Marvel Comics by the Numbers

by: Bruno Ustariz

Summary

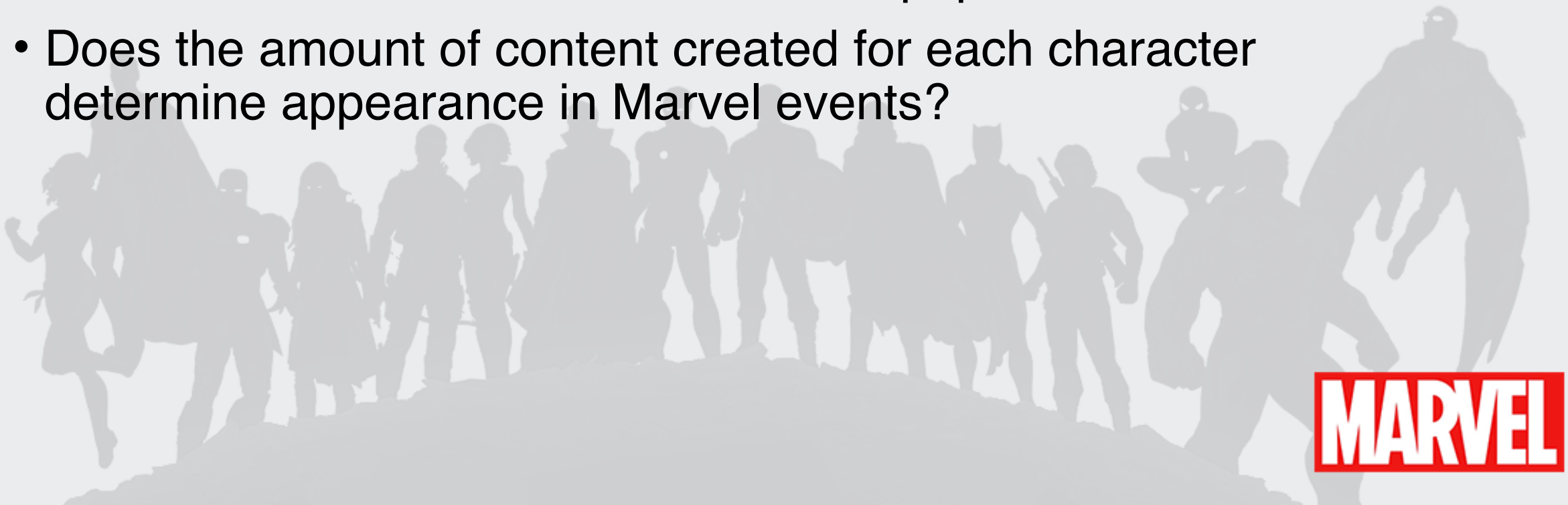
Goals, Questions, Findings.

Summary: Goals

- This project is to uncover any trends in the content created by marvel comics with respect to all its characters. The goals are:
  - To analyze and draw conclusions about the popularity of characters based on the amount of material created for each.
  - To analyze the amount of content created for the most popular characters vs characters in general.
  - To determine if the amount of comic books created for each character determines their appearance in events(universe changing story lines)

Summary: Questions

- How much content does Marvel create for all of its characters?
- Which Marvel characters are the most popular?
- Does the amount of content created for each character determine appearance in Marvel events?



Summary: Findings

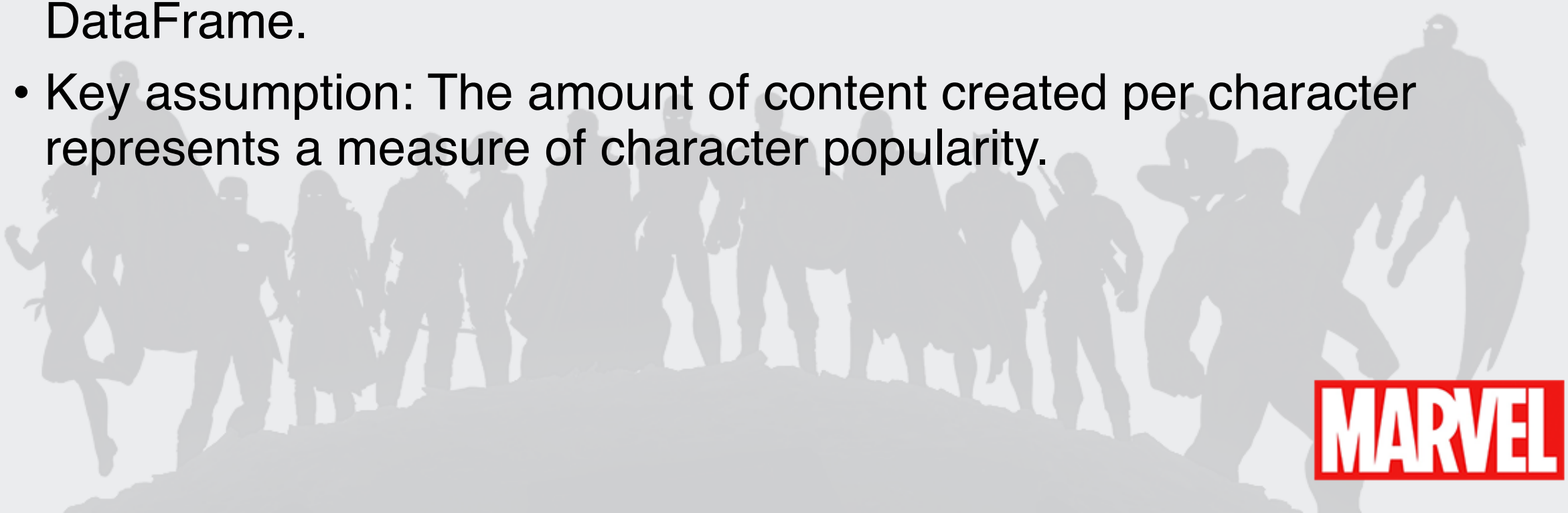
- How much content does Marvel create for all of its characters?
  - The data shows that the majority of marvels characters have very low amounts of content. In fact, the most popular characters are outliers.
- Which Marvel characters are the most popular?
  - For those who are Marvel fans and have either read comics, or watched the cartoons or movies, the most popular characters are expected.
- Does the amount of content created for each character determine appearance in Marvel events and series?
  - The data shows that there is a strong positive correlation between the amount of content created for each character and the amount of appearances in marvels events and series.

Data Set

Marvels API

Data Set

- The data required was the amount of content created for each marvel character in regards to comic books, events and series.
- Data was pulled from Marvel API and stored into a Pandas DataFrame.
- Key assumption: The amount of content created per character represents a measure of character popularity.



DataFrame Statistics

- Data shows that 75% of the data has less than 40 comic books made compared to a max of 3,918.
- In addition, 75% of the data shows up to 2 event appearances.

to hide

	character_id	num_of_comics	num_of_series	num_of_stories	num_of_events
count	1.493000e+03	1493.000000	1493.000000	1493.000000	1493.000000
mean	1.010906e+06	59.006698	19.029471	75.734092	1.817816
std	1.911881e+03	220.402596	59.423395	318.568612	4.286681
min	1.009144e+06	0.000000	0.000000	0.000000	0.000000
25%	1.009548e+06	1.000000	1.000000	1.000000	0.000000
50%	1.010812e+06	10.000000	4.000000	11.000000	0.000000
75%	1.011188e+06	38.000000	15.000000	43.000000	2.000000
max	1.017840e+06	3918.000000	989.000000	5898.000000	42.000000

DataFrame Filtering

- DataFrame was filtered to show only characters with at least 1 comic book and one event.
- Characters with 0 comic books were determined to be supporting characters in comics of other characters.
- The result was a reduction of characters from 1493 to 660.

	character_id	character_name	num_of_comics	num_of_series	num_of_stories	num_of_events
1185	1009610	Spider-Man	3918	989	5898	38
1472	1009726	X-Men	3432	828	5092	42
573	1009368	Iron Man	2561	626	3880	31
1453	1009718	Wolverine	2470	663	3457	42
185	1009220	Captain America	2362	651	3503	31
...	...	...	...	...	...	...
706	1011145	Lord Hawal	1	1	1	1
1372	1010987	Unus (Age of Apocalypse)	1	1	1	1
1068	1010970	Sabretooth (House of M)	1	1	1	1
1073	1011156	Salo	1	1	1	1
67	1010827	Armory	1	1	2	1

660 rows × 6 columns

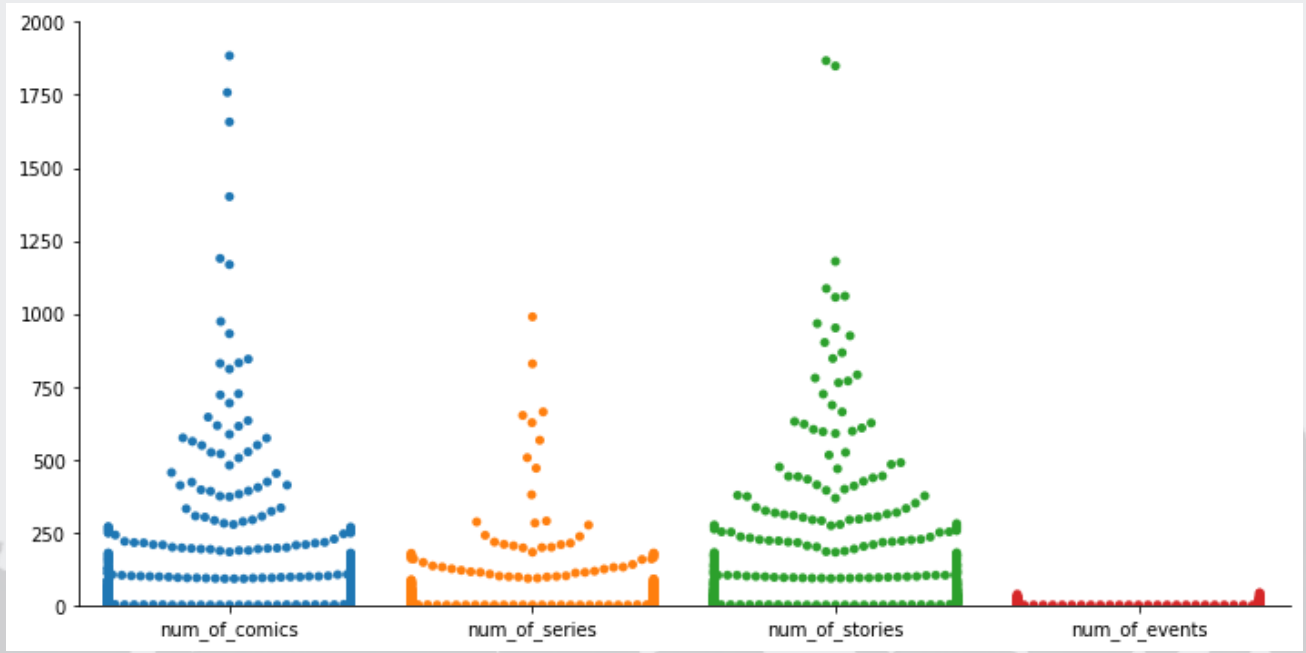
Content per Character?

How much content has marvel created for each character?

Content per character

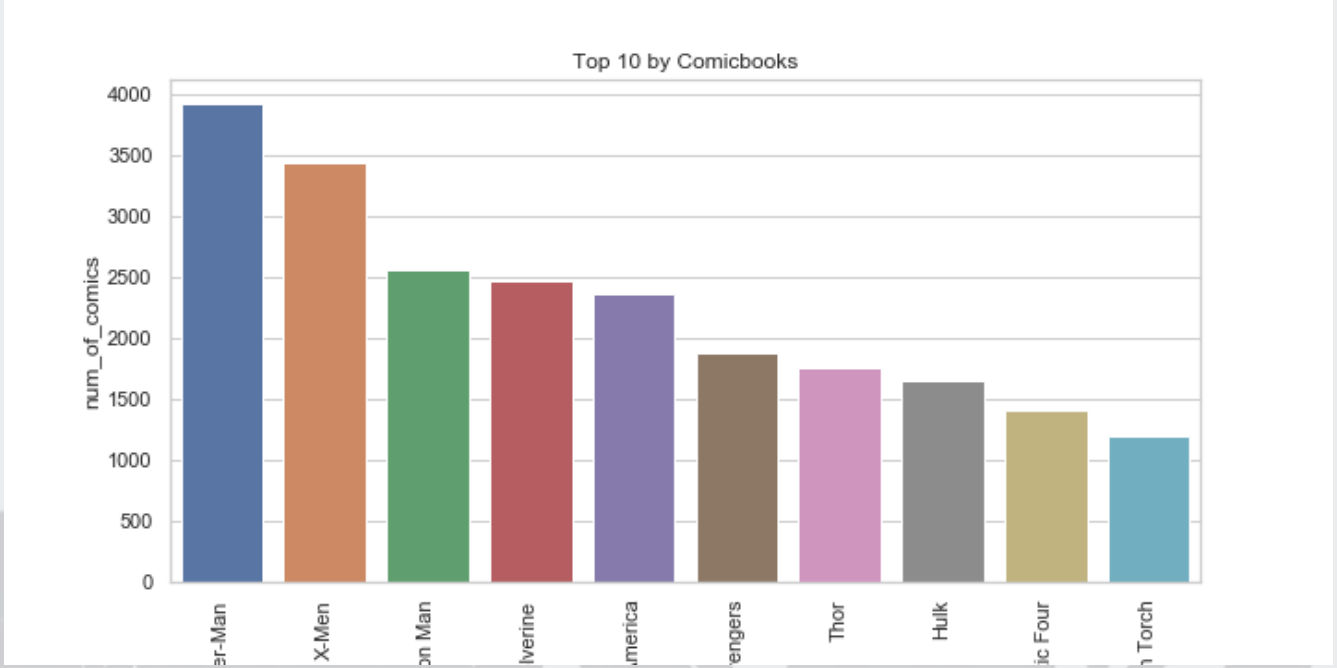
- The majority of characters have a very low amount of content.
- Note: this representation is based on our reduced DataFrame.

Data Distribution



Top 10 Characters by Comic books

- The top 10 characters have over 1,000 comic books.
- In comparison to our data, the most popular characters have significantly more content vs the rest of the characters.
-

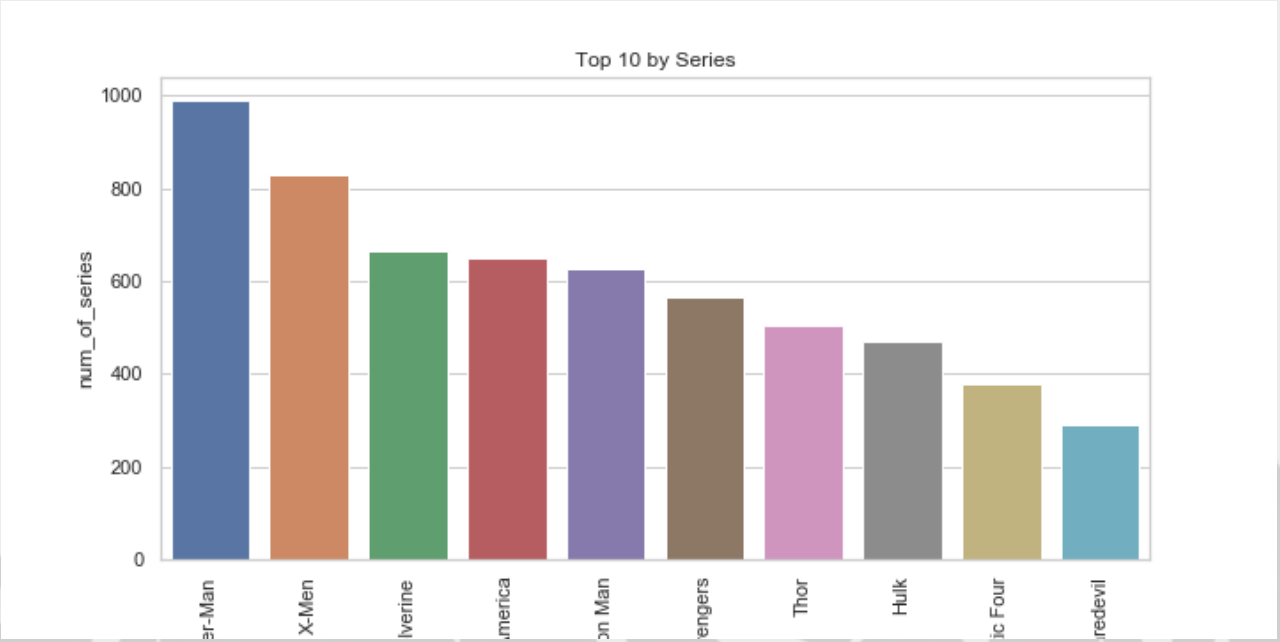


Characters by Series

Top characters by amount of series and correlation vs comic books.

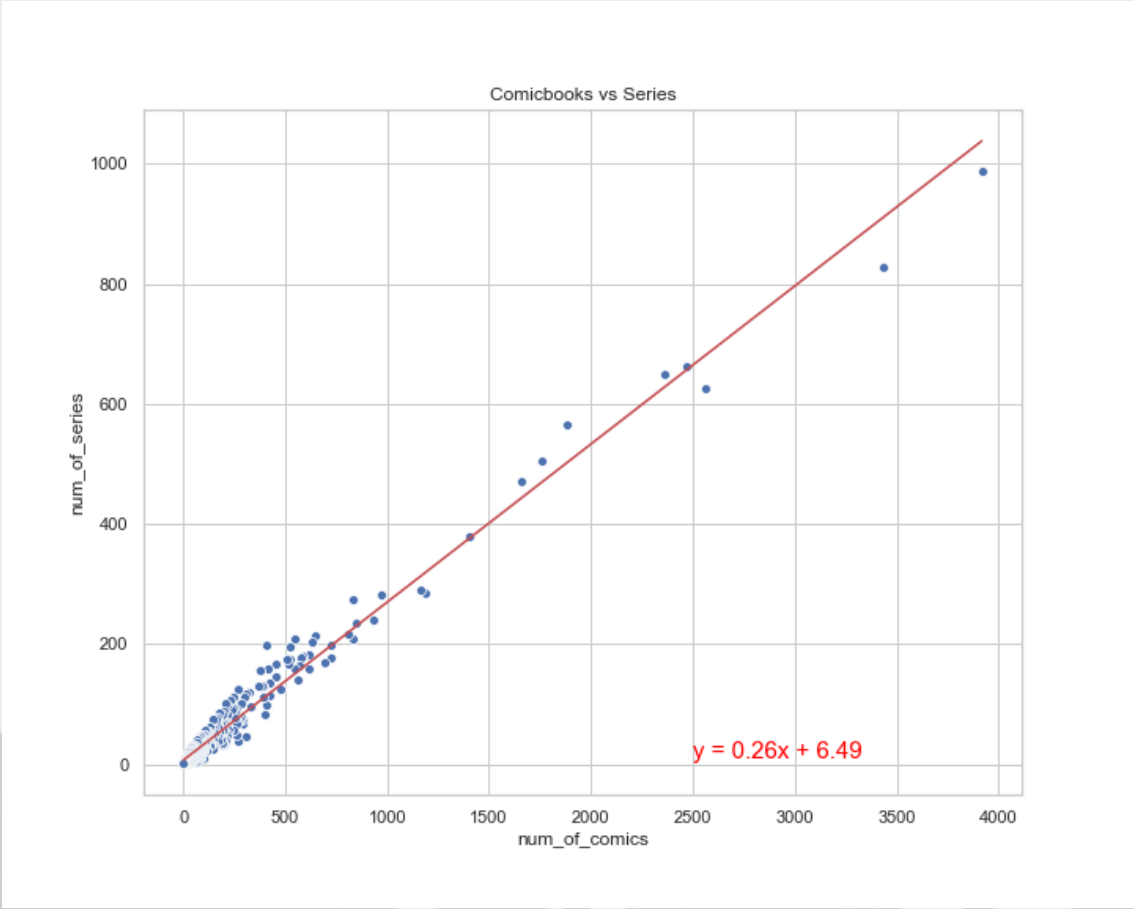
Top 10 Characters by Series

- Top 10 are very similar to our previous bar chart of top 10 characters by comic book count.
- An assumption can be made that the amount of comics made for each character is positively correlated to the amount of series they are in.



Character Comic books vs Series

- The data shows that our previous assumption is correct. In fact, there is a nearly perfect correlation between both data points.
- The r-squared is: 0.9887123917887364

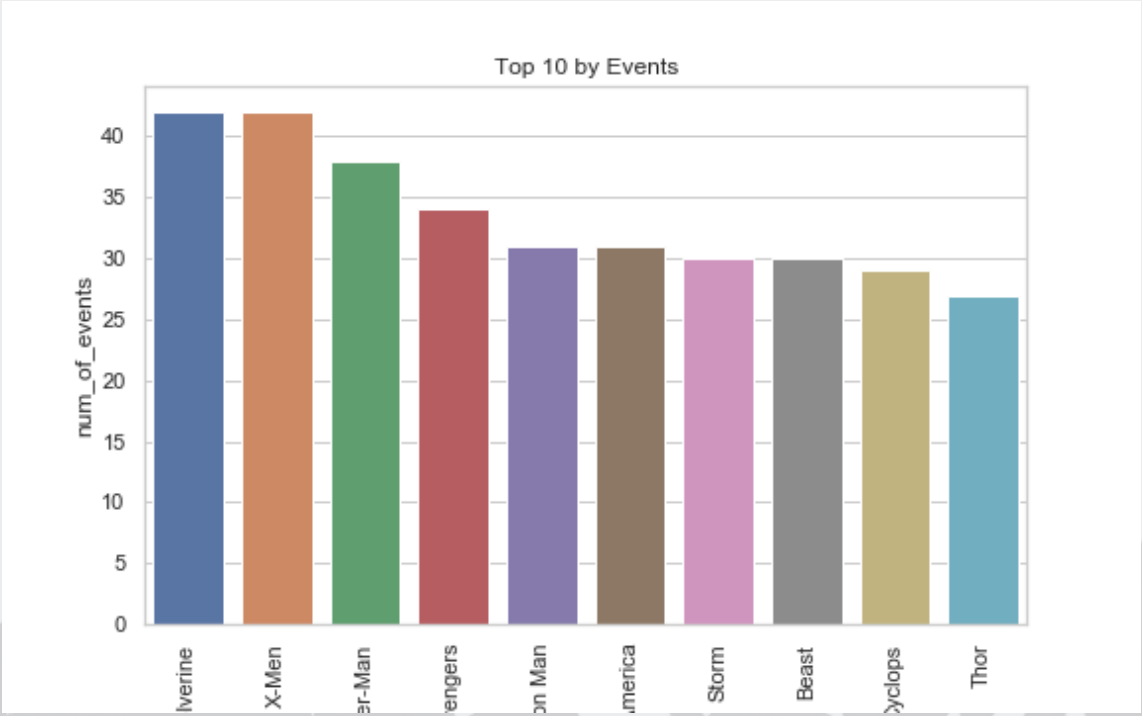


Characters by Events

Top characters by amount of events and correlation vs comic books.

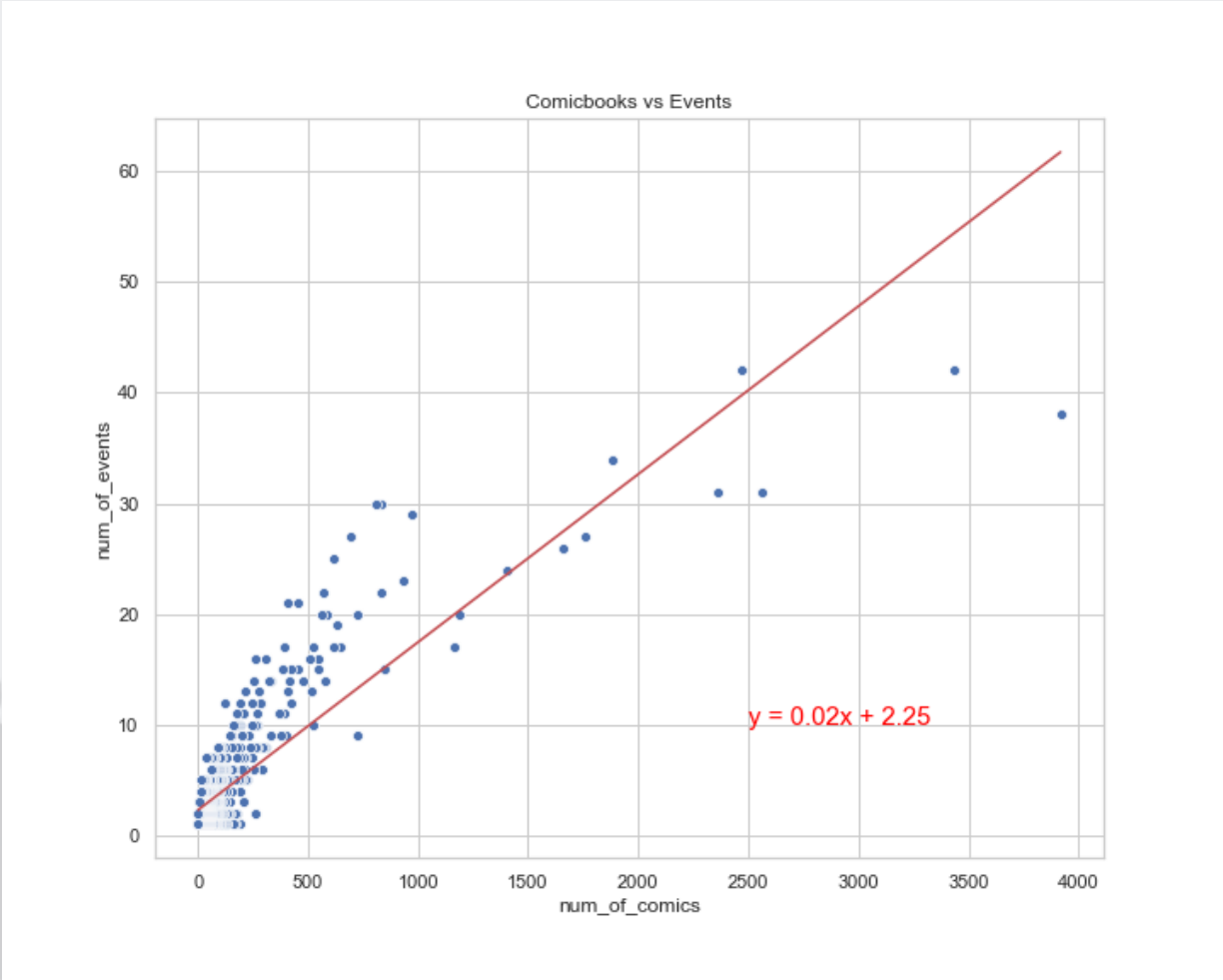
Top 10 characters by Events

- Top 10 are also very similar to our previous bar chart of top 10 characters by comic book count, although less so than series top 10.
- An assumption can be made that the amount of comics made for each character is positively correlated to the amount of events they are in.



Characters Comic books vs Events

- The data shows that In there is a positive correlation between both data points. Although less so than series vs comic books.
- The r-squared is: 0.8568303798251373

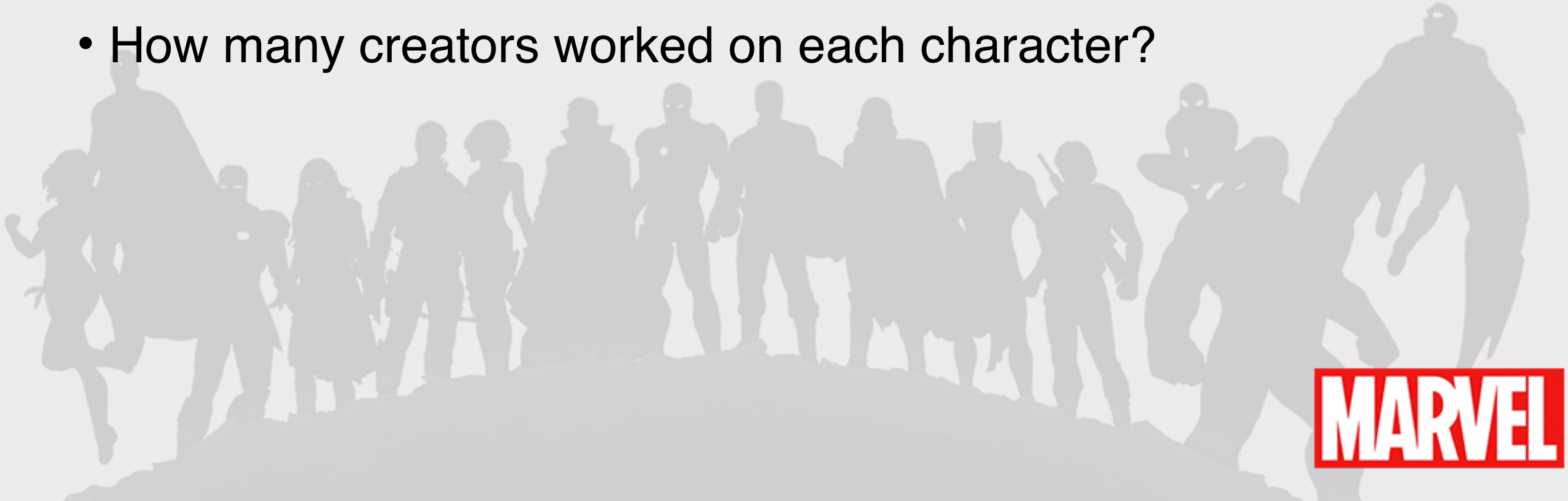


Conclusion

- Discoveries:
  - There are very few popular characters who are popular. In fact, they are outliers of the data.
  - The proportion of investment in content is very high for few characters.
- Data Shortcomings:
  - Quantity sold per comic book was not available.
  - Some characters are part of a group (x-men, avengers, etc) and could not be separated.
  - Time component of comic books was not available.

Conclusion contd.

- Additional questions:
  - Popularity of characters over time(by year/month)?
  - Total comics marvel creates per year?
  - How many creators worked on each character?



Thanks!