



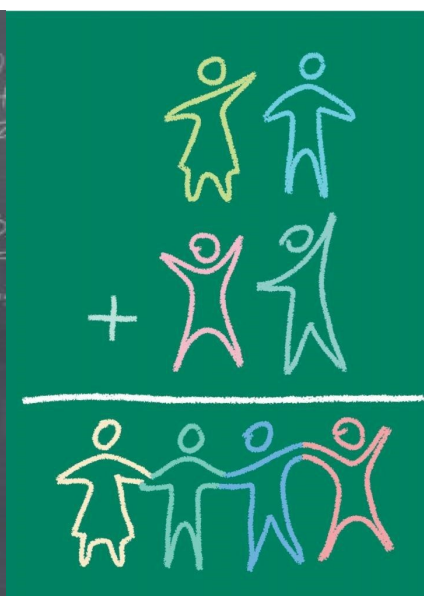
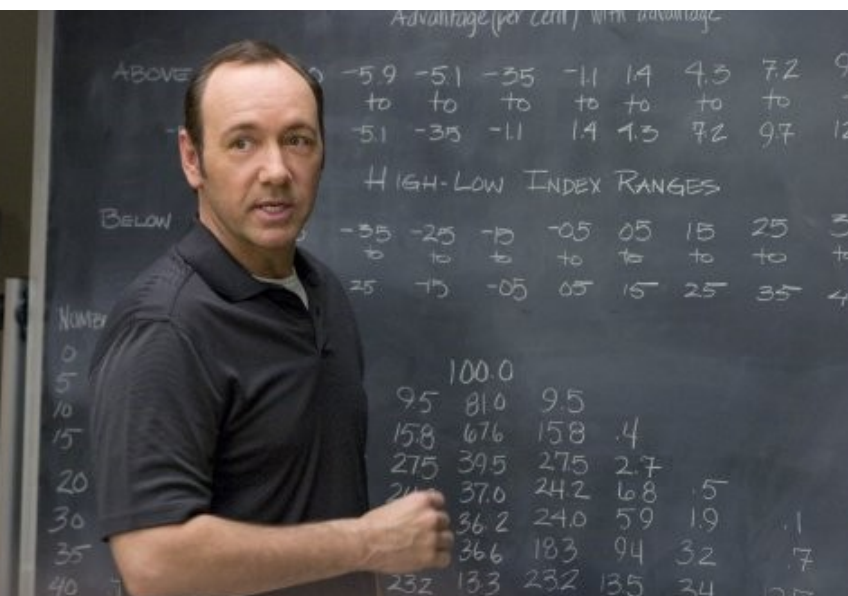
# Wikistats: New Patterns (and some Old Metrics Nuanced)

Annotated Edition (presented at Wikimania London on Aug 9, 2014)

<http://infodisiac.com/blog>

Erik Zachte CC-BY-SA-4.0

[ezachte@wikimedia.org](mailto:ezachte@wikimedia.org)



# Erik Zachte

- Volunteer since 2002
- WMF Data Analyst since 2008

Any views expressed are personal

# Topics

Metrics for public and press

Editor Trends

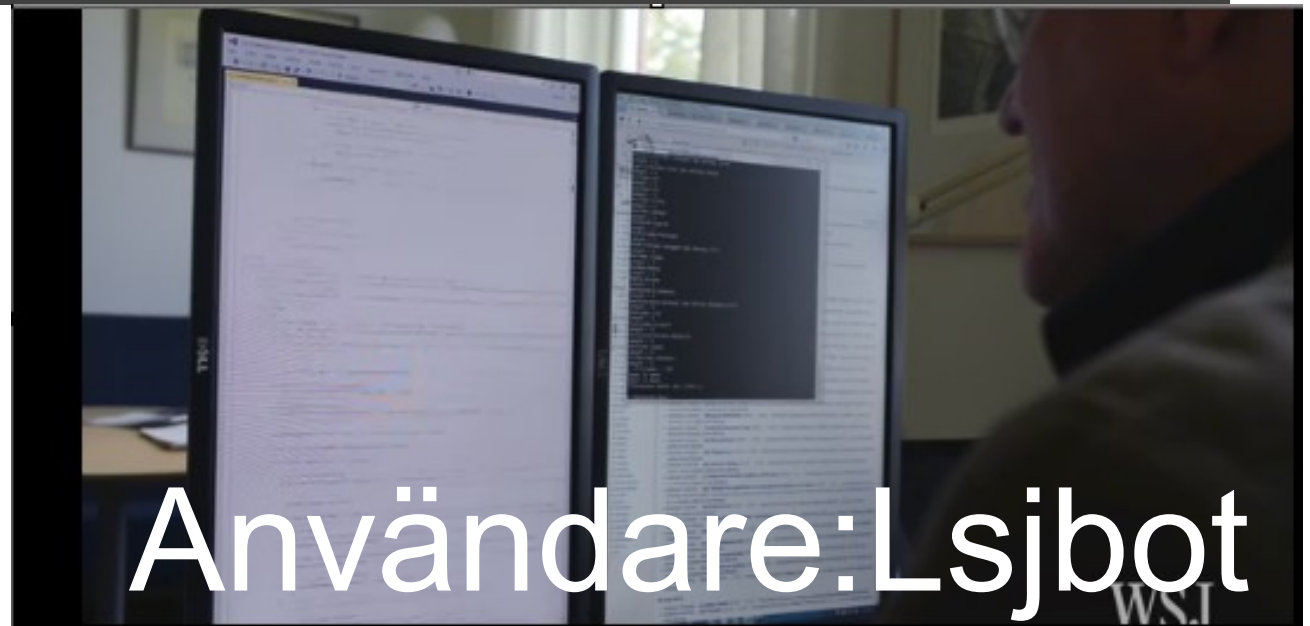
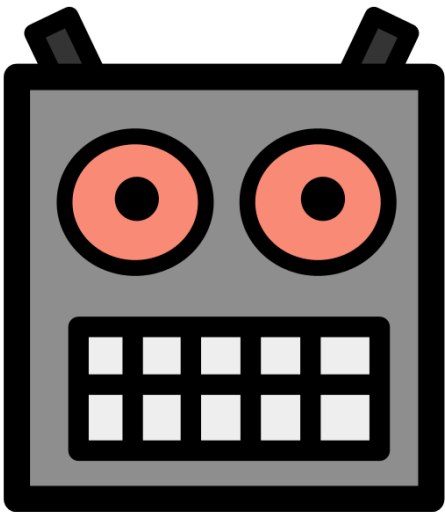
# Metrics for public and press

- 32 million Wikipedia articles
- 5<sup>th</sup> web site (comScore)
- Wikipedia in ~284 languages

# Reality Check 1: Articles

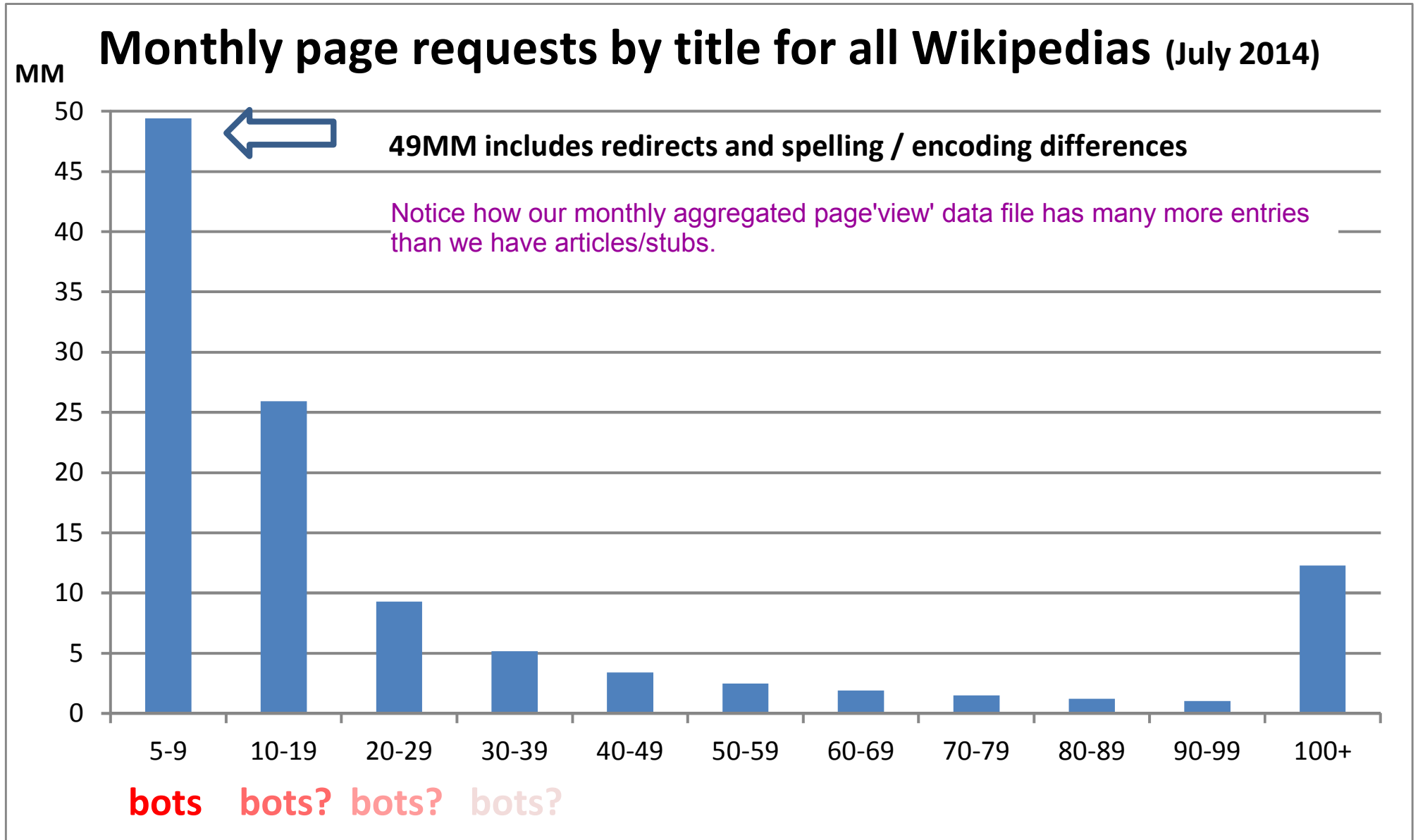
32 M 'articles' → lemmas  
(many are 'proto articles' aka stubs)

THE WALL STREET JOURNAL.



There has been substantial coverage on our article counts in the press lately (WSJ, but also Washington Post). Particularly the large share of bot created articles. Many of those articles are not full articles yet. They serve a purpose, but the public may have a different expectation when they hear '32 million articles'.

# Reality Check 1: Articles 'viewed' (?)



Can we use page views stats to count articles/stubs which are really *viewed* (= requested by humans)? Not yet. Our upcoming hadoop infrastructure will help us filter out bot requests, and add views from mobile devices.

# Reality Check 2: comScore rank

5<sup>th</sup> web site in U<sub>nique</sub> V<sub>isitors</sub> (comScore)

27<sup>th</sup> web site in P<sub>age</sub> V<sub>iews</sub> (comScore)

27<sup>th</sup> in May 2014, down from 20<sup>th</sup> a year earlier

After Facebook, Google, Tencent, Yahoo, Alibaba, Baidu, Odnoklassiki, Microsoft, Qihoo, 360.cn, eBay, Mail.ru, Yandex, Sohu, Sina, Ifeng, Amazon, AOL, PCPOP, NetEase, Schibsted, Youko&Tudou, KuGou, UOL, Craigslist, NHN Corp

Wikimedia is 5<sup>th</sup> and 27<sup>th</sup> web property, according to same web research firm. It depends on which metric one uses. Both are common metrics for comparison of web properties. 27<sup>th</sup> is **not** truer than 5<sup>th</sup>, just a different perspective. And both can be nuanced further, see next sheet.



# comScore ranks nuanced

Our privacy policy forbids web beacons, which hampers comScore's accuracy

comScore excludes visitors with age below 15

This may affect Wikimedia more than other top web properties

comScore U<sub>nique</sub>V<sub>isitors</sub>

- double counts? (multiple devices, browsers/apps per user)
- doesn't include much mobile traffic

ComScore presents mobile traffic in separate system, which we don't use

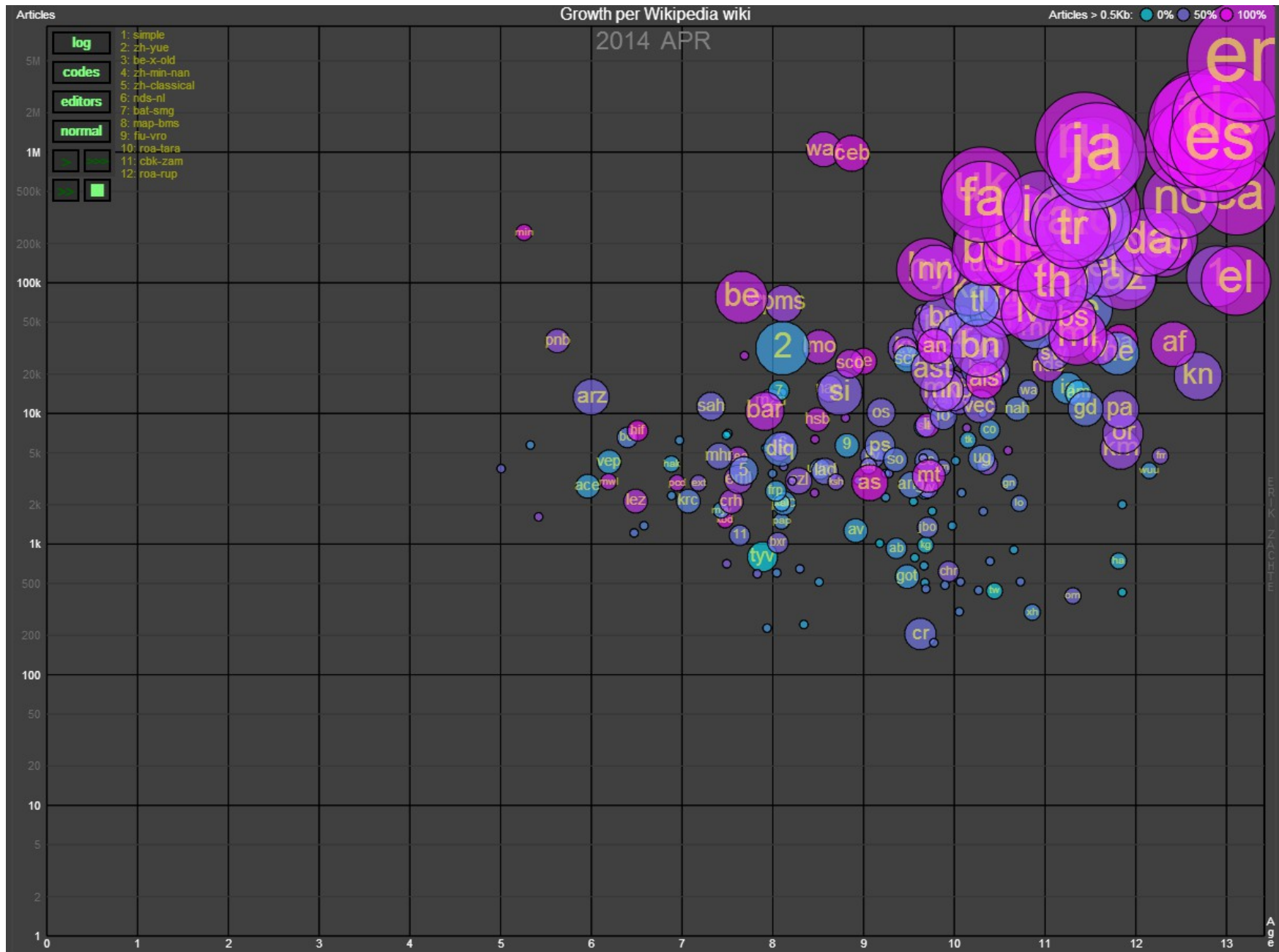
comScore P<sub>age</sub>V<sub>iews</sub>

- we count 3x as much, but
- we don't filter crawler traffic yet (20% ?)

Thanks much to comScore for providing WMF with free subscription to MyMetrix data!  
Several factors need to be taken into account to interpret comScore's numbers.

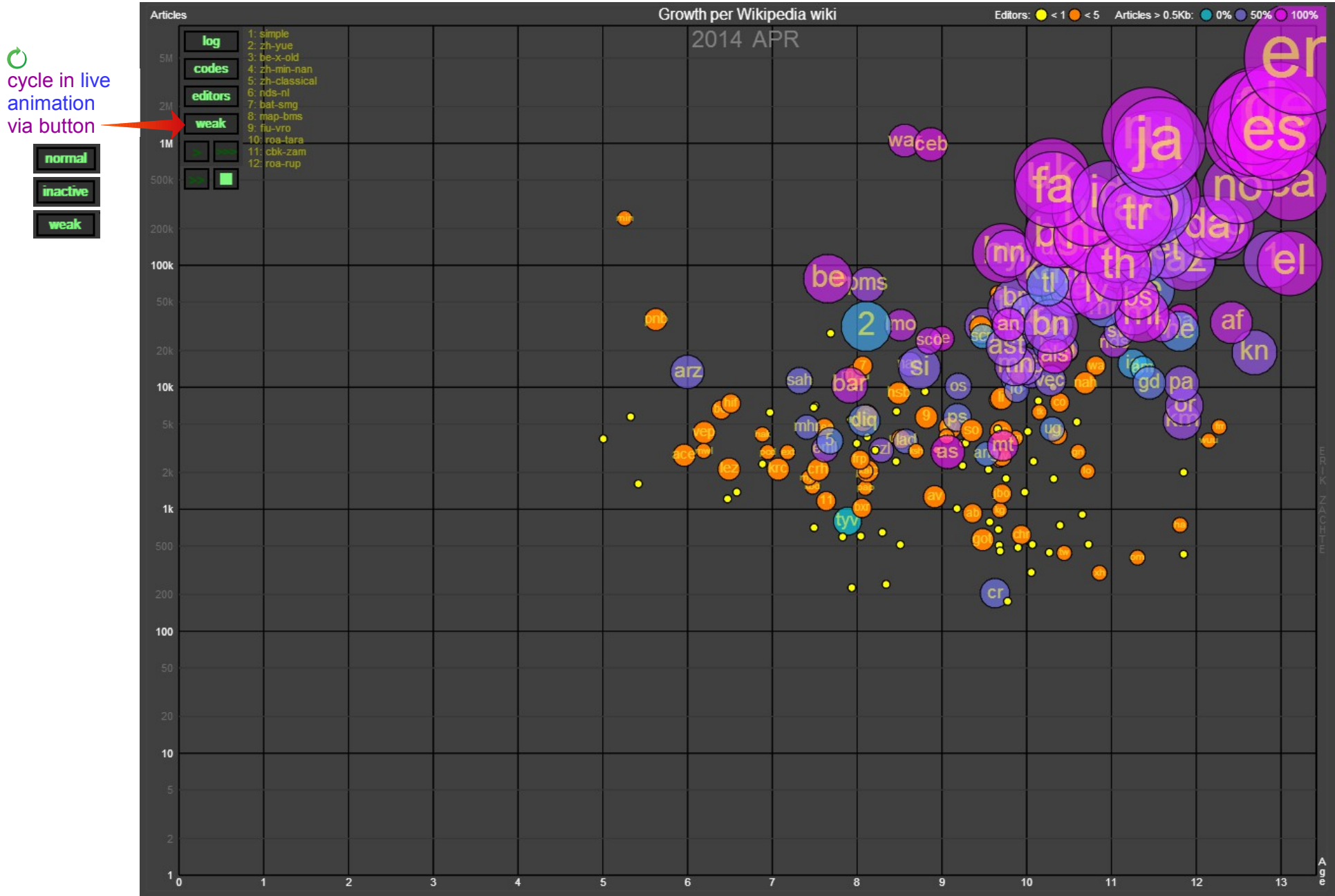


# Reality Check 3: 284 Wikipedias



This animation may seem to suggest all Wikipedias start to pick up momentum sooner or later. That observation seemed true when this was first published (2008), but needs to be nuanced now.

# Reality Check 3: 284 Wikipedias

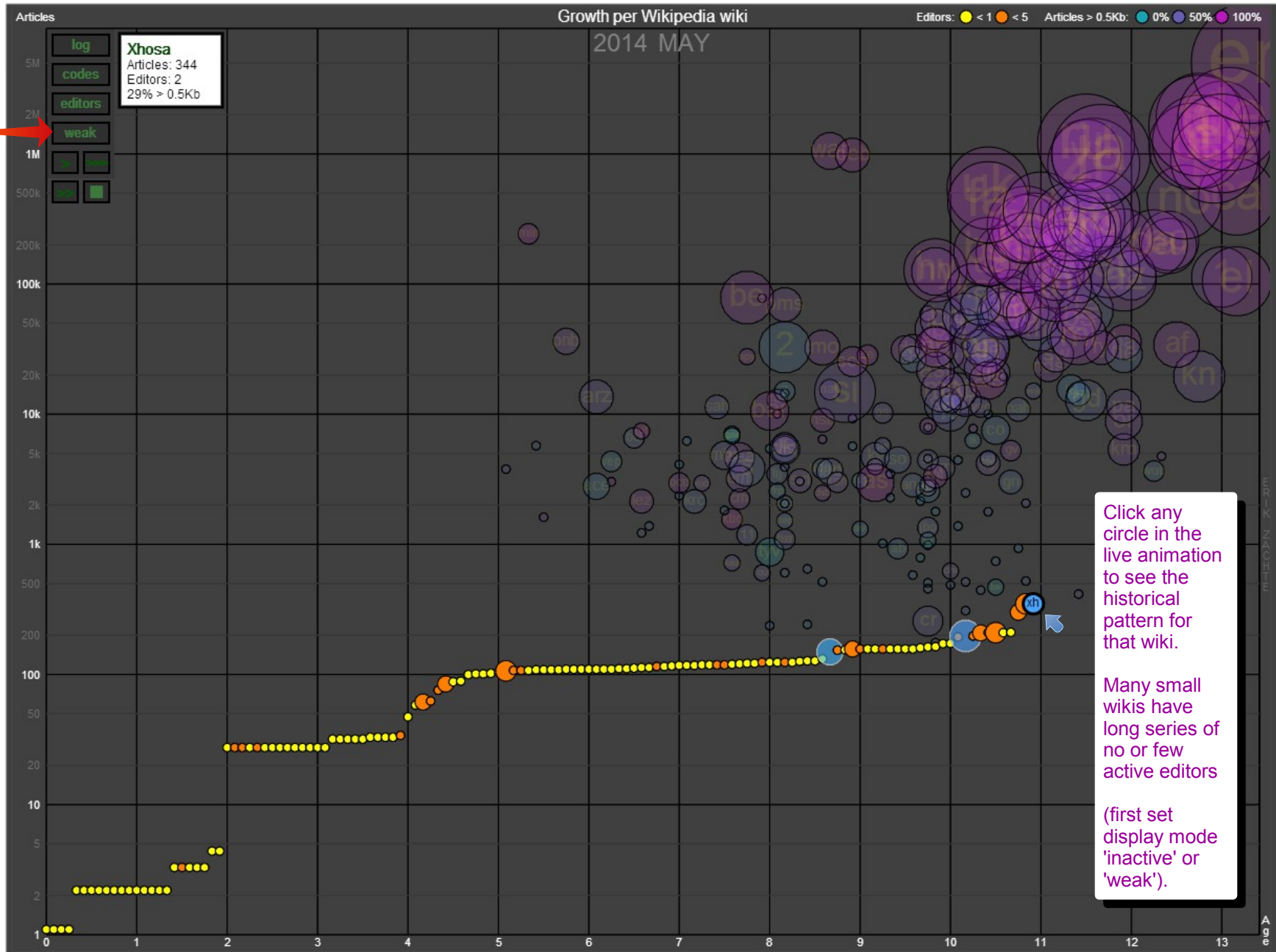


Wikipedias shown in yellow have no active editors in this month, orange: less than 5 active editors. Active editor defined as registered user making 5 or more edits in this month in 'content namespaces' (think articles).

# Reality Check 3: 284 Wikipedias

cycle in live animation via button

- normal
- inactive
- weak



Many small Wikipedias have long series of no or few active editors in any given month.

# Reality Check 3: 284 Wikipedias

12 dead (locked)

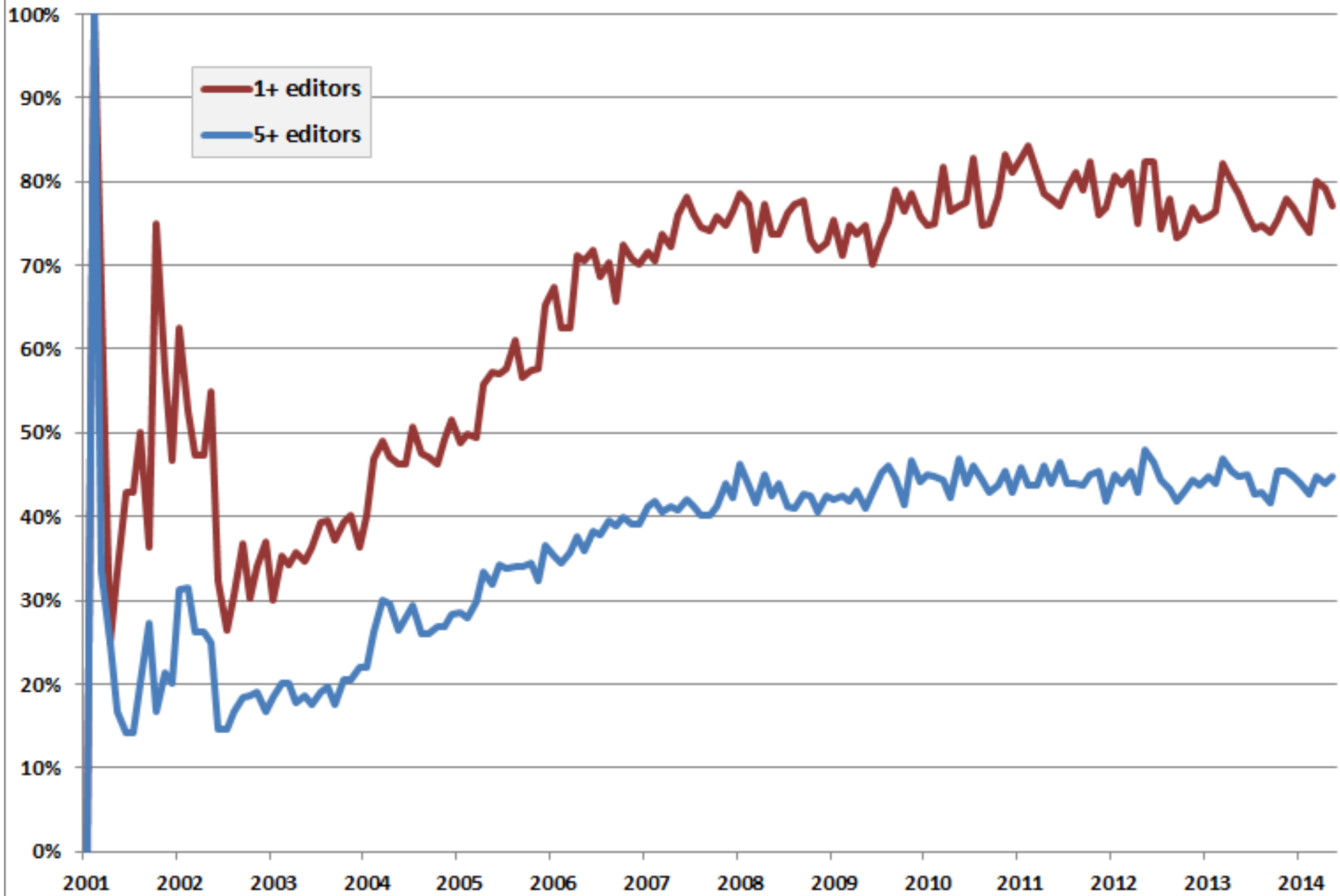
53 zombies (open, no editors)

94 struggling (open, < 5 editors)

125 in good or excellent health

If we breakdown our Wikipedias by number of active editors we see a more nuanced picture.  
Of course the threshold of 5 active editors in any month is arbitrary (but IMO conservative).

## Percentage of Wikipedia wikis with 1+/5+ active editors per month (with 5+ edits each)



The number of active Wikipedias is stable for many years. In early years many projects were created without proven substantial user base. Nowadays we use the [incubator project](#) to give new Wikipedia languages a place to gain credibility.

# Wikipedias with 10,000+ articles and < 5 active (registered) editors

	5+ edits	100+ edits	articles
Minangkabau	1	0	221,452
Haitian	1	0	53,756
Malagasy	2	0	47,280
Chechen	4	2	39,390
Yoruba	3	0	31,026
Ido	4	1	25,759
Bishnupriya Manipuri	2	0	25,231
Buginese	0	0	14,094
Walloon	2	1	13,807
Samogitian	1	0	13,672
Mazandarani	3	0	11,857

Some Wikipedias with a substantial amount of articles are really short of editors to maintain these.



# Instead of 284 Wikipedias we might say

We have  $X$  Wikipedias with

100,000 page views a month

again *views*, not bot requests

or

3+ active human editors

or

1000+ articles created by humans

Could we come up with a simple composite metric that better represents our installed base of active wikis?  
Some suggestions to start the conversation. (BTW same holds for our other projects)



# Let's err on the side of humbleness

Reporter:

“Can I ask about the 284 Wikipedias?”

We:

“You counted them? Interesting!

We knew it was above 100” (-;

Symbol (-; aka ;- ) means: please take this statement in a light-hearted manner.  
I am not really proposing to stop counting, just to report more meaningful numbers.

Let's err on the side of humbleness

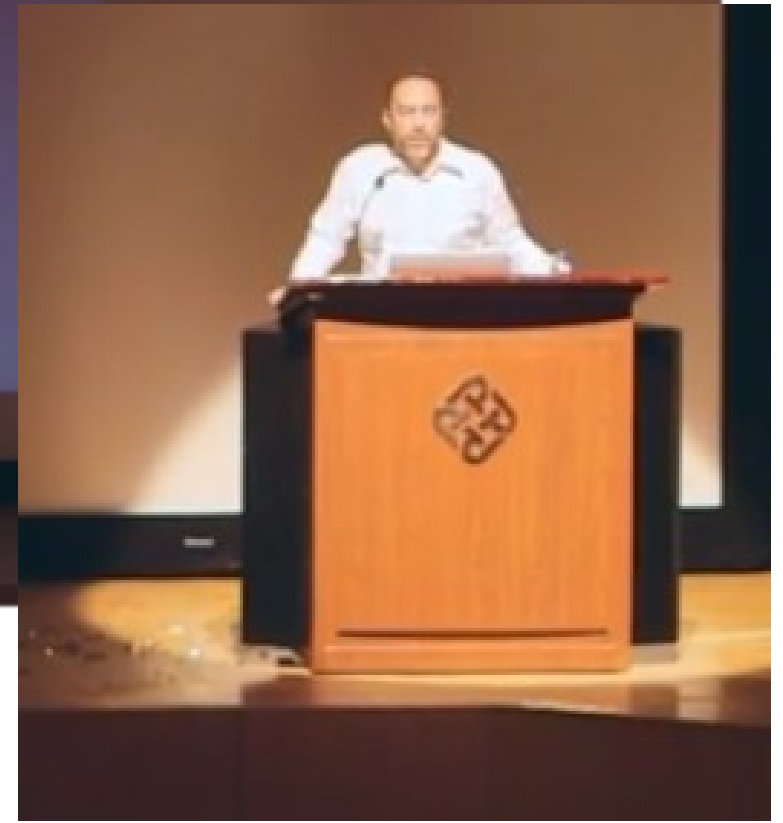
But

Let's not overdo it !

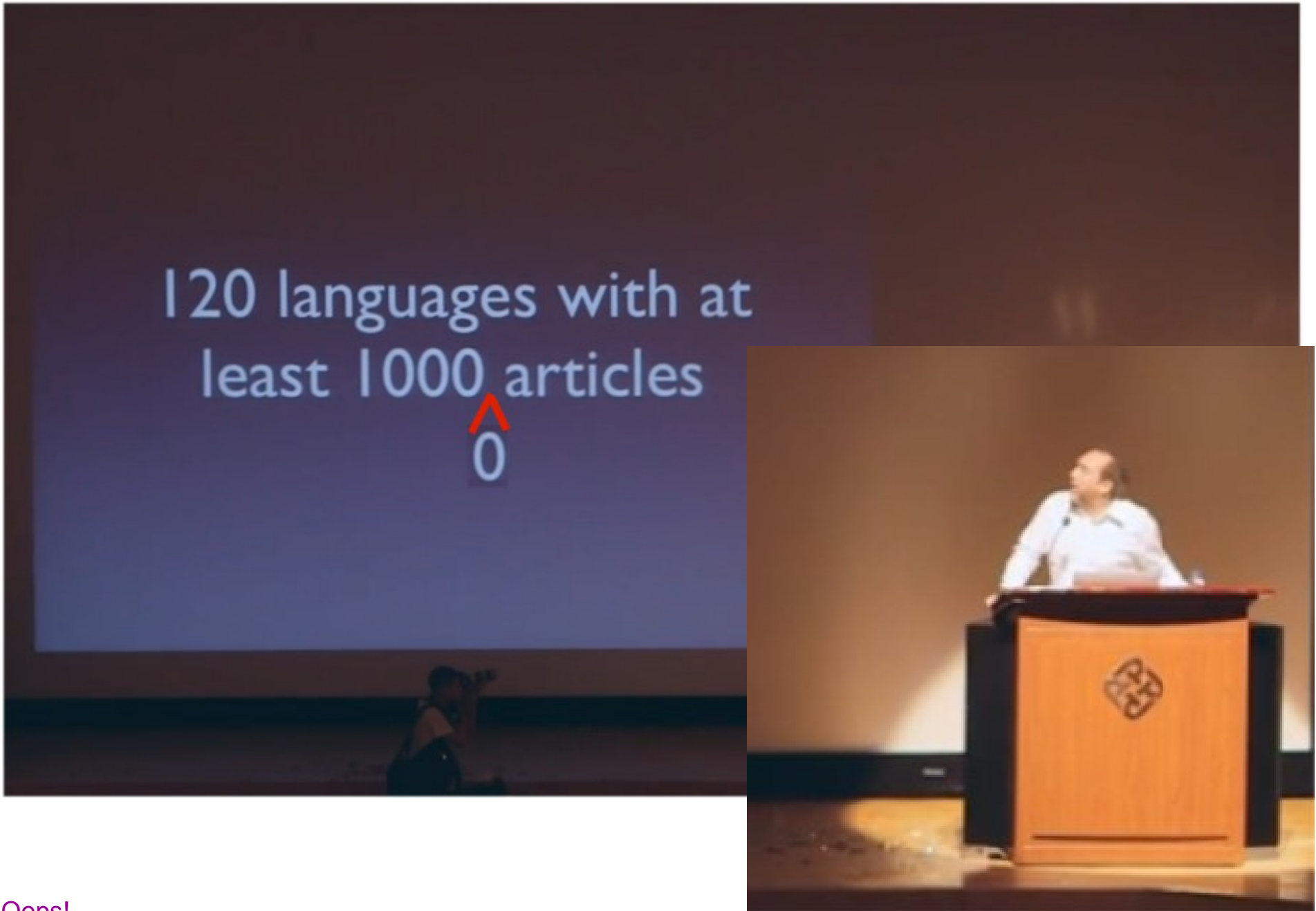
# Wikimania 2013

120 languages with at  
least 1000 articles

Jimmy showed this metric in his 2013 keynote.

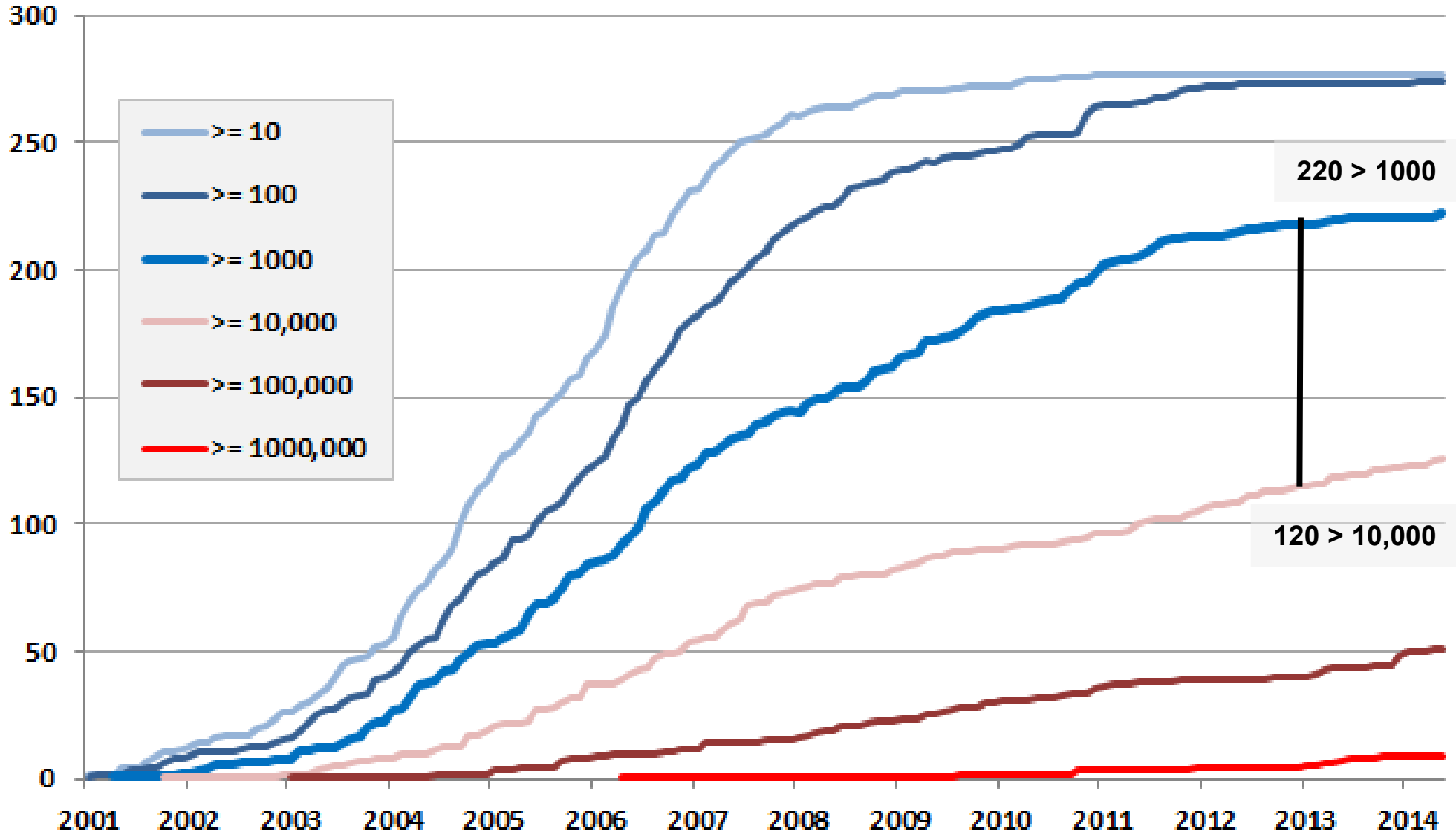


# Wikimania 2013



Oops!

# Wikipedias with at least x articles in content namespace(s)



We have about 100 more Wikipedias with 1000+ articles than Jimmy inadvertently showed us. Ironically many Wikipedias almost jumped into existence by bot created articles: 2000 stubs for all years after 0 A.D.

# Awesomeness Recalibrated

$$\infty / 2 = \infty$$

To be sure, I do not want to take away from our 'infinite' awesomeness at all!  
I just want us to quote metrics which better represent our accomplishments.

# Topics

Metrics for public and press

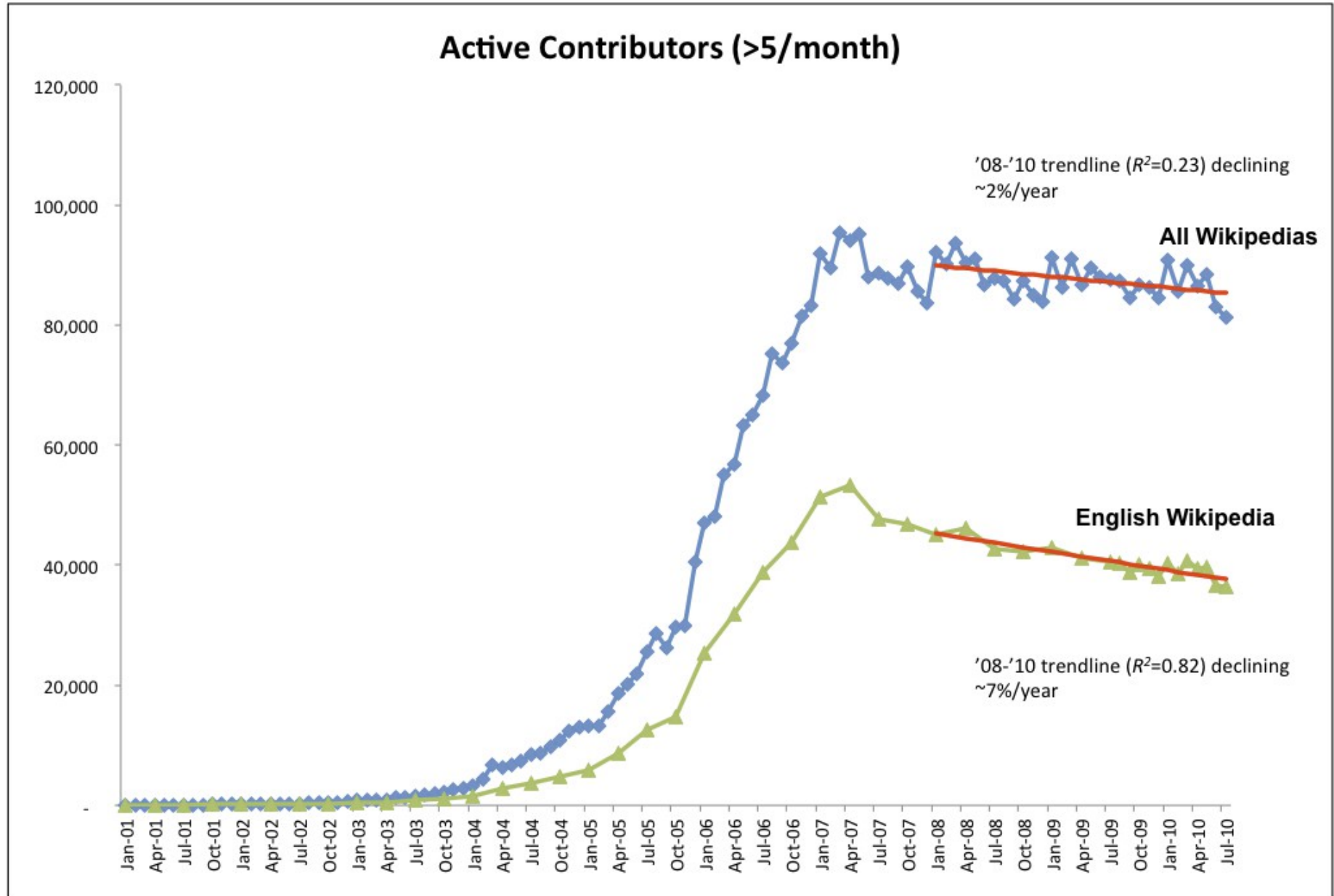
## Editor Trends

- Unique Editors
- Migrations
- Retention



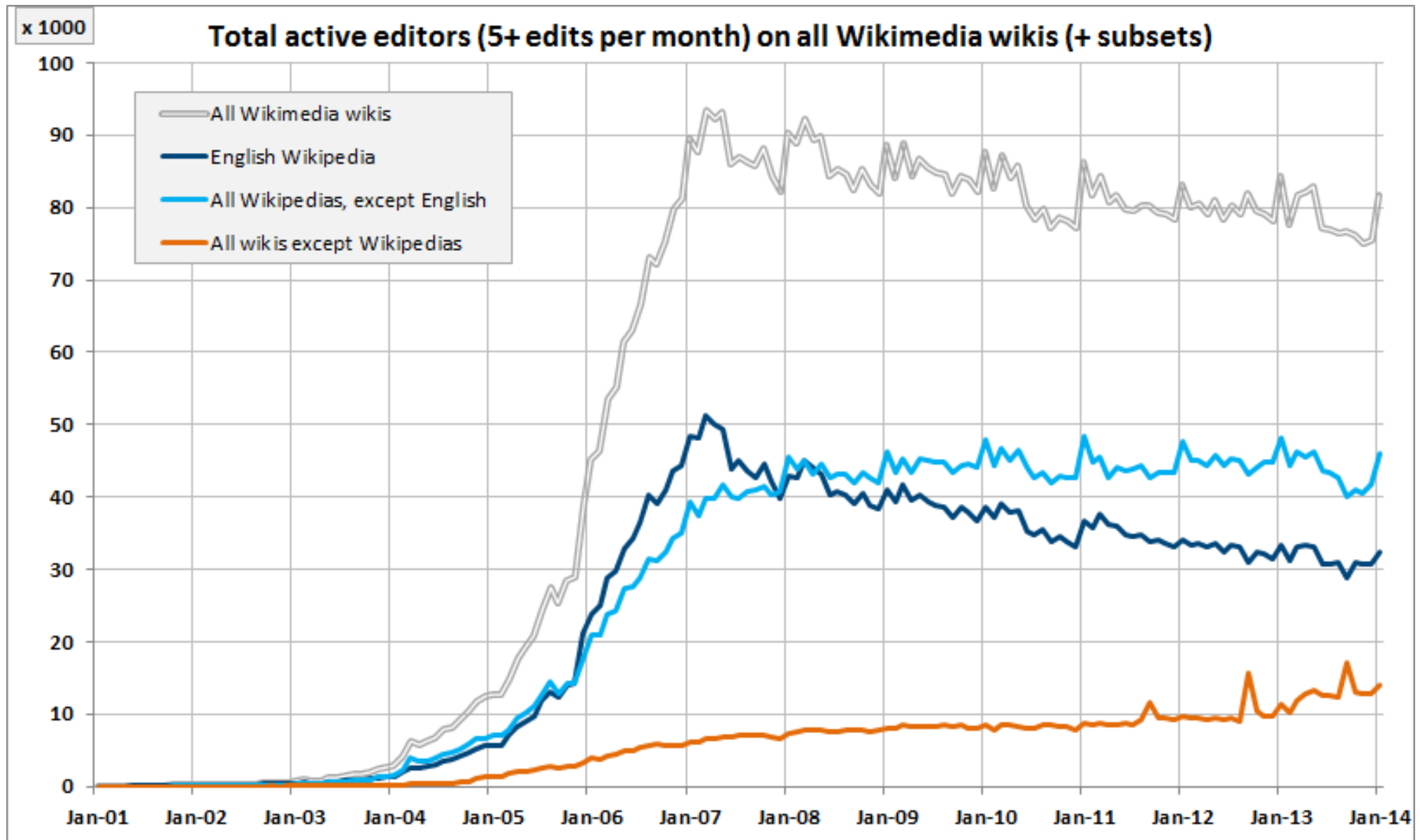
# Active Editor Trends Wikipedia

## English / All



This chart dates from our strategy project. The direction of the trend lines after 2010 only changed slightly.

# Editor Decline mostly on WP:EN



However if we break down trends differently, singling out English Wikipedia, other Wikipedias \*on average\* showed no decline in number of active editors. Peaks in 11/12/13 on other projects are Wiki Loves Monuments contests. [More charts at my blog.](#)

# Migration between projects

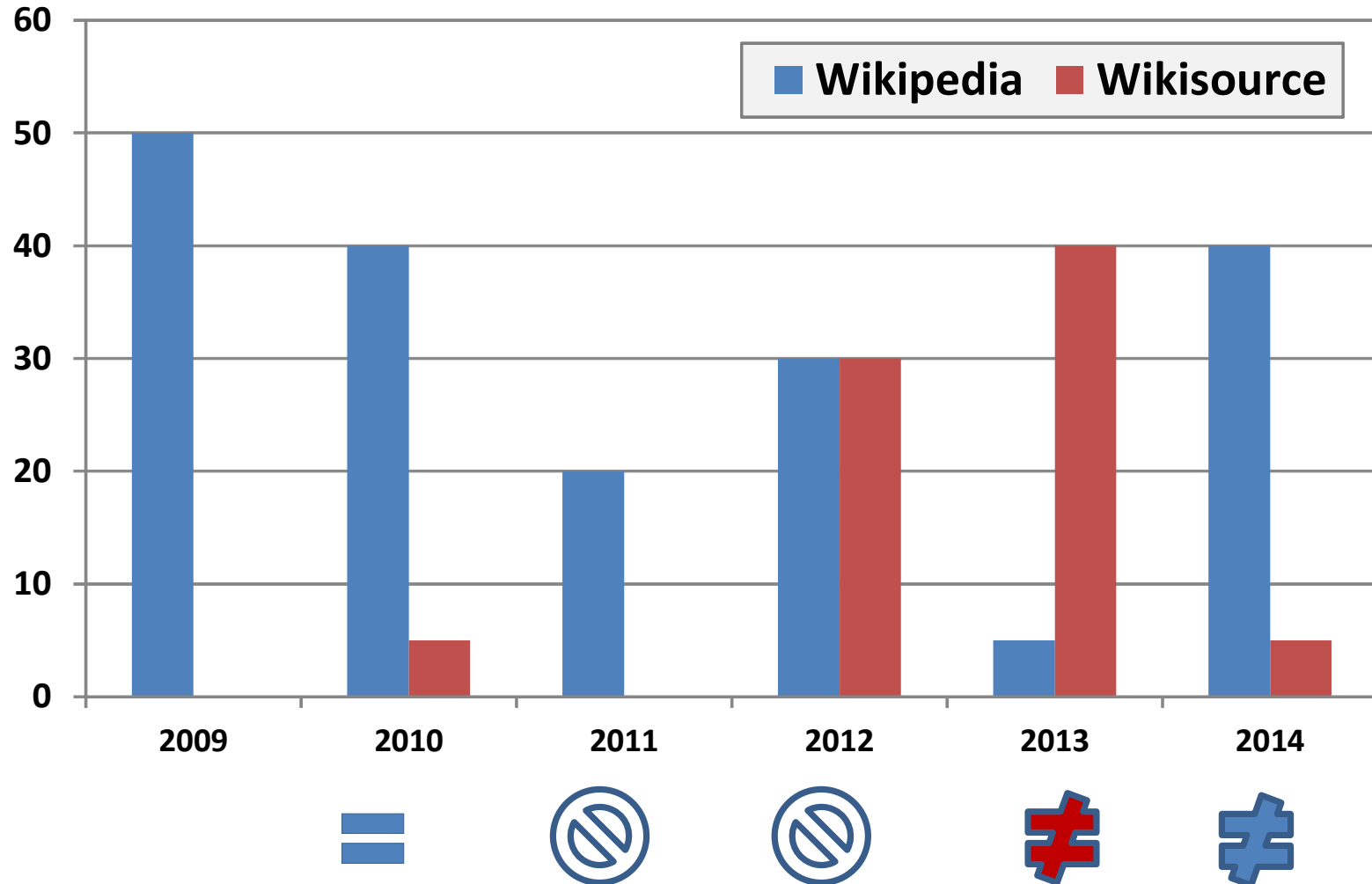
## John's edits per year per project

namespace 0, registered users only

thresholds:

> 25 overall edits/year

main > 60% on one project



Could it be that some editors who left English Wikipedia did not abandon our projects all-together, but shifted focus to other projects or languages?

Based on two thresholds this hypothetical user migrated in 2013 and 2014.  
For 2011 he had too less edits to be counted, for 2012 he had no 'main' project.

# Editors migrating between projects

Registered editors with 25+ article edits per year and 60+% edits in 'main' project

	wp => wk	wk => wp	wp => wb	wp => ws	wb => wp	wp => wd	wp => wx	ws => wp	wp => wq	wx => wp
<b>total</b>	<b>941</b>	<b>627</b>	<b>570</b>	<b>530</b>	<b>509</b>	<b>463</b>	<b>363</b>	<b>307</b>	<b>284</b>	<b>194</b>
<b>2002</b>										
<b>2003</b>	<b>3</b>									
<b>2004</b>	<b>8</b>	<b>4</b>	<b>3</b>	<b>4</b>	<b>1</b>		<b>3</b>	<b>1</b>	<b>4</b>	
<b>2005</b>	<b>40</b>	<b>32</b>	<b>44</b>	<b>16</b>	<b>13</b>		<b>12</b>	<b>2</b>	<b>6</b>	<b>4</b>
<b>2006</b>	<b>69</b>	<b>57</b>	<b>50</b>	<b>43</b>	<b>62</b>		<b>18</b>	<b>12</b>	<b>22</b>	<b>19</b>
<b>2007</b>	<b>111</b>	<b>66</b>	<b>96</b>	<b>55</b>	<b>97</b>		<b>25</b>	<b>29</b>	<b>38</b>	<b>18</b>
<b>2008</b>	<b>117</b>	<b>70</b>	<b>79</b>	<b>83</b>	<b>61</b>		<b>27</b>	<b>39</b>	<b>38</b>	<b>20</b>
<b>2009</b>	<b>146</b>	<b>83</b>	<b>74</b>	<b>75</b>	<b>63</b>		<b>31</b>	<b>40</b>	<b>44</b>	<b>24</b>
<b>2010</b>	<b>115</b>	<b>94</b>	<b>83</b>	<b>54</b>	<b>56</b>		<b>35</b>	<b>48</b>	<b>45</b>	<b>19</b>
<b>2011</b>	<b>109</b>	<b>83</b>	<b>58</b>	<b>68</b>	<b>64</b>		<b>53</b>	<b>38</b>	<b>34</b>	<b>17</b>
<b>2012</b>	<b>108</b>	<b>69</b>	<b>47</b>	<b>61</b>	<b>49</b>	<b>51</b>	<b>76</b>	<b>54</b>	<b>34</b>	<b>26</b>
<b>2013</b>	<b>115</b>	<b>69</b>	<b>36</b>	<b>71</b>	<b>43</b>	<b>412</b>	<b>83</b>	<b>44</b>	<b>19</b>	<b>47</b>

Users with 25+ edits per year = 11% of total. For these editors 60% of edits in one project (= 'main project') → rejects 0.6% of editor-years)

wp : Wikipedia, wk: Wiktionary, wb: Wikibooks, ws: Wikisource, wd: Wikidata, wq:Wikiquote, wx: Other projects

# % editors migrating between projects

Registered editors with 25+ article edits per year and 60+% edits in 'main' project

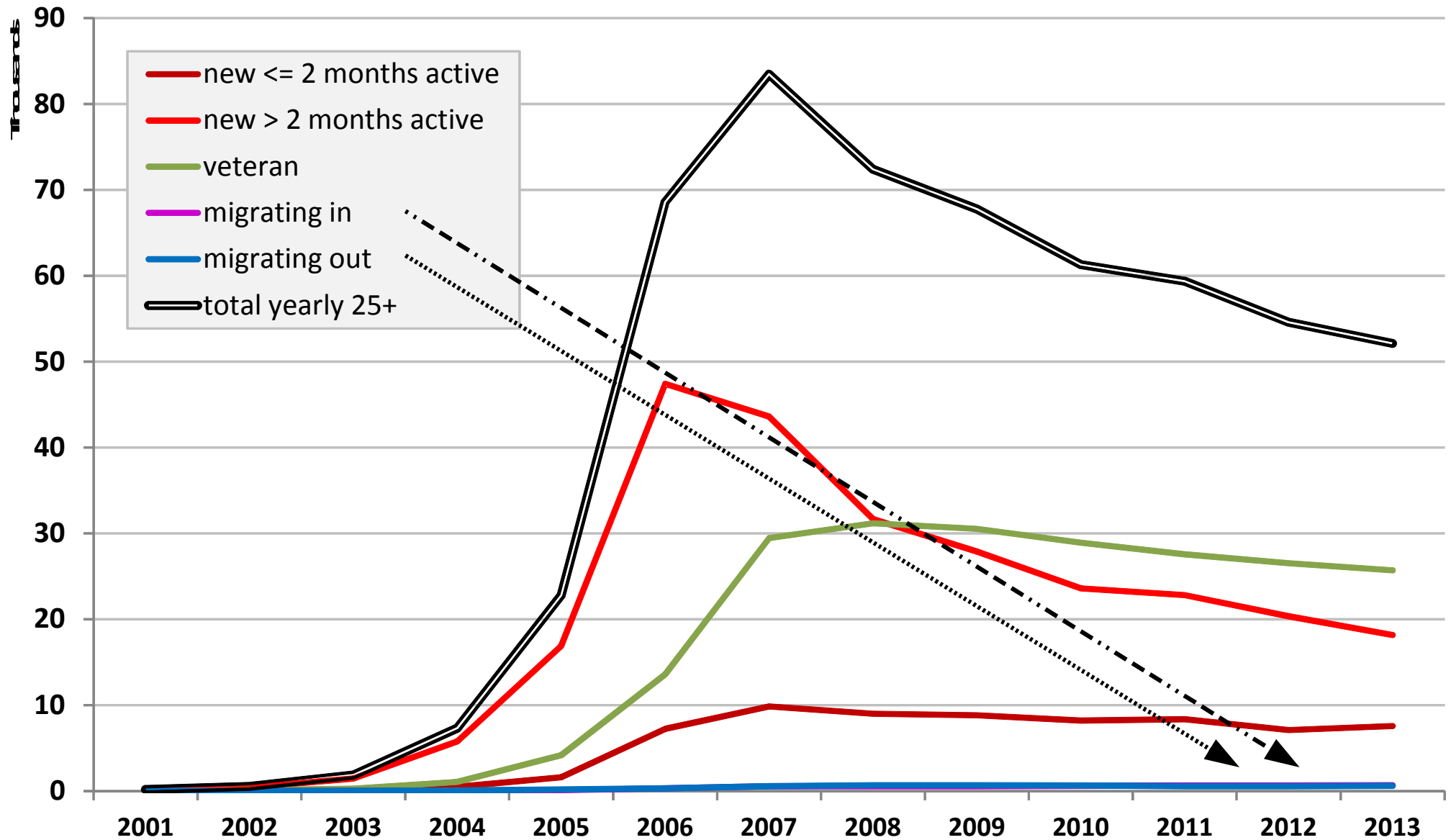
	Wikipedia		Wikidata		Wiktionary		Wikibooks		Wikisource		Commons	
	in	out	in	out	in	out	in	out	in	out	in	out
2003					9							
2004		0.1			5	2	2	1	18	5		
2005	0.1	0.3			10	8	8	3	22	2	11	2
2006	0.2	0.2			10	8	5	7	18	5	13	8
2007	0.2	0.3			12	8	8	8	19	10	13	9
2008	0.2	0.3			14	9	6	6	21	10	18	9
2009	0.2	0.3			17	10	7	6	18	10	16	20
2010	0.2	0.3			14	11	10	7	16	14	19	18
2011	0.2	0.3			13	11	8	9	18	10	22	24
2012	0.2	0.3	62		13	9	6	5	17	17	13	19
2013	0.2	0.7	55	2	12	8	5	7	22	17	20	17

Users with 25+ edits per year = 11% of total. For these editors 60% of edits in one project (= 'main project') → rejects 0.6% of editor-years)

These migration stats are more meaningful when expressed as share of total editors who had that same project as main project in that year.

# % editors migrating to/from Wikipedia

Active editors (25+ edits/year) with Wikipedia as main project ( $\geq 60\%$  edits)



So not much migration from/to Wikipedia projects really. Breakdown of editors into veterans, and new editors, new again broken down by fleeting and more committed, is to spice up the chart, but unrelated.

# % net migration between Wikipedias

Registered editors with 25+ article edits per year and 60+% edits in 'main' wiki

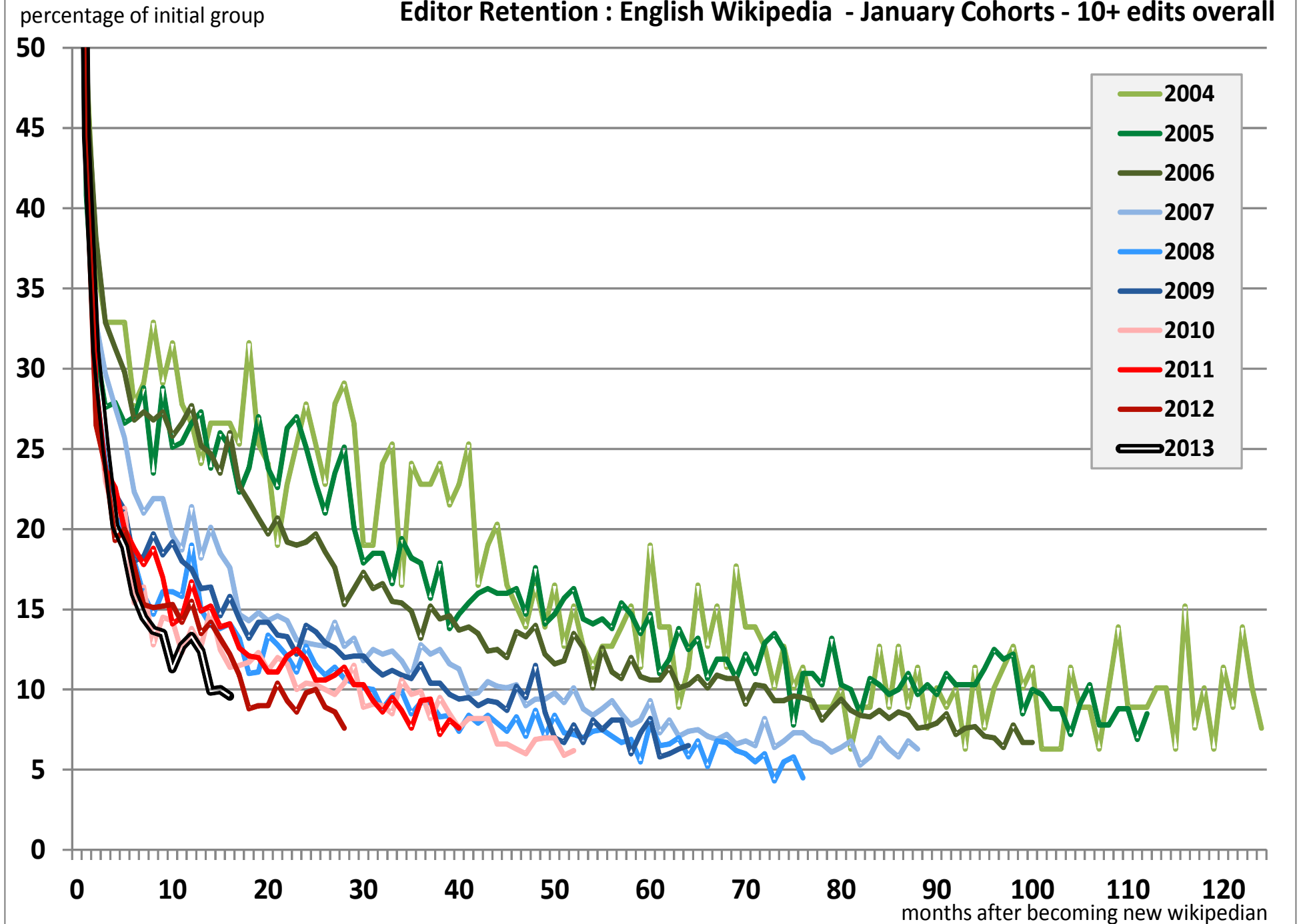
	English	German	French	Japanese	Russian	Spanish	Italian	Chinese	Polish	Portug.	Dutch
2002	-0.5	2.5	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0
2003	-0.8	0.0	0.5	0.4	0.0	0.0	3.6	4.8	3.4	10.0	0.0
2004	-0.5	0.4	-0.3	-0.3	1.9	0.5	1.2	0.8	-0.5	-0.7	0.6
2005	-0.3	0.0	0.4	-0.7	2.9	-0.2	0.6	-0.1	-0.2	0.3	0.4
2006	0.0	-1.1	-0.5	-0.1	1.7	0.7	0.5	0.0	0.5	0.7	-0.4
2007	0.0	-1.1	-0.6	0.0	1.2	0.2	0.5	-0.3	0.2	-0.1	0.0
2008	-0.1	-0.4	0.0	-0.4	1.4	-0.1	0.0	-0.1	-0.5	-0.3	-1.4
2009	-0.1	-0.6	0.0	-0.4	0.9	-0.4	0.2	0.0	0.0	0.0	-0.4
2010	0.0	-0.7	-0.1	0.0	0.4	-0.2	0.3	0.7	0.4	-0.3	-0.7
2011	0.1	-0.5	-0.3	-0.3	-0.2	-0.2	-0.4	-0.1	-0.4	-0.4	-1.2
2012	0.2	-0.2	-0.5	0.0	0.1	-0.5	0.2	0.0	-0.7	-0.5	0.4
2013	0.2	-0.4	0.0	0.2	-0.3	-0.3	-0.3	-0.2	-0.7	-0.7	-0.1

Negative net migration: more outgoing than incoming editors

Similarly, migration between Wikipedia languages only applies to small fraction of overall editors.



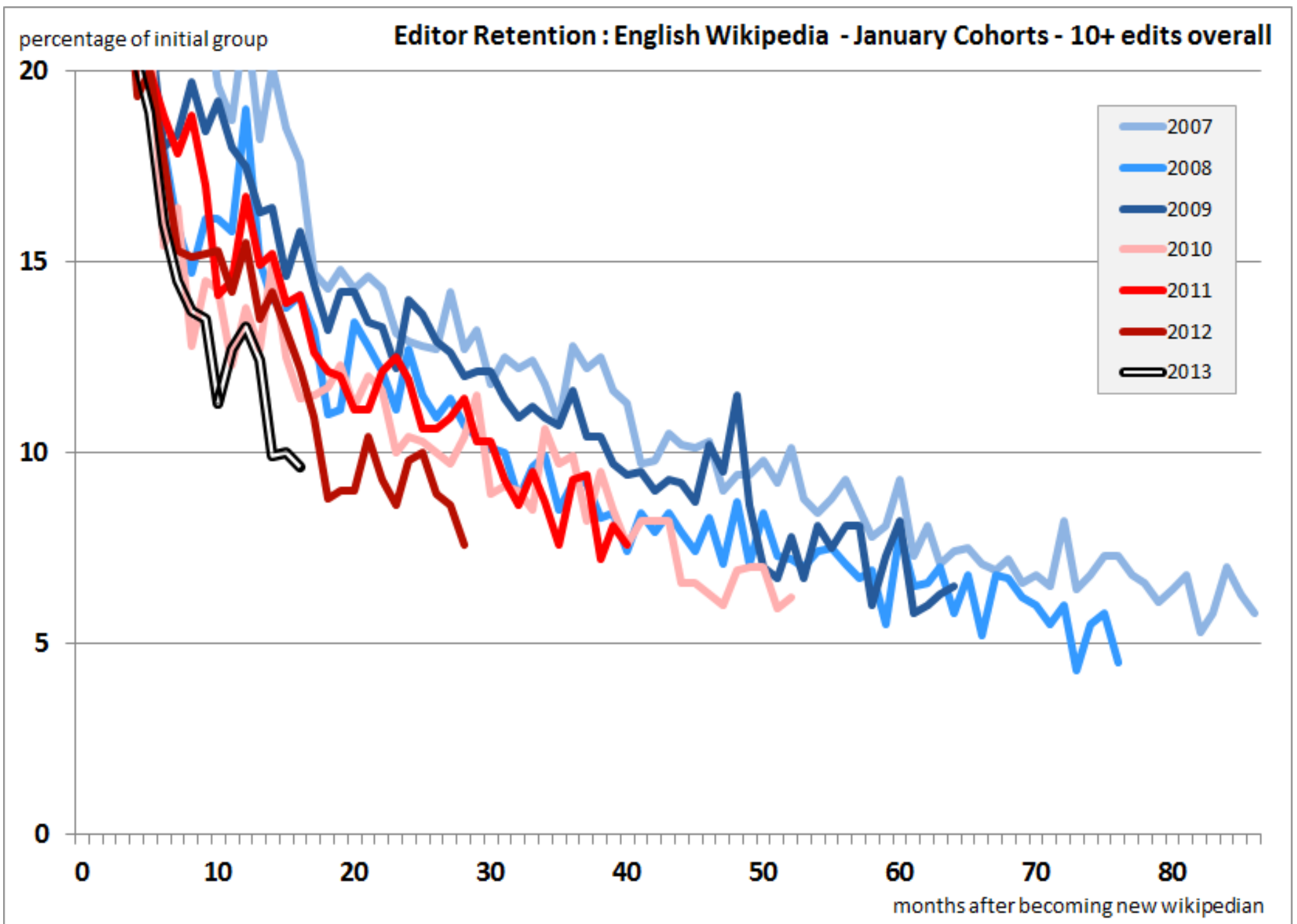
# Editor Retention : English Wikipedia - January Cohorts - 10+ edits overall



Not much point in refreshing migration stats regularly.

Fortunately same data can also be used to revive one-time chart from 2010 on editor retention stats.

This might be updated for major Wikipedias on a regular basis. Expect blog post.



Same chart, zoomed in, and for later years only.

# Recap

- Editor decline:  
Mostly on English Wikipedia
- Editor retention rate:  
Minor decrease since 2007
- Editor loyalty:  
Few migrations away from  
Wikipedia or to other language

# Questions?



Photo at Wikimania 2014 by Pierre-Selim Huard (CC-BY-SA 3.0)