

# Comic Charcter GSM Representation

William Lovejoy

3/18/2022

```
library(tidyverse)
library(dplyr)
library(ggthemes)
library(scales)
```

## Data Sourcing and Cleaning

The data used in this brief analysis is from:

<https://www.kaggle.com/fivethirtyeight/fivethirtyeight-comic-characters-dataset>

To start with, we'll load in the csv files, and assign a new column to each with the source company as "Universe". We'll then take a look at the structure, and change the name of the year column in the Marvel dataset from "year" to "YEAR", so it matches with the DC Dataset. Afterwards the datasets are combined into one larger one called "comics".

```
dc <- read.csv("C:\\Users\\William Lovejoy\\Documents\\Codes\\R\\DataScience\\
\\comics\\dc-wikia-data.csv", na.strings = "")
marvel <- read.csv("C:\\Users\\William Lovejoy\\Documents\\Codes\\R\\DataScie
nce\\comics\\marvel-wikia-data.csv", na.strings = "")
dc$Universe <- "DC"
marvel$Universe <- "Marvel"
glimpse(dc)
```

```
## Rows: 6,896
## Columns: 14
## $ page_id      <int> 1422, 23387, 1458, 1659, 1576, 1448, 1486, 1451,
7176~
## $ name         <chr> "Batman (Bruce Wayne)", "Superman (Clark Kent)",
"Gre~
## $ urlslug      <chr> "\\wiki\\Batman_(Bruce_Wayne)", "\\wiki\\Supe
rman~
## $ ID           <chr> "Secret Identity", "Secret Identity", "Secret Ide
ntit~
## $ ALIGN        <chr> "Good Characters", "Good Characters", "Good Chara
cter~
## $ EYE          <chr> "Blue Eyes", "Blue Eyes", "Brown Eyes", "Brown Ey
es",~
## $ HAIR         <chr> "Black Hair", "Black Hair", "Brown Hair", "White
Hair~
## $ SEX          <chr> "Male Characters", "Male Characters", "Male Chara
cter~
## $ GSM          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
```

```

A, N~
## $ ALIVE          <chr> "Living Characters", "Living Characters", "Living
Cha~
## $ APPEARANCES    <int> 3093, 2496, 1565, 1316, 1237, 1231, 1121, 1095, 1
075,~
## $ FIRST.APPEARANCE <chr> "1939, May", "1986, October", "1959, October", "1
987,~
## $ YEAR           <int> 1939, 1986, 1959, 1987, 1940, 1941, 1941, 1989, 1
969,~
## $ Universe       <chr> "DC", "DC", "DC", "DC", "DC", "DC", "DC", "DC", "
DC",~

```

```
glimpse(marvel)
```

```

## Rows: 16,376
## Columns: 14
## $ page_id        <int> 1678, 7139, 64786, 1868, 2460, 2458, 2166, 1833,
2948~
## $ name           <chr> "Spider-Man (Peter Parker)", "Captain America (St
even~
## $ urlslug        <chr> "\\Spider-Man_(Peter_Parker)", "\\Captain_Ameri
ca_(~
## $ ID             <chr> "Secret Identity", "Public Identity", "Public Ide
ntit~
## $ ALIGN          <chr> "Good Characters", "Good Characters", "Neutral Ch
arac~
## $ EYE            <chr> "Hazel Eyes", "Blue Eyes", "Blue Eyes", "Blue Eye
s", ~
## $ HAIR           <chr> "Brown Hair", "White Hair", "Black Hair", "Black
Hair~
## $ SEX            <chr> "Male Characters", "Male Characters", "Male Chara
cter~
## $ GSM            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, N~
## $ ALIVE          <chr> "Living Characters", "Living Characters", "Living
Cha~
## $ APPEARANCES    <int> 4043, 3360, 3061, 2961, 2258, 2255, 2072, 2017, 1
955,~
## $ FIRST.APPEARANCE <chr> "Aug-62", "Mar-41", "Oct-74", "Mar-63", "Nov-50",
"No~
## $ Year           <int> 1962, 1941, 1974, 1963, 1950, 1961, 1961, 1962, 1
963,~
## $ Universe       <chr> "Marvel", "Marvel", "Marvel", "Marvel", "Marvel",
"Ma~

```

```

names(marvel)[names(marvel) == "Year"] <- "YEAR"
comics <- rbind(dc, marvel)
glimpse(comics)

```

```

## Rows: 23,272
## Columns: 14

```

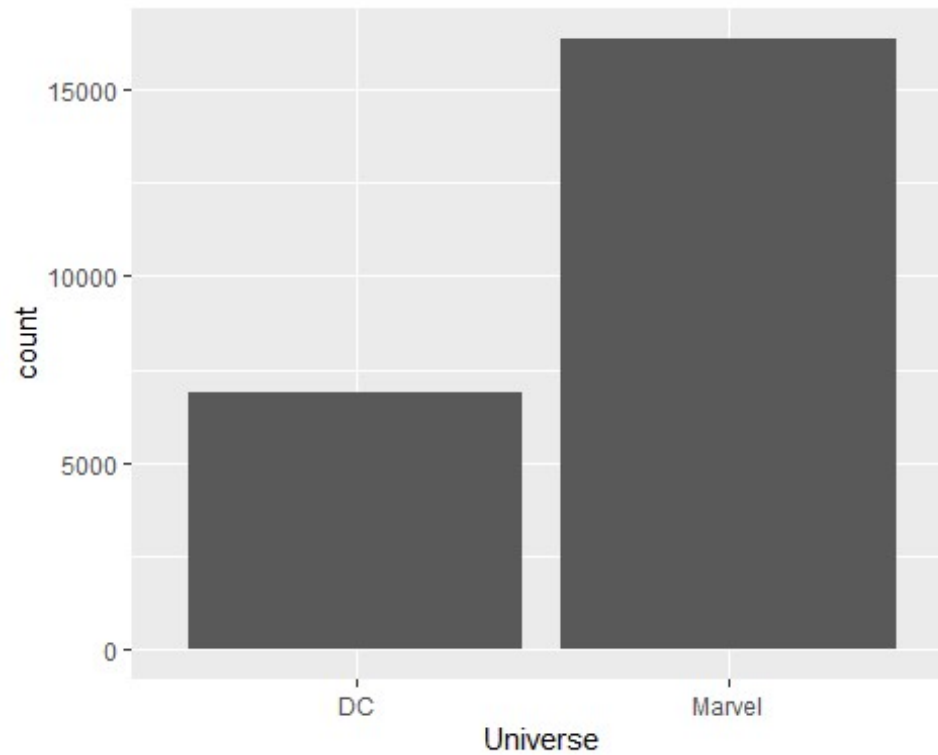
```
## $ page_id      <int> 1422, 23387, 1458, 1659, 1576, 1448, 1486, 1451,
7176~
## $ name         <chr> "Batman (Bruce Wayne)", "Superman (Clark Kent)",
"Gre~
## $ urlslug      <chr> "\\wiki\\Batman_(Bruce_Wayne)", "\\wiki\\Supe
rman~
## $ ID           <chr> "Secret Identity", "Secret Identity", "Secret Ide
ntit~
## $ ALIGN        <chr> "Good Characters", "Good Characters", "Good Chara
cter~
## $ EYE          <chr> "Blue Eyes", "Blue Eyes", "Brown Eyes", "Brown Ey
es",~
## $ HAIR         <chr> "Black Hair", "Black Hair", "Brown Hair", "White
Hair~
## $ SEX          <chr> "Male Characters", "Male Characters", "Male Chara
cter~
## $ GSM          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, N~
## $ ALIVE        <chr> "Living Characters", "Living Characters", "Living
Cha~
## $ APPEARANCES  <int> 3093, 2496, 1565, 1316, 1237, 1231, 1121, 1095, 1
075,~
## $ FIRST.APPEARANCE <chr> "1939, May", "1986, October", "1959, October", "1
987,~
## $ YEAR         <int> 1939, 1986, 1959, 1987, 1940, 1941, 1941, 1989, 1
969,~
## $ Universe     <chr> "DC", "DC", "DC", "DC", "DC", "DC", "DC", "DC", "
```

## Questions to ask

Which comic universe is more accurate in terms of female and queer representation, how consistent is said representation, and how often does the comic's universe fall back on bad tropes such as "bury your gays" or "villainous queer"?

## Plots and Observations

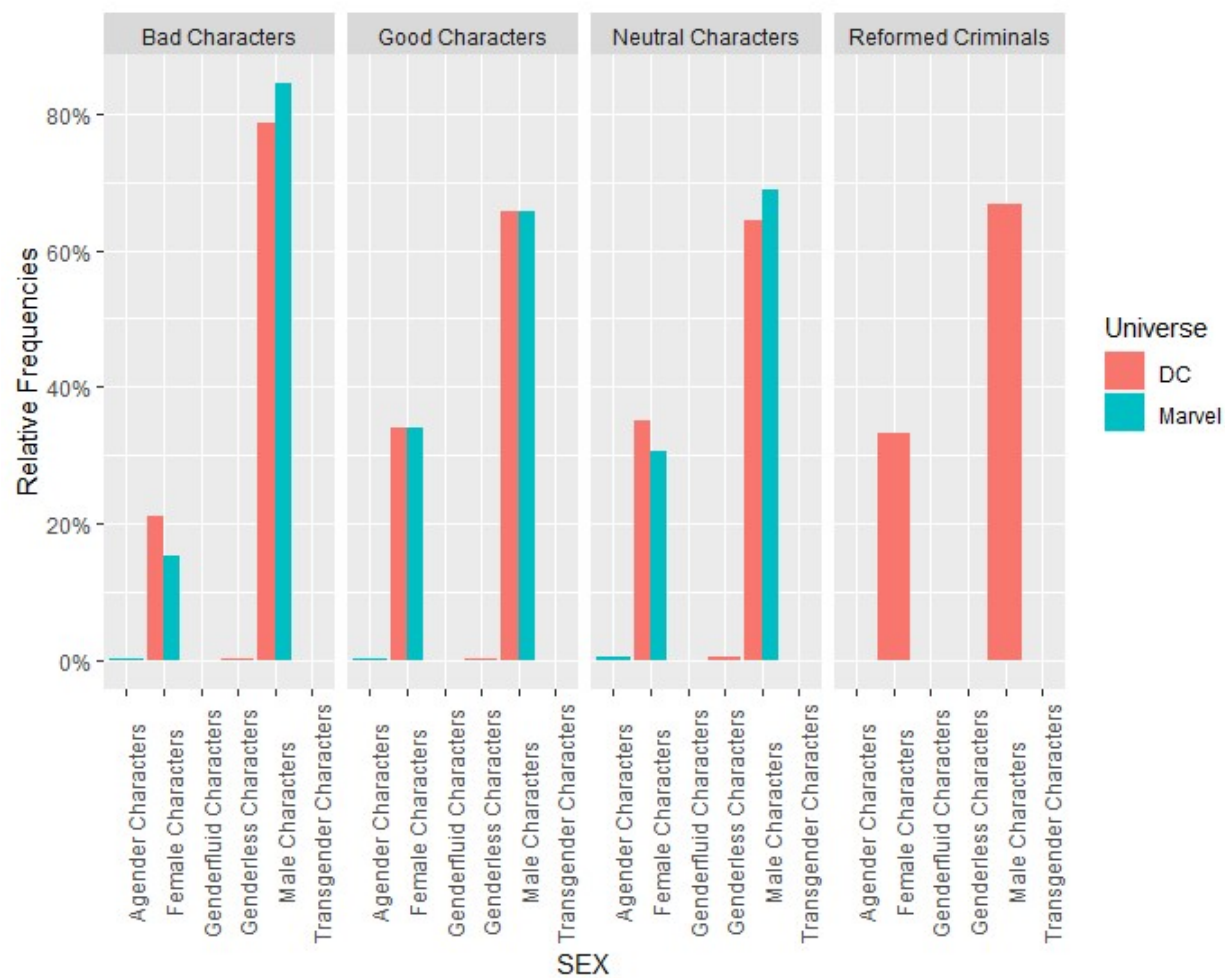
```
ggplot(comics, aes(x = Universe)) + geom_bar()
```

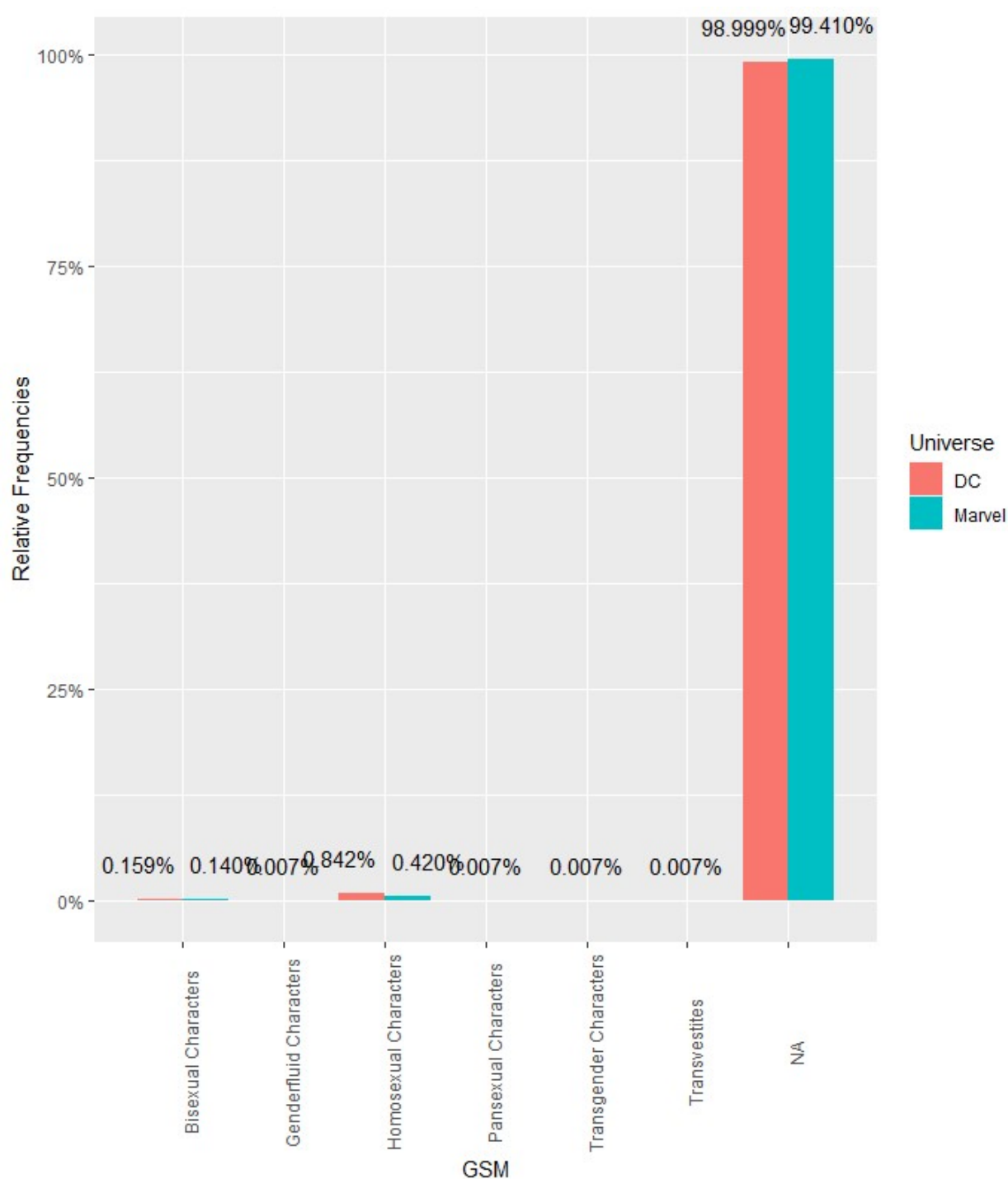


```
nrow(marvel)/nrow(dc)
```

```
## [1] 2.37471
```

First off, Marvel has a lot more characters than DC. Roughly 2.4 times as many to be precise. This means that our graphs will be using proportions rather than counts. Otherwise our analysis will be heavily biased around Marvel.



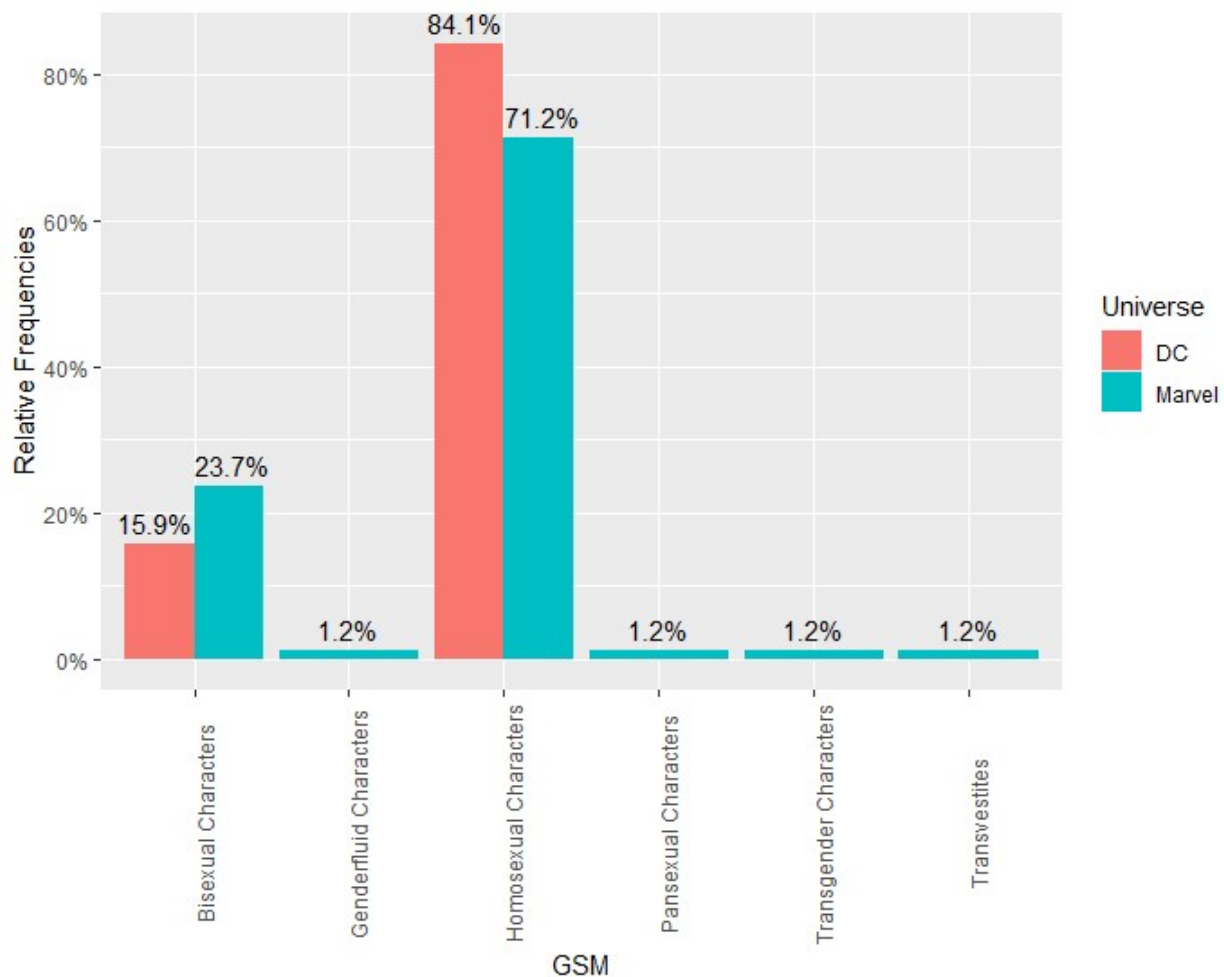


Both companies have a lot more male characters than anything else, which is nothing new. While unrelated to our data, it is interesting to note that DC utilizes reformed criminals for characters, while Marvel tends to be more concrete about their character's alignment.

In terms of character sexuality, all NA's are assumed to be heterosexual and cisgendered because the data states what gender or sexual minority the character belongs to, if they are a part of one. Using this we can see that 98.999% of DC characters are straight-cis, while

99.410% of Marvel characters are straight-cis. This means that all non-cis and non-straight characters in DC comics make up a combined 1.001% of their total cast. Marvel is lower with only 0.59%. According to a gallup poll in 2021, 7.1% of Americans identified as “Something Other Than Heterosexual” (<https://news.gallup.com/poll/332522/percentage-americans-lgbt.aspx>). This means that both comic creators are inaccurately representing social groups.

```
ggplot(subset(comics, is.na(ALIGN) == FALSE & is.na(GSM) == FALSE),
  aes(x = GSM, group = Universe)) +
  geom_bar(aes(y = ..prop.., fill = Universe),
    stat = "count", position = "dodge") +
  geom_text(aes( label = scales::percent(..prop..), y= ..prop.. ),
    stat= "count", vjust = -.5, position = position_dodge(1)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 90)) + ylab("Relative Frequencies"
)
```



```
table(comics$GSM, comics$Universe)
```

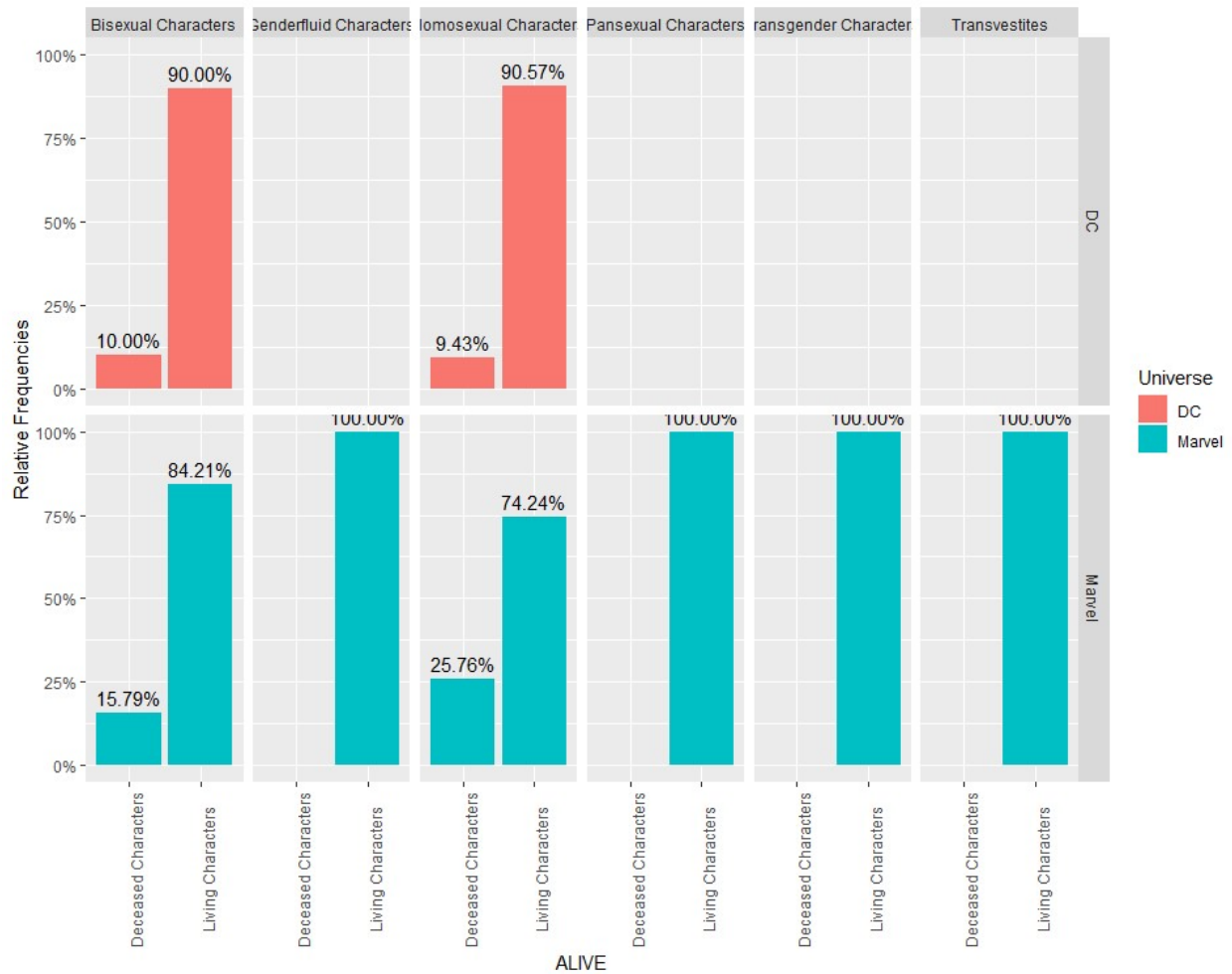
```
##
##
##      Bisexual Characters    10    19
##      Genderfluid Characters  0     1
##      Homosexual Characters  54    66
##      Pansexual Characters   0     1
##      Transgender Characters  0     2
##      Transvestites         0     1
```

To start with, we'll take a subset of the data where characters have a moral alignment, and are a member of a gender or sexual minority. Once straight-cis characters are removed from the data, we can see that homosexual characters are the most common, followed by bisexual. Note that DC only creates these kinds of characters, whereas Marvel will also make genderfluid, pansexual, transgender, and transvestite characters. It's important to be aware that while the graph shows 84.1% of DC characters are homosexual, that's 84.1% of 1.001%: so 0.0842%. Looking at a table breakdown of the number of GSM characters for each comic company, we can see that that 0.0842% gay DC characters equates to roughly 54 total characters.

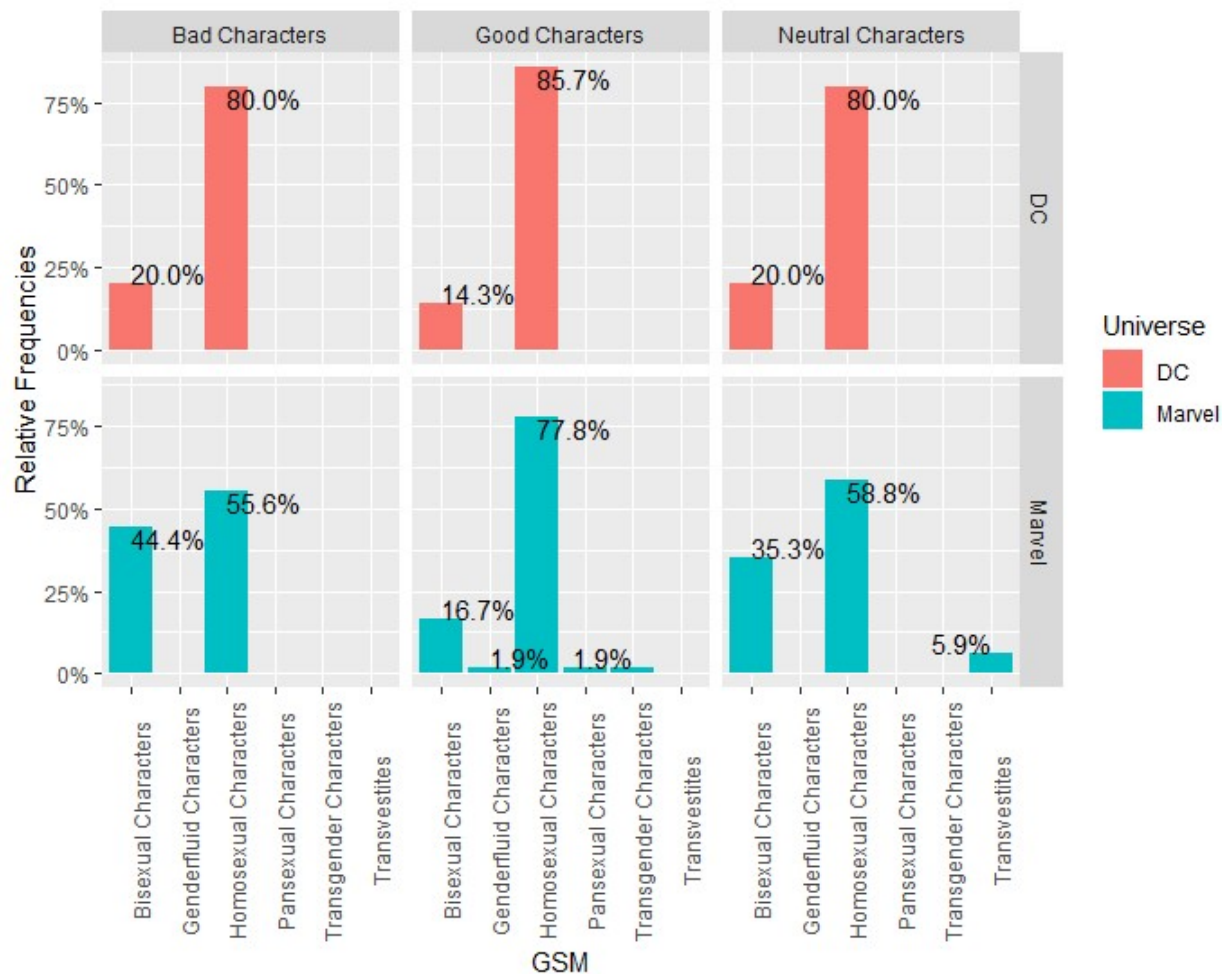
## Tropes

```
ggplot(subset(comics,
               is.na(ALIVE) == FALSE & is.na(GSM) == FALSE & is.na(ALIGN) == F
               ALSE),
       aes(x = ALIVE, group = Universe)) +
  geom_bar(aes(y = ..prop.., fill = Universe),
           stat = "count", position = "dodge") +
  geom_text(aes( label = scales::percent(..prop..), y= ..prop.. ),
            stat= "count", vjust = -.5, position = position_dodge(1)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 90)) + ylab("Relative Frequencies"
) +
  facet_grid(Universe ~ GSM)
```





```
ggplot(subset(comics,
  is.na(ALIVE) == FALSE &
  is.na(GSM) == FALSE &
  is.na(ALIGN) == FALSE),
  aes(x = GSM, group = Universe)) +
  geom_bar(aes(y = ..prop.., fill = Universe),
    stat = "count", position = "dodge") +
  geom_text(aes(label = scales::percent(..prop..), y = ..prop..),
    stat = "count", vjust = "inward", hjust = "inward",
    position = position_dodge(1)) +
  scale_y_continuous(labels = scales::percent) +
  theme(axis.text.x = element_text(angle = 90)) + ylab("Relative Frequencies")
) +
  facet_grid(Universe ~ ALIGN)
```



```
table(marvel$GSM)

##
##      Bisexual Characters Genderfluid Characters Homosexual Characters
##              19                1                66
##      Pansexual Characters Transgender Characters Transvestites
##              1                2                1

GSM_YEAR <- comics %>% filter(GSM %in% c("Transvestites",
                                         "Pansexual Characters",
                                         "Genderfluid Characters",
                                         "Transgender Characters"))

table(GSM_YEAR$GSM, GSM_YEAR$YEAR)

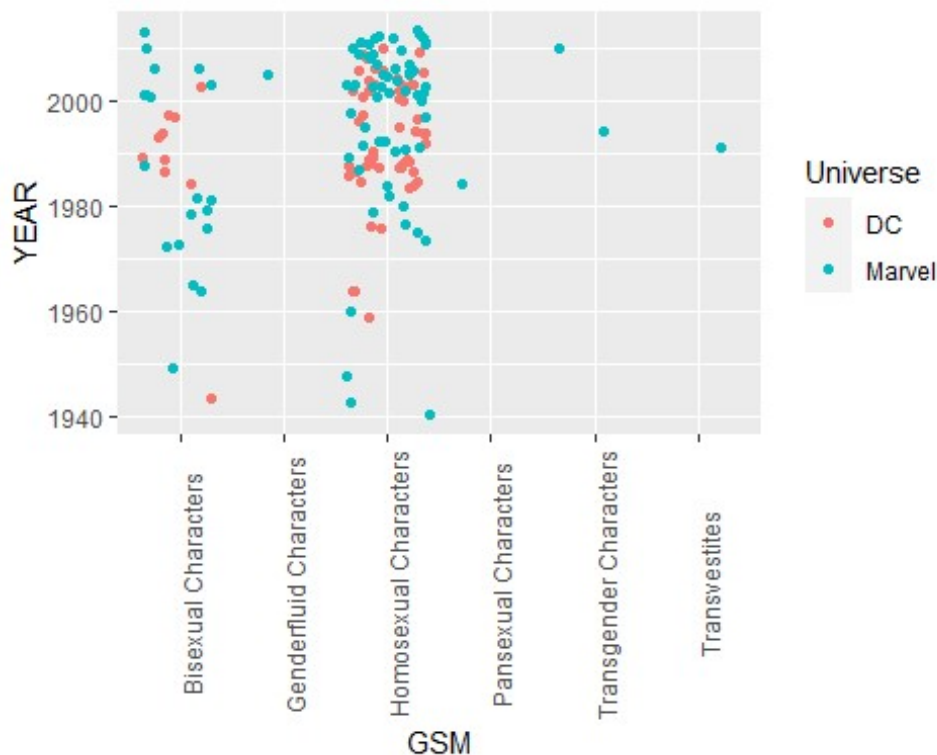
##
##              1984 1991 1994 2005 2010
## Genderfluid Characters    0    0    0    1    0
## Pansexual Characters     1    0    0    0    0
## Transgender Characters    0    0    1    0    1
## Transvestites            0    1    0    0    0
```

To discuss the first trope of “burying gays”, it’s important to know what the trope refers to. The act of “burying your gays” in writing refers to the idea that LGBTQ+ characters are viewed as more expendable than heterosexual their counterparts. This has roots in what is colloquially known as the “depraved homosexual” trope, which caused non-hetero characters to be either villains or characters undeserving of respect. For more information on either of these tropes, (<https://tvtropes.org/pmwiki/pmwiki.php/Main/BuryYourGays>) is an easily approachable starting point. We can see quite readily that Marvel tends to follow the “bury your gays” trope a lot more than DC. Almost 26% of their gay characters are dead, and ~16% of their bisexual ones. It’s worht noting that they haven’t killed off other minorities. However, they might not have had time to do so as their oldest member of t his subset is from the 1980s (Julie Power (Earth-616)), and there are only a few in their stories (1 pan, 1 fluid, 2 transexual, and 1 transvestive).

As for whether or not queer characters tend to be more evil in comics, we can see from the graphs that they don’t. Unless they’re bisexual. Both creators try to cast their queer characters as heros, or at least as neutral, however social stigma attached to bisexual people are also visible in the characters.

```
ggplot(subset(comics, is.na(GSM) == FALSE),
  aes(x = GSM, y = YEAR, color = Universe)) +
  geom_point(position = "jitter") +
  theme(axis.text.x = element_text(angle = 90))
```

## Warning: Removed 6 rows containing missing values (geom\_point).



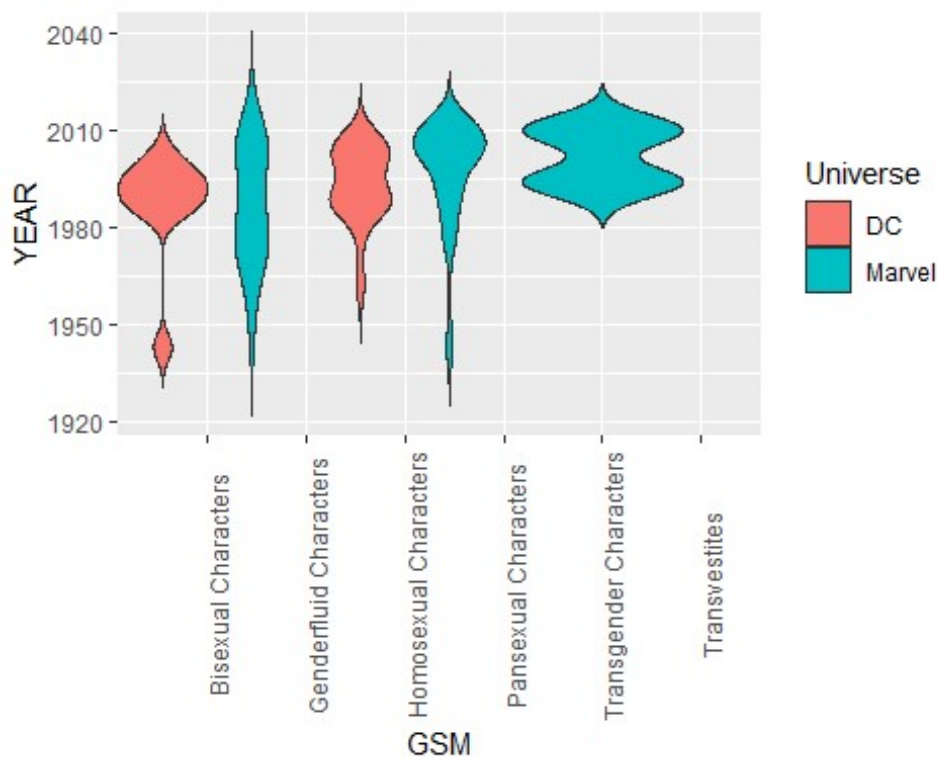
```
ggplot(subset(comics, is.na(GSM) == FALSE),
       aes(x = GSM, y = YEAR, fill = Universe)) +
  geom_violin(trim = FALSE) +
  theme(axis.text.x = element_text(angle = 90))
```

## Warning: Removed 6 rows containing non-finite values (stat\_ydensity).

## Warning: Groups with fewer than two data points have been dropped.

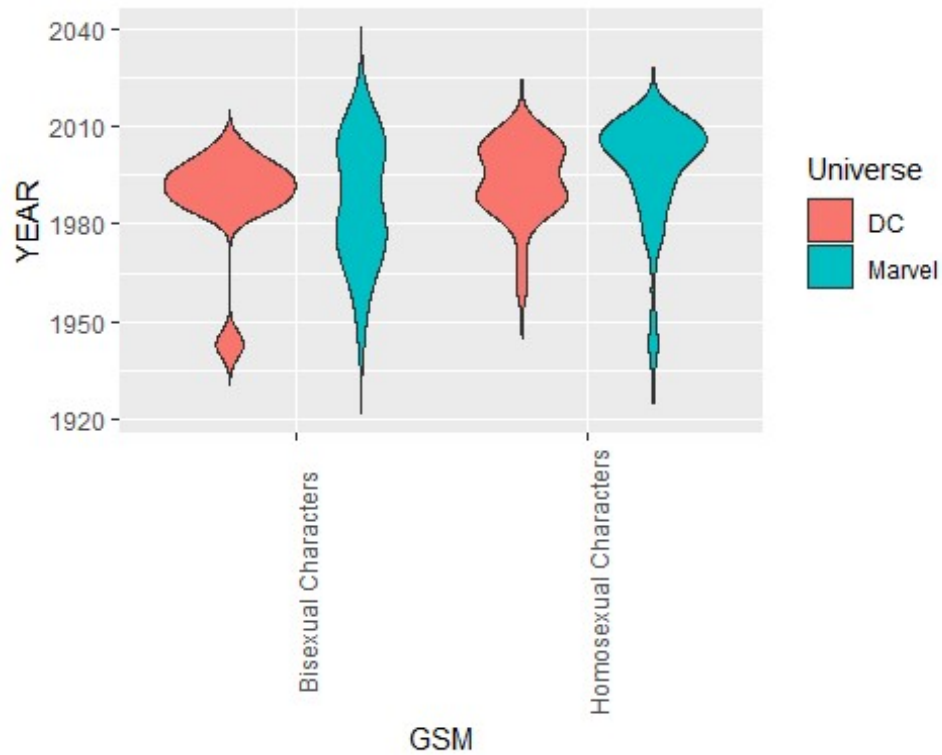
## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.



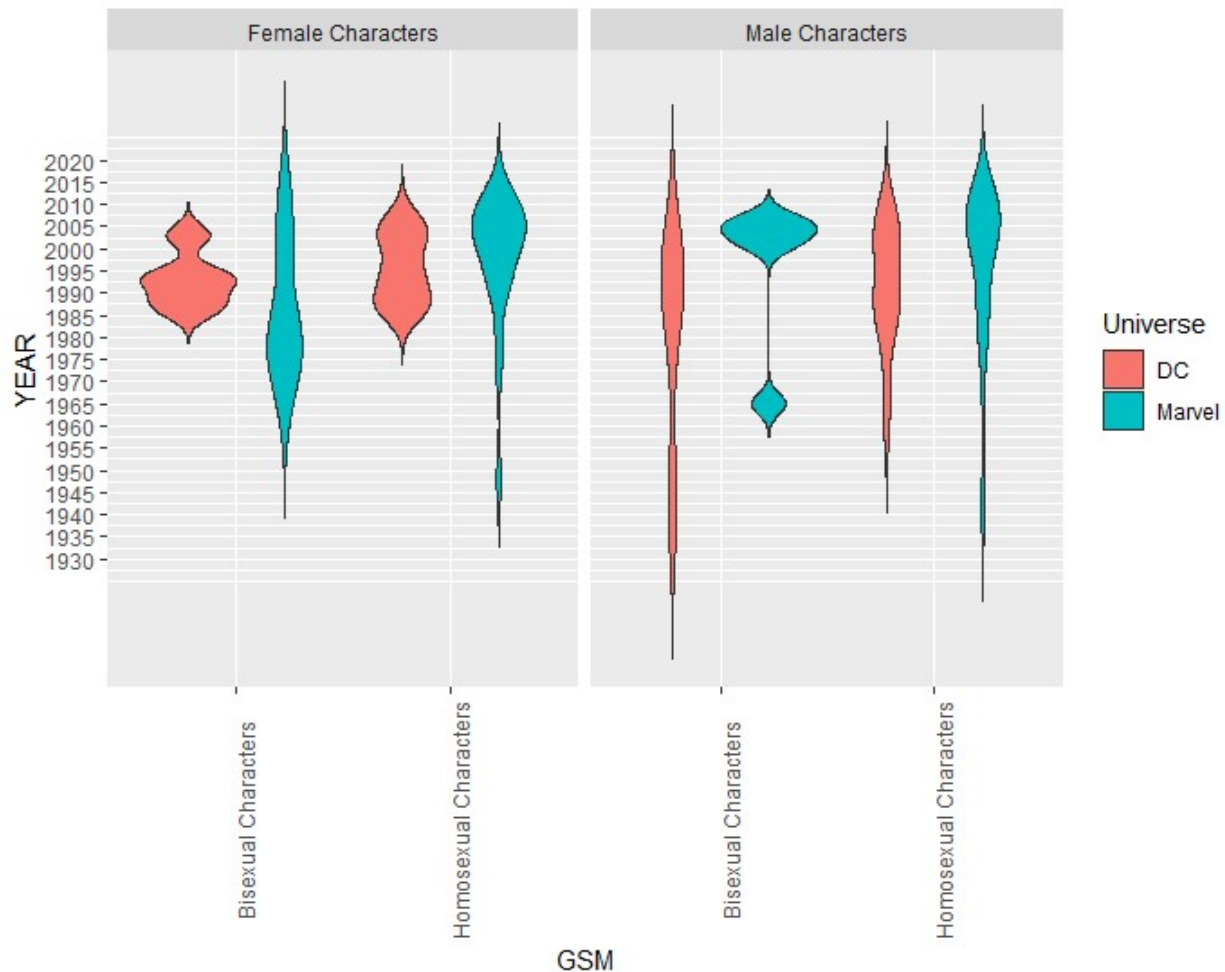
```
ggplot(subset(comics, GSM == "Bisexual Characters" |
              GSM == "Homosexual Characters"),
       aes(x = GSM, y = YEAR, fill = Universe)) +
  geom_violin(trim = FALSE) + theme(axis.text.x = element_text(angle = 90))
```

## Warning: Removed 6 rows containing non-finite values (stat\_ydensity).



```
ggplot(subset(comics,
              GSM %in% c("Bisexual Characters", "Homosexual Characters") &
              SEX %in% c("Female Characters", "Male Characters")),
       aes(x = GSM, y = YEAR, fill = Universe)) +
  geom_violin(trim = FALSE) + theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(breaks = c(1930, 1935, 1940, 1945, 1950, 1955, 1960, 1965,
                                1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005,
                                2010, 2015, 2020)) + facet_grid(~ SEX)

## Warning: Removed 6 rows containing non-finite values (stat_ydensity).
```



Marvel has the first openly gay character in one of their comics, Jack Casey, who first appeared in 1940. The character had a total of 23 appearances before the character committed suicide. It's worth noting that the character had no superpowers or abilities. Overall, you can see that Marvel Comics were more open and accepting of non-hetero characters than DC. Both companies show a marked uptick in their GSM character representation after the 1980s.

Now with just the two variables with more than 2 datapoints, we can see that Marvel has been fairly consistent in their (rather low) level of support for sexual minorities. DC was less open minded and actually saw almost no new GSM characters during the civil rights era. However, when faceted by character sex (still dropping those with less than 2 data points) it gets a little strange. Marvel's limited approval of GSM characters was limited to females, with a preference for bisexual ones. They also drastically reduced the amount of new bisexual male characters during the AIDS epidemic (starting in 1981). This is likely due to the cultural idea that bisexual men could spread the disease to women, who could then spread it to straight men. It was a stigma at the time that made bisexual men seem more unpleasant than their female counterparts, and one that is still difficult for them to shake.

## Conclusions

From the look of it, Marvel tends to have higher levels of representation of women and sexual minorities, but they also tend to follow problematic tropes such as “burying their gays”. It’s also worth noting that Marvel tends to prefer bisexual female characters to lesbian ones, and that they seemed to abandon important members of the queer community at the start of the American AIDS epidemic. DC on the other hand, is late in creating female and queer characters, but they’ve been more consistent in their treatment of them. However, both companies don’t create an accurate level of representation of non-straight male characters. In all areas, they do not reflect national percentages. It’s slow going, but at least it’s going. The long term goal for both comic companies would be to continually create new characters with narrative depth that abandon problematic tropes, and to do so in large enough numbers to at least accurately represent the real world.