

Decoding Gender Stereotypes:

Gender Portrayal on the Cover of Time Magazine from 1923 to 2022

1. Introduction

1.1 Overview

This project uses a series of machine learning models to extract the characteristics of people on the covers of Time magazine from 1923 to 2022. Various statistical methods, including chi-square test, multiple correspondence analysis, time series analysis and random forest model, were used to explore how Time magazine portrayed the men and women on its covers over the course of nearly a century. The results of the study show that the frequency of women's appearance on the cover has been gradually increasing, revealing a milestone in the media's efforts to promote gender equality. However, men still dominate and the gender representation on covers reveals some gender stereotyping issues. The project lays the foundation for future research and discussion on the impact of the media in shaping public perceptions of gender.

1.2 Background

Magazine cover, as the "face" of a magazine (Johnson & Christ, 1988), carries the important responsibility of conveying the first impression to the public. The design and content selection of the cover not only attracts readers' attention, but also carries the important responsibility of conveying social values and concepts. The choice of images and texts on the cover reflects the editor's values and the cultural tendency of the current society, among which the gender performance of the cover characters is particularly noteworthy. As part of the social and cultural construction of society, the portrayal of gender in the media tends to subconsciously influence the public's expectations and perceptions of gender roles.

Founded in 1923, Time magazine is an American news magazine based in New York City. It was published weekly for nearly a century (Wikipedia, 2024). It has played a key role in the dissemination of news and the shaping of public opinion. The selection of the people featured on its covers was an important channel for observing global news developments and cultural trends. The way magazines present men and women on their covers not only reflects societal attitudes towards gender roles in different historical periods, but also has the potential to influence public perceptions and discussions about gender equality. Therefore, by comprehensively analysing the depiction of gender on the cover of Time magazine, it is possible to reveal the gender stereotypes of society hidden behind the visual representations.

1.3 Objectives

1. Analyse the trends over time in the frequency of gender appearances of people on the cover of Time magazine.
2. Identify the existence and characteristics of gender stereotypes on the cover of Time magazine
3. Assess the impact of different factors on the gender representation of people on the cover of Time magazine.

2. Methodology

2.1 Data Collection

The data collection process for this project involved extracting cover images and their corresponding metadata from Time Magazine's official website (TIME, n.d.). This was accomplished using a web scraping methodology, utilizing Python and various libraries designed to facilitate the extraction of web content.

Step 1: Accessing and Parsing Web Page:

Utilizing Python's Requests library in conjunction with BeautifulSoup, I retrieved all the page links under each cover topic category. For pages featuring dynamic content, Selenium was deployed to emulate browser behaviour, ensuring complete loading of all elements prior to the commencement of data extraction.

Step 2: Extracting Relevant Data:

By iterating over each web link with a for loop, I used BeautifulSoup to parse HTML content to locate and extract pertinent data. The primary focus was to extract the 'src' attribute from each tag to secure the URL of the cover image. Additionally, metadata, including the publication date and subject category, was collected.

Step 3: Data Munging:

The year was extracted from the date details initially, as the project did not necessitate specific daily and monthly data. The dataset extended over an expansive temporal range from 1923 to 2022. To help with subsequent analysis and visualization, I introduced a new "decade" feature, classifying the data into ten-year brackets, such as the 1920s, 1930s, etc. Furthermore, to avoid redundancies and ensure the integrity of the ensuing analyses, duplicates within each subject category were identified and removed, preserving only the initial occurrence of each cover.

Step 4: Data Storage:

With a predefined function, I downloaded the 3258 gathered images into a local directory. The corresponding metadata is stored in CSV format, which includes "image index", "image URL", "year", "decade", and "category" for each image.

2.2 Data Pre-processing

At this stage, I used several pre-trained models on Hugging Face to recognise and extract the character features of each image, including gender, age, and body display. With these classification results, the magazine cover images were converted into data that could be used for analysis.

Human Detection

The 3258 images and their corresponding metadata obtained by crawling the TIME website in the previous section included many covers without characters or with multiple characters present. The research in this project focuses on the gender representation of single person covers, so the first step is to identify and quantify the characters in the images. Detectron2 is Facebook AI Research's next generation library that provides state-of-the-art detection and segmentation algorithms (Wu, et al.,

2019). I use this model to identify and derive the number of characters per image, retaining 1,944 images of single covers and their corresponding metadata.

Gender Identification

The gender of the person in the image is recognised by using a combination of four gender classification models from the Hugging Face (Rizvandwiki, 2023) (Rizvandwiki, 2023) (Cledoux42, 2023) (Cahoidangngu, 2024). Each model votes on the gender identified in the cover image, and if more than one model identifies the same gender, then that gender is the final result. If the models were unable to identify it, it will be corrected manually. This integrated approach enhances the robustness of gender recognition and reduces the risk of misclassification by relying on the consensus of multiple specialised models.

Age Estimation

The ages of the individuals on the cover were estimated using a model initially trained for demographic analysis (Raw, 2023). The model provided age estimates that were then classified into broader age groups (e.g., youth, young adults, middle-aged, older adults) to facilitate analyses of age-related trends in gender descriptions.

Body Display Analysis

A novel adaptation was employed where a model originally trained to classify film shots (Pszemraj, 2024) was repurposed to categorize the type of body display in the cover images. This was achieved by mapping the original film scene categories to the relevant body display types. For instance, transforming “close-up” labels into “face”, “full shot” into “full body”, and “medium close-up” into “portrait”. This adaptation is crucial for analysing how body depictions change with gender.

The pre-processed features were then cleaned and standardized to prepare the dataset for in-depth analysis. The following is a variable description table (Table 1).

Variable	Description	Levels
Category	Theme of the Magazine	arts, business, government, influence, military politics, science, society, sports, world
Year	Year of Publication	1923,1924,1925...2022
Decade	Decade of Publication	1920s, 1930s,1940s...2010s
Gender	Gender of the Cover Figure	male, female
Age	Predicted Age of the Cover Figure	Youth (0-19 years), Young Adult (20-39 years), Middle-Aged (40-59 years), Senior (60 years and above)
Body display	Body Display of the Cover Figure	face, full body, portrait, waist-up (Coutler, 2015)

Table 1 Variables Description

2.3Data Analysis Methods

Exploratory Data Analysis (EDA)

EDA is a fundamental step in data analysis, enabling the review of raw data and initial hypothesis assessment. In this project, EDA was conducted to examine the frequency distribution of categorical variables “gender”, “age”, “decade”, “body display”, and “category” within the dataset. Bar graphs, produced using the ggplot2 library in R, facilitated an intuitive understanding of the data by unveiling potential trends and disparities among categories.

Chi-Square Tests

Chi-square tests were performed to examine the independence of categorical variables as it assesses the consistency of the distribution of categorical variables. The chi-square test in this project focused on the relationship between gender and other variables such as “age”, “decade”, “body display”, and “category”. The null hypothesis is that there is no significant relationship between gender and other categorical variables. To see the results of the chi-square test more intuitively, I created mosaic plots visualising the relationship between these variables and gender.

Multiple Correspondence Analysis (MCA)

Multiple Correspondence Analysis (MCA) is conducted by assigning numerical quantitative calibration (categorical) data to cases (objects) and categories to keep objects of the same category close together and objects of different categories far apart (IBM, 2021). This visualisation helps to analyse the relationships of the variables in the dataset in more depth, exploring which categories men and women are closer to, respectively.

Time-Series Analysis

Time-Series Analysis is commonly used to study the pattern of change of a phenomenon over time. To explore the trend of the frequency of gender on magazine covers over time, I conducted this analytical analysis. Firstly, I calculated the number of male and female representatives from 1923 to 2022 to assess the time trend. Line graphs were then created using the loess smoothing method to show the trends over time more clearly.

Random Forest Modelling

Random forest is an integrated learning method that works well with categorical variables and can model complex interactions between variables. I built a random forest model to predict gender on magazine covers. Since the data set is very unbalanced between males and females, I oversampled the female data before modelling. This step improves the model's ability to learn from the dataset so that the classification results do not disproportionately represent the gender with the larger amount of data. The package “randomForest” in R was used to construct the model, assessing the variable importance to determine which factors had the greatest impact on the gender description.

3. Results

3.1 Exploratory Data Analysis (EDA)

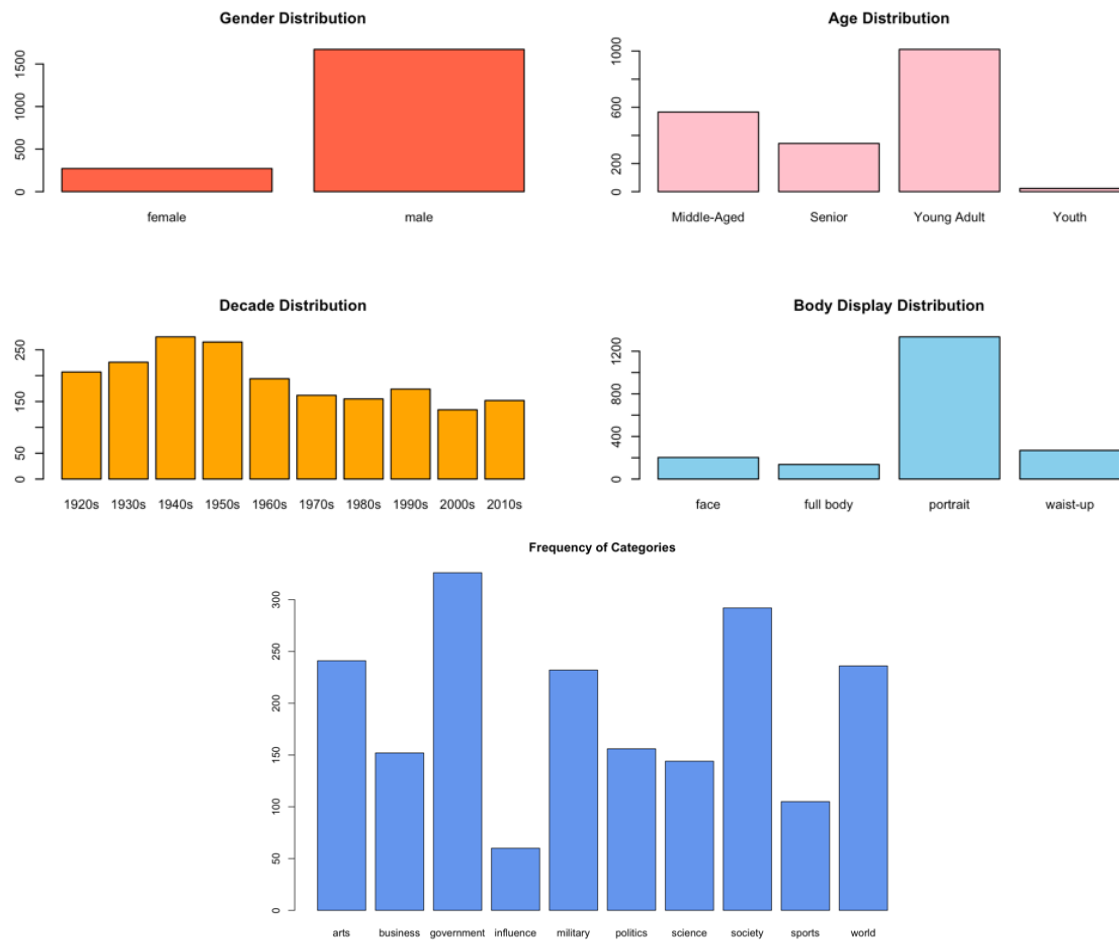


Figure 1. Variables Distribution

As depicted in Figure 1, the variable distribution graph shows a marked gender imbalance, with a significantly higher occurrence of males compared to females. This suggests a need for oversampling female data in future gender prediction modelling to enhance the model's capacity for learning. Regarding the "age" variable, a large majority are categorized as "Young Adults", followed by the "Middle-Aged" demographic, while "Youths" are least represented. The distribution of "decade" is relatively balanced; however, there are slight fluctuations in the 1940s and 1950s. In terms of "body display", a predominant number of images fall into the "portrait" category. The "category" bar chart indicates a prevalence of "government" and "society" themes, whereas "influence" is least represented.

3.2 Chi-Square Tests

In the analysis of 1944 images, 86% of the cover subjects were male (N=1672), while 14% were female (N=272).

Category and Gender

It was hypothesized that no association would exist between the cover category and the gender of the individuals depicted. Contrary to this, the chi-square test revealed a significant association, $\chi^2(9, N=933) = 192.81, p < 0.001$. Figure 2 shows the results of the Category and Gender test, which

illustrates that men are predominantly represented in categories traditionally associated with male dominance, such as "Business", "Government", and "Military". Conversely, women are better represented in "Arts" and "Society", where the gender balance is more pronounced, yet still displays a male bias.

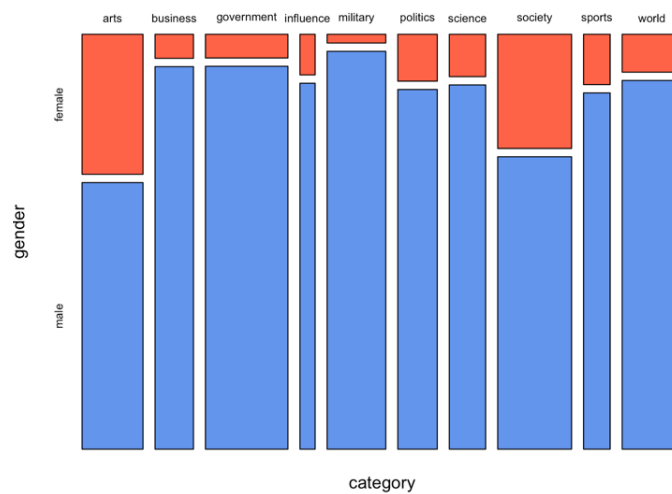


Figure 2. Mosaic Plot of Category and Gender

Decade and Gender

The null hypothesis was that the representation of genders would be independent of the decade of the cover. However, the analysis indicated a significant change over time in gender portrayal, $X^2 (9, N = 1012) = 61.597, p < 0.001$. As can be seen from Figure 3, earlier decades such as the 1920s and 1940s were characterized by male-dominated covers, while the 2010s show progress towards a more balanced gender distribution, although male images remain more prevalent.

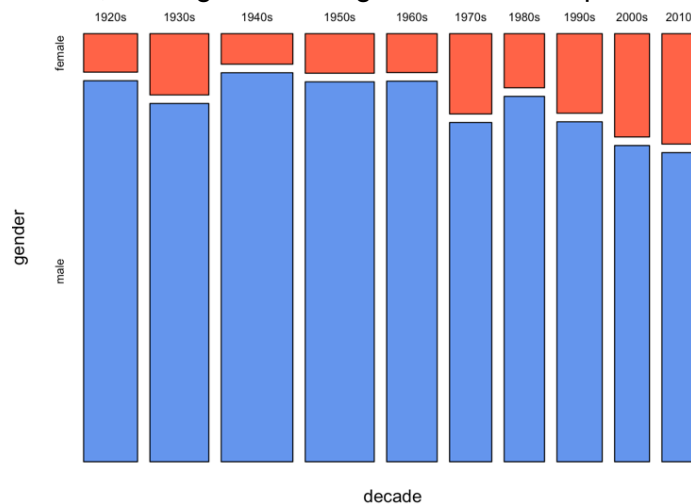


Figure 3. Mosaic Plot of Decade and Gender

Age and Gender

The null hypothesis was that the age group of the person on the cover would be independent of their gender. This was disproven by the results, which indicated a strong association, $X^2 (3, N = 1597) = 192.9, p < 0.001$. As shown in Figure 4, the "Middle-Aged" and "Senior" categories are primarily represented by males, while the "Young Adult" category includes more females, though their numbers are still lower than those of males.

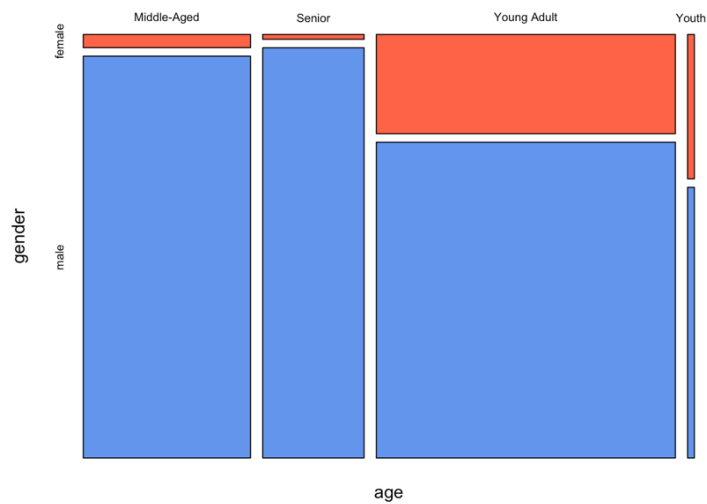


Figure 4. Mosaic Plot of Age and Gender

Body Display and Gender

The test was based on the null hypothesis that the type of body display on the cover would not be associated with the gender of the depicted individual. This hypothesis was rejected, as significant differences were found, $X^2(3, N = 1645) = 65.349, p < 0.001$. As illustrated in Figure 5, males are significantly more represented in the "Portrait" images, while "Full Body" and "Waist-Up" images tend to be closer to gender balance. Females are better represented in "Full Body" displays.

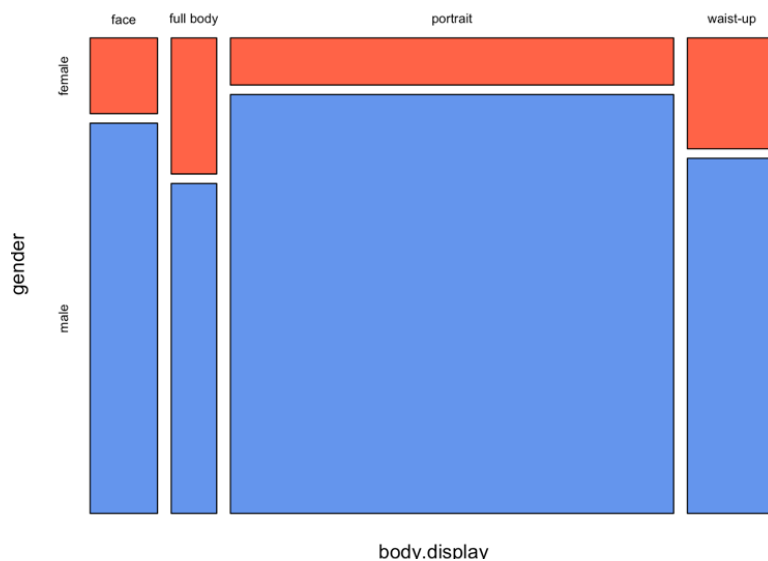


Figure 5. Mosaic Plot of Age and Gender

These results provide evidence to reject the null hypothesis in each case, confirming the presence of significant associations between "gender" and the variables of "category", "year", "decade", "age" and "body display".

3.3 Multiple Correspondence Analysis (MCA)

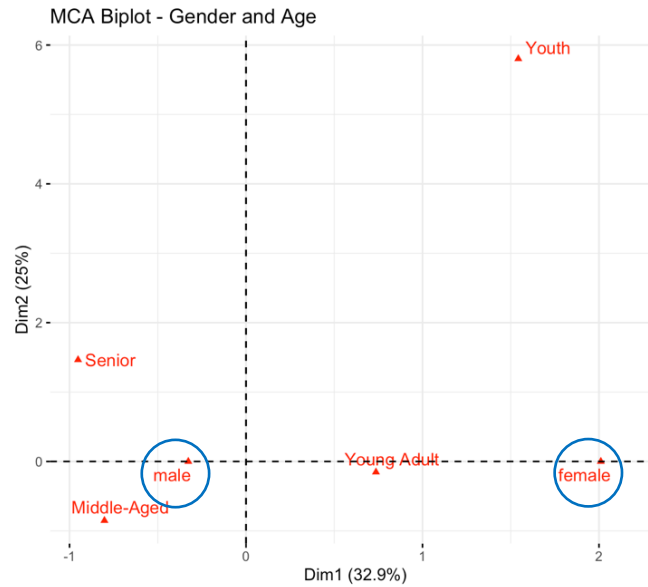


Figure 6. MCA Biplot of Gender and Age

Figure 6 visualizes the relationship between the gender and age. The "Middle-Aged" and "Senior" categories are closer to "male," suggesting that these age groups are more likely to be represented by males. Conversely, the "female" point is predominantly positioned in the "Young Adult" category, indicating a tendency to depict young adult females.

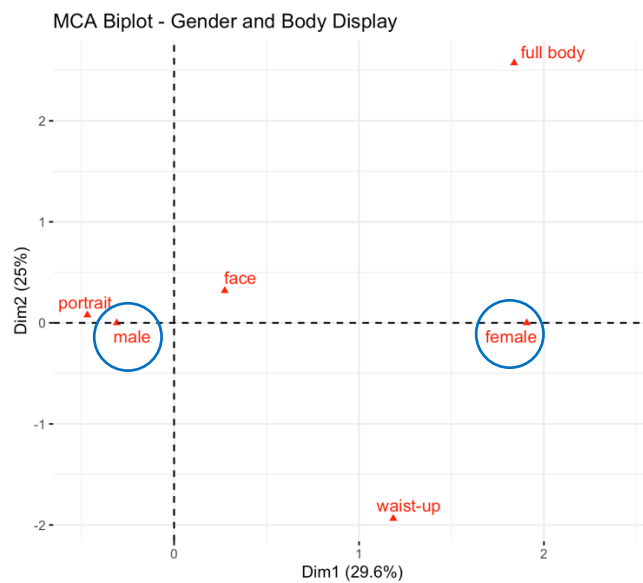


Figure 7. MCA Biplot of Gender and Body Display

From Figure 7, the "male" category is closely associated with the "portrait" display, signifying that male figures are commonly portrayed in portrait mode. Additionally, the proximity of the "face" category to "male" may suggest a focus on male faces. Meanwhile, the "female" category is closer to the "full body" display, implying that female figures may appear more frequently in full body displays.

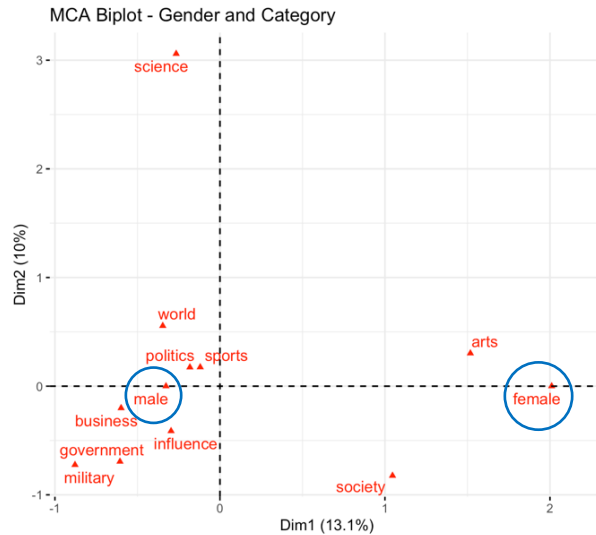


Figure 8. MCA Biplot of Gender and Category

Figure 8 reveals that the categories "business," "government," and "military" are near "male," indicating that these themes tend to be linked with male figures on the cover. Conversely, "Arts" and "Society" are closer to "female," suggesting these topics may feature more prominently female figures.

3.4 Time-Series Analysis

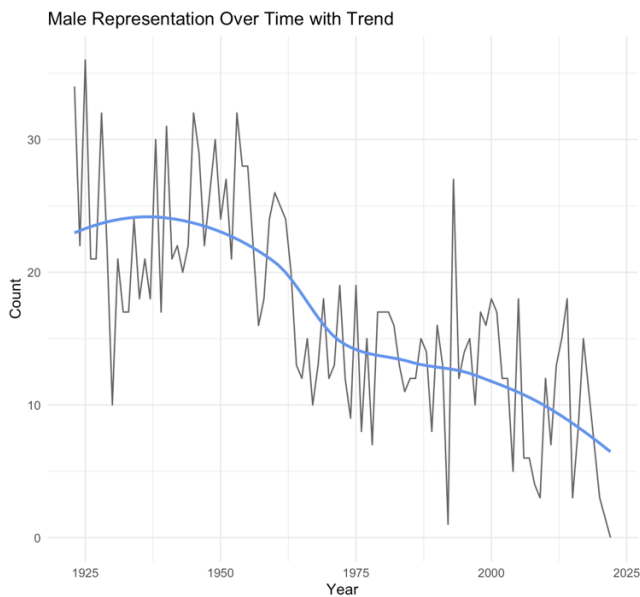


Figure9. Male Representation Over Time with Trend

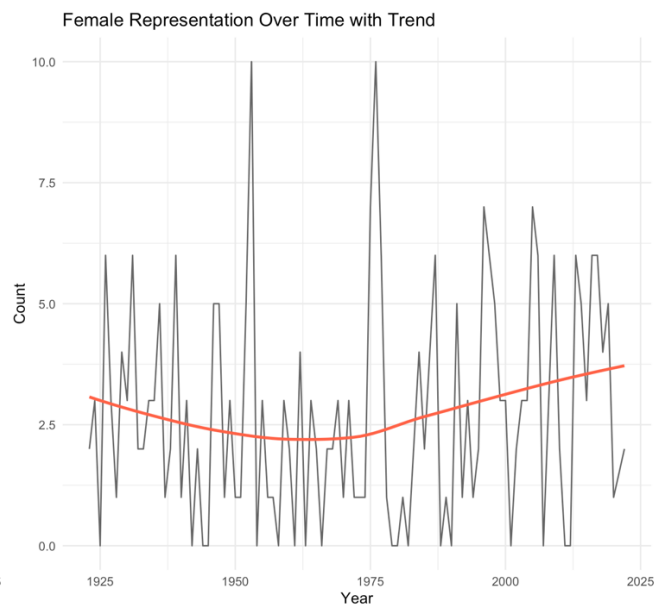


Figure10. Female Representation Over Time with Trend

Time-series analysis (Figure 9 and Figure 10) indicates a decline in the representation of men and an increase in the representation of women on Time Magazine covers over the studied period, particularly from the mid-20th century onwards. These trends demonstrate a shift towards a more balanced gender representation in recent decades.

3.5 Random Forest Modelling

The Random Forest model demonstrated strong predictive performance, achieving an accuracy of

83.93%. According to Table 2, the model categorizes female with significantly higher accuracy than male, exhibiting a lower class error rate for females. This enhanced performance for females may stem from the oversampling of female data, which likely tunes the model more closely to the characteristics of the oversampled class. Figure11 shows that the variable importance indicator identifies "category" and "age" as key characteristics, which significantly affects the model's classification decision.

	male	female	class.error
male	1007	331	0.24738416
female	95	1234	0.07148232

Table 2. Confusion Matrix

Variable Importance in Random Forest Model

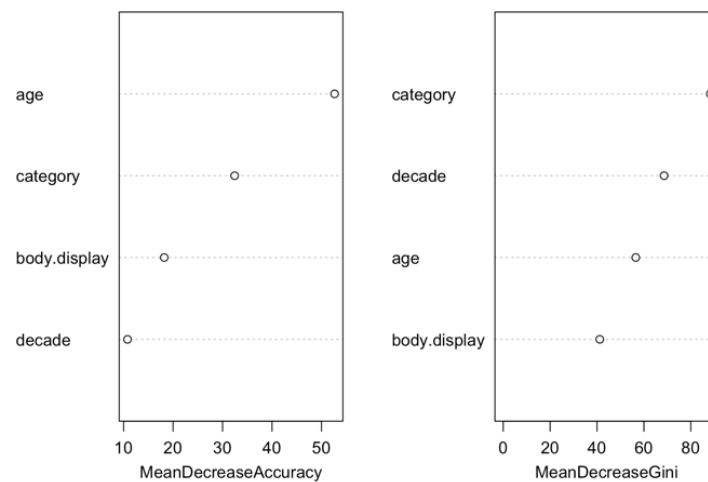


Figure 11. Variable Importance in Random Forest

4. Discussion

This section further discusses and unpacks the findings of the previous section's analysis of features on the covers of Time magazine from 1923 to 2022 and reveals gender stereotypes in the field of magazine media in the context of gender norms and historical background.

4.1 Temporal Trends in Gender Representation

The gradual increase in the frequency of women's appearances on the covers of Time magazines and the further development of gender equality from the mid-20th century onwards coincided with major social changes. In recent decades, the women's liberation movement and the gender equality movement have led to the emergence of many outstanding women from all walks of life on the one hand, and on the other hand, more and more people have become more aware of the gender imbalance that exists in the field of magazines and media. However, the frequency of male characters on magazine covers is still dominant, and the gender representation on magazine covers has not yet achieved full parity.

4.2 Stereotypes and Bias in Gender Portrayal

Firstly, the combination of gender and age on magazine covers is strongly associated with traditional

stereotypical gender roles. The categories “Middle Aged” and “Senior” are predominantly represented by men, which may reflect past social norms in which middle-aged and older men were often seen as authority figures. On the other hand, the “Young Adult” category is more relevant to women, which may reflect the fact that younger women are considered more attractive. This difference highlights how societal expectations of age and gender influence media portrayals.

Secondly, in terms of body display, the significant association of females with “full body” displays, compared to males who mostly focus on “portrait” displays, may reflect an implicit bias in the media, which emphasises women's bodies and appearances over other attributes, which may be a reflection of female objectification. This may be a manifestation of the objectification of women.

In addition to this, traditional gender roles persist in terms of magazine themes, with certain categories such as “Government” and “Business” dominated by male, while “Art” and “Society” are dominated by female. These differences in representation may solidify gender stereotypes.

4.3 Influence of Variables on Gender Representation

The importance of variables derived from the Random Forest model suggests that “category” and “age” are key factors in the modelling decision-making process, highlighting the importance of these attributes in gender representation. There is an urgent need for magazine media organisations to use these findings to reflect on and reassess the way cover characters are presented, particularly in terms of the combination of character gender, magazine theme and age stage. By positively altering the imbalance of cover characters in terms of thematic categories and age groups, Time magazine can spearhead a shift towards more equitable and diverse gender portrayals. This will not only help to enrich the miscellanea of the magazine's narrative, but it can also contribute to the reforms across the magazine media industry and the development of a wider movement for gender equality.

4.4 Limitations

There are several limitations to this project. Firstly, the project focuses primarily on the cover of Time Magazine and is not fully representative of gender representation in other types of media or cultural industries. In addition to this, there are limitations with the dataset used, such as the sample size of the dataset is not large enough and the sample sizes of males and females vary greatly, which may have an impact on the interpretation of the findings. Lastly, this project only examined the gender representation of single covers, there are still many co-existing multi-person covers of both genders and these deserve further analysis.

5. Conclusion

Through a comprehensive analysis of Time magazine covers from 1923 to 2022, this project first illustrates the major trends in gender presentation, with the frequency of female figures on magazine covers evolving significantly over time, yet more work is needed to achieve gender equality. It is then made clear that there are multiple gender biases and stereotypes on magazine covers, and that this differential depiction is often influenced by societal expectations of gender. Finally, age and subject categories were highlighted as key factors in the depiction of gender on magazine covers. These findings provide a basis for further research into the causes of these disparities and point the way for magazine media organisations to make interventions to promote gender equality.

6. Bibliography

- Johnson, S. & Christ, W., 1988. Women through Time: Who gets covered?. Journalism Quarterly, pp. 889-897.
- Wikipedia, 2024. Time (magazine). [Online] Available at: [https://en.wikipedia.org/wiki/Time_\(magazine\)](https://en.wikipedia.org/wiki/Time_(magazine))
- [Accessed 28 April 2024]. TIME, n.d. The TIME Value. [Online] Available at: <https://time.com/vault/> [Accessed 28 April 2024].
- Wu, Y. et al., 2019. Github. [Online] Available at: <https://github.com/facebookresearch/detectron2> [Accessed 28 April 2024].
- Rizvandwiki, 2023. Hugging Face. [Online] Available at: <https://huggingface.co/rizvandwiki/gender-classification> [Accessed 28 April 2024].
- Rizvandwiki, 2023. Hugging Face. [Online] Available at: <https://huggingface.co/rizvandwiki/gender-classification-2> [Accessed 28 April 2024].
- Cledoux42, 2023. Hugging Face. [Online] Available at: https://huggingface.co/cledoux42/GenderNew_v002 [Accessed 28 April 2024].
- Cahoidangngu, 2024. Hugging Face. [Online] Available at: <https://huggingface.co/cahoidangngu/convnext-fine-tune-gender-classification> [Accessed 28 April 2024].
- Raw, N., 2023. Hugging Face. [Online] Available at: <https://huggingface.co/nateraw/vit-age-classifier> [Accessed 28 April 2024].
- Pszemraj, 2024. Hugging Face. [Online] Available at: <https://huggingface.co/pszemraj/beit-large-patch16-512-film-shot-classifier> [Accessed 28 April 2024].
- Coutler, J., 2015. A Matter of Time: An Exploration of Gender in Time Magazine Covers Since 1965. Honors Projects, p. 400.
- IBM, 2021. SPSS Statistics. [Online] Available at: <https://www.ibm.com/docs/zh/spss-statistics/26.0.0?topic=categories-multiple-correspondence-analysis> [Accessed 28 April 2024].

7. Appendix

The full code for this project has been uploaded to GitHub.

Data set:

<https://github.com/Aastra01/TIME-magazine/blob/main/TIME%20dataset.csv>

Data Collection:

<https://github.com/Aastra01/TIME-magazine/blob/main/TIME%20Data%20Collection.ipynb>

Data Pre-processing:

<https://github.com/Aastra01/TIME-magazine/blob/main/TIME%20Data%20Pre-processing.ipynb>

Data analyse:

<https://github.com/Aastra01/TIME-magazine/blob/main/Time%20Magazine%20analyse.R>