This is a story of silences: of silent—perhaps silenced—women, of dead ends, and of work that was important but unmentioned in the creation of a research article.

Last February, Ted Underwood (UIUC), David Bamman (UC Berkeley), and I published “The Transformation of Gender in English-Language Fiction.” Using a program called BookNLP we read 104,000 works of fiction published between, roughly, 1780 and 2007, and available in HathiTrust Digital Library and Chicago Text Lab. One of the patterns we found is that from about 1800 to 1960 the space taken up in fiction by women characters and in the real-life publishing world by women authors steadily shrunk.¹ We were surprised: we had expected to find patterns of progress.

Description of women, as a percentage of characterization in fiction

¹Note: The source for the data is not specified in the text.
We, of course, wondered: Why were there fewer and fewer women writers and women characters? The character question can partially be answered by our data. We found that, perhaps unsurprisingly, women write about women more often than men write about women: fewer women authors means fewer women characters. So, what about the women authors?

In our article, we admit that we’re still not sure what caused the decline. Searching this silence—trying to detect its parameters and looking for obvious omissions—turned out to be quite tricky and quite labor intensive. We hypothesized that the culprit might actually be our dataset and wondered if the fault might lie in what academic libraries had deemed worthy enough to purchase. To test this hypothesis, we took a couple of paths that were necessary to our research but that didn’t make it into the article:

1. The presence of romance fiction in HathiTrust
2. The status of women’s writing in review publications that guided library acquisition
We began by investigating the presence of romance fiction in HathiTrust against the presence of westerns, two genres that have traditionally been geared toward women and men, respectively. After gathering a sampling of romance and western titles from a few bibliographies, which was manual (not digital), labor-intensive work, we found that HathiTrust seems to treat these genres relatively equally: just over half (56%) of our sample of romances and just under half (46%) of our westerns, published between 1900 and 1950, were found in HathiTrust. So, HathiTrust actually contains more romances than we had expected. On the one hand, this data vouches for HathiTrust’s representation of genre fiction, but it also shows that the gap of women writers and characters can’t be explained by a serious neglect of romance fiction. We had reached a partial dead end.

When we extended our inquiry further into the 20th century, we found that the percentages of both westerns and romances drop precipitously. So, although HathiTrust represents both romances and westerns pretty equally, its representation of genre fiction declines dramatically.

The possibility of a relationship between genre fiction and women’s authorship prompted us to investigate library acquisitions of genre fiction more broadly, and we turned to review publications that guided these purchases. These publications showed us that librarians were not encouraged to purchase genre fiction. For example, at the end of a short review of one of the “problematical books” from 1940, Virginia Kirkus, intrepid reviewer and founder of *Kirkus* reviews, sums up her judgment of a romance as “Rental tripe” (284, vol. 8). According to Kirkus, this particular novel should be relegated to the rental shelf of a bookstore. Although this research confirmed our suspicions that genre fiction wasn’t privileged, it still didn’t provide a satisfying answer to our questions. It’s even possible that men were better represented in these
reviews considering that Krikus dedicates sections to westerns but not to romances. Another dead end.

We still aren’t sure why the percentage of women authors declines for about 150 years, but nevertheless, this work did teach (or remind) us a few things. For one, it can be really difficult trying to explain silences and holes. But, what also struck me about this project were the silences we created as we crafted an article. I have spreadsheets and write-ups—the products of weeks of work—that only Ted and David (and now you) know about. I was also amazed by other overlooked labor: Virginia Kirkus dedicated her life to advising books sellers and librarians, and she was certainly influential—Kirkus Reviews lives on. But, now, the reviews she wrote gather dust on a basement shelf of the library. Digital methods may illuminate gaps in our understanding and holes in our own narratives, but they are often reliant on invisible or partially visible labor. Moving forward, as we embrace digital methods more and more, we shouldn’t forget that labor because it also might be something worth listening to.
Notes

1 We have shared our data (including images) here: https://github.com/tedunderwood/character.

2 We used Geoff Sadler’s *Twentieth-century Western Writers* (1991) to collect western titles. The corresponding volume for romance titles is Leslie Henderson and D.L. Kirkpatrick’s *Twentieth-century Romance and Historical Writers* (1990), but the inclusion of historical fiction makes the bibliography too capacious for our purposes, so we opted instead for James Vinson and D.L. Kirkpatrick’s *Twentieth-century Romance and Gothic Writers* (1982).

3 We looked at both *Kirkus* and *Choice*. *Kirkus* proved more interesting given that *Choice* hardly paid any attention to fiction.

4 The copy I accessed is from Urbana, but was originally from Gary Public Library. There are initials penciled in beside the titles that (I assume) the Gary librarians wished to purchase. Both “S” and “H” wanted Maysie Greg’s *A Fortune in Romance*, and “H” even wrote “2 cop”—presumably short for “two copies”—next to their initial. So, librarians, especially those of public libraries, may not have strictly followed Kirkus’s advice, even though they subscribed to her publication.
Works Cited


