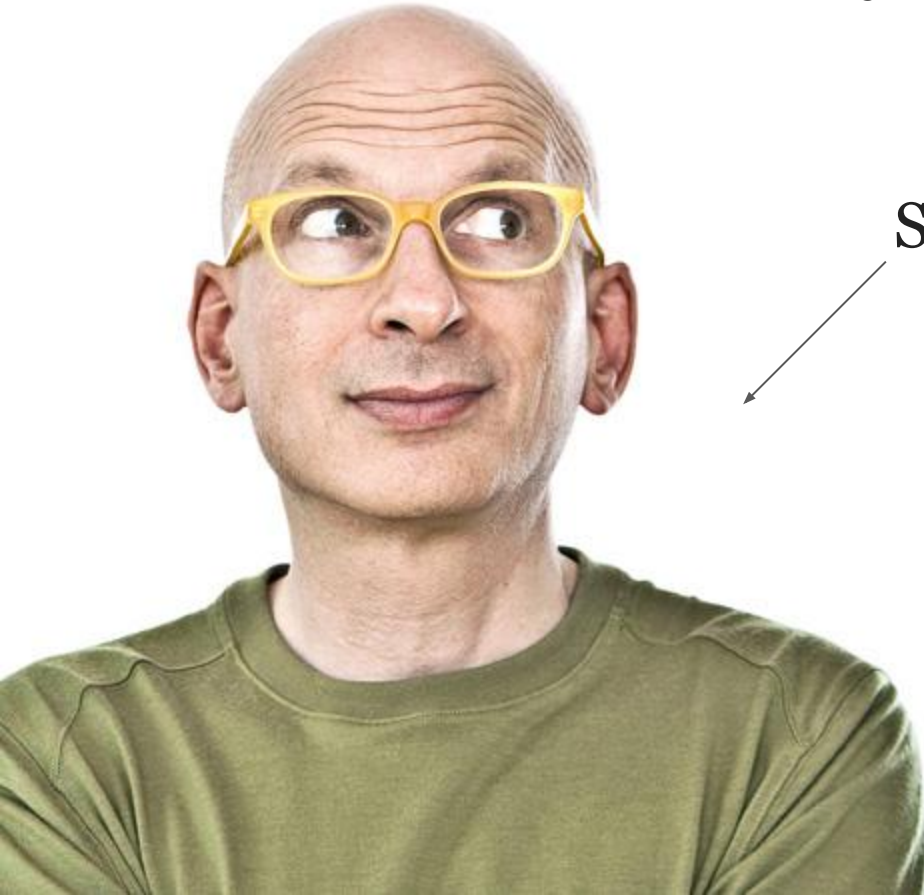


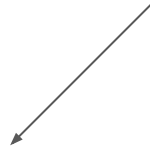
The New York Times Bestseller List is stupid.



The New York Times Bestseller List is stupid.



Seth Godin



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

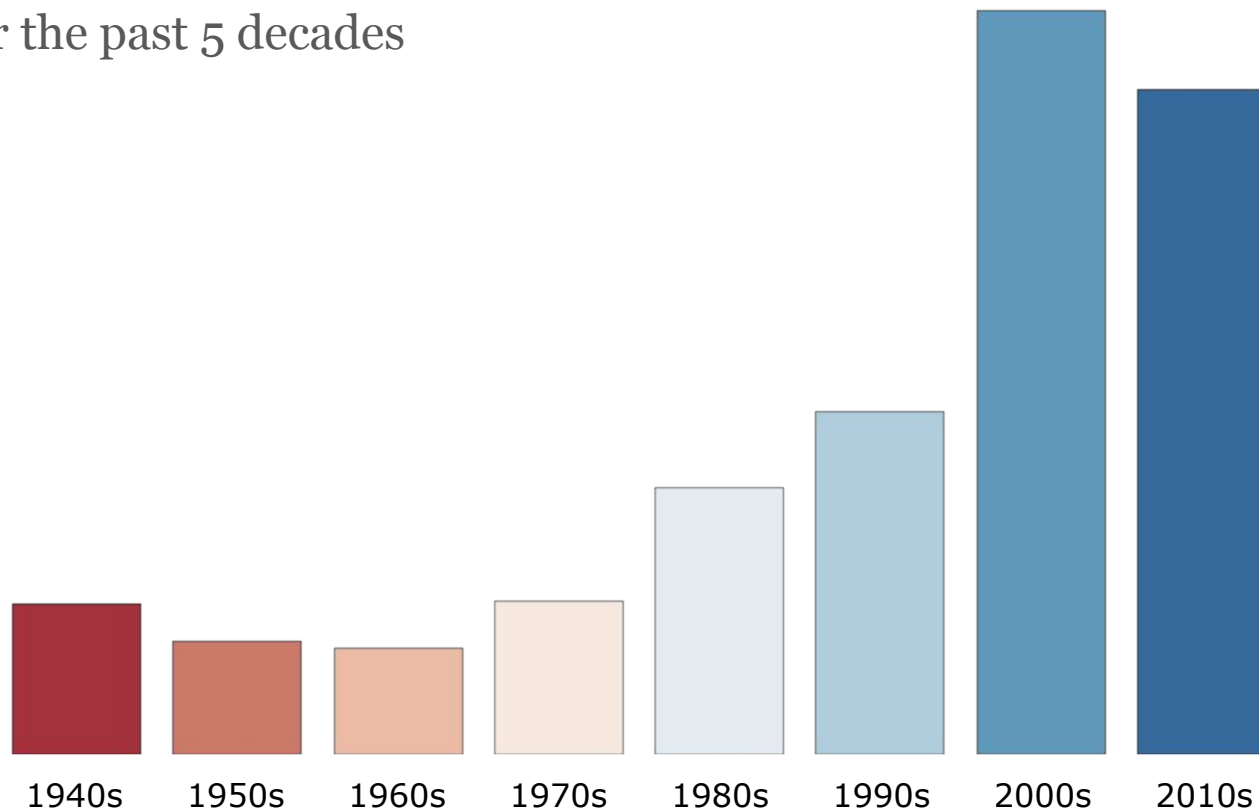
75 years

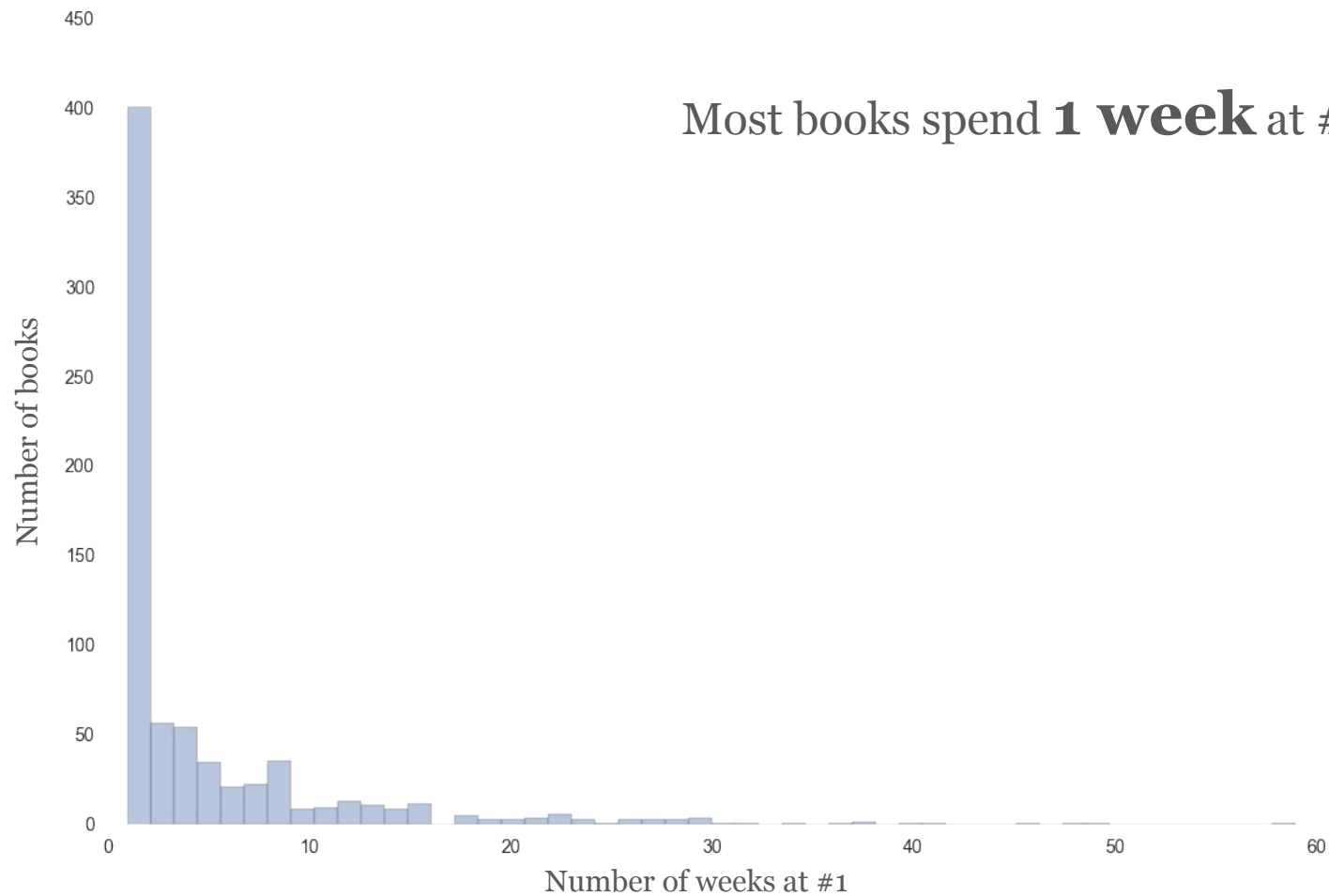
3900 weeks

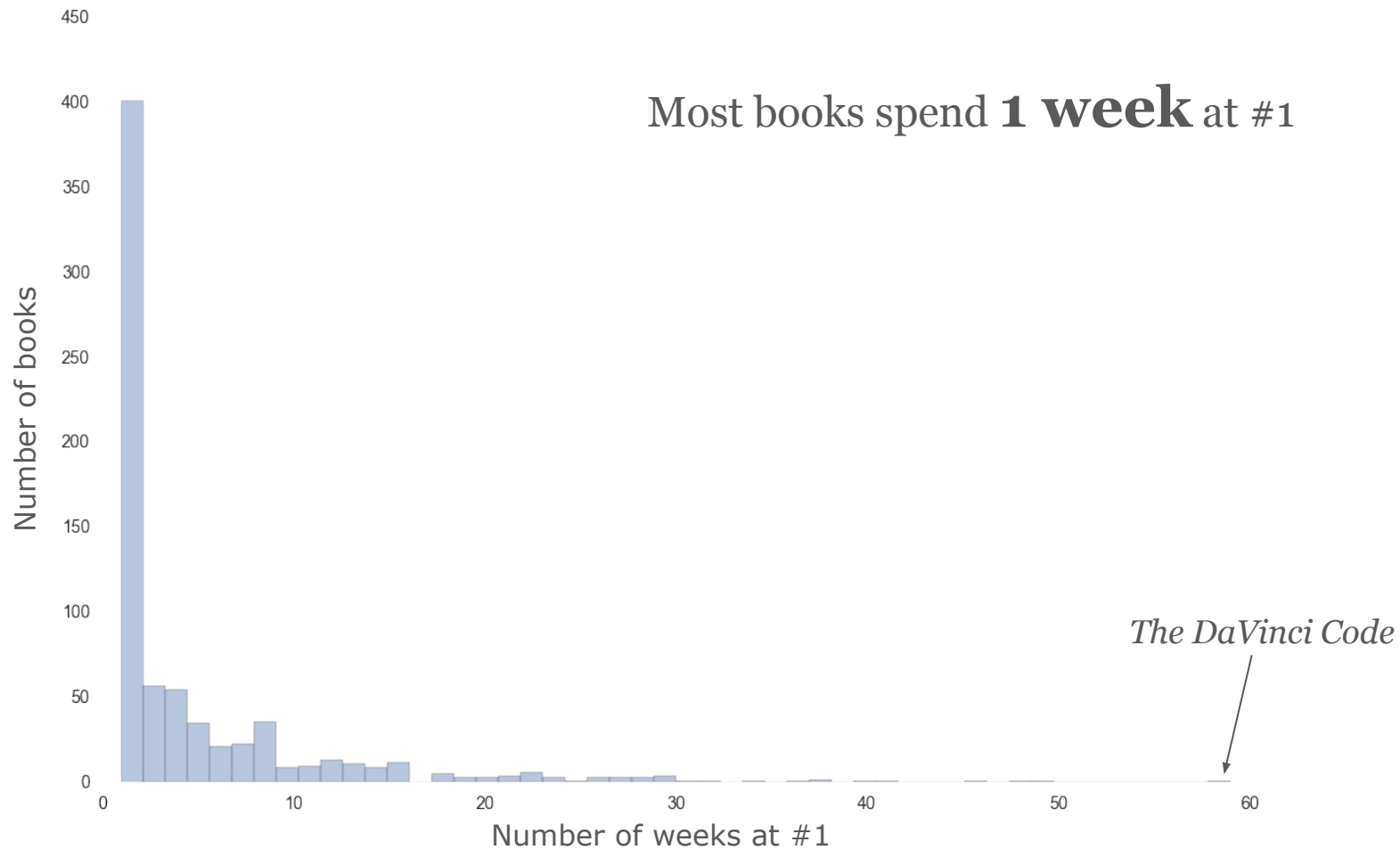
746 unique books

241 unique authors

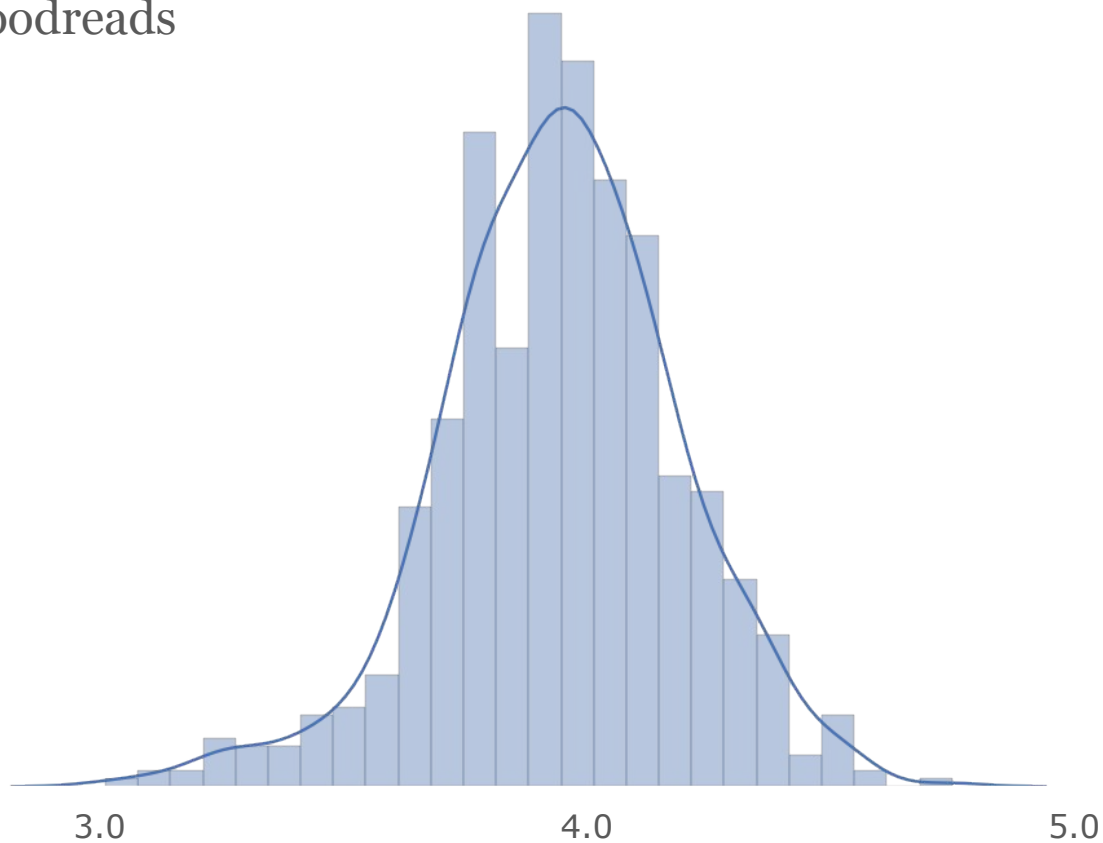
The number of #1 NYT bestsellers has been **increasing** over the past 5 decades





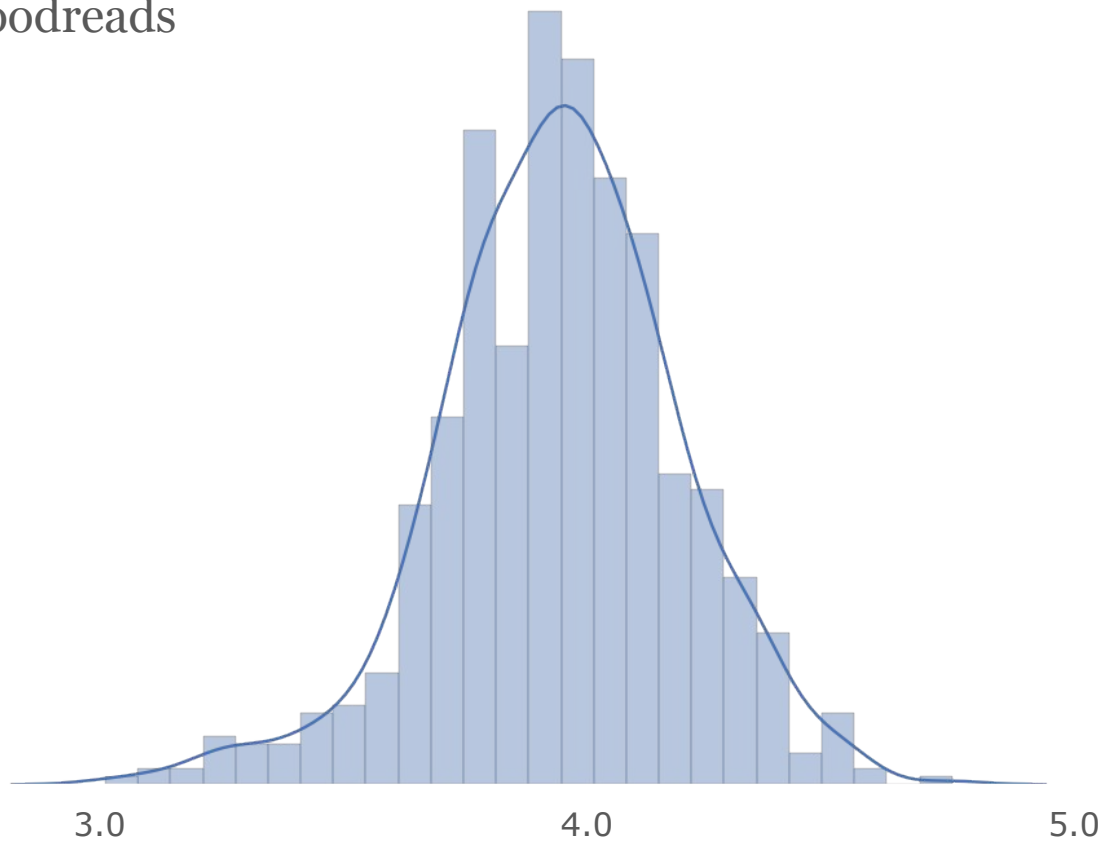


Mean rating for a **book** on Goodreads

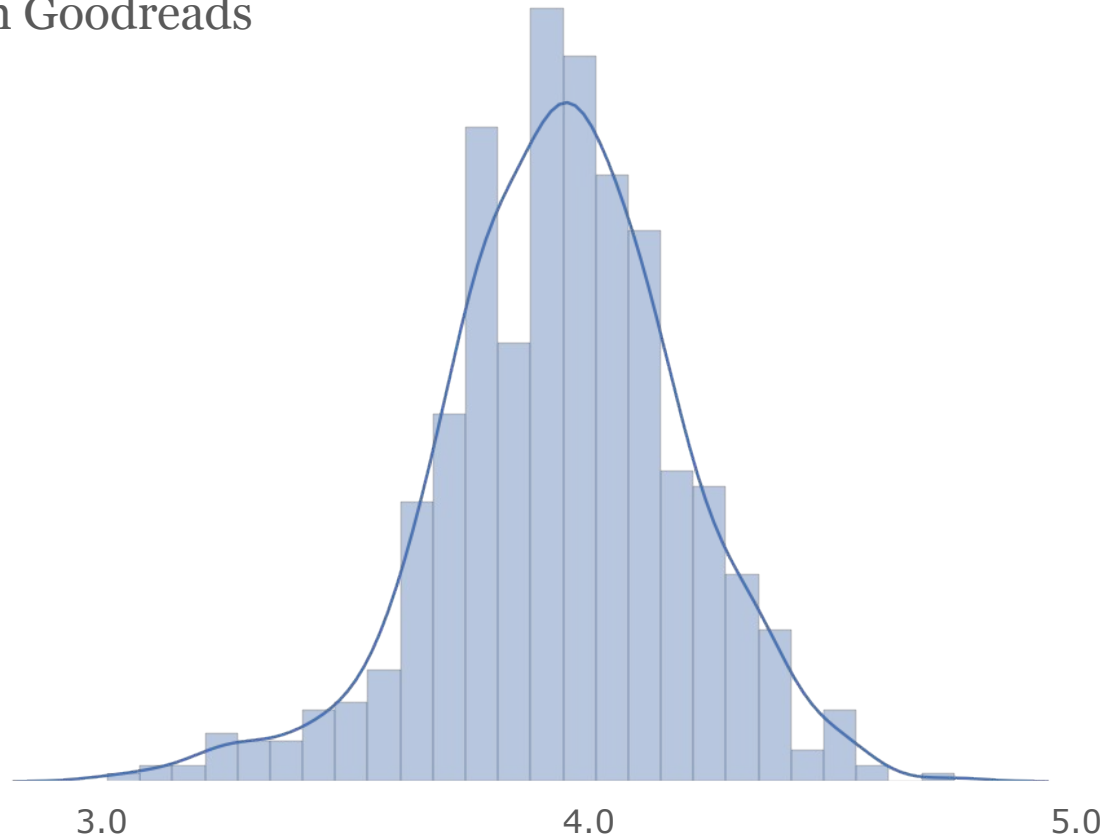


Mean rating for a **book** on Goodreads

3.95

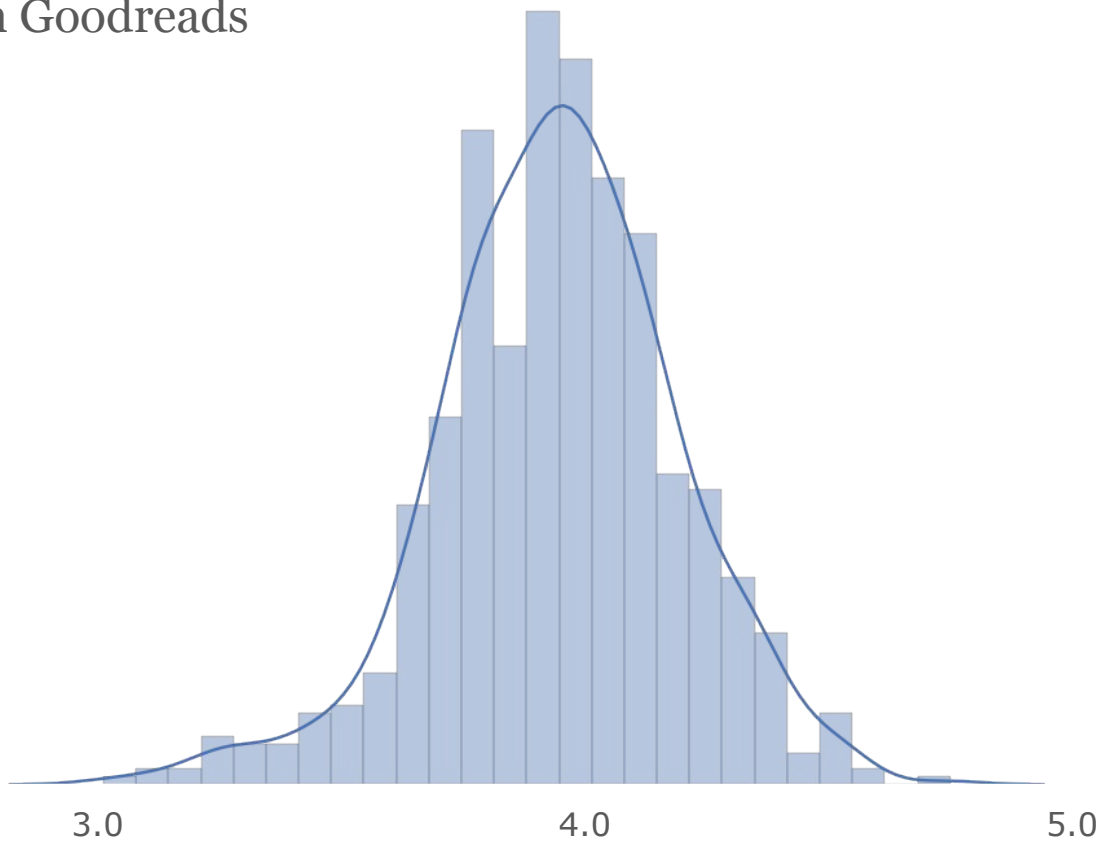


Mean rating for an **author** on Goodreads

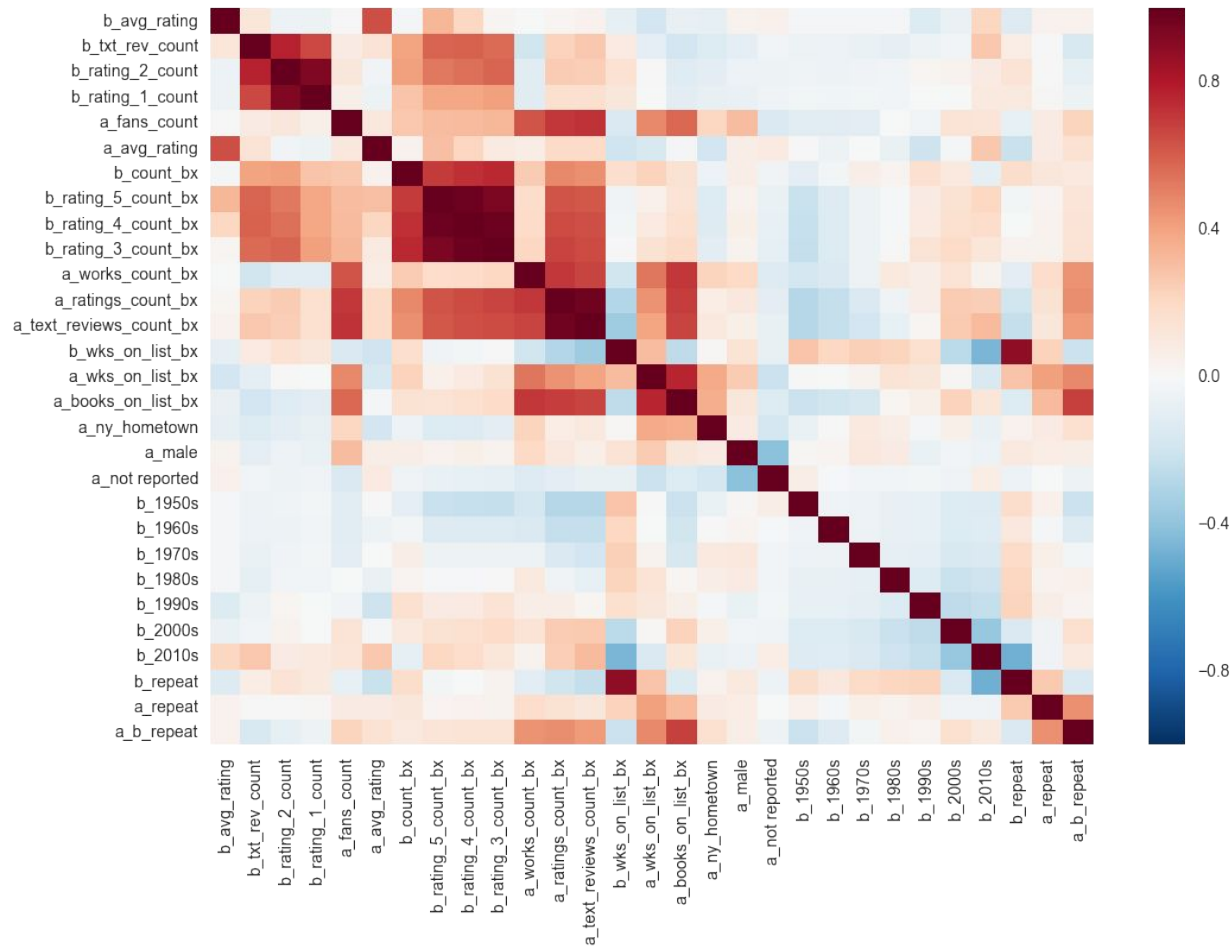


Mean rating for an **author** on Goodreads

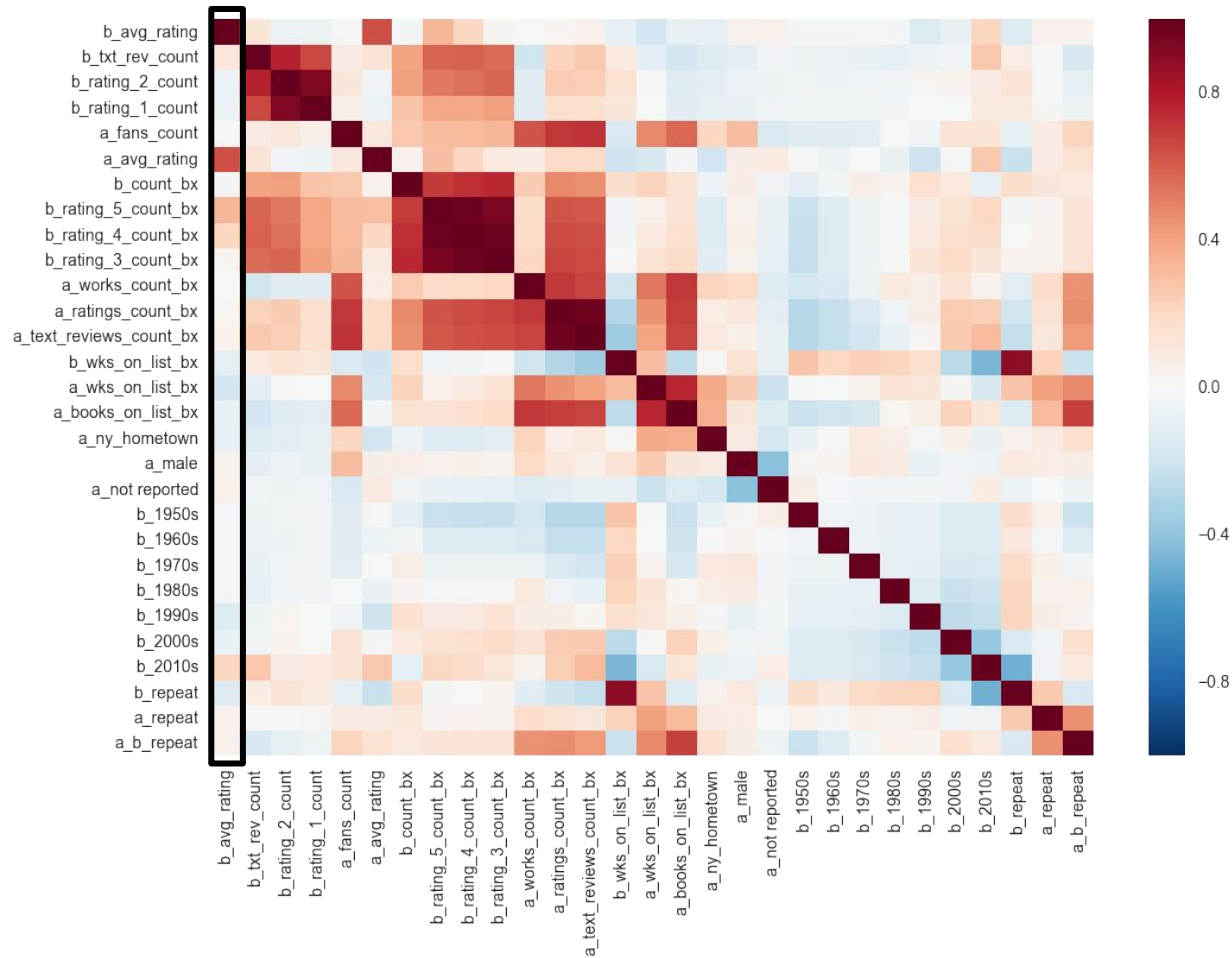
3.96



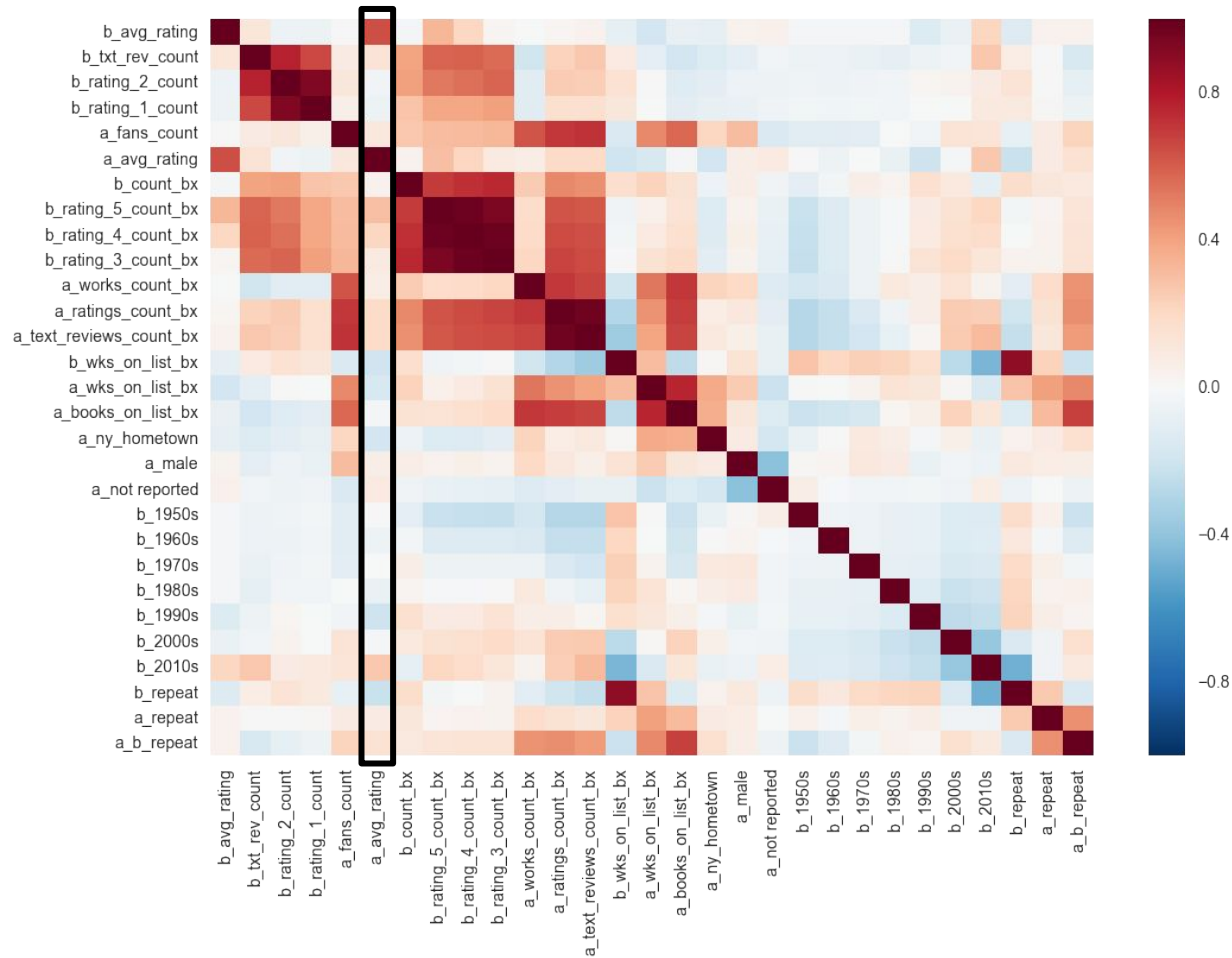
What might I predict?

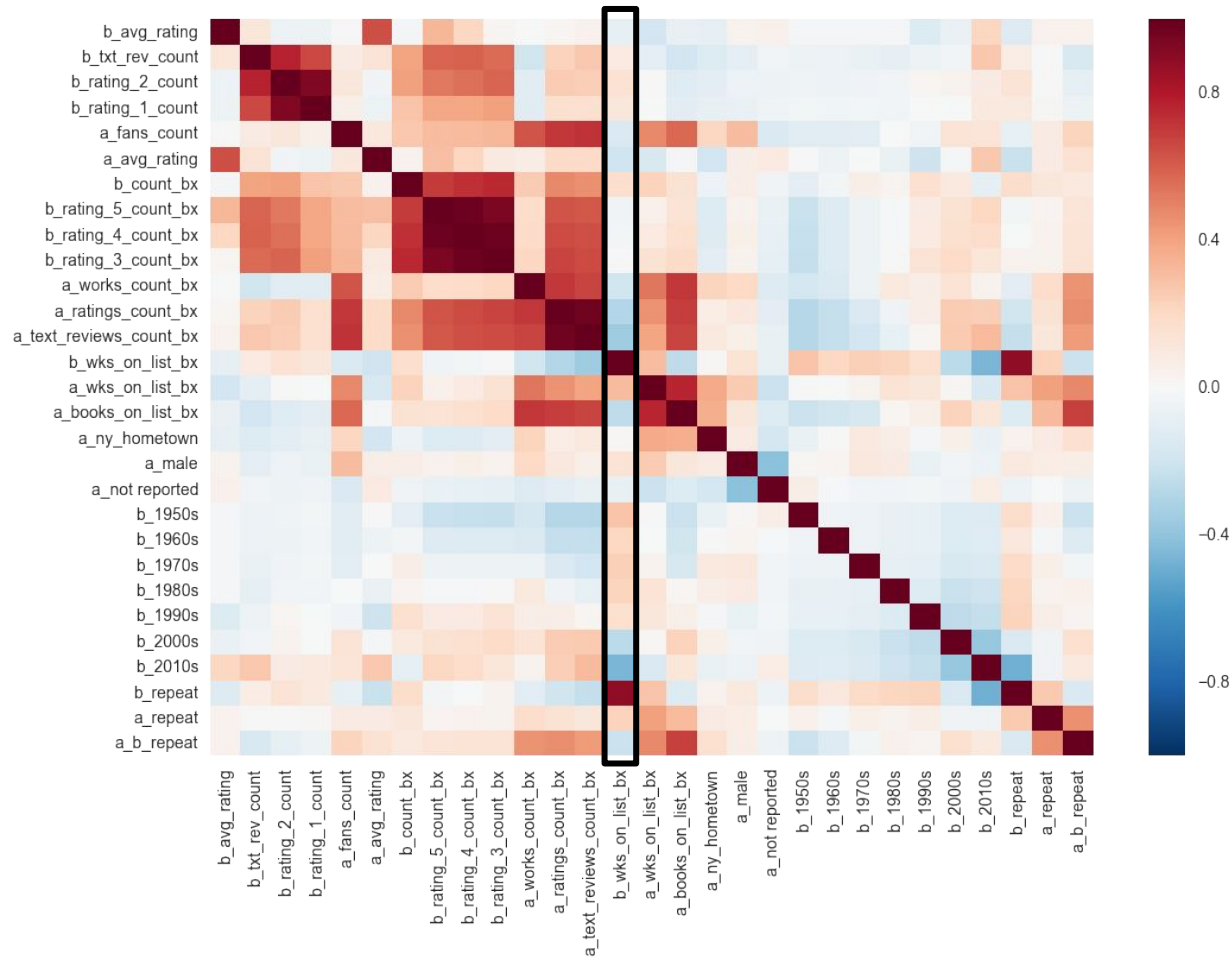


Correlation plot



Book's average rating





Weeks as #1 best seller

Let's try to predict a
book's rating

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

Regression Results

Model		
Adjusted R ²	0.441	Low
AIC	1694	High
BIC	1786	High
Residuals		
Omnibus P(Omnibus)	139.2 0	0 probability random residuals
Skew	-0.941	Left
Kurtosis	5.979	2x normal

Cross Validation

Root Mean Squared Errors

Linear Regression: 0.39

Lasso: 0.33

Ridge: 0.36

Elastic Net: 0.35

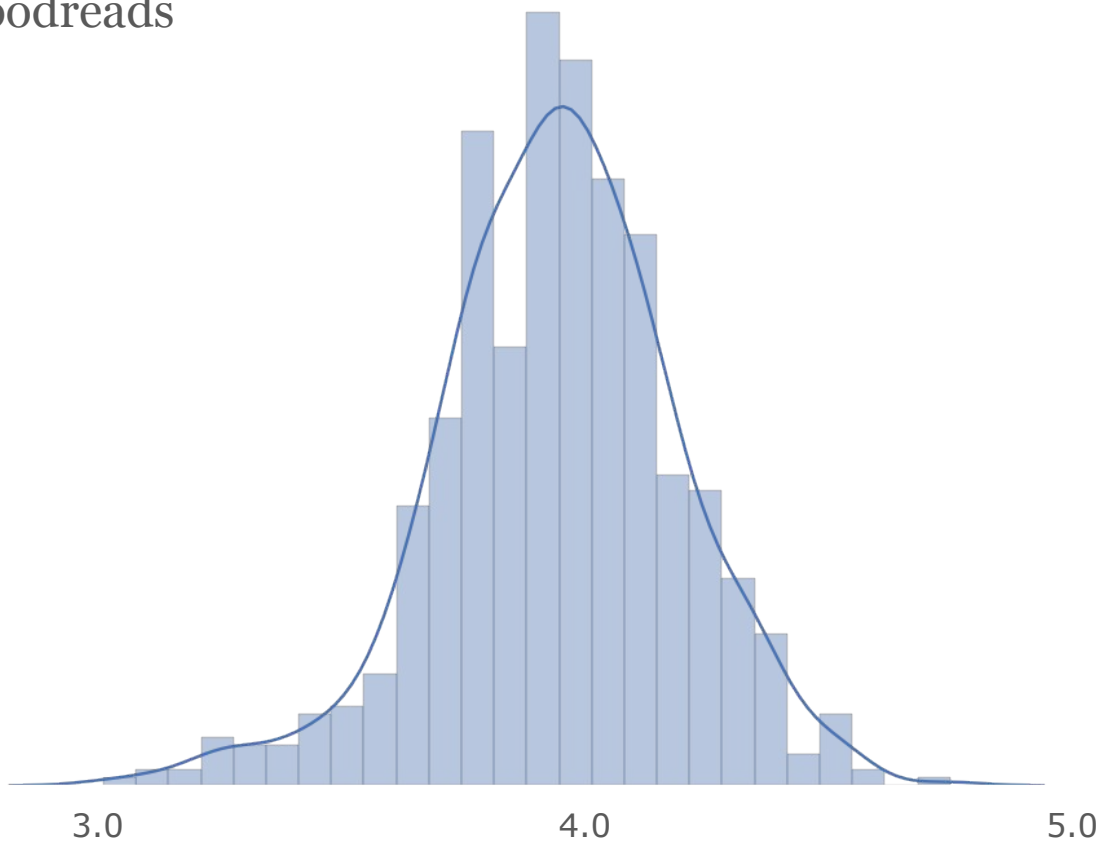
Cross Validation

Root Mean Squared Errors

Linear Regression:	0.39
Lasso:	0.33
Ridge:	0.36
Elastic Net:	0.35

Mean rating for a **book** on Goodreads

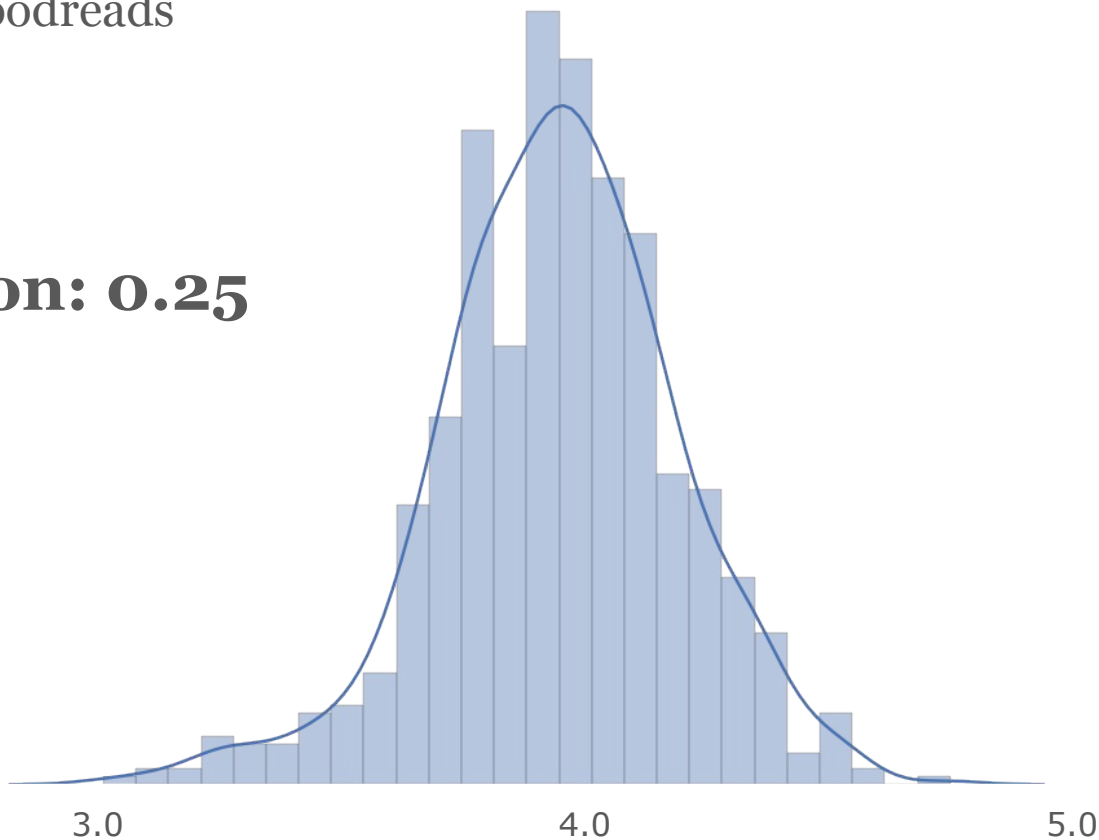
3.95



Mean rating for a **book** on Goodreads

3.95

Standard deviation: 0.25

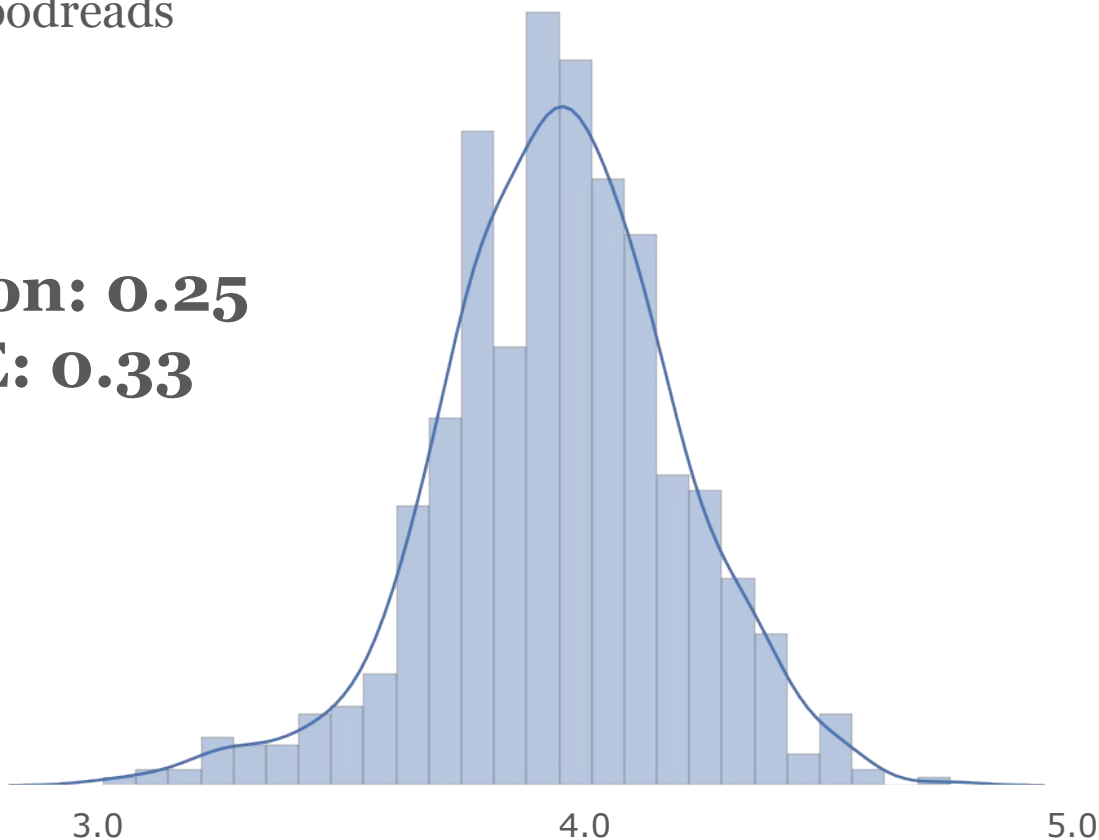


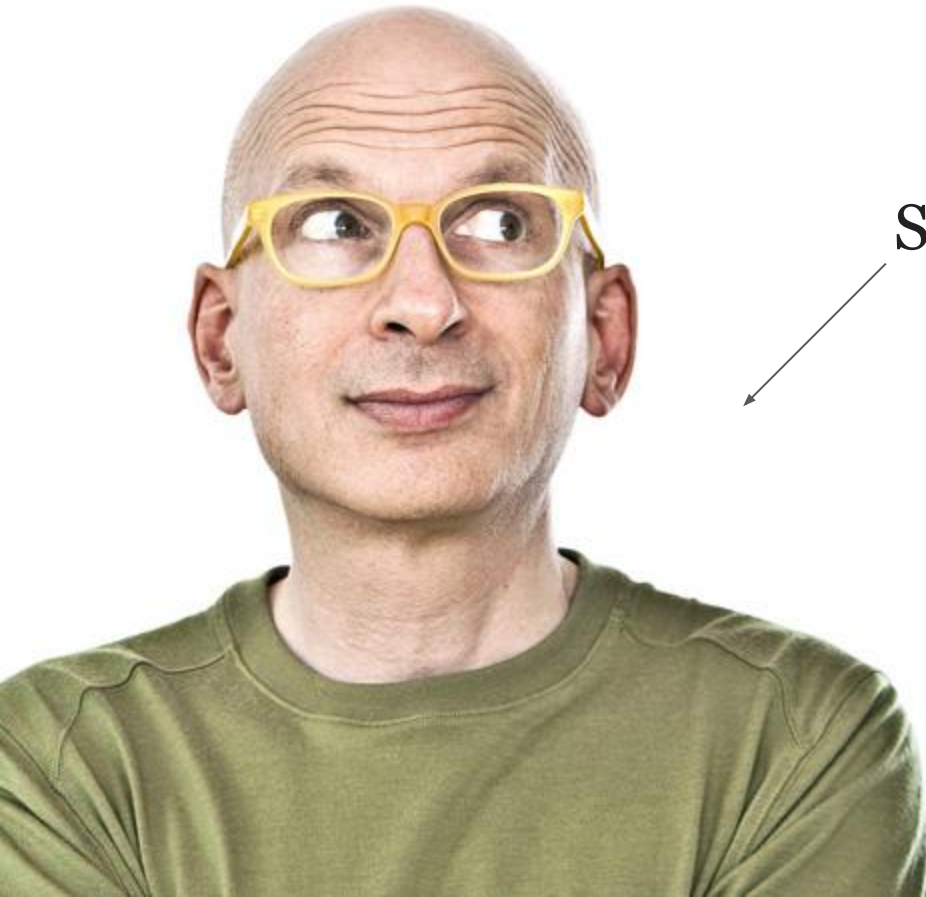
Mean rating for a **book** on Goodreads

3.95

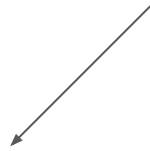
Standard deviation: 0.25

Lasso CV RMSE: 0.33





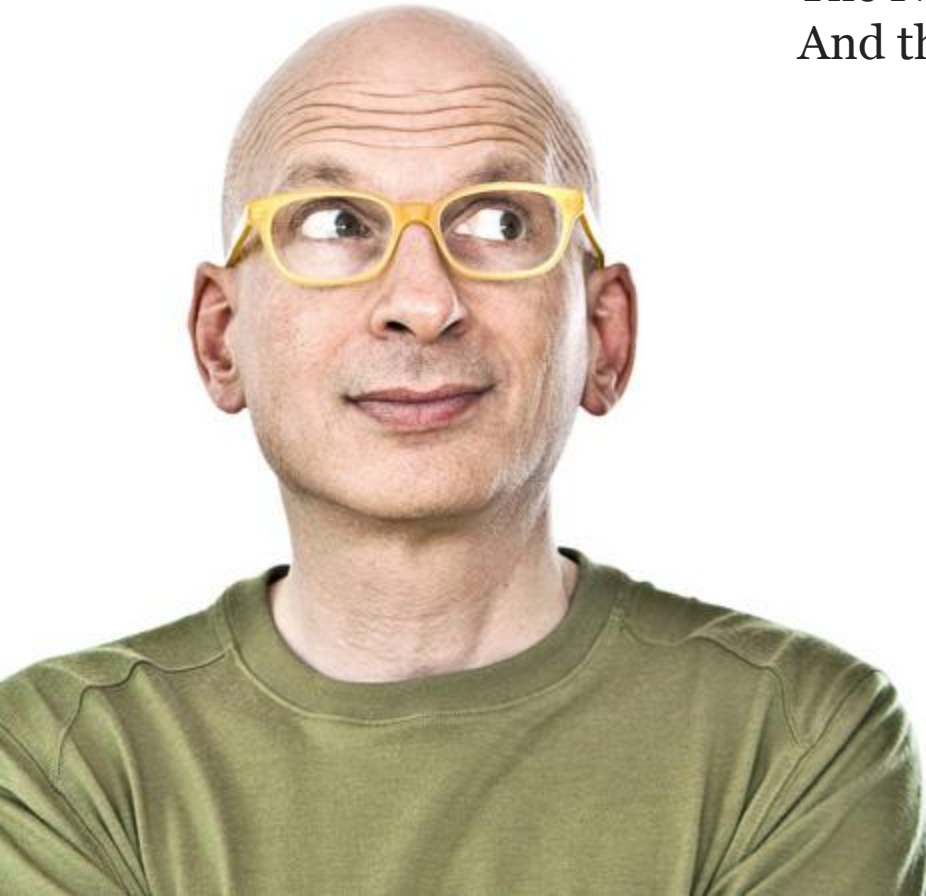
Seth Godin

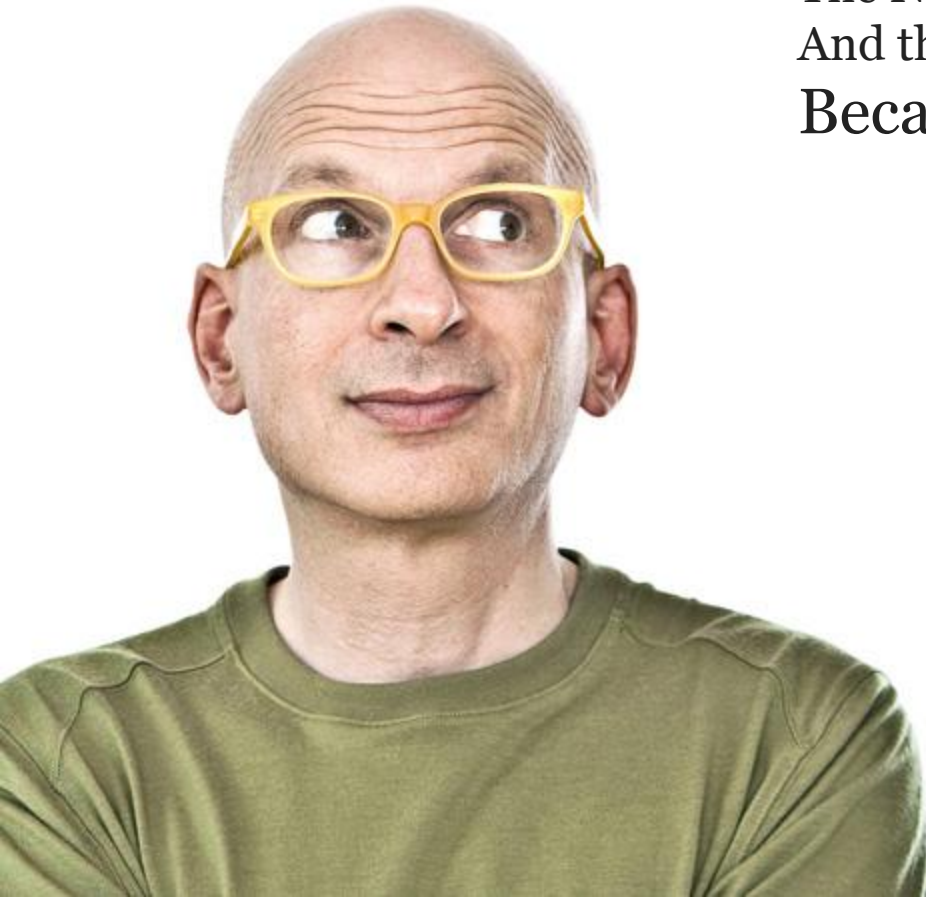


The New York Times Bestseller List is stupid.



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





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