Final\_Project\_Part1

Zhuofei Xie

10/13/2019

The article I choose for final project part 1 is: <http://fivethirtyeight.com/features/women-in-comic-books/>

The first graph is “New Comic Book Characters Introduced Per Year”. In order to creat a graph with both marvel data and DC data, I use full\_joint function to join to data sets together. I realzie the the year column for marvel and DC data sets were different. I change one of them in order to let the names are corresponding. I also only keep the distinct name appear each year. Then the graph has been generated indicates both companies created more and more characters from 1940 to 2010, except 1940 to 1950. Both companies created less characters during that time.

head(marvel\_data)

## page\_id name  
## 1 1678 Spider-Man (Peter Parker)  
## 2 7139 Captain America (Steven Rogers)  
## 3 64786 Wolverine (James \\"Logan\\" Howlett)  
## 4 1868 Iron Man (Anthony \\"Tony\\" Stark)  
## 5 2460 Thor (Thor Odinson)  
## 6 2458 Benjamin Grimm (Earth-616)  
## urlslug ID  
## 1 \\/Spider-Man\_(Peter\_Parker) Secret Identity  
## 2 \\/Captain\_America\_(Steven\_Rogers) Public Identity  
## 3 \\/Wolverine\_(James\_%22Logan%22\_Howlett) Public Identity  
## 4 \\/Iron\_Man\_(Anthony\_%22Tony%22\_Stark) Public Identity  
## 5 \\/Thor\_(Thor\_Odinson) No Dual Identity  
## 6 \\/Benjamin\_Grimm\_(Earth-616) Public Identity  
## ALIGN EYE HAIR SEX GSM  
## 1 Good Characters Hazel Eyes Brown Hair Male Characters   
## 2 Good Characters Blue Eyes White Hair Male Characters   
## 3 Neutral Characters Blue Eyes Black Hair Male Characters   
## 4 Good Characters Blue Eyes Black Hair Male Characters   
## 5 Good Characters Blue Eyes Blond Hair Male Characters   
## 6 Good Characters Blue Eyes No Hair Male Characters   
## ALIVE APPEARANCES FIRST.APPEARANCE Year  
## 1 Living Characters 4043 Aug-62 1962  
## 2 Living Characters 3360 Mar-41 1941  
## 3 Living Characters 3061 Oct-74 1974  
## 4 Living Characters 2961 Mar-63 1963  
## 5 Living Characters 2258 Nov-50 1950  
## 6 Living Characters 2255 Nov-61 1961

names(marvel\_data)[13] <- "YEAR"  
add\_company\_dc <- dc\_data %>% mutate(company = "DC")  
add\_company\_marvel <- marvel\_data %>% mutate(company = "Marvel")  
join\_dc\_marvel <- full\_join(add\_company\_dc, add\_company\_marvel)

## Joining, by = c("page\_id", "name", "urlslug", "ID", "ALIGN", "EYE", "HAIR", "SEX", "GSM", "ALIVE", "APPEARANCES", "FIRST.APPEARANCE", "YEAR", "company")

head(join\_dc\_marvel)

## page\_id name  
## 1 1422 Batman (Bruce Wayne)  
## 2 23387 Superman (Clark Kent)  
## 3 1458 Green Lantern (Hal Jordan)  
## 4 1659 James Gordon (New Earth)  
## 5 1576 Richard Grayson (New Earth)  
## 6 1448 Wonder Woman (Diana Prince)  
## urlslug ID ALIGN  
## 1 \\/wiki\\/Batman\_(Bruce\_Wayne) Secret Identity Good Characters  
## 2 \\/wiki\\/Superman\_(Clark\_Kent) Secret Identity Good Characters  
## 3 \\/wiki\\/Green\_Lantern\_(Hal\_Jordan) Secret Identity Good Characters  
## 4 \\/wiki\\/James\_Gordon\_(New\_Earth) Public Identity Good Characters  
## 5 \\/wiki\\/Richard\_Grayson\_(New\_Earth) Secret Identity Good Characters  
## 6 \\/wiki\\/Wonder\_Woman\_(Diana\_Prince) Public Identity Good Characters  
## EYE HAIR SEX GSM ALIVE  
## 1 Blue Eyes Black Hair Male Characters Living Characters  
## 2 Blue Eyes Black Hair Male Characters Living Characters  
## 3 Brown Eyes Brown Hair Male Characters Living Characters  
## 4 Brown Eyes White Hair Male Characters Living Characters  
## 5 Blue Eyes Black Hair Male Characters Living Characters  
## 6 Blue Eyes Black Hair Female Characters Living Characters  
## APPEARANCES FIRST.APPEARANCE YEAR company  
## 1 3093 1939, May 1939 DC  
## 2 2496 1986, October 1986 DC  
## 3 1565 1959, October 1959 DC  
## 4 1316 1987, February 1987 DC  
## 5 1237 1940, April 1940 DC  
## 6 1231 1941, December 1941 DC

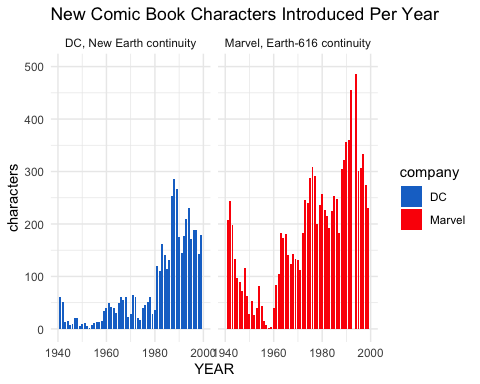
after\_selection <- join\_dc\_marvel %>%   
 group\_by(YEAR, company) %>%   
 summarise(  
 characters = n\_distinct(name),   
 )   
head(after\_selection)

## # A tibble: 6 x 3  
## # Groups: YEAR [5]  
## YEAR company characters  
## <int> <chr> <int>  
## 1 1935 DC 1  
## 2 1936 DC 9  
## 3 1937 DC 4  
## 4 1938 DC 10  
## 5 1939 DC 18  
## 6 1939 Marvel 69

company.labs<- c("DC, New Earth continuity", "Marvel, Earth-616 continuity")  
names(company.labs) <- c("DC", "Marvel")  
  
ggplot(data = after\_selection, mapping = aes(x = YEAR, fill = company)) + geom\_bar(aes(y = characters), stat = "identity", width = 0.8) +ggtitle("New Comic Book Characters Introduced Per Year") + theme\_minimal() + facet\_grid(~company, labeller = labeller(company = company.labs)) + scale\_fill\_manual( values = c( "dodgerblue3", "#FC0505" ) ) + ylim(0,500) + xlim(1940,2000)

## Warning: Removed 35 rows containing missing values (position\_stack).

## Warning: Removed 4 rows containing missing values (geom\_bar).

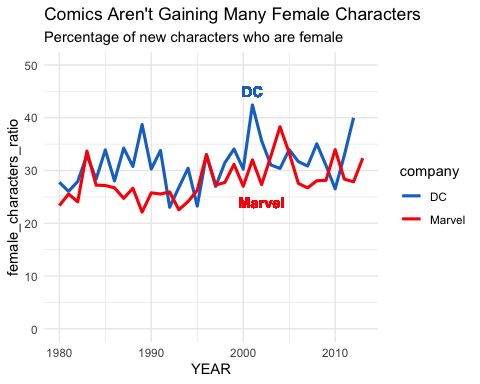


The next graph I plot is about the “Comic Arn’t Gaining Many Female Characters”. In order to make this graph, I only select the time after 1980. Then using the group\_by and summarise functions to get the percentage of the female characters. As can be seen from the graph. The percentage of female characters were less than 50% from 1980 to 2010. The highest ratio was around 45% at 2002. Therefore, this result indicates the female characters were less than male characters.

new\_female\_characters <- join\_dc\_marvel %>%   
 filter(YEAR >= 1980) %>%   
 group\_by(YEAR, company) %>%   
 summarise(  
 number\_characters = n\_distinct(name),  
 number\_female\_characters = sum(ifelse(SEX == "Female Characters",1, 0)),   
 female\_characters\_ratio = (number\_female\_characters/number\_characters) \* 100,   
 ) %>%   
 filter(female\_characters\_ratio>0)  
head(new\_female\_characters)

## # A tibble: 6 x 5  
## # Groups: YEAR [3]  
## YEAR company number\_characters number\_female\_charac… female\_characters\_…  
## <int> <chr> <int> <dbl> <dbl>  
## 1 1980 DC 36 10 27.8  
## 2 1980 Marvel 257 60 23.3  
## 3 1981 DC 119 31 26.1  
## 4 1981 Marvel 227 58 25.6  
## 5 1982 DC 111 31 27.9  
## 6 1982 Marvel 216 52 24.1

ggplot(data = new\_female\_characters, mapping = aes(x = YEAR, y = female\_characters\_ratio, color = company)) + geom\_line(lwd = 1.1) + ylim(0,50) + ggtitle("Comics Aren't Gaining Many Female Characters", subtitle = "Percentage of new characters who are female") + theme\_minimal() + scale\_color\_manual(values=c("dodgerblue3", "#FC0505")) + geom\_text( aes(fontface = 2), x = 2001, y = 45, label = "DC", color = "dodgerblue3") + geom\_text( aes(fontface = 2), x = 2002, y = 24, label = "Marvel", color = "#FC0505")



The last graph I created is “Comics Are Gaining A Few LGBT Characters”. In order to generate this graph, I also use filter select, group\_by functions to generate the suitable data sets. The result graph indicates also from 1940 to 2010, the number of LGBT characters increased, however, the number of those characters were still minirities comparing with the other normal characters.

with\_LGBT <- join\_dc\_marvel %>%   
 select(YEAR, GSM) %>%   
 filter(GSM == "Bisexual Characters" | GSM == "Homosexual Characters" | GSM == "Transgender Characters" | GSM == "Pansexual Characters" | GSM == "Transvestites") %>%   
 mutate(n = 1) %>%   
 arrange(desc(YEAR))  
  
after\_count <- with\_LGBT %>%   
 group\_by(YEAR) %>%   
 summarise(  
 n = sum(n),   
 )   
head(after\_count)

## # A tibble: 6 x 2  
## YEAR n  
## <int> <dbl>  
## 1 1940 1  
## 2 1943 2  
## 3 1948 1  
## 4 1949 1  
## 5 1959 1  
## 6 1960 1

ggplot(data = after\_count, mapping = aes(x = YEAR, y = n)) + geom\_bar(stat = "identity", fill = "#6AA121") + ylim(0,15) + ggtitle("Comics Are Gaining A Few LGBT Characters" , subtitle = "LGBT characters introduced into DC and Marvel comics per year, including retroactive continuity changes") + theme\_minimal()

## Warning: Removed 1 rows containing missing values (position\_stack).

