

# Big data analysis reveals staggering extent of gender inequality in creative industries

28 August 2019, by Cath Sleeman



Credit: Shutterstock

The term "big data" may bring to mind swaths of private information held by tech companies. But lots of big data is, in fact, visible to all—we just may not think of it as "data".

If you've been to the movies recently, you will have seen a dataset of credits—listing the cast and crew members alongside their roles. While the credits from any one film may not be that useful, the credits from every film can form a big dataset. At [Nesta](#) and [the PEC](#) (a [new policy](#) and evidence centre for the creative industries), we have been exploring how these types of non-confidential big datasets can shine new light on gender representation in the creative industries.

Gender representation has traditionally been gauged using surveys of workers. But most surveys haven't been going for that long and it can take several years (after launching a new survey) before we can tell how the gender mix is changing. Also, surveys often don't go beyond counting the number of [women](#) and men—and so can't shed light on how prominent each group was in the [creative process](#), or how they were portrayed in a particular art form.

## Digging deep

We looked recently at the [media's reporting of women in the creative industries](#) using more than half a million articles from The Guardian newspaper, published between 2000 and 2018, from sections of the paper relating to the creative industries (such as Books, Film, Fashion and Games).

In the past five years, there has been a large increase in references to women. From 2000 to 2013, less than one-third of gendered pronouns within articles (for example, "he" and "she") referred to women. But this began to change in 2014—and by 2018 the percentage of gendered pronouns that were female had reached 40%. By contrast, the gender mix among workers in the UK's creative industries has remained flat in recent years, and [sits at around 37%](#).

We also studied the words that followed the pronouns "he" and "she", to gain insight into the media's portrayal of creative workers. This led us to discover that, compared to men, there was greater focus on particular sounds made by women, such as "laughs", "cries", "giggles", and "coos", and non-verbal reactions, such as "smiles", "grins" and "nods". These words were never used frequently, but when they were used, they were more likely to be referring to women than men (compared to other words).

In contrast, words relating to past creative achievements and leadership activities more frequently referred to men. For example, you're much more likely to see "he directed" than "she directed", and similarly "he performed", "he designed", "he managed" and "he founded". This finding is consistent with the long-running gender imbalances in the creative industries.



Big data from The Guardian offers a valuable insight into gender equality in the media. Credit: [Shutterstock](#)

[In another study](#), we used a dataset from the [British Film Institute](#) (BFI) that contained the credits from every UK feature-length film released to cinema.

After the BFI inferred people's gender from their first names, we found that the on-screen gender mix hasn't changed meaningfully since the end of World War II—and in 2017 women still only made up around 30% of cast members and 34% of crew members.

This dataset also showed gender-based differences in the jobs of on-screen characters. Since 2005, for example, only 16% of on-screen "doctors" (in unnamed roles) have been played by women, which jars with the fact that women make up [46% of doctors in the UK](#).

### Creative fairness

We are by no means the only researchers showing the potential of non-confidential sources of big data to inform gender metrics in the creative industries. [Researchers at Google](#), in collaboration with the Geena Davis Institute, used facial and speech recognition technology to show that in the 100 highest-grossing live-action films in the US, in each year from 2014 to 2016, women occupied just 36% of screen time and 35% of the speaking time.

While big data studies can enrich diversity

measures, there are two important sources of potential bias. First, we're almost always inferring gender—from a face, a first name or a single pronoun—and so we may get a person's gender wrong. Second, these inference methods typically only detect "male" and "female", excluding or misclassifying anyone who identifies with a non-binary gender. For these reasons, big data methods are not a replacement for surveys—as surveys allow people to self-identify and opt out entirely.

Even bearing in mind these potential biases, there are still many [big data](#) sources that could shed new light on gender imbalances, if only they were made available to researchers. For example, access to the stills and subtitles of films and television programmes could be used to evaluate diversity schemes, while access to the content of more newspapers would enable a broader study on the media's reporting of creative workers.

To realise the potential of these new methods, we need to encourage and support creative organisations to securely share their non-confidential data. That will hopefully allow researchers to get a little more creative about measuring [gender](#) equality in the UK's [creative industries](#).

This article is republished from [The Conversation](#) under a Creative Commons license. Read the

[original article](#).

Provided by The Conversation

APA citation: Big data analysis reveals staggering extent of gender inequality in creative industries (2019, August 28) retrieved 20 October 2019 from <https://phys.org/news/2019-08-big-analysis-reveals-staggering-extent.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*