

Timelines as Summaries of Popular Scheduled Events

Omar Alonso - Kyle Shiells

Microsoft

13 May 2013



Disclaimer

The views, opinions, positions, or strategies expressed in this talk are mine and do not necessarily reflect the official policy or position of Microsoft.

Agenda

- Introduction
- The World Cup data set
- Framework
- Games
- Evaluation
- Conclusions and future work

Introduction

- Media coverage for known events
- Certain actions, facts and opinions are omitted from objective summaries
- This is usually the case in sports
- Metaphor: “global bar atmosphere”

Research questions

- Would it be possible to follow a game via a microblogging site?
- Can we present a game summary in a way that combines the main events along with salient comments and the dynamics of the game in a timeline?

Contributions

- A new timeline display designed for sports that allow observers to compare games within the same sport and to a lesser extent across sports
- Annotation of the timeline with popular terms that reflect what the crowd was discussing at key moments
- A study of a few games of the World Cup using our techniques and some statistical insight into the dataset

World Cup data set

- FIFA 2010 World Cup
- #worldcup
- June 5 to July 17 2010
- 2,901,840 tweets
- 50 languages
- 32 teams participated

Top terms

Keyword	Frequency	Keyword	Frequency
Watch	306,880	Team	86,294
World	301,856	Today	77,837
Cup	284,089	South	74,069
Free	209,129	Final	74,063
Fifa	190,417	Live	72,098
game	182,824	Good	70,405
vuvuzela	178,453	soccer	69,519
match	159,631	:)	57,968
online	131,528	football	55,893
2010	127,268	Great	43,380
Twitter	101,413	Best	38,894

Language distribution

Language	Count	World Cup Rank	Overall Rank
English	1,799,136	1	1
Japanese	454,617	2	2
Spanish	195,011	3	4
Portuguese	168,689	4	3
German	63,739	5	11
Indonesian	40,156	6	5
Dutch	20,823	7	8
Italian	20,088	8	9
Malay	19,171	9	6
French	18,395	10	12

Frequency distribution by hashtag

Country	# tweets	Country	# tweets
#GER	302,903	#DEN	63,467
#JPN	275,474	#KOR	56,572
#USA	246,413	#AUS	54,612
#BRA	232,075	#CIV	47,379
#ENG	202,203	#CMR	46,542
#ESP	197,023	#ALG	37,845
#ARG	174,638	#RSA	37,005
#URU	108,393	#PRK	36,540
#NED	107,502	#GRE	34,809
#PAR	105,585	#NGA	32,867
#POR	97,275	#SRB	30,789
#MEX	95,705	#SVK	30,127
#ITA	79,306	#SVN	28,881
#FRA	72,004	#NZL	26,794
#CHI	68,582	#HON	24,927
#GHA	66,903	#SUI	24,697

Framework

- Data granularity: at the minute level
- Advantages: known date of the game, start and end time.
- Event e as rules of the game

Display design

- x-axis is time with marks
- 2 teams; each team with a color
- Gray color to show totals
- Variables encoded:
 - Game sequence
 - Rules
 - Game outcome
 - Activity per team
 - Noise level of the virtual bar

Timeline generation

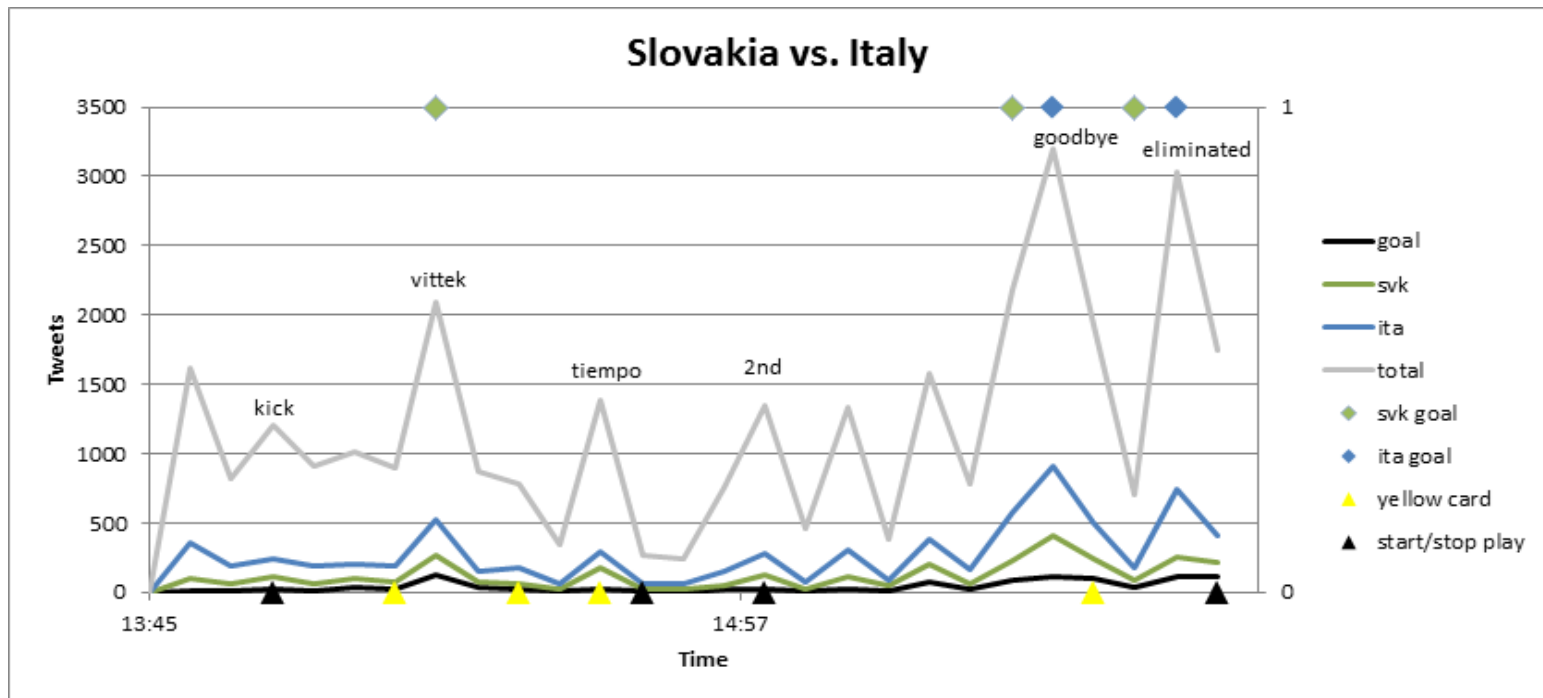
- Collect all tweets
- Group tweets into 5min buckets
- Within each bucket
 - Volume counts
 - Term counts

A few games

- Let's take a look at 6 games
 - #22 (Slovenia-USA)
 - #41 (Slovakia-Italy)
 - #51 (England-Germany)
 - #61 (Uruguay-Netherlands)
 - #62 (Germany-Spain)
 - #63 (Germany-Uruguay)
- Present timeline and few examples
- Evaluation results

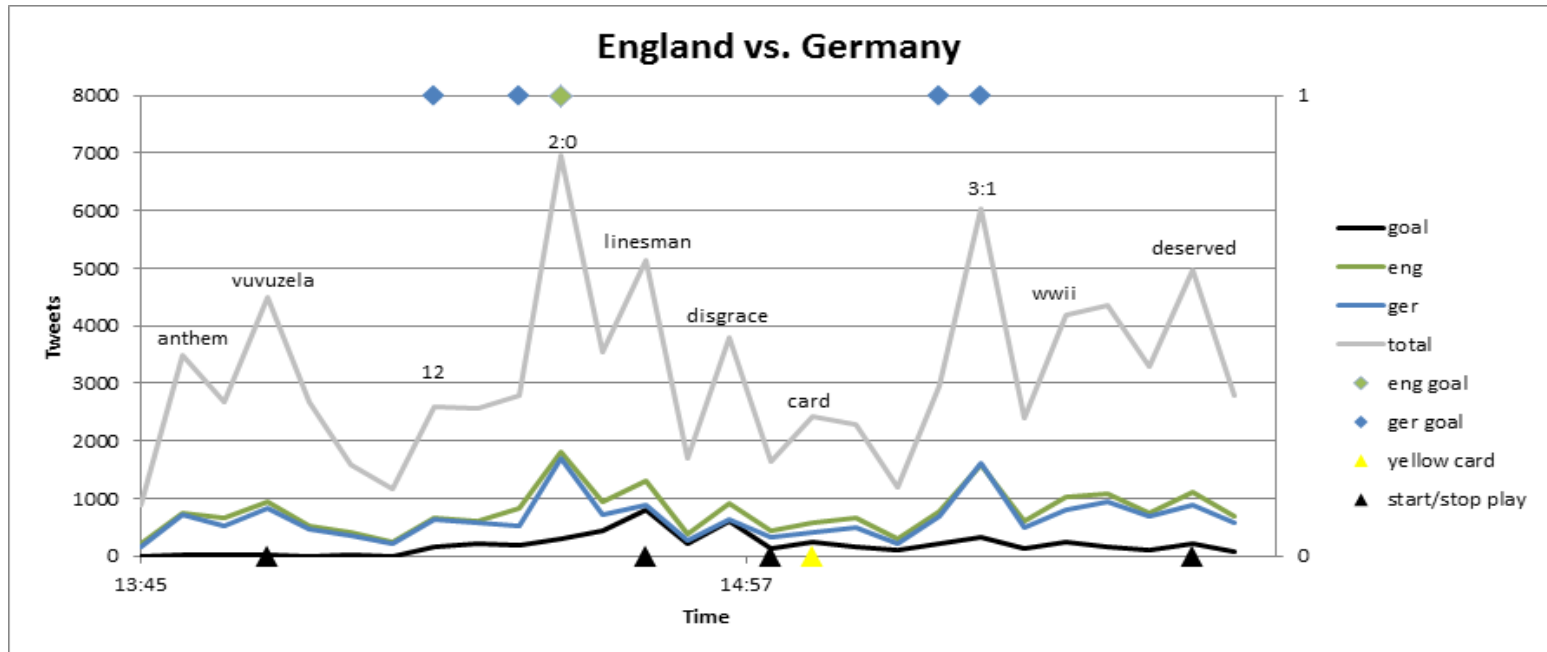
Slovakia vs. Italy

- Major upset
- Salient keywords: “goodbye”, “eliminated”



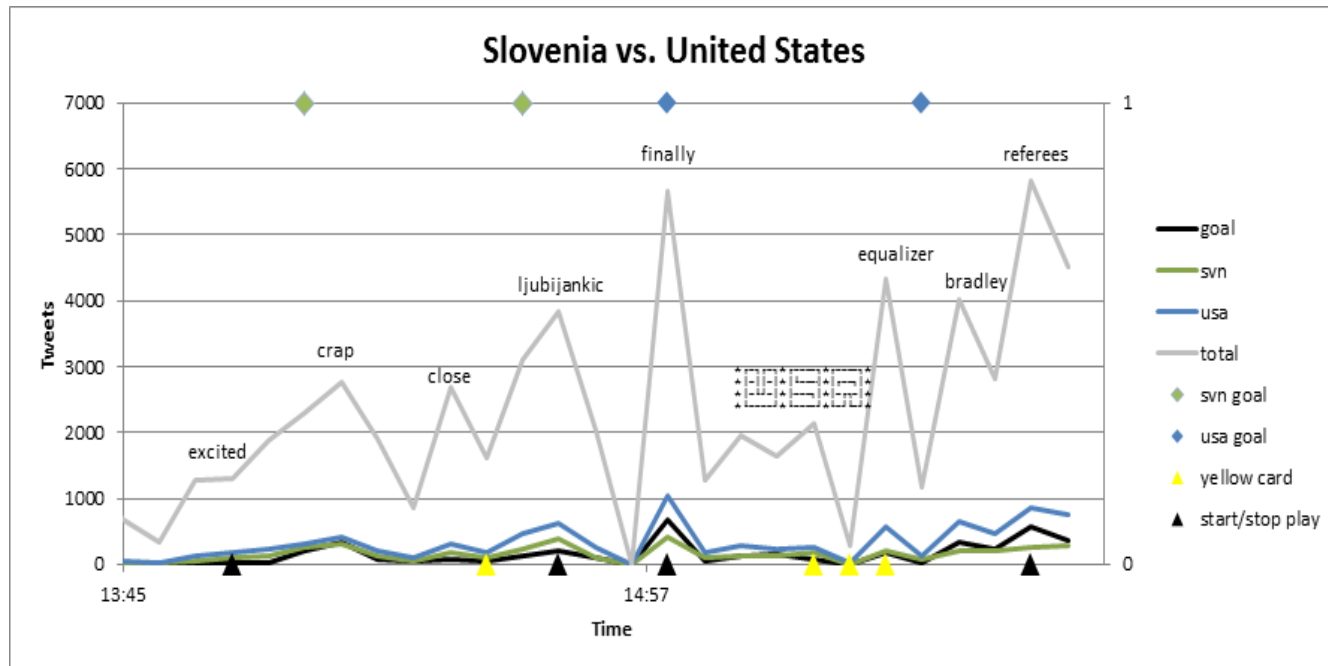
England vs. Germany

- Game rhythm
- #12 meaning



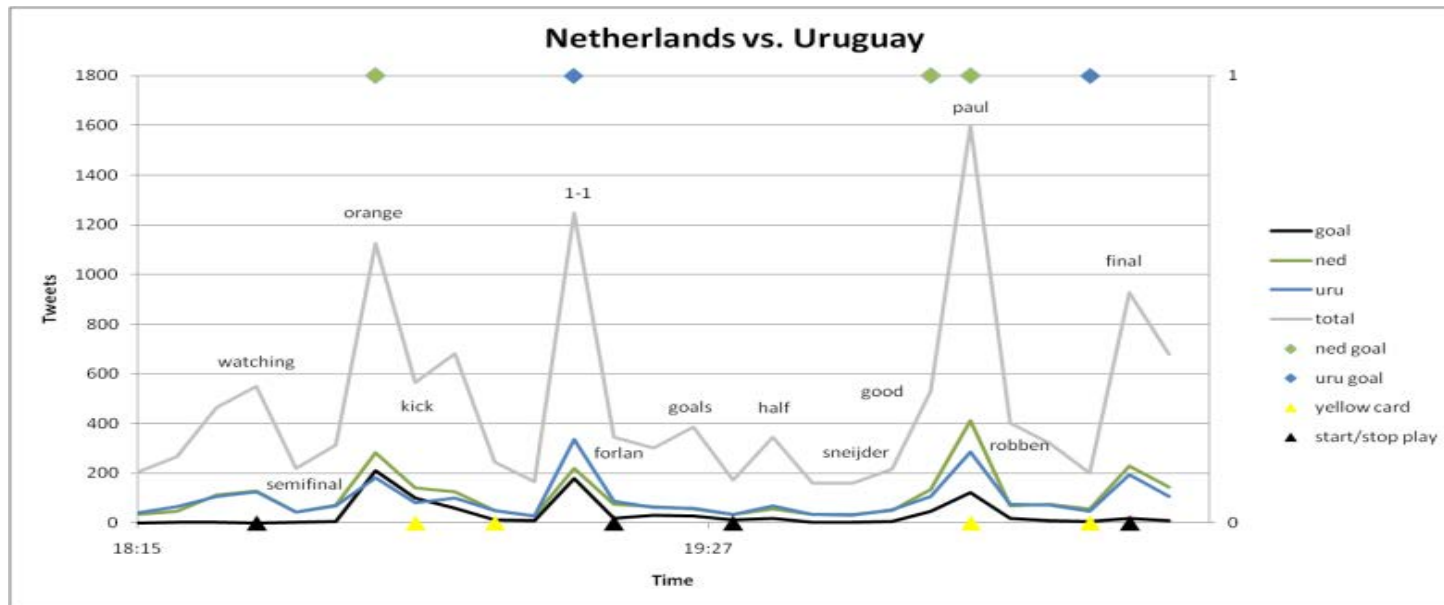
Slovenia vs. USA

- USA fans more active
- When things don't work, you blame the refs



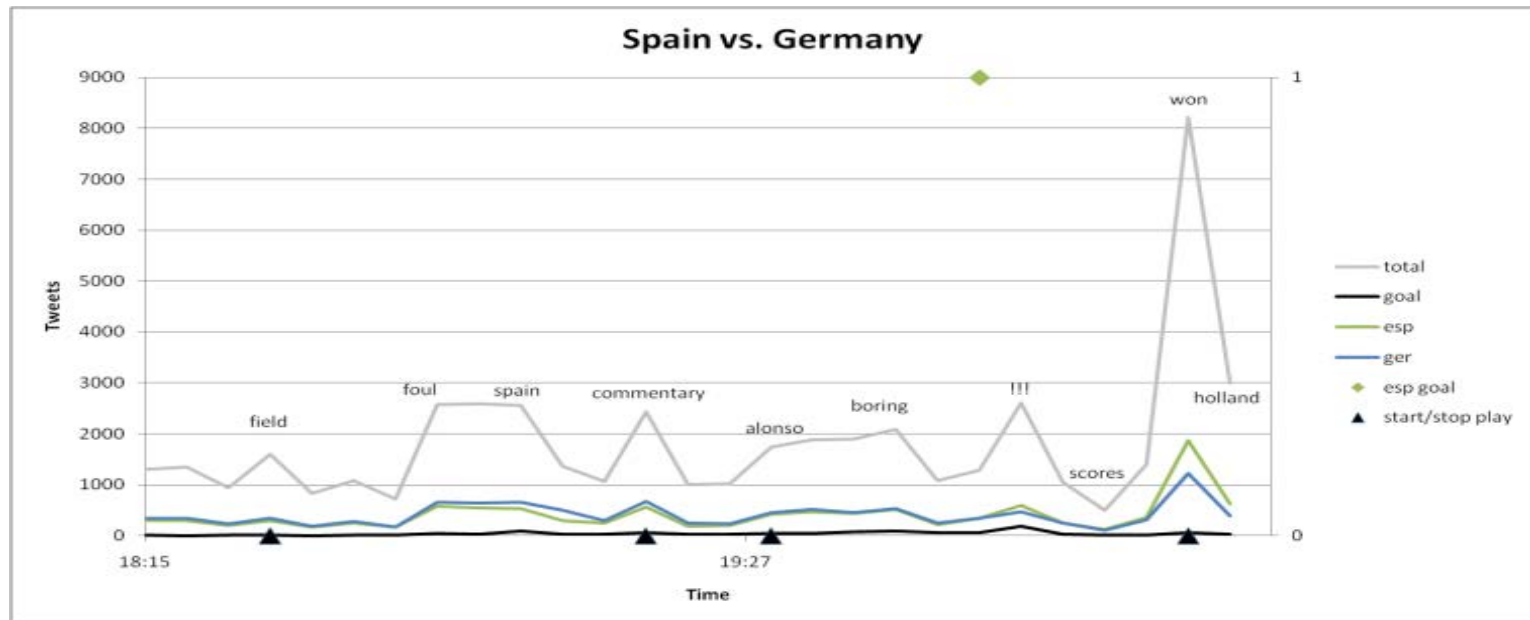
Netherlands vs. Uruguay

- Paul is the psychic Octopus
- Salient keywords: players, “final”



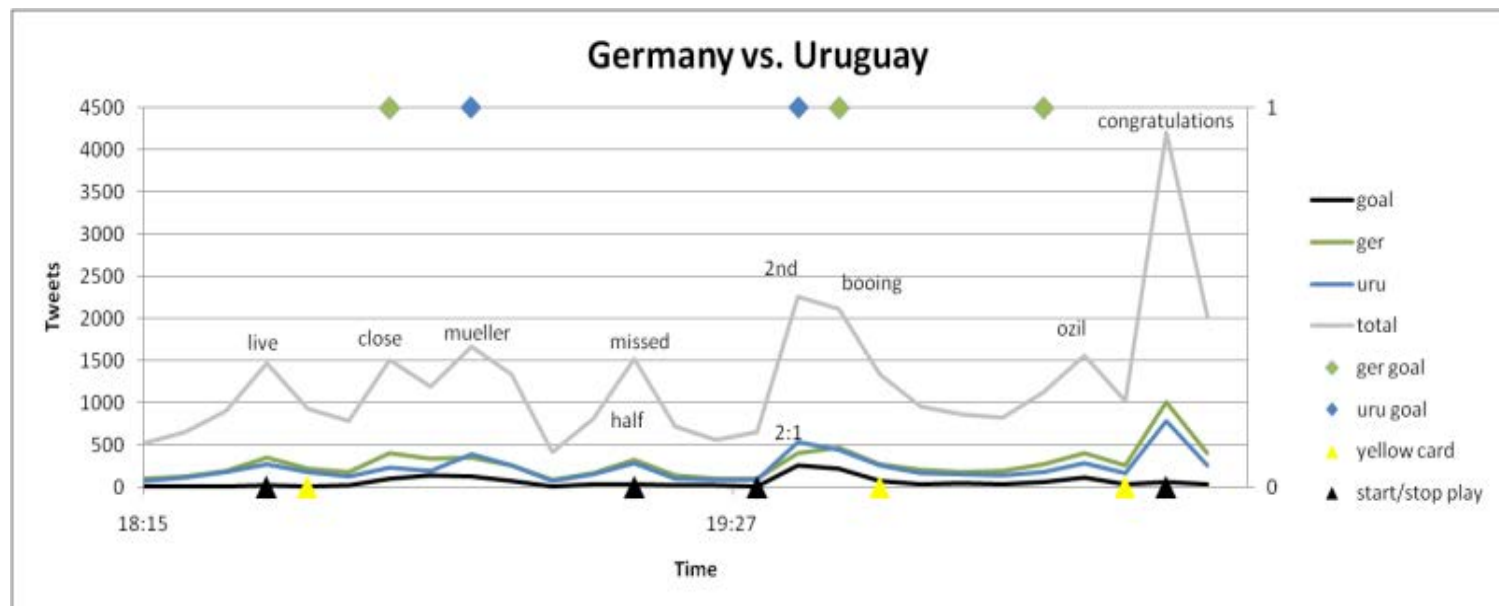
Spain vs. Germany

- Boring game
- Salient keywords: “won”, “holland”



Germany vs. Uruguay

- By now, players are well known
- Salient keywords: “booing”, “congratulations”



Evaluation

- BBC Expert editors

Timestamp	Comment
14:00	KICK-OFF Germany v England Germany and England's last-16 World Cup clash is under way in Bloemfontein. Here we go.
14:20	GOAL Germany 1-0 England That's among the worst England defending I've seen at a World Cup, I'm gobsmacked. Manuel Neuer 's goalkick sails over John Terry 's head and Miroslav Klose shows great strength and coolness to hold off Matthew Upson and poke past David James. Klose could have gone down there and Upson would have been sent off, but he bagged his 50th international goal instead.
15:14	Our friends at Infostrada Sports tell us that Germany have played 30 previous World Cup matches in which they took a 2-0 lead. They have won 29 of them and lost one: in the last 16 round in 1938 against Switzerland (2-4). That is 72 years ago.

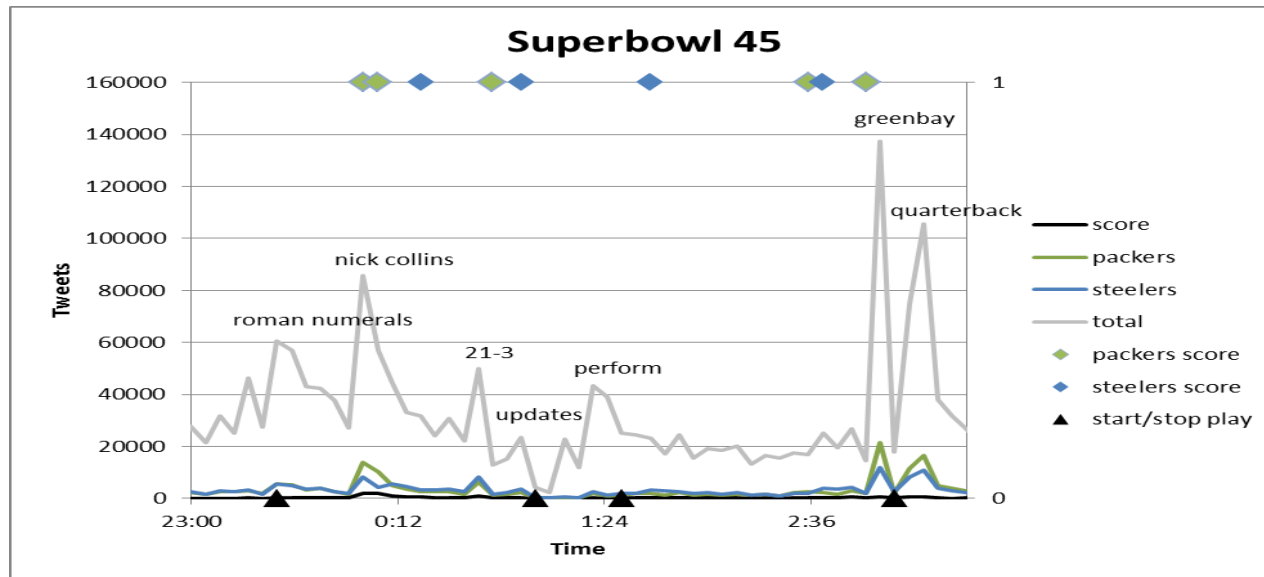
Evaluation

- Crowdsourcing with Mechanical Turk
 - What was the final score of the game?
 - Based on the summary, please rate the game on a scale 1 to 5 (1=boring, 5=exciting)
 - Is there anything in particular that you found interesting in the timeline summary?
 - Please rate how informative the timeline summary is on a scale 1 to 5 (1=irrelevant, 5=outstanding)
- 7 workers per HIT
- \$0.10 per assignment

Game	q1 (percentage of correct answers	q2	q4
#61	71%	4.14	3.14
#62	86%	2.86	3.43
#22	100%	4.43	3.86
#51	86%	4.14	3.71
#63	100%	3.71	4
#41	83%	3.5	3.67

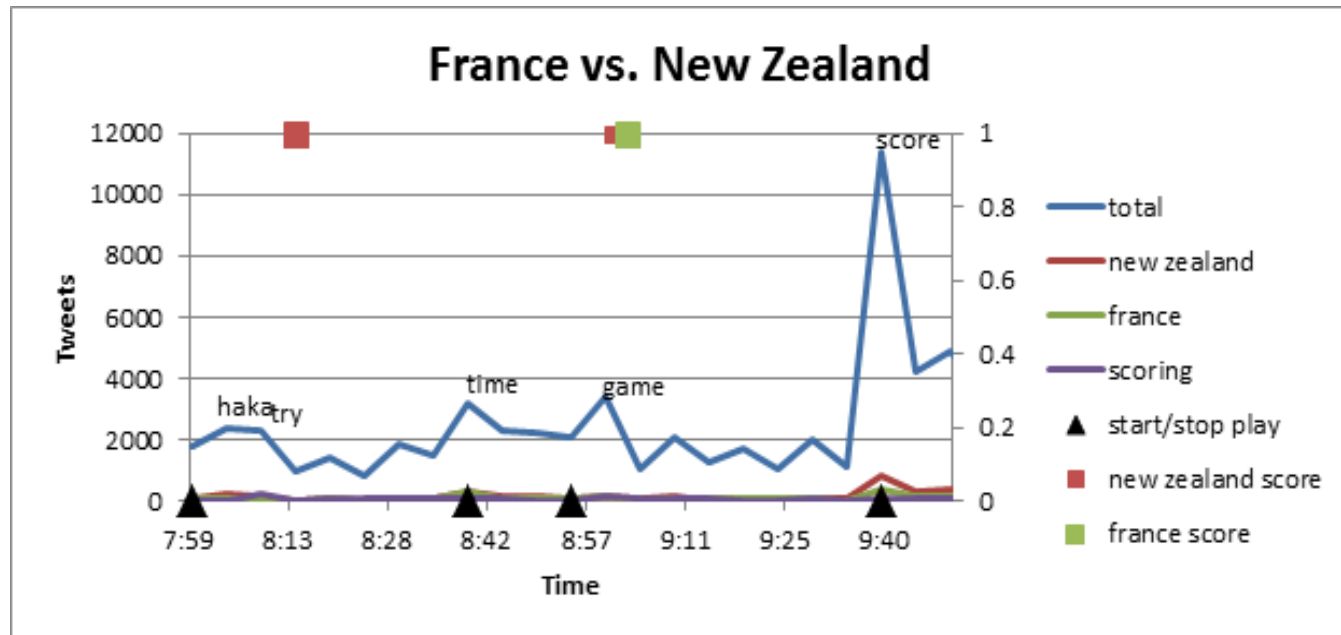
Generalization

- #superbowl



Generalization

- #rwc2011



Conclusions

- It is possible to follow games on Twitter
- Generate visual summaries that capture the salient aspects of the game
- Good evaluation results with the new presentation
- People found it very useful
- Next steps
 - Other sports
 - Product announcements

Thank you