CS 506 Final Report

See Jane | Public Opinion Analysis

Yanjie Chen, Blake Abel, Houjun Tang

Dec. 2021

**Abstract and Introduction**:

Geena Davis Institute on Gender in Media (GDI hereafter) collaborated with Boston University’s Spark team to build their Public Opinion Analysis project. Our team’s goal is to help our client the GDI use data they’ve collected to analyze their impact on accomplishing their mission, to make a positive impact on gender representation in media. The data we had access to includes news, broadcast and social media throughout the year 2004 to 2021. Our goal was to find public opinion trends towards more gender parity on screen in relation with the Institute’s efforts and research.

**Background:** Geena Davis Institute(GDI) :

Geena Davis Institute was founded in 2004 by Actor Geena Davis.  
GDI is a research-based organization working collaboratively within the entertainment industry to create gender balance, foster inclusion and reduce negative stereotyping in family entertainment media.

Meltwater:

Meltwater is a Media & Social Media Data Monitoring Platform that provides raw data and analysis for organizations to make informed decisions.

**Goal of the project:**

Our goal in this project is to interpret the organic data and evaluate the impact of the Institute on the salience of gender in media. Another main goal was to create analytical strategies that can be replicated for years to come by the institute, easily replicating our results when they have access to more years’ worth of data in the future. To measure impact the SPARK team will look at media and news coverage of GDIGM since its founding in 2004, to evaluate how often the Institute is reported on and how it is associated with gender parity onscreen (and “gender in media”).

After our team’ s review of our client’ s proposal and requests, we formed five major questions to be answered:

* What correlations do we see between “Geena Davis” and the “Geena Davis Institute (on Gender in Media)” with mention of “gender in media” (or “gender parity in media,” “gender balance in media,” etc.) in news/social/media coverage since 2004? Are there any other trends associated with Geena Davis and GDIGM since 2004?
* How many articles have been written about GDIGM and Geena Davis’ work toward gender parity and better representation on screen since its founding in 2004?
* Is there any relationship between the publishing of our reports on google searches for topics in those reports?
* What are people’s overall sentiments toward GDIGM on social media? How do people respond to coverage of GDIGM and our published reports/findings?
* How well-known is GDIGM in the Entertainment Industry (via media coverage, searches, or social media mentions, etc.)? Is GDIGM associated with improving gender portrayals in television and Hollywood?

We set out to answer each one of these questions, and the results can be found in our Analysis and Data section.

**Technology & Procedure**:

Our team for this project has only three people on it, fewer people than other projects’ teams, so we decided to all work collaboratively and divided the work evenly between us most of the time. We did the bulk of our work on a group call together, so we could collaborate and help each other eas we completed our portions of the work.

We set up meetings with our clients, and did our preliminary data collection off of Meltwater. We gathered data from all three datasets Meltwater has available, from the different sources (News, Radio/TV Broadcasts, and Social Media) it collects. Each dataset includes detailed information about the source and how it mentions the Geena Davis Institute.

Once we had our data, we worked backwards from our key questions to try and find the best ways of answering them and the best ways of demonstrating our findings to the people at the GDI. We found concise ways to visualize different aspects of the dataset, like the media’s sentiments toward GDI overtime. Also, combining the overall number of articles that mentioned GDI and their sentiments toward GDI, we studied the overall impact of GDI and its research.

After collecting and visualizing the data, we continued to investigate our findings and try to find explanations and reasons that could be causing every peak and negative spike in our data. In addition, oher minor factors, like the connection between Geena Davis and GDI, were studied.

**Analysis and Data:**

In this section, we will explain our discoveries and observations based on the data we obtained.

First, we confirmed the number of articles since 2004: We observed that there are 4645 articles in total with regards to the topic “Gender parity” and “Better representation on screen”. The earliest article source we could find was on Oct.4 2010 by NewsBlaze.com and the article was talking about Gender in Media Search for kids in the General Entertainment Industry. *(From D1)*

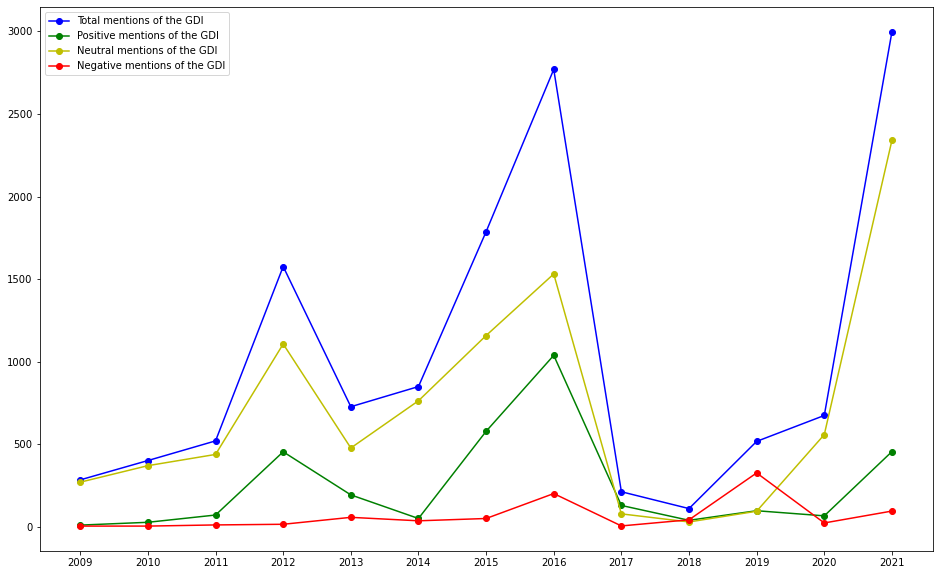
Next, the correlations we see between “Geena Davis” and the “Geena Davis Institute (on Gender in Media)” with mention of “gender in media” (or “gender parity in media,” “gender balance in media,” etc.) in news/social/media coverage since 2004 has three spikes in our findings. The first small spike is in 2012 and the other larger two spikes are in 2016 and 2021. One of our hypotheses is that this is a result from the presidential election. Media tends to publish the Gender topics to gain attention from voters, and presidential candidates.

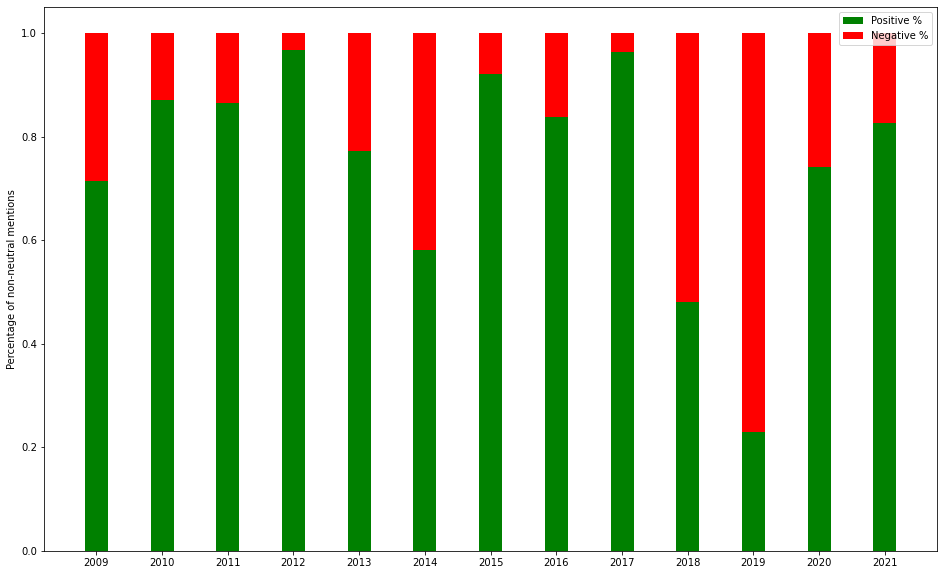
We continued to investigate with raw data from Google Trends to try and quantify how strongly correlated the GDI is with terms like “gender in media”, “gender roles”, “gender parity”, and “gender equality”. The results for each of these terms was vastly different in size, with more popular searches like gender equality eclipsing the searches for the Geena Davis Institute on Gender in Media. To get around this, we use the pearson product-moment correlation coefficients to compare the GDI against the other searches, effectively normalizing our data and providing a straightforward way to compare them all against each other. However, possibly due to the searches for the Geena Davis Institute that are tracked by Google Trends being so low, we did not find a particularly strong correlation with the GDI and any of those terms with the raw Google Trends data (the highest was Geena Davis Institute on Gender in Media and gender equality, with a value of slightly over 0.2).

Furthermore, there is a parallel relationship between the publication of GDI’s work and the google searches. Whenever there are full research reports published, there is a local peak on the google searches for Geena Davis Institute on Gender in Media. *(From D4, Page 1 of this doc)*

Now focusing on the sentiment, we looked towards other parts of the data we collected. The overall sentiment toward GDIGM is Neutral, with non-neutral opinions heavily favoring positive sentiment. Between 2020 and 2021, about 67 percent of news is neutral about GDIGM, while almost 80 percent of Social posts are Neutral. The overall neutral sentiments in news are reasonable considering that news are aimed to stay neutral. However, besides neutral sentiments, the proportion of positive sentiments is much higher than the proportion of negative sentiments. There were 6 times more positive sentiments than negative sentiments in social media, and a total of 1300 more positive sentiments than negative sentiments.*(From D2)* There was an interesting spike in negative sentiment in 2019, but we found after manual review that this was because of an NPR interview the actor Geena Davis took part in where she spoke with fervor about the reasons why she founded the GDI. During the interview, some of her appropriate language languishing the awful state of gender representation in children’s media at the time caused the sentiment analysis to flag that broadcast as negative (when in reality, it was positive given how many stations rebroadcasted that interview and how the rest of the interview was praising the decision to found the GDI). Taking that into account and removing the incorrectly flagged broadcasts, the data is in line with other years. An alternative explanation could be that the research topic of GDI in 2019 is about women in STEM, which might not be highly correlated with the previous research that has been done about the film industry and lead to a less welcoming reception.

Overall sentiment is positive and searches have increased so it seems good that it is having a good effect. And this effect is continuously rising. *(From D3)* In other words, the overall impact of GDI into the media industry since 2009 (the first publication of GDI’s research) is applaudable and progressive. The impact of GDI is likely to continue to positively influence the industry.





**Conclusions:**

The overall public’s opinion toward GDI is neutral from 2009 to now, with most of the time positive sentiments higher than negative sentiments.

**Future works:**

**References and Links:**

* [Github Repositary](https://github.com/BU-Spark/ds-gdi-gdigm-report)  
  https://github.com/BU-Spark/ds-gdi-gdigm-report
* [Historic GDI Media Data(2004-2021)](https://drive.google.com/drive/u/2/folders/1ubN9aBJuNkpi4lA_-pskEzl9ncwMDnqk)https://drive.google.com/drive/u/2/folders/1ubN9aBJuNkpi4lA\_-pskEzl9ncwMDnqk
* [Geena Davis Institute Main Website](https://seejane.org/)  
  https://seejane.org/
* [Meltwater](https://app.meltwater.com/home)  
  https://app.meltwater.com/home
* [Shared Google Drive from BU Spark Team](https://drive.google.com/drive/folders/1LPnlkQ_RRKm1IlU33OTdY_Et8bLpM8Ai?usp=sharing)  
  <https://drive.google.com/drive/folders/1LPnlkQ_RRKm1IlU33OTdY_Et8bLpM8Ai?usp=sharing>

Code used to generate the graphs and work with the data output from Meltwater is in the repo, but we also attached a version to this report for convenience! Thank you for working with us this semester.