

Group Name: SuperWranglers

Group Members: Madison Lusane, Jordan Moses, Marquesia Atwater

## Data set overview and preparation:

### Data set overview:

Filename: dc-wikia-data      Extension: csv

Filename: marvel-wikia-data      Extension: csv

### Data set summary:

Our team will be using the datasets acquired from [data.world](#). The characters were scrapped on August 24th, the appearance count was scrapped on September 3rd, and the month and year of the first issue each character appeared in was pulled on October 6. The data will be stored in non-relational databases. The variables range from their hair color, eye color, the year they first arrived as characters, and etc.

	page_id	name	urlslug	ID	ALIGN	EYE	HAIR	SEX	GSM	ALIVE	APPEARANCES	FIRST APPEARANCE	YEAR
0	1422	Batman (Bruce Wayne)	Vwiki/Batman_(Bruce_Wayne)	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters	3093.0	1939, May	1939.0
1	23387	Superman (Clark Kent)	Vwiki/Superman_(Clark_Kent)	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters	2496.0	1986, October	1986.0
2	1458	Green Lantern (Hal Jordan)	Vwiki/Green_Lantern_(Hal_Jordan)	Secret Identity	Good Characters	Brown Eyes	Brown Hair	Male Characters	NaN	Living Characters	1565.0	1959, October	1959.0
3	1659	James Gordon (New Earth)	Vwiki/James_Gordon_(New_Earth)	Public Identity	Good Characters	Brown Eyes	White Hair	Male Characters	NaN	Living Characters	1316.0	1987, February	1987.0
4	1576	Richard Grayson (New Earth)	Vwiki/Richard_Grayson_(New_Earth)	Secret Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters	1237.0	1940, April	1940.0

	page_id	name	urlslug	ID	ALIGN	EYE	HAIR	SEX	GSM	ALIVE	APPEARANCES	FIRST APPEARANCE	Year
0	1678	Spider-Man (Peter Parker)	VSpider-Man_(Peter_Parker)	Secret Identity	Good Characters	Hazel Eyes	Brown Hair	Male Characters	NaN	Living Characters	4043.0	Aug-62	1962.0
1	7139	Captain America (Steven Rogers)	VCaptain_America_(Steven_Rogers)	Public Identity	Good Characters	Blue Eyes	White Hair	Male Characters	NaN	Living Characters	3360.0	Mar-41	1941.0
2	64786	Wolverine (James "Logan" Howlett)	VWolverine_(James_%22Logan%22_Howlett)	Public Identity	Neutral Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters	3061.0	Oct-74	1974.0
3	1868	Iron Man (Anthony "Tony" Stark)	VIron_Man_(Anthony_%22Tony%22_Stark)	Public Identity	Good Characters	Blue Eyes	Black Hair	Male Characters	NaN	Living Characters	2961.0	Mar-63	1963.0

Table of variable definitions:

Variable	Identification
page_id	<ul style="list-style-type: none"> <li>The unique identifier for that characters page within the wikia</li> </ul>
Name	<ul style="list-style-type: none"> <li>The name of the character</li> </ul>
Urlslug	<ul style="list-style-type: none"> <li>The unique url within the wikia that takes you to the character</li> </ul>
Align	<ul style="list-style-type: none"> <li>If the character is Good, Bad or Neutral</li> </ul>
Eye	<ul style="list-style-type: none"> <li>Eye color of the character</li> </ul>
Hair	<ul style="list-style-type: none"> <li>Hair color of the character</li> </ul>
Sex	<ul style="list-style-type: none"> <li>Sex of the character (e.g. Male, Female, etc.)</li> </ul>
GSM	<ul style="list-style-type: none"> <li>If the character is a gender or sexual minority (e.g. Homosexual characters, bisexual characters)</li> </ul>
Alive	<ul style="list-style-type: none"> <li>If the character is alive or deceased</li> </ul>
Appearances	<ul style="list-style-type: none"> <li>The number of appearance(s) of the character in comic books (as of Sep. 2, 2014. Number will become increasingly out of date as time goes on.)</li> </ul>
First Appearances	<ul style="list-style-type: none"> <li>The month and year of the character's first appearance in a comic book, if available</li> </ul>

Year	<ul style="list-style-type: none"> <li>The year of the character's first appearance in a comic book, if available</li> </ul>
ID	<ul style="list-style-type: none"> <li>The identity status of the character (Secret Identity, Public identity, [on marvel only: No Dual Identity])</li> </ul>

### Data Preparation methods:

The DC and Marvel datasets contain the year of appearances which we will use to extrapolate several findings. The data was downloaded via CSV and uploaded into a pandas data frame. By reviewing, the data frames, we found that we only needed the Name, Id, Alignment, Sex, Alive, Appearances, and Year columns in both CSV files. Thus, we dropped the other columns to be able to begin our analysis. In addition, we had to append both tables together using Pandas. Lastly, we uploaded them into our local databases via sqlalchemy.

dc_db		marvel_db	
Name	varchar(200)	Name	varchar(200)
Id	int	Id	int
Alignment	varchar(200)	Alignment	varchar(200)
Sex	varchar(200)	Sex	varchar(200)
Alive	varchar(200)	Alive	varchar(200)
Appearances	varchar(200)	Appearances	varchar(200)
Year	date	Year	date

### Findings:

- We wanted to see if there was a rising trend in male characters within the last 10 years of our data set.
  - According to our transformed data set there wasn't a rising trend. The sex was the same within the 10 years. With the data being consistent, in the next 10 years we predict that the male characters will continue to dominate the comic field.

2. We also wanted to review if there is a remarkable distinction between good/bad characters from DC and Marvel comic books.

- Our findings show that there is not a remarkable distinction. Most of the DC and Marvel characters were categorized as “good characters”. Both data sets were equal in the alignment of the characters.