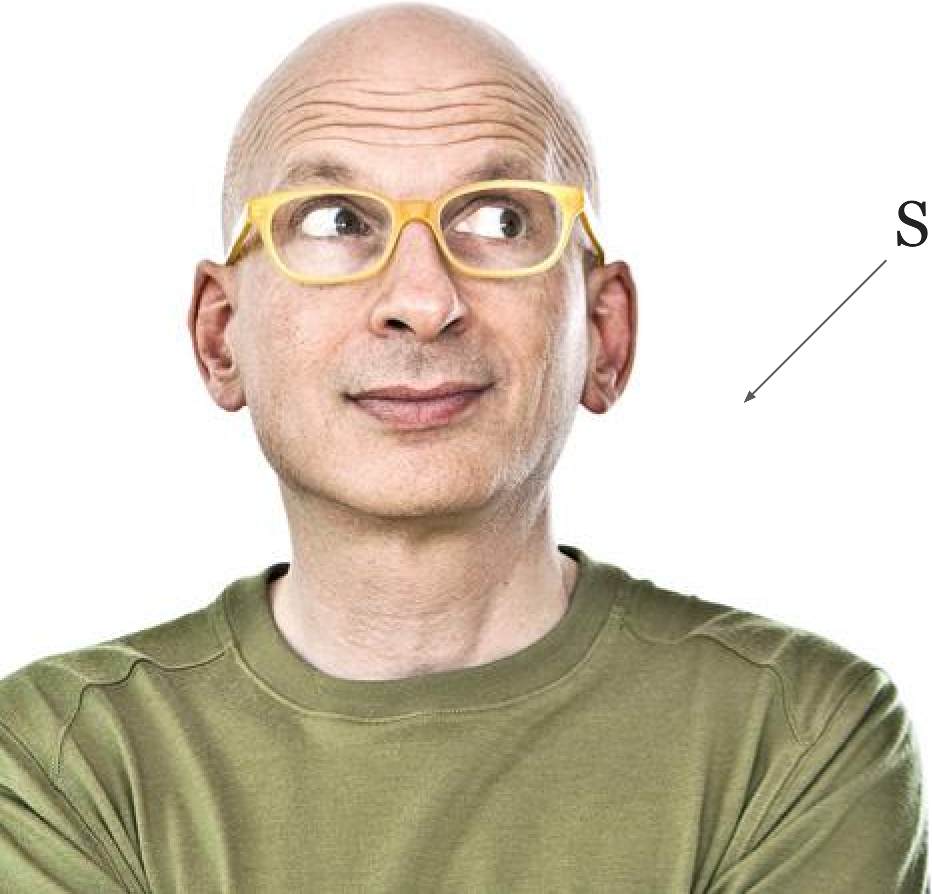
The New York Times Bestseller List is stupid.

Source: http://bookfans.net/wp-content/uploads/images/Seth_Godin_18682.jpg?c3d821

The New York Times Bestseller List is stupid.



Seth Godin

Source: http://bookfans.net/wp-content/uploads/images/Seth_Godin_18682.jpg?c3d821

Is Seth Godin right?

How might I find out?

Wikipedia

https://www.wikipedia.org/

Lists of #1 NYT bestsellers

1942-2016

Scraped using Beautiful Soup

Goodreads

https://www.goodreads.com/

APIs

- General Search
- Author
- Book

Parsed xml using ElementTree

Totals

75 years

3900 weeks

746 unique books

241 unique authors

Numeric Features

Books

- Number of ratings
- Number of reviews

Authors

- Number of ratings
- Average rating
- Number of reviews
- Number of fans
- Number of books as #1
- Number of weeks as #1

Categorical Features

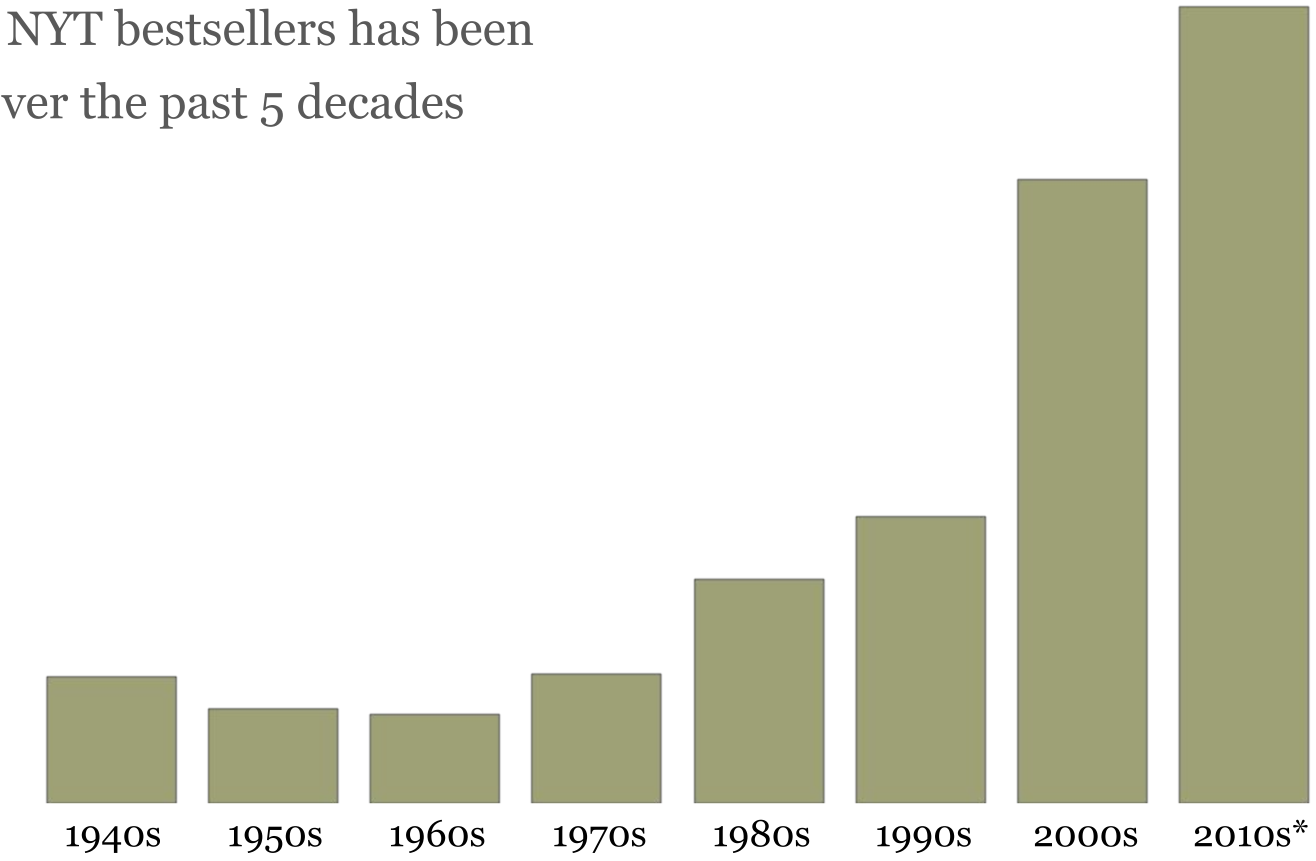
Books

- Repeat books

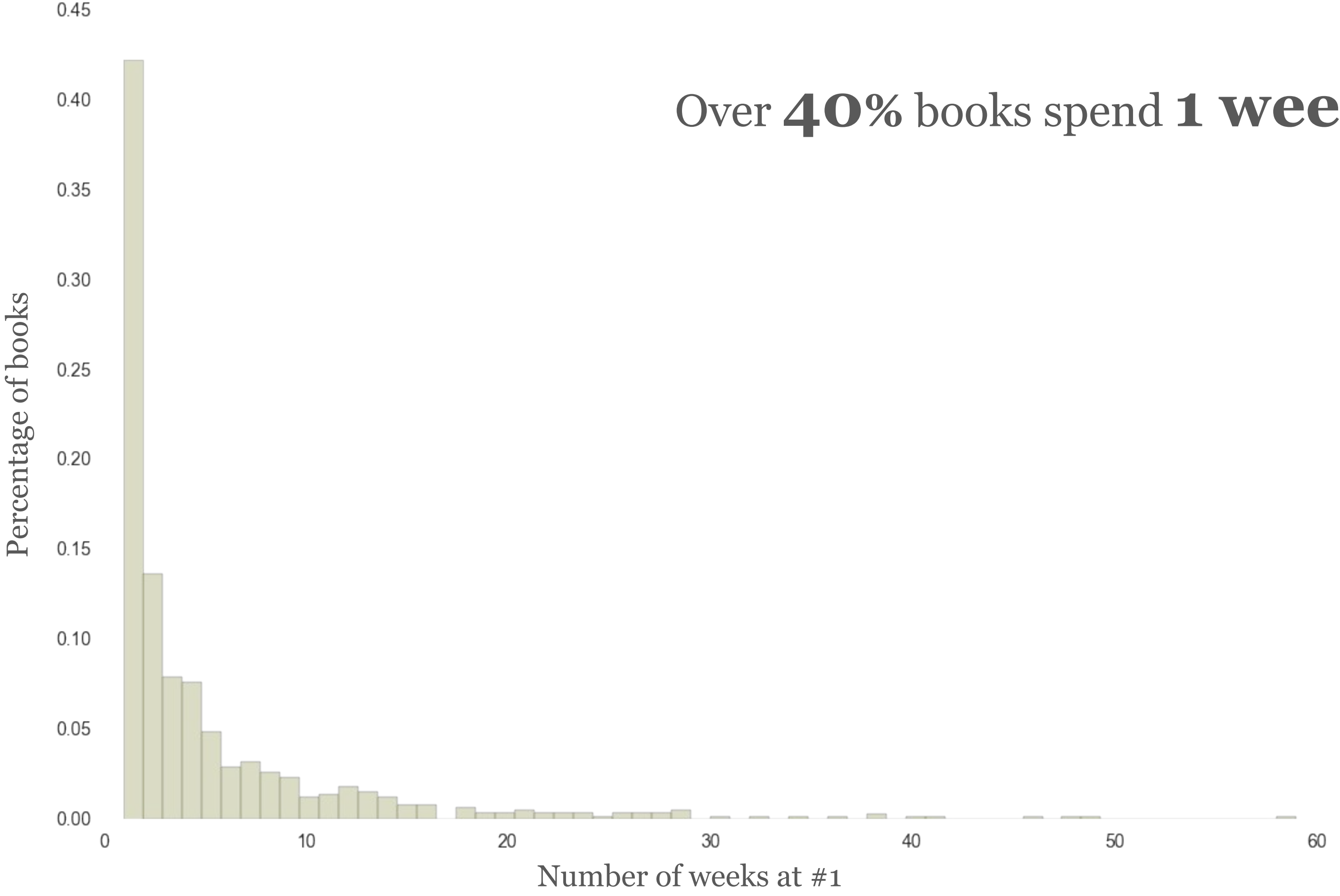
Authors

- Repeat author
- Hometown in NY
- Gender

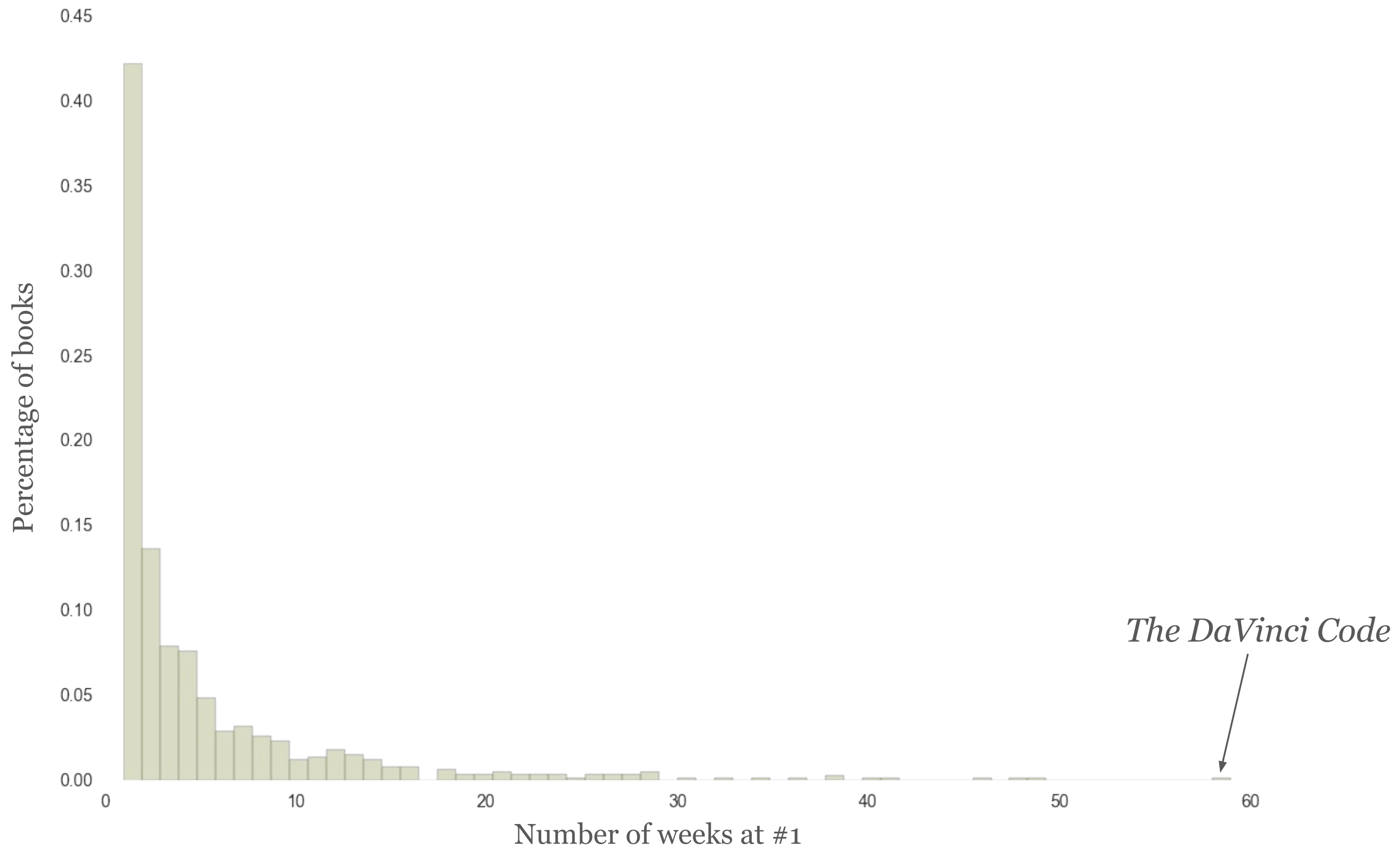
The number of #1 NYT bestsellers has been **increasing** over the past 5 decades



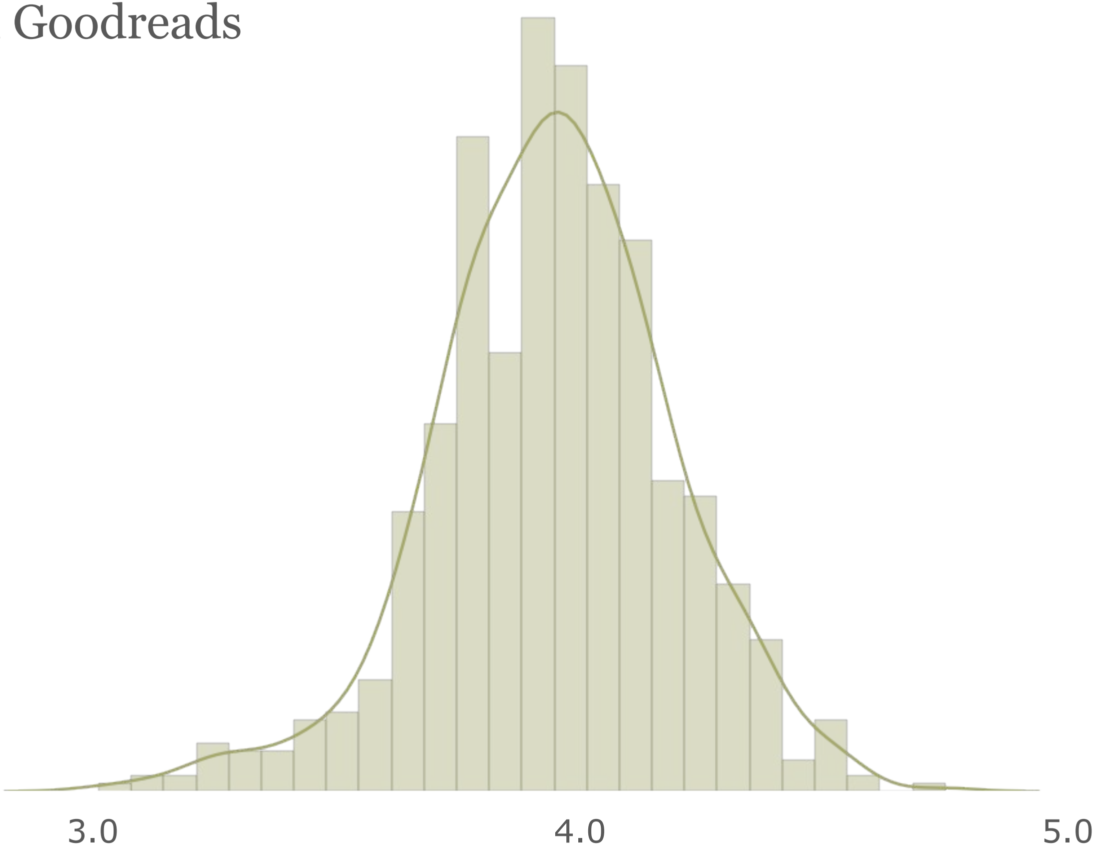
*projected based on data through 2016



Over **40%** books spend **1 week** at #1

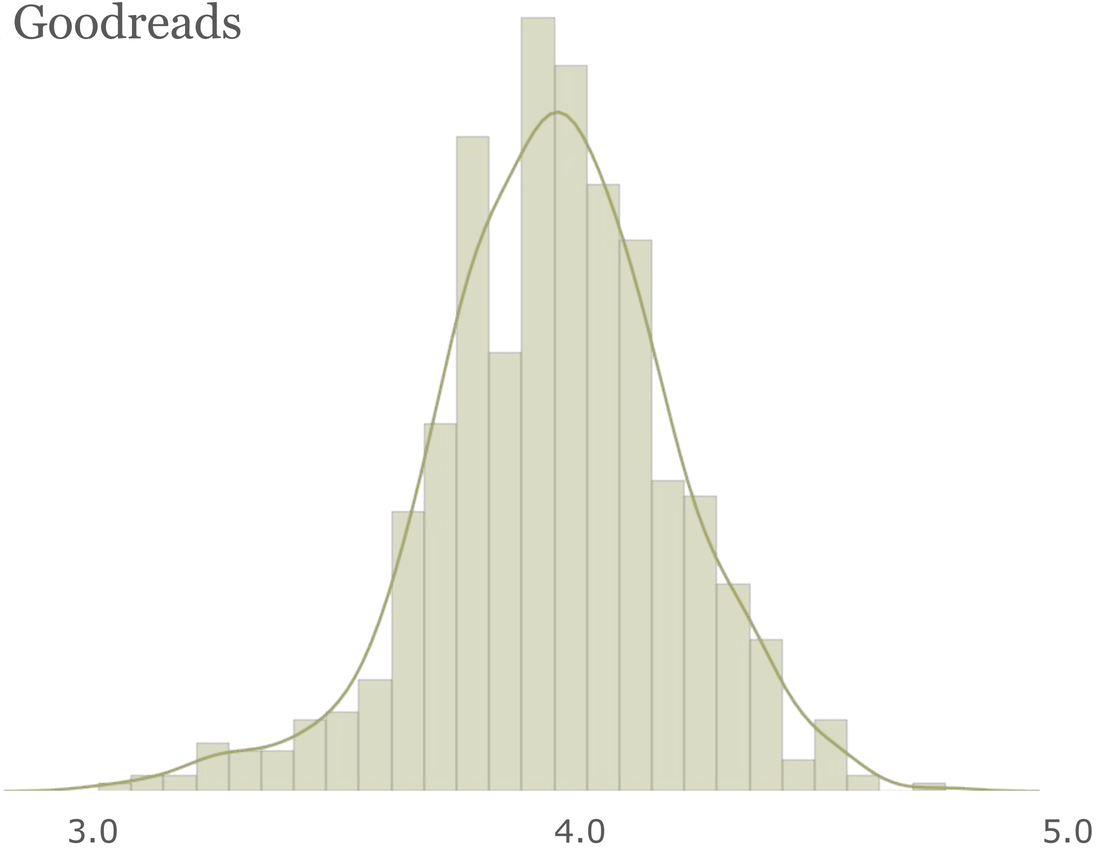

0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Percentage of books
0
10
20
30
40
50
60
Number of weeks at #1
The DaVinci Code

Mean rating for a **book** on Goodreads

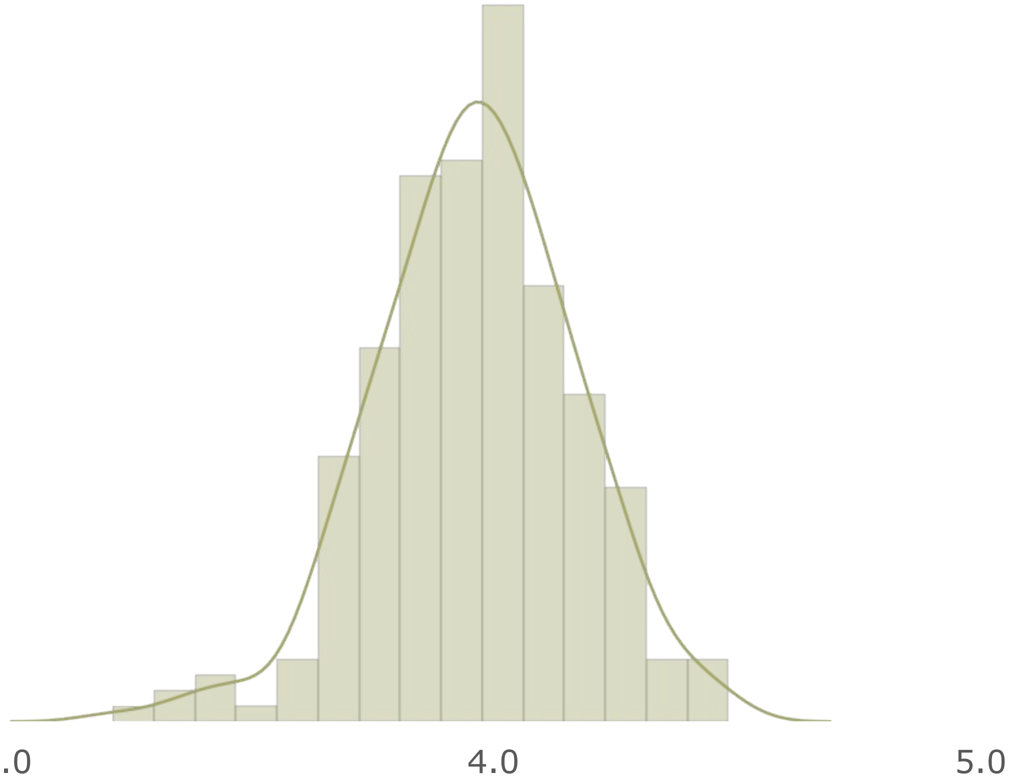


Mean rating for a **book** on Goodreads

3·95

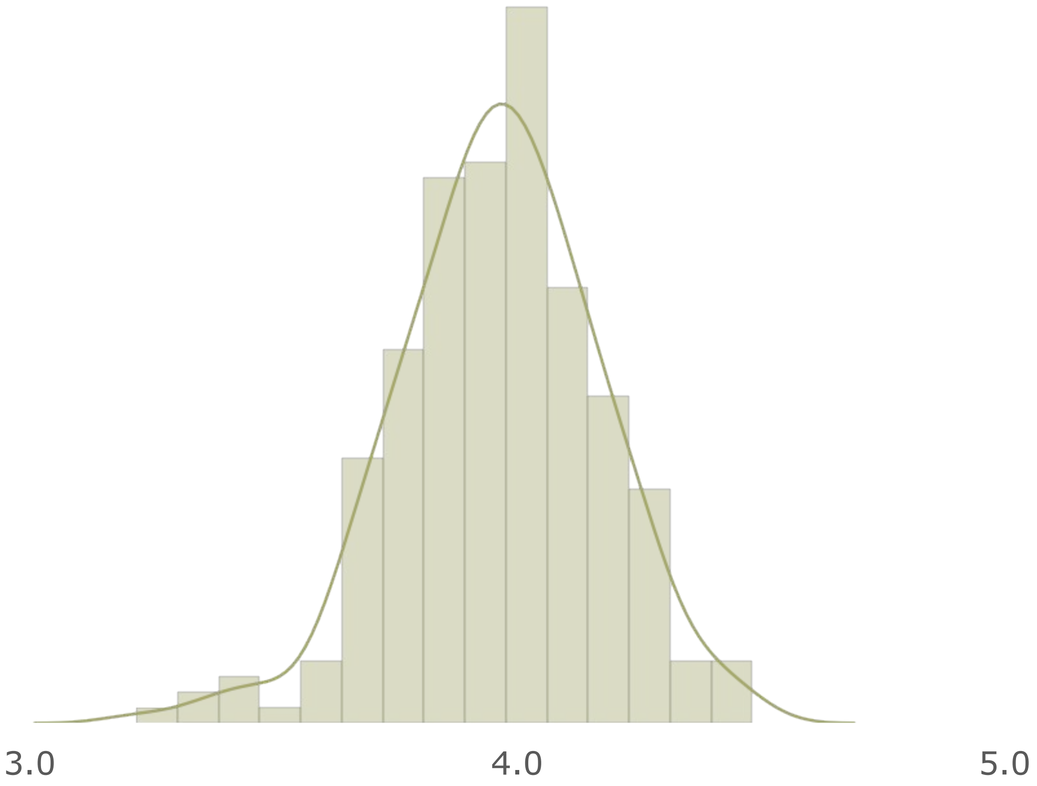


Mean rating for an **author** on Goodreads

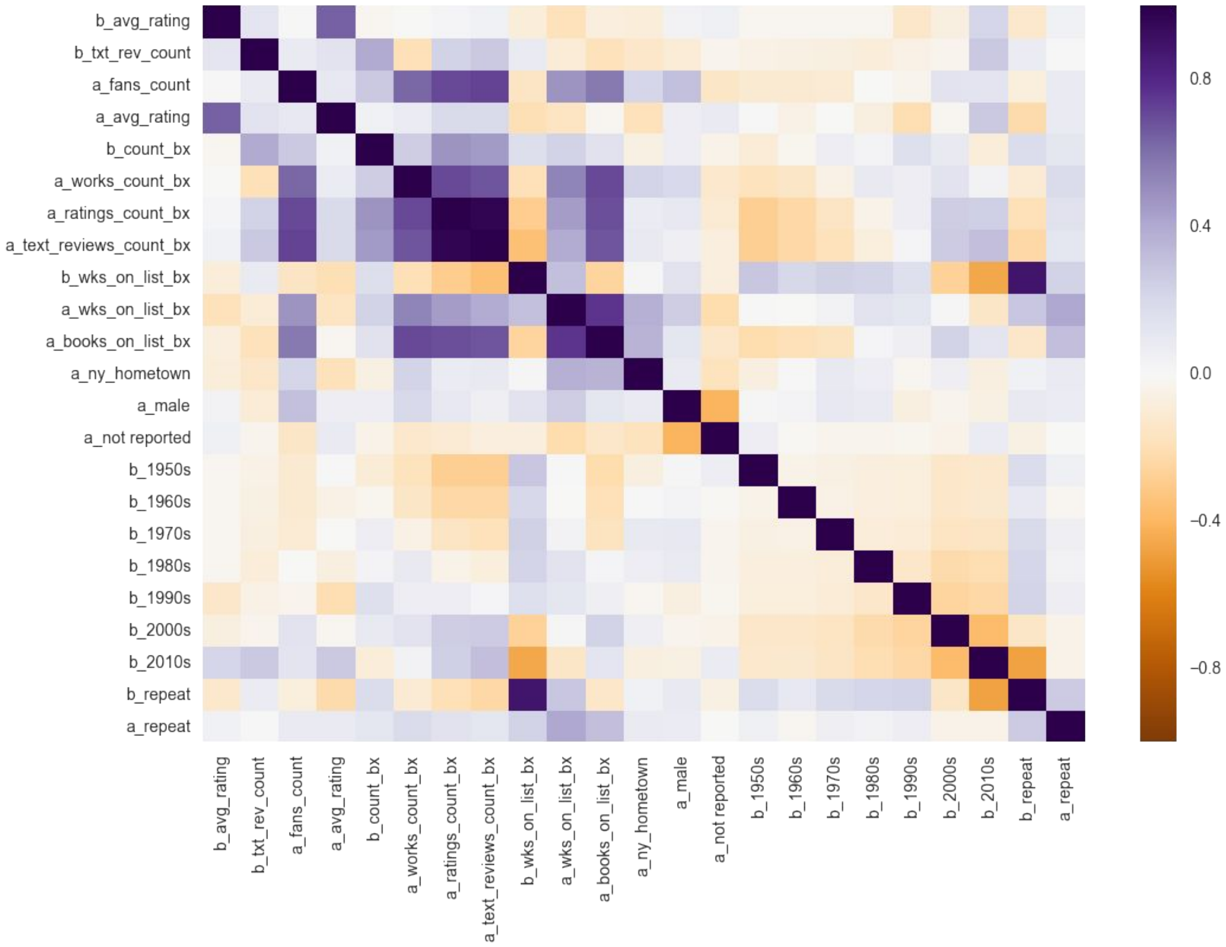


Mean rating for an **author** on Goodreads

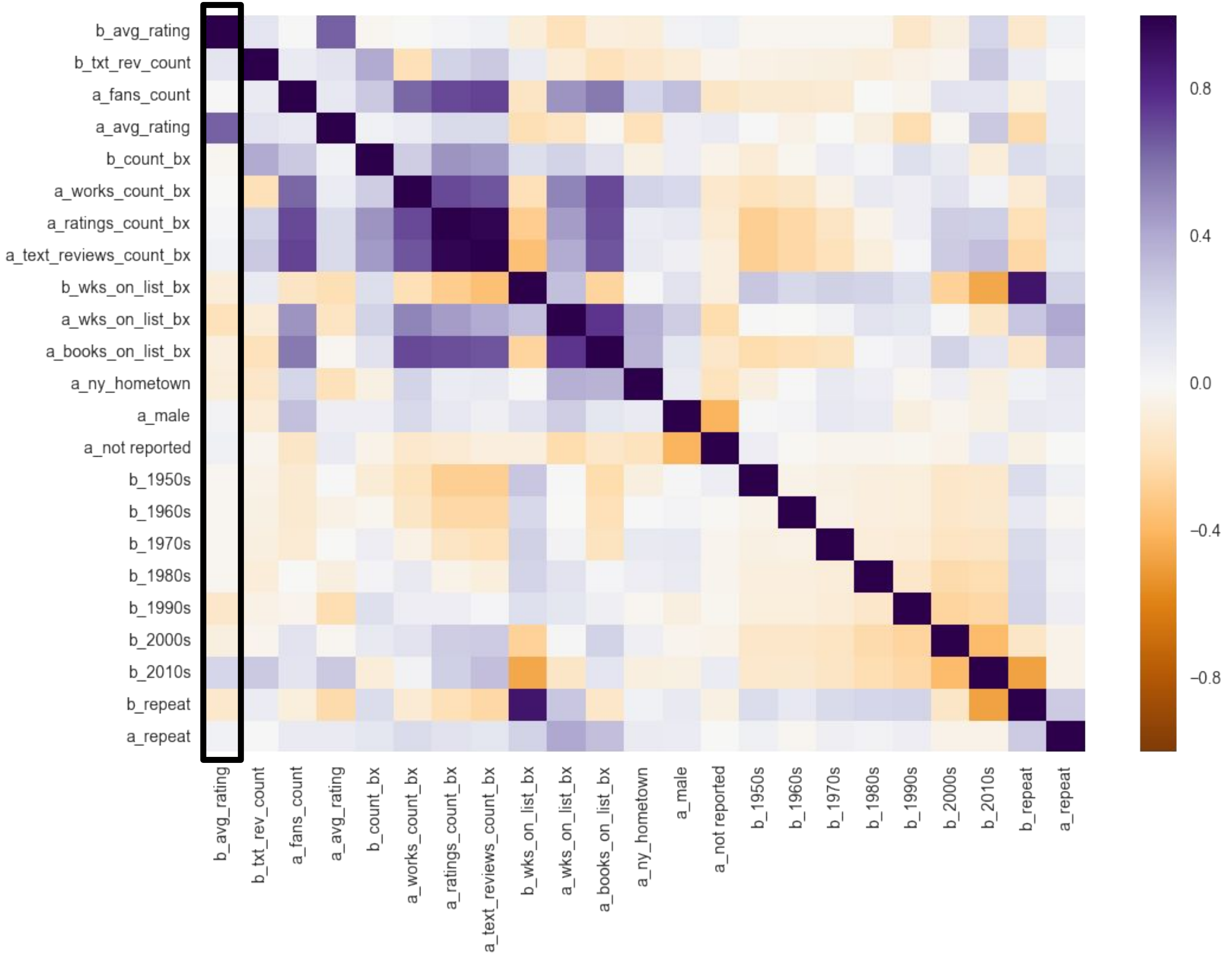
3.96



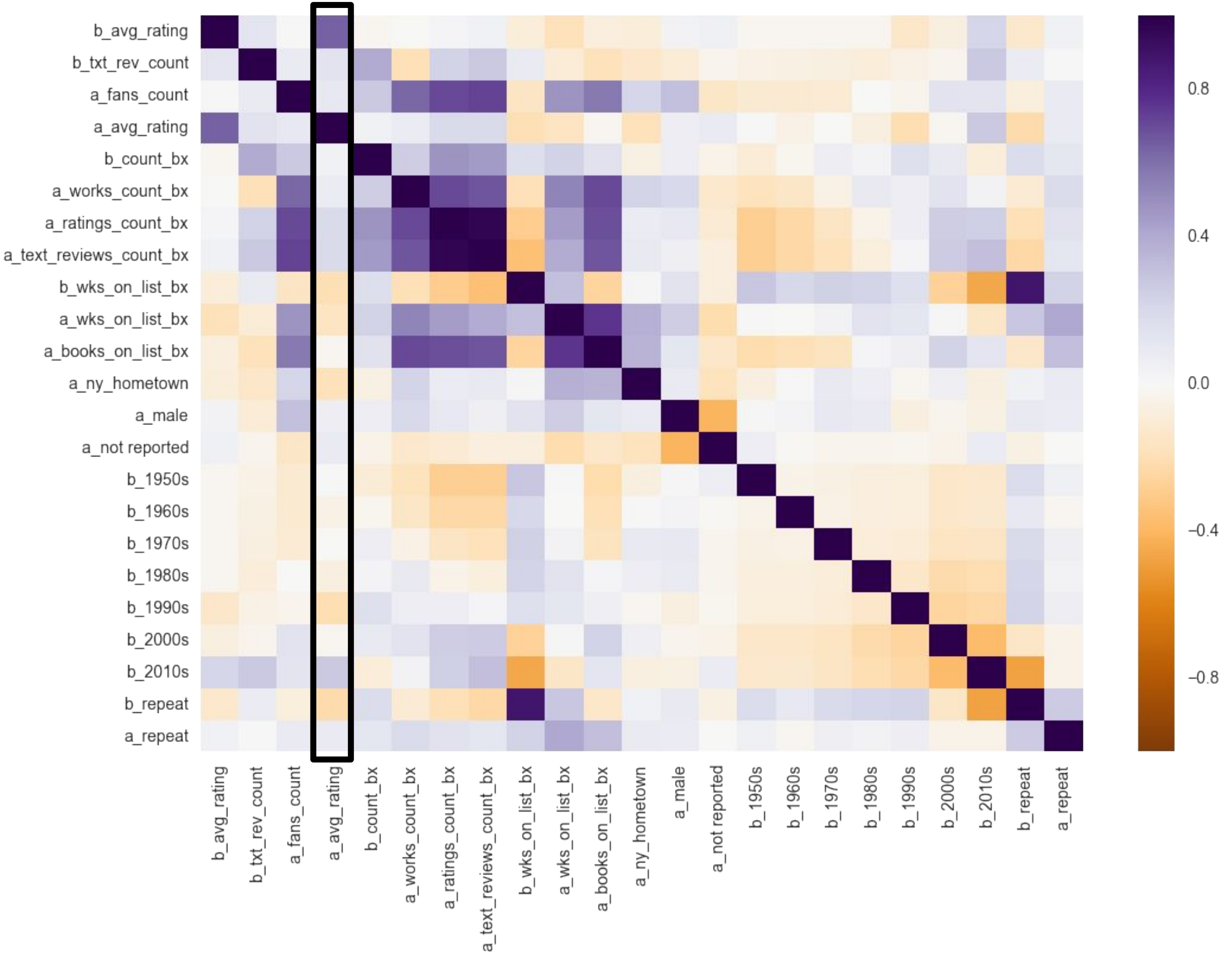
What might I predict?



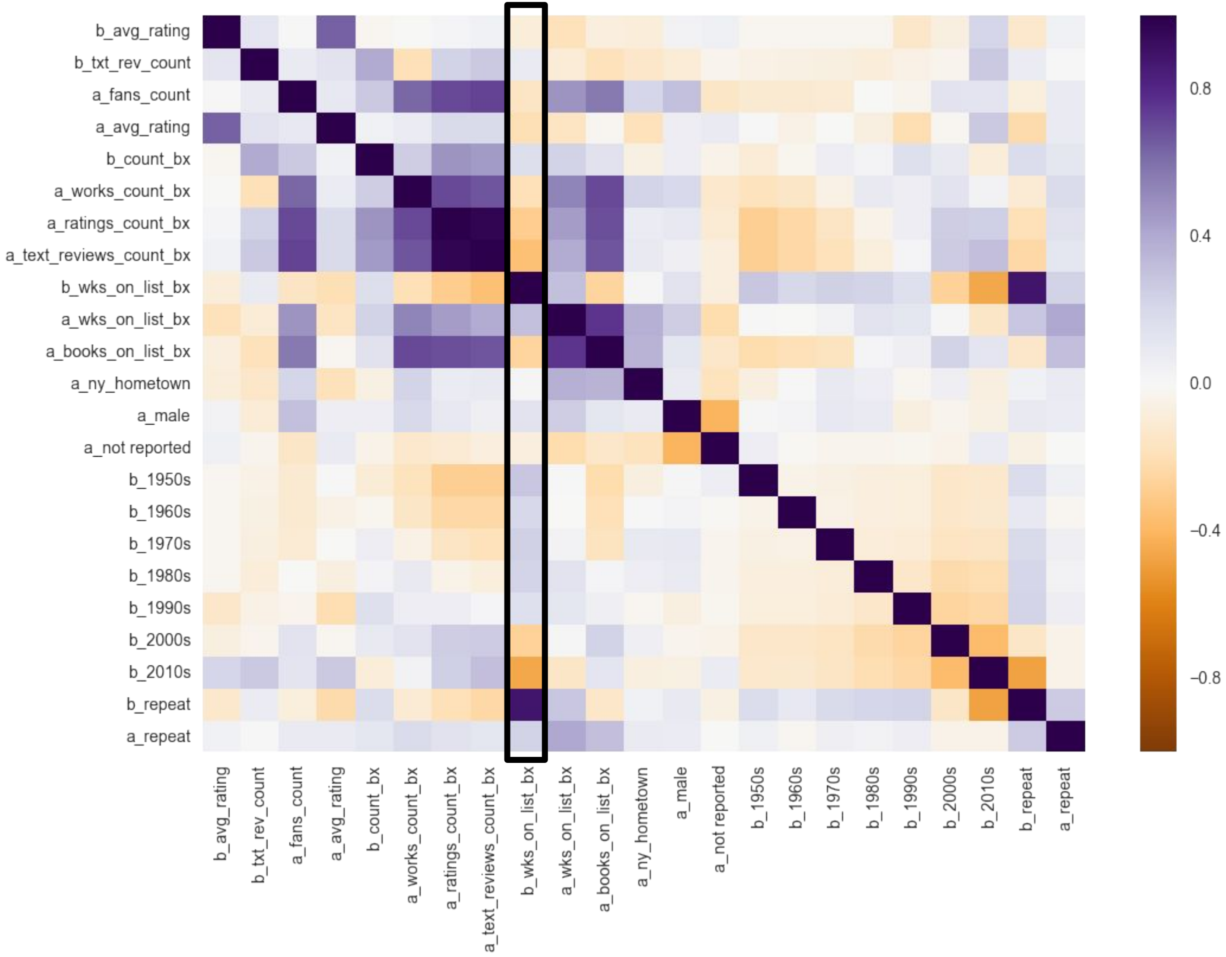
Correlation plot



Book’s average rating



Author's average rating



Weeks as #1 best seller

Let’s try to predict a book’s rating

Regression Results

Model		
Adjusted R^2	0.453	Low
AIC	1670	High
BIC	1700	High
Residuals		
Omnibus P(Omnibus)	137 0	0 probability random residuals
Skew	-0.925	Left
Kurtosis	5.97	2x normal

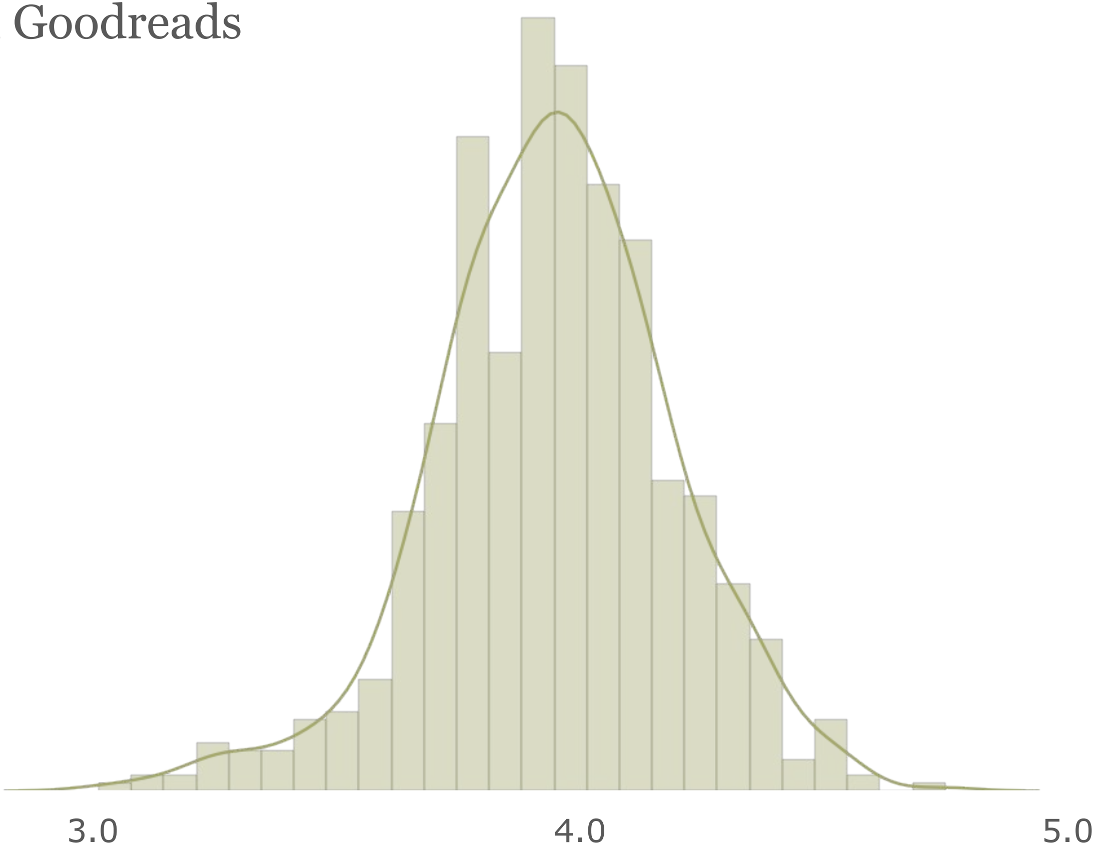
Cross Validation

Root Mean Squared Errors

Linear Regression:	0.746
Lasso:	0.746
Ridge:	0.746
Elastic Net:	0.746

Mean rating for a **book** on Goodreads

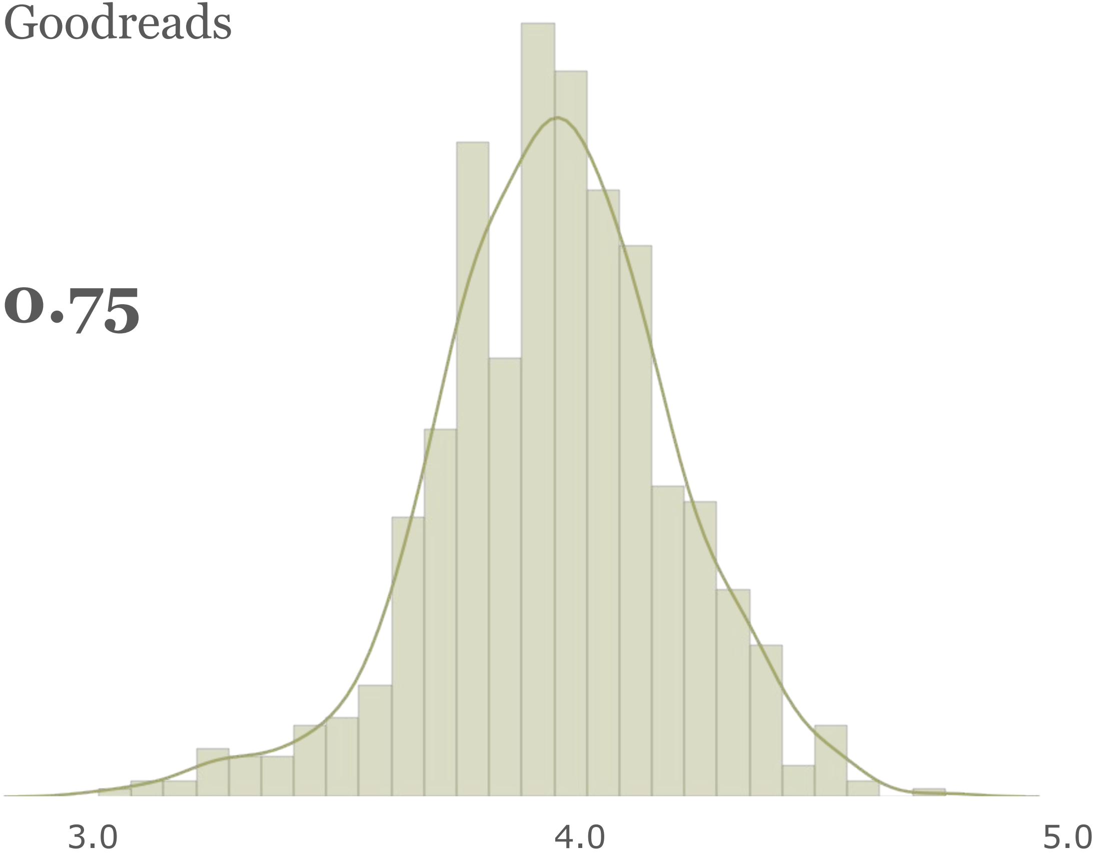
3·95



Mean rating for a book on Goodreads

3.95

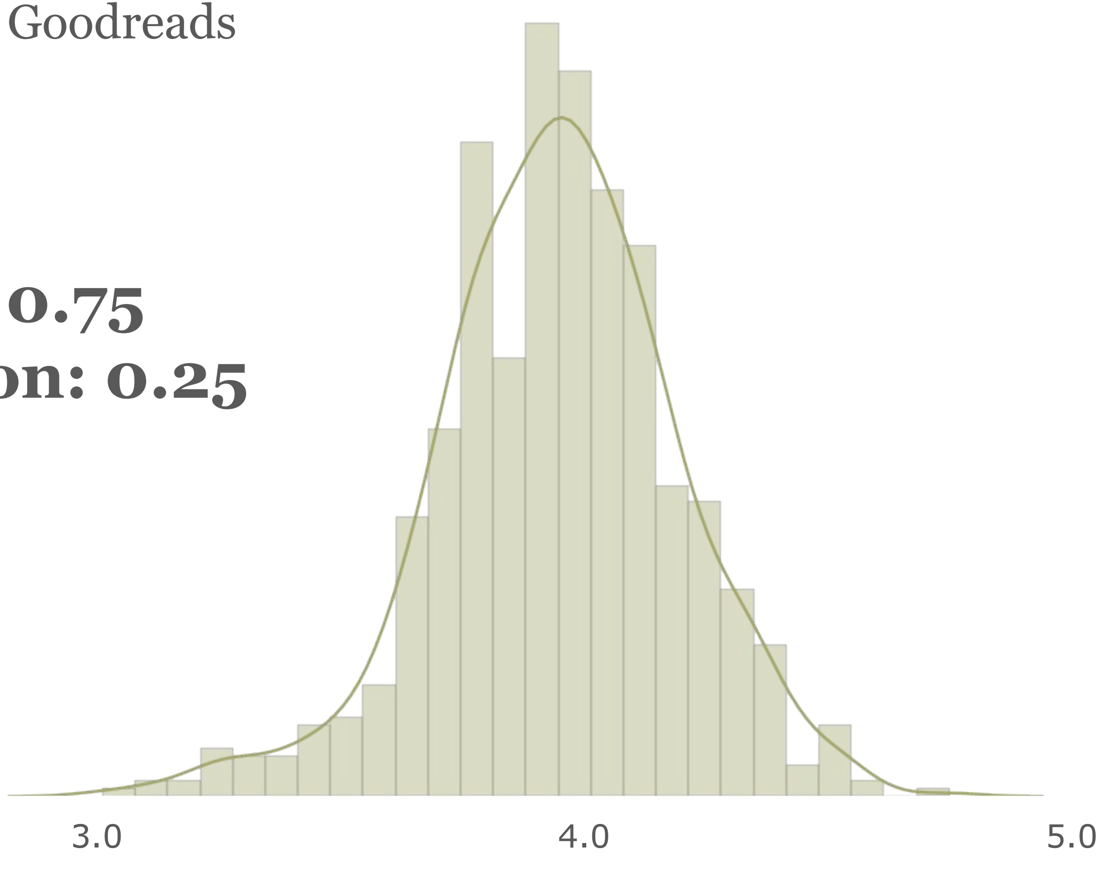
Model RMSE: 0.75

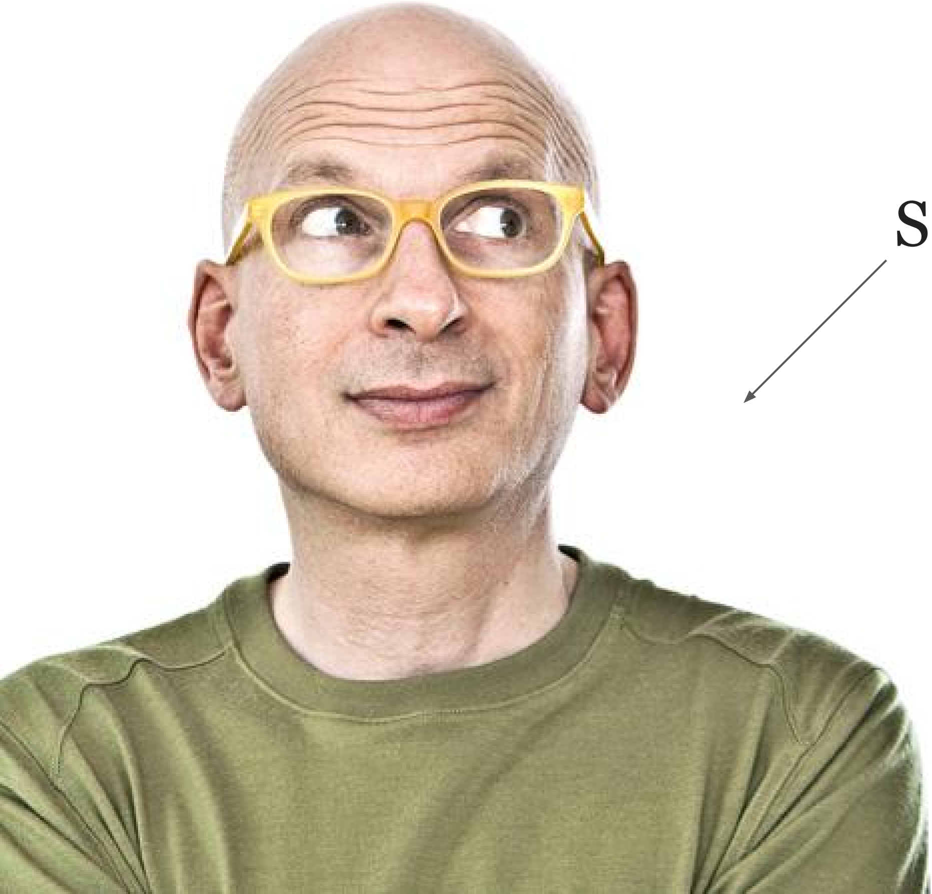


Mean rating for a book on Goodreads

3.95

Model RMSE: 0.75
Standard deviation: 0.25



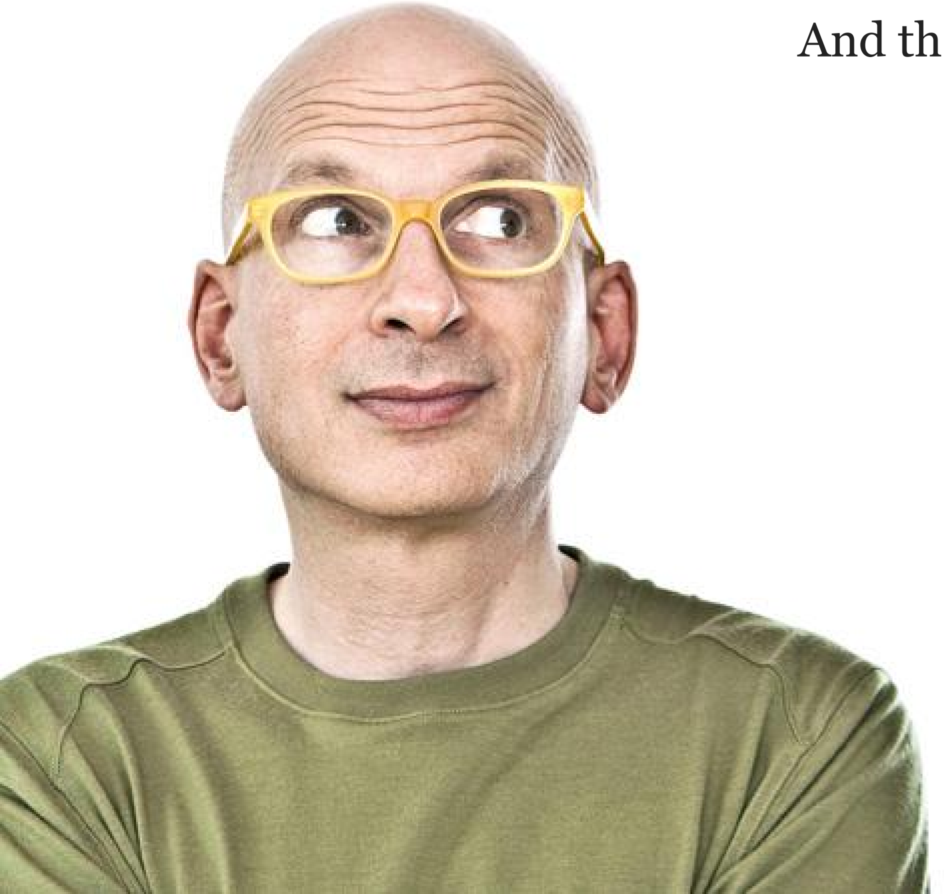


Seth Godin

Source: http://bookfans.net/wp-content/uploads/images/Seth_Godin_18682.jpg?c3d821

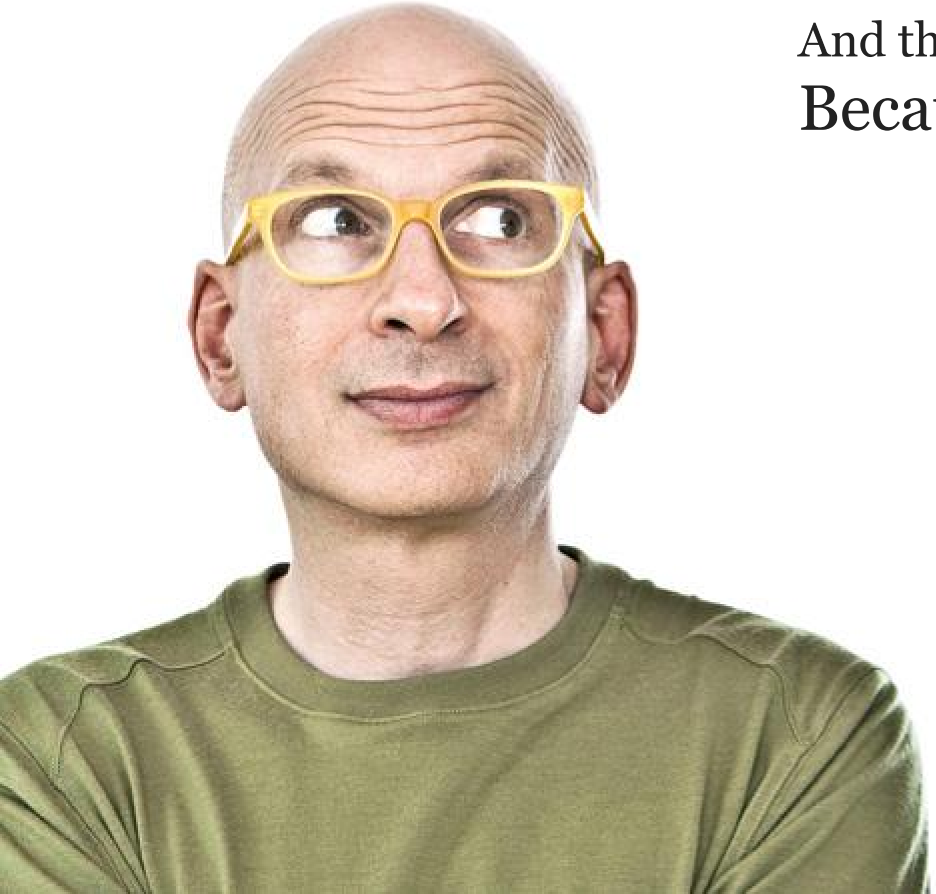
The New York Times Bestseller List is stupid.

Source: http://bookfans.net/wp-content/uploads/images/Seth_Godin_18682.jpg?c3d821



The New York Times Bestseller List is stupid.
And they should stop publishing it.

Source: http://bookfans.net/wp-content/uploads/images/Seth_Godin_18682.jpg?c3d821



The New York Times Bestseller List is stupid.
And they should stop publishing it.
Because it doesn’t mean anything.

Source: http://bookfans.net/wp-content/uploads/images/Seth_Godin_18682.jpg?c3d821