Sushmita V Gopalan
Dr. James Evans
SOCI 40133
March 16, 2018

Judge a Magazine by its Cover (Story)

In May 2017, Hadley Freeman wrote about how male journalists have been sexualizing their female interviewees for many years now. In 1999, a cover article by Steven Daly about the singer Britney Spears (who was, at the time, a minor) in Rolling Stone Magazine opens with "Britney Spears extends a honeyed thigh across the length of the sofa, keeping one foot on the floor as she does so." A 2014 profile of Scarlett Johansson, by Anthony Lane, for the New Yorker says “Johansson was, indeed, gilded to behold. She seemed to be *made* from champagne” before making reference to “Johansson’s backside, barely veiled in peach-colored underwear”.

Rich Cohen wrote for Rolling Stone in 1995 that “Alicia Silverstone is a kittenish 18-year-old movie star whom lots of men want to sleep with.” Twenty years later, in a Vanity Fair interview of Margot Robbie, described by the actress herself as ‘really weird’, he wrote that “She can be sexy and composed even while naked, but only in character.” Rob Haskell, another male journalist, wrote in his interview of Selena Gomez on Vogue, where the singer speaks with great clarity and insight about mental health that “As I….tie an apron around her tiny waist, I wonder whether her legions have felt for years the same sharp pang of protectiveness that I’m feeling at the present.”

This paper is an attempt to analyze the prevalence and nature of this anecdotally observed condescension and sexualization by employing computational content analysis tools. Do men interviewing women spent inordinate column space on physical descriptions? Are the topics discussed in each interviewer-interviewee gender pair category different? Do some categories have more diversity of subjects, more camaraderie, more hostility than others? Are the images that are chosen to appear with interviews of women more sexualized than those that appear with interviews of men?

Section 1 describes the data collection process and the composition of the data used in this analysis. Section 2 delves into an analysis of the actual words found in the corpus – conditional frequencies, parts-of-speech tagging, statistically significant collocations and document divergences. Section 3 explores an unsupervised clustering approach to the data and finds that organic clusters evolve centered around themes and networks, rather than gender. Section 4 is an analysis of the sexual content of the images associated with each interview using a residual neural net and finally, Section 5 validates the analysis in Section 4 using human coders.

1. Data

The corpus used for this analysis consists of cover stories from Vanity Fair, an English-language monthly published by Condé Nast Publications in the United States, from April, 2012 to February, 2018. Each cover story is a profile and interview of a prominent celebrity – typically in cinema, fashion or sports. This data was gathered by automating a Google search with the keywords 'Vanity Fair cover story (month) (year)'. The first five links found were scoured to find the link to the relevant cover story and then the page with the story was scraped to collect information on the name of the journalist, photographer, stylist, headline, text of the interview, tags associated with the interview and date of publication. The URLs pointing to the top five images in each interview was also recorded. The gender of each journalist and interviewee was hand-coded. Note that for the purposes of this paper, gender is treated as binary. One of the interviewees Caitlyn Jenner (formerly Bruce Jenner) was coded as female after her transition and as male before.

This gives us a total of 58 interviews, accounting for a handful of months that had more than one volume with different cover stories and a few months where the cover story was not about a single individual. Table 1 summarizes the split across genders for journalists and interviewees. 27 interviews are conducted by female journalists and 31 by male journalists. Of the individuals interviewed, 40 are women and 18 are men.

Table 1: Composition of Data

	Female Interviewee	**Male Interviewee**
Female Journalist	**22**	**5**
Male Journalist	**18**	**13**

2. Corpus Linguistics

Distribution of Length

First, we explore the distribution of the lengths of the interviews. Figure 1 (a) shows that on average, male journalists write longer articles than female journalists do. Male journalists write articles that are, on average, 800 words longer than female journalists. Male journalists' articles are on average, 4775 words long while female journalists write articles that are 3900 words long, on average. As Figure 1(b) depicts, the differences in the distribution of length across male and female interviewees are narrower. Female interviewees are on average, written about in 4370 words and male interviewees in 4330 words.

Buzz Bissinger's profile of Caitlyn Jenner in April, 2015 is a clear outlier in terms of length at over 11,000 words. This profile, with photographs by Annie Leibovitz was iconic because it was her first interview after transitioning from male to female and delves deeply into her journey as a former Olympian and transgender woman in the Kardashian family, which consists of some of America's most visible – both loved and reviled, public figures.

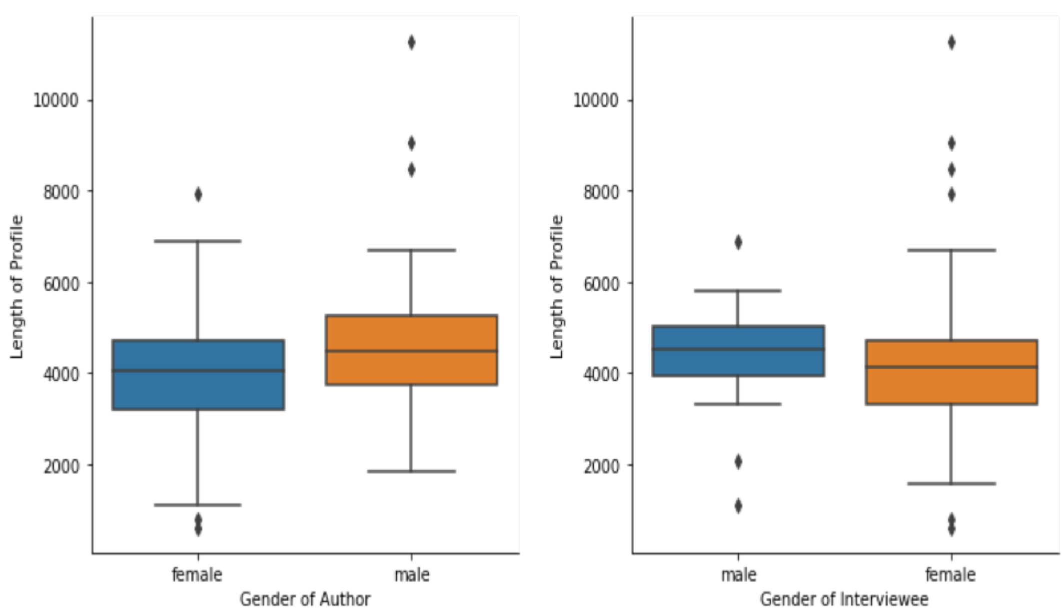

Figure 1: (a) Distribution of Length of Profile across Gender of Author, (b): Distribution of Length of Profile across Gender of Interviewee

Breaking this up further by gender pairs across journalists and interviewees in Figure 2, reveals a curious pattern. Both male and female journalists write shorter articles when they're interviewing individuals of the same gender than when they're interviewing someone of the opposite gender. Average length is highest for men writing about women (5111 words) and lowest for women writing about women (3766 words). Further exploration into the differences in the content of these interviews could reveal interesting insights about the nature of interaction between members of the same gender and those in opposite genders. The politics of power in these interactions are also specific in this context – the interviewee is typically someone who is far more famous, visible, wealthy, and given the industries targeted by Vanity Fair, conventionally attractive than the journalist, but the journalist has the power to mediate the representation of this famous person to their audiences. While a lot of research has been conducted on the effects of gender-based dynamics in job interviews, less attention has been paid to the context of interviews in journalism.

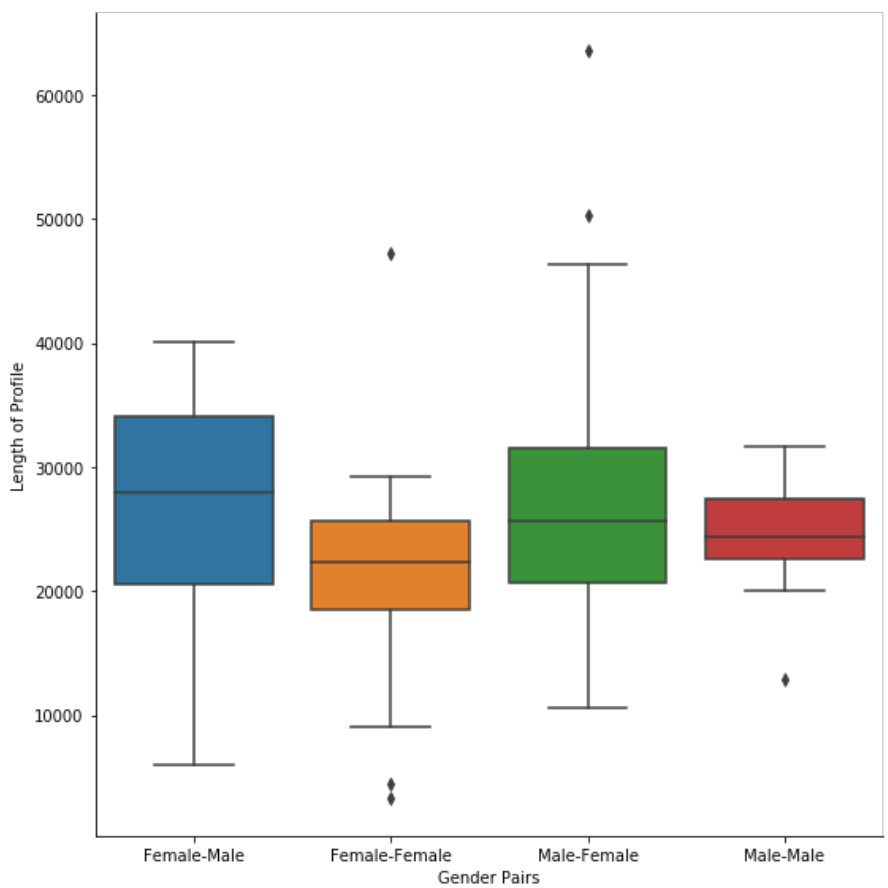

Figure 2: Distribution of Length of Profile across Gender Pairs

Frequently Occurring Words

Figure 3 shows the twenty most frequently occurring words in each gender pair. Most of the words that show up in this list are not very surprising – they are mostly words such as ‘like’, ‘said’, ‘going’, ‘she’ – verbs, adverbs and pronouns that do not illuminate differences across these categories.

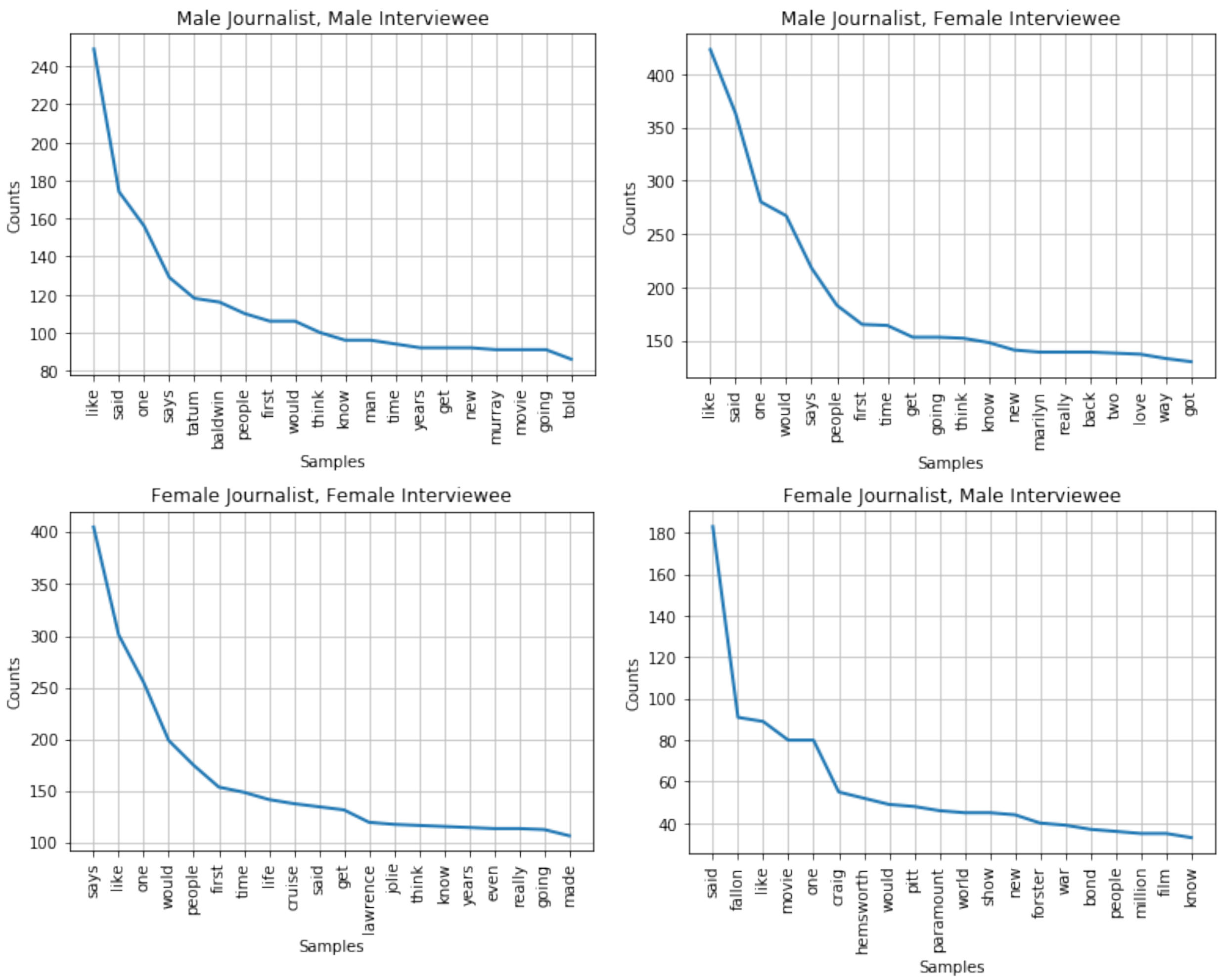

Figure 3: Most Frequently Occurring Words

For a better understanding of the differences in the issues discussed in interviews, we examine the twenty most frequently occurring *nouns* in each category in Table 2. It’s interesting here that ‘love’ only shows up in the Male Journalist + Female Interviewee category. Notable also, is the fact that the idea of motherhood or fatherhood seems to come up more in interviews where the interviewee and journalist belong to the same gender. The word ‘house’ only shows up when the interviewee is a woman – suggesting that more domestic matters are discussed when the interviewee is a woman, regardless of the journalist's gender. Also, the word 'man' comes up very often in male journalist – male interviewee articles. Naturally, words related to movies like 'studio', 'film', 'movie' and 'show' show up across all categories.

Table 2: Most Frequently Occurring Nouns, across Gender Pair Categories

FJ + FI	FJ + MI	MJ + FI	MJ + MI
time	movie	time	time
life	film	way	movie
way	kind	day	way
film	way	life	man
day	studio	movie	life
movie	time	thing	something
something	person	something	father
mother	thing	kind	actor
thing	something	show	guy
family	star	part	film
world	day	story	thing
year	show	love	show
star	year	film	day
person	story	someone	world
woman	action	world	year
actress	crew	woman	lot
kind	bit	year	kind
someone	program	lot	part
house	script	house	point

In order to probe the idea that male journalists tend to employ more sexualizing language, more references to physical characteristics of women they are interviewing, I hand-coded a list of words that imply references to parts of the body – such as 'hair', 'skin', 'lips', 'breasts', and 'waist', and then calculated the average number of times these words are used in each interview.

It turns out that men interviewing women reference parts of the body almost twice as often as women interviewing men – an average of 6.8 times per article as opposed to 3.6. Men interviewing men and women interviewing women use these words with similar frequency – about 4.6 and 4.9 times each per article.

Examining the collocations present in the corpus is not particularly insightful as most collocations across gender pair categories are of cities like 'New York' or 'Beverly Hills', names like 'Meryl Streep' or 'Jennifer Lawrence' and movies or shows like 'Mary Poppins', 'Mad Men' or 'Hunger Games'. Frequently occurring bigrams and trigrams in the corpus are also rather similar with the notable exception of the trigram (according, knowledgeable, sources) being the fifth most common trigram in the interviews of women, whereas it does not even show up in the top twenty most common trigrams in the interviews of men – suggesting that perhaps speculation is more common in popular media when the subject of an article is a woman. It could also suggest that women are not seen as primary, authoritative sources of the telling of their own stories.

Closely related to this is an inspection of the number of times an interviewee is quoted directly in an article – this was measured by just counting the number of double-quotes employed in each article. Surprisingly, articles where the interviewee is a man employ far fewer direct quotes (14 and 16 times per article for male and female journalists, respectively) than those where the interviewee is a woman, irrespective of the gender of the journalist. When a woman interviews a woman, she uses direct quotes 28 times on average, while a man interviewing a woman uses them 34 times on average. A logical next step to this would be to further examine who those direct quotes are attributed to – the interviewee themselves or other people.

Finally, we examine the degree to which these categories of text diverge from each other. The Kullback-Leibler (KL) divergence is a measure of 'relative entropy' – the degree to which one probability distribution diverges from another. It is an asymmetric measure – i.e., the KL divergence of A from B is the same as the KL divergence of B from A. This makes it possible to visualize document divergence using a heatmap. Figure 4 shows the KL divergences among the four gender pair categories. Lighter colours indicate greater divergence. The very dark coloured centre of the map shows that the male journalist- female interviewee and female journalist- female interview subsets are very similar to each other. Interviews of men by men are a little different, and interviews of women interviewing men are very divergent from the rest of the corpus. It could be a result of the fact that our corpus only has 5 interviews of men by men that this category doesn't register as very divergent from the interviews of women.

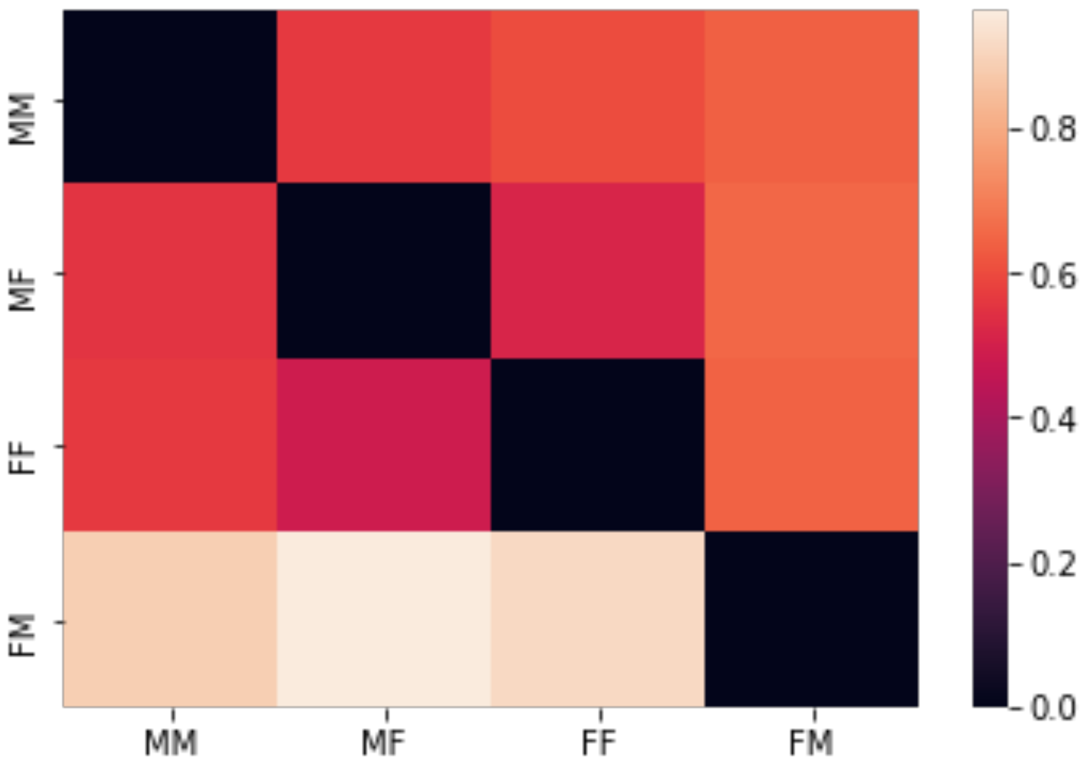

Figure 4: KL-Divergence

3. Clustering

k-means clustering is an unsupervised machine learning method used to partition data iteratively into k clusters. Feature similarity is used to cluster data points around 'centroids'. Each centroid is a collection of feature values that defines the resulting cluster. k-means clustering allows us to find and analyze groups that form organically. Since our analysis focuses on four subsets of the data, we first try and cluster the corpus of interviews into four groups. We first use Principal Component Analysis (PCA) to reduce the dimensionality of the text data and then apply the k-means clustering algorithm to the selected set of features. We see from the Figure 5 that the k-means algorithm does not do a good job of predicting the right category for each interview.

Figure 5: k-Means Clustering (k=4)

However, it's worth noting that clustering sometimes happens along axes we do not anticipate. While our analysis focuses on gender as an axis that cuts through the data – the algorithm could be picking up on a different feature that creates these organic clusters. One thing that we can do to examine this is look at the distinguishing terms identified by the algorithm. The terms in Table 3 show that clustering has happened along a different axis – each of the clusters contain words and people that we know to be intimately related with each other.

For instance, Cluster 1 seems to pick out the interviews that talk about music with words such as 'music', 'songs', 'Nashville' (the home of American country music) and 'Taylor Swift'. Cluster 3 has 'cruise', 'kidman', 'scientology' and 'marriage' – this makes sense because Nicole Kidman was married to Tom Cruise, who is a well-known follower of Scientology. Interviews of both Kidman and Cruise are present in the corpus and it is easily conceivable that they both talked about these issues. Cluster 4 has 'lawrence', 'like', 'schumer', 'hemsworth', 'jennifer' and 'hunger'. Jennifer Lawrence has often been documented talking about Amy Schumer and Liam Hemsworth – whom she considers best friends, and starred in the Hunger Games trilogy. Schumer and Lawrence are also rumoured to be writing a comedy film together.

Table 3: Distinguishing Terms in k-Means Clustering (k=4)

CLUSTER	DISTINGUISHING TERMS
Cluster 1	swift, taylor, said, like, nashville, apple, kennedy, music, songs, conor
Cluster 2	like, says, said, just, people, tatum, jolie, don, time, think
Cluster 3	cruise, kidman, scientology, kashner, miscaviage, boniadi, tom, marriage, river
Cluster 4	lawrence, like, schumer, larson, hemsworth, jennifer, kate, just, says, hunger

Silhouette Analysis suggests that the optimal number of clusters is 2, but this too has a very low average Silhouette score and poor predictive ability. It is reasonable to conclude that the corpus considered in this analysis has too few observations and too many features to cluster effectively.

4. Image Analysis

The latter part of my analysis is an exploration of the nature of the images that are chosen to accompany each interview. Anecdotal evidence would suggest that on average, the photos accompanying interviews of female celebrities are more revealing, 'sexy' and often intended for a male gaze, which is interesting for a magazine whose readership is primarily female. In an attempt to quantify the degree to which an image is considered sexy, I take recourse to supervised classification using a deep residual neural net. In 2016, two engineers from Yahoo – Jay Mahadeonkar and Gerry Pesavento, released to Github, a model used by Yahoo to recognize images that were 'Not Suitable For Work'. The algorithm is a thin 'ResNet 50' model – a residual neural net with 50 layers and 25 filters in each layer that takes in an image as input and outputs the probability that the image contains pornographic content. An image with an 'NSFW score' of over 0.8 is tagged as 'NSFW' by Yahoo and will not display on webpages without the viewer's consent.

Since the algorithm is designed to identify sexual content, it would be useful to calculate NSFW scores for the images that appear in these interviews and examine the gender dynamics that come into play. For this part of the analysis we only use the gender of the interviewee as an independent variable because the article's journalist has no say in the selection of these images – it is an editorial decision and we do not have access to the gender composition of the editorial team.

Algorithmia is a cloud-based application that allows you to make API calls to machine learning algorithms hosted on its servers. It hosts an implementation of Yahoo's NSFW algorithm using the Caffe library, pre-trained on the ImageNet 1000 classifier dataset and Yahoo's proprietary NSFW dataset. Once we obtain a private API key, we can access the algorithm using an image URL as an input and procure a score between 0 and 1 as an output.

Figure 6 shows the distribution of NSFW scores across male and female interviewees and the results are rather stark. The scores plotted here are the average score for the first five images associated with each interview. The median NSFW score for images of women is almost 0.1 higher than that of men, with the third quartile also being much higher. Recall that the NSFW score is meant to weed out pornographic content – so it stands to reason that majority of the photos from Vanity Fair fall under a score of about 0.3. Notable outliers are Sofia Vergara, Amy Schumer and Serana Williams among women, and Jimmy Fallon, among men. Sofia Vergara's sexualization in popular media is well-documented and points to larger issues of race, representation and sexuality in Hollywood. Serena Williams' profile with a very high NSFW score was her iconic, nude, pregnancy photoshoot. It might appear counter-intuitive that comedian Jimmy Fallon's photos would be outlier on a 'sexiness' scale - however, it turns out that the image that raised Fallon's average NSFW score was one of him in a tuxedo flanked by three women in bikinis.

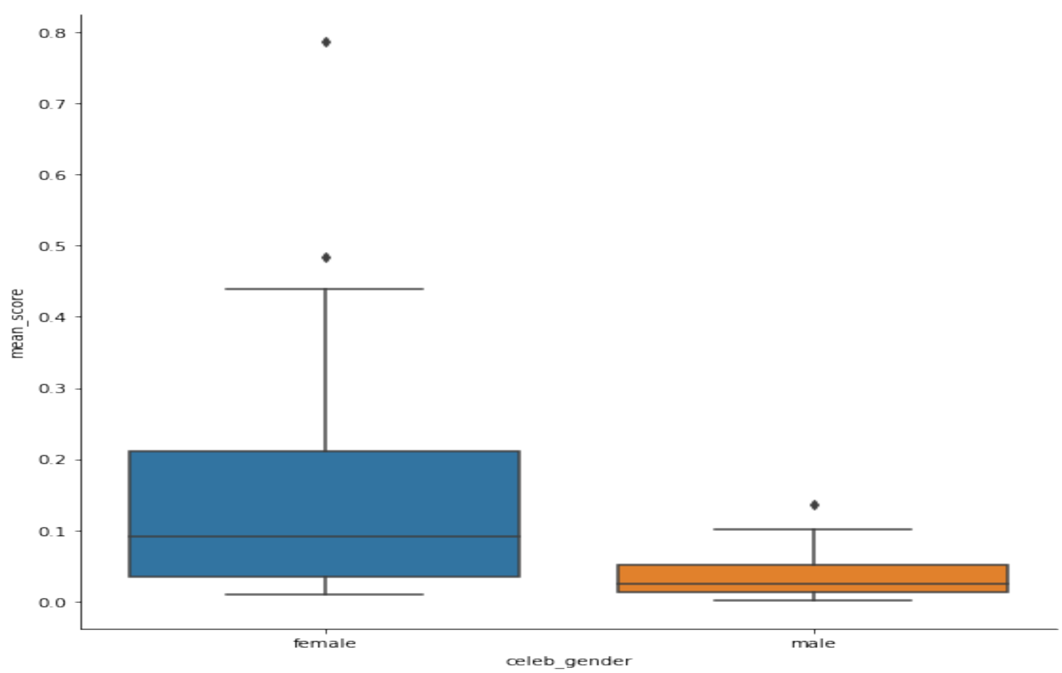

Figure 6: Distribution of NSFW Scores Across Gender

5. Reliability

Since the NSFW scale was created specifically to flag pornographic content, and what readers subjectively see as 'sexy' might not correlate with positions on the scale of 'not pornography' to 'pornography', it is useful to conduct a validation exercise to understand whether we can actually use NSFW scores as a proxy for the degree to which an image is generally considered 'sexy'.

In order to do this, we designed a toy experiment. First, each image's NSFW score was converted to a scale from 1 to 5, with scores less than 0.1 corresponding to 1, scores less than 0.2 corresponding to 2, scores less than 0.3 corresponding to 3, scores less than 0.4 corresponding to 4 and anything greater than that corresponding to 5. Then, five different coders of varying demographic and cultural backgrounds, were asked to classify 100 images each, on a scale from 1 to 5 – with 1 being 'not sexual' and 5 being 'extremely sexual'. The word 'sexual' was used instead of the word 'sexy' in an attempt to disentangle the coder's aesthetic preferences from whether or not they perceive the content in an image to be evocative of sexuality.

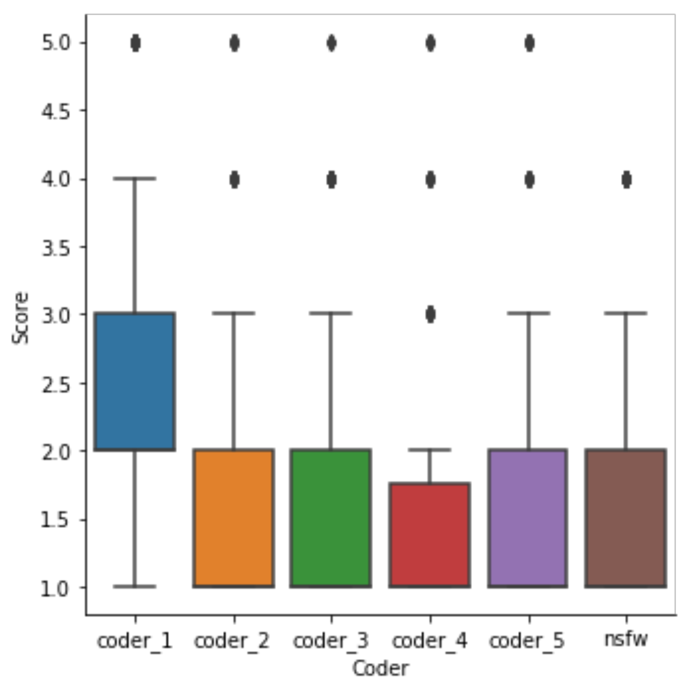

Figure 7: NSFW Scores: Coders and NSFW Algorithm

Figure 7 shows the distribution of scores provided by each coder and converted scores from the algorithm. We see that Coders 2, 3 and 5 largely agree with the algorithm. Coder 1 appears to have a lower threshold for what is concerned sexy than Yahoo's algorithm, while Coder 4 appears to have a higher threshold. Note that there is some loss of nuance while converting a range of continuous NSFW scores to a discrete scale ranging from 1 to 5.

Conclusion

Interviews are an important way for successful people in several fields, particularly those where celebrities are created, to narrate their stories, document their struggles, contextualize their accomplishments and present themselves as 'real people' to those who follow their work. Interviewers play a key role as story-tellers here - the questions they ask, the stories they choose to highlight, the context and framing they provide are instrumental in shaping the message that's ultimately communicated to the reader.

We set out to explore the differences in the ways that men and women are interviewed in popular media – with Vanity Fair being the source of interviews for this study. An exploration of the linguistics of the dataset shows that there are indeed some differences in the nature of interviews across the gender pair categories of female/male journalist and female/male interviewees. Male journalists write longer articles on average, than female journalists, but both write longer articles while interviewing someone of the opposite gender. Topics of discussion centered around family and the home are more likely to come up with same-gender pairs. Male journalists interviewing women are twice as likely to reference a part of the interviewee's body as are female journalists interviewing men. Interviews with interviewers and interviewees of the same gender have higher KL divergence with opposite-gender pairs than amongst themselves.

This points to a need to further explore the specific dynamics that come into play in gendered interactions, with a constantly shifting power imbalance. The interviewee is typically someone who is far more famous, visible, wealthy, and given the industries targeted by Vanity Fair, conventionally attractive than the journalist, but the journalist has the power to mediate the representation of this famous person to their audiences.

Since the articles in this corpus are all drawn from the same magazine within a relatively short period of five years, we can consider historical context, intended audience and influence of sponsorship as exogenous to the analysis. However, for more generalizable research, we would need a larger corpus, drawn from a variety of sources. It would also be interesting to observe the changes in these patterns across the popular-elite spectrum of magazines.

Yahoo's NSFW algorithm was used to score images associated with these interviews on a scale of 0 to 1 on sexual content. Images of women consistently scored higher on this scale than men. The algorithm's scores were cross-validated with 5 different human coders and found to be within reasonable range. Further cross-validation could be useful. It is also important to note that much like the training dataset used to training the deep neural net used by Yahoo, our set of images also lean heavily towards conventionally attractive white people – it would be interesting to see both the algorithm's scores and their validation across different cultural contexts.