

‘Gender Hierarchy in Fiction: A Replication of Kraicer and Piper, 2019: Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction’

Hannah Hardenbergh

5/10/2019

This file contains code that replicates the first table and first four figures of Kraicer and Piper’s report on how gender bias has found its way into contemporary fiction. Published in the Journal of Cultural Analytics on January 30, 2019, Kraicer and Piper’s research is at the intersection of both the humanities and a quantitative analysis of the social sciences. Their work is extensive and impressive. The scope of this replication is to enhance the first four figures and first two tables, which will give the paper as a whole a more understandable visual for its readers. Kraicer and Piper establish several statistically significant assertions in this paper; my goal is to simply provide readers a more accurate visual for the contextualization of the underrepresentation of female-gendered characters in contemporary novels so that Kraicer and Piper’s foundation for their statistical analyses is crystal-clear. Their paper is strong, and their dataset is massive. It could benefit from continued cleaning and simplifying. This is certainly an area for future research.

Below I replicate the table that describes the dataset that Kraicer and Piper are working with. They selected novels from seven different popular novels lists, based on selected books for the Readers’ Choice Awards, books that received prominent literary awards, books that appeared most frequently in the New York Times’s bestsellers list since 2000, and novels that were reviewed by the NYT between 2010 and 2015. Note the differences in Percentage of women authors for each category.

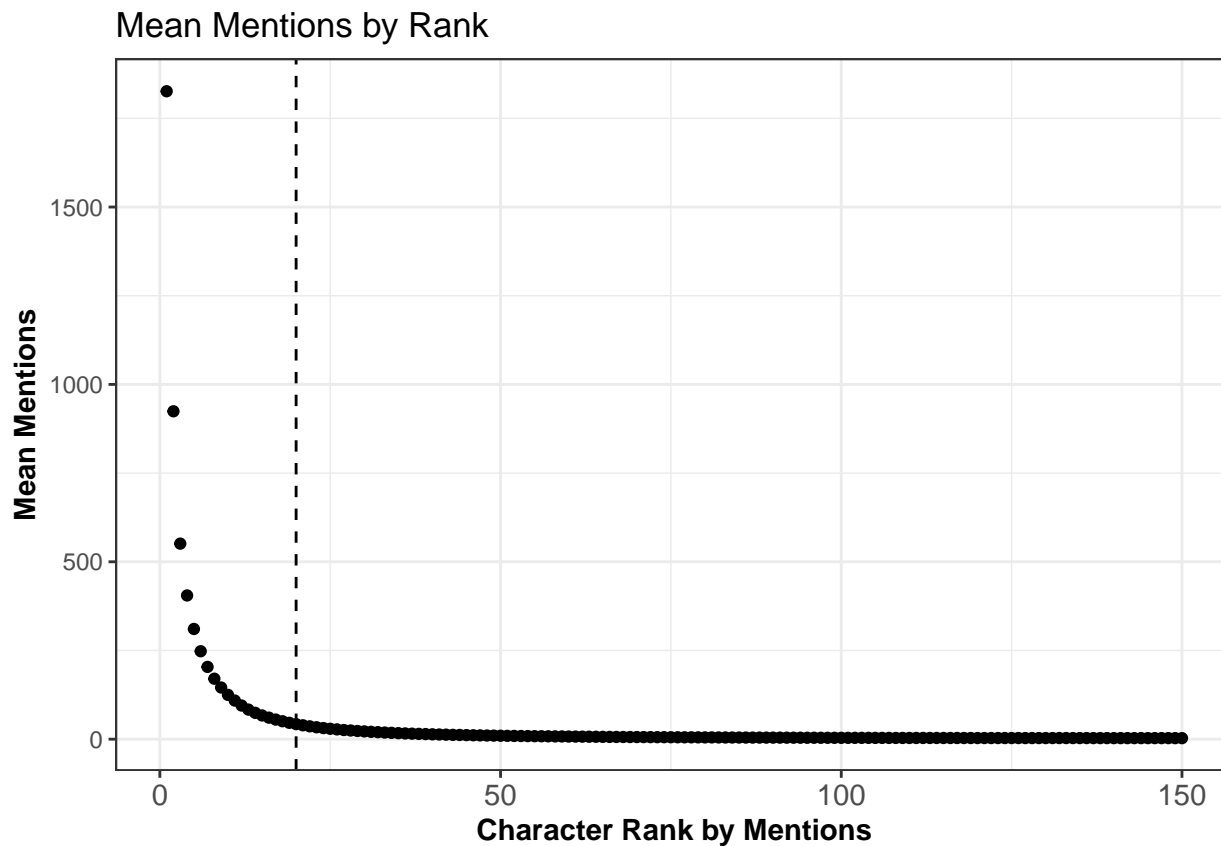
Interestingly, even though the authors report the effects that genre and author gender can have on gender bias in contemporary novels, the authors do not include the Romance category in their discussion of author gender effects - the data has 98% female authors for this category. Their paper mentions this once, but does not analyze the reasons for this or the implications it may have on Romance novel readers.

Genre ¹	Code	Novels	# Authors	% Women Authors
Science Fiction	SF	192	155	31.77
Prizewinners	PW	208	188	41.35
Bestsellers	BS	195	96	41.54
NYT Reviewed	NYT	180	179	48.89
Mystery	MY	188	140	51.60
Young Adult	YA	174	144	84.48
Romance	ROM	196	172	98.47

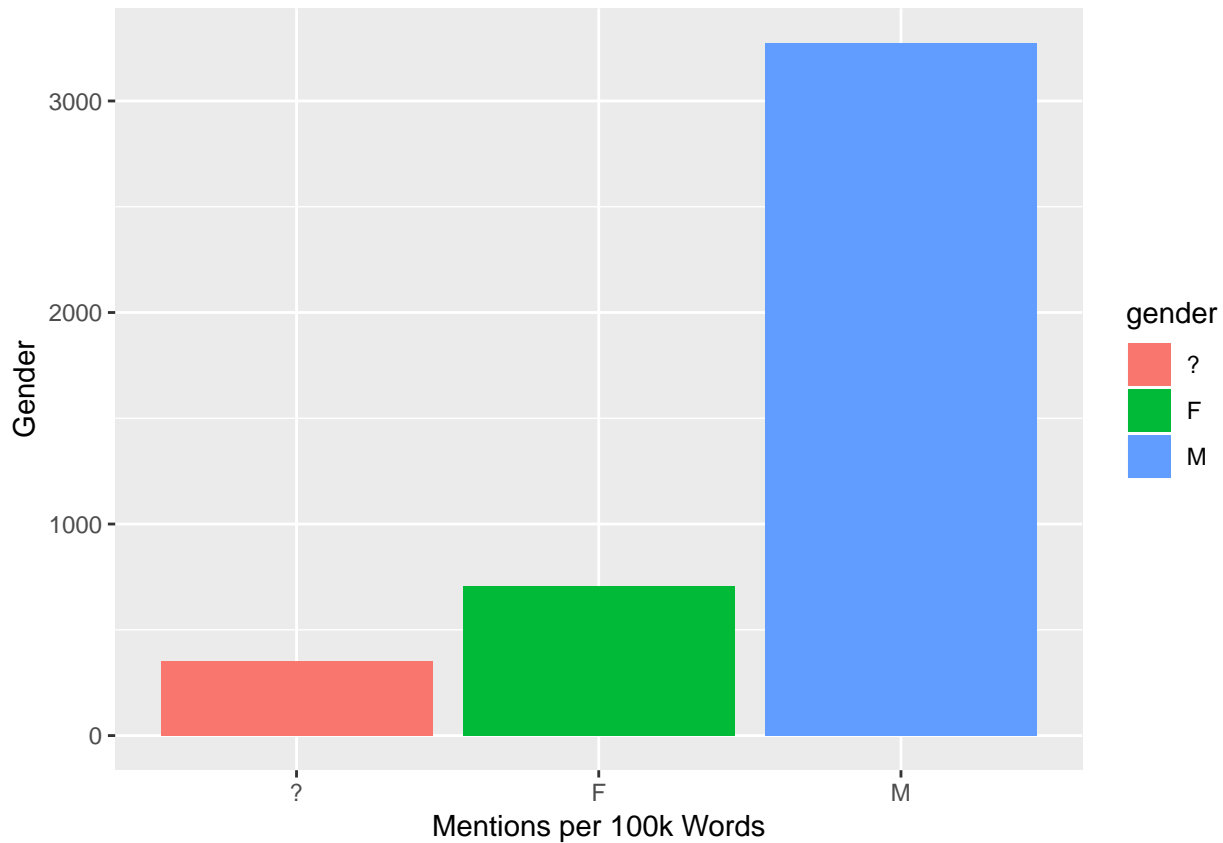
¹Table 1. Summary of our data, with the number of works per genre, number of authors. Genres are ranked by percent of women authors per genre.

The authors use the first set of figures to contextualize their research and set up their dataset for later analysis. They simply explain how their character rankings method correlates to the number of times a character is mentioned within 100,000 words of a selected text in any of their novels in the first two figures. The figures are replicated below, however it would be much simpler for whoever is reading the original paper just to get a clearer visual of something more interesting other than a black line that simply shows the nature of a ranked list. The correlation between a ranked list and the number of times a character is mentioned in a novel would be easy to understand with a short explanation included in the paper. Therefore, if this paper were to be revised, I would include simple bar graphs to show the imbalance of gender in characters who are mentioned more often within text. It becomes clear that 1) male characters are more prominent in contemporary text,

and 2) several characters in the authors' dataset are not assigned male or female. *Visualizing* these details would be a great addition to contextualizing the type of gender bias that the authors begin to work with in their later tests to find more detailed conclusions about the effect of gender in contemporary fiction.

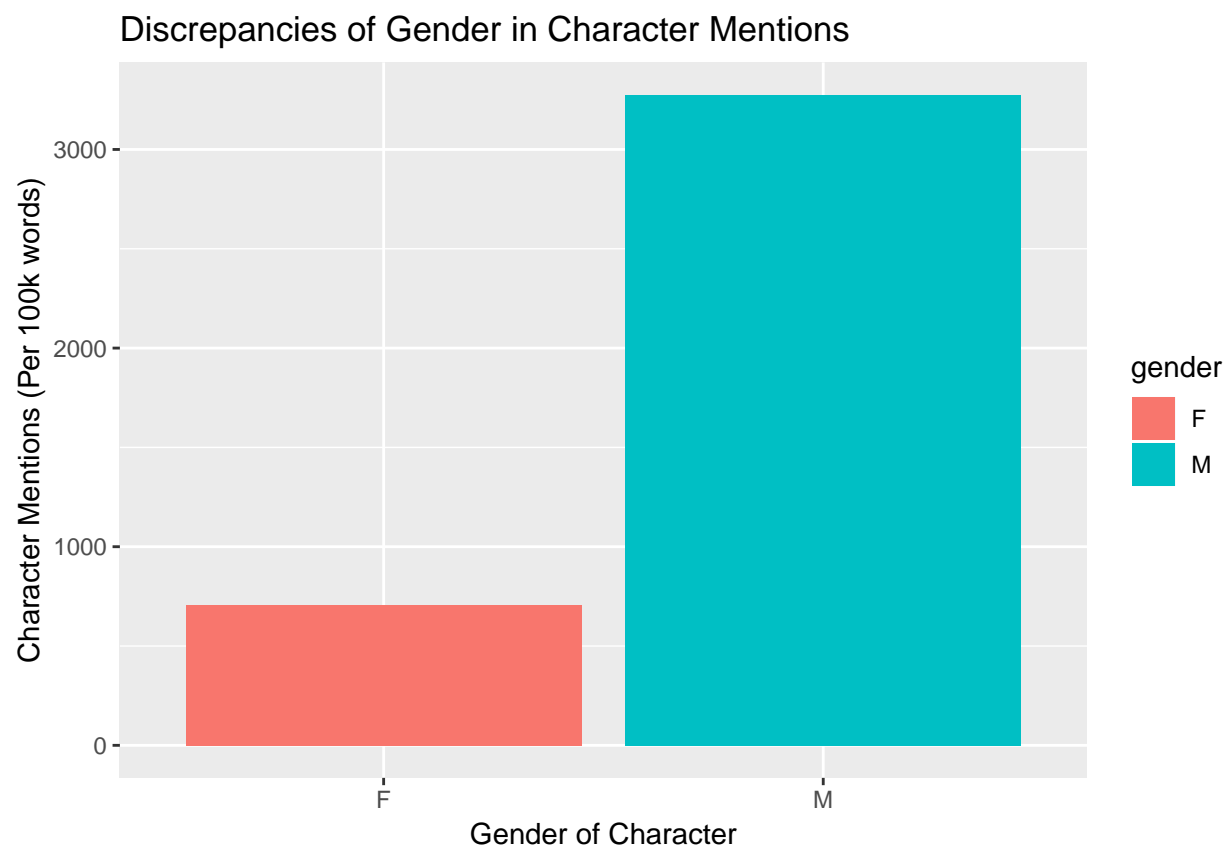
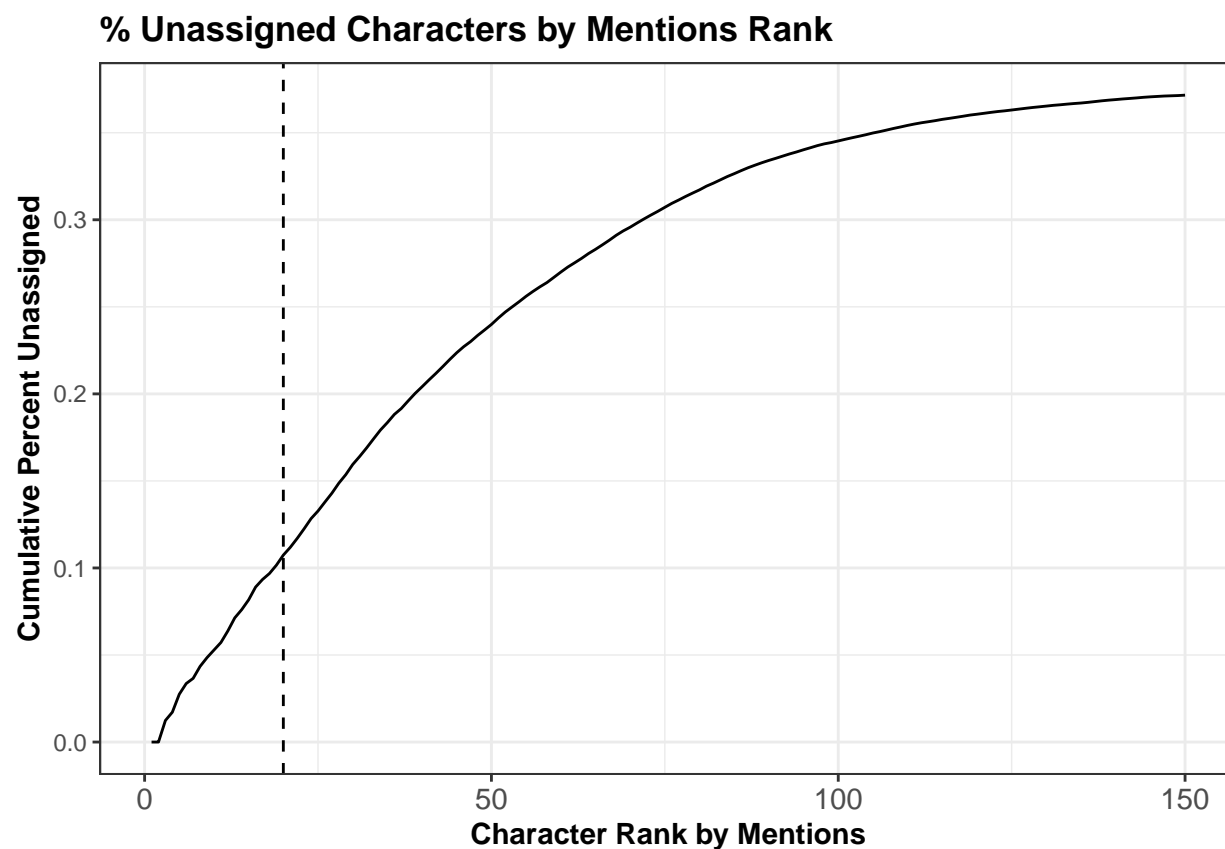


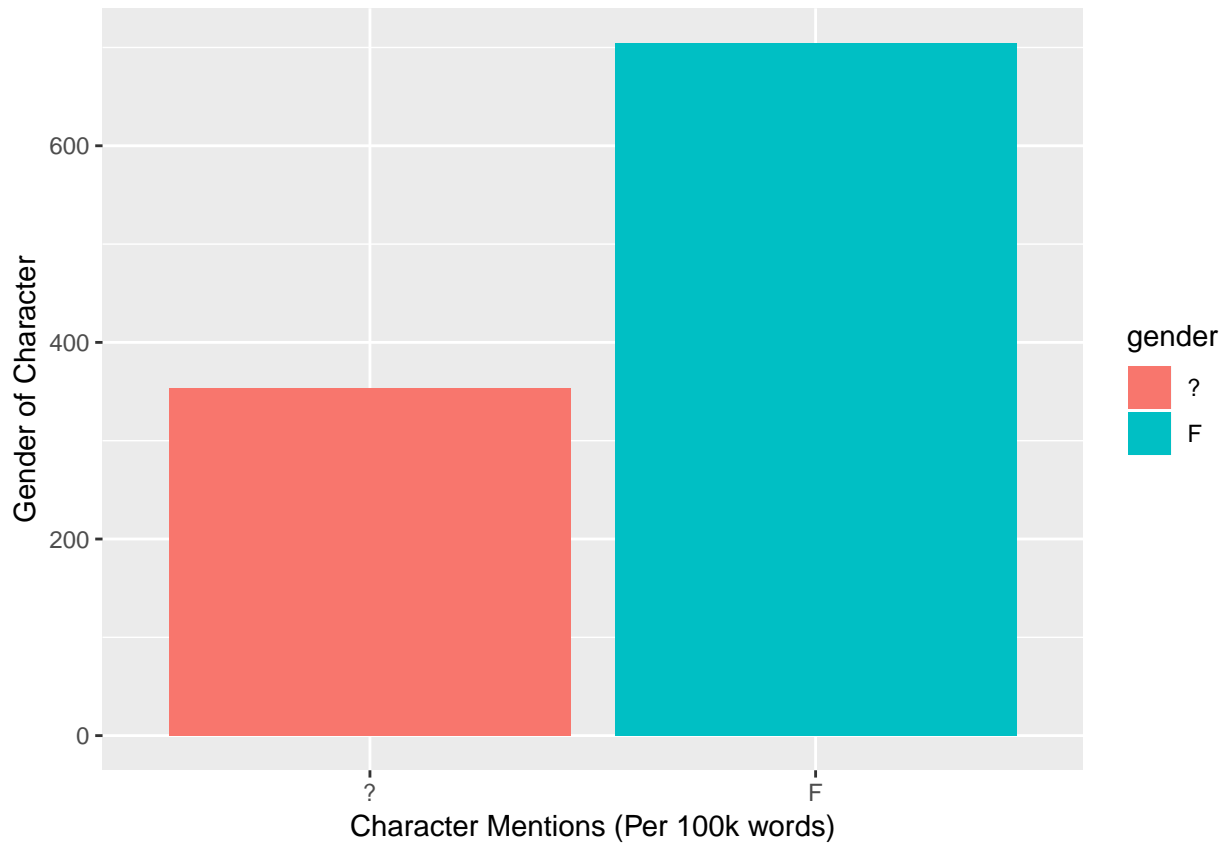
Below is an extension and improvement of the above figure.



The authors describe their “manual” process of assigning gender ID’s to each character once they acquired the data. This process included an error based on how their methods failed to assign some (about fifteen) characters either male or female. Below is, first, the replication of the authors’ graph that shows that as character mentions increase, the number of “unassigned” gendered characters decreases. However, it would be more interesting to simply focus on how significantly the unassigned characters effect the number of times a character is mentioned.

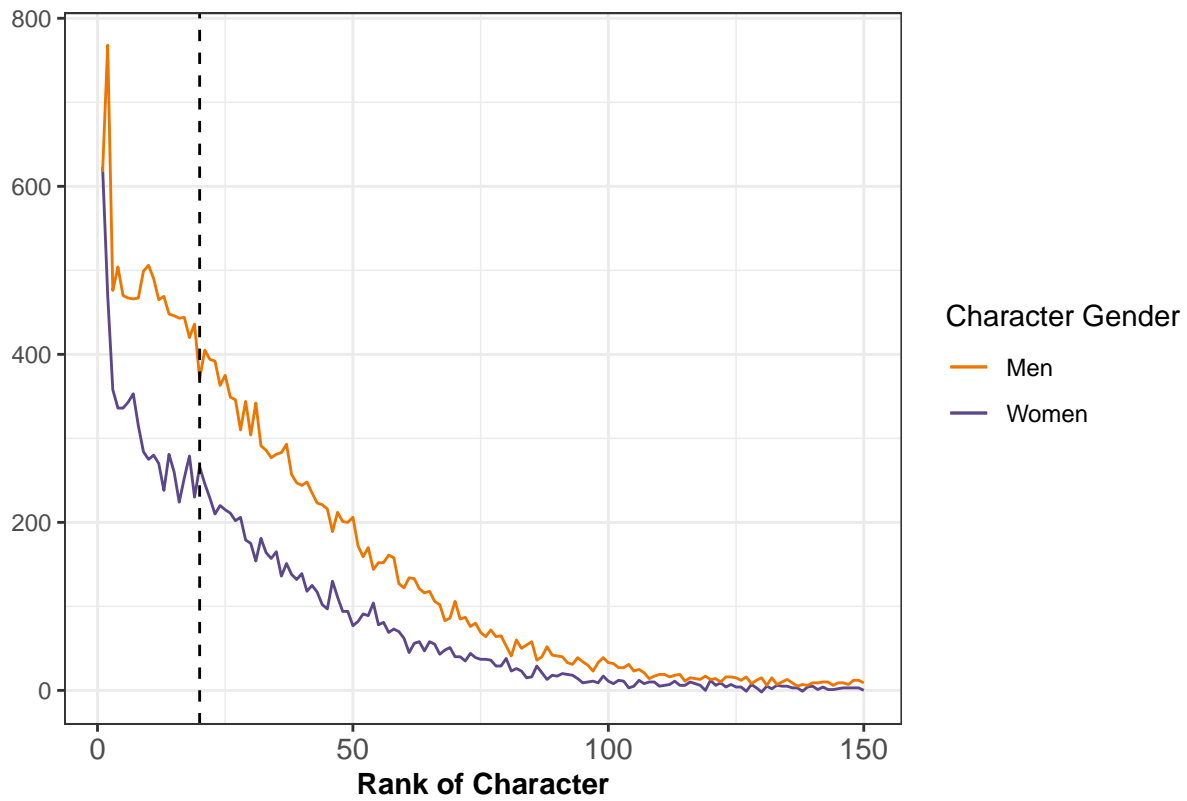
I then include another table and bar graphs to visualize this error. It is easier to understand that, by this bar graph, should the number of unassigned (marked with a “?” in the data) gendered characters be female, the number of female character mentions would be greater. Additionally, visualizing this error also calls more attention to the question of using a gender binary in contemporary fiction. What if there were other categories besides male and female?



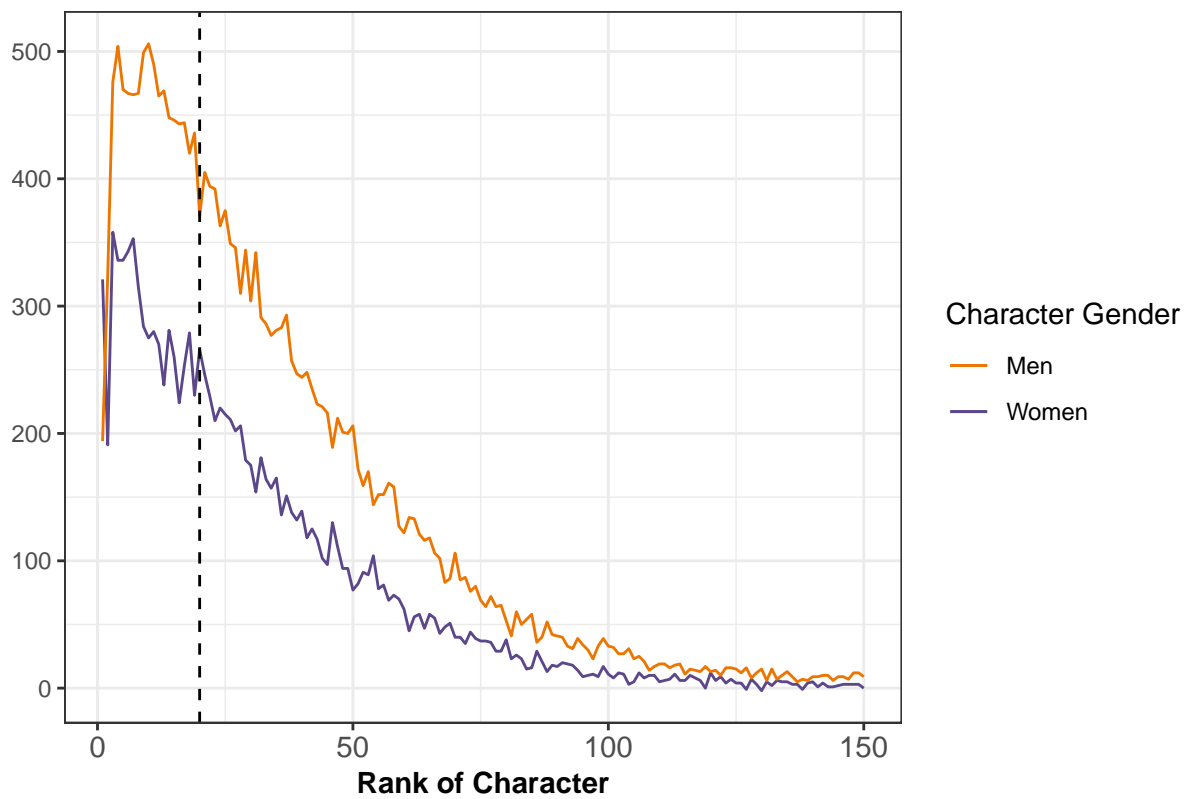


The following graphics show how gender in authors effects the prominence of male and female characters across genres. I replicate both of the author's plots as they were originally published. Interestingly, as the authors find, protagonist characters have about a 50:50 ratio of male to female characters, while all other main characters remain at about a 62:38 ratio of male to female. This is displayed in the plots. Additionally, based on their statistical methods and tests, Kraicer and Piper find that if the author is a woman, they are more likely to include a woman protagonist. The data points to the left of the dotted line show the top-20 ranked characters (i.e. most mentioned) in each novel.

Men and Women by Rank Position (All Authors)



Men and Women by Rank Position (Women Authors)



References

REFERENCES 1. David Bamman, Ted Underwood and Noah Smith, “A Bayesian Mixed Effects Model of Literary Character,” ACL 2014. 2. Eve Kraicer and Andrew Piper, “Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction,” Journal of Cultural Analytics. January 30, 2019.