Marvel vs DC Visualizations

Maya Reese Farmer

October 29, 2021

Introduction I'm going to be querying and visualizing data on Marvel and DC movies to determine which Universe has performed the best in terms of revenue, profits, and movie ratings. I'm using the Marvel vs DC dataset from kaggle.com

This dataset originally only included movies released up to 2019. I updated the dataset and added additional movies up to 2021. It now contains 44 observations and has 10 columns with data on:

- title
- company (Marvel or DC)
- rate
- metascore
- duration in minutes
- release year
- budget
- opening weekend USA (gross revenue)
- gross USA (gross domestic revenue)
- gross worldwide (gross revenue worldwide)

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## importing dataset into r
marveldc <- read_csv("marveldc.csv")
```

```
## Rows: 44 Columns: 10

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Title, Company
## dbl (8): Rate, Metascore, Minutes, Release, Budget, OpeningWeekendUSA, Gross...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## view data
head(marveldc,10)
```

```
## # A tibble: 10 x 10
##    Title         Company  Rate Metascore Minutes Release Budget OpeningWeekendU~
##    <chr>         <chr>   <dbl>     <dbl>   <dbl>   <dbl>  <dbl>            <dbl>
##  1 Iron Man      Marvel    7.9        79     126    2008  1.4e8         98618668
##  2 The Incredib~ Marvel    6.7        61     112    2008  1.5e8         55414050
##  3 Iron Man 2    Marvel    7          57     124    2010  2  e8        128122480
##  4 Thor          Marvel    7          57     115    2011  1.5e8         65723338
##  5 Captain Amer~ Marvel    6.9        66     124    2011  1.4e8         65058524
##  6 The Avengers  Marvel    8          69     143    2012  2.2e8        207438708
##  7 Iron Man 3    Marvel    7.2        62     130    2013  2  e8        174144585
##  8 Thor: The Da~ Marvel    6.9        54     112    2013  1.7e8         85737841
##  9 Captain Amer~ Marvel    7.7        70     136    2014  1.7e8         95023721
## 10 Guardians of~ Marvel    8          76     121    2014  1.7e8         94320883
## # ... with 2 more variables: GrossUSA <dbl>, GrossWorldwide <dbl>
```

The data are pretty clean and this is a small data set so no data prep is necessary.

Queries and Visualizations I already performed most of my queries in SQL, but I will repeat them here and show how I might visualize these data.

```
library(cowplot)
library(scales)
```

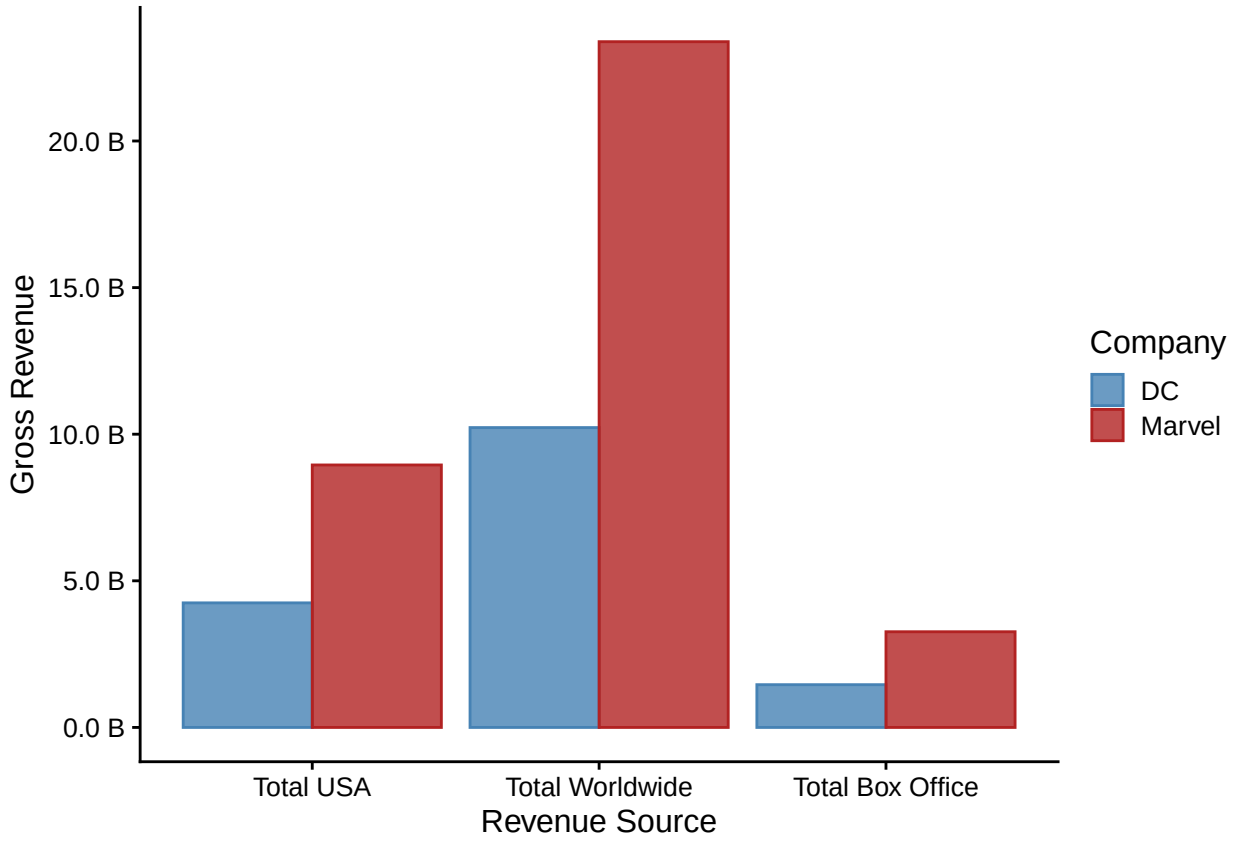
task 1: summary data on revenue, profit, and ratings

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
## calculate total gross USA, worldwide, and opening
## weekend
marveldc %>%
    group_by(Company) %>%
    summarise(total.grossUSA = sum(GrossUSA), total.grossworldwide = sum(GrossWorldwide),
        total.opening = sum(OpeningWeekendUSA))
```

```
## # A tibble: 2 x 4
##   Company total.grossUSA total.grossworldwide total.opening
##   <chr>            <dbl>                <dbl>         <dbl>
## 1 DC          4248391442          10225915033    1457976968
## 2 Marvel      8950642892          23387207891    3262976463
```

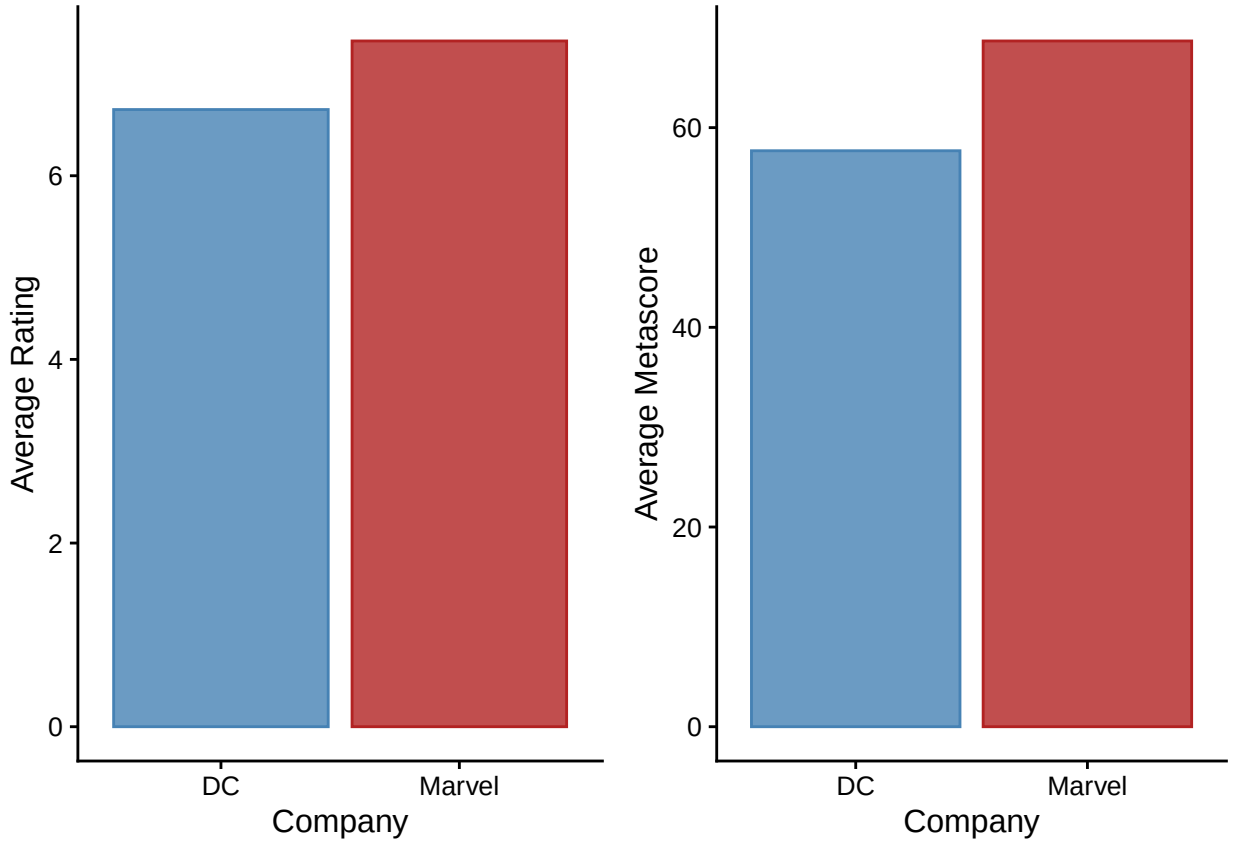
```
marveldc %>%
    group_by(Company) %>%
    summarise(total.grossUSA = sum(GrossUSA), total.grossworldwide = sum(GrossWorldwide),
        total.opening = sum(OpeningWeekendUSA)) %>%
    pivot_longer(2:4, names_to = "revenue.source", values_to = "gross.revenue") %>%
    ggplot(aes(x = revenue.source, y = gross.revenue, fill = Company,
        color = Company)) + geom_col(position = "dodge") + theme_cowplot(12) +
    scale_fill_manual(values = alpha(c("steelblue", "firebrick"),
        0.8)) + scale_color_manual(values = c("steelblue", "firebrick")) +
    labs(x = "Revenue Source", y = "Gross Revenue") + scale_x_discrete(labels = c(total.grossUSA = "Tota
    total.grossworldwide = "Total Worldwide", total.opening = "Total Box Office")) +
    scale_y_continuous(labels = unit_format(unit = "B", scale = 1e-09))
```



We can see that Marvel has accumulated more than double the amount that DC has in terms of gross revenue.

```
## calculate average movie rating and metascore for each
## company
rate.plot = marveldc %>%
    group_by(Company) %>%
    summarise(avg.rate = mean(Rate)) %>%
```

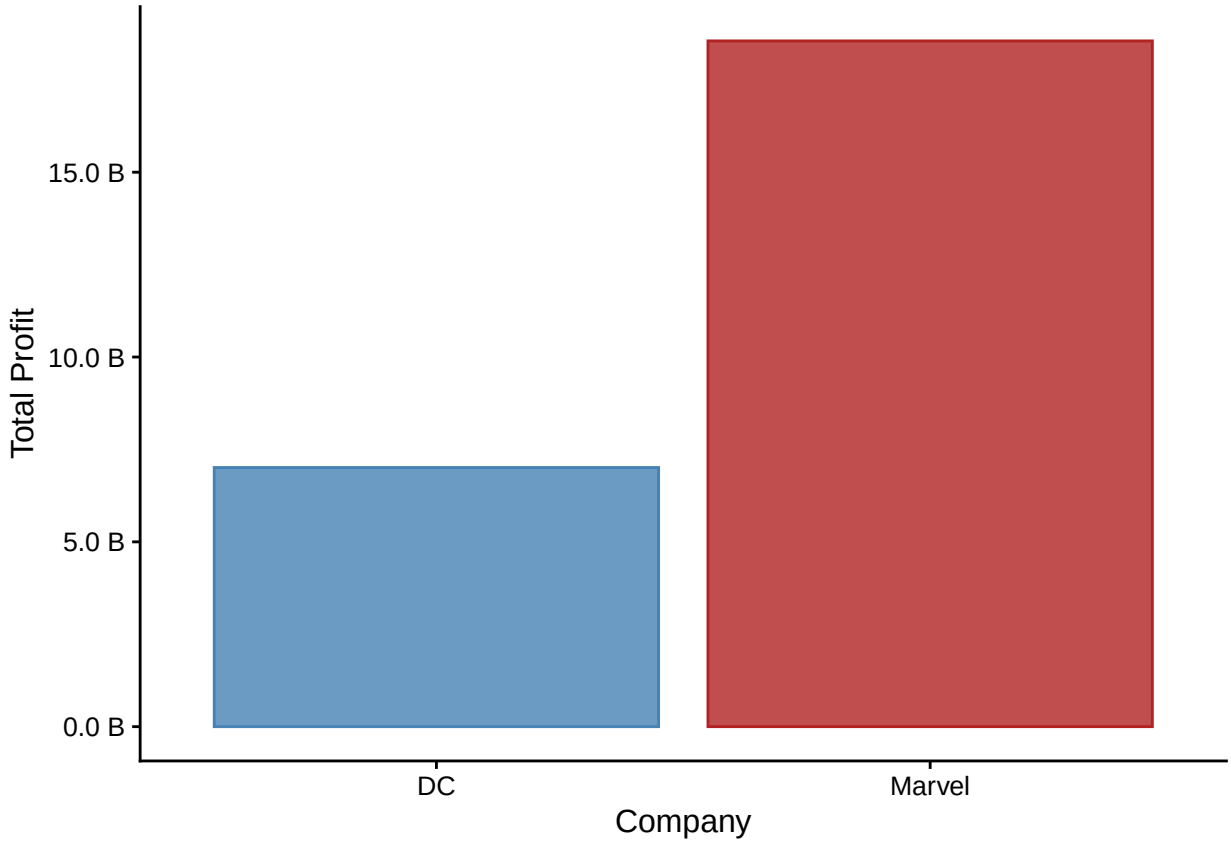
```
    ggplot(aes(x = Company, y = avg.rate, fill = Company, color = Company)) +
    geom_col() + theme_cowplot(12) + scale_fill_manual(values = alpha(c("steelblue",
    "firebrick"), 0.8)) + scale_color_manual(values = c("steelblue",
    "firebrick")) + labs(y = "Average Rating") + theme(legend.position = "none")

meta.plot = marveldc %>%
    group_by(Company) %>%
    summarise(avg.meta = mean(Metascore)) %>%
    ggplot(aes(x = Company, y = avg.meta, fill = Company, color = Company)) +
    geom_col() + theme_cowplot(12) + scale_fill_manual(values = alpha(c("steelblue",
    "firebrick"), 0.8)) + scale_color_manual(values = c("steelblue",
    "firebrick")) + labs(y = "Average Metascore") + theme(legend.position = "none")

plot_grid(rate.plot, meta.plot, nrow = 1)
```



We can see that both the average movie rating and metascore are higher in Marvel movies.

```
## calculate total profit generated by each company
marveldc %>%
    mutate(Profit = GrossWorldwide - Budget) %>%
    group_by(Company) %>%
    summarise(total.profit = sum(Profit)) %>%
    ggplot(aes(x = Company, y = total.profit, fill = Company,
        color = Company)) + geom_col() + theme_cowplot(12) +
    scale_fill_manual(values = alpha(c("steelblue", "firebrick"),
        0.8)) + scale_color_manual(values = c("steelblue", "firebrick")) +
    labs(y = "Total Profit") + scale_y_continuous(labels = unit_format(unit = "B",
```

```
    scale = 1e-09)) + theme(legend.position = "none")
```



Finally, Marvel has generated over 2x the profit that DC has.

```
## top 10 movies by rating
marveldc %>%
    arrange(desc(Rate)) %>%
    top_n(10, Rate) %>%
    select(Title, Company, Rate)
```

task 2: Best Movies Overall

```
## # A tibble: 11 x 3
##    Title                                Company  Rate
##    <chr>                                <chr>   <dbl>
##  1 The Dark Knight                      DC        9
##  2 Joker                                DC        8.7
##  3 Avengers: Infinity War               Marvel    8.5
##  4 Avengers: Endgame                    Marvel    8.5
##  5 The Dark Knight Rises                DC        8.4
##  6 Batman Begins                        DC        8.2
##  7 The Avengers                         Marvel    8
##  8 Guardians of the Galaxy              Marvel    8
##  9 Iron Man                             Marvel    7.9
```

```
## 10 Thor: Ragnarok                              Marvel    7.9
## 11 Shang Chi and the Legend of the Ten Rings Marvel    7.9
```

```
## top 10 movies by opening weekend
marveldc %>%
    arrange(desc(OpeningWeekendUSA)) %>%
    top_n(10, OpeningWeekendUSA) %>%
    select(Title, Company, OpeningWeekendUSA)
```

```
## # A tibble: 10 x 3
##    Title                             Company OpeningWeekendUSA
##    <chr>                             <chr>               <dbl>
##  1 Avengers: Endgame                 Marvel          357115007
##  2 Avengers: Infinity War            Marvel          257698183
##  3 The Avengers                      Marvel          207438708
##  4 Black Panther                     Marvel          202003951
##  5 Avengers: Age of Ultron           Marvel          191271109
##  6 Captain America: Civil War        Marvel          179139142
##  7 Iron Man 3                        Marvel          174144585
##  8 Batman v Superman: Dawn of Justice DC             166007347
##  9 The Dark Knight Rises             DC              160887295
## 10 The Dark Knight                   DC              158411483
```

```
## top 10 movies by gross worldwide
marveldc %>%
    arrange(desc(GrossWorldwide)) %>%
    top_n(10, GrossWorldwide) %>%
    select(Title, Company, GrossWorldwide)
```

```
## # A tibble: 10 x 3
##    Title                       Company GrossWorldwide
##    <chr>                       <chr>            <dbl>
##  1 Avengers: Endgame           Marvel      2797800564
##  2 Avengers: Infinity War      Marvel      2048359754
##  3 The Avengers                Marvel      1518812988
##  4 Avengers: Age of Ultron     Marvel      1402805868
##  5 Black Panther               Marvel      1346913161
##  6 Iron Man 3                  Marvel      1214811252
##  7 Captain America: Civil War  Marvel      1153296293
##  8 Aquaman                     DC          1148161807
##  9 Spider-Man: Far from Home   Marvel      1131927996
## 10 Captain Marvel              Marvel      1128274794
```

```
## top 10 movies by profit
marveldc %>%
    mutate(Profit = GrossWorldwide - Budget) %>%
    arrange(desc(Profit)) %>%
    top_n(10, Profit) %>%
    select(Title, Company, Profit)
```

```
## # A tibble: 10 x 3
##    Title                       Company     Profit
```

```
##    <chr>                       <chr>        <dbl>
##  1 Avengers: Endgame           Marvel  2441800564
##  2 Avengers: Infinity War      Marvel  1727359754
##  3 The Avengers                Marvel  1298812988
##  4 Avengers: Age of Ultron     Marvel  1152805868
##  5 Black Panther               Marvel  1146913161
##  6 Iron Man 3                  Marvel  1014811252
##  7 Joker                       DC      1005504580
##  8 Aquaman                     DC       988161807
##  9 Spider-Man: Far from Home   Marvel   971927996
## 10 Captain Marvel              Marvel   953274794
```

When evaluating the top 10 movies by rating, DC has 4 movies in the top 10 while Marvel has 7 (there are a couple ties). The highest rated film is The Dark Knight (rightfully so) and the other DC films in the top 10 are from the Batman Trilogy and Joker movie.
When considering best opening weekend revenue in the USA, Marvel movies take up the top 7 spots with DC movies in the bottom 3. However, DC only has one film, Aquaman, in the top 10 when looking at gross worldwide revenue. Marvel films take up the other 9 spots. And finally, when we examine movies that generated the most profit, 8 of the top 10 are Marvel movies and only 2 were from DC.
Overall, Marvel has outperformed DC in almost every metric.

```
## top 5 movies by grossworldwide
marveldc %>%
    group_by(Company) %>%
    top_n(5, GrossWorldwide) %>%
    arrange(desc(GrossWorldwide)) %>%
    select(Title, Company, GrossWorldwide)
```

task 3: Top 5 Movies by Revenue, Profit, and Rating for each Studio

```
## # A tibble: 10 x 3
## # Groups:   Company [2]
##    Title                              Company GrossWorldwide
##    <chr>                              <chr>            <dbl>
##  1 Avengers: Endgame                  Marvel      2797800564
##  2 Avengers: Infinity War             Marvel      2048359754
##  3 The Avengers                       Marvel      1518812988
##  4 Avengers: Age of Ultron            Marvel      1402805868
##  5 Black Panther                      Marvel      1346913161
##  6 Aquaman                            DC          1148161807
##  7 The Dark Knight Rises              DC          1081041287
##  8 Joker                              DC          1060504580
##  9 The Dark Knight                    DC          1004934033
## 10 Batman v Superman: Dawn of Justice DC           873634919
```

```
## bottom 3 movies based on gross worldwide
marveldc %>%
    group_by(Company) %>%
    top_n(-3, GrossWorldwide) %>%
    arrange(GrossWorldwide) %>%
    select(Title, Company, GrossWorldwide)
```

```
## # A tibble: 6 x 3
## # Groups:   Company [2]
##   Title                                Company GrossWorldwide
##   <chr>                                <chr>            <dbl>
## 1 Jonah Hex                            DC            10903312
## 2 Catwoman                             DC            82102379
## 3 Wonder Woman 1984                    DC           166534027
## 4 The Incredible Hulk                  Marvel       263427551
## 5 Captain America: The First Avenger Marvel       370569774
## 6 Black Widow                          Marvel       379631351
```

```
## top 5 movies by profit (different method using rank)
marveldc %>%
    group_by(Company) %>%
    mutate(Profit = GrossWorldwide - Budget, ranking = rank(desc(Profit),
        ties.method = "first")) %>%
    select(Title, Company, Profit, ranking) %>%
    arrange(ranking) %>%
    filter(ranking < 6)
```

```
## # A tibble: 10 x 4
## # Groups:   Company [2]
##    Title                   Company     Profit ranking
##    <chr>                   <chr>        <dbl>   <int>
##  1 Avengers: Endgame       Marvel  2441800564       1
##  2 Joker                   DC      1005504580       1
##  3 Avengers: Infinity War  Marvel  1727359754       2
##  4 Aquaman                 DC       988161807       2
##  5 The Avengers            Marvel  1298812988       3
##  6 The Dark Knight Rises   DC       831041287       3
##  7 Avengers: Age of Ultron Marvel  1152805868       4
##  8 The Dark Knight         DC       819934033       4
##  9 Black Panther           Marvel  1146913161       5
## 10 Wonder Woman            DC       672847012       5
```

```
## bottom 3 movies by profit for each studio
marveldc %>%
    group_by(Company) %>%
    mutate(Profit = GrossWorldwide - Budget, ranking = rank(desc(Profit),
        ties.method = "first")) %>%
    select(Title, Company, Profit, ranking) %>%
    arrange(Company, ranking) %>%
    top_n(-3, Profit)
```

```
## # A tibble: 6 x 4
## # Groups:   Company [2]
##   Title                                    Company    Profit ranking
##   <chr>                                    <chr>       <dbl>   <int>
## 1 Catwoman                                 DC      -17897621      17
## 2 Wonder Woman 1984                        DC      -33465973      18
## 3 Jonah Hex                                DC      -36096688      19
## 4 Shang Chi and the Legend of the Ten Rings Marvel  222364506      23
## 5 Black Widow                              Marvel  179631351      24
## 6 The Incredible Hulk                      Marvel  113427551      25
```

```r
## top 5 movies by rating
marveldc %>%
    group_by(Company) %>%
    top_n(5, Rate) %>%
    arrange(desc(Rate)) %>%
    select(Title, Company, Rate, Metascore)
```

```
## # A tibble: 12 x 4
## # Groups:   Company [2]
##    Title                                    Company  Rate Metascore
##    <chr>                                    <chr>   <dbl>     <dbl>
##  1 The Dark Knight                          DC        9          84
##  2 Joker                                    DC        8.7        59
##  3 Avengers: Infinity War                   Marvel    8.5        68
##  4 Avengers: Endgame                        Marvel    8.5        78
##  5 The Dark Knight Rises                    DC        8.4        78
##  6 Batman Begins                            DC        8.2        70
##  7 The Avengers                             Marvel    8          69
##  8 Guardians of the Galaxy                  Marvel    8          76
##  9 Iron Man                                 Marvel    7.9        79
## 10 Thor: Ragnarok                           Marvel    7.9        74
## 11 Shang Chi and the Legend of the Ten Rings Marvel    7.9        71
## 12 Watchmen                                 DC        7.6        56
```

```r
## bottom 3 movies by metascore
marveldc %>%
    group_by(Company) %>%
    mutate(meta.rank = rank(desc(Metascore), ties.method = "first")) %>%
    select(Title, Company, Metascore, meta.rank) %>%
    arrange(Company, meta.rank) %>%
    top_n(-3, Metascore)
```

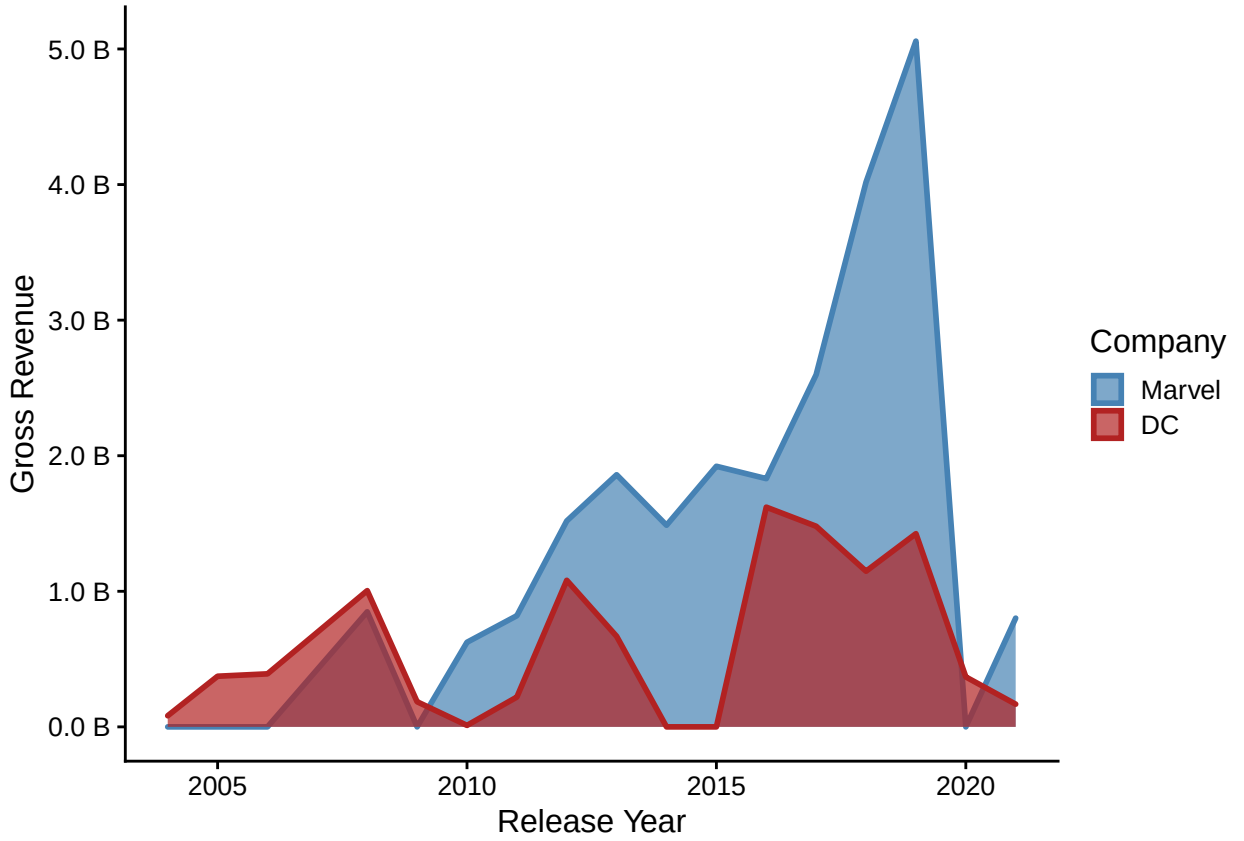
```
## # A tibble: 6 x 4
## # Groups:   Company [2]
##   Title                Company Metascore meta.rank
##   <chr>                <chr>       <dbl>     <int>
## 1 Green Lantern        DC             39        17
## 2 Jonah Hex            DC             33        18
## 3 Catwoman             DC             27        19
## 4 Iron Man 2           Marvel         57        23
## 5 Thor                 Marvel         57        24
## 6 Thor: The Dark World Marvel         54        25
```

```r
marveldc %>%
    group_by(Release, Company) %>%
    summarise(year.gross = sum(GrossWorldwide)) %>%
    ungroup() %>%
    complete(Release, Company, fill = list(year.gross = 0)) %>%
    ggplot(aes(x = Release, y = year.gross, fill = fct_relevel(Company,
        "Marvel", "DC"), color = fct_relevel(Company, "Marvel",
```

```
    "DC"))) + geom_area(size = 1, position = "identity") +
theme_cowplot(12) + scale_y_continuous(labels = unit_format(unit = "B",
scale = 1e-09)) + labs(x = "Release Year", y = "Gross Revenue") +
scale_fill_manual(values = alpha(c("steelblue", "firebrick"),
    0.7), name = "Company") + scale_color_manual(values = c("steelblue",
"firebrick"), name = "Company")
```

task 4: Changes in Revenue over Time

```
## `summarise()` has grouped output by 'Release'. You can override using the `.groups` argument.
```



Although the lines aren't very smooth, I still think this graph is fairly informative. First, it shows that Marvel (as we know it today) was able to make more movies and generate more revenue over a shorter amount of time, compared to DC. It's important to note that although Spiderman is a part of Marvel (or the MCU), the trilogy featuring Tobey Maguire that released in the early 2000's was not produced by Marvel, but instead by Sony (who own the rights to the Spiderman Universe). That's why those movies weren't included in this dataset. Second, there seemed to be a significant increase in revenue over time for Marvel until 2020-2021 where there's a sharp decline. This is because of movie theaters closing and releases being pushed back due to the COVID 19 pandemic. Overall, Marvel has absolutely crushed DC in terms of gross revenue and profit.