Painting (Inequality) By Numbers: Visualizing Gender Disparity in Literary Bibliographies

Jonathan D. Fitzgerald, Regis College

Prepared for MLA 2019

To begin, let’s consider the canon. Among the nearly 100 selections—stories and excerpts from longer works—in a prominent anthology of literary journalism…and I’m not going to name any names…just four were written by women. This glaring omission can be easily ascertained by the curious researcher simply by knowing two things: first, traditional gender associations of Western names and, second, how to count to at least 4.

The ease of access to this information, however, does not in any way lessen its impact. Additionally, the simplicity of this method means it is readily scalable—counting was, after all, among the first tools in humanities computing. I first got the idea to attempt to visualize these kinds of gender disparities while—what else—working, or not working, on my dissertation. Before I get into the details of the project, however, three very brief disclaimers:

• Computational analysis at scale is a blunt instrument

• Counting four women writers in an anthology can be done by hand—one in hand in fact—counting in a corpus of hundreds or even thousands calls for automation. And while automation is great at scale, it’s not great at details. So, that said…

• In an effort to include some, others are excluded

• The very first issue here is that I discuss gender in binary terms. Not a complete picture I know. And, ultimately…

• There are probably better ways to do this
• And by probably I mean definitely. But a part of what I’m going to attempt to do here is oversimplifying to make a point. Sometimes an obvious problem requires an obvious method.

Okay, so my first go of this as I mentioned was an effort to illustrate the foundational argument of my then dissertation now manuscript—that women writers are underrepresented in the history and the canon of my field, literary journalism. So, I created a data visualization based on a comprehensive bibliography of the genre, which I compiled by mining and combining already extant bibliographies.

I derived my data from five prominent source. I tried to pull together sources that would be recognizable and respected by scholars in the field. Then, I extracted the relevant information from these five sources. I used a series of complex regular expressions, which are really just advanced text searches, to extract elements such as author name, title of the work, and date of publication. Then, I used Lincoln Mullen’s “gender” package for R, which uses historical datasets to predict gender based on first names for particular countries and time periods. Lincoln and Cameron Blevins discuss their method in an excellent DHQ article in which they acknowledge, “Inferring gender from personal names is a blunt tool to study a complex subject”—again, to underscore that point.

That said, the first thing you’ll notice is that male authors far outnumber the women authors that are listed in the bibliographies. And, as the bar graph makes clear, this pattern is seen across the timeline. What was perhaps most striking to me is that from all of the source I’ve compiled, the earliest publication by a woman writer appears as late as 1936—it’s Martha
Gelhorn’s _The Trouble I’ve Seen_. Now, certainly, far fewer women than men were being published in the nineteenth and early twentieth centuries, but where, I wondered, is Catherine Williams, Margaret Fuller, or Nellie Bly, or Fanny Fern, or Nell Nelson, just to name a few?

But, I was so pleased with the way that this rather simple visualization made a rather simple truth plain, I wanted to do more.

So I can briefly show one other from a field that should be a bit more recognizable to folks here…DH! This visualization was created by combining two DH bibliographies that are publicly available on Zotero. The first is owned by Dan Cohen, who many of you know, and the other is owned by the Digital Research Infrastructure for the Arts and Humanities Consortium, a DH network in Europe. Here, we’re looking at a period between 1980 and 2012, and though there has certainly been an uptick in the number of publications by women, the disparity very much continues.

What I think is particularly important to note is that I’m not claiming that these visualizations represent a comprehensive record of all publications in a field—that would be impossible. Rather, in both cases what I wanted to show are bibliographies that matter, that are actually being used by scholars in a field. At the risk of appearing to overly flatter our host, as Lauren so eloquently writes in the concluding paragraph of her essay “The Image of Absence," we need to think of the archive—and I’m suggesting bibliographies are a kind of first draft of archives—we need to think of the archive “not as a neutral repository of knowledge, but instead as a tool for exposing the limits of our knowledge.” So, by simply counting authors in these
bibliographies and visualizing the results, I’m attempting to re-present the bibliographies as depictions of their own disparities.

Works Cited

