



Phase 3 – News Analysis

MARVEL VS DC CHARACTERS

RAHUL BAHADUR

Recently, there has been lots of furor regarding the inadequate representation of women and other genders in the movies and comics. The same can be said about the representation of non-heterosexual characters in the movies or comics. Movies have, since they fall under the main stream entertainment channels, been readily embracing such changes in public opinion about non-binary genders and non-heterosexual characters. However, the same cannot be said about comic book series (excluding the movies which are based on comic books). A recent analysis by fivethirtyeight.com revealed the poor representation of women in the comic books.

The purpose of this report was to extend this analysis to other genders as well as to characters having non-heterosexual preferences. Furthermore, an analysis has also been done to see to what degree are the “Bad” characters stereotyped. Most of the analysis has been done in a way to provide a comparative view of both the Marvel and DC universes.

All this analysis has been done on RStudio.

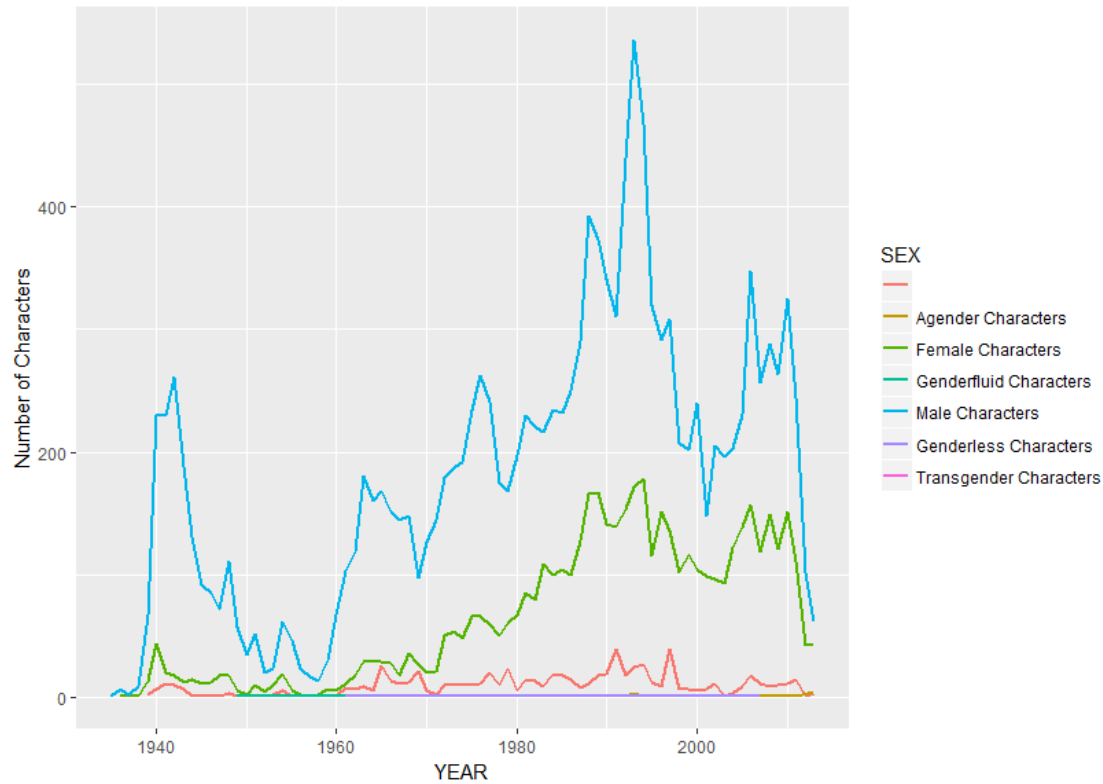
About the Dataset

This dataset describes all the DC and Marvel comic book characters that have appeared till 2013. Originally available on DC and Marvel Wikias, they were scrapped for similar analysis purposes by the website fivethirtyeight.com. Its APA citation is mentioned below.

F. (2015, March 23). *Women-in-comic-books*. Retrieved May 07, 2017, from <https://github.com/fivethirtyeight/data/tree/master/comic-characters>.

Variation of the gender of characters over the years

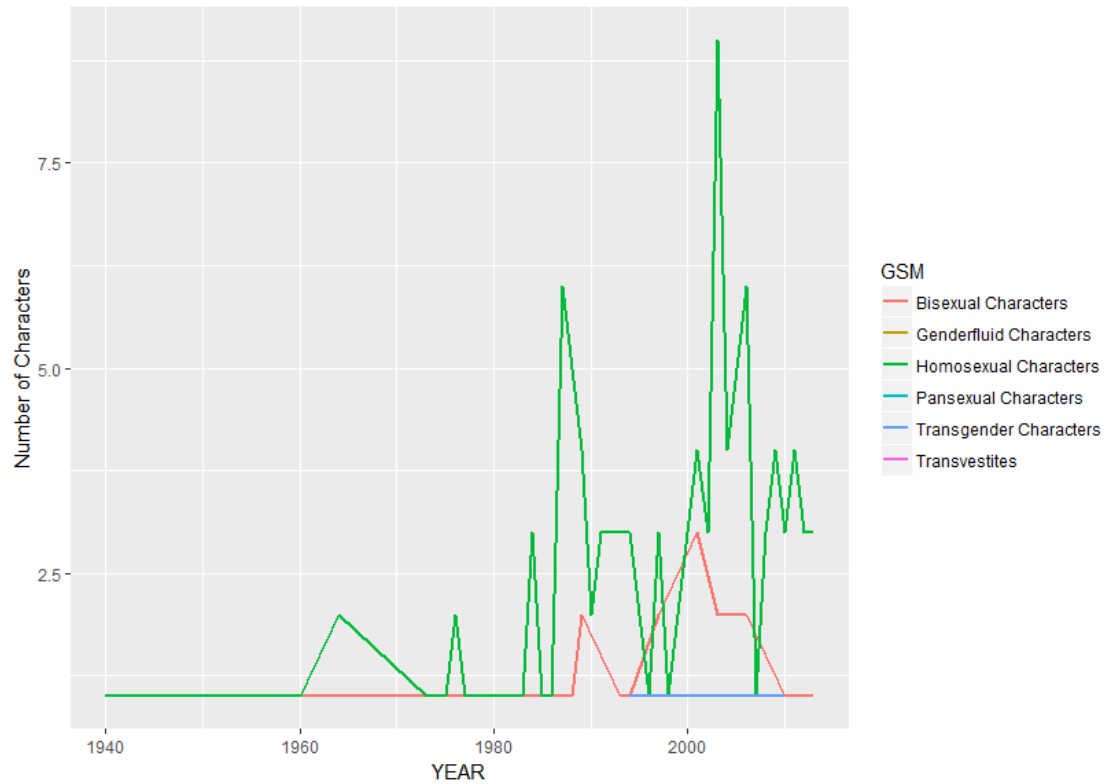
The plot below gives the number of characters, belonging to different sexes that were introduced each year since 1935.



Clearly, and as expected, there is ratio of males to females is heavily skewed with the males highly outnumbering the females. It should be noted here that only the “Male Characters”, “Female Characters” and “Transgender Characters” should be analyzed here. Other genders are fictional .

Sexual Preference of Characters

Below is a plot that describes the sexual preferences of the characters introduced each year since 1935.



It was expected that the creators of the comic series would become increasingly more open towards introducing non-heterosexual characters with time. Again, it should be noted that only “Bisexual Characters”, “Homosexual Characters” and “Transgender Characters” should be analyzed here. Other sexual preferences are fictional.

Representation of the most prominent characters from both the Universes.

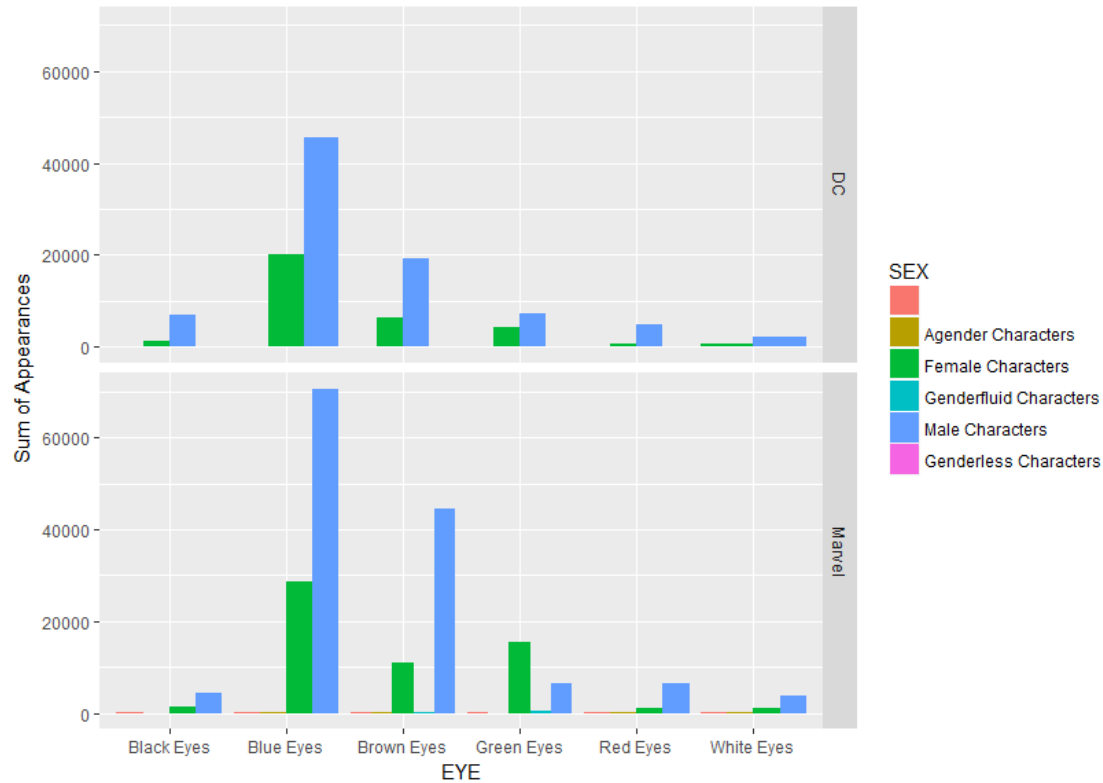
Following is a Word Cloud of the most important characters in both the universes. This representation suits better than a histogram since it can incorporate far many characters in a much smaller space.



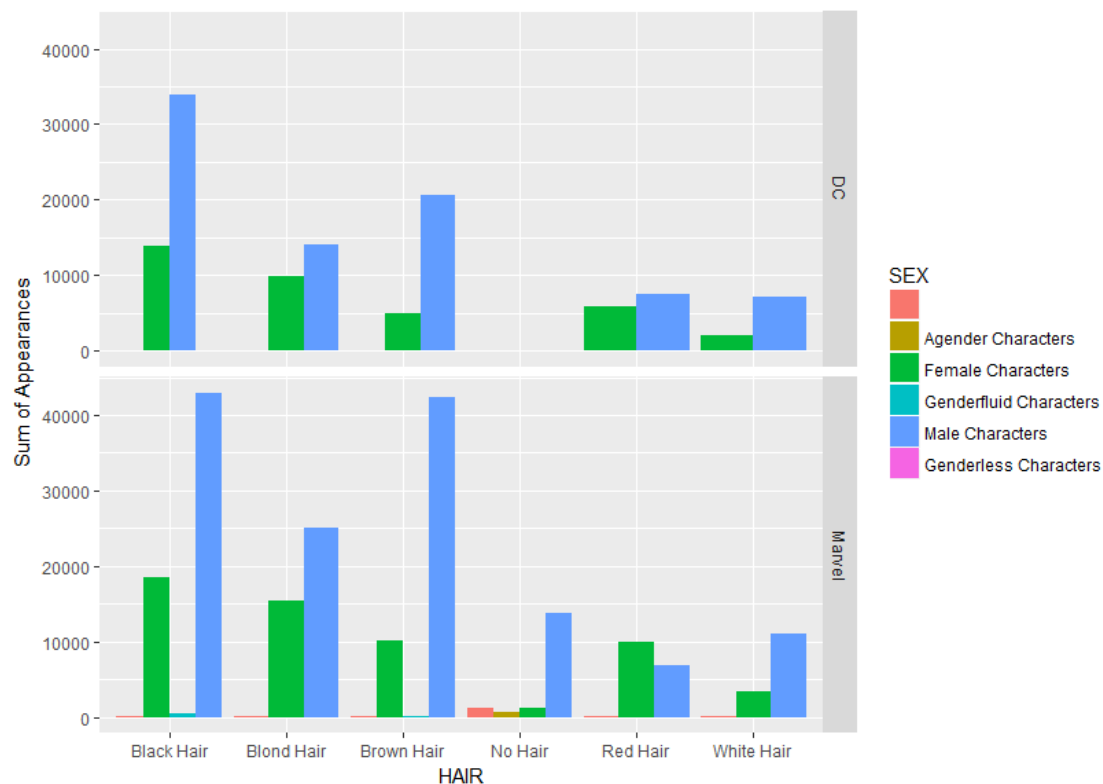
As evident from the Word Cloud, the DC Universe rides heavily on the shoulders of Batman and Superman characters followed by Green Lantern and Wonder Woman. Whereas, in case of Marvel it's more evenly distributed among 4 to 5 characters.

Prominent Facial Features

The following plots highlight the most desirable eye and hair color in any comic book character. For this plot only the five most prominent EYE and HAIR have been selected.



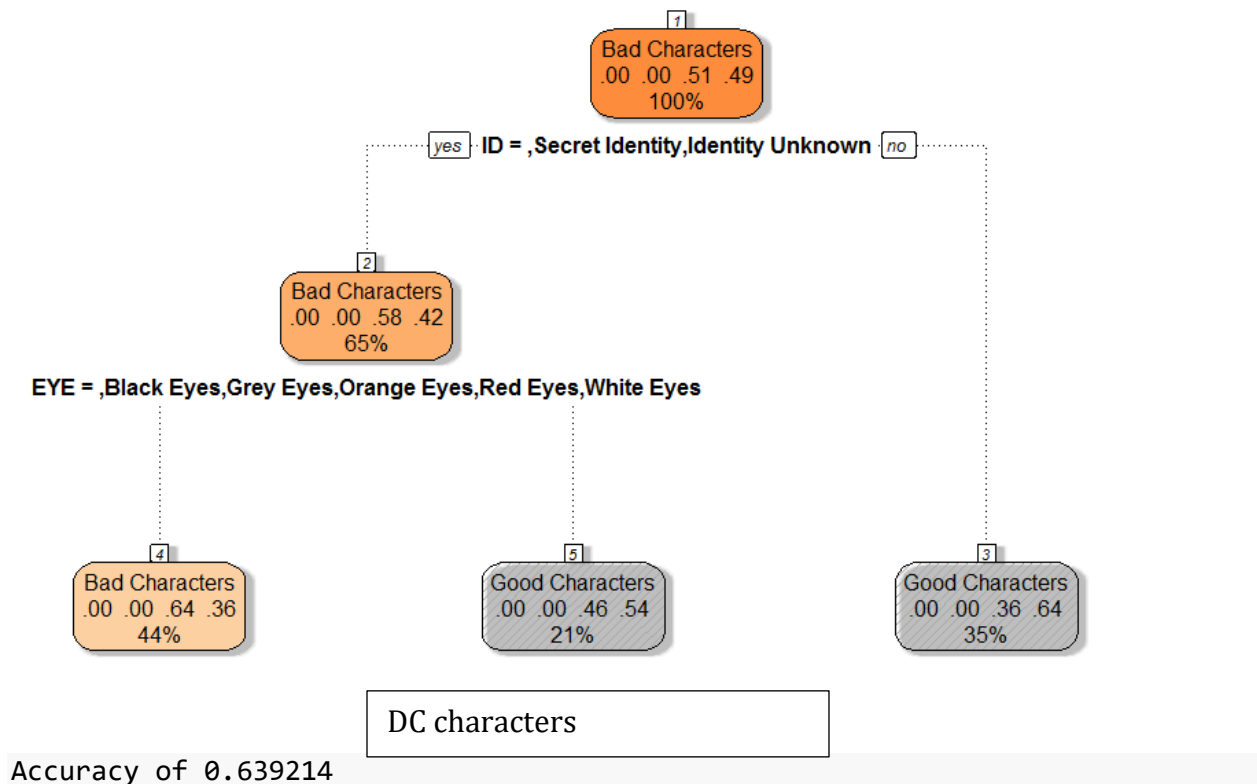
For males, in decreasing order of appeal, the eye colors are - Blue, Brown, Black, Red and White. For Females they are - Blue, Green, Brown, Black, White and Red. This distribution is same for both the production houses.



For males, in decreasing order of appeal, the hair colors are - Black, Brown, Blond, White, Red and No Hair. For Females they are -Black, Blond, Red, Brown and White with a very few without hair.

Predicting character alignment based on facial features and identity.

Following is an analysis of how prominent are the facial features (eye and hair color), the character's gender and the type of identity (secret or public) in determining whether a character is good or bad. This analysis can help to identify how stereotyped are characters' alignment. For this analysis the dataset was segregated into the two different production houses - Marvel and DC so that a comparative analysis could be done. A decision tree analysis was done on R after which the decision tree was graphically plotted.



Interpreting this tree diagram:

1. The title of each node tells us the type of character that has the highest probability of being found in the set given that the conditions stated just above the node are true. For example - in the case of the first node there is no condition stated prior to the first node. The first node is titled "Bad Characters" indicating that on the whole bad characters outnumber the good characters.
2. The numbers on the line next to the title of each node give the relative probability of all the different type of characters in that subset, with the highest number corresponding to the type of character mentioned as the title of the node. For example in the first node the "Bad Characters" have a probability of 0.50 which is equal to that of the "Good Characters" - also 0.50. In the case of the first node the probability of a character being Bad exceeds that of being Good only by a small fraction - which is lost while rounding off.
3. The number expressed as a percent on the line next to the set of numbers representing the probabilities gives the percent of the data-set that is explained by that node. For example the first node has this number as 100% indicating that it represents all the characters of the data-set.
4. The statement written next to each node indicates the condition on which the next split is made. It has two outcomes - "Yes" or a "No" - Based on which the next node is selected for analysis. For example - for the first node the statement is "ID = Secret

Identity, Identity Unknown". If this condition is TRUE i.e "Yes" one goes to the node on the left - "Bad Characters", if not, i.e. "No" one goes to the right node - "Good Characters".

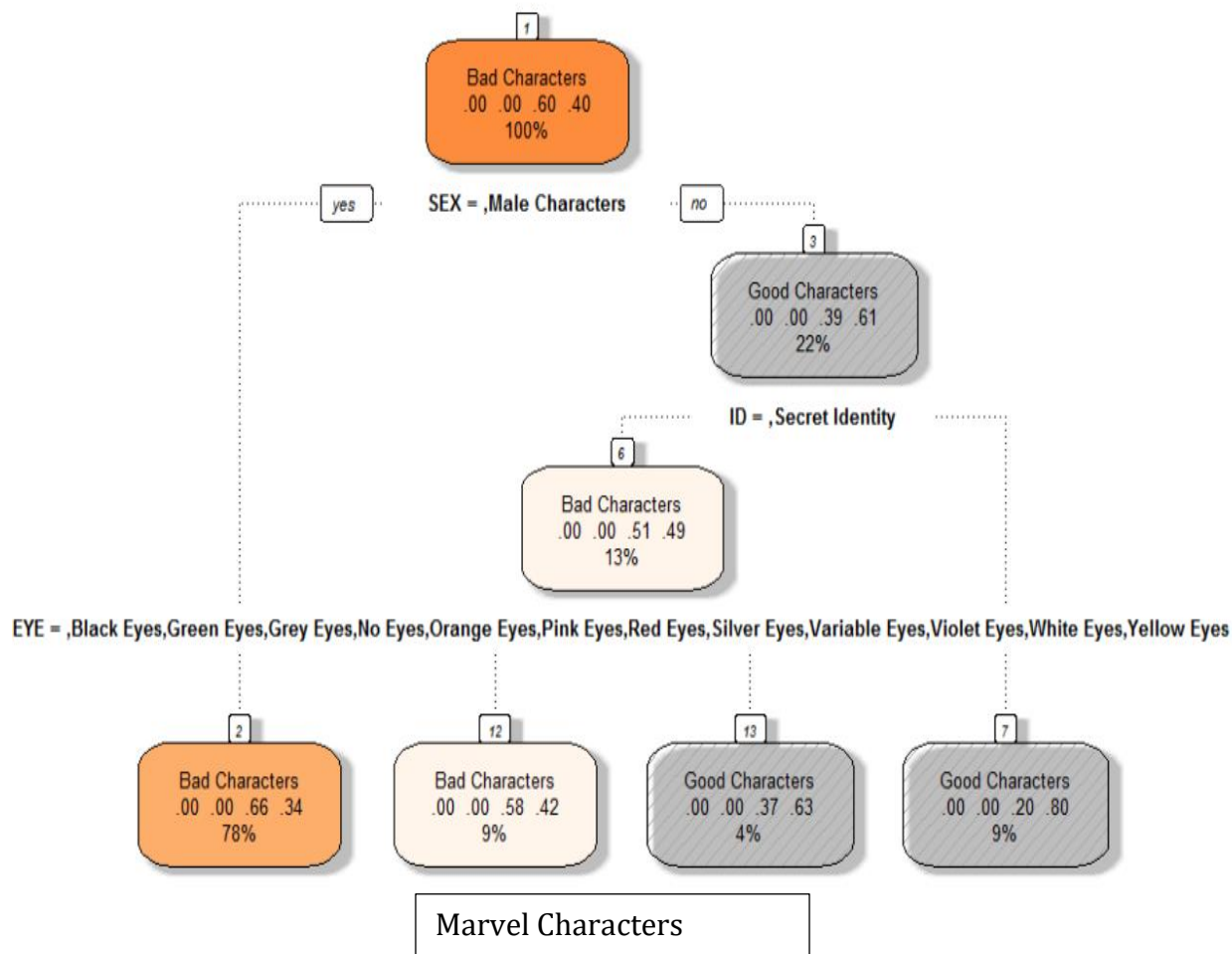
Conclusions from this tree diagram:

1. Node.1. The "Bad Characters" slightly outnumber the "Good Characters" - both having a probability of 50-50.
2. Node.2. If only those characters having a "Secret Identity" or an identity which is unknown are selected - a total of 65% of the characters come under this group out of which there is a high probability 0.59 of the characters being "Bad Characters".
3. Node.3. Characters whose identities are not secret or unknown comprise of 35% of the total and in them there is a high probability that any specific character is good (0.65).
4. Node.4. If only those characters are selected which are having either Black, Grey, Hazel, Orange, Red, White such characters comprise a total of 44% of the total characters and out of them a strong possibility is that such characters are bad(0.64) as compared to good (0.36).
5. Node.5. Characters which do not have the above mentioned eye colors, i.e. those having Blue, Brown or Green eyes constitute a total of 20% of all DC characters and there is a 53% probability that the character is "Good".

Takeaways from the analysis of DC Characters:

1. As expected, the Bad guys generally have a secret or an unknown identity whereas the good guys mostly don't.
2. Except for the "Black" eye color most of the other eye colors are very rare if not impossible to find naturally. This, probably, has been done to give an element of exclusivity or other-worldly appearance to the Bad Characters.

A fact to be noted is that the accuracy of this model is around 60% when only Hair color, Eye color and Sex is taken into account and around 63% when "Identity" is also added to the model.



Accuracy of 0.6077984

Takeaways from the analysis of Marvel Characters:

1. The Bad guys greatly outnumber the Good Characters.
2. Unlike DC, the gender of the character is one of the most important factors when deciding whether a character is good or bad with the majority of the Male characters being bad.
3. Similar to DC Universe, the “Bad” characters have facial characteristics that are very rare if not impossible to find in nature – perhaps to give them some other-worldly appearance.

Conclusions:

1. The comic books are still heavily skewed in favor of the male characters, and this effect is even more pronounced in case of Marvel where the ratio of male to characters with other genders is almost 80-20.

2. The number of characters having non-heterosexual preferences, though very small, is on the rise.
 3. The number of “Bad” characters always exceed that of the “Good” characters – perhaps to show that the hero can tackle many villains by himself or herself.
 4. Marvel is much more into gender stereotyping with the majority of male characters being “Bad”.
 5. The “Bad” characters are generally portrayed having facial features that are not that common, perhaps to give them some other worldly appearance.
-