

Thomas Hamming

Assignment 3

This data set came from the wikias of all of the Marvel and DC characters. The data was collected by Walt Hickey. It contains 23272 rows representing the characters, with 16 columns for variables. Each recorded observation is one character, and has its own ID. Each column is a trait about the character. The variables include data about alignment, eye and hair color, sex, gay/bi characters, whether they are alive, appearances, date of first appearance.

Categorical Variables:

- Publisher – which publisher. DC or Marvel.
- Align – The characters alignment. Good, bad, or neutral.
- Eye – Character's eye color
- Hair – Character's hair color
- Sex – sex of the character. Male, female, something else.
- Alive – whether character is dead or not. Living character or dead character.

There are a number of questions I could get from this data. Does DC or Marvel have more female characters? What percentage of the total characters are female? Are male or female characters more likely to stay alive? Was more female characters a recent trend, or always the same? Are the LGBTQ characters more likely to be male or female?

Three Questions:

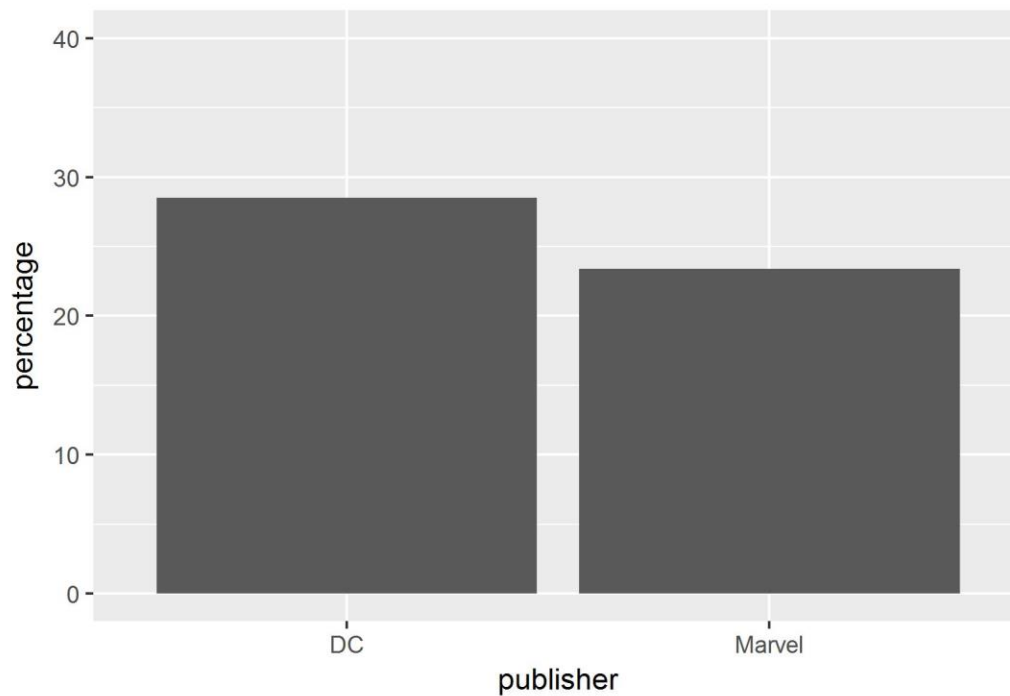
1. Does DC or Marvel have a higher percentage of female characters? I will need a count of the female and male characters, grouped by publisher. I can calculate the percentage with those numbers. Then I would like to create a simple bar graph showing which has more.
2. Are the LGBTQ characters more likely to be male or female? I need a count of all the characters identified as gay/bi/trans etc. and then I need to find number of male or female. Then I can use a bar graph with a bar for each sexual identification.
3. What date did female characters first appear? Are most female characters older or newer? Need to sort all the female characters, and then get a data set of when the first appeared. Then create a distribution based on date.

I chose these questions because they seemed most interesting on determining the situation with female comic characters. I wanted to know if there was any significant difference between DC and marvel in terms of female representation. Knowing the gender of the non-heterosexual characters could teach us a lot about why certain characters are written to be gay/bi in the first place. Finally, I wanted to know if female heroes have always been in comics, or if it is a recent thing due to societal pressures.

Question 1

For this I needed to find out what percentage of characters are female in both marvel and DC. I used dplyr to find the total number of marvel and DC characters. Then I used dplyr to find the total number of female characters for each publisher. I used these numbers to find the percentages for each

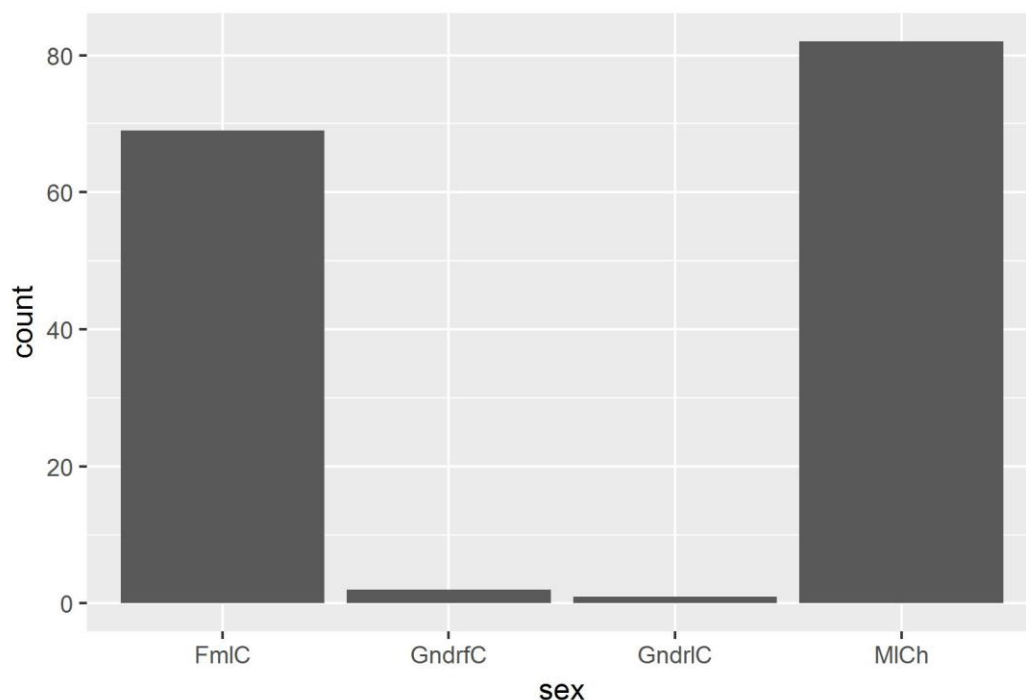
publisher. I then put these values into a new data frame, and created a basic bar graph in ggplot.



DC does have a slightly higher percentage of female characters, even with many less characters overall. However, there isn't a major difference, and both have a noticeably low percentage of female characters.

Question 2

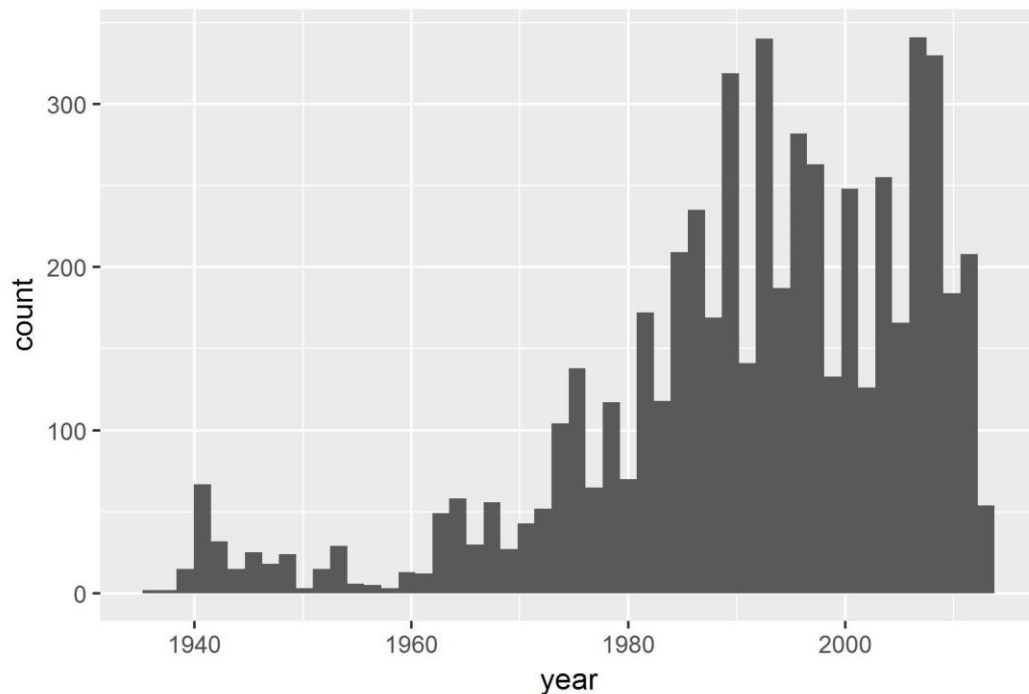
In order to find how many males and females were gay/bi, I needed a clear graph comparing the numbers of them. I filtered out all the NA values for the "gsm" variable, to get only the gay/bi characters. I then used dplyr to count the number of them and created a bar graph comparing females, males, genderfluid, and gender-neutral characters.



Not only are there a very small number of LGBTQ characters, but there are almost as many females as males. This means that a much higher percentage of the gay/bi characters are female, since there are far less female characters.

Question 3

This question was pretty simple to figure out. I just needed to graph the distribution of when female characters first appeared by year. I used dplyr to filter by female characters, and then graphed a histogram using year as the x axis.



This graph makes it very clear that almost all of the female characters in comics started after 1980. Most female characters were introduced in the 1990's or very recently.

Conclusion

It is clear that females are underrepresented in comic books. In both Marvel and DC, females make up only about 25% of the total characters. Most of those characters were not introduced until very recently, as well. I could see either two reasons for the early draught and the recent abundance (comparatively) of female characters in comics. It could be social influences, forcing creators to write more diverse comics. There could also be more women working in the comic book industry then there used to be. Female characters are also much more likely than male characters to be gay or bisexual. This is probably due to comics being a male dominated industry. I think that if more women started working on comics, and comics started trying to gain a less male dominated fan base, these problems will probably sort themselves out. Until then, however, comics will probably continue to be about men.