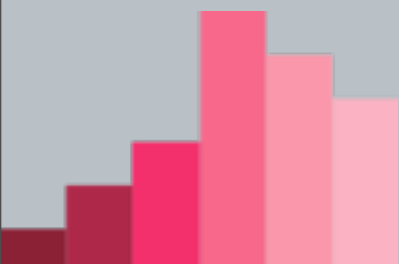


Measuring Article Quality

Peer Review and the Article Feedback Tool

Adam Hyland
protonk @ en-wp



Look Familiar?

Rate this page

[What's this?](#)

[View page ratings](#)

Trustworthy



Objective



Complete



Well-written

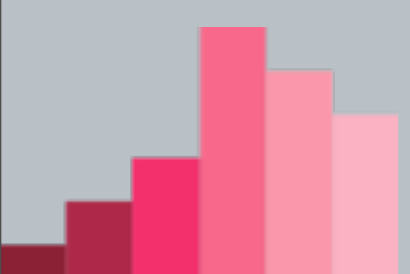


- I am highly knowledgeable about this topic (optional)
- I have a relevant college/university degree
 - It is part of my profession
 - It is a deep personal passion
 - The source of my knowledge is not listed here
 - I would like to help improve Wikipedia, send me an e-mail (optional)

We will send you a confirmation e-mail. We will not share your e-mail address with outside parties as per our [feedback privacy statement](#).

Your ratings have not been submitted yet

[Submit ratings](#)



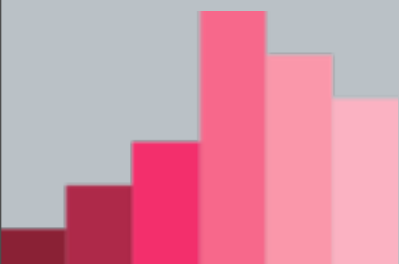
Maybe This Version?

Help improve this page

Did you find what you were looking for?

Yes

No





Rate this page [View page ratings](#)

[What's this?](#)

Trustworthy Objective Complete Well-written

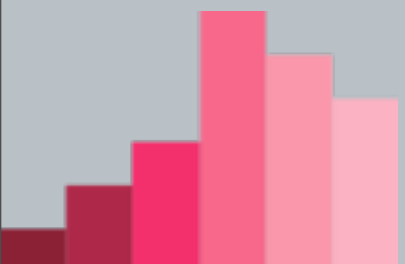
★★★★★ ★★★★★ ★★★★★ ★★★★★

I am highly knowledgeable about this topic (optional)

- I have a relevant college/university degree
- It is part of my profession
- It is a deep personal passion
- The source of my knowledge is not listed here
- I would like to help improve Wikipedia, send me an e-mail (optional)

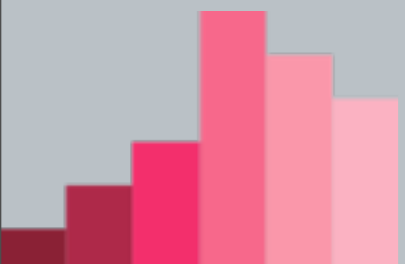
We will send you a confirmation e-mail. We will not share your e-mail address with outside parties as per our [feedback privacy statement](#).

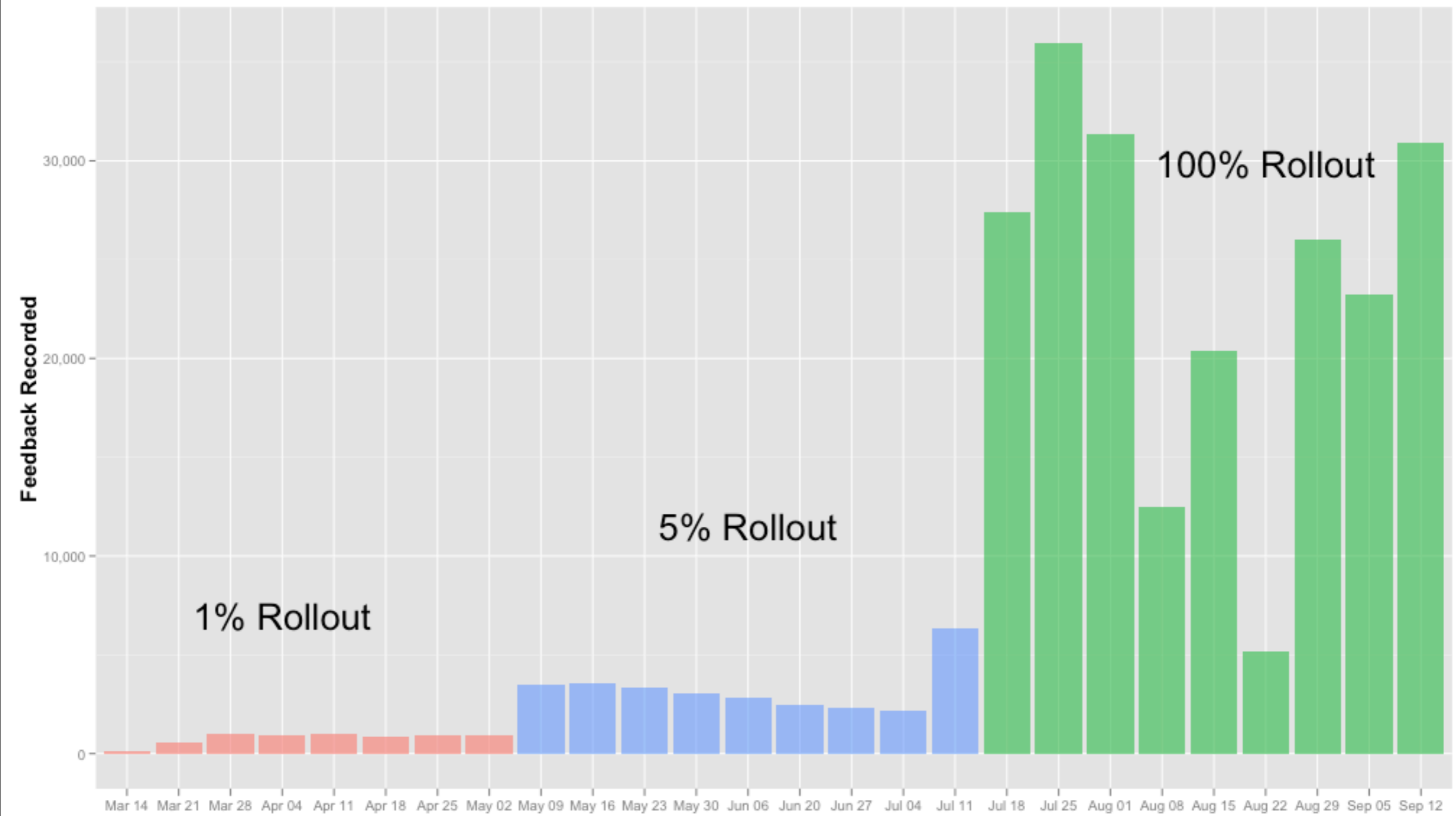
Your ratings have not been submitted yet



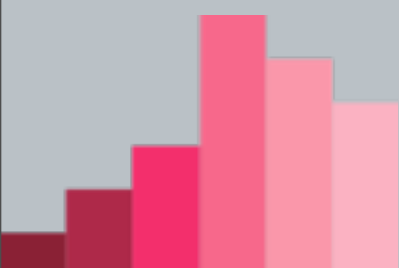
Article Feedback Tool

- Deployed in 2010
- Version 4 (the current version) ramped up in 2011
- Designed to offer an avenue for reader feedback
- High volume of reader feedback





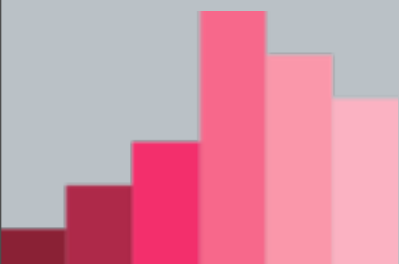
- 6 months of public data
- 795,353 articles -- 2,487,522 responses



Featured Articles (FA)



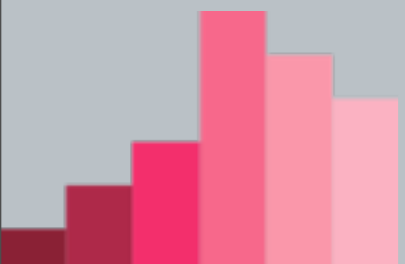
- 3,599 articles (0.09% of all articles)
- 2,267 Featured Lists (FL)
- Most rigorous peer review process on the English Wikipedia
- Very sensitive to editor preferences
- Some idiosyncrasies



Good Articles (GA)

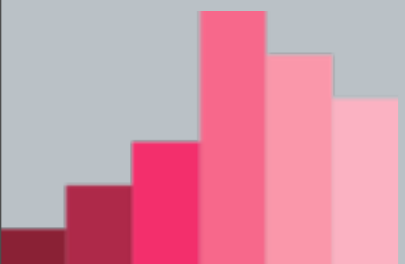


- 15,357 articles
- Relatively rigorous peer review (yes I know reasonable minds may disagree)
- Less idiosyncratic than FA in some ways
- Perhaps less dependent on editor preference

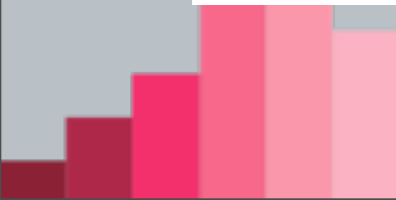
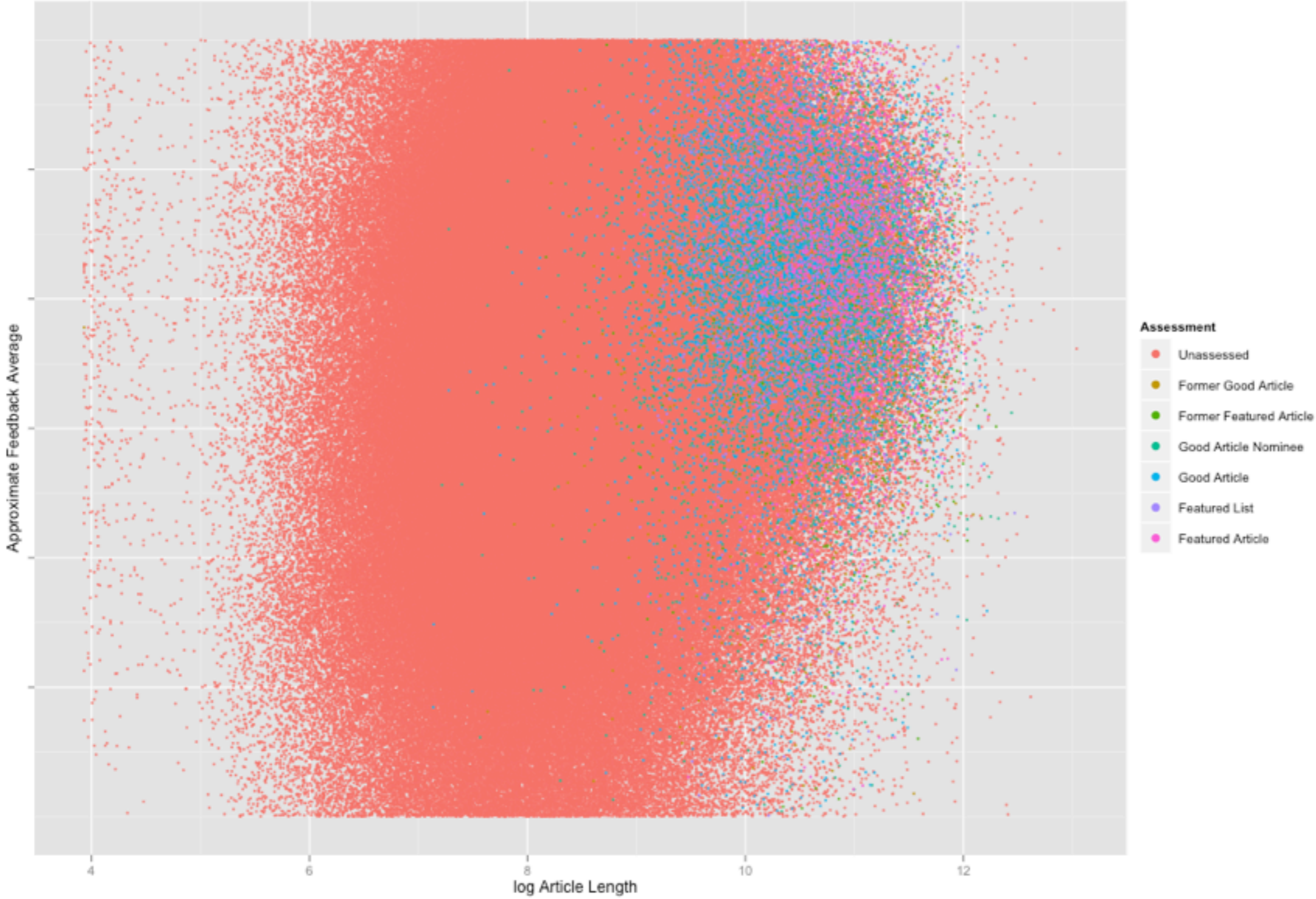


Data

- Article name
- Length (in bytes)
- GA/FA status (including former/not-promoted)
- Some user data

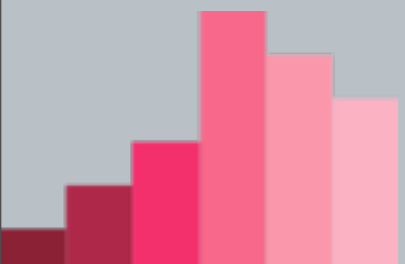


Article Ratings by Length and Project Quality Measure

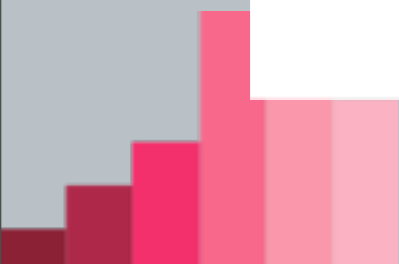
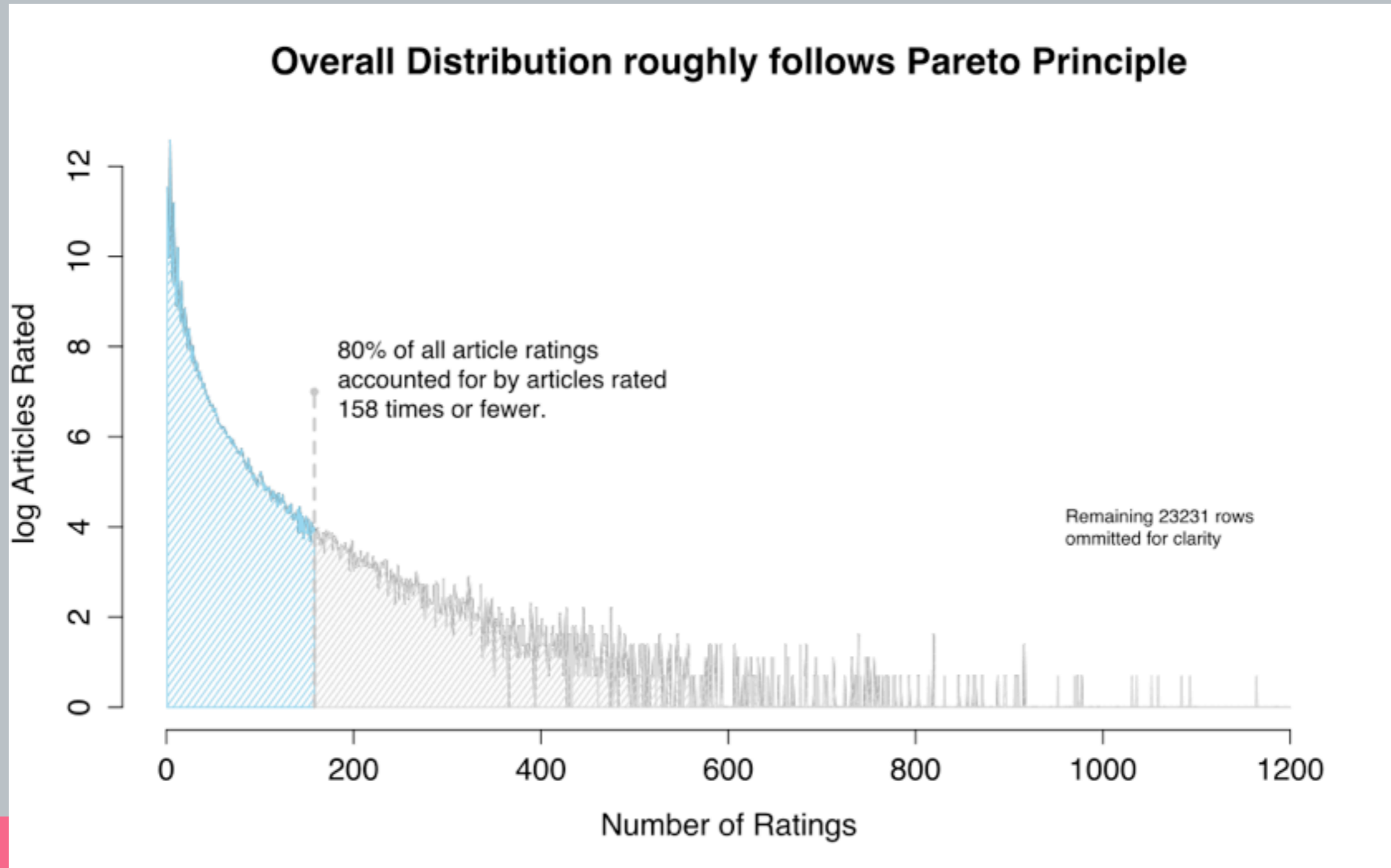


Beyond Summaries

- Reader ratings follow pageviews
- Predominantly non-editors
- Popular articles:
 - Call of Duty
 - Justin Bieber
 - Jimmy Wales (avg. rating: 1.10585)

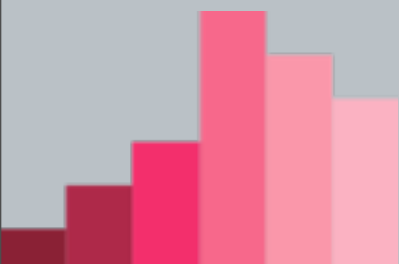


Power Laws Everywhere!



Classical(ish) Models

- Logistic regression model supports a relationship between rating and likelihood of FA/GA
- Linear model does, but with a twist
- Can't escape Cambridge Endogeneity Police!



Rate this page [View page ratings](#)

What's this?

Trustworthy Objective Complete Well-written

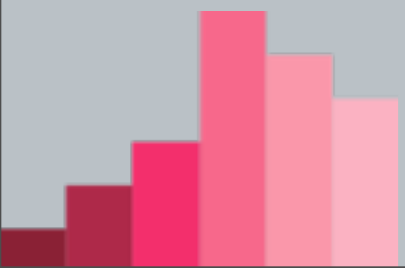
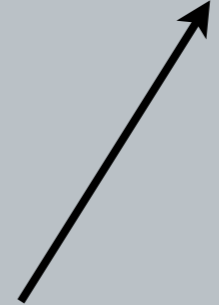
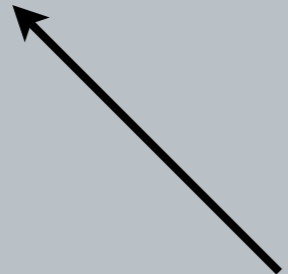
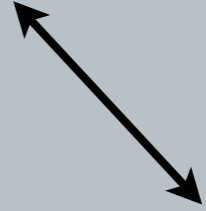
★★★★★ ★★★★★ ★★★★★ ★★★★★

I am highly knowledgeable about this topic (optional)

- I have a relevant college/university degree
- It is part of my profession
- It is a deep personal passion
- The source of my knowledge is not listed here
- I would like to help improve Wikipedia, send me an e-mail (optional)

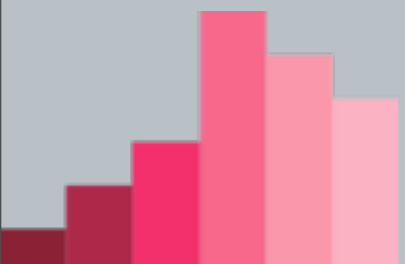
We will send you a confirmation e-mail. We will not share your e-mail address with outside parties as per our [feedback privacy statement](#).

➔ Your ratings have not been submitted yet [Submit ratings](#)



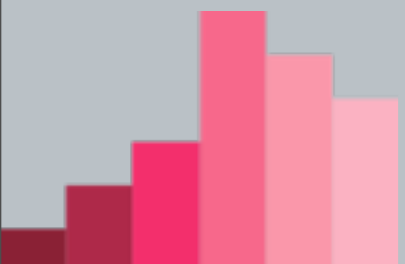
Data Mining

- Predicting featured status from reader ratings and minimal meta-data.
- Bayesian classifier able to roughly predict featured status (with a high false positive rate)

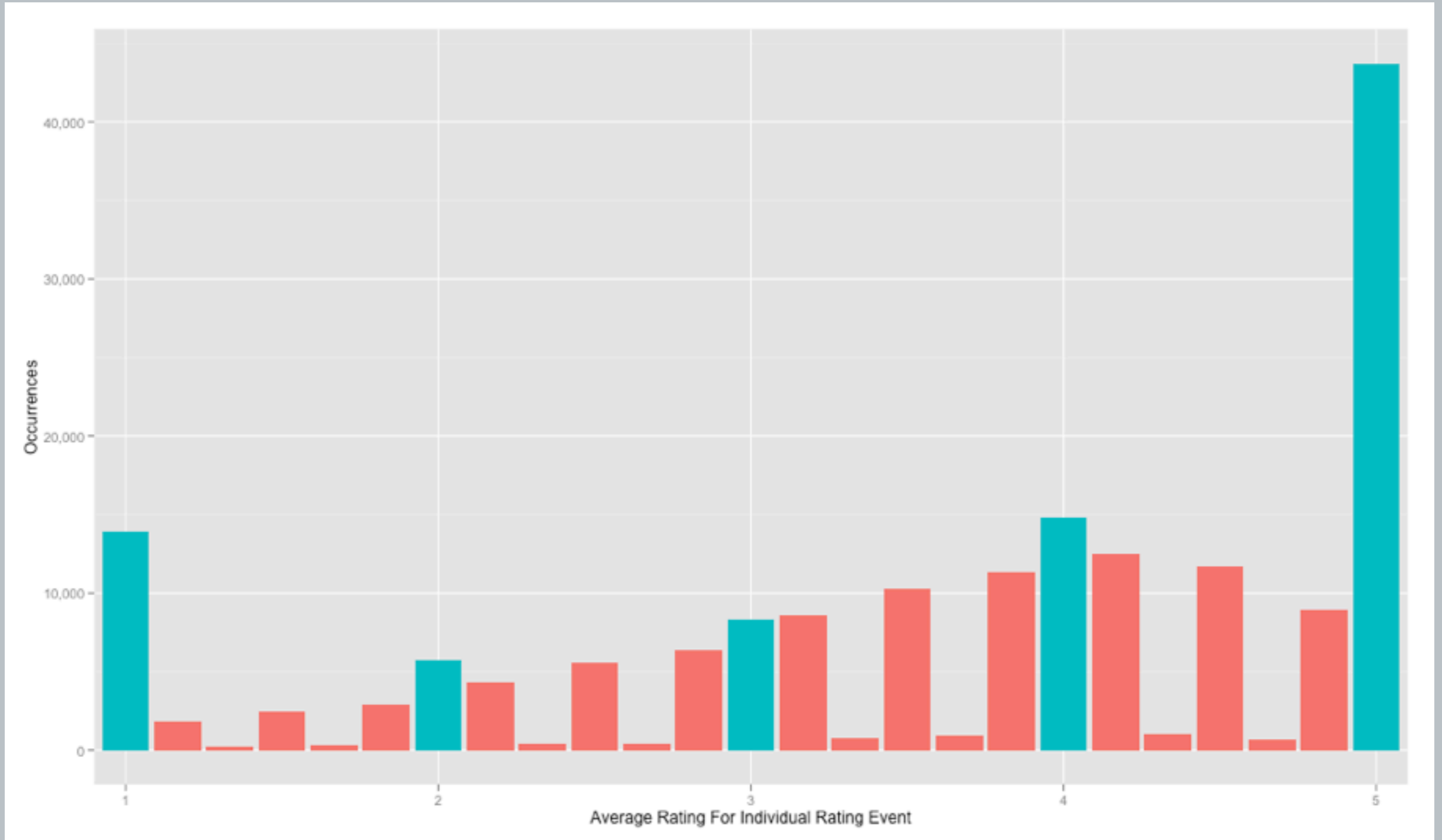


But the system's changing!

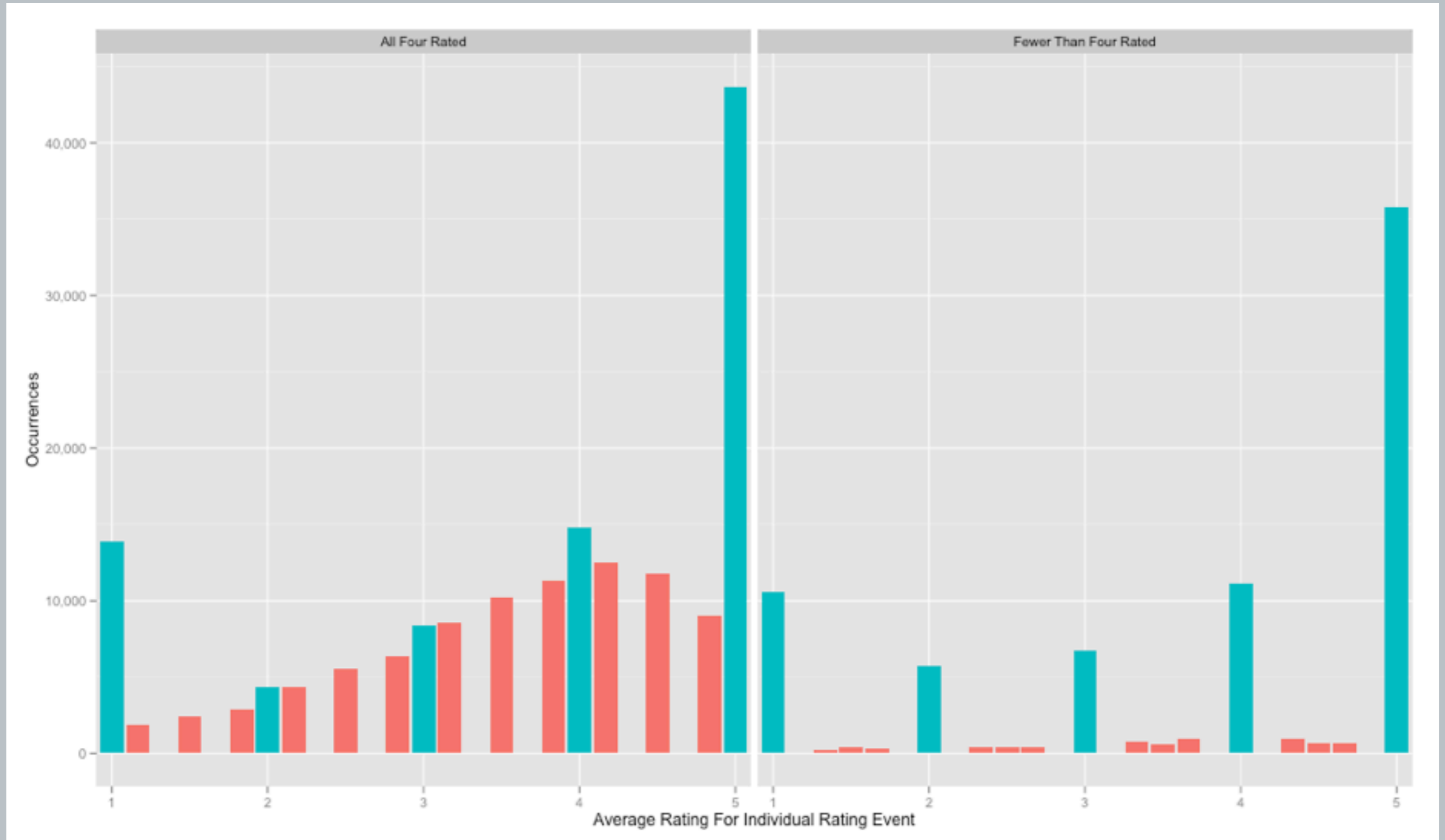
- AFT v4 is a multi-category quantitative measure
- AFT v5 is, roughly, YES/NO
- Is this a problem?
- Frank Harrell and the perils of dichotomization.



Actual Reader Ratings

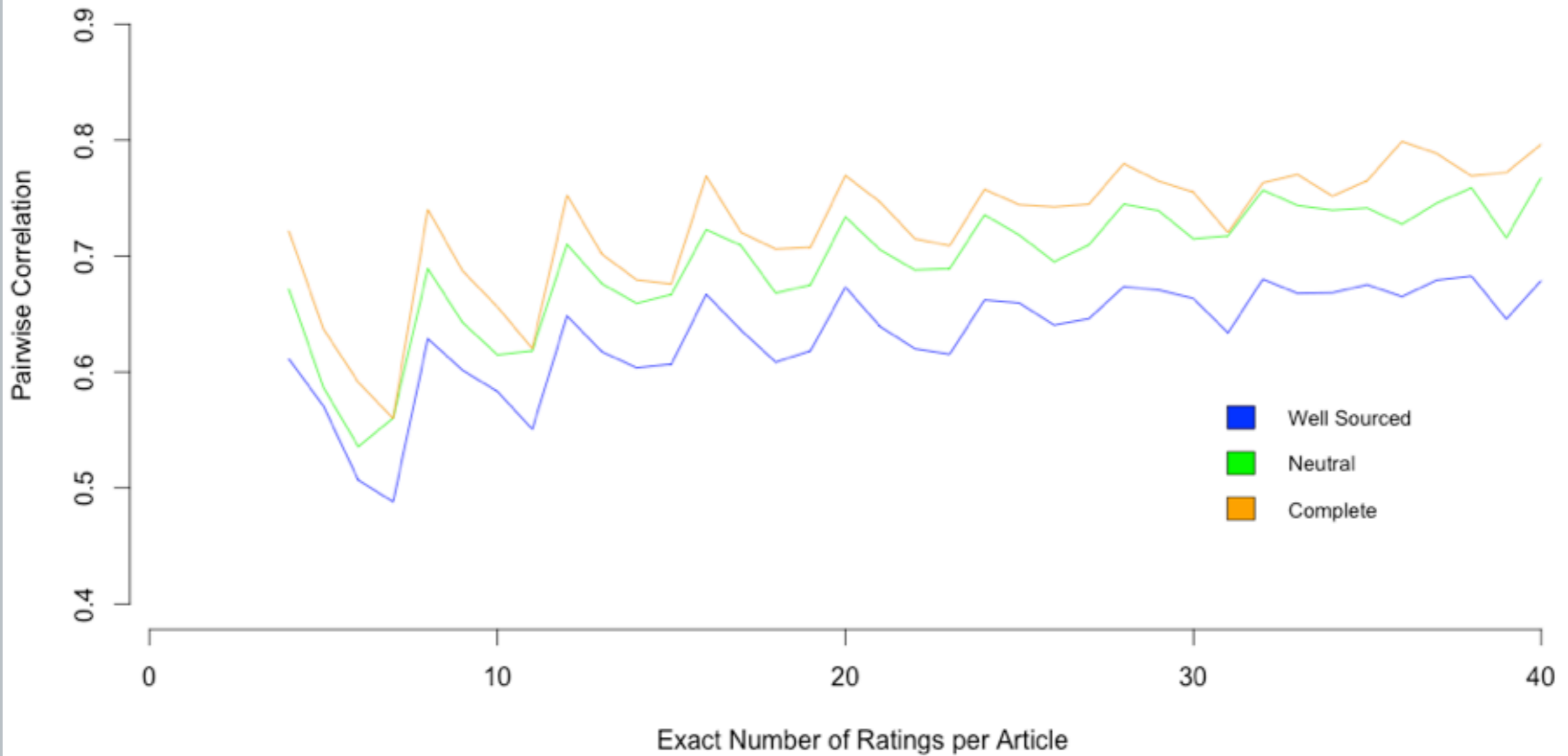


Another Look



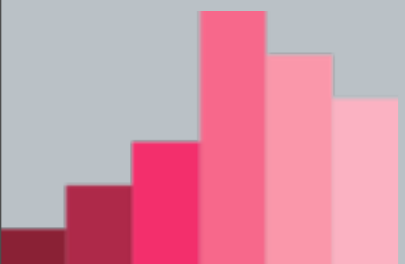
For the skeptics

Leave-one-out Correlation Between Ratings and Average



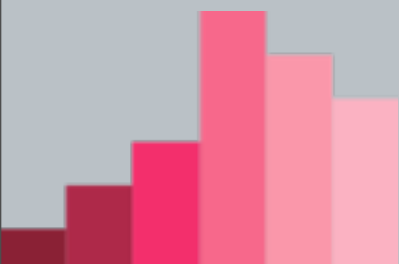
Information

- We can imagine we might not lose information in shifting to v5
- This is born out by the classifier, to some degree.
- We don't lose a lot of power when dichotomizing individual ratings



A Look Ahead

- Really exciting!
- Great compliment to current research methods
- Long exposures can help discover reader/editor divergence
- Predictive analytics
- Need more open data



Questions?

- Of course you have questions!
- All work is or soon will be available on github under a free license
- Full writeup on en-wp forthcoming

