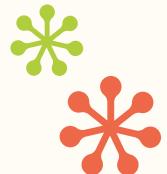




Marvel & DC Analysis

By: Susan Lin
STA 3000





Dataset

- ❖ Marvel VS DC
- ❖ X= Budget
- ❖ Y1= Rate, Y2= Gross Worldwide

	A	B	C	D	E	F	G	H	I	J	K	L
1		Original Title	Company	Rate	Metascore	Minutes	Release	Budget	Opening V	Gross USA	Gross	Worldwide
2	1	Iron Man	Marvel	7.9	79	126	2008	1.4E+08	98618668	3.19E+08	5.85E+08	
3	2	The Incredibles	Marvel	6.7	61	112	2008	1.5E+08	55414050	1.35E+08	2.63E+08	
4	3	Iron Man 2	Marvel	7	57	124	2010	2E+08	1.28E+08	3.12E+08	6.24E+08	
5	4	Thor	Marvel	7	57	115	2011	150000000	65723338	1.81E+08	4.49E+08	
6	5	Captain America: The First Avenger	Marvel	6.9	66	124	2011	1.4E+08	65058524	1.77E+08	3.71E+08	
7	6	The Avengers	Marvel	8	69	143	2012	2.2E+08	2.07E+08	6.23E+08	1.52E+09	
8	7	Iron Man 3	Marvel	7.2	62	130	2013	200000000	1.74E+08	4.09E+08	1.21E+09	
9	8	Thor: The Dark World	Marvel	6.9	54	112	2013	1.7E+08	85737841	2.06E+08	6.45E+08	
10	9	Captain America: Civil War	Marvel	7.7	70	136	2014	1.7E+08	95023721	2.6E+08	7.14E+08	
11	10	Guardians of the Galaxy	Marvel	8	76	121	2014	1.7E+08	94320883	3.33E+08	7.73E+08	
12	11	Avengers: Age of Ultron	Marvel	7.3	66	141	2015	2.5E+08	1.91E+08	4.59E+08	1.4E+09	
13	12	Ant-Man	Marvel	7.3	64	117	2015	1.3E+08	57225526	1.8E+08	5.19E+08	
14	13	Captain America: Civil War	Marvel	7.8	75	147	2016	2.5E+08	1.79E+08	4.08E+08	1.15E+09	
15	14	Doctor Strange	Marvel	7.5	72	115	2016	1.65E+08	85058311	2.33E+08	6.78E+08	
16	15	Guardians of the Galaxy Vol. 2	Marvel	7.6	67	136	2017	2E+08	1.47E+08	3.9E+08	8.64E+08	
17	16	Spider-Man: Homecoming	Marvel	7.4	73	133	2017	175000000	1.17E+08	3.34E+08	8.8E+08	
18	17	Thor: Ragnarok	Marvel	7.9	74	130	2017	1.8E+08	1.23E+08	3.15E+08	8.54E+08	

Marvel VS DC

```
db.sorted <- db[order(db$Rate,decreasing=TRUE),]
head(db.sorted, 10)
```

```
##          Original.Title Company Rate   Budget Gross.Worldwide ##
## 27      The Dark Knight      DC  9.0 1.85e+08    1004934033 ##
## 39          Joker           DC  8.7 5.50e+07    1060504580 ##
## 19  Avengers: Infinity War  Marvel 8.5 3.21e+08    2048359754 ##
## 22      Avengers: Endgame  Marvel 8.5 3.56e+08    2797800564 ##
## 31  The Dark Knight Rises      DC  8.4 2.50e+08    1081041287 ##
## 25      Batman Begins       DC  8.2 1.50e+08    373413297 ##
## 6       The Avengers        Marvel 8.0 2.20e+08    1518812988 ##
## 10  Guardians of the Galaxy  Marvel 8.0 1.70e+08    772776600 ##
## 1       Iron Man           Marvel 7.9 1.40e+08    585366247 ##
## 17  Thor:Ragnarok         Marvel 7.9 1.80e+08    853977126 ##
```

```
db.sorted <- db[order(db$Gross.WORLDWIDE,decreasing=TRUE),]
head(db.sorted, 10)
```

```
##          Original.Title Company Rate   Budget Gross.Worldwide ##
## 22      Avengers: Endgame  Marvel 8.5 3.56e+08    2797800564 ##
## 19  Avengers: Infinity War  Marvel 8.5 3.21e+08    2048359754 ##
## 6       The Avengers        Marvel 8.0 2.20e+08    1518812988 ##
## 11  Avengers: Age of Ultron  Marvel 7.3 2.50e+08    1402805868 ##
## 18          Black Panther     Marvel 7.3 2.00e+08    1346913161 ##
## 7       Iron Man Three      Marvel 7.2 2.00e+08    1214811252 ##
## 13  Captain America: Civil War  Marvel 7.8 2.50e+08    1153296293 ##
## 37          Aquaman            DC  7.0 1.60e+08    1148161807 ##
## 23  Spider-Man: Far from Home  Marvel 7.6 1.60e+08    1131927996 ##
## 21      Captain Marve       Marvel 6.9 1.75e+08    1128274794 ##
```

The Dark Knight (DC)
have the highest rating

```
db.sorted <- db[order(db$Budget,decreasing=TRUE),]
head(db.sorted, 10)
```

	Original.Title <chr>	Compa... <chr>	Rate <dbl>	Budget <dbl>	Gross.Worldwide <dbl>
22	Avengers: Endgame	Marvel	8.5	3.56e+08	2797800564
19	Avengers: Infinity War	Marvel	8.5	3.21e+08	2048359754
36	Justice League	DC	6.4	3.00e+08	657924295
26	Superman Returns	DC	6.0	2.70e+08	391081192
11	Avengers: Age of Ultron	Marvel	7.3	2.50e+08	1402805868
13	Captain America: Civil War	Marvel	7.8	2.50e+08	1153296293
31	The Dark Knight Rises	DC	8.4	2.50e+08	1081041287
33	Batman v Superman: Dawn of Justice	DC	6.5	2.50e+08	873634919
32	Man of Steel	DC	7.1	2.25e+08	668045518
6	The Avengers	Marvel	8.0	2.20e+08	1518812988

Avengers: Endgame
(Marvel) have the
highest gross
worldwide and budget



Continue...

```
tail(db, 16)  
DC <- tail(db, 16)  
summary(DC)
```

Rate	Budget
Min. :3.300	Min. : 47000000
1st Qu.:6.000	1st Qu.:122500000
Median :7.050	Median :167500000
Mean :6.806	Mean :171625000
3rd Qu.:7.750	3rd Qu.:231250000
Max. :9.000	Max. :300000000

Gross.Worldwide
Min. :1.090e+07
1st Qu.:3.284e+08
Median :6.630e+08
Mean :6.056e+08
3rd Qu.:9.065e+08
Max. :1.148e+09

```
summary(db)
```

Rate	Budget
Min. :3.300	Min. : 47000000
1st Qu.:6.900	1st Qu.:150000000
Median :7.300	Median :175000000
Mean :7.203	Mean :184102564
3rd Qu.:7.850	3rd Qu.:210000000
Max. :9.000	Max. :356000000

```
## Gross.Worldwide  
## Min. :1.090e+07  
## 1st Qu.:4.843e+08  
## Median :7.468e+08  
## Mean :8.276e+08  
## 3rd Qu.:1.105e+09  
## Max. :2.798e+09
```

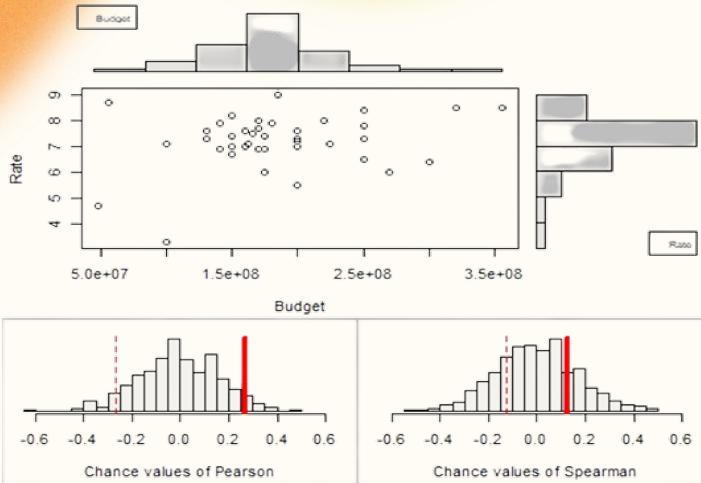
```
head(db, 23)  
Marvel <- head(db, 23)  
summary(Marvel)
```

Rate	Budget
Min. :6.700	Min. :130000000
1st Qu.:7.050	1st Qu.:161000000
Median :7.400	Median :175000000
Mean :7.478	Mean :192782609
3rd Qu.:7.850	3rd Qu.:200000000
Max. :8.500	Max. :356000000

```
## Gross.Worldwide  
## Min. :2.634e+08  
## 1st Qu.:6.233e+08  
## Median :8.540e+08  
## Mean :9.820e+08  
## 3rd Qu.:1.184e+09  
## Max. :2.798e+09
```

Association

`associate(Rate~Budget, data= db)`



Association between Budget (numerical) and Rate (numerical)
using 39 complete cases
Permutation procedure:



	Value	Estimated	p-value
Pearson's r	0.2656554	0.09	
Spearman's rank correlation	0.1238763	0.46	

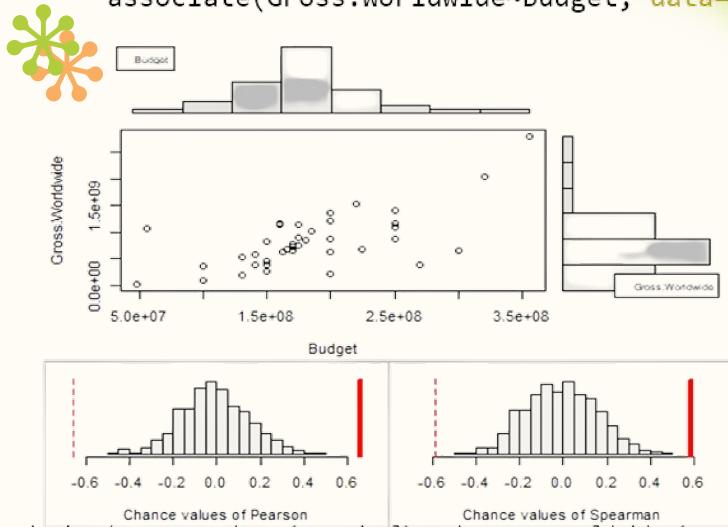
With 500 permutations, we are 95% confident that:

the p-value of Pearson's correlation (r) is between 0.066 and 0.119
the p-value of Spearman's rank correlation is between 0.416 and 0.505

Note: If 0.05 is in this range, increase the permutations= argument.

$$R-sq = 0.26^2 = \sim 7\%$$

`associate(Gross.Worldwide~Budget, data= db)`



Association between Budget (numerical) and Gross.Worldwide (numerical)
using 39 complete cases
Permutation procedure:

	Value	Estimated	p-value
Pearson's r	0.6563322	0	
Spearman's rank correlation	0.5858797	0	

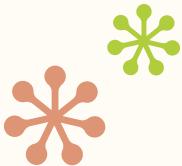
With 500 permutations, we are 95% confident that:

the p-value of Pearson's correlation (r) is between 0 and 0.007
the p-value of Spearman's rank correlation is between 0 and 0.007

Note: If 0.05 is in this range, increase the permutations= argument.

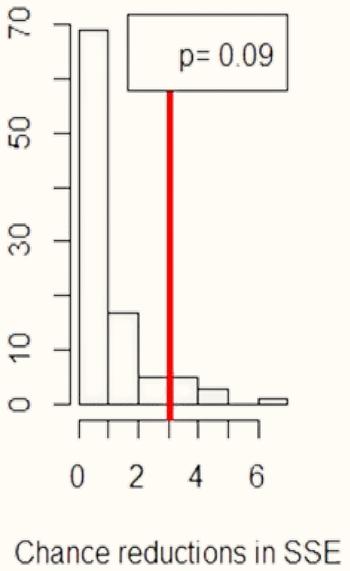
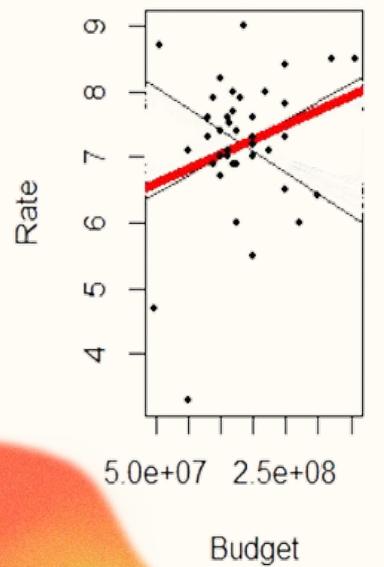
$$R-sq = 0.66^2 = \sim 44\%$$

Linear Regression



Budget Vs Rate

```
M <- lm(Rate~Budget, data= db)  
possible_regressions(M)
```



Call:
lm(formula = Rate ~ Budget, data = db)

Coefficients:

(Intercept)	Budget
6.387e+00	4.433e-09

$$y = 4.433e-09x + 6.387$$

summary(M)

Residuals:

Min	1Q	Median	3Q	Max
-3.5298	-0.2676	0.2378	0.5986	2.0697

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.387e+00	5.151e-01	12.399	9.61e-15 ***
Budget	4.433e-09	2.644e-09	1.676	0.102

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.05 on 37 degrees of freedom
Multiple R-squared: 0.07057, Adjusted R-squared: 0.04545
F-statistic: 2.809 on 1 and 37 DF, p-value: 0.1021

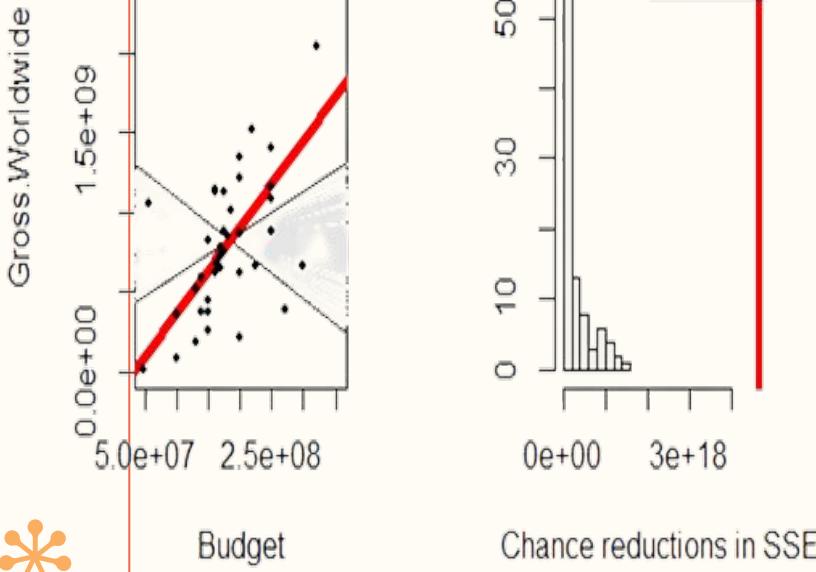
confint(M, level=0.95)

	2.5 %	97.5 %
## (Intercept)	5.342885e+00	7.430155e+00
## Budget	-9.256954e-10	9.790796e-09



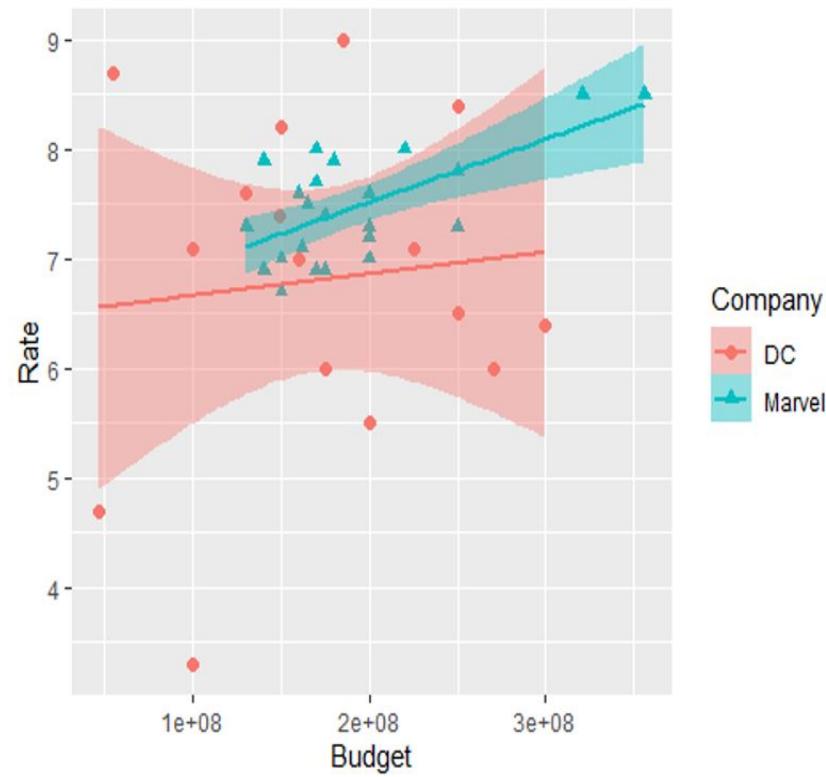
Budget Vs Gross Worldwide

```
B <- lm(Gross.WORLDWIDE~Budget, data= db)  
possible_regressions(B)
```

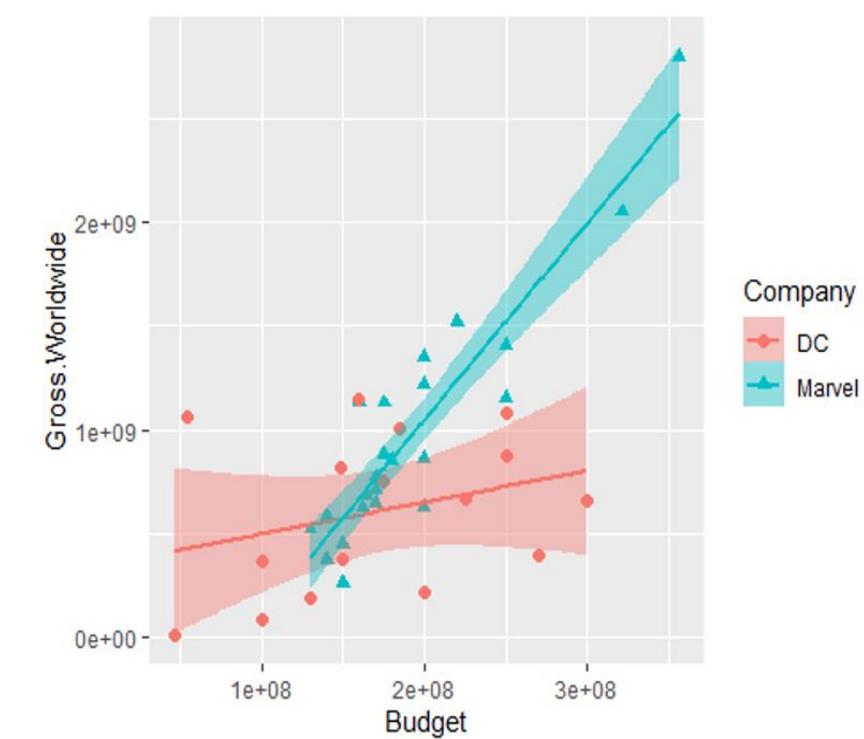


```
Call:  
lm(formula = Gross.WORLDWIDE ~ Budget, data = db)  
  
Coefficients:  
(Intercept) Budget  
-1.728e+08 5.434e+00  
  
 $y = 5.34x - 1.728e+08$   
  
summary(B)  
Residuals:  
Min 1Q Median 3Q Max  
-903221102 -243114673 -32334628 201090738 1036211362  
  
Coefficients:  
Estimate Std. Error t value Pr(>|t|)  
(Intercept) -1.728e+08 2.000e+08 -0.864 0.393  
Budget 5.434e+00 1.027e+00 5.292 5.71e-06 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 407700000 on 37 degrees of freedom  
Multiple R-squared: 0.4308, Adjusted R-squared: 0.4154  
F-statistic: 28 on 1 and 37 DF, p-value: 5.713e-06  
  
confint(B, level=0.95)  
## 2.5 % 97.5 %  
## (Intercept) -5.779995e+08 2.324770e+08  
## Budget 3.352989e+00 7.514148e+00
```

```
ggplot(db, aes(x= Budget, y= Rate, color= Company, pch= Company)) +  
  geom_point(size=2) +  
  geom_smooth(method="lm", aes(fill=Company))
```

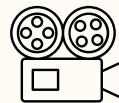


```
ggplot(db, aes(x= Budget, y= Gross.WORLDWIDE, color= Company, pch= Company)) +  
  geom_point(size=2) +  
  geom_smooth(method="lm", aes(fill=Company))
```

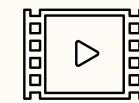


In conclusion

Marvel VS DC

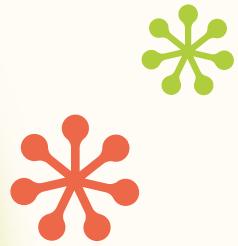


~~Budget VS
Rate~~



Budget VS
Gross
Worldwide





Thanks!

Resource:

[Marvel vs DC | Kaggle](#)