The Prevalence of Political Discourse in Non-Political Blogs
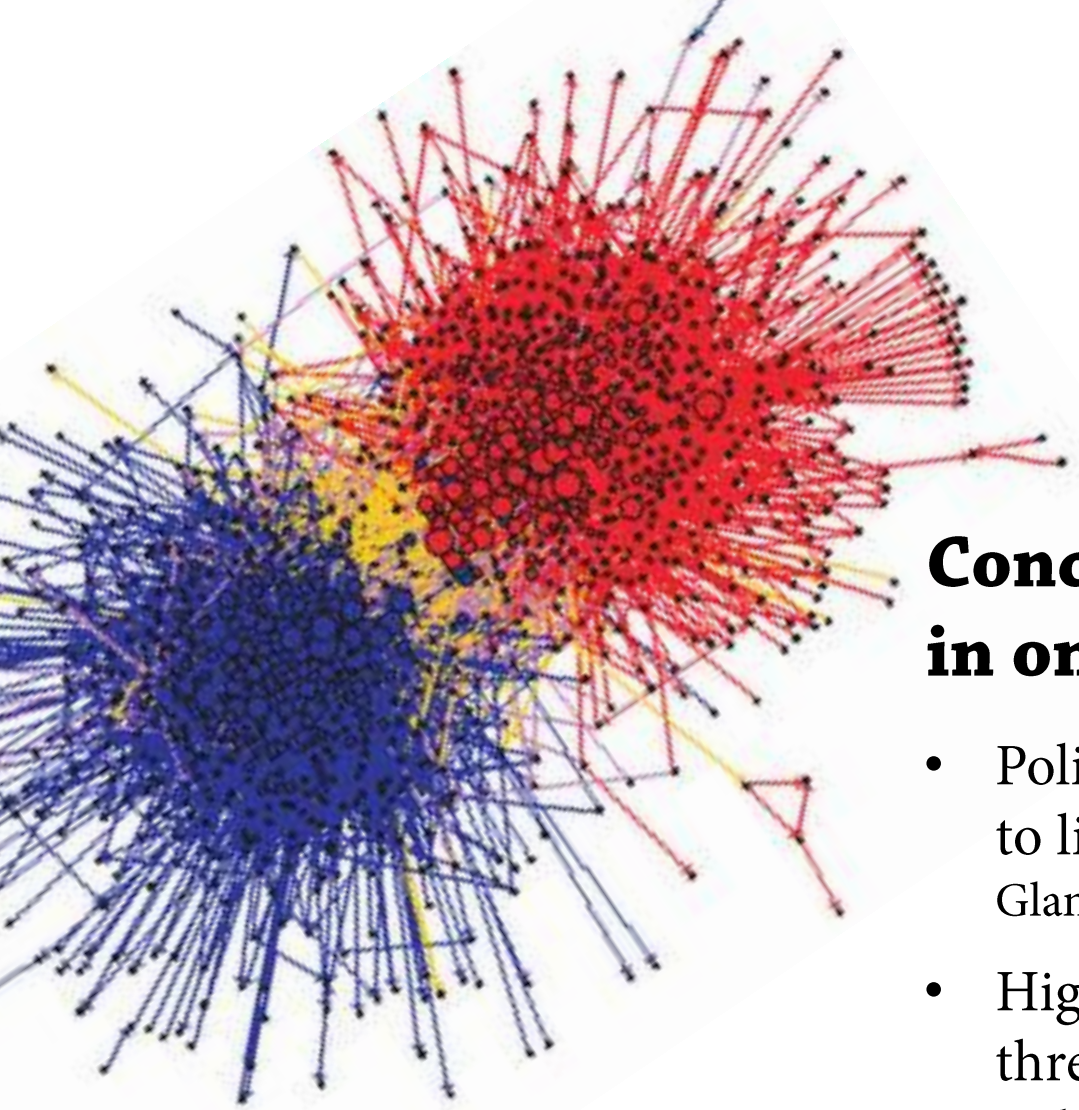
Sean A. Munson, Paul Resnick
School of Information, University of Michigan

MISC
Michigan Interactive & Social Computing

UNIVERSITY OF MICHIGAN

Adamic & Glance 2005

Concerns about homophily in online political discourse

- Political bloggers predominantly link to like-minded bloggers (Adamic & Glance 2005)

- High levels of agreement in comment threads on many blogs, including political blogs (Gilbert et al 2009)

- Challenge aversion prevalent in preferences for political news aggregators (Munson & Resnick 2010)

the White House

jmtimages

Mr. Wright

Akron Kiwanis

ladybugbkt

El Caganer

RobinJP

People who come together for other reasons

- may be more politically diverse

- listen and frame arguments to protect their relationships

… but they may also avoid political discussion or disagreement, also to protect relationships.

ladybugbkt

El Caganer

RobinJP

signs that politics is not taboo in non-political spaces online.

- Discussion of political candidates by Twitter users (Tumasjan et al 2010; Diakopoulos & Shamma 2010; Conover et al 2011)

- 8% of US adults posted political content to a social network site during 2010 midterm elections, 11% discovered for whom their friends voted (Smith/Pew 2011)

- People say they encounter cross-cutting political discussion online, but in non-political spaces (Wojcieszak & Mutz 2009)

questions for our study

- How prevalent are political blog posts on non-political blogs?

- What is the distribution of political blog posts across different categories of blogs?

- When readers of non-political blogs encounter political posts, do they treat them as taboo, or do they engage with the political content of the post?

sample of blogs

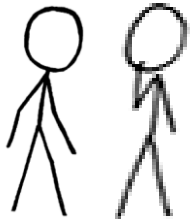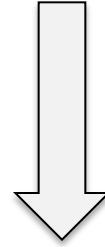Poll Blogger.com's recently updated list.   23,904 blogs

6-20 January 2008

Kept only blogs that:

- *had at least 5 posts,*

- *were written in English, and*

- *had existed since at least 31 August 2007.*

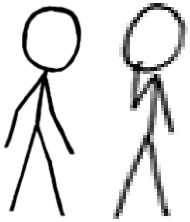Poll Blogger.com's recently updated list.   23,904 blogs
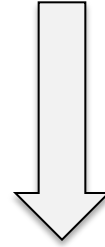
Research team inspects & eliminates spam blogs and blogs only partially in English.   8,861 blogs
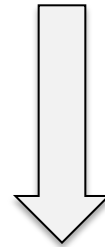
Poll Blogger.com's recently updated list. 23,904 blogs

Research team inspects & eliminates spam blogs and blogs only partially in English. 8,861 blogs

Code posts as political / nonpolitical, classify blogs by genre. Eliminate additional spam & partially English blogs. 8,765 blogs 2.3M posts

classifying posts as political or not

Broad definition of political: any mention of public policy, campaigns, and elected or appointed officials, and did not restrict this definition to US politics.

classifying posts as political or not

Broad definition of political: any mention of public policy, campaigns, and elected or appointed officials, and did not restrict this definition to US politics.

Research team categorized 6,691 posts as political or not. (oversampled political posts; $\kappa = 0.969$)
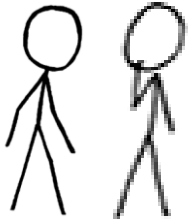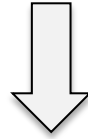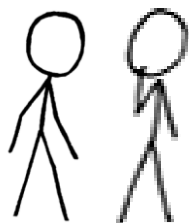
classifying posts as political or not

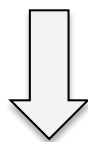Broad definition of political: any mention of public policy, campaigns, and elected or appointed officials, and did not restrict this definition to US politics.

Research team categorized 6,691 posts as political or not. (oversampled political posts; κ = 0.969)
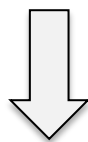
Used to train multinomial naïve Bayes classifier, classify remaining posts. (κ = 0.902)

217,727 political posts and 2,136,551 non-political posts.

estimating prevalence

Simple tally is problematic:

- overestimate percent political on blogs with few political posts (more opportunities for false positives)
- Underestimate political posts on blogs with many political posts (more opportunities for false negatives)

estimating prevalence

Simple tally is problematic:

- overestimate percent political on blogs with few political posts (more opportunities for false positives)
- Underestimate political posts on blogs with many political posts (more opportunities for false negatives)

Consider a blog that is *always* political:

Political posts

False negatives (12.6%)

Consider a blog that is *never* political:

Non-political posts

False positives (0.5%)

estimating prevalence

Simple tally is problematic:

- overestimate percent political on blogs with few political posts (more opportunities for false positives)
- Underestimate political posts on blogs with many political posts (more opportunities for false negatives)

When reporting about blogs or bins, we handle this with revised estimates ($p^\star$):

$$p^* = \frac{prevalence - (1 - specificity)}{sensitivity - (1 - specificity)}$$

(Zhou et al 2002)

categorizing blogs

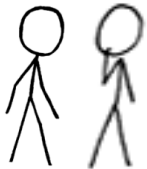Classify blogs into seven categories: diary, hobby & fan, professional & sales, politics, religion, civic & issue, health &wellness, and ethnic / cultural.

amazonmechanical turk
Artificial Artificial Intelligence
beta

+

get-another-label

At least 5 categorizations per blog, determine final category with get-another-label (Sheng, Provost, Ipeirotis 2008)

Check against ratings from research team (overall κ = 0.72)

coding comments

244 threads classified by research team.

	Decision	% agree	κ
Post	Had some political content?	100%	1.00
	Had other, non-political topics?	100%	1.00
Comment	Spam?	98%	0.87
	Engaged political content of post?	93%	0.79
	Agreed, disagreed, or neither with post's author?	91%	0.84
	Said the blog post's political content did not belong?	100%	1.00

Inter-rater reliability on 56 comments on 42 blog posts.

coding comments

244 threads classified by research team.

	Decision	% agree	κ
Post	Had some political content?	100%	1.00
	Had other, non-political topics?	100%	1.00
Comment	Spam?	98%	0.87
	Engaged political content of post?	93%	0.79
	Agreed, disagreed, or neither with post's author?	91%	0.84
	Said the blog post's political content did not belong?	100%	1.00

Inter-rater reliability on 56 comments on 42 blog posts.

coding comments

244 threads classified by research team.

	Decision	% agree	κ
Post	Had some political content?	100%	1.00
	Had other, non-political topics?	100%	1.00
Comment	Spam?	98%	0.87
	Engaged political content of post?	93%	0.79
	Agreed, disagreed, or neither with post's author?	91%	0.84
	Said the blog post's political content did not belong?	100%	1.00

Inter-rater reliability on 56 comments on 42 blog posts.

results

overall prevalence | prevalence by category | engagement

results
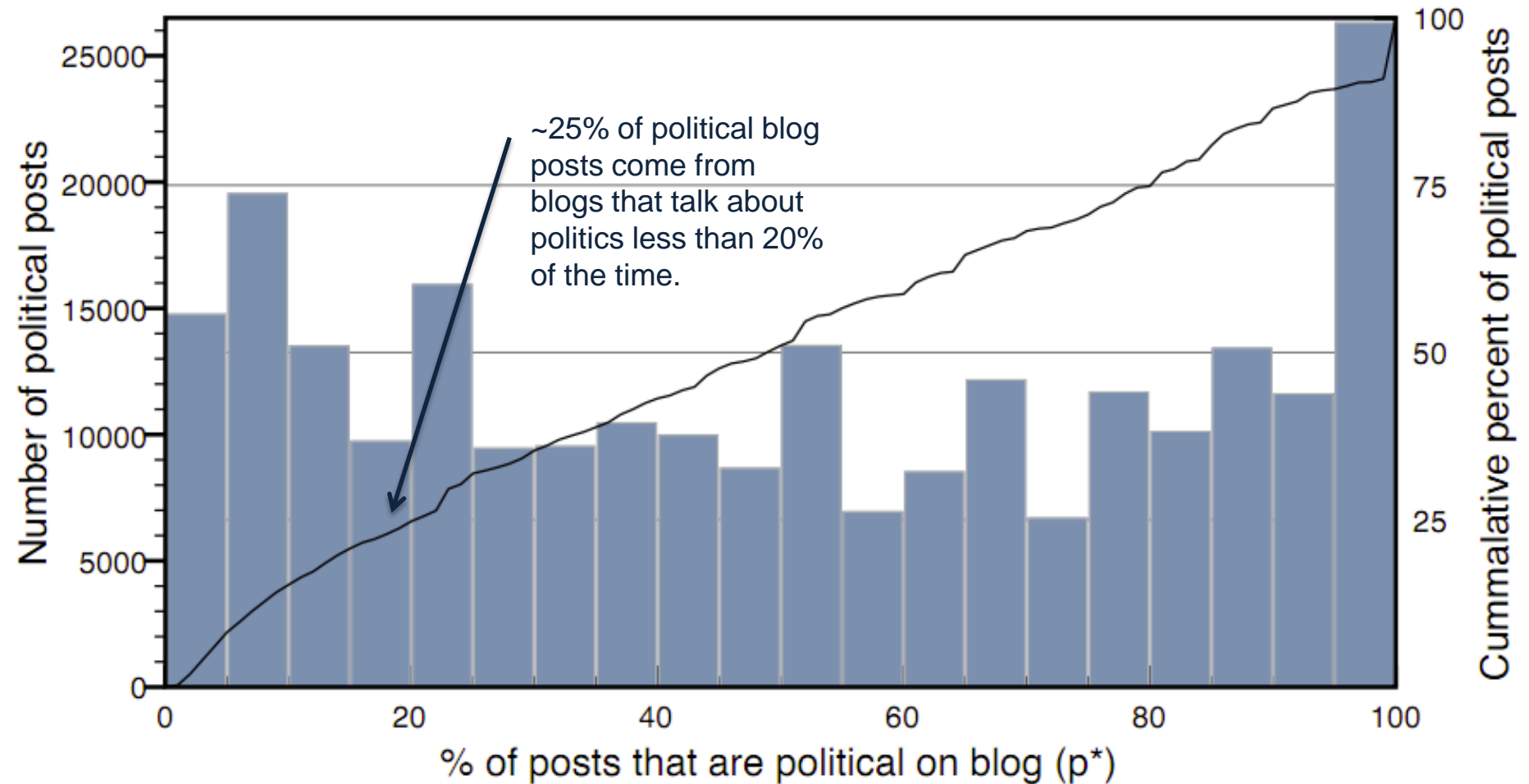
overall prevalence | prevalence by category | engagement
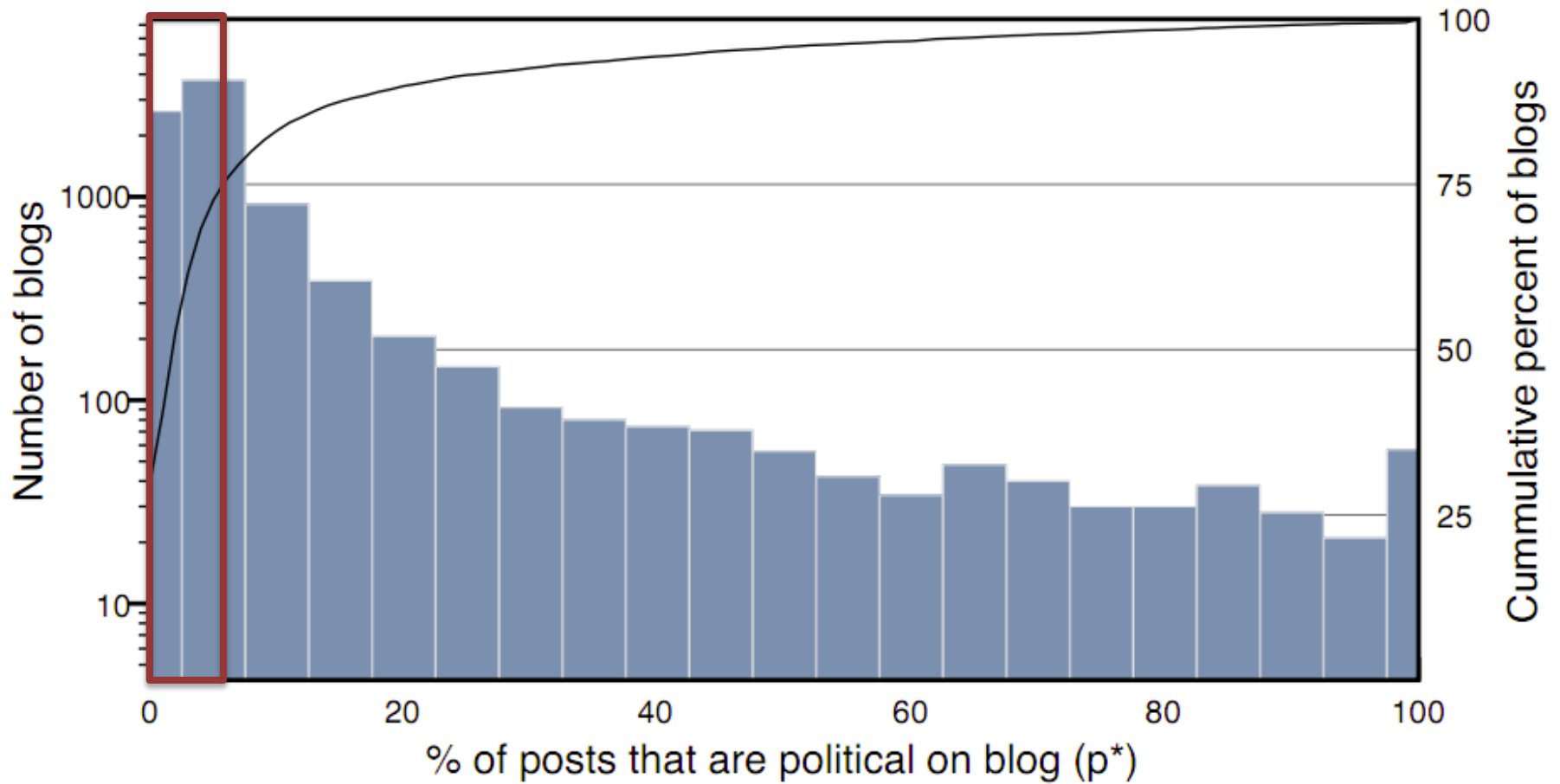
prevalence of political posts

prevalence of political posts

~25% of political blog posts come from blogs that talk about politics less than 20% of the time.

prevalence of political posts by day

Limited to posts from non-political blogs.

US election day

Start of US presidential primary season

perpol

results

categories: descriptions & summary statistics

| Category	Description	n	%	Post per week per blog Mean (stdev)	Comments per post Mean (stdev)
Diary	Individual, group, or family blog with news about life. Includes blogs that are collections of links, images, or videos that interested the author if the collection does not fit into one of the other categories.	5307	60.5%	3.62 (5.17)	2.33 (6.10)
Hobby & Fan	Blog about a particular hobby, interest, or activity (such as crafts, photography, programming, or cooking). Also includes blogs by enthusiasts of a particular TV show, celebrity, actor, movie, musical group, or sports team. Incudes travel and exercise diaries (e.g. someone who writes about running or cycling as a hobby).	2148	24.5%	5.81 (5.81)	2.20 (8.30)
Professional & sales	Blog for a trade, educational, or professional association, or containing news, tips, or advice for people in a particular career or line of work, or an official blog to promote a product, service, or event, to interact with customers, or to provide news about a business or other organization.	519	5.9%	40.1 (567.42)	2.20 (8.81)
Politics	Blog with commentary or news on issues or controversies in politics and government	422	4.8%	11.89 (18.47)	3.11 (12.88)
Religion	Blog by/about religious organizations, daily devotionals, or meditations. Does not include life diaries by people for whom religion is a big part of their life.	200	2.3%	4.06 (4.49)	2.06 (5.75)
Civic & issue	Blog that promotes a *particular* social or political change, such as an environmental organization	81	0.9%	7.85 (13.12)	1.35 (3.71)
Health & Wellness	Blog with tips, suggestions, support, or advice for health and/or wellness. Includes patient diaries and blogs with advice about exercise for health.	66	0.8%	3.75 (3.87)	2.16 (5.91)
Ethnic / cultural	Blog about a particular culture or heritage.	22	0.3%	2.27 (5.54)	1.35 (2.92)

categories: descriptions & summary statistics

| Category	Description	n	%	Post per week per blog Mean (stdev)	Comments per post Mean (stdev)
Diary	Individual, group, or family blog with news about life. Includes blogs that are collections of links, images, or videos that interested the author if the collection does not fit into one of the other categories.	5307	60.5%	3.62 (5.17)	2.33 (6.10)
Hobby & Fan	Blog about a particular hobby, interest, or activity (such as crafts, photography, programming, or cooking). Also includes blogs by enthusiasts of a particular TV show, celebrity, actor, movie, musical group, or sports team. Incudes travel and exercise diaries (e.g. someone who writes about running or cycling as a hobby).	2148	24.5%	5.81 (5.81)	2.20 (8.30)
Professional & sales	Blog for a trade, educational, or professional association, or containing news, tips, or advice for people in a particular career or line of work, or an official blog to promote a product, service, or event, to interact with customers, or to provide news about a business or other organization.	519	5.9%	40.1 (567.42)	2.20 (8.81)
Politics	Blog with commentary or news on issues or controversies in politics and government	422	4.8%	11.89 (18.47)	3.11 (12.88)
Religion	Blog by/about religious organizations, daily devotionals, or meditations. Does not include life diaries by people for whom religion is a big part of their life.	200	2.3%	4.06 (4.49)	2.06 (5.75)
Civic & issue	Blog that promotes a *particular* social or political change, such as an environmental organization	81	0.9%	7.85 (13.12)	1.35 (3.71)
Health & Wellness	Blog with tips, suggestions, support, or advice for health and/or wellness. Includes patient diaries and blogs with advice about exercise for health.	66	0.8%	3.75 (3.87)	2.16 (5.91)
Ethnic / cultural	Blog about a particular culture or heritage.	22	0.3%	2.27 (5.54)	1.35 (2.92)

categories: descriptions & summary statistics

Category	Description	n	%	Post per week per blog Mean (stdev)	Comments per post Mean (stdev)
Diary	Individual, group, or family blog with news about life. Includes blogs that are collections of links, images, or videos that interested the author if the collection does not fit into one of the other categories.	5307	60.5%	3.62 (5.17)	2.33 (6.10)
Hobby & Fan	Blog about a particular hobby, interest, or activity (such as crafts, photography, programming, or cooking). Also includes blogs by enthusiasts of a particular TV show, celebrity, actor, movie, musical group, or sports team. Incudes travel and exercise diaries (e.g. someone who writes about running or cycling as a hobby).	2148	24.5%	5.81 (5.81)	2.20 (8.30)
Professional & sales	Blog for a trade, educational, or professional association, or containing news, tips, or advice for people in a particular career or line of work, or an official blog to promote a product, service, or event, to interact with customers, or to provide news about a business or other organization.	519	5.9%	40.1 (567.42)	2.20 (8.81)
Politics	Blog with commentary or news on issues or controversies in politics and government	422	4.8%	11.89 (18.47)	3.11 (12.88)
Religion	Blog by/about religious organizations, daily devotionals, or meditations. Does not include life diaries by people for whom religion is a big part of their life.	200	2.3%	4.06 (4.49)	2.06 (5.75)
Civic & issue	Blog that promotes a *particular* social or political change, such as an environmental organization	81	0.9%	7.85 (13.12)	1.35 (3.71)
Health & Wellness	Blog with tips, suggestions, support, or advice for health and/or wellness. Includes patient diaries and blogs with advice about exercise for health.	66	0.8%	3.75 (3.87)	2.16 (5.91)
Ethnic / cultural	Blog about a particular culture or heritage.	22	0.3%	2.27 (5.54)	1.35 (2.92)

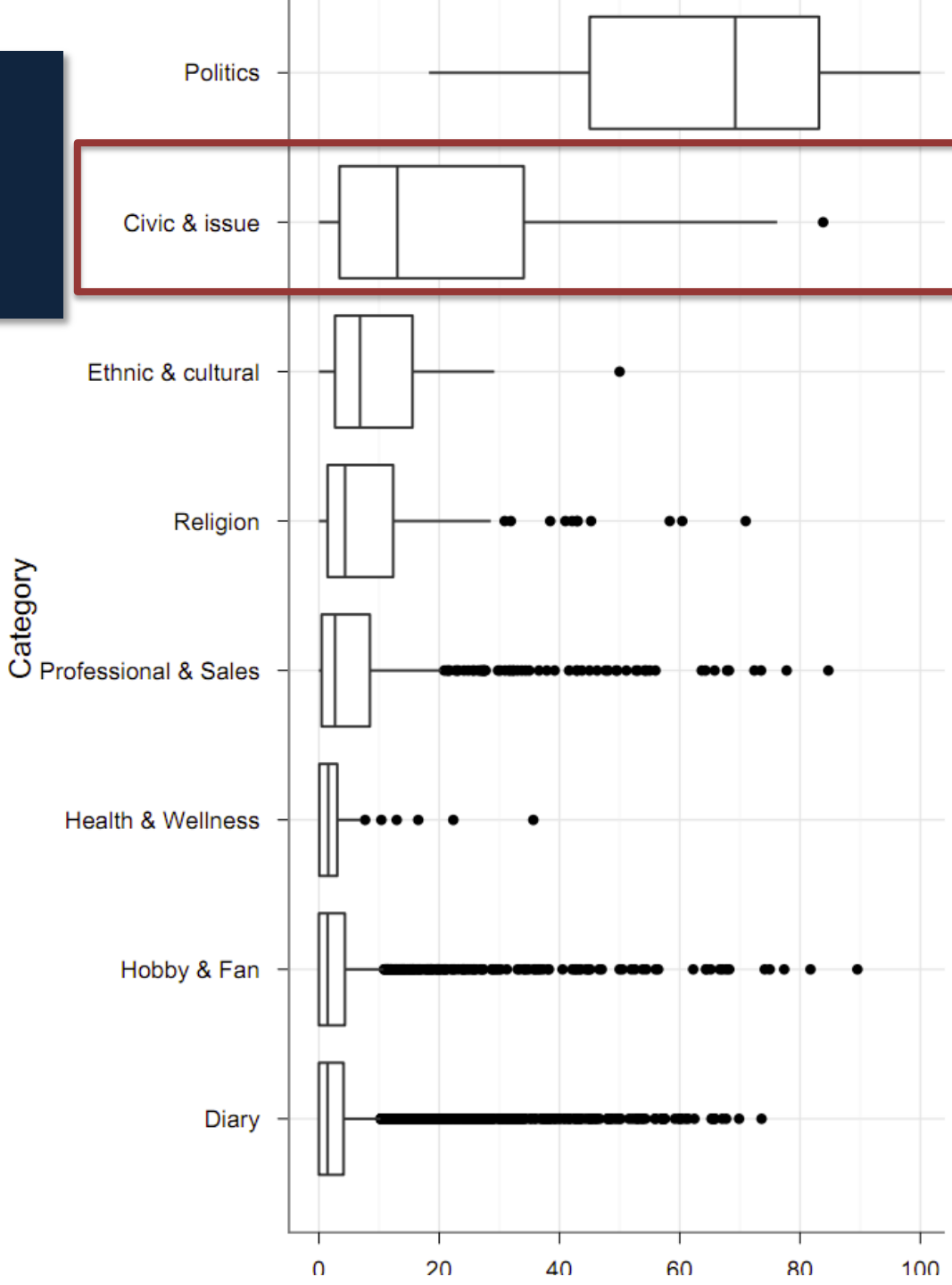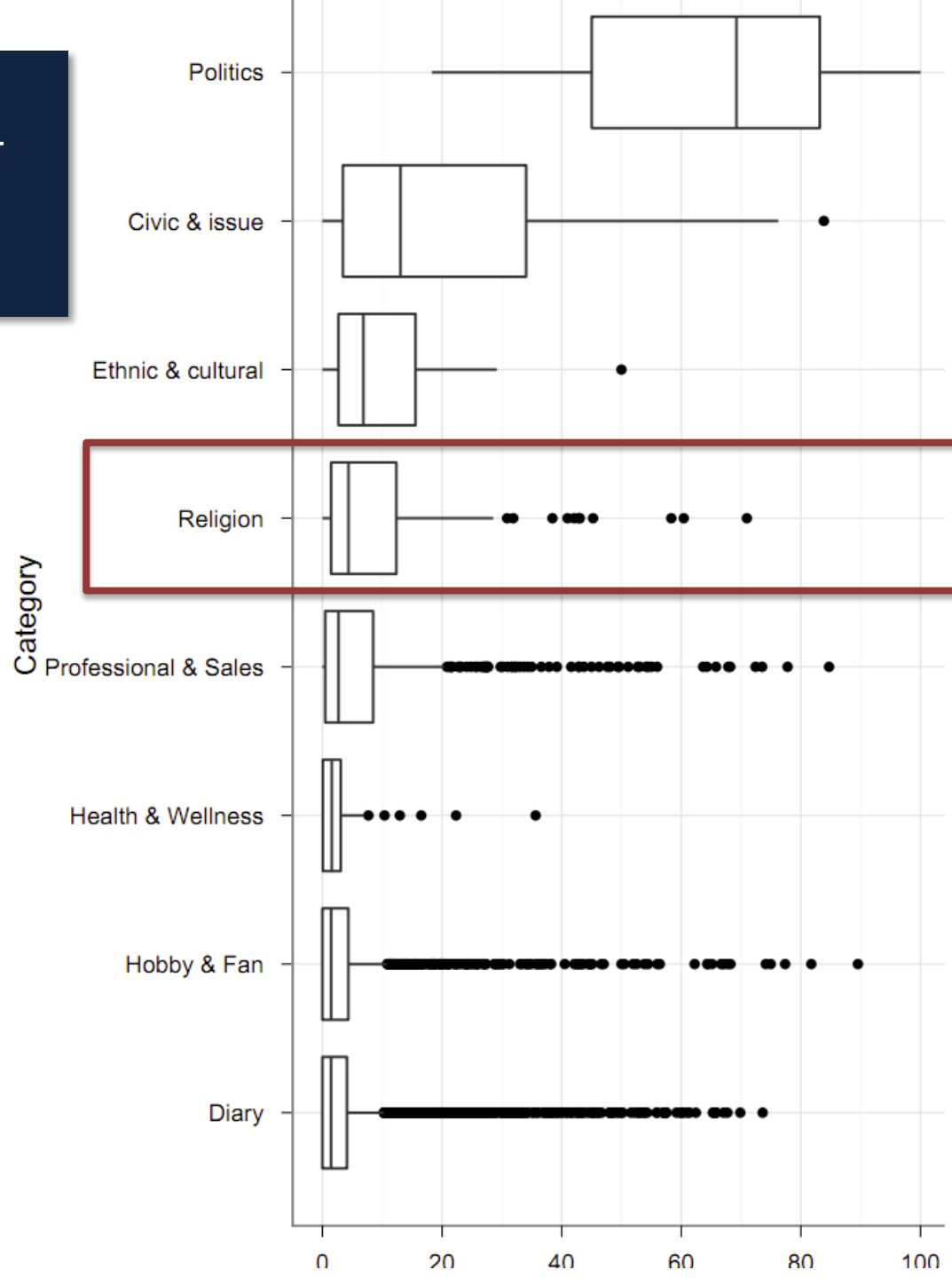# prevalence of political posts by category
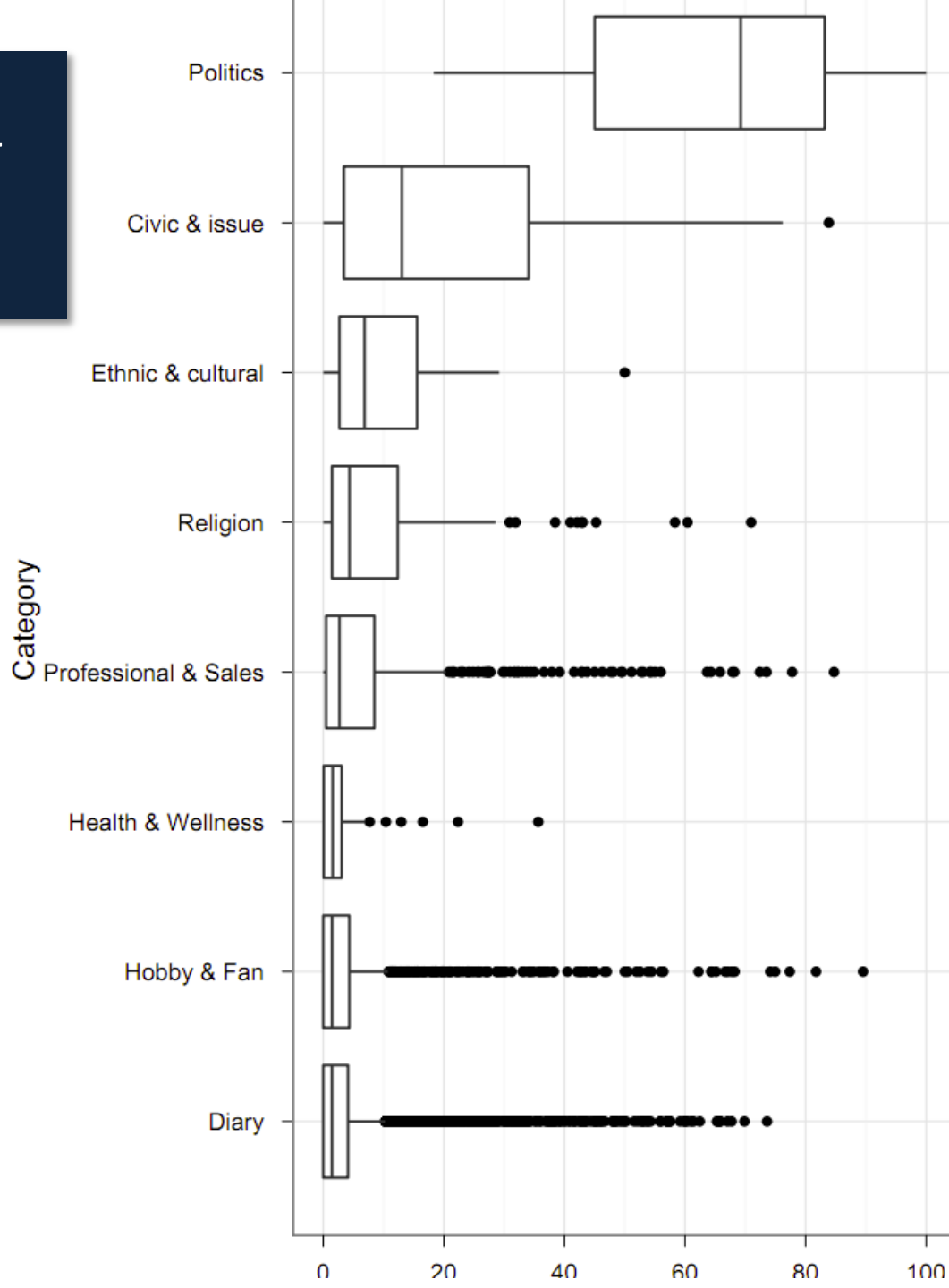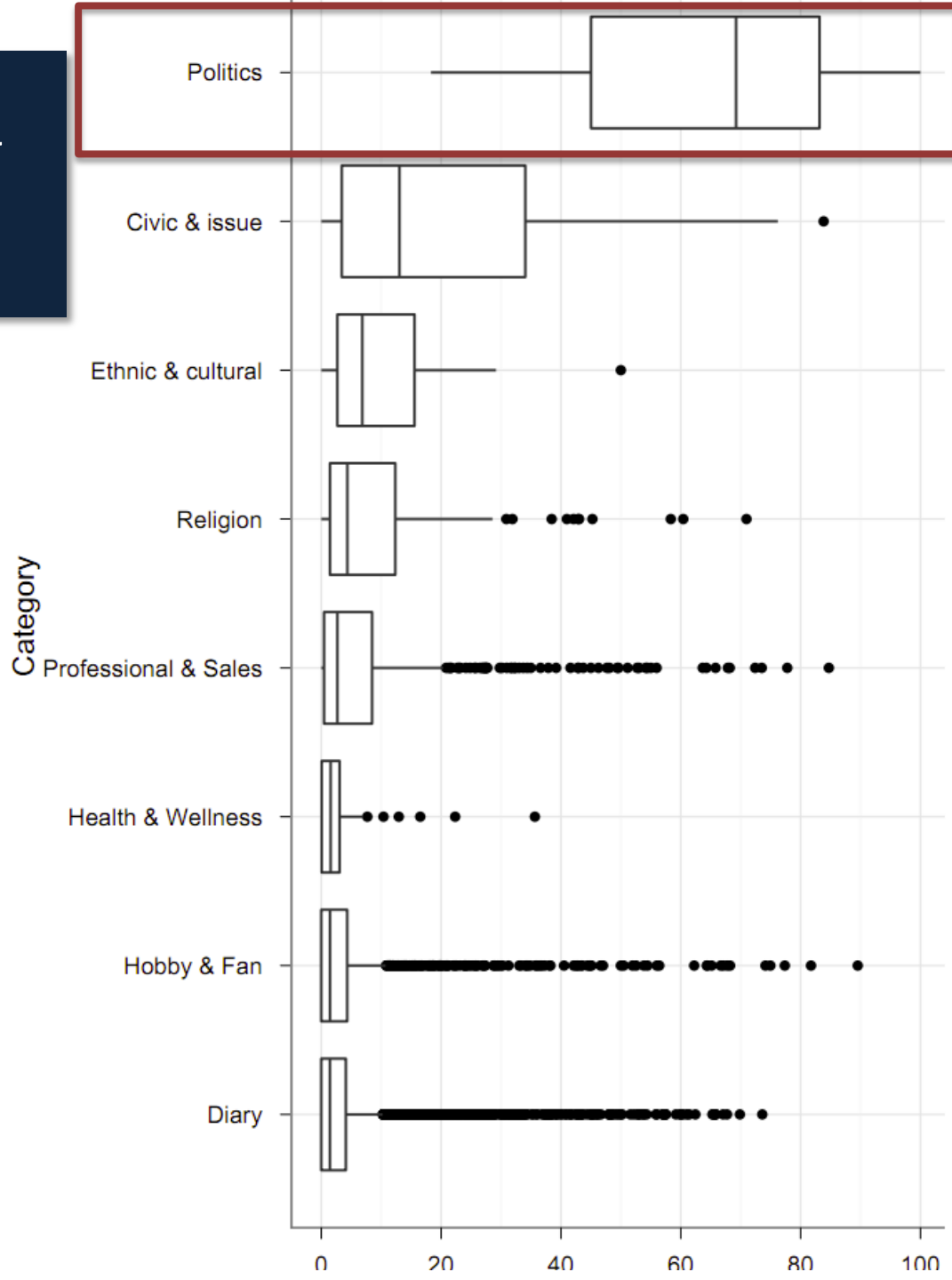
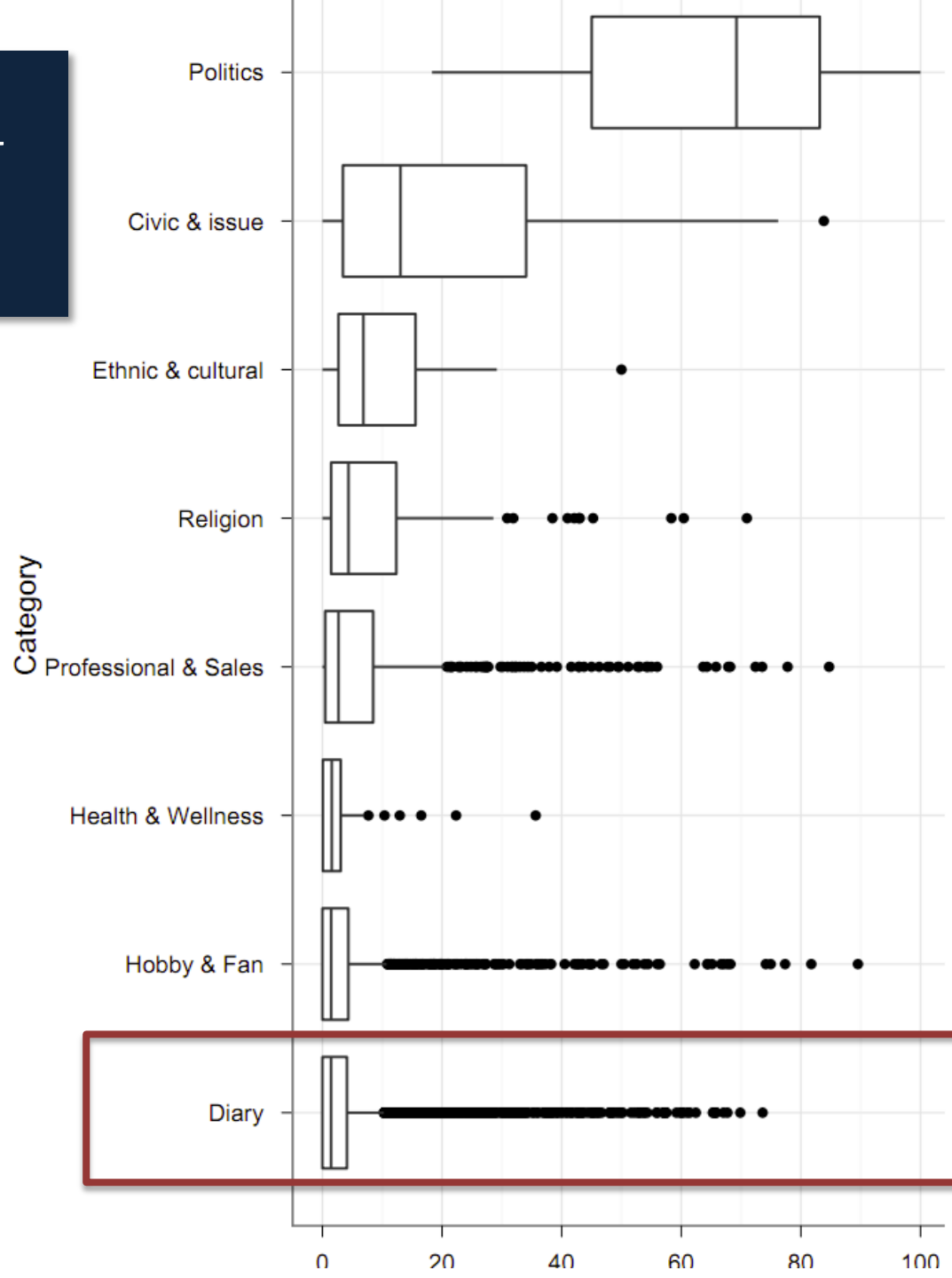# prevalence of political posts by category

prevalence of political posts by category

prevalence of political posts by category

prevalence of political posts by category

results

overall prevalence | prevalence by category | engagement

mean comments per post

	n	Mean comments on	
		Political Posts	**Nonpolitical posts**
Non-political blogs	6181	2.506	2.250
Political blogs	414	3.489	2.675

mean comments per post

	n	Mean comments on	
		Political Posts	**Nonpolitical posts**
Non-political blogs	6181	2.506	2.250
Political blogs	414	3.489	2.675

expected comments per post

	β	**Std Error**	*p*-value
Intercept	-0.012	0.0015	<0.001
Mean comments for blog	1.000	0.0001	<0.001
Political post	0.207	0.0298	<0.001
Political blog	-0.158	0.0468	<0.001
Political post * political blog	-0.138	0.0777	0.075

$F(4, 8764) = 3.862e{+}05$ ($p<0.0001$); adjusted $R^2 = 0.3962$.

engagement & agreement

Among 990 comments on 244 political posts:

	n	%
Engaged political content of post	823	83%

Among comments engaging with political content

Expressed		n	%
	agreement	334	41%
	neither / balanced	370	45%
	disagreement	119	14%

engagement & agreement

Among 990 comments on 244 political posts:

	n	%
Engaged political content of post	823	83%

Among comments engaging with political content			
Expressed	agreement	334	41%
	neither / balanced	370	45%
	disagreement	119	14%

this is a start. (and a challenge.)
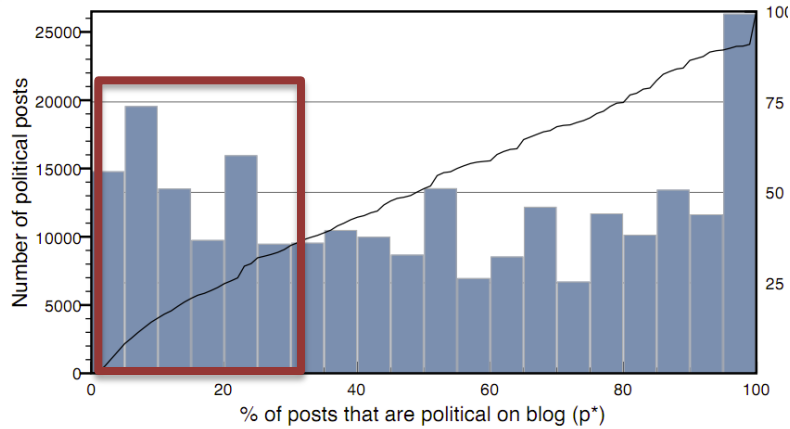
There's a lot of political talk happening on non-political blogs.
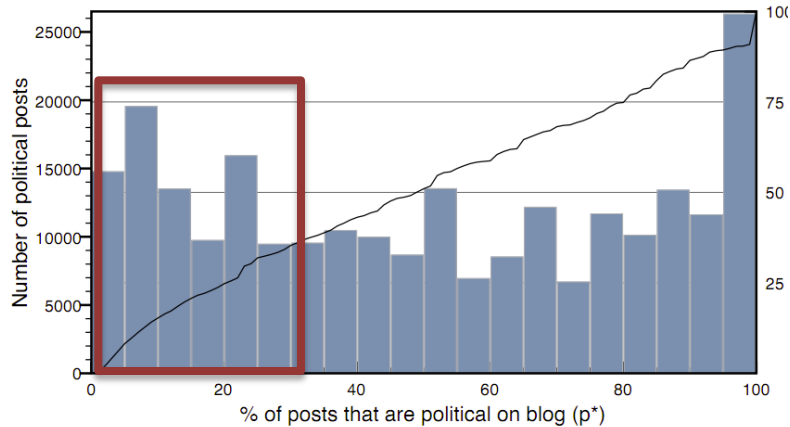
There's a lot of political talk happening on non-political blogs.

Not taboo: readers engage with this political content in replies.
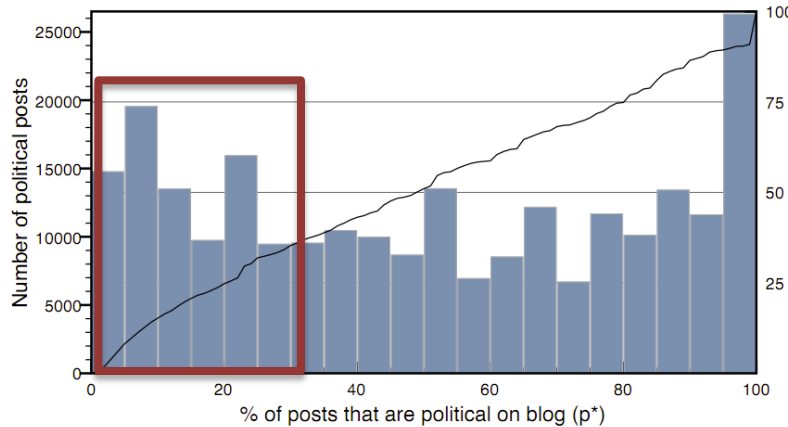
this is a start. (and a challenge.)

There's a lot of political talk happening on non-political blogs.

Not taboo: readers engage with this political content in replies.

What is the actual discourse quality? Civility? Arguments expressed?

How does this stack up against other spaces?

The Prevalence of Political Discourse in Non-Political Blogs

Sean A. Munson   @smunson   samunson@umich.edu

Paul Resnick   @presnick   presnick@umich.edu

balance.projects.si.umich.edu

balance

MISC
Michigan Interactive & Social Computing

UNIVERSITY OF MICHIGAN