MARVEL MOVIES

GROUP MEMBERS:-

AVNI SHARMA 20/49047

PARAS KHAMPA 20/49041

Marvel Movie

The Marvel Cinematic Universe (MCU) is an American media franchise and shared universe centered on a series of superhero films produced by Marvel Studios. The films are based on characters that appear in American comic books published by Marvel Comics. The franchise also includes television series, short films, digital series, and literature. The shared universe, much like the original Marvel Universe in comic books, was established by crossing over common plot elements, settings, cast, and characters.

CONTENT

This dataset includes data of all the movies produced by marvel studios . The data include info like distributor name , budget , release date , rotten tomatoes percentage , metacritic , cinema score

DETAILS ABOUT THE MARVEL DATA SET :

marvel<-read.csv("marvel_movies.csv")

marvel1<-read.csv("marvel_reviews.csv")

- **summary(marvel) :** This shows the summary of the contents of the dataset containing details like budget , revenues and distributorof the marvel movies released so far.

```
> summary(marvel)
    Title           Distributor        ReleaseDateUS          Budget
 Length:64          Length:64          Length:64          Min.   :        0
 Class :character   Class :character   Class :character   1st Qu.:100000000
 Mode  :character   Mode  :character   Mode  :character   Median :150000000
                                                          Mean   :148156250
                                                          3rd Qu.:200000000
                                                          Max.   :356000000
 OpeningWeekendNorthAmerica  NorthAmerica       OtherTerritories
 Min.   :  1500000          Min.   :  1521787   Min.   :1.330e+06
 1st Qu.: 54132044          1st Qu.:132457528   1st Qu.:1.623e+08
 Median : 82693904          Median :213532936   Median :3.000e+08
 Mean   : 90148236          Mean   :247027107   Mean   :3.820e+08
 3rd Qu.:115389963          3rd Qu.:333432735   3rd Qu.:4.985e+08
 Max.   :357115007          Max.   :858373000   Max.   :1.938e+09
   Worldwide
 Min.   :2.852e+06
 1st Qu.:2.881e+08
 Median :5.316e+08
 Mean   :6.291e+08
 3rd Qu.:7.972e+08
 Max.   :2.798e+09
>
```

- summary(marvel1) : This shows the summary of the dataset containing the reviews of the marvel reviews.

```
> summary(marvel1)
    Film           Rotten.Tomatoesin.   Metacritic     CinemaScore        CinemaScore.1
 Length:67          Min.   : 9.00      Min.   : 0.0   Length:67          Length:67
 Class :character   1st Qu.:37.00      1st Qu.:44.0   Class :character   Class :character
 Mode  :character   Median :76.00      Median :63.0   Mode  :character   Mode  :character
                    Mean   :64.37      Mean   :56.7
                    3rd Qu.:90.00      3rd Qu.:70.0
                    Max.   :97.00      Max.   :88.0
```

Dimensions shows the number of observations and the variables(Columns that we have in our dataset).

- dim(marvel) :

```
> dim(marvel)
[1] 64  8
>
```

- dim(marvel1) :

```
> dim(marvel1)
[1] 67  5
>
```

- **str(marvel) :**

```
> str(marvel)
'data.frame':   64 obs. of  8 variables:
 $ Title                    : chr  "Howard the Duck" "Blade" "X-Men" "Blade II" ...
 $ Distributor              : chr  "Universal Pictures" "New Line Cinema" "20th Century Fox" "New Line Cinema"
 ...
 $ ReleaseDateUS            : chr  "01-08-1986 00:00" "21-08-1998 00:00" "14-07-2000 00:00" "22-03-2002 00:00"
 ...
 $ Budget                   : int  37000000 45000000 75000000 54000000 139000000 78000000 110000000 137000000 3
3000000 200000000 ...
 $ OpeningWeekendNorthAmerica: int  5070136 17073856 54471475 32528016 114844116 40310419 85558731 62128420 1383
4527 88156227 ...
 $ NorthAmerica             : int  16295774 70087718 157299717 82348319 403706375 102543518 214949694 132177234
 33810189 373585825 ...
 $ OtherTerritories         : int  21667000 61095812 139039810 72661713 418002176 76636200 192761855 113183246
 20889916 415390628 ...
 $ Worldwide                : num  3.80e+07 1.31e+08 2.96e+08 1.55e+08 8.22e+08 ...
> |
```

- **str(marvel1) :**

```
> str(marvel1)
'data.frame':   67 obs. of  5 variables:
 $ Film             : chr  "Howard the Duck" "The Punisher (1989)" "Captain America (1990)" "The Fantastic Four (1994)"
 ...
 $ Rotten.Tomatoesin.: int  14 28 13 30 57 82 57 90 44 85 ...
 $ Metacritic       : int  28 63 0 0 47 64 52 73 42 68 ...
 $ CinemaScore      : chr  "B" NA NA NA ...
 $ CinemaScore.1    : chr  "B-" NA NA NA ...
> |
```

- **names(marvel) :**

```
> names(marvel)
[1] "Title"                     "Distributor"               "ReleaseDateUS"
[4] "Budget"                    "OpeningWeekendNorthAmerica" "NorthAmerica"
[7] "OtherTerritories"          "Worldwide"
```

- **names(marvel1) :**

```
> names(marvel1)
[1] "Film"              "Rotten.Tomatoesin." "Metacritic"        "CinemaScore"       "CinemaScore.1"
> |
```

Number of columns can be retrieved as:-

- **length(marvel) :**

```
> length(marvel)
[1] 8
```

- **length(marvel1) :**

```
> length(marvel1)
[1] 5
> |
```

The details of the first six observations:-

- **head(marvel) :**

```
> head(marvel)
            Title        Distributor    ReleaseDateUS    Budget OpeningWeekendNorthAmerica NorthAmerica
1 Howard the Duck Universal Pictures 01-08-1986 00:00  37000000                    5070136     16295774
2           Blade    New Line Cinema 21-08-1998 00:00  45000000                   17073856     70087718
3           X-Men   20th Century Fox 14-07-2000 00:00  75000000                   54471475    157299717
4        Blade II    New Line Cinema 22-03-2002 00:00  54000000                   32528016     82348319
5      Spider-Man      Sony Pictures 03-05-2002 00:00 139000000                  114844116    403706375
6       Daredevil   20th Century Fox 14-02-2003 00:00  78000000                   40310419    102543518
  OtherTerritories Worldwide
1         21667000  37962774
2         61095812 131183530
3        139039810 296339527
4         72661713 155010032
5        418002176 821708551
6         76636200 179179718
```

- **head(marvel1) :**

```
> head(marvel1)
                      Film Rotten.Tomatoesin. Metacritic CinemaScore CinemaScore.1
1           Howard the Duck                14         28          B            B-
2       The Punisher (1989)                28         63       <NA>          <NA>
3    Captain America (1990)                13          0       <NA>          <NA>
4 The Fantastic Four (1994)                30          0       <NA>          <NA>
5                     Blade                57         47          A            A-
6                     X-Men                82         64          A            A-
```

The details of the last six observations:-

- **tail(marvel) :**

```
> tail(marvel)
                                    Title                        Distributor    ReleaseDateUS    Budget
59                          The New Mutants                 20th Century Studios 28-08-2020 00:00  67000000
60                              Black Widow Walt Disney Studios Motion Pictures 09-07-2021 00:00 200000000
61 Shang-Chi and the Legend of the Ten Rings Walt Disney Studios Motion Pictures 03-09-2021 00:00 150000000
62               Venom: Let There Be Carnage                       Sony Pictures 01-10-2021 00:00 110000000
63                                  Eternals Walt Disney Studios Motion Pictures 05-11-2021 00:00 200000000
64                    Spider-Man: No Way Home                       Sony Pictures 17-12-2021 00:00 200000000
   OpeningWeekendNorthAmerica NorthAmerica OtherTerritories  Worldwide
59                    7037017     23855569         24819497   48675066
60                   80366312    183651665        195979696  379631351
61                   75388688    224543292        207700000  432243292
62                   90033210    213550366        288500000  502050366
63                   85021497    164870234        237194665  402064899
64                  260138569    780418859       1072000000 1852418859
> |
```

- **tail(marvel1) :**

```
> tail(marvel1)
                                    Film Rotten.Tomatoesin. Metacritic CinemaScore CinemaScore.1
62                          The New Mutants                36         43       <NA>          <NA>
63                              Black Widow                79         67          A            A-
64 Shang-Chi and the Legend of the Ten Rings                91         71          A             A
65               Venom: Let There Be Carnage                58         49          B            B+
66                                  Eternals                47         52          B             B
67                    Spider-Man: No Way Home                93         72          A            A+
> |
```

- **marvel$Title :**

```
> marvel$Title
 [1] "Howard the Duck"                          "Blade"
 [3] "X-Men"                                    "Blade II"
 [5] "Spider-Man"                               "Daredevil"
 [7] "X2"                                       "Hulk"
 [9] "The Punisher"                             "Spider-Man 2"
[11] "Blade: Trinity"                           "Elektra"
[13] "Fantastic Four"                           "X-Men: The Last Stand"
[15] "Ghost Rider"                              "Spider-Man 3"
[17] "Fantastic Four: Rise of the Silver Surfer" "Iron Man"
[19] "The Incredible Hulk"                      "Punisher: War Zone"
[21] "X-Men Origins: Wolverine"                 "Iron Man 2"
[23] "Thor"                                     "X-Men: First Class"
[25] "Captain America: The First Avenger"       "Ghost Rider: Spirit of Vengeance"
[27] "The Avengers"                             "The Amazing Spider-Man"
[29] "Iron Man 3"                               "The Wolverine"
[31] "Thor: The Dark World"                     "Captain America: The Winter Soldier"
[33] "The Amazing Spider-Man 2"                 "X-Men: Days of Future Past"
[35] "Guardians of the Galaxy"                  "Big Hero 6"
[37] "Avengers: Age of Ultron"                  "Ant-Man"
[39] "Fantastic Four"                           "Deadpool"
[41] "Captain America: Civil War"               "X-Men: Apocalypse"
[43] "Doctor Strange"                           "Logan"
[45] "Guardians of the Galaxy Vol. 2"           "Spider-Man: Homecoming"
[47] "Inhumans"                                 "Thor: Ragnarok"
[49] "Black Panther"                            "Avengers: Infinity War"
[51] "Deadpool 2"                               "Ant-Man and the Wasp"
[53] "Venom"                                    "Spider-Man: Into the Spider-Verse"
[55] "Captain Marvel"                           "Avengers: Endgame"
[57] "Dark Phoenix"                             "Spider-Man: Far From Home"
[59] "The New Mutants"                          "Black Widow"
[61] "Shang-Chi and the Legend of the Ten Rings" "Venom: Let There Be Carnage"
[63] "Eternals"                                 "Spider-Man: No Way Home"
> 
```

- **unique(marvel$Distributor) :**

```
> unique(marvel$Distributor)
[1] "Universal Pictures"                  "New Line Cinema"
[3] "20th Century Fox"                    "Sony Pictures"
[5] "Lionsgate Films"                     "Paramount Pictures"
[7] "Walt Disney Studios Motion Pictures" "IMAX Entertainment"
[9] "20th Century Studios"
> 
```

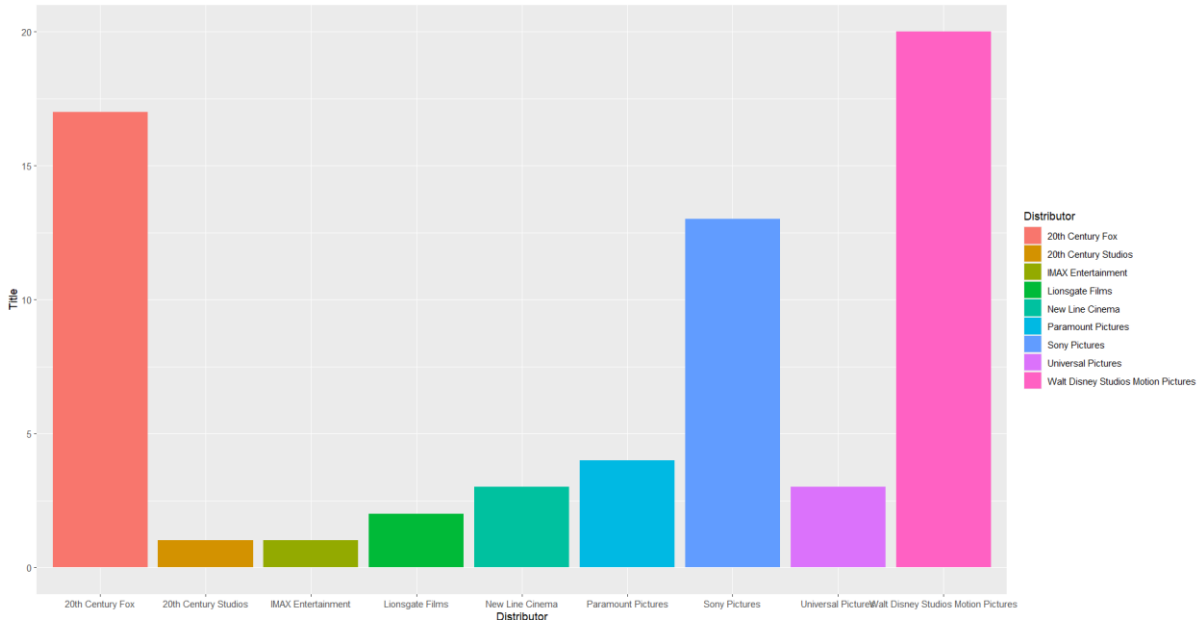
SOME QUERIES :

- Number of movies produced by different distributors :

X<-aggregate(Title~Distributor,mavel,length)

```
                         Distributor Title
1                    20th Century Fox    17
2                  20th Century Studios     1
3                  IMAX Entertainment     1
4                     Lionsgate Films     2
5                     New Line Cinema     3
6                   Paramount Pictures     4
7                        Sony Pictures    13
8                   Universal Pictures     3
9 Walt Disney Studios Motion Pictures    20
> 
```

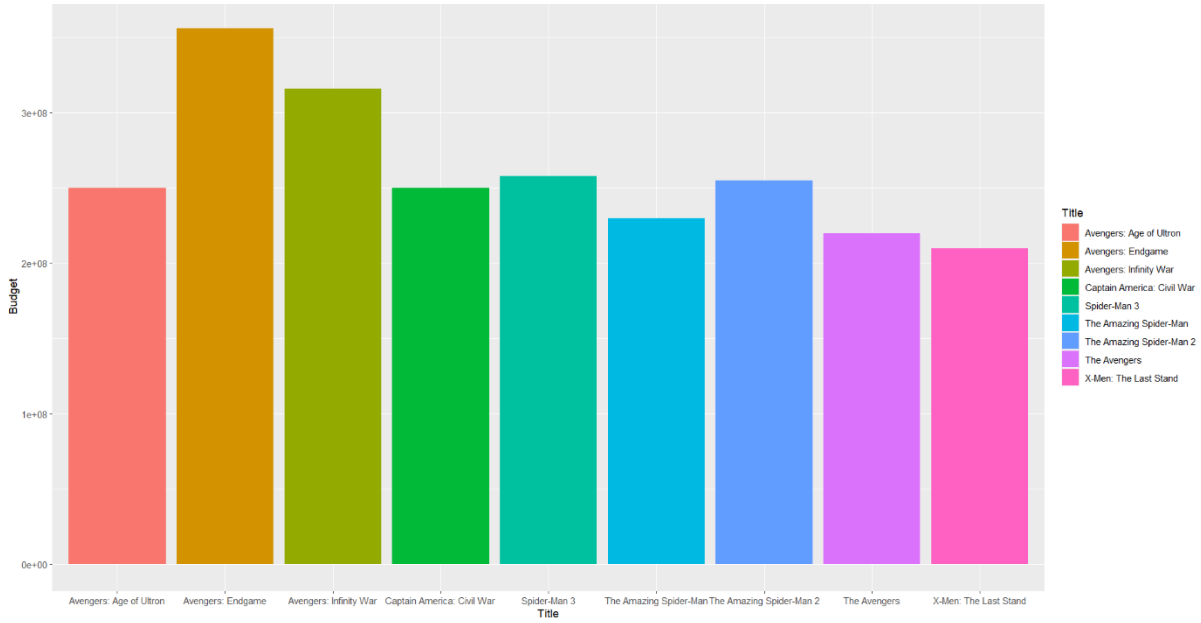
ggplot(a,aes(x=Distributor,y=Title,fill=Distributor))+geom_bar(stat="identity")



- Names and budgets of movies having more than 2.00e+08 budget :

```
> a1<-subset(marvel,Budget>2.00e+08,select = c(Title,Budget))
> a1
                        Title    Budget
14       X-Men: The Last Stand 210000000
16                Spider-Man 3 258000000
27                The Avengers 220000000
28      The Amazing Spider-Man 230000000
33    The Amazing Spider-Man 2 255000000
37     Avengers: Age of Ultron 250000000
41 Captain America: Civil War 250000000
50      Avengers: Infinity War 316000000
56           Avengers: Endgame 356000000
```

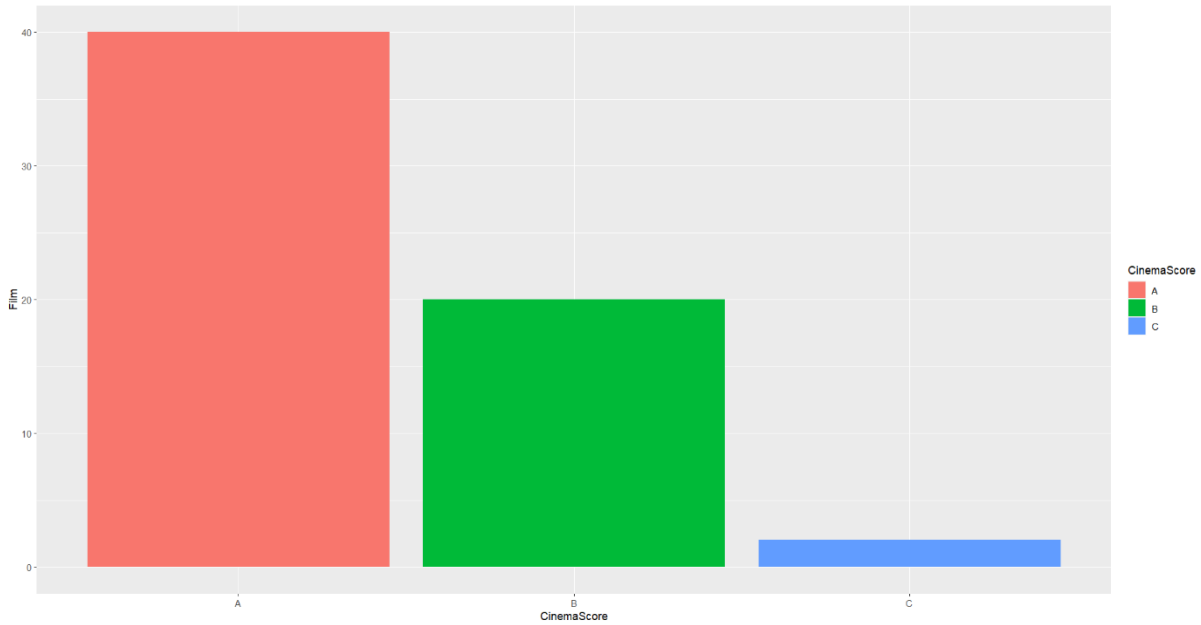
ggplot(a1,aes(x=Title,y=Budget,fill=Title))+geom_bar(stat="identity")



- Movies according to their CinemaScore :

```
> a2<-aggregate(Film~CinemaScore,marvel1,length)
> a2
  CinemaScore Film
1           A   40
2           B   20
3           C    2
```

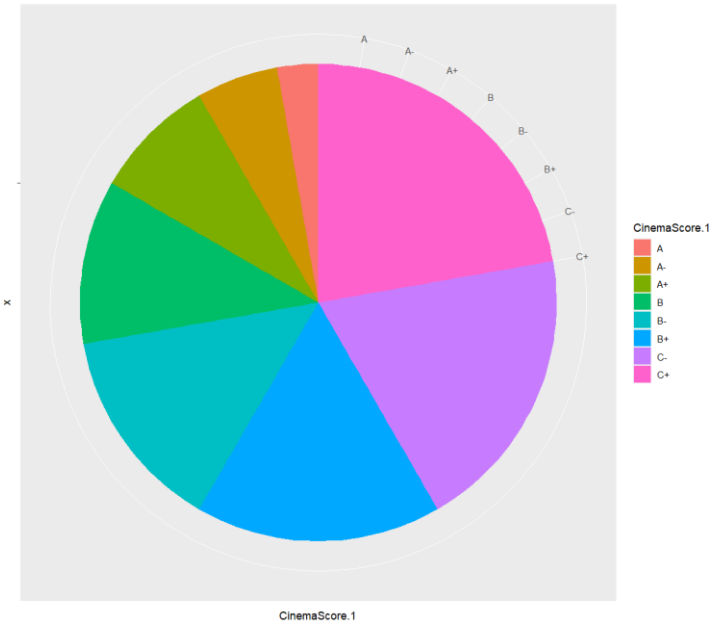
```
ggplot(a2,aes(x=CinemaScore,y=Film,fill=CinemaScore))+geom_bar(stat="identity")
```



- Movies according to their CinemaScore.1 :

```
> a3<-aggregate(Film~CinemaScore.1,marvel1,length)
> a3
  CinemaScore.1 Film
1             A   21
2            A-   14
3            A+    5
4             B    6
5            B-    6
6            B+    8
7            C-    1
8            C+    1
```

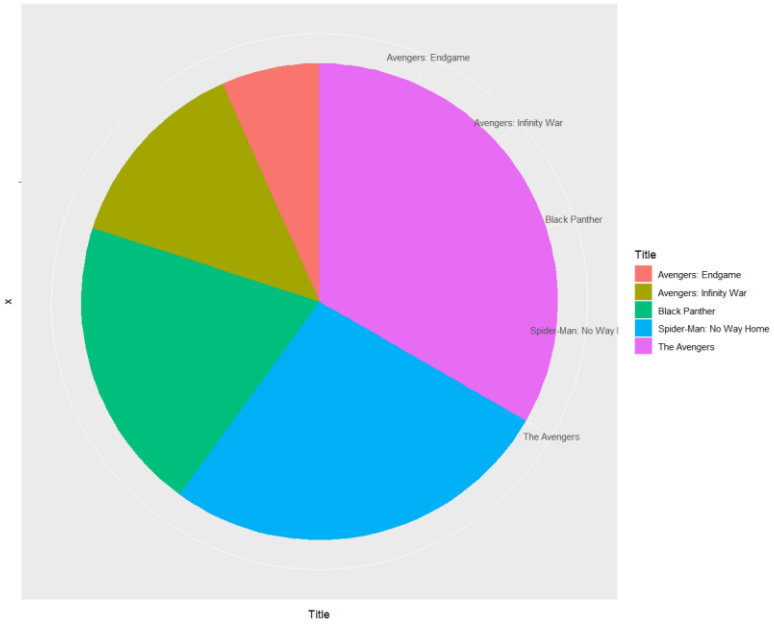
```
ggplot(a3,aes(x="",y=CinemaScore.1,fill=CinemaScore.1))+geom_bar(stat="identity")+coord_polar("
y")
```



- Movies where gross revenue > 500000000 in North America :

```
                   Title NorthAmerica
27           The Avengers    623357910
49          Black Panther    700059566
50  Avengers: Infinity War    678815482
56       Avengers: Endgame    858373000
64 Spider-Man: No Way Home    780418859
> ggplot(a4,aes(x="",y=Title,fill=Title))+geom_bar(stat="identity")+coord_polar("y")
>
```

```
ggplot(a4,aes(x="",y=Title,fill=Title))+geom_bar(stat="identity")+coord_polar("y")
```



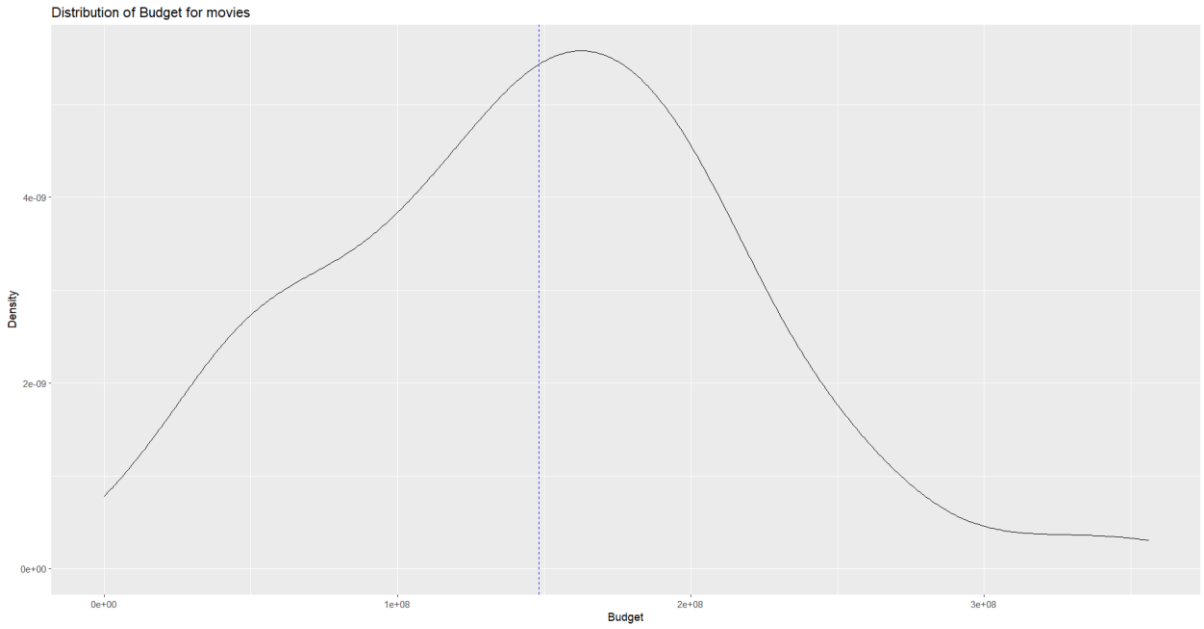
- Movies which are produced by “20th Century Fox” :

```
> a5<-marvel[which(marvel$Distributor=="20th Century Fox"),"Title"]
> a5
 [1] "X-Men"                                    "Daredevil"
 [3] "X2"                                       "Elektra"
 [5] "Fantastic Four"                           "X-Men: The Last Stand"
 [7] "Fantastic Four: Rise of the Silver Surfer" "X-Men Origins: Wolverine"
 [9] "X-Men: First Class"                       "The Wolverine"
[11] "X-Men: Days of Future Past"               "Fantastic Four"
[13] "Deadpool"                                 "X-Men: Apocalypse"
[15] "Logan"                                    "Deadpool 2"
[17] "Dark Phoenix"
> |
```

- What is the distribution of budget for movies :

```
ggplot(data=marvel, aes(x=Budget)) +
 geom_density(color="black") +
 geom_vline(xintercept = mean(marvel$Budget), color = "green", linetype = "dashed") +
 labs(title="Distribution of Budget for movies") +
 xlab(label="Budget") +
 ylab(label="Density")
```

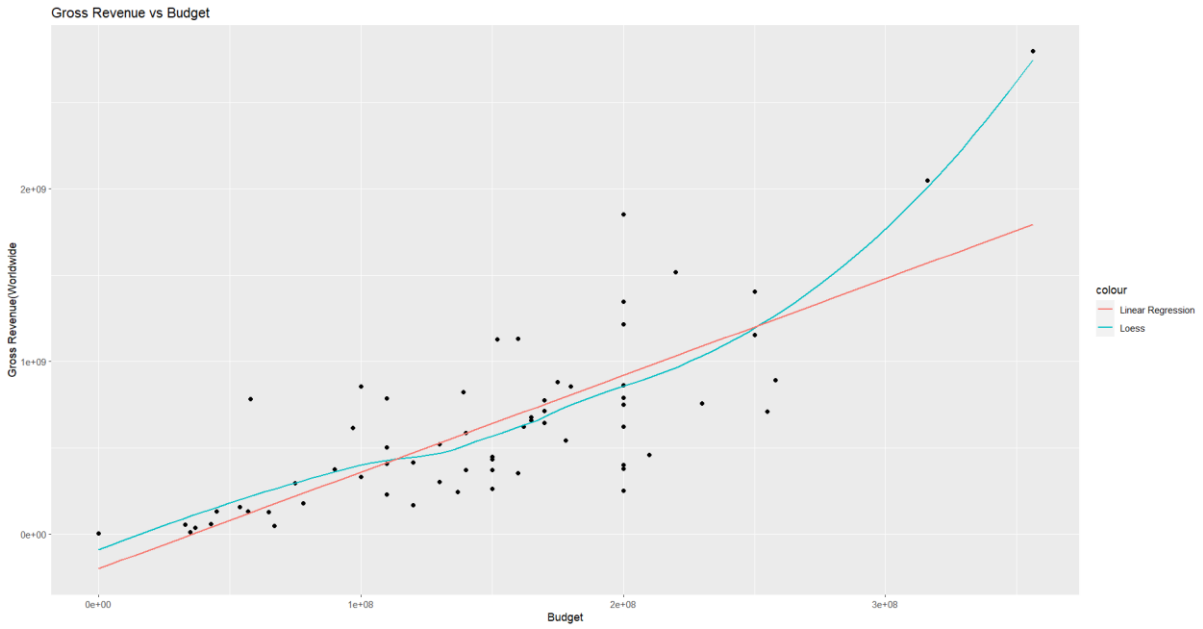
The blue dashed line represents the mean revenue and the black surve represents the density curve. The distribution appears to be positively skewed to the right. After the transformation, the distribution appears to be closer to a normal distribution.



- Relationship of Budget of the movie and its Gross Revenue(Worldwide) :

```
ggplot(marvel, aes(x=Budget, y=Worldwide)) +
 geom_point() +
```

```
geom_smooth(aes(colour="Loess"), method="loess",size=.6, se=F) +
geom_smooth(aes(colour="Linear Regression"), method="lm", se=F, size=.6) +
xlab(label="Budget") +
ylab(label="Gross Revenue(Worldwide)") +
labs(title="Gross Revenue vs Budget")
```



The summary of the simple linear regression with Budget as the predictor and Revenue as the response shows that Budget is a a significant factor that influences the revenue positively.

- Which movie has topped the revenue charts?
  - Its avengers: endgame

```
> marvel$Title[marvel$Worldwide==max(marvel$Worldwide)]
[1] "Avengers: Endgame"
>
```

- Which distributor has produced the most marvel movies?

➔ Walt disney

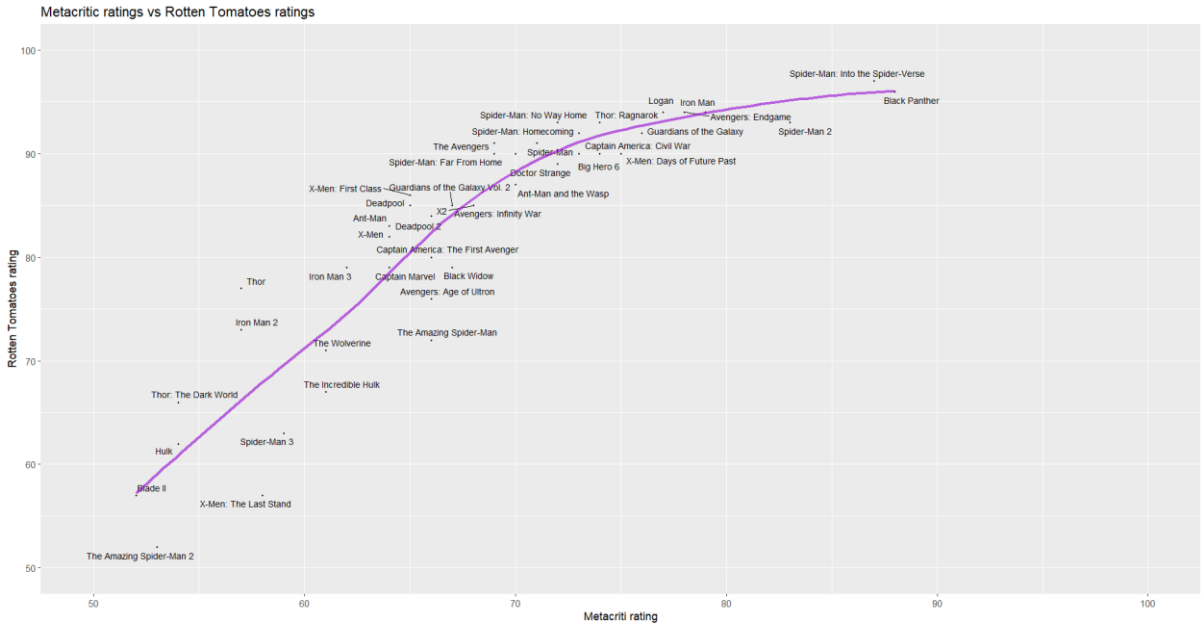
```
> table(marvel$Distributor)[table(marvel$Distributor)==max(tab
le(marvel$Distributor))]
Walt Disney Studios Motion Pictures
                                 20
```

- Relationship between Rotten Tomatoes rating and the Metacritic rating of the movies :

```
> marvel1[,c("Film","Rotten.Tomatoesin.","Metacritic")]
                                        Film Rotten.Tomatoesin. Metacritic
1                                Howard the Duck                14         28
2                            The Punisher (1989)                28         63
3                         Captain America (1990)                13          0
4                     The Fantastic Four (1994)                 30          0
5                                         Blade                 57         47
6                                         X-Men                 82         64
7                                      Blade II                 57         52
8                                    Spider-Man                 90         73
9                                     Daredevil                 44         42
10                                           X2                 85         68
11                                         Hulk                 62         54
12                           The Punisher (2004)                29         33
13                                  Spider-Man 2                93         83
14                               Blade: Trinity                 25         38
15                                      Elektra                 11         34
16                                    Man-Thing                 17          0
17                         Fantastic Four (2005)                28         40
18                         X-Men: The Last Stand                57         58
19                                   Ghost Rider                26         35
20                                  Spider-Man 3                63         59
21 Fantastic Four: Rise of the Silver Surfer                    37         45
22                                      Iron Man                94         79
23                           The Incredible Hulk                67         61
24                            Punisher: War Zone                29         30
25                     X-Men Origins: Wolverine                 37         40
26                                   Iron Man 2                 73         57
27                                         Thor                 77         57
28                           X-Men: First Class                 86         65
29          Captain America: The First Avenger                 80         66
30            Ghost Rider: Spirit of Vengeance                 18         34
31                               The Avengers                   91         69
32                         The Amazing Spider-Man               72         66
33                                  Iron Man 3                  79         62
34                               The Wolverine                  71         61
```

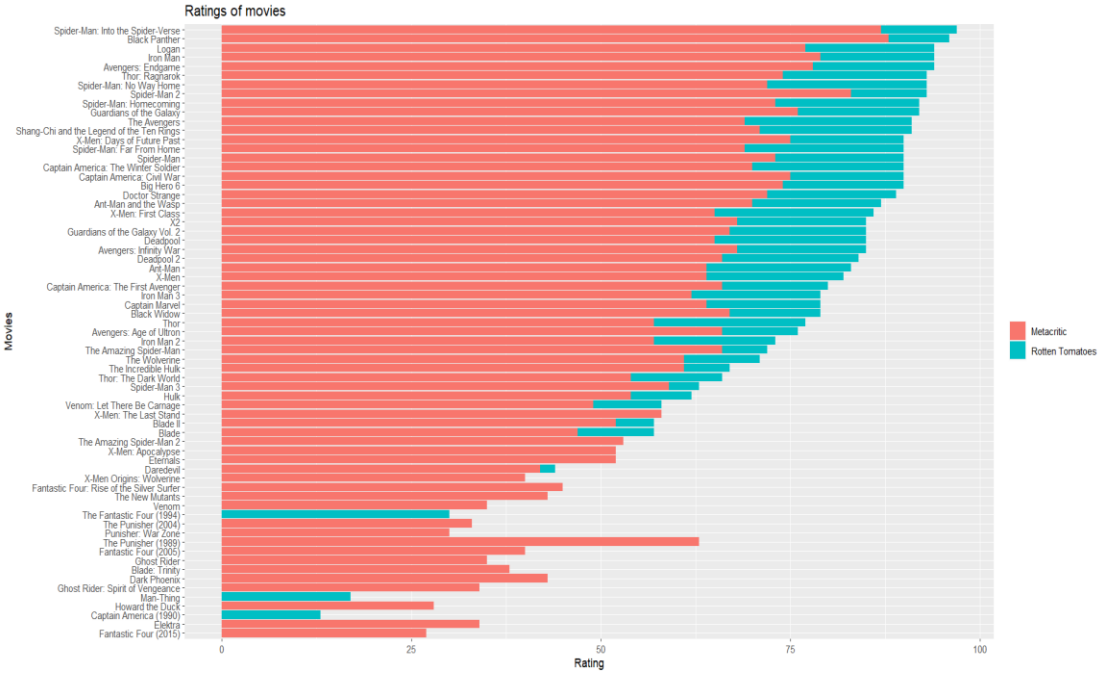
35	Thor: The Dark World	66	54
36	Captain America: The Winter Soldier	90	70
37	The Amazing Spider-Man 2	52	53
38	X-Men: Days of Future Past	90	75
39	Guardians of the Galaxy	92	76
40	Big Hero 6	90	74
41	Avengers: Age of Ultron	76	66
42	Ant-Man	83	64
43	Fantastic Four (2015)	9	27
44	Deadpool	85	65
45	Captain America: Civil War	90	75
46	X-Men: Apocalypse	47	52
47	Doctor Strange	89	72
48	Logan	94	77
49	Guardians of the Galaxy Vol. 2	85	67
50	Spider-Man: Homecoming	92	73
51	Thor: Ragnarok	93	74
52	Black Panther	96	88
53	Avengers: Infinity War	85	68
54	Deadpool 2	84	66
55	Ant-Man and the Wasp	87	70
56	Venom	30	35
57	Spider-Man: Into the Spider-Verse	97	87
58	Captain Marvel	79	64
59	Avengers: Endgame	94	78
60	Dark Phoenix	22	43
61	Spider-Man: Far From Home	90	69
62	The New Mutants	36	43
63	Black Widow	79	67
64	Shang-Chi and the Legend of the Ten Rings	91	71
65	Venom: Let There Be Carnage	58	49
66	Eternals	47	52
67	Spider-Man: No Way Home	93	72

```
ggplot(marvel1, aes(x=Metacritic, Rotten.Tomatoesin.)) +
 geom_point( size=.5) +
 geom_text_repel(label=marvel1$Film, size=3) +
 xlim(50,100) +
 ylim(50,100) +
 geom_line(colour="darkviolet", alpha=.6, method="loess", stat="smooth", size=1.3) +
 labs(title="Metacritic ratings vs Rotten Tomatoes ratings") +
 ylab(label="Rotten Tomatoes rating") +
 xlab(label="Metacriti rating")
```



They seem to have a positive relationship, however it is not proportional. Metacritic appears to give lower ratings than Rotten Tomatoes.

```
ggplot(marvel1) +
  geom_bar(aes(x=reorder(Film, Rotten.Tomatoesin.), y=Rotten.Tomatoesin., fill="Rotten
Tomatoes"), stat="identity", alpha=1) +
  geom_bar(aes(x=reorder(Film, Metacritic), y=Metacritic, fill="Metacritic"),
stat="identity",alpha=1) +
  coord_flip() +
  scale_fill_discrete("") +
  labs(title="Ratings of movies") +
  ylab("Rating") +
  xlab("Movies")
```



How have the ratings evolved over time?

In general, Rotten Tomatoes reviewers appears to give more generous ratings than Metacritic reviewers. The graphs suggests that the ratings are increasing the but have fallen a bit since 2018.

Conclusion :

1. Budget appears to be somewhat constant while Gross is on an upward trend.
2. Rotten tomatoes tends to give higher ratings than Metacritic.
3. The highest by Rotten Tomatoes and Metacritic is Spiderman-Into the spider verse followed by Black Panther.
4. The movie that brought in the most revenue was Avengers: Endgame in 2019.
5. Budget is a significant predictor of Revenue.