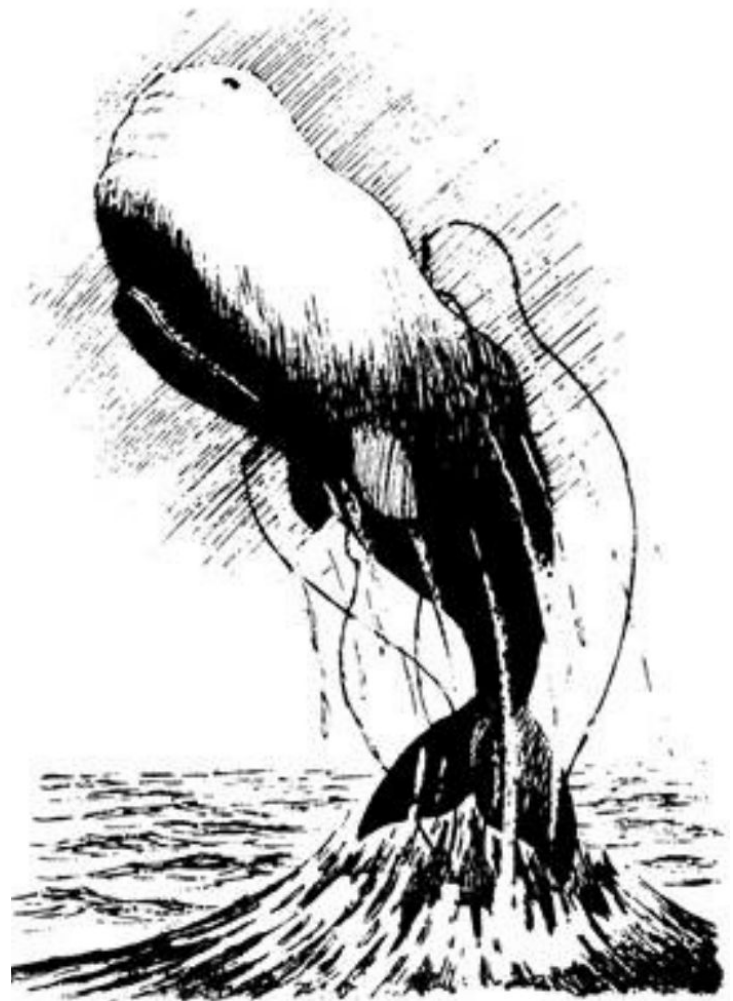


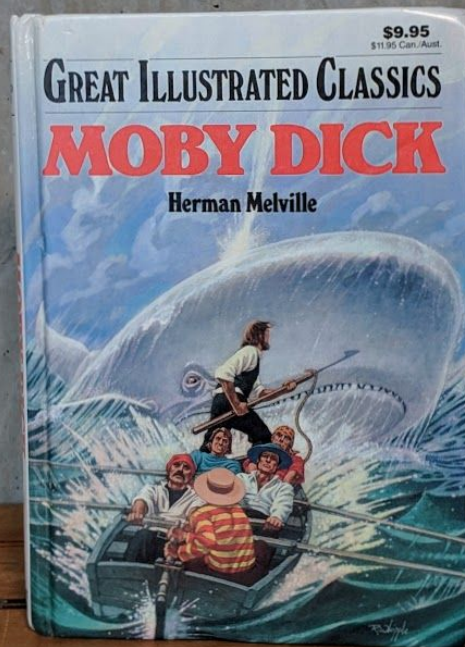
Cory Dominguez  
[twitter.com/@corydominguez](https://twitter.com/@corydominguez)  
[github.com/c11z](https://github.com/c11z)

# The Great Illustrated Classics, A Text Analysis of Moby Dick



The Great  
Illustrated  
Classics,  
A Text Analysis  
of Moby Dick;  
**or, The Whale**







**and Whale Oil.**

**2,000** GALLONS Spring pressed  
SPERM OIL;

1,000 Galls. refin'd WHALE OIL,  
superior quality, now landing direct from  
Nantucket, for sale at reduced prices by

**JOHN G. CLASBY.**

June 15.

31st.

Bow-st.





Moby Dick Had Cut Off Ahab's Leg.



Going Aboard the *Pequod* to Sign Up















	Abridged	Original	
Chapters	18	135	<b>8x</b>
Sentences	1,866	9,822	<b>5x</b>
Words	16,815	211,956	<b>13x</b>
Avg words/sentence	9	22	<b>2x</b>
Read time (265 wpm)	1 hour	13 hours	

	Abridged	Original	
Chapters	18	135	<b>8x</b>
Sentences	1,866	9,822	<b>5x</b>
Words	16,815	211,956	<b>13x</b>
Avg words/sentence	9	22	<b>2x</b>
Read time (265 wpm)	1 hour	13 hours	
Illustrations	122	0	

# Reading Indexes





# Flesch Kincaid Reading Age

$$\text{fkra} = (0.39 * \text{asl}) + (11.8 * \text{asw}) - 15.59$$

asl: Average sentence length.

asw: Average syllables per word.

Abridged: 3rd Grade

Original: 9th Grade

# Flesch Kincaid Reading Age

```
import cmudict

leviathan = (
    ["L", "EH0", "V", "AY1", "AH0", "TH", "AH0", "N"]
)

pattern = r"[aiouy]+e*|e(?!d$|ly)\.|[td]ed|le$"
```

# Flesch Kincaid Reading Age

```
import cmudict

leviathan = (
    ["L", "EH0", "V", "AY1", "AH0", "TH", "AH0", "N"]
)

pattern = r"[aiouy]+e*|e(?!d$|ly)\.|[td]ed|le$"
```

# Flesch Kincaid Reading Age

```
import cmudict

leviathan = (
    ["L", "EH0", "V", "AY1", "AH0", "TH", "AH0", "N"]
)

pattern = r"[aiouy]+e*|e(?!d$|ly)\.|[td]ed|le$"
```

# Automated Readability Index

$$\text{ari} = (0.5 * \text{asl}) + (4.17 * \text{awl}) - 21.43$$

asl: Average sentence length.

awl: Average word length.

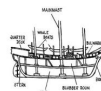
Abridged: 2.4th Grade

Original: 10th Grade

# Proper Nouns

one-of-a-kind

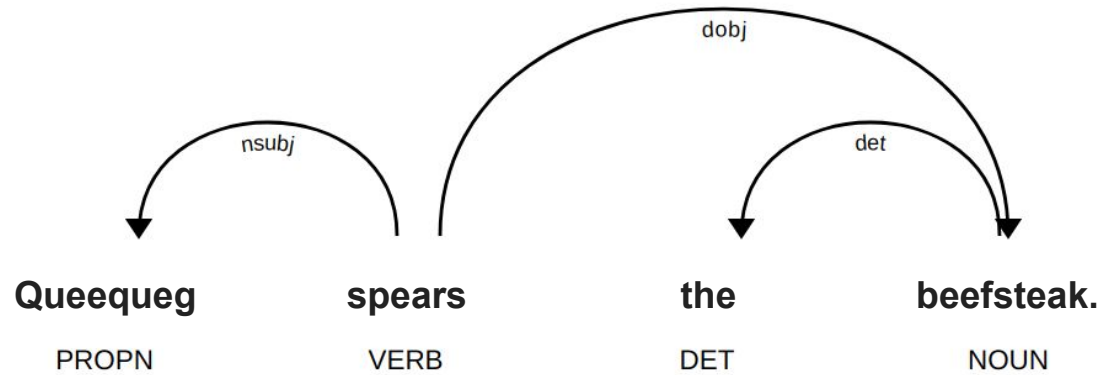
people, animals, places, things, and ideas

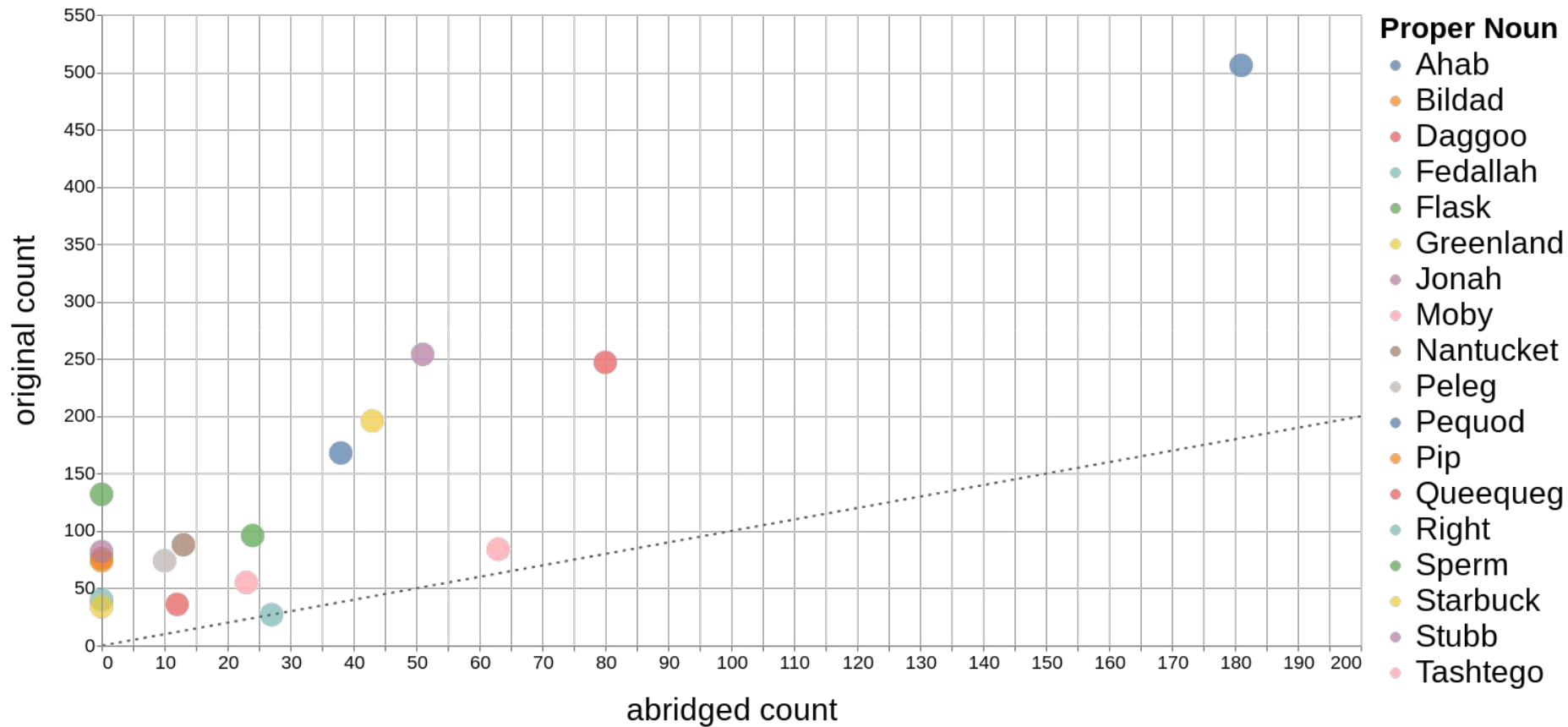


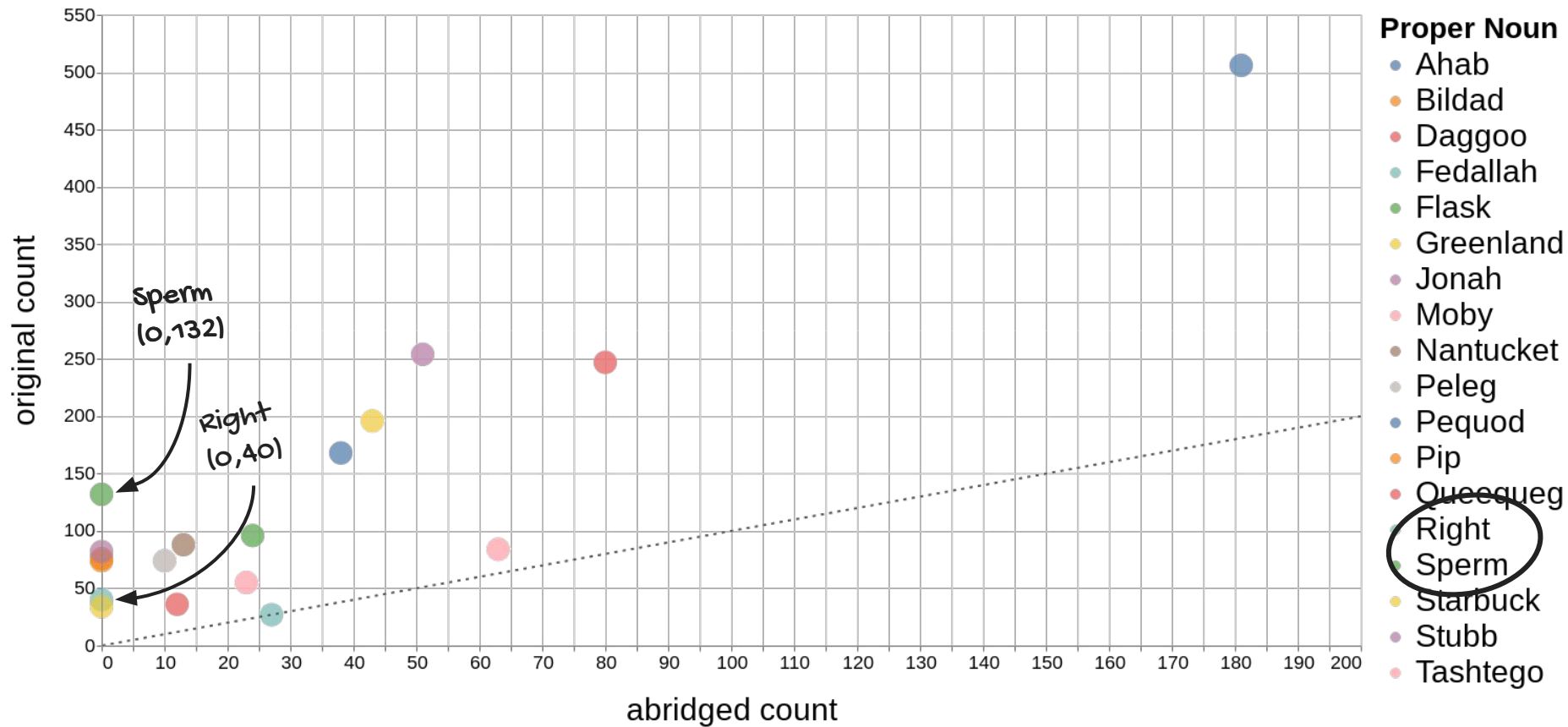
# spaCy

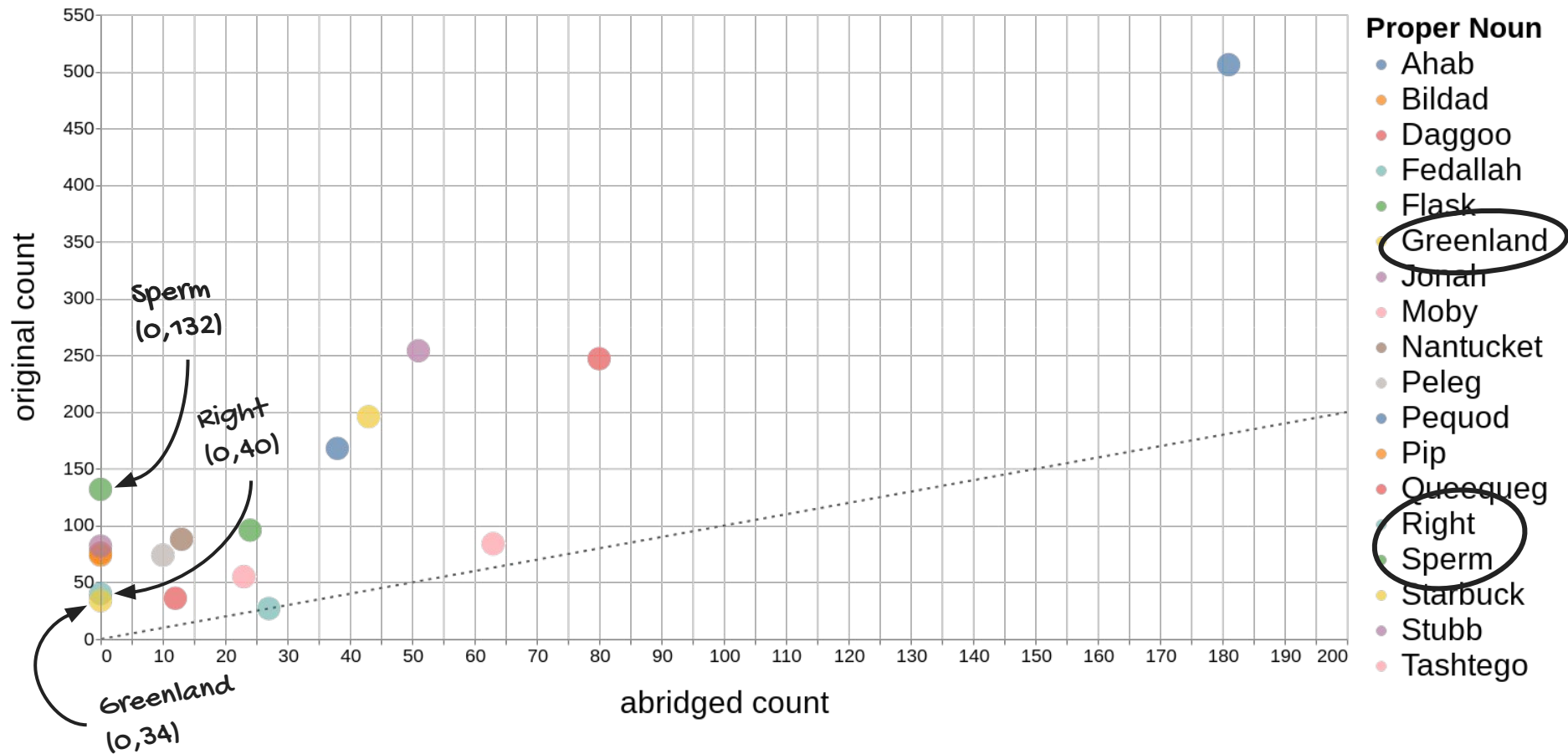
Parts of speech tagging  
with  
Average Perceptron











Sperm  
Right  
Greenland



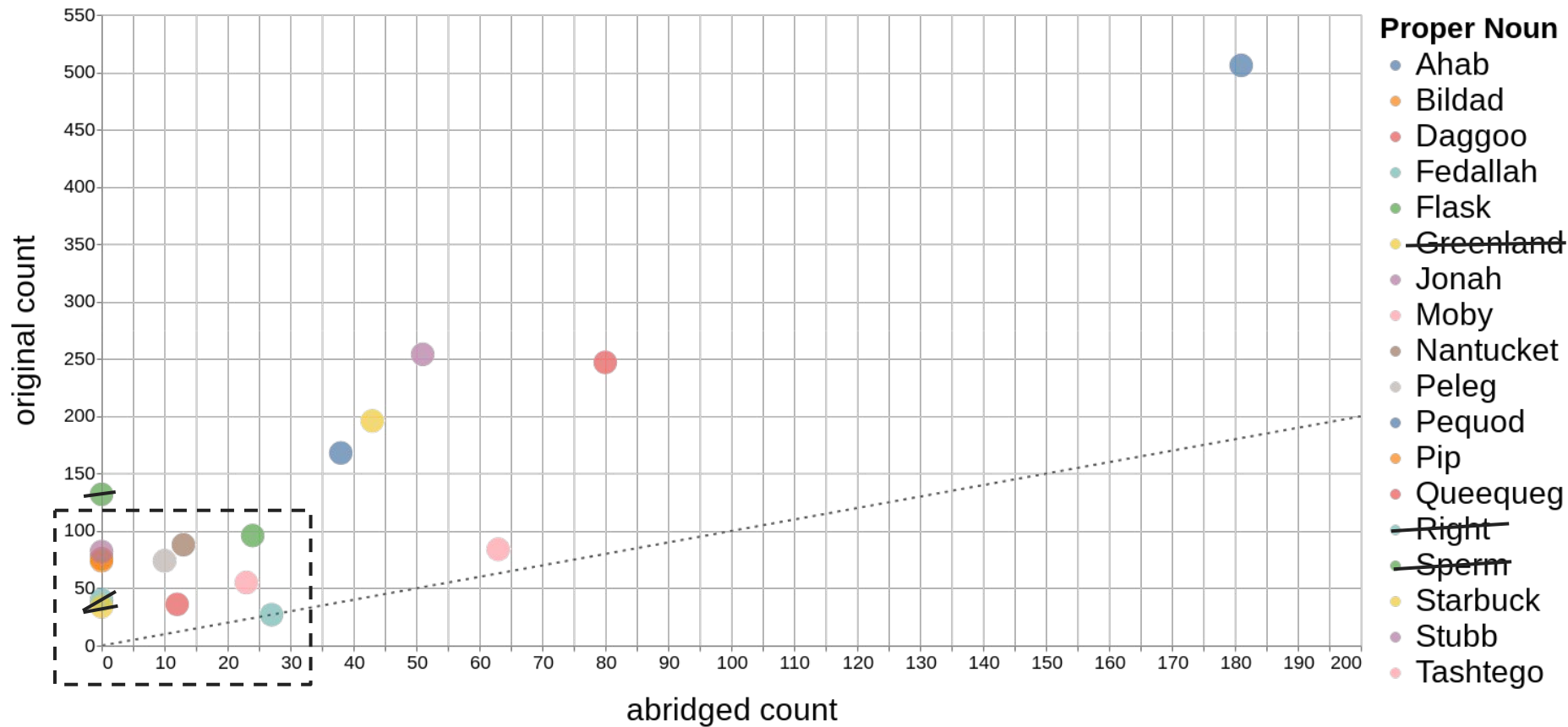
# Whales

Sperm

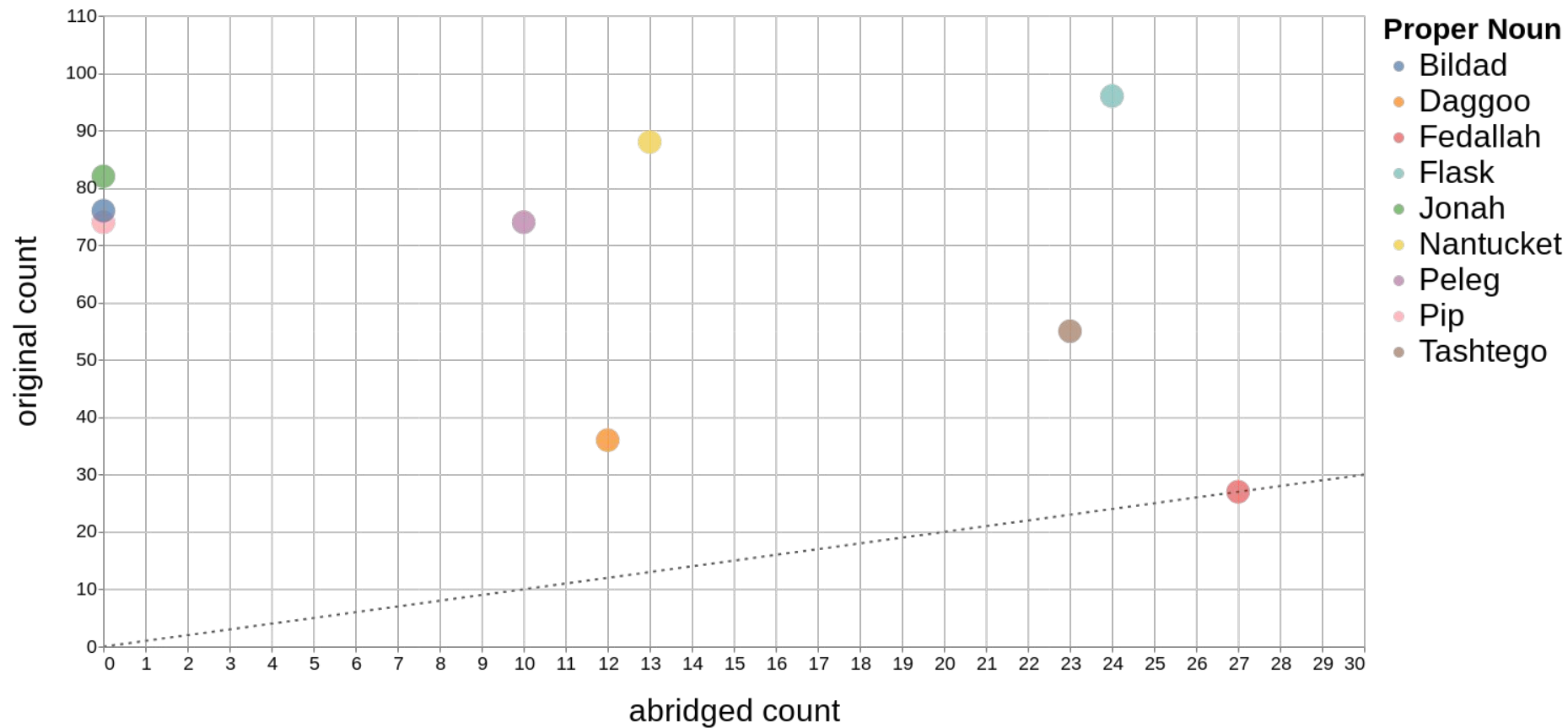
Right

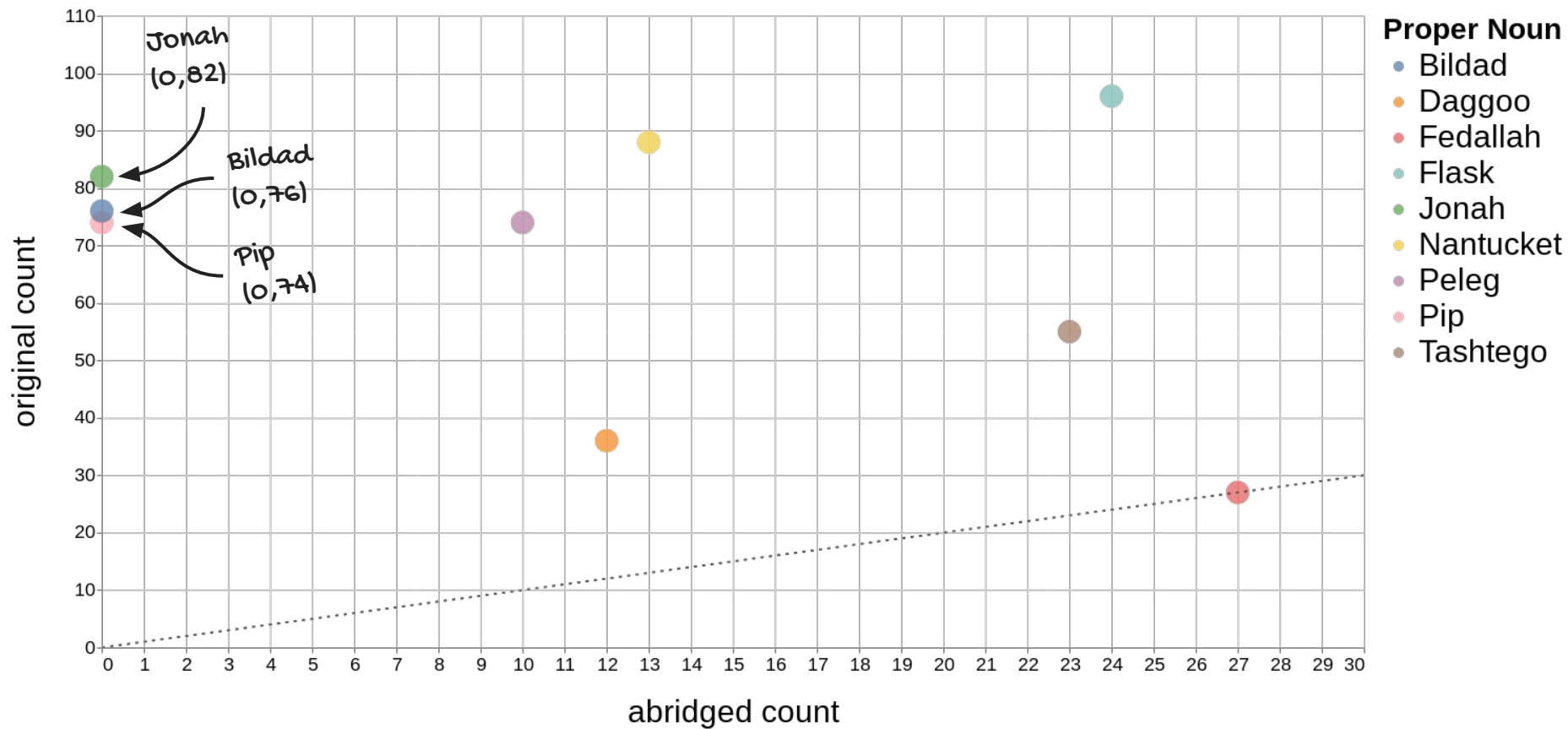
Greenland



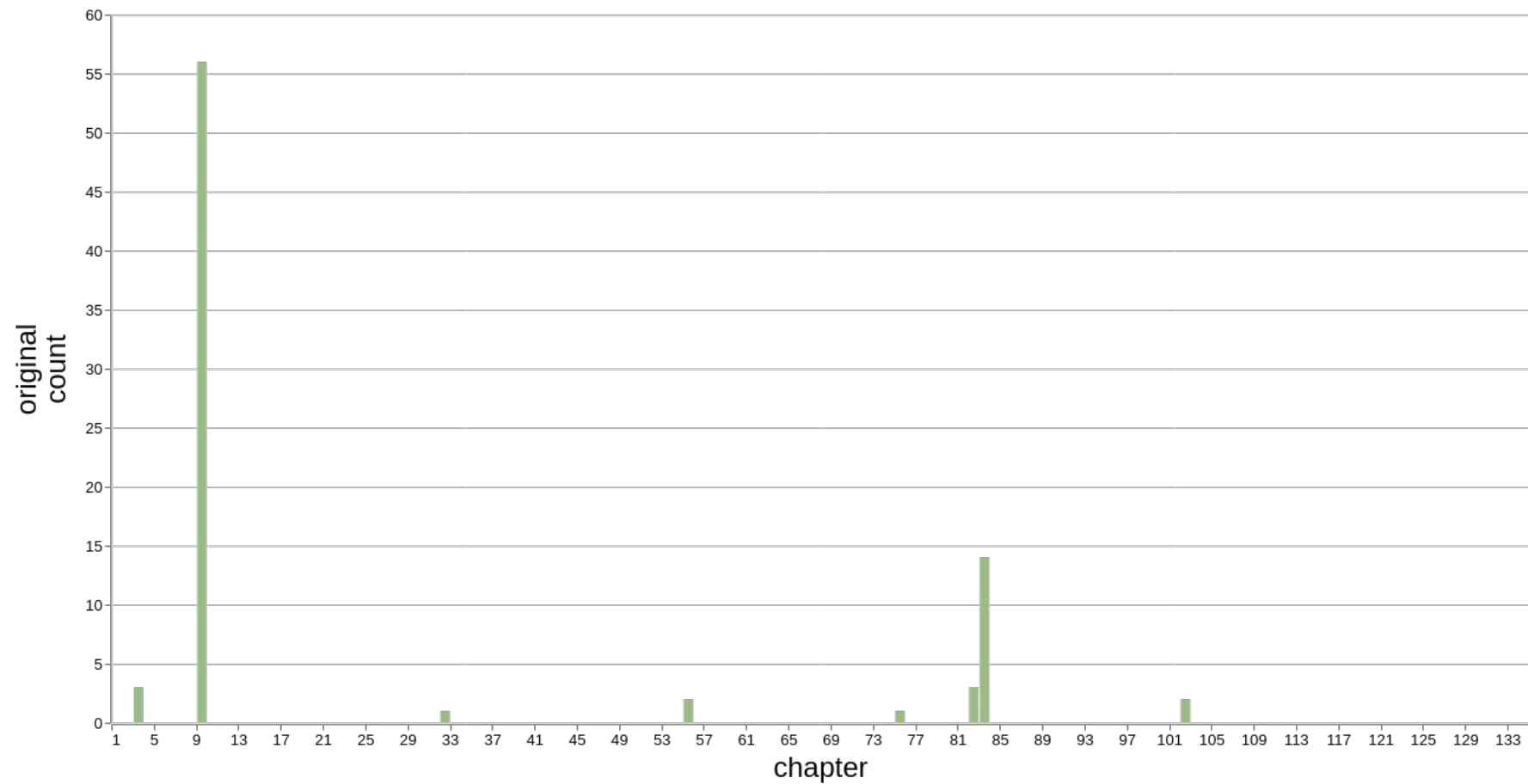




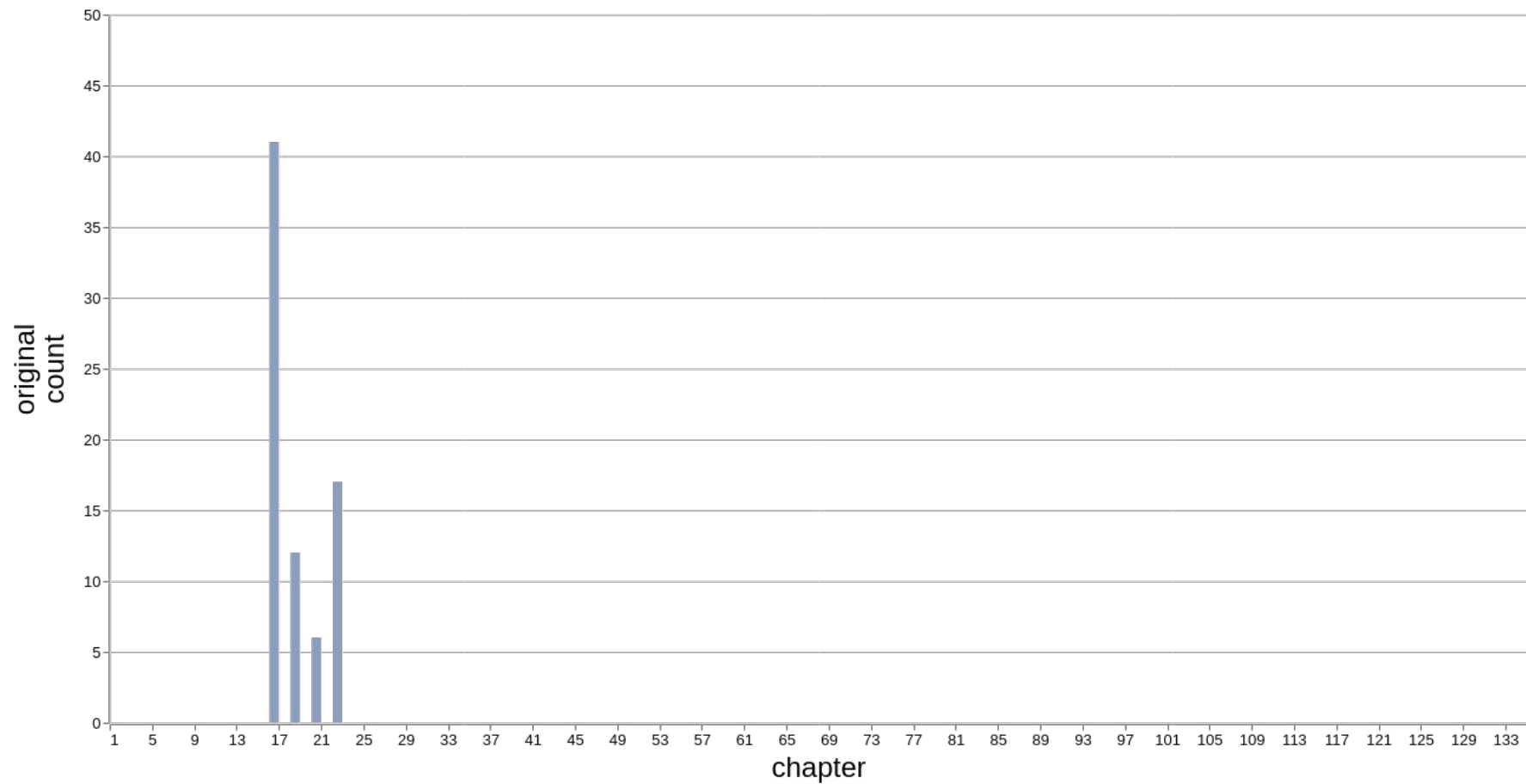




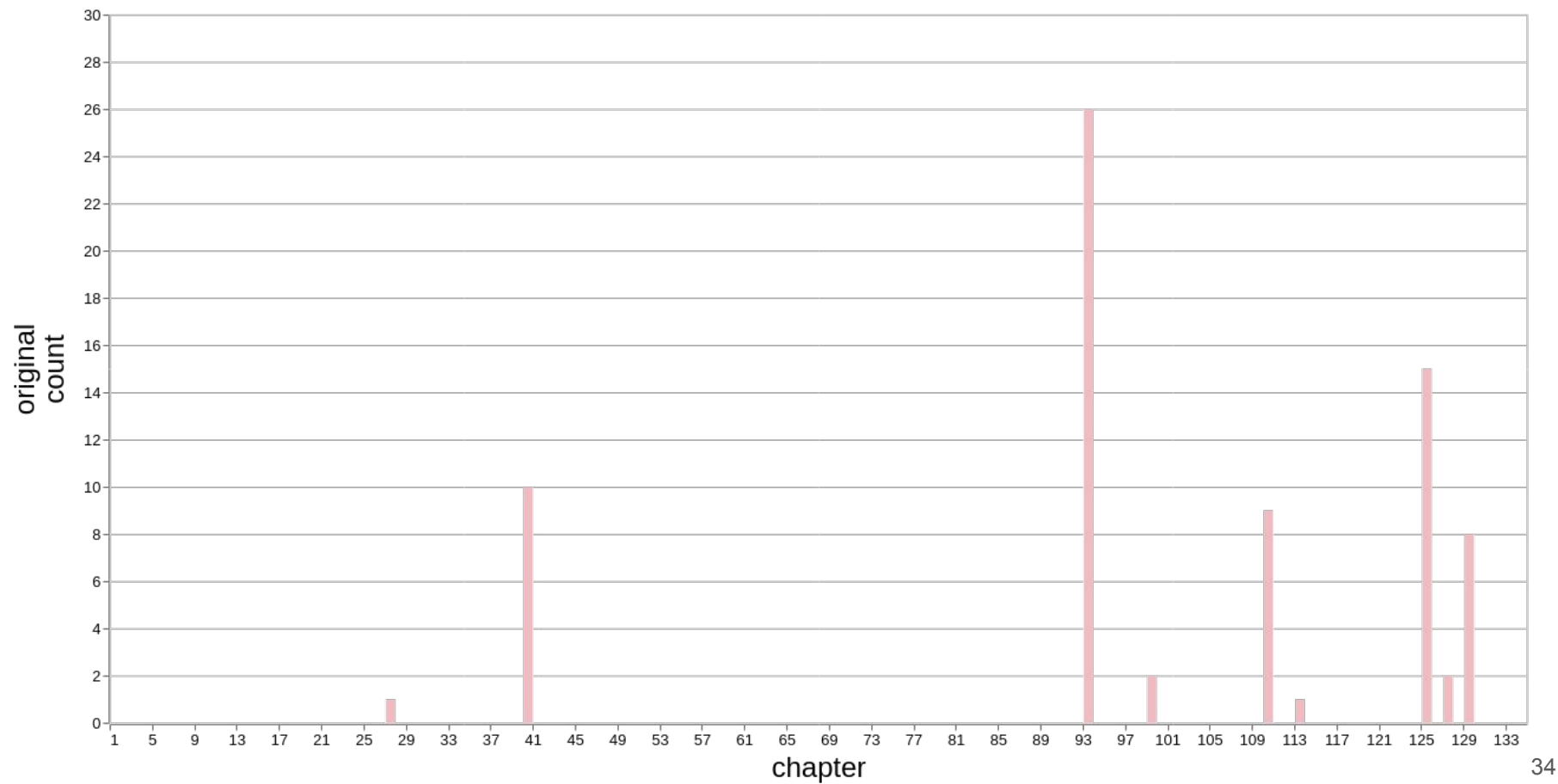
# Jonah



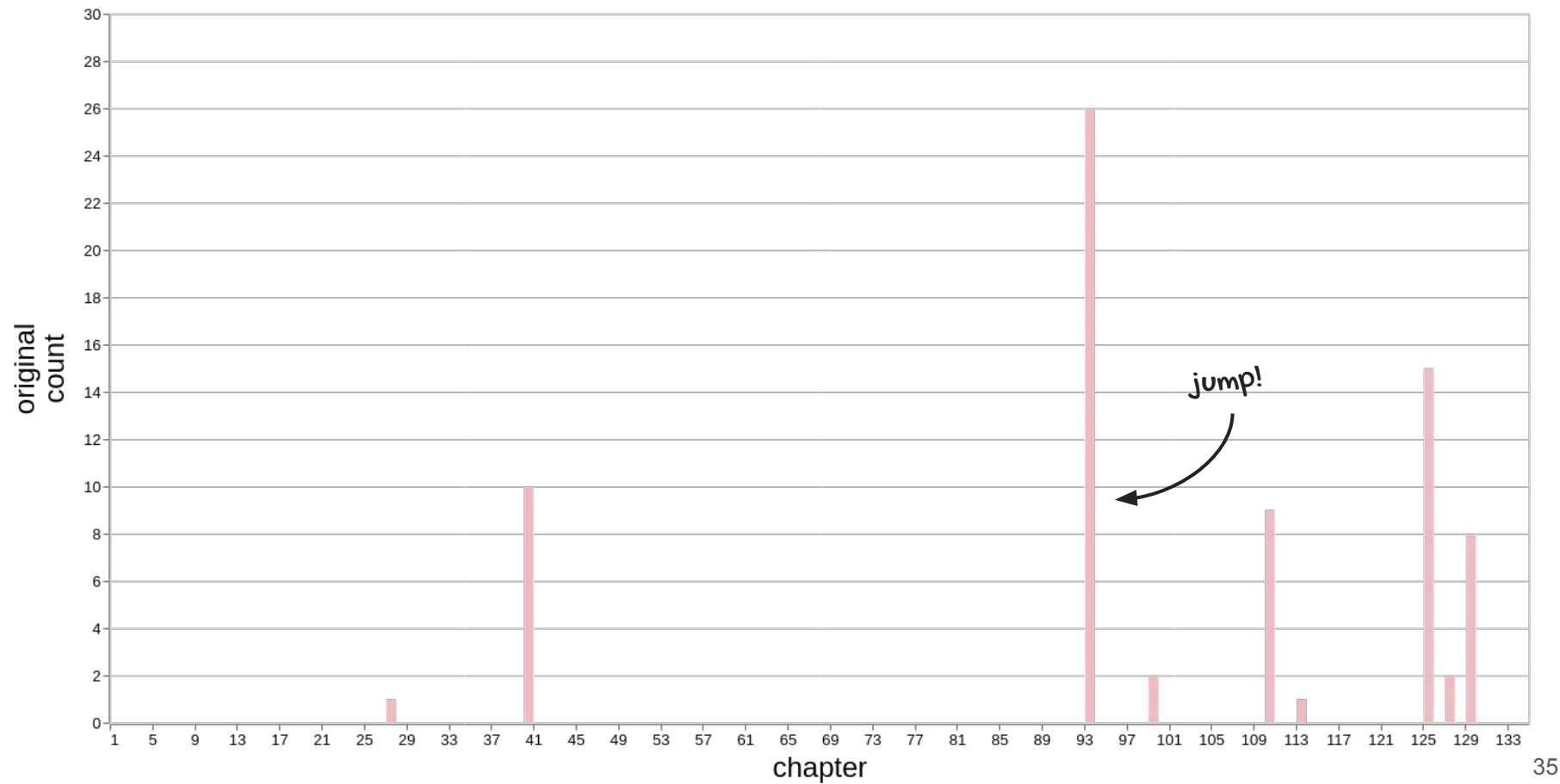
# Bildad



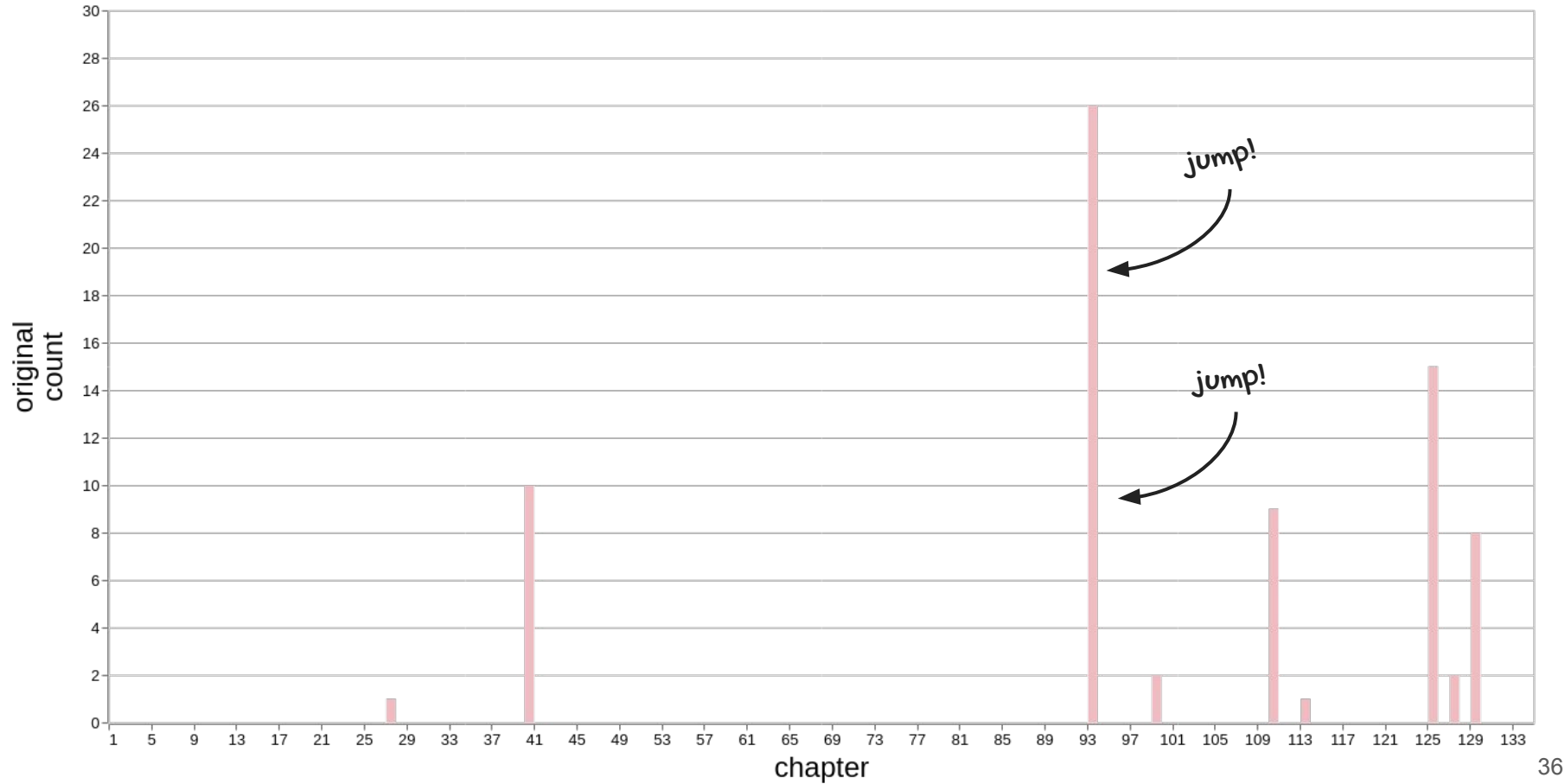
# Pip



# Pip

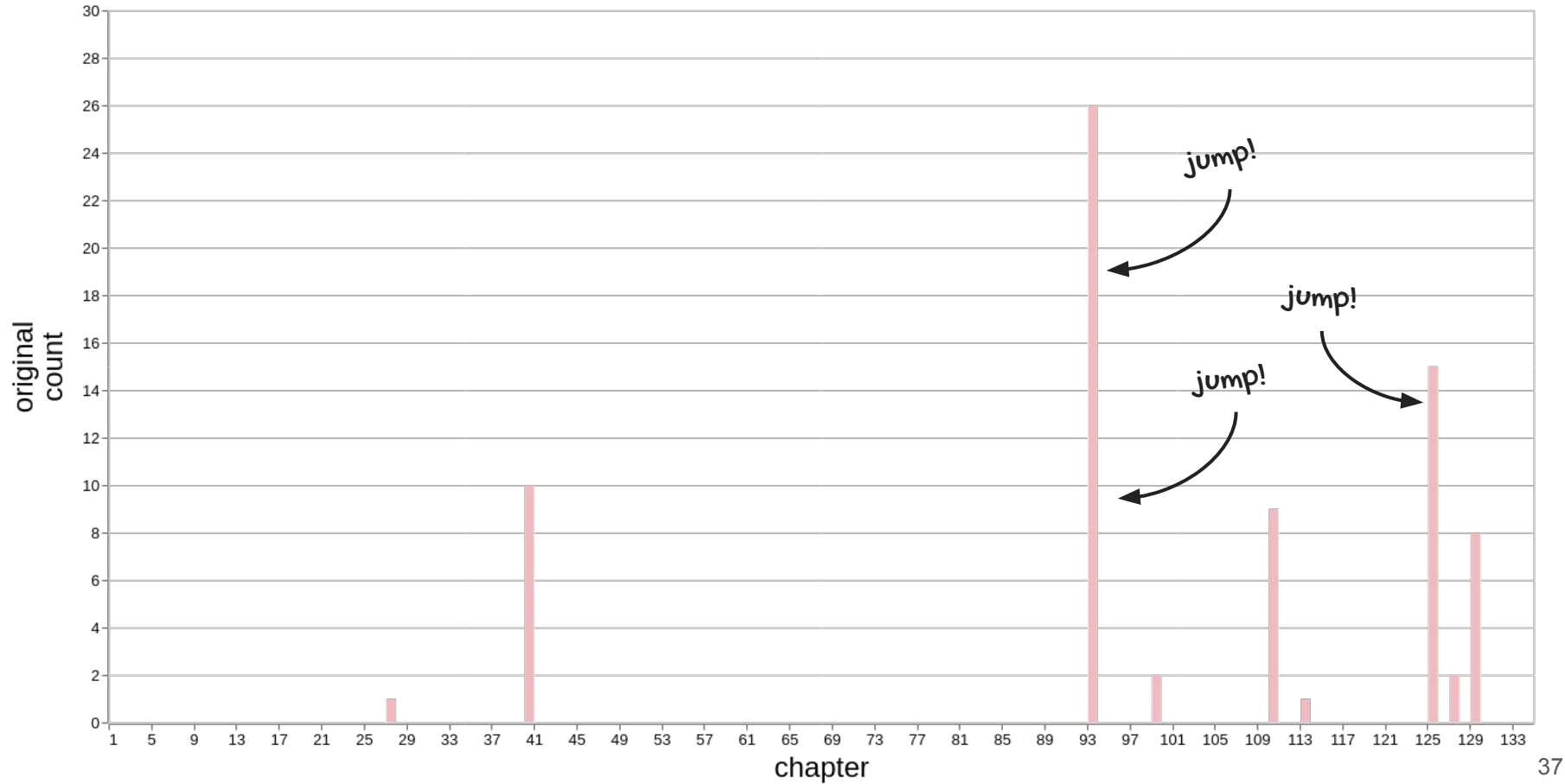


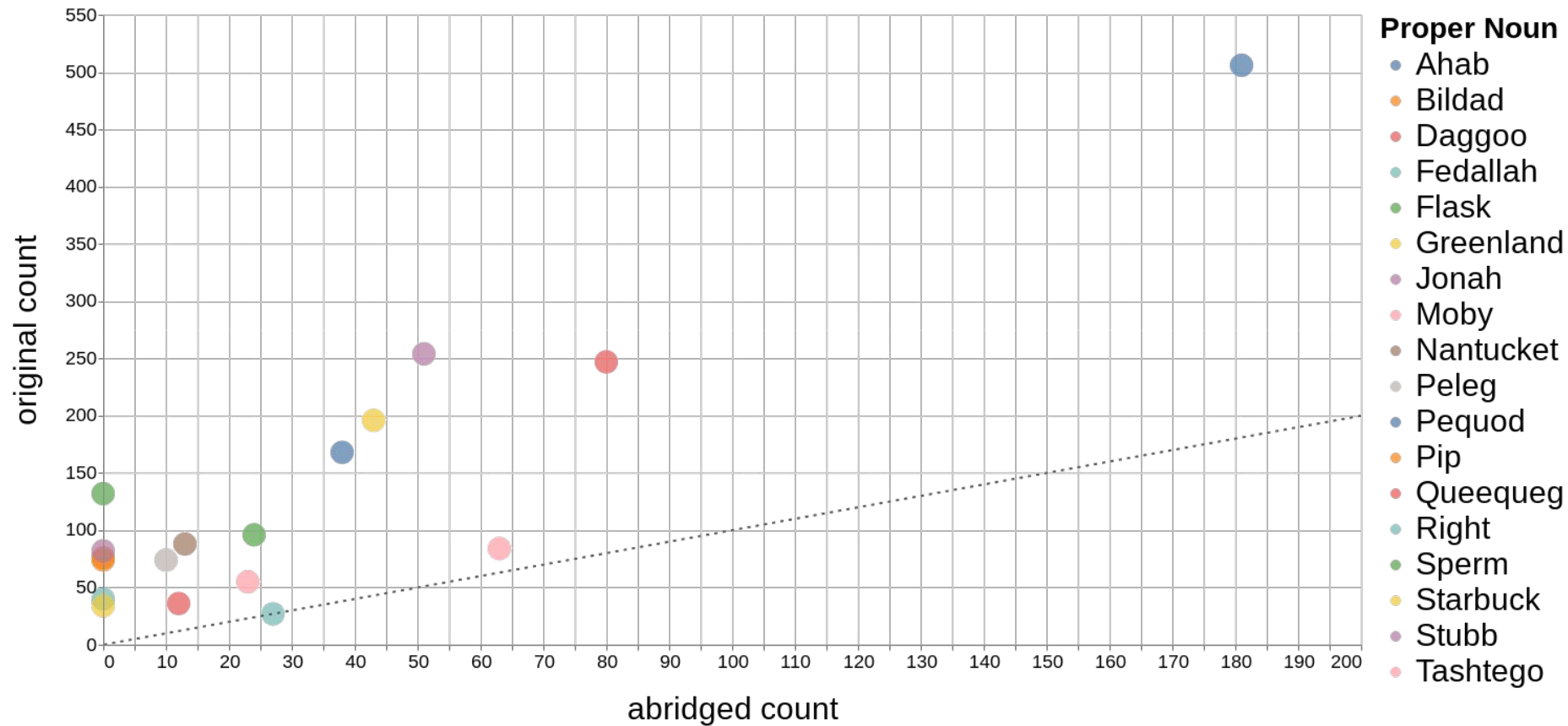
# Pip

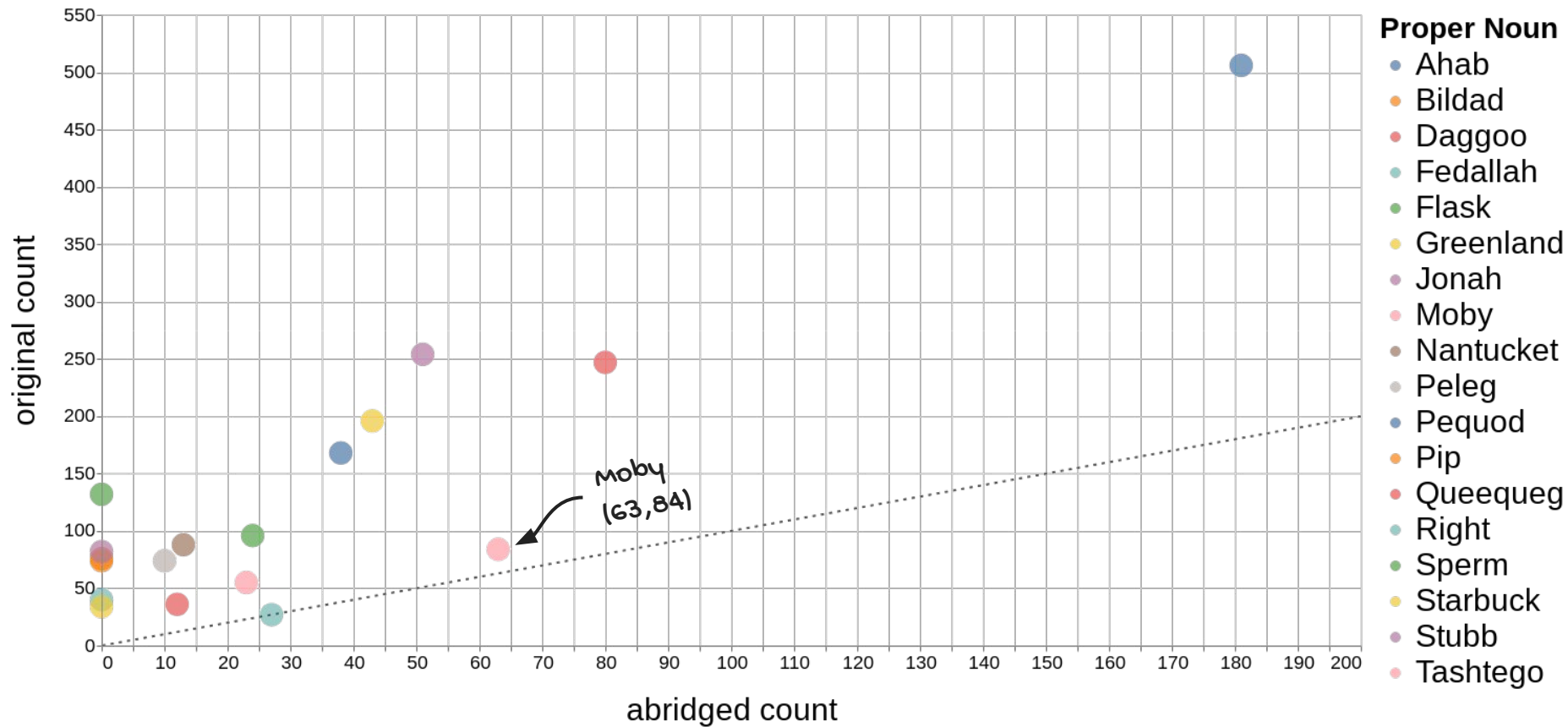




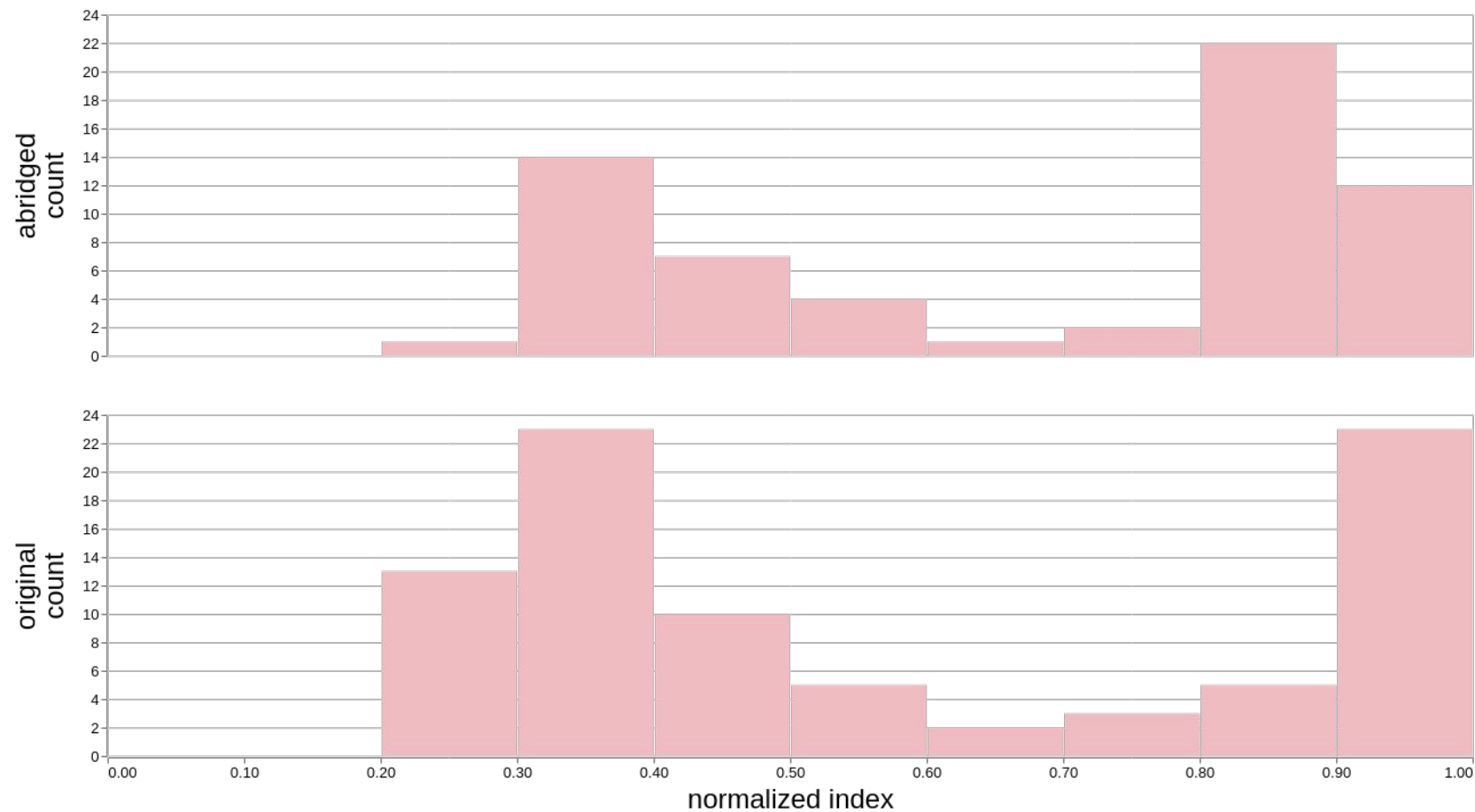
# Pip

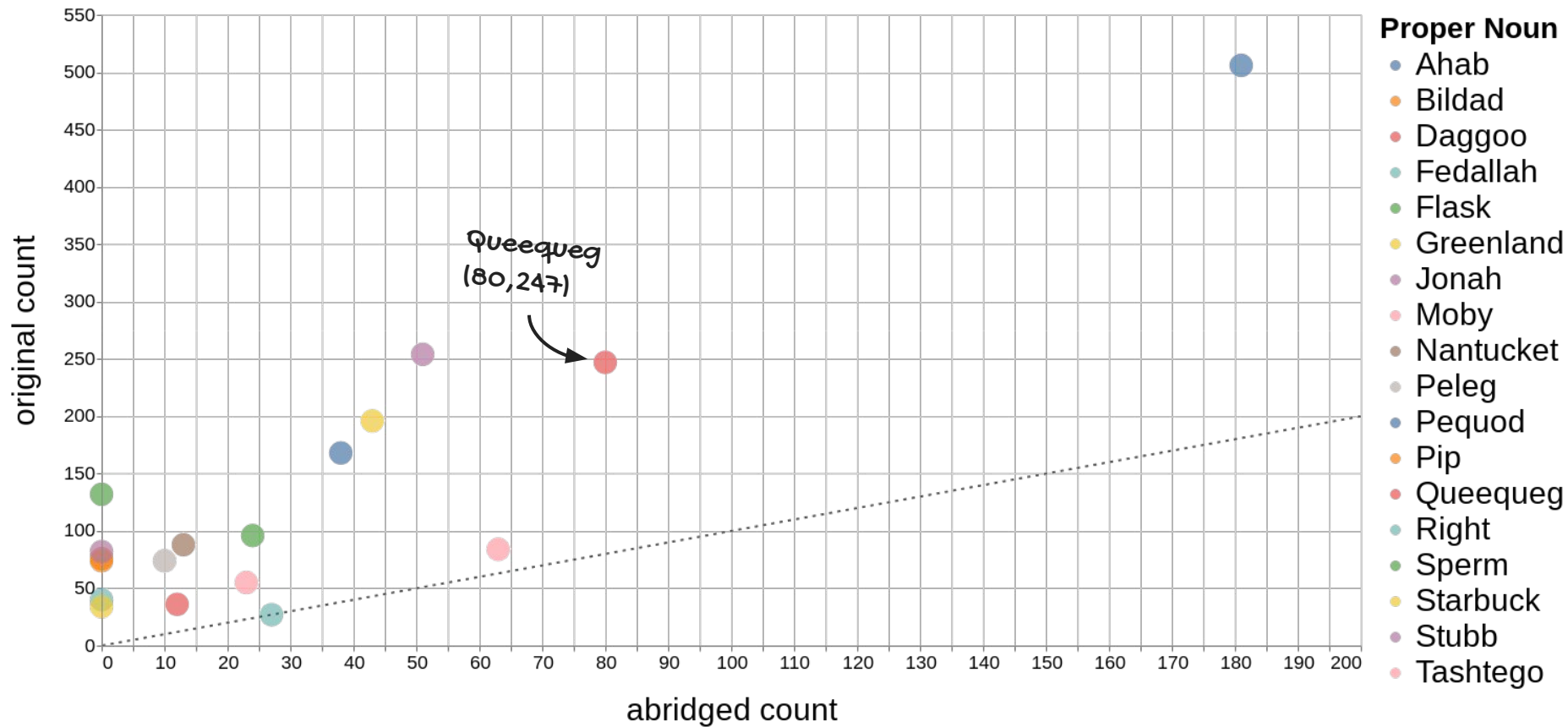




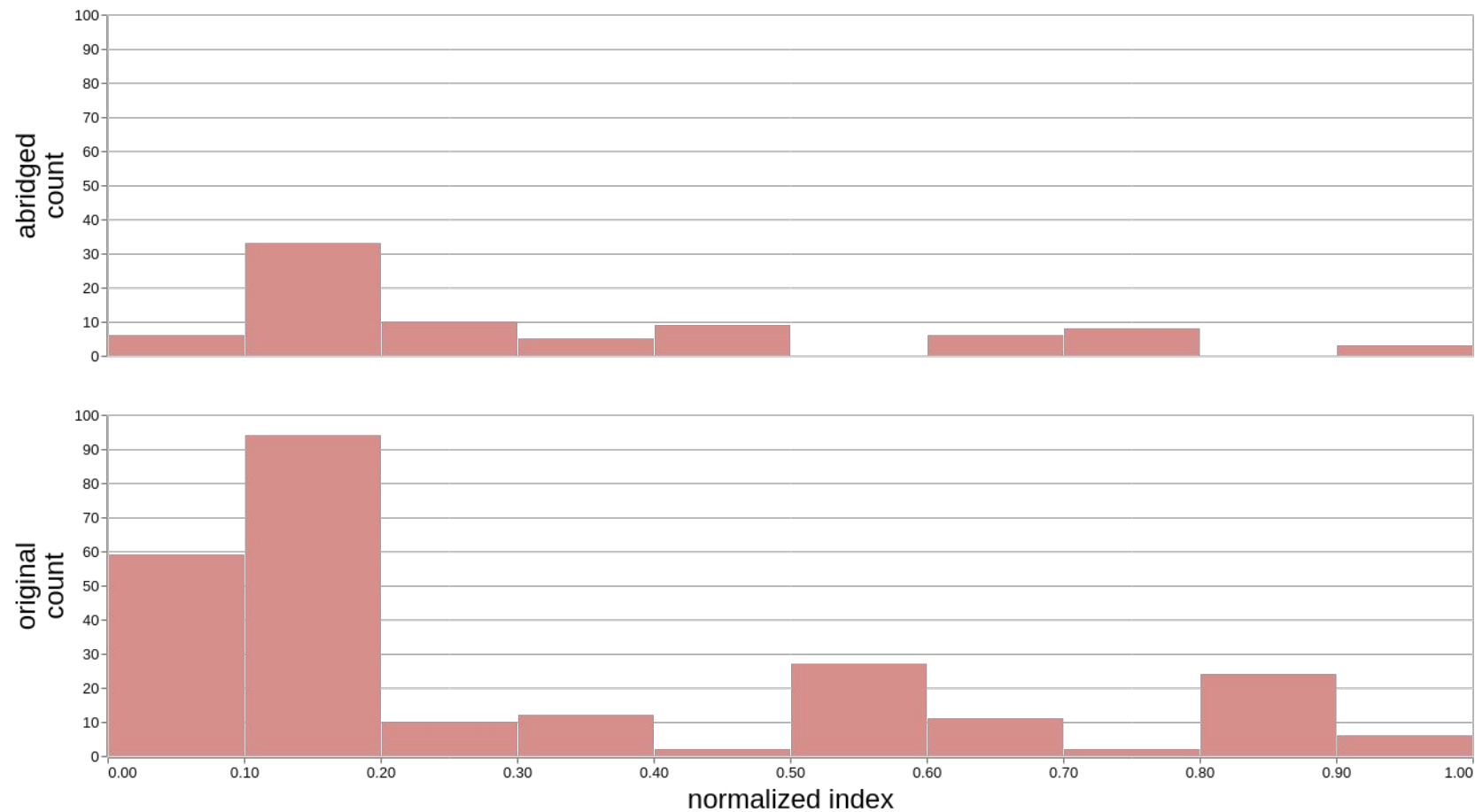


# Moby

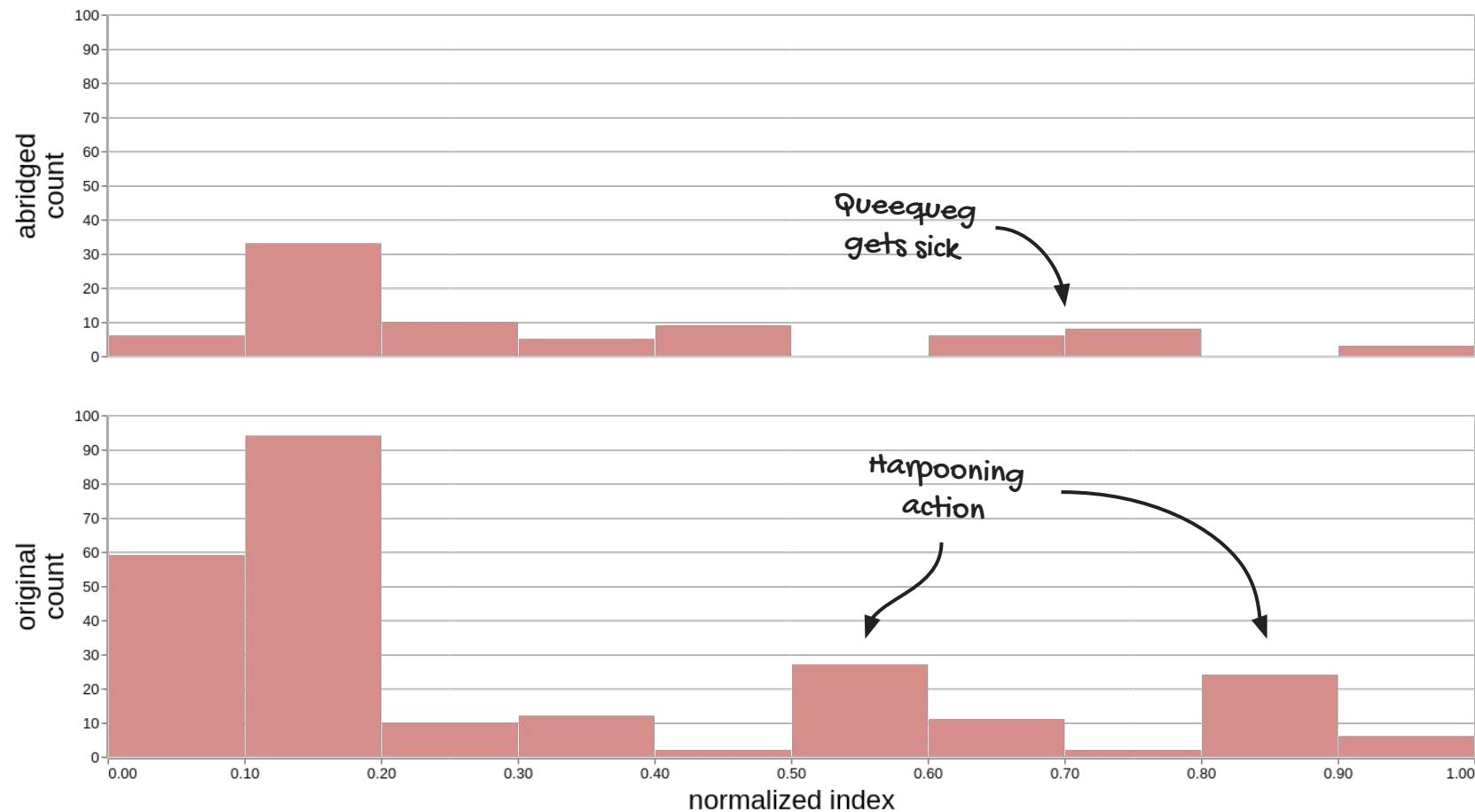


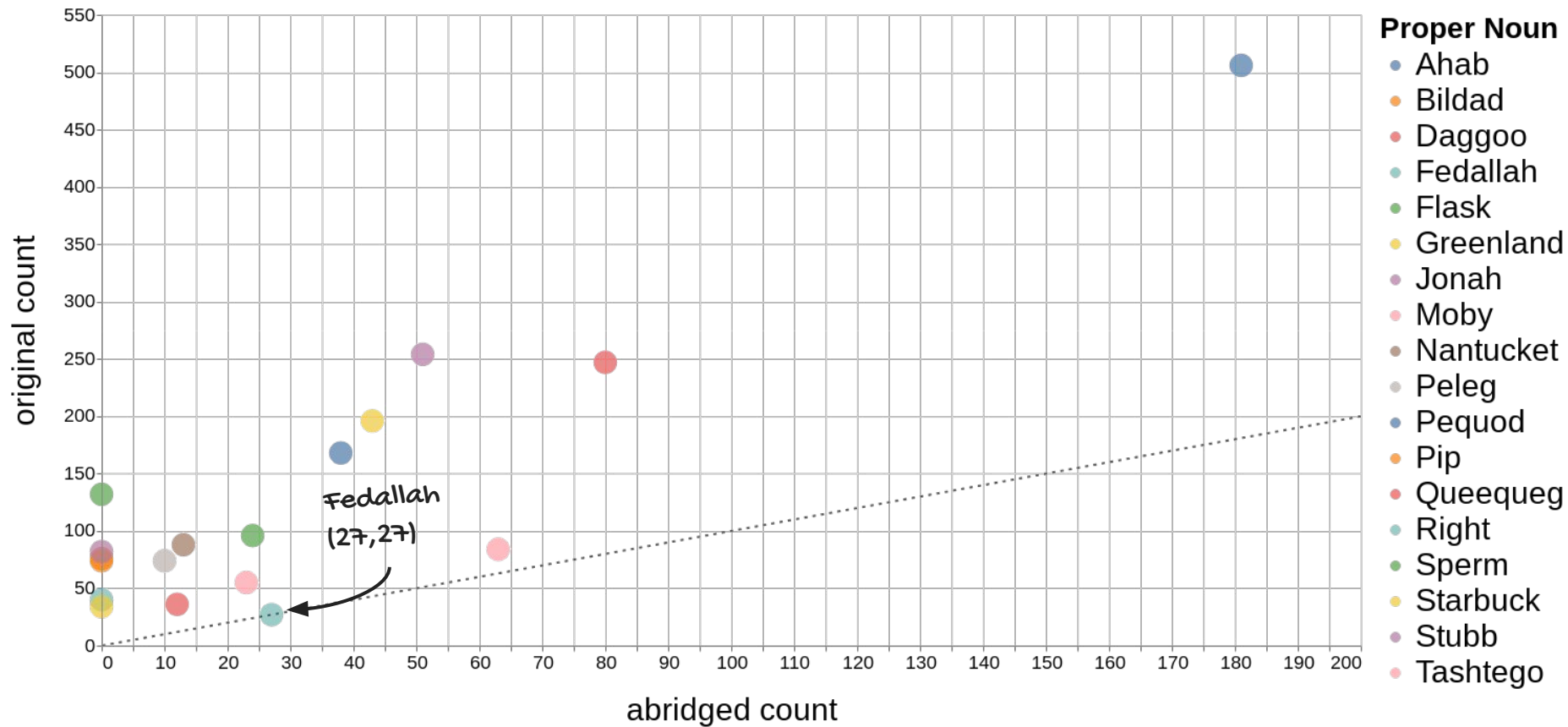


# Queuequeg



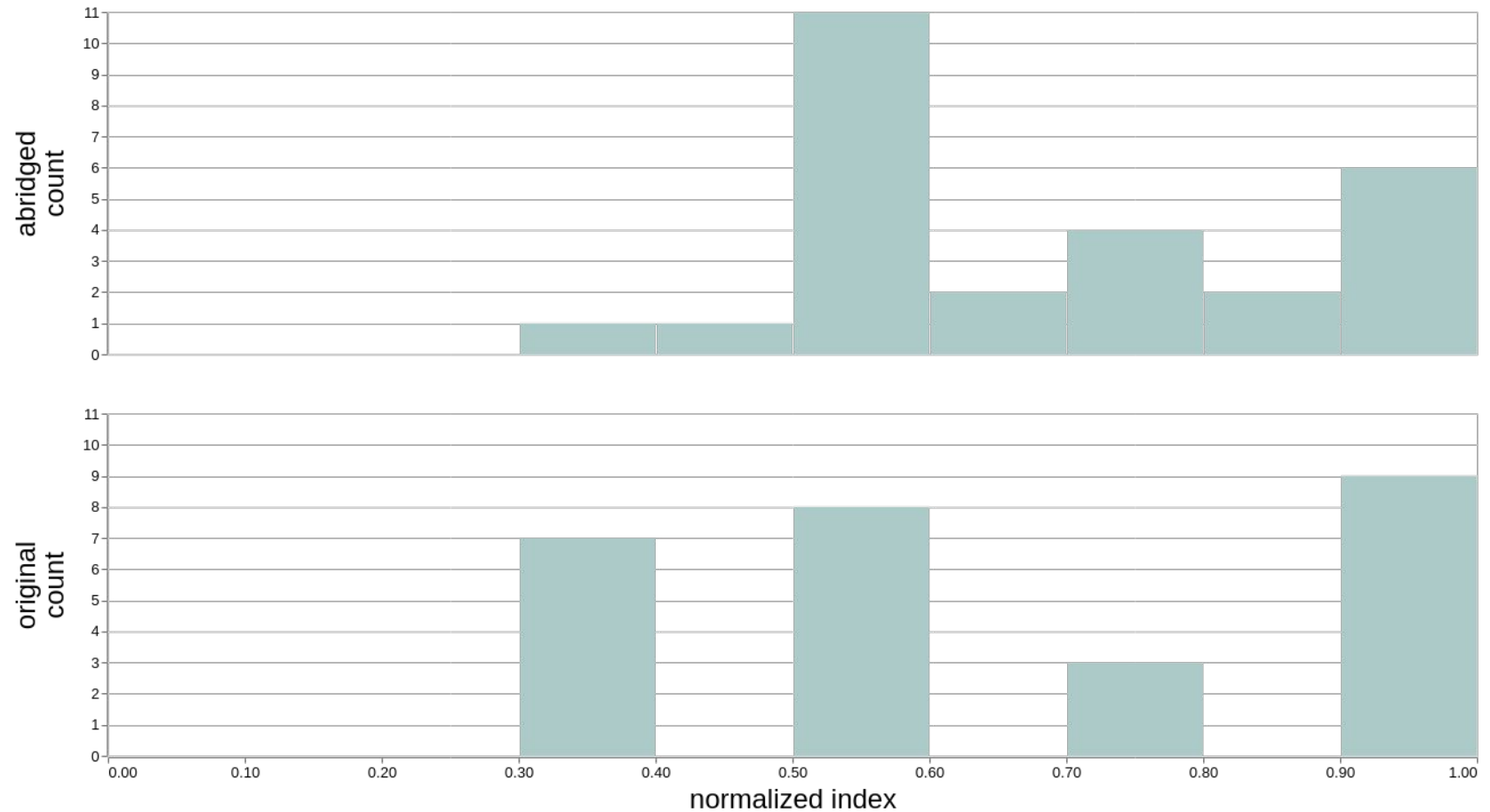
# Queequeg

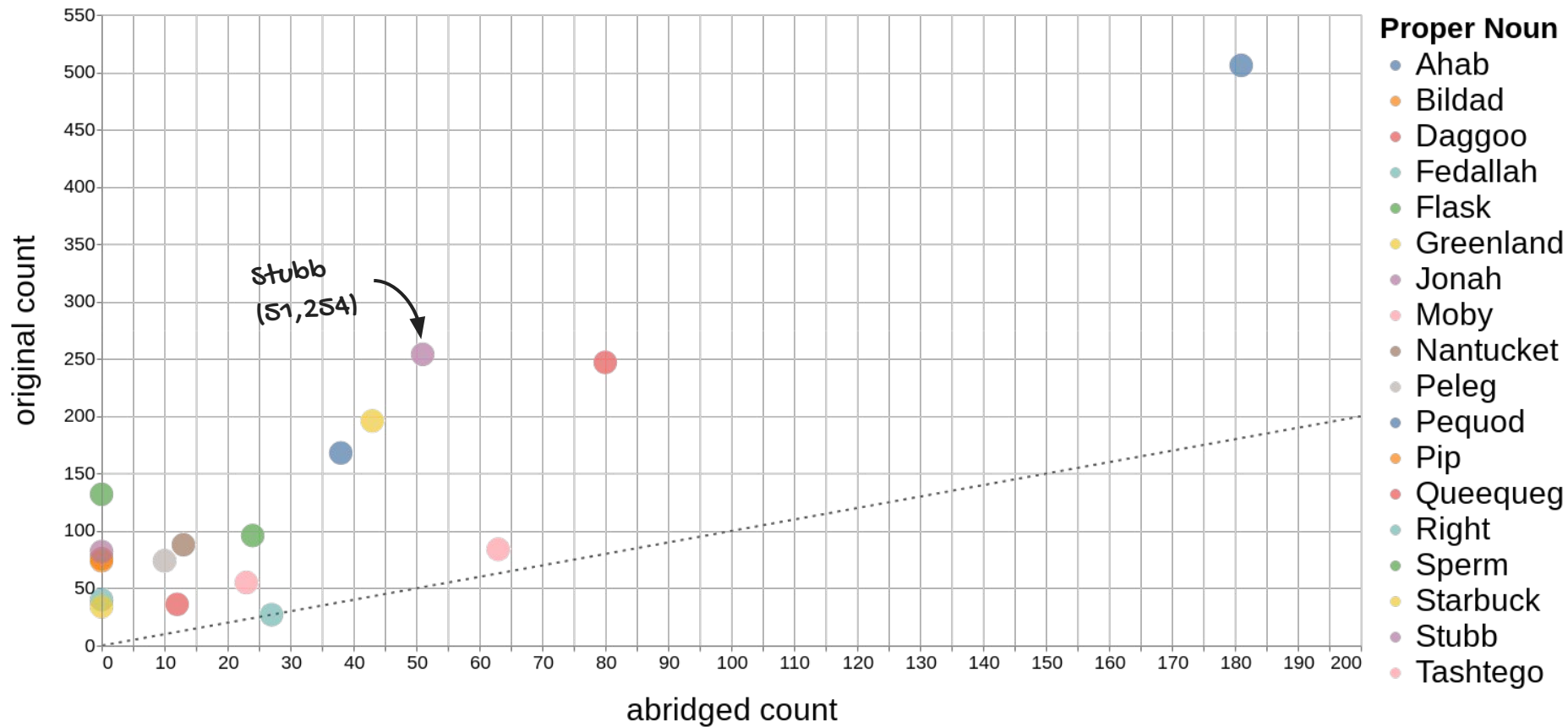




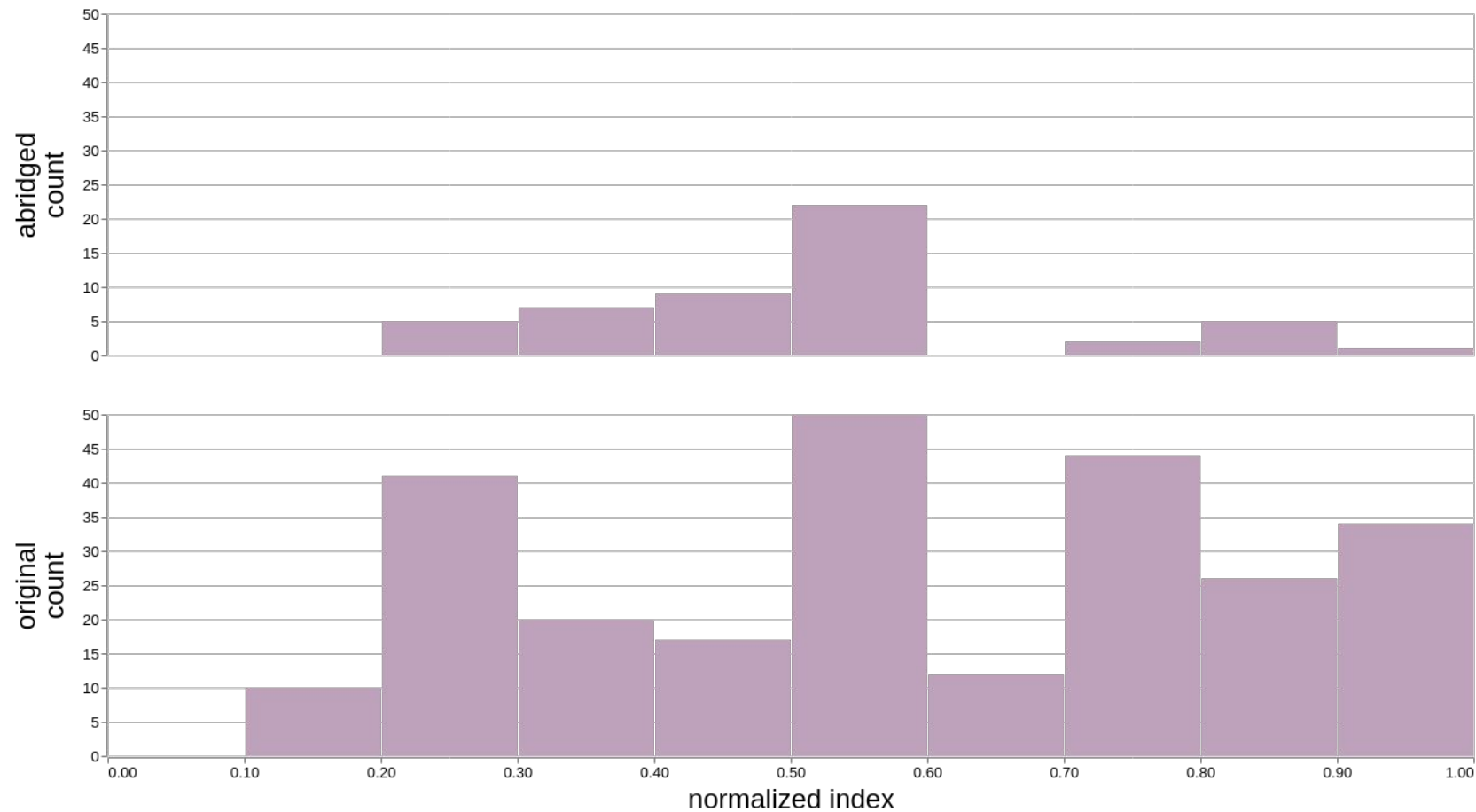


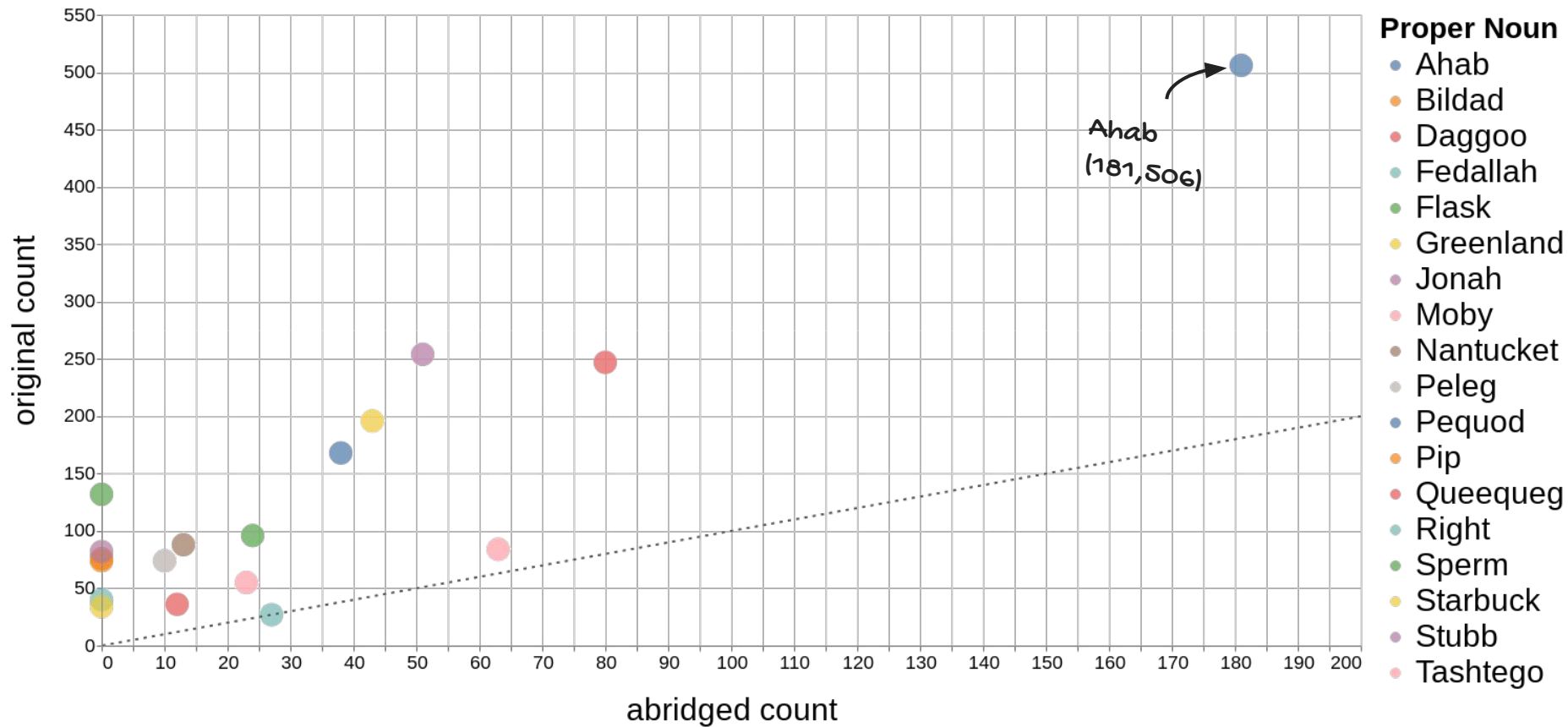
# Fedallah



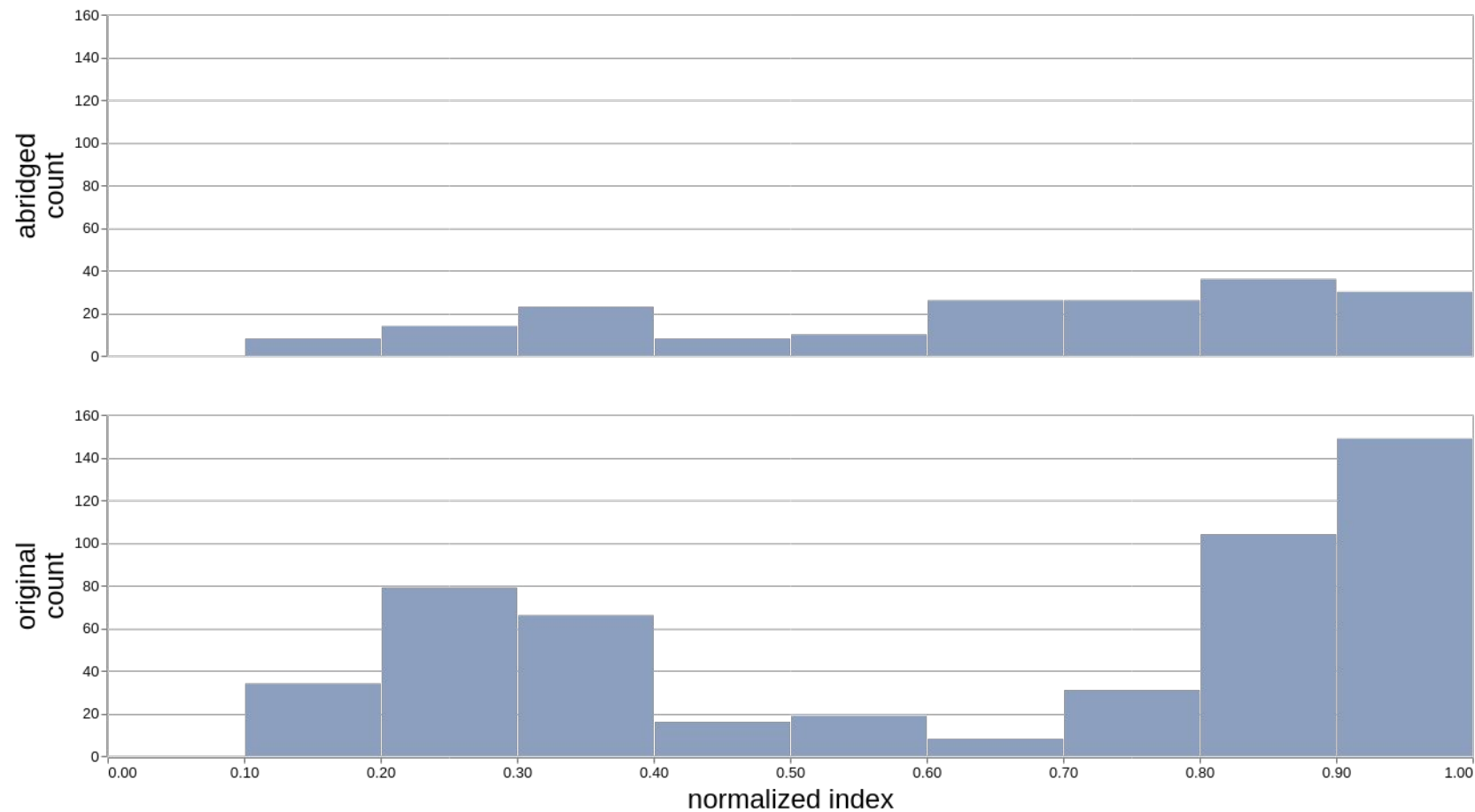


# Stubb





# Ahab



- ❖ One biblical story
- ❖ Two missing characters
- ❖ Three not so one-of-a-kind whales
- ❖ Five trellis histograms
- ❖ And a regex that matches syllables



# Thank You

Cory Dominguez

[twitter.com/@corydominguez](https://twitter.com/@corydominguez)

[github.com/c11z](https://github.com/c11z)