Final Report

Warren Woo Zachary Lakin Justin Mai Daniel Tsemekhman

6/10/24

Exploring the spotlight in superhero sagas: What determines which X-Men characters stand out in the pages of Marvel Comics? This study delves into the factors influencing the appearance frequency, or total issues, of X-Men characters in comic books published from 1963 to 1992. Motivated by the intriguing variance in character prominence, we analyze a dataset encompassing 26 X-Men, examining 44 variables per character, including total appearances, demographic details, and estimated financial values. Employing statistical techniques such as linear regression and t-tests, we identify the significant predictors of visibility, with a particular focus on a character's estimated monetary value. Our findings shed light on the complex dynamics of narrative and market influences in the X-Men series, offering insights into how profitability might shape character development in serialized media.

Table of contents

1 Introduction

The X-Men, a cornerstone of Marvel Comics introduced in 1963, showcases a variety of mutants with unique abilities, yet their visibility in the comic series varies significantly. This study investigates what drives the prominence of certain characters over others, analyzing data on 26 X-Men characters

from 1963 to 1992 to uncover the underlying factors influencing their appearance frequency, or total issues.

Using a dataset compiled by Anderson Evans from the Marvel Fandom Database, we examine variables such as total appearances, appearance percentages by decade, estimated monetary values, and gender. The findings, based on rigorous statistical analysis, aim to offer a deeper understanding of the dynamics shaping narrative decisions in the X-Men comics. This report will begin by providing an overview of the data set and variables, followed by a detailed description of the statistical methods used. The results section will present the findings in the form of graphs and tables, and the discussion will interpret these results as well as discuss any issues or limitations with the data. Finally, the conclusion will summarize the key takeaways and suggest future directions for research on factors influencing X-Men character appearances.

2 Data and Methods

2.1 Data Collection

The data collection took place in 2022 by Anderson Evans for his article on the website Rallyrd.com (Evans, 2022). Rally is an asset investment company which allows users to buy and sell equity shares in collectible assets. The data collection took place at the Rally company, looking at websites such as UncannyXMen.net and the Marvel Fandom Database. From the data collected there are only 26 X-men in this dataset, and in the context of this study they are the observational units, Inclusion and . A computer was used to collect the data, from looking on the internet, and including keywords in Ebay search such as included issue title, issue number, and "VG," which indicates that the seller believed their comic book to be in "Very Good" condition. The original author did not state the criteria for variable exclusion, so it is unknown how observations were excluded.

The response variable that we are interested in testing is the total issues of X-men appearances, to see if the variables we choose below have an effect on an X-men's total issues. The variables we used were

TotalValue_heritage: The total value of each X-Men team member's issues as reflected by Heritage's highest sale, in dollars TotalValue60s, TotalValue70s, TotalValue80s, TotalValue90s: The total values for each decade (60s, 70s, 80s, and 90s) from Heritage. TotalValue_ebay60s, TotalValue_ebay70s, TotalValue_ebay80s, TotalValue_ebay90s: The total value of each X-Men team member's issues as reflected by Ebay's highest sale. Members: The X-Men characters. Total Issues: The total number of issues each character appeared in from the 60s to the 90s. TotalValue, TotalValue60s, TotalValue70s, TotalValue80s, TotalValue90s: Combined values from Heritage and Ebay. Sex: The X-men's sex Decade Introduced: The decade of an X-men's first appearance

2.2 Statistical Analysis

To make the total value easier to understand we merged values from Heritage and Ebay to create TotalValue, along with TotalValue60s, TotalValue70s, TotalValue80s, TotalValue90s, and are using that instead of Heritage and Ebay separately. For the methods in this section the original dataset did not include sex, and decade introduced, our group had to manually add a characters sex based on a google search, and appending the sex using the tidyverse's mutate feature. For the decade introduced, we mutated the appearance percentage by decade, and found the first instance of a character's appearance that was not zero percent, and assigned them the decade that corresponded to that non-zero percentage.

We first wanted to test if sex had an effect on total issues. So we ran a two-sample t-test, with male as one group, and female as the other group. The two-sample t-test was conducted to compare the means of total issues between male and female characters. Null Hypothesis (H0): There is no significant difference in the mean number of total issues between male and female X-Men characters. Alternative Hypothesis (H1): There is a significant difference in the mean number of total issues between male and female X-Men characters.

Furthermore the data was analyzed using multiple linear regression twice. The first time was to determine if there were significant differences in the total number of issues based on the decade a character was introduced. While the second time was with a total number of issues, based on the total value of the decade. Checking model assumptions for both models. it seems like all the FINE

(Form, Independence, Normality, Equal Variance) to determine if appropriate assumptions are met: normality is good, equal variance seems okay, and independence is met through the assumption that the the values of the comics don't affect each other, but for context there is a light issue as the sample size is small, it is only 26 X-men.

3 Results

With a total of 26 X-Men members to go off of, 15 being male and 11 being female, the mean amount of total issues was 85.27. The mean for males was 94.73 (sd = .151) and the mean for females was 72.36 (sd = 358). The highest male was Scott Summers with 197 appearances and the lowest male was John Proudstar with only 3 appearances. The highest female was Ororo Munroe with 190 appearances and the lowest females was Remy Lebeau with 17 appearances.

Member	TotalIssues	Sex
scottSummers	197	Male
ororoMunroe	190	Female
peterRasputin	169	Male
charlesXavier	169	Male
loganHowlett	167	Male
jeanGrey	164	Female
warrenWorthington	139	Male
bobbyDrake	123	Male
kurtWagner	120	Male
hankMcCoy	119	Male
annaMarieLeBeau	85	Female
seanCassidy	74	Male
kittyPryde	73	Female
alexSummers	68	Female
betsyBraddock	59	Female
lornaDane	48	Female
alisonBlaire	45	Female
ericMagnus	41	Male

jonathanSilvercloud	39	Male
longshot	35	Male
rachelSummers	24	Female
jubilationLee	23	Female
remyLeBeau	17	Female
lucasBishop	16	Male
shiroYoshida	10	Male
johnProudstar	3	Male

Table 1: The table shown displays all of the X-Men members total issues, ordered from greatest to least. 8 out of the top 10 characters are male

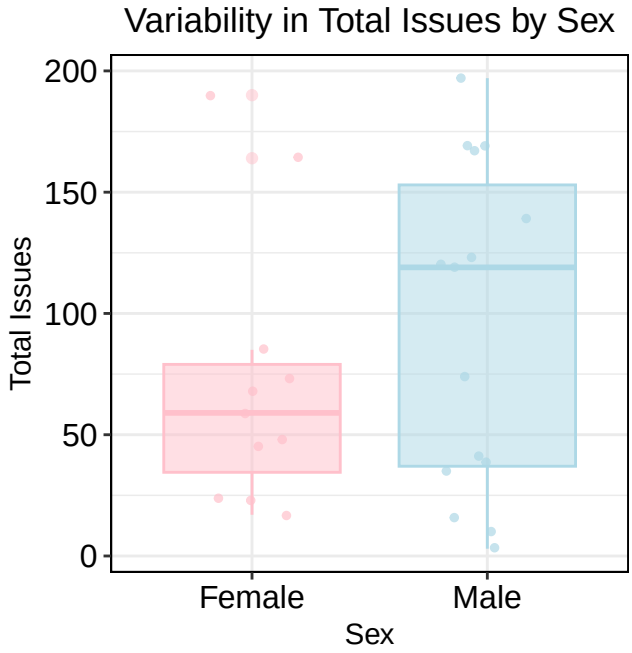

Figure 1: The two boxplots show the variability in total issues for males and females X-Men characters. The male character group had higher variability and a higher median, however the two groups had the same amount of outliers. Each dot represents one member.

Based off of Table 1 and Figure 1, it would seem that male characters typically appear in more issues than female characters. However, after running a two-sample t-test , we determined that we did not have enough evidence to say that there is a significant difference in total issue means between the two sexes (p-value = .365, df = 23, t = -.924) Despite the fact that 8 out of the top 8 X-Men with the highest amount of total issues are males (Table 1), the test revealed that there was not a significant difference in means. A 95% confidence interval of (-72.41, 27.67) would

also further support this claim as from this we can say that we are 95% confident that the true difference in means of total issues between male and female X-Men characters is not conclusively different from 0, which means that we cannot confidently say that there is a significant difference between male and female characters when it comes to total issues. While in Figure 1 we can see that male characters are more varied and do tend to be more in more issues than female characters, it is not by a large enough margin to say that the sex of a character is a significant factor in how many issues a character appears in. So in conclusion we that sex does not play a significant role in determining the total amount of issues of a X-Men.

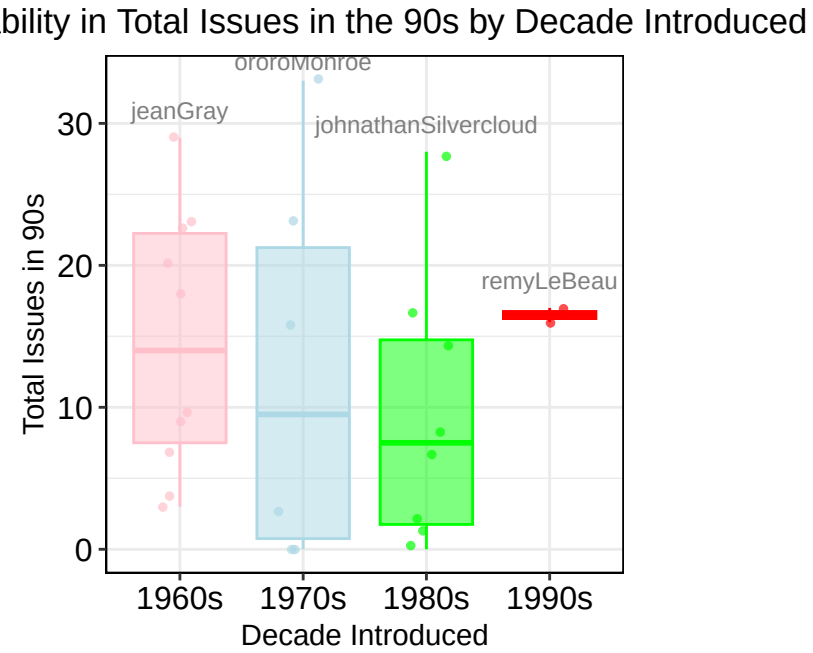

Figure 2: These boxplots display the variability in total issues for X-Men characters introduced in different decades. Characters introduced in the 1960s and 1970s show higher median total issues compared to those introduced in the 1980s, with the 1970s group exhibiting the most significant variability. All three of these groups contain multiple outliers. The 1990s group has only two characters, resulting in a narrow range of total issues. The analysis focuses on issues from the 1990s to avoid bias, as older characters would naturally have appeared in more issues over time. Each dot represents one character. The characters labeled have the highest amount of total issues in the 90s for each decade

After running a linear regression test, we determined that we did not have enough evidence to say that there is a significant difference in total issue means between decade introduced. (p-value = .723, df = 22, f-statistic = .4417) Figure 2 shows that each box plot from every decade group overlaps each other. This, along with linear regression test indicate that the total issues for characters across different decades share a considerable amount of common ground and that there is little to no evidence that the decade of introduction is a determining factor of a character's total issues.

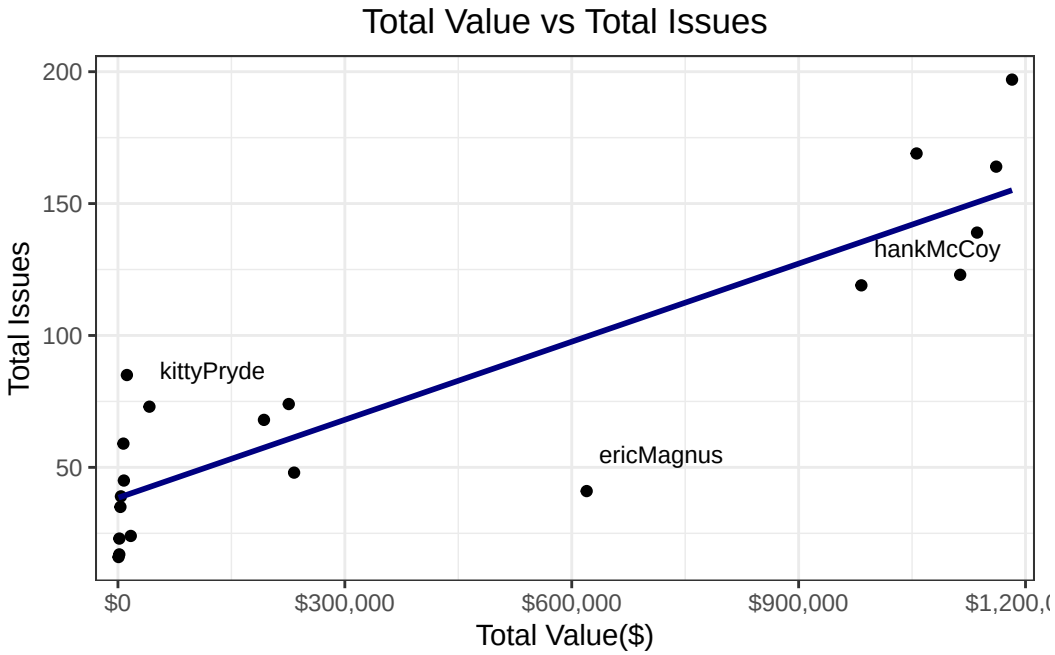

Figure 3: This graph illustrates the correlation between total value and total issues for each X-Man character. The plot reveals a positive correlation between these variables, indicating that characters with higher total values tend to have more appearances. Additionally, notable outliers are observed, representing a high amount of appearances despite relitvely smaller total value. The X-Men labeled are some of the more popular X-Men in today's pop culture landscape today

After running a regression analysis, we found that there was statistically significant evidence that the total value of a character is a strong indicator of total issues (p-value = <.01, df = 18, f-statistic = 66.1). This means that there is strong evidence that characters with a higher amount of total appearances tend to have a higher total value, and characters with a lower amount of total appearances tend to have a lower total value. As shown in Figure 3, characters with a higher total value appeared in more issues than characters with a lower total value. Using these findings, we can create more accurate predictions on which characters, based on their total value, will have more appearances going forward in the X-Men comics.

4 Discussion

Our report aimed to uncover the key variables that influence the total appearance count of X-Men characters. After conducting our analysis, we learned that the more profitable an X-Men character is, the more they're going to appear in X-Men comic-book issues. This analysis aimed to identify the key factors that influence the total appearance count of X-Men characters in comic book issues from the 1960s to the 1990s. The results suggest that a character's profitability, as measured by their estimated monetary value, is positively correlated with their total number of appearances. This finding aligns with expectations, as publishers are likely to feature popular and lucrative characters more frequently to capitalize on their success and appeal to readers. The effect of profitability on appearance count was found to be statistically significant, with a moderate to large effect size based on the regression analysis. This indicates that the relationship is unlikely due to chance and has practical significance in understanding character prominence. However, it is important to consider that while profitability is a strong predictor, it is not the sole determinant of appearance frequency. Other factors, such as narrative importance, character development, and creative decisions, also may play a role in shaping character appearances. So while the other factors we analyzed, those being sex and decade of introduction, resulted in us finding insufficient evidence that they were significant factors, we believe that there are other confounding variables that could influence a characters total number of appearances. The strengths of this study lie in the comprehensive data set, which covers a substantial period of X-Men history and includes a diverse range of characters. However, the observational nature of the data limits the ability to draw definitive cause-and-effect conclusions. Limitations of this study include the potential for selection bias, as the data set only includes a subset of X-Men characters and may not fully represent the entire franchise. While profitability and appearance count are correlated, it is also possible that other unmeasured variables are influencing this relationship. This study was not generated from a random sample, so caution must be taken into consideration when generalizing the results. Additionally, the reliance on estimated monetary values as a proxy for profitability may not capture all aspects of a character's financial success or appeal. This research contributes to the existing knowledge base by providing empirical evidence for the link between character profitability and appearance frequency in the context of X-Men comics. It highlights the importance of considering financial factors when analyzing character prominence and opens avenues for further research on the interplay between commercial success and creative decisions in the comic book industry. Future research could explore alternative measures of profitability and expand the scope of the analysis to include more characters and a wider range of publication years.In conclusion, this study reveals that profitability is a significant factor influencing the total appearance count of specific X-Men characters in comic books from the 1960s to the 1990s. While this finding provides valuable insights into character prominence, it is important to consider the limitations and potential biases of the analysis when interpreting the results. Further research is needed to fully understand the complex dynamics shaping character appearances in the comic book industry.

5 Conclusion

Ultimately, our data suggests that if you're a profitable X-Men character, you get put in more X-Men comic book issues. Within the confines of the dataset, outside of monetary value, there aren't any other statistically significant contributing factors to a X-Men's total appearances. We cannot generalize our results to other X-Men comic book characters because the data only looked was not a random sample. Drawing a conclusion to other Marvel characters would be unwise because our sample isn't representative of the Marvel population as a whole either. In other words, our research has very low generalizability; our results are only representative of our sample. Furthermore, we cannot infer causality either because there simply exists many more variables that contribute to an X-Men's total appearances, which are not included in the dataset; variables such as: an X-Men character's writer, an X-Men's amount of media appearances outside of the comics, and how many plots/storylines does an X-Men character appear in. For those interested and curious - these variables necessitate more investigation. Theoretically, there are infinitely many variables that could contribute to an X-Men's number of appearances in comic book issues, and if one was extremely curious, they could run a study including some of these variables to continue our research and attempt to answer the question: "What drives the prominence of certain X-Men characters over others? What are the underlying factors influencing their appearance frequency?" Although, from our data set, we were only able to determine that one factor, that being a character's total value, was a strong indicator in their total amount of issues, we are confident that there are more confounding factors in what goes into how many issues a character appears in. By furthering our research, we believe that these researchers will discover some groundbreaking results that could change the way comic book writers write their characters.

6 References

Marvel database. Marvel Fandom Database. (n.d.). https://marvel.fandom.com/wiki/Marvel_Database

Evans, A. (2023, June 16). Mutant Moneyball: A Data Driven Ultimate X-Men. Rally. https://rallyrd.com/mutant-moneyball-a-data-driven-ultimate-x-men/

The R project for statistical computing. R. (n.d.). https://www.r-project.org/

Zhu H (2024). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.4.0, https://CRAN.R-project.org/package=kableExtra.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wilke C, Wiernik B (2022). *ggtext: Improved Text Rendering Support for 'ggplot2'*. R package version 0.1.2, https://CRAN.R-project.org/package=ggtext.

Wickham H (2023). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0, https://CRAN.R-project.org/package=forcats.

Wickham H, Vaughan D, Girlich M (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1, https://CRAN.R-project.org/package=tidyr.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.