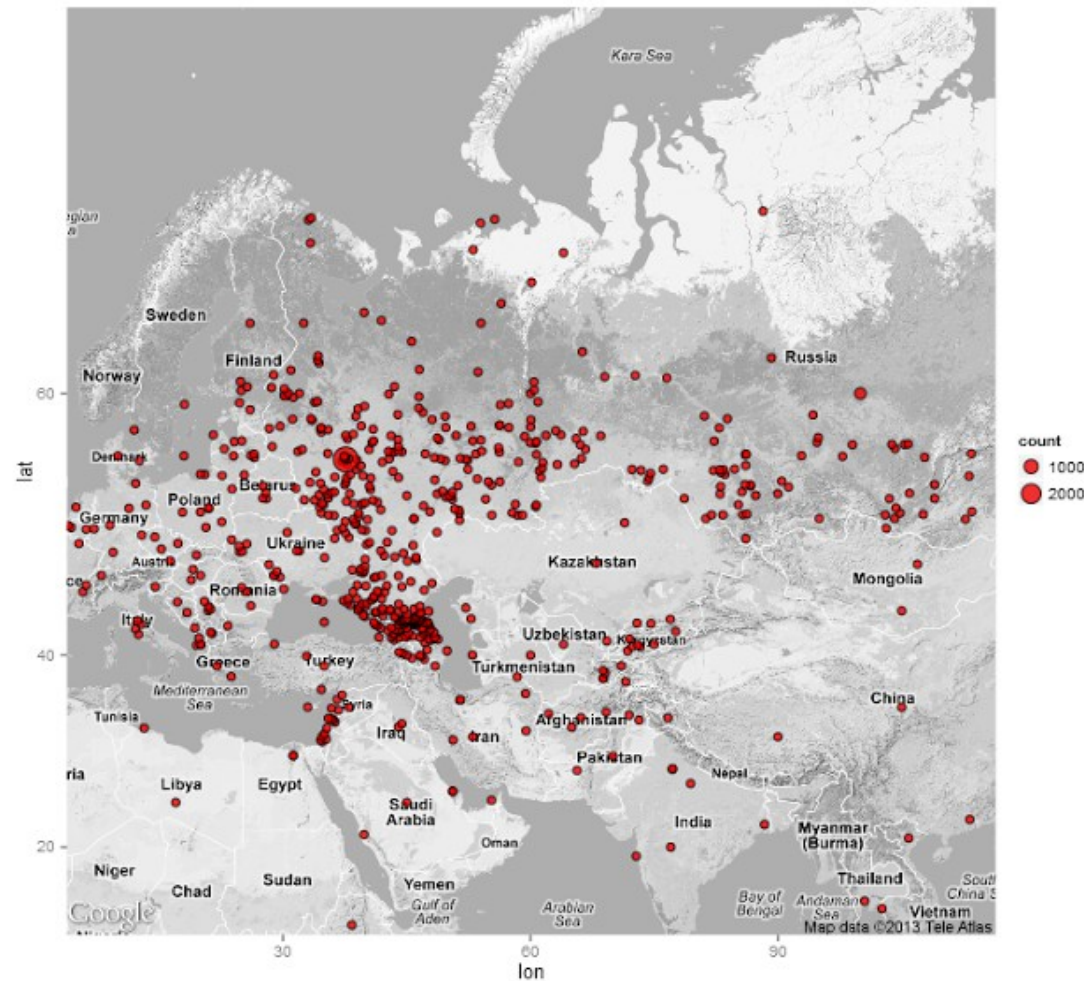




# GDELT Quad Counts: China - Taiwan

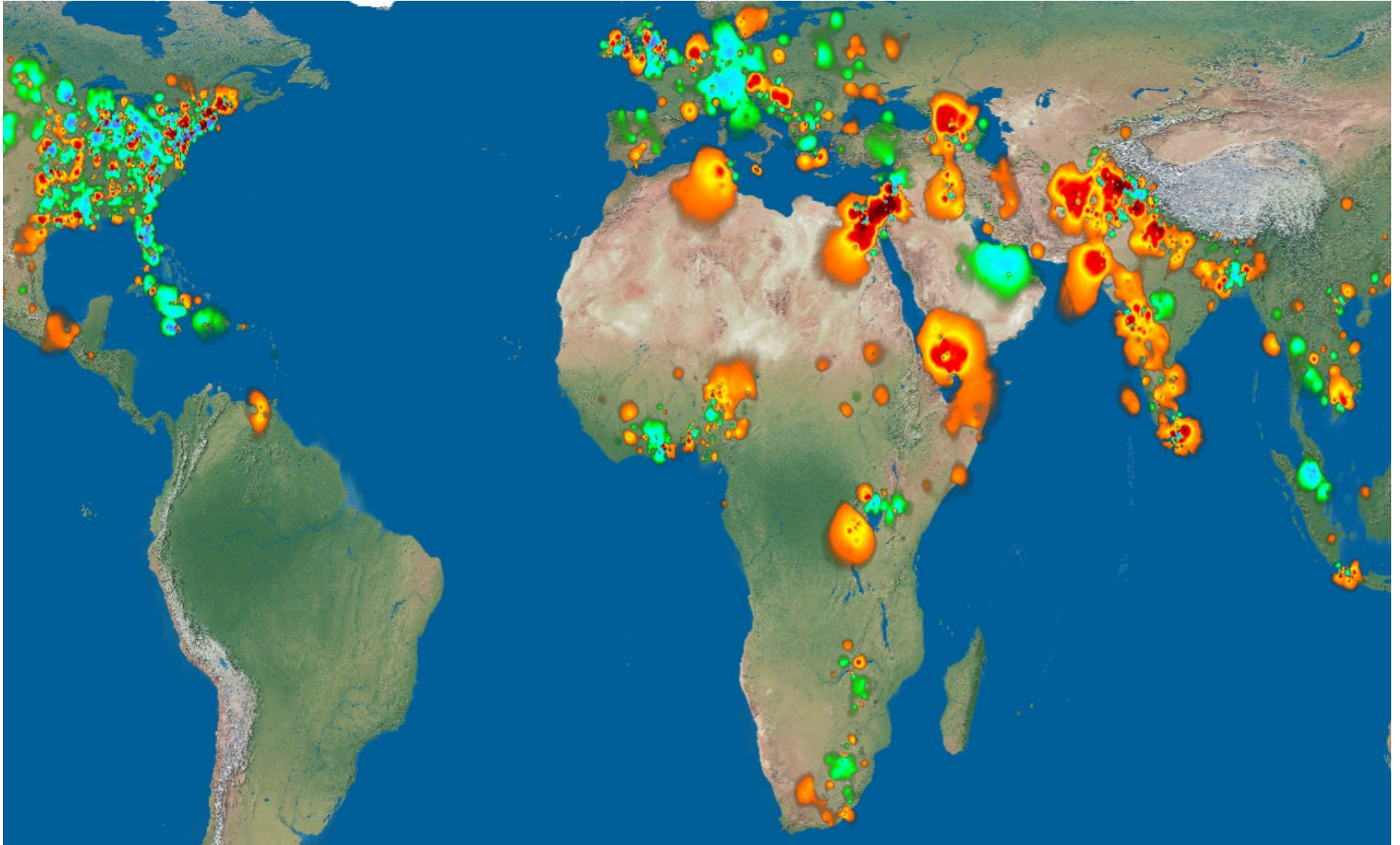


# GDELT: Civil Protests



Source: Rolf Fredheim, <http://quantifyingmemory.blogspot.co.uk/2013/04/mapping-gdelt-data-in-r-and-some.html>

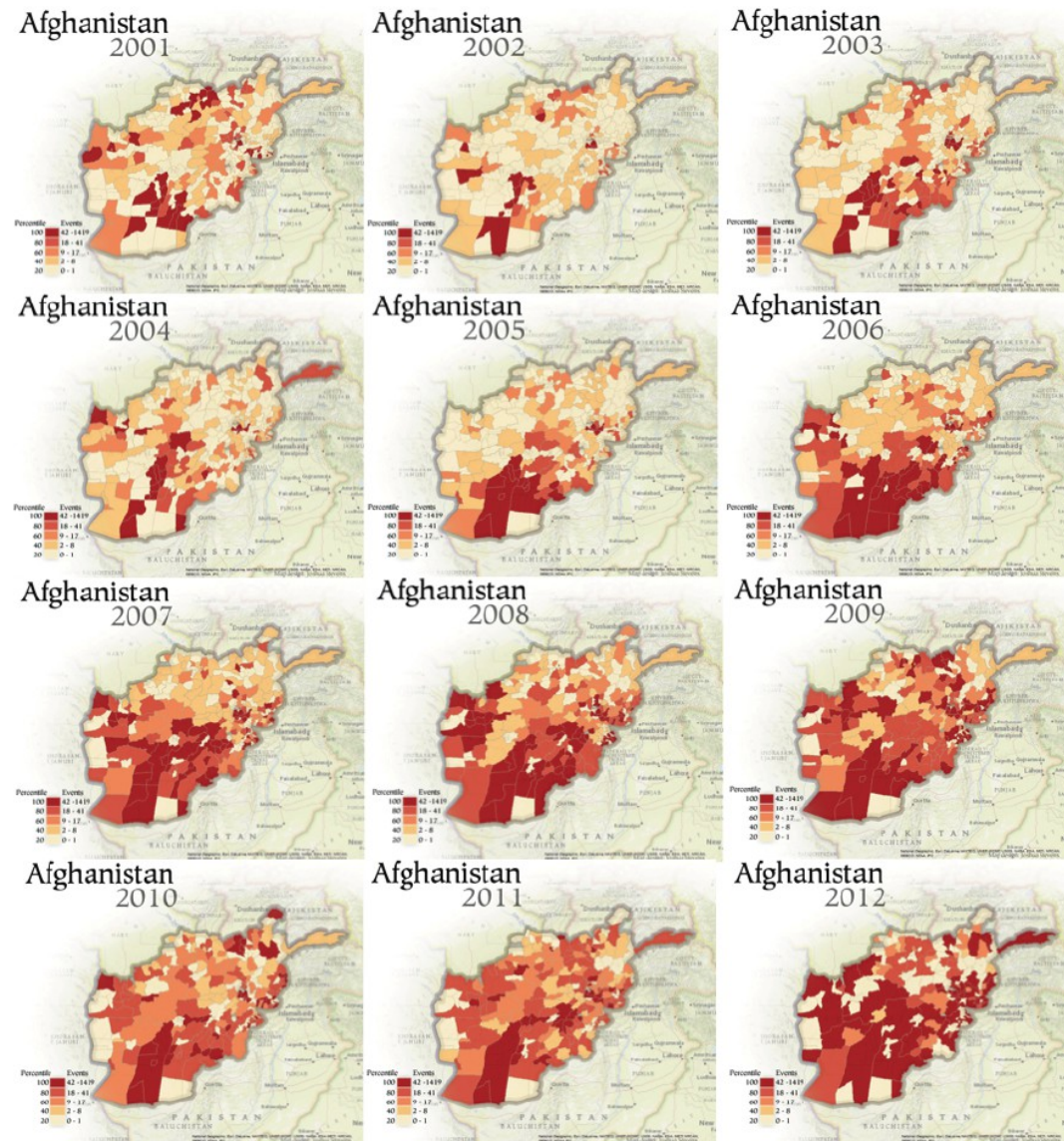
# GDELT: News intensity 29 Jan 2011



Source: Kalev Leetaru



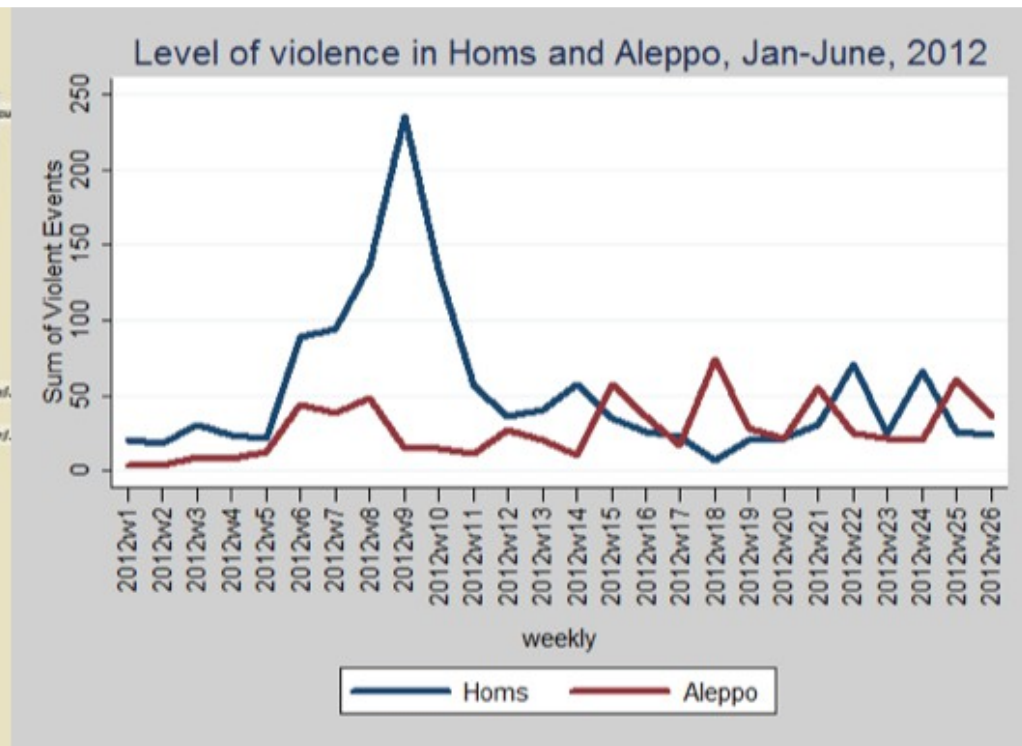
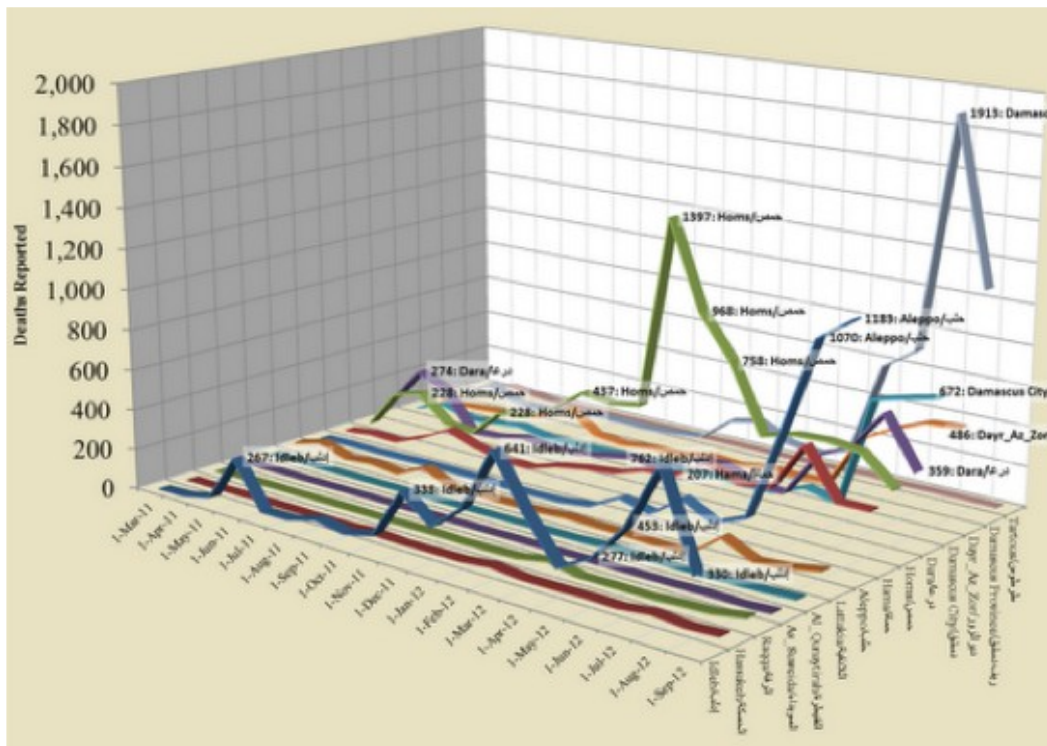
# GDELT: Afghanistan by District, 2001-2012



[This is *not* [!] Wikileaks!]

Source: Jay Yonamine dissertation, chapter 5

# Violence in Syria: Ushahihi and GDELT

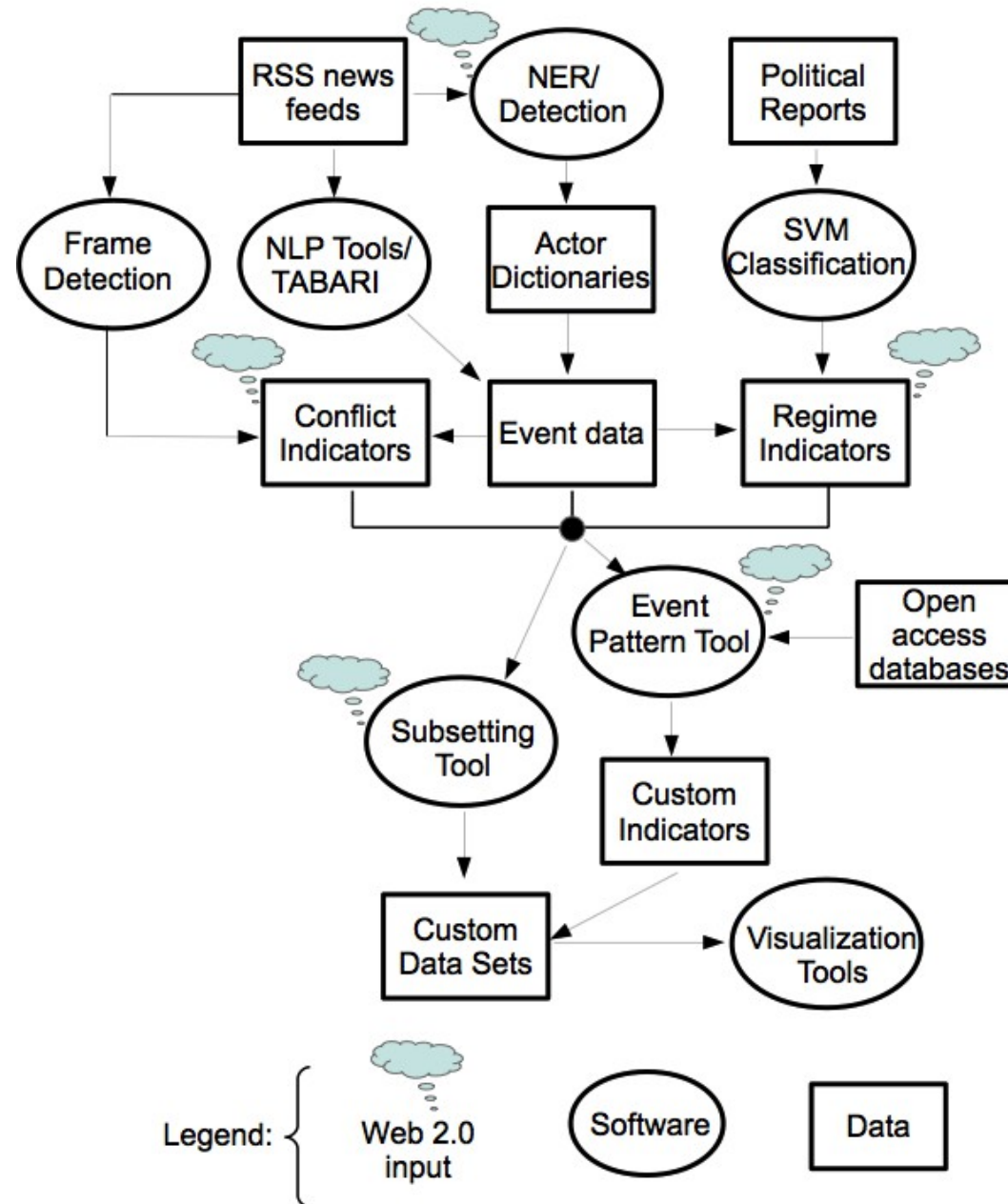




# Event Model: Core Innovation

- Once calibrated, real-time event forecasting models can be run entirely without human intervention
  - Web-based news feeds provide a rich multi-source flow of political information in real time
  - Statistical models can be run and tested automatically, and are 100% transparent
- In other words, for the first time in human history—quite literally—we have a system that can provide real-time measures of political activity without any human intermediaries

# Integrated open, real time data generation



# The Debate



► **ARGUMENT**

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

## Why the World Can't Have a Nate Silver

The quants are riding high after Team Data crushed Team Gut in the U.S. election forecasts. But predicting the Electoral College vote is child's play next to some of these hard targets.

BY JAY ULFELDER | NOVEMBER 8, 2012

# Vs.



► **ARGUMENT**

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

## Predicting the Future Is Easier Than It Looks

Nate Silver was just the beginning. Some of the same statistical techniques used by America's forecaster-in-chief are about to revolutionize world politics.

BY MICHAEL D. WARD, NILS METTERNICH | NOVEMBER 16, 2012

# Factors encouraging technical political forecasting

- Success of forecasting models in other behavioral domains
  - Macroeconomic forecasting [maybe...]
  - Elections: Nate Silver effect
  - Demographic and epidemiological forecasting
  - Famine forecasting: USAID FEWS model
  - Example: statistical models for mortgage repayment were quite accurate
  - *Moneyball*
- Technological imperative
  - Increased processing capacity
  - Information available on the web
  - “Moore’s Law states that computing power doubles every 18 months. Human cognitive ability is pretty much a constant. This leads to some interesting and not always desirable substitution effects”  
Larry Bartels, Princeton University

# Factors encouraging technical political forecasting

- Demonstrated utility of existing methods, which tend to converge on about 80% accuracy
  - Political Instability Task Force
  - ICEWS
  - “Big Data” analytical methods
- Decision-makers now expect visual displays of analytical information, which in turn requires systematic measurement
  - “They won’t read things any more”



# Contemporary Technical Political Forecasting

- State Failures Project 1994
- Joint Warfare Analysis Center 1997
- FEWER [Davies and Gurr 1998]
- Various UN and EU forecasting projects
- Center for Army Analysis 2002-2005
- Swiss Peace Foundation FAST 2000-present
- Political Instability Task Force 2002-present
- DARPA ICEWS 2007-present
- Peace Research Center Oslo (PRIO) and Uppsala University UCDP political forecasting models

# Integrated Conflict Early Warning System

- Unclassified project funded by DARPA Information Processing Techniques Office
- Funding at \$40-million for 2007-2011
  - Largest quantitative conflict analysis project since the 1970s
- Objective is real-time forecasting of indicators of political instability in Asia with 6-24 month leads, 70%-80% accuracy
- Machine-coded event data has proven to be the core methodology for accurate forecasts
- Data covers 1997-present with 8.5-million stories from 27 sources
- Model accuracy has been assessed with a strict split-sample design

## Reference:

Sean O'Brien. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87-104, 2010.

# ICEWS “Events of Interest”

Domestic Political Crisis—Significant opposition to the government, but not to the level of rebellion or insurgency (for example, power struggle between two political factions involving disruptive strikes or violent clashes between supporters)

Rebellion—Organized opposition where the objective is to seek autonomy or independence

Insurgency—Organized opposition where the objective is to overthrow the central government

Ethnic/Religious Violence—Violence between ethnic or religious groups that is not specifically directed against the government

International Crisis—Conflict between two or more states or elevated tensions between two or more states that could lead to conflict

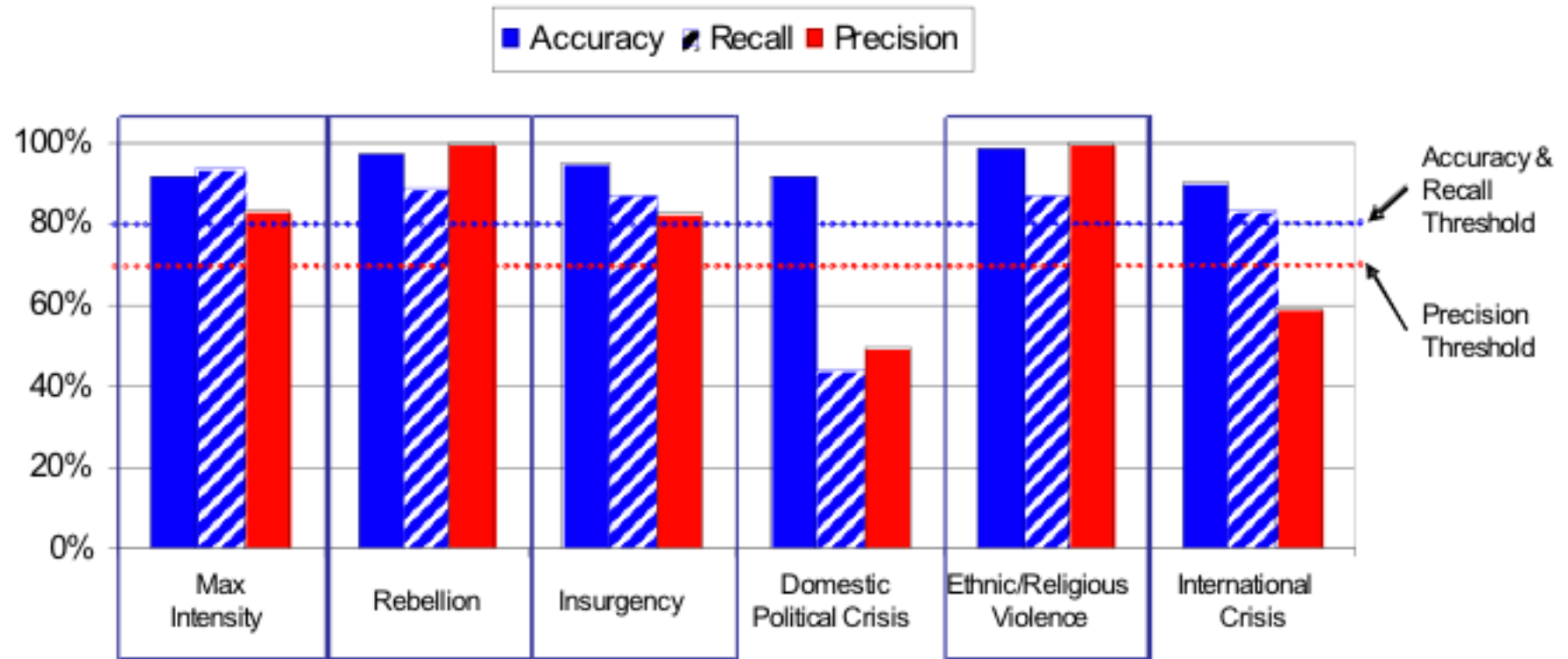
# ICEWS Evaluation Criteria

$$\textit{Accuracy} = \frac{\textit{number of correct predictions}}{\textit{total predictions made}}$$

$$\textit{Recall} = \frac{\textit{number of correctly predicted conflicts}}{\textit{total conflicts that occurred}}$$

$$\textit{Precision} = \frac{\textit{number of correctly predicted conflicts}}{\textit{total conflicts predicted}}$$

# ICEWS Phase 1 Results: LM-ATL Out-of-Sample Results (DARPA Chart)



- Exceeds metrics for the maximum intensity index and 3 instability events: Rebellion, Insurgency, and Ethnic/Religious Violence – Passes Phase 1 gates
- By integrating improved versions of best models from multiple perspectives, team achieves more accurate, precise forecasts than any one model alone



# Predicting ICEWS EOIs with GDELT Data

| EOI                    | Out of sample test cases | Acc | Prec | Recall |
|------------------------|--------------------------|-----|------|--------|
| Rebellion              | 2005-2006                | 91% | 87%  | 67%    |
| International Conflict | 2005-2006                | 92% | 65%  | 66%    |
| Ethnic/<br>Religious   | 2007-2008                | 98% | 75%  | 84%    |
| International conflict | 2007-2008                | 95% | 92%  | 64%    |
| ICEWS benchmarks       | 2005-2006 [quarterly]    | 80% | 75%  | 80%    |

Method: Monthly data; training cases 1998 to beginning of test cases; 6-month lag; random forest models

# Political Instability Task Force

- US government, multi-agency: 1995-present
- Statistical modeling of various forms of state-level instability
- Forecasting models actively used since about 2005
  - Two year probability forecasts with roughly 80% accuracy (AUC)
  - Predominantly logistic models with a simple “standard PITF” set of variables; shifting to Bayesian approaches
- Finances a variety of data sets, including Polity IV, Worldwide Atrocities Dataset, Institutions and Elections Data

# But these models don't work! *Wired* magazine tells me so!



[PREVIOUS POST](#)

[NEXT POST](#)

## Pentagon's Prediction Software Didn't Spot Egypt Unrest

By [Noah Shachtman](#) February 11, 2011 | 7:00 am | Categories: [DarpaWatch](#)

[Follow @dangerroom](#)

518

0

172

[Tweet](#)

[+1](#)

[Share](#)

[Like](#)

[Send](#)

[497 people like this. Sign Up to see what your friends like.](#)



In the last three years, America's military and intelligence agencies have spent more than \$125 million on computer models that are supposed to forecast political unrest. It's the latest episode in Washington's four-decade dalliance with future-spotting programs. But if any of these algorithms saw the upheaval in Egypt coming, the spooks and the generals are keeping the predictions very quiet.

# The Forecaster's Quartet

- Nassim Nicholas Taleb. *The Black Swan*  
[most entertaining]
- Daniel Kahneman. *Thinking Fast and Slow*  
[30 years of research which won Nobel Prize]
- Philip Tetlock. *Expert Political Judgment*  
[most directly relevant]
- Nate Silver. *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*  
[easiest read; examples from multiple domains]

# The Forecaster's Quartet

- Nassim Nicholas Taleb. *The Black Swan*  
[most entertaining...or most obnoxious...]
- Daniel Kahneman. *Thinking Fast and Slow*  
[30 years of research which won Nobel Prize]
- Philip Tetlock. *Expert Political Judgment*  
[most directly relevant]
- Nate Silver. *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*  
[easiest read; examples from multiple domains]

# Irreducible sources of error

- Specification error: no model of a complex, open system can contain all of the relevant variables;
- Measurement error: with very few exceptions, variables will contain some measurement error
  - presupposing there is even agreement on what the “correct” measurement is in an ideal setting;
  - Predictive accuracy is limited by the square root of measurement error: if your reliability is 80%, your accuracy can't be more than 90%
- Free will
  - Rule-of-thumb from our rat-running colleagues: “A genetically standardized experimental animal, subjected to carefully controlled stimuli in a laboratory setting, will do whatever it wants.”
- Quasi-random structural error: Complex and chaotic deterministic systems behave as if they were random under at least some parameter combinations



# Statistical challenges

- Rare events
  - Incorporate much longer historical time lines? — Schelling used Caesar's *Gallic Wars* to analyze nuclear deterrence
  - Calibration can be very tricky
- Analysis of event sequences, which are not a standard data type
- Causality
  - *Oxford Handbook of Causation* is 800 pages long
- Integration of qualitative and SME information
  - Bayesian approaches are promising but to date they have not really been used

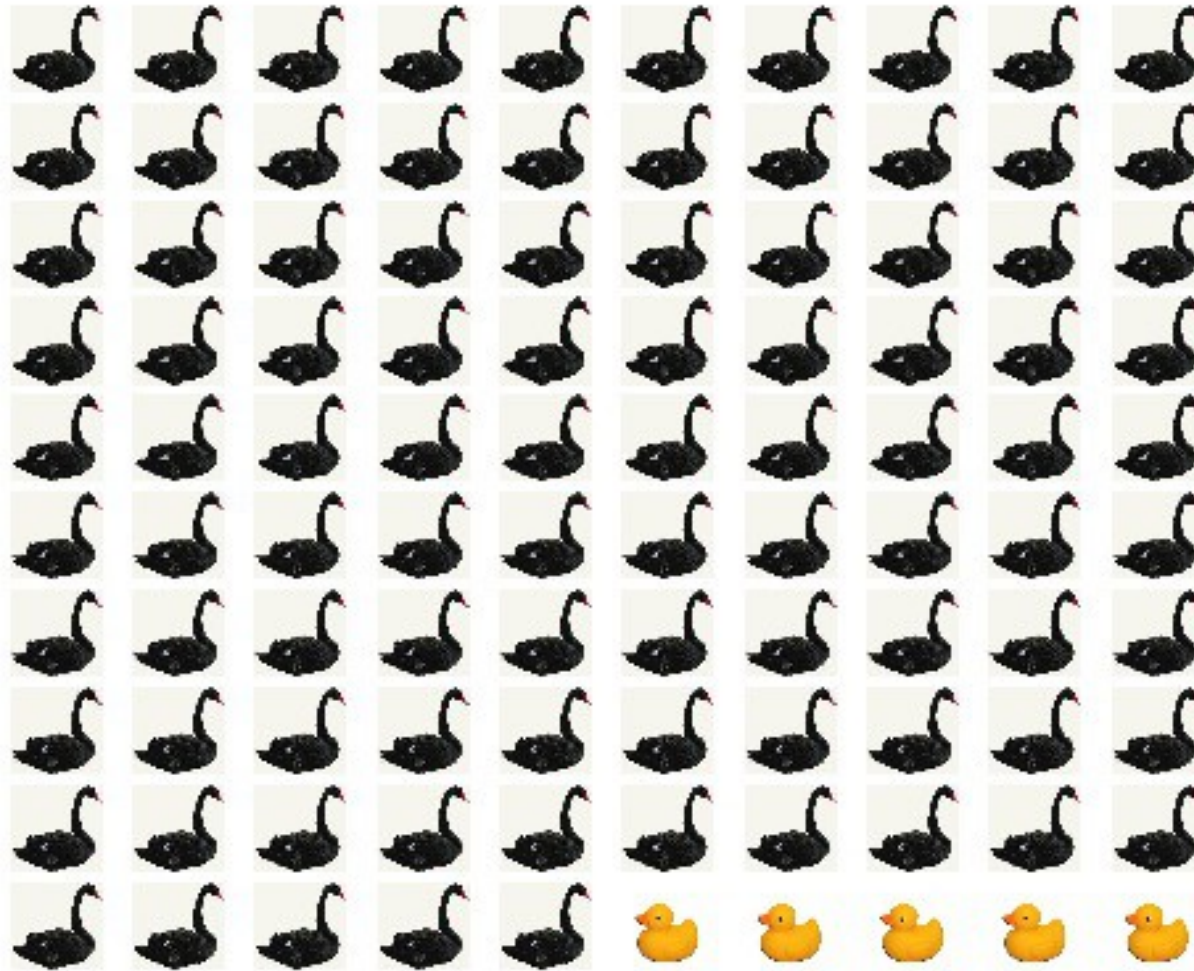
# Choice of model matters

- Arab Spring is an unprecedented product of the new social media
  - Model used by Chinese censors of NSM: King, Peng, Roberts 2012
  - Next likely candidates: Africa
- Arab Spring is an example of an instability contagion/diffusion process
  - Eastern Europe 1989-1991, OECD 1968, CSA 1859-1861, Europe 1848, Latin America 1820-1828
  - Next likely candidates: Central Asia
- Arab Spring is a black swan
  - There is no point in modeling black swans, you instead build systems robust against them

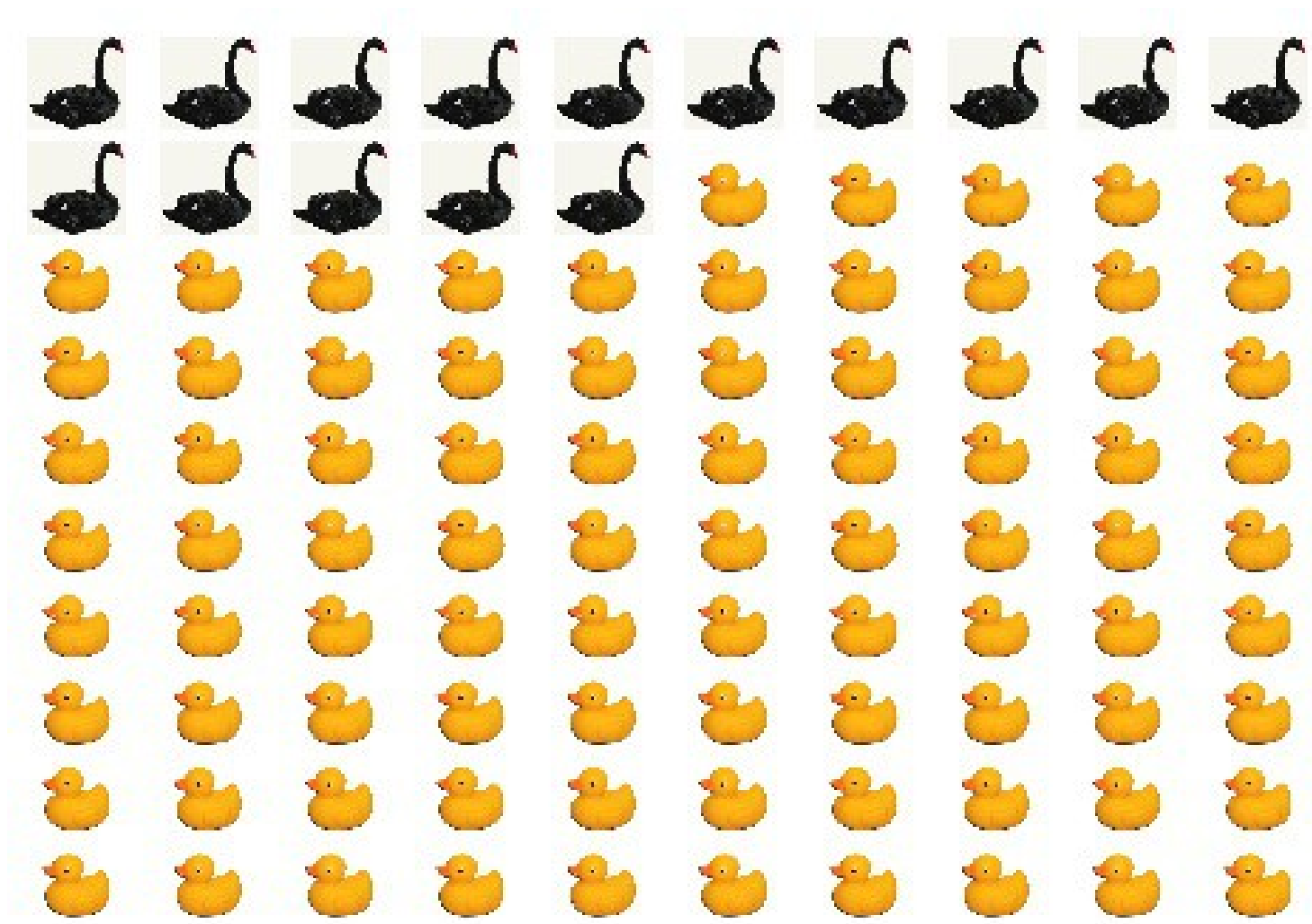
# Why do data based methods fail?

- Clever anecdotes/charm/rank trump data
  - *Moneyball*, managed mutual funds
- Ideology trumps data
  - 2012 Romney campaign
- “It is difficult to get a man to understand something, when his salary depends on his not understanding it.”— Upton Sinclair
  - Sub-prime mortgage crisis
- The model was incorrect
  - Reinhart-Rogoff debt/growth model (using Excel also doesn't help)
- Black swans: very low probability events which by definition cannot be predicted
  - Taleb: don't use normal-distribution models in a power-law world
- Difficulties of incorporating probabilistic reasoning by individuals and organizations
  - You aren't stupid, you're human

# DARPA-World



# NNT-World



# Why should we care about ducks?



Size



Variety



Quantity



Suspicious behaviors



# Where should new efforts go?

- More precise data?
  - We may be at a point of diminishing returns here
  - There is no point in precisely measuring noise
- Better models?
  - Machine learning and Bayesian models present new opportunities
  - But only if the results can be correctly understood
- Analyzing how probabilistic forecasts can be incorporated into the decision-making process
  - Preferably before the Chinese do...

# Questions?

Philip A. Schrodtt

Political Science

Pennsylvania State University

State College, PA 16802

Phone: 814-863-8978

Email: [schrodtt@psu.edu](mailto:schrodtt@psu.edu)

Project Web Site: <http://eventdata.psu.edu>

