

Mengting Guo
Yuxiao Ning
Haig Shishmanian
Ma 331 – Final Project

Investing in Major Film Productions

Analyzing Profitability Determined by Film Characteristics

Abstract:

The film industry continually produces very successful box-office hits. Where there is profit, there is investment opportunity. A study was done for the purpose of advising potential investors in what films to become financiers of. Specific aspects of films were chosen as factors potentially contributing to profitability. Recent films from the past 5 years were sorted based on 3 factors: The source of the story, the genre of the film, and the MPAA rating given to the film.

First, the analysis showed that there is no direct relationship between budget size and profitability. Next, two very profitable combinations were determined: PG-rated Adventure Sequels and PG-rated Dramas based on original stories. Lastly, G-rated films were analyzed separately because of their being rare and consistently profitable.

Introduction:

The multi-billion dollar film industry is apparently very successful. Why? Major films became an integral part of the international community in the 20th century and, currently, the population is increasingly hungry for entertainment. A high stake is put on ideas that will lead people to buy tickets. What contributes to the profitability of films released in theatres? What details of a film's production contribute (consciously or subconsciously) to a person's paying money for increasingly overpriced tickets? Which of these characteristics can be known during the time of production so that people can choose to invest in a film?

The group chose to analyze the effect of three explanatory variables on profitability. Profitability was defined as the International Gross of Ticket Sales – Film Budget. The variables were chosen out of a plethora of distinct characteristics that films have. Some of the variables that we rejected are the films distributor, the location of its filming, and the director. Throughout the analysis, the three variables chosen differed in their validity to the study. For instance, the genre of films proved to have a greater effect on non-G-rated films. G-rated films were analyzed separately and were observed to not really depend on genre for profitability.

The first variable chosen was the genre. The genre was considered as an obvious determinant of the popularity of a film because certain genres seemed to be consistently profitable just by observing the data. The most popular genres of major film productions in the past 5 years have been comedy, action, adventure, and drama.

Second was the MPAA rating because of the obvious way that it determines the film's audience as well as the type of content that is expected in the film. The MPAA is the official organization that decides what ages are suited for viewing the content of a film. It gives four different ratings: G (For all audiences), PG (For most audiences, suggesting parental guidance for children), PG-13 (For viewers aged 13 and over), and R (Restricted – admitting only those viewers aged 17 or older). The MPAA rating of a film can be assumed before production because of the very standardized rating methods of the organization. Up to a certain amount of violence and sex will be tolerated for a PG-13, and after a very standardized amount of such inappropriate material, a film will be rated R. In the same way, a G-rated film is very clean and

appropriate for even toddlers, while a PG-rated film will have some darkness, negativity, unpleasantness, or angst.

The third variable was the source. The source of a film is what the idea for the storyline has come from. In most cases, the source is either an original screenplay, a sequel to a famous movie, or it is based on a book, television program or real-life event. It was found that this is an important aspect of a movie and only considered whether the films were original, sequels, or based on something. Original films tend to be most original and creative in plot. Sequels are often very popular because nobody makes a sequel of an unsuccessful movie (and people are always happy to see more enjoyable material). Films based on stories or television programs interest those people who follow such forms of entertainment.

In order to be able to offer valuable advice to investors, 3 questions are raised by the study and are answered separately:

1. Is there a direct correlation between budget and profitability?
2. What combination of factors is most profitable when investing in a movie?
3. Is there a reason for G-rated movies' consistent high profitability?

Methodology:

To answer the first question, two linear regression models and a plot are performed to see the relationship between the budget and the profit of the movie as well as the relationship between the budget and the profit-percentage.

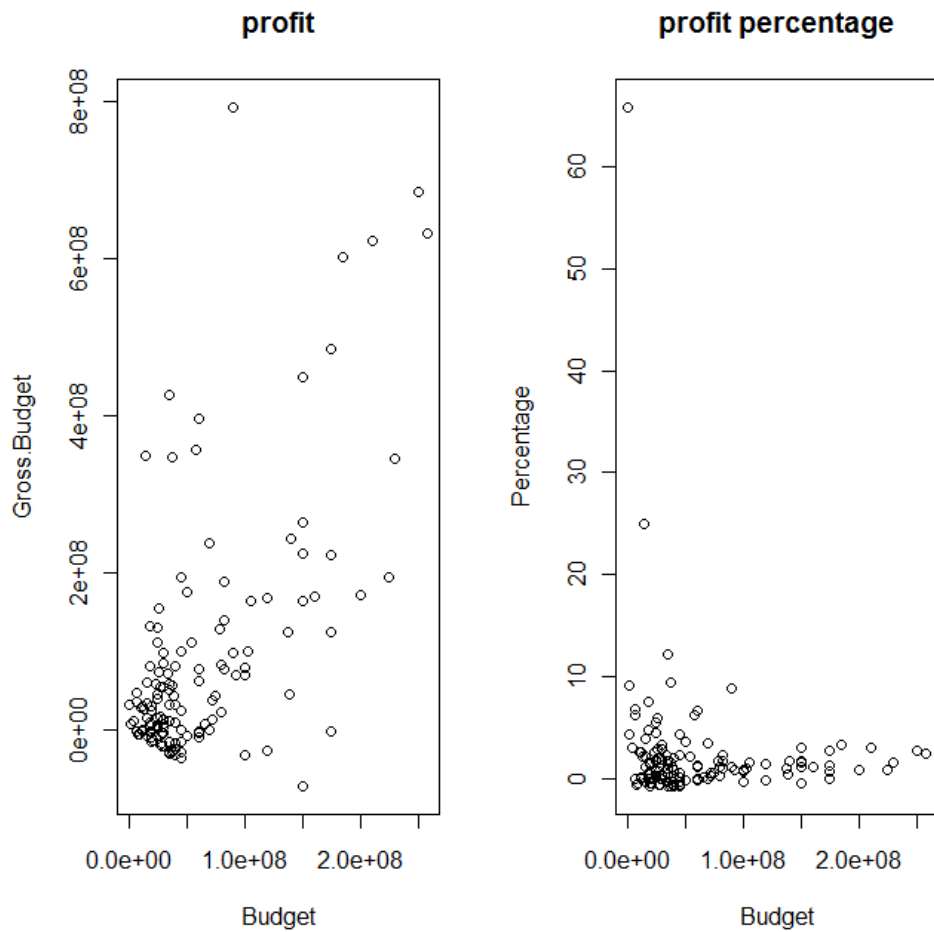
Then, the second question, which is the main question in the analysis, could be answered. In order to answer this question, two-way ANOVA, three-way ANOVA and Interaction plot were used to pick out the features for a highly-profitable movie.

For the G-rated movies, Interaction plot and one way ANOVA were included to help the authors to analyze the result.

Results:

1. In this part, the two relationships between profit vs. budget. Gross and profit.percentage vs. budget were obtained separately by applying the plot and the linear regression model.

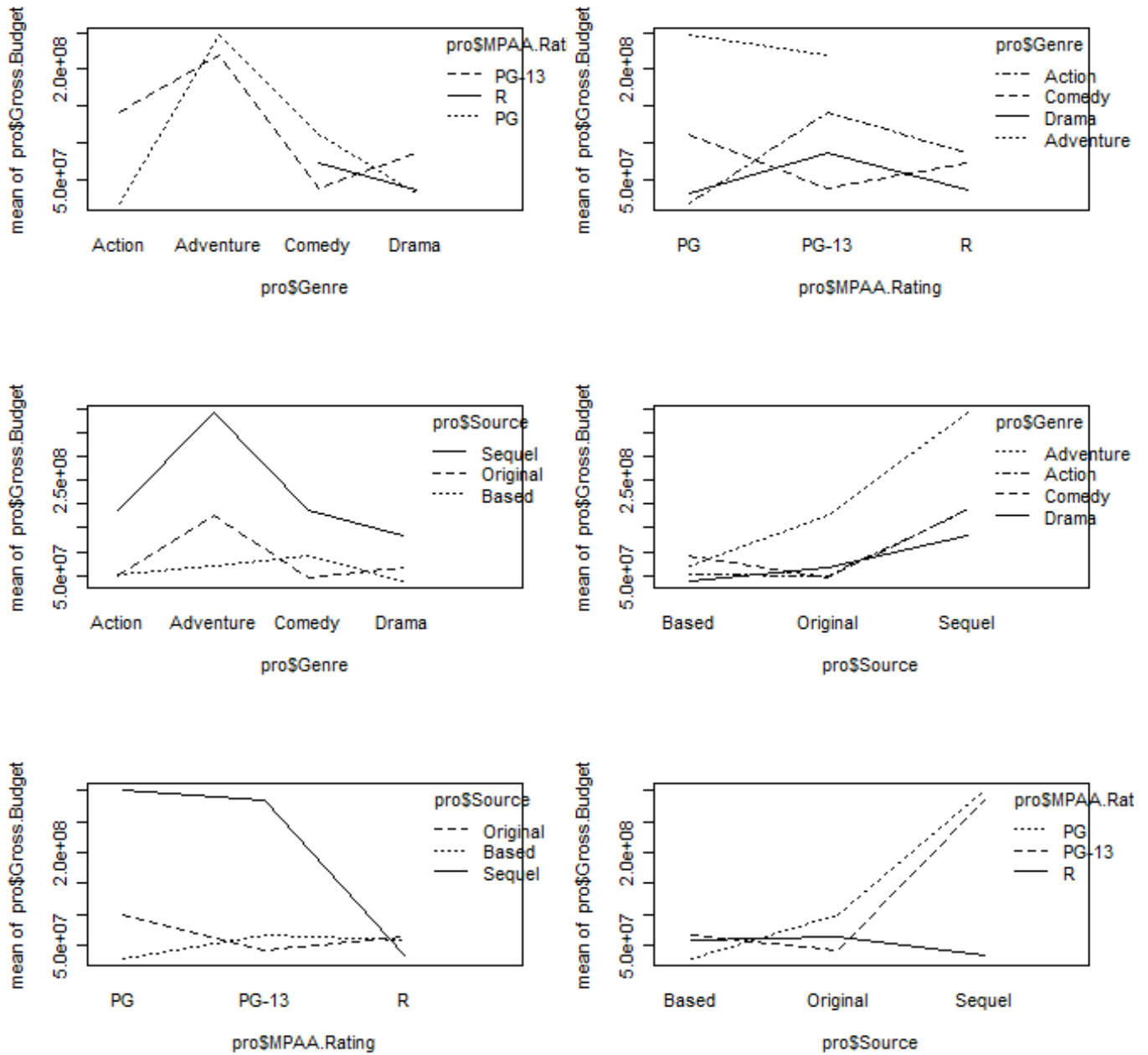
Plots:



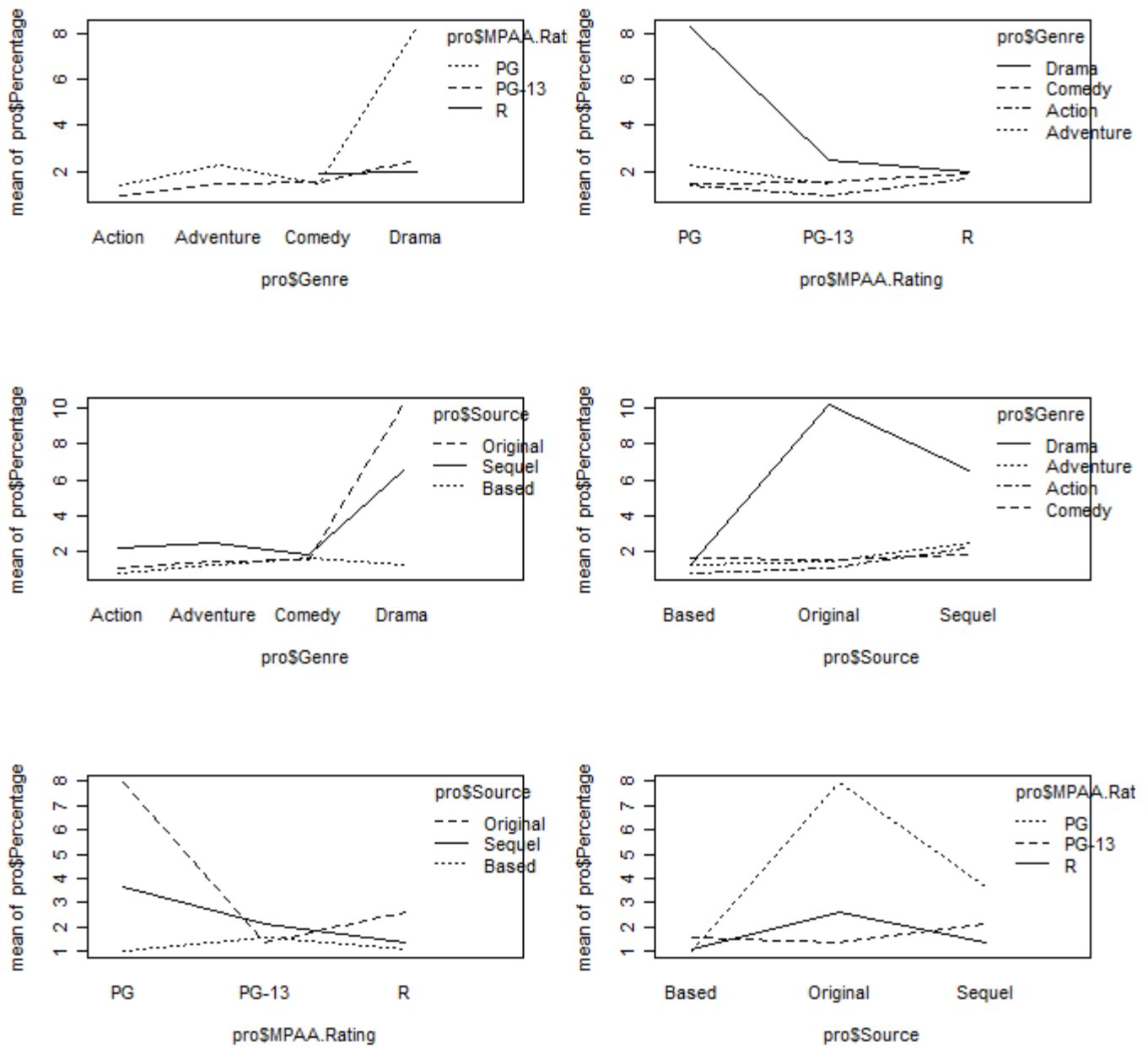
After the linear regression model, the R-squared was obtained for the two of them: 0.3686 and 0.01385 respectively.

2. In the second part, 2-way interaction plots were performed for each of the 6 combinations. (Genre vs. Mpaa rating, the reverse, Genre vs. Source, the reverse, and Source and Mpaa rating, the reverse.) Then 2-way ANOVA was performed and the best combination was selected. The result was put into a 3-way ANOVA by profit and by percentage.

Interaction plots (profit interactions):



(Percentage interactions):



Significant results for two-way ANOVA:

By profit:

GenreAdventure:SourceSequel	414882963	77538863	5.351	4.76e-07	***
MPAA.RatingR:SourceSequel	-302240616	111250232	-2.717	0.00765	**

By percentage:

GenreAction:SourceOriginal	13.0861	3.7369	3.502	0.000667	***
GenreComedy:SourceOriginal	14.9207	3.5537	4.199	5.44e-05	***
GenreDrama:SourceOriginal	25.9063	3.4837	7.436	2.32e-11	***
MPAA.RatingPG-13:SourceOriginal	-17.5007	3.4794	-5.030	1.90e-06	***
MPAA.RatingR:SourceOriginal	-18.1290	3.5868	-5.054	1.72e-06	***
GenreComedy:SourceSequel	12.3671	4.4170	2.800	0.006031	**
MPAA.RatingR:SourceSequel	-12.9904	4.7980	-2.707	0.007852	**

Significant results for three-way ANOVA:

By profit:

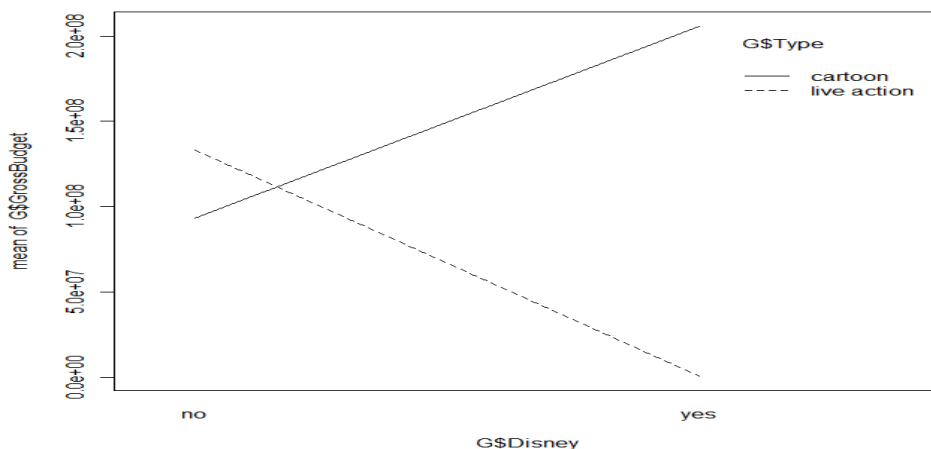
GenreAdventure:MPAA.RatingPG:SourceSequel	526295892	144952668	3.631	0.000442	***
---	-----------	-----------	-------	----------	-----

By percentage:

GenreDrama:MPAA.RatingPG:SourceOriginal	63.3003	4.4272	14.298	<2e-16	***
---	---------	--------	--------	--------	-----

3.

Interaction plot.



The results for ANOVA strengthened the interaction plot and confirmed the results.

Conclusion:

Based on the results received from analysis, it was possible to answer the questions raised in the beginning of the study.

From results 1, it can be concluded that there seems to be no direct relationship between the budget and the profitability. It is true that there seems to be a weak linear relationship between the budget and the profit (R squared equals to 0.3686). However, after analyzing the relationship between budget and profit percentage, the results clearly suggest that there seems to have no direct relationship between these two. (R squared equals to 0.01385)

The main part of this analysis can be explained by part two of the results which involved one, two, and three-way ANOVA with genre, MPAA rating, and source as the explanatory variables, being compared to the response variable: Profit Percentage and Profit. It was concluded that the best combination for the most profit is the combination of Adventure, PG and Sequel and the best for the highest profit percentage is the combination of Drama, PG and Original screen play.

Finally, it was concluded that G-rated movies are very profitable. This was seen as a product of the overwhelming presence of Disney as a producer, distributor, or studio (or any combination of these) of G-rated movies and based on the data we gathered, almost all the movies are somewhat “related” to Disney. The 3rd part of the results strongly supported this conclusion.

Limitation and Discussion:

Due to the resource of the data, only the movies with available budgets were selected and thus the result concluded might be different than the behavior of all the movies as a whole. Also, because of the design of the film database, it is very hard to export data and that caused a limitation on the amount of resource the paper included. These two limitations would be eliminated if better source and design of the database were provided.

In the future, computer science scholars and other people who are interested can further expand the ideas mentioned in this paper into a program. The program could use the same

methods to help the investors to make the best decision on investing based on the money they are willing to put in, the amount of movies they want to invest and their special interests.

Appendix:

Code:

```
pro = read.csv("D:\\Study\\MA331\\project\\NewDataFinal.csv", header=TRUE)
pro
1.##the relationship between (Budget), (Gross.Budget), and (Profit.Percentage)
1.1
## we can check if a variable is a factor or not using:
is.factor(pro$Genre)
is.factor(pro$MPAA.Rating)
is.factor(pro$Source)
is.factor(pro$Budget)
is.factor(pro$Gross.Budget)
is.factor(pro$Profit.Percentage)
1.2
##explain why we should focus on both profit and profit.percentage
par(mfrow=c(1,2))
boxplot(Budget)
plot(Gross.Budget~Budget, data = pro, main="profit")
plot(Percentage~Budget, data = pro, main="profit percentage")
linear = lm(Gross.Budget~Budget, data = pro)
linear2 = lm(Percentage~Budget, data = pro)
summary(linear)
summary(linear2)
2.##seperate to Gross.Budget and Profit.Percentage
2.1##focus on Profit(Gross.Budget)
2.1.1
##do interactions between the two factors in Profit
par(mfrow=c(3,2))
interaction.plot(pro$Genre,pro$MPAA.Rating,pro$Gross.Budget)
interaction.plot(pro$MPAA.Rating,pro$Genre,pro$Gross.Budget)
interaction.plot(pro$Genre,pro$Source,pro$Gross.Budget)
interaction.plot(pro$Source,pro$Genre,pro$Gross.Budget)
interaction.plot(pro$MPAA.Rating,pro$Source,pro$Gross.Budget)
interaction.plot(pro$Source,pro$MPAA.Rating,pro$Gross.Budget)
##to see the combination of two factors that produce a high profit in investment
movie.profit2=lm(Gross.Budget~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro)
anova(lm(Gross.Budget~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro))
summary(movie.profit2)
2.1.2
##to see the combination of three factors that produce a high profit in investment
movie.profit3=lm(Gross.Budget~Genre:MPAA.Rating:Source,data=pro)
anova(lm(Gross.Budget~Genre:MPAA.Rating:Source,data=pro))
summary(movie.profit3)
2.2##focus on (Profit.Percentage)
2.2.2
##do interactions between the two factors in Percentage
par(mfrow=c(3,2))
interaction.plot(pro$Genre,pro$MPAA.Rating,pro$Percentage)
interaction.plot(pro$MPAA.Rating,pro$Genre,pro$Percentage)
interaction.plot(pro$Genre,pro$Source,pro$Percentage)
interaction.plot(pro$Source,pro$Genre,pro$Percentage)
interaction.plot(pro$MPAA.Rating,pro$Source,pro$Percentage)
interaction.plot(pro$Source,pro$MPAA.Rating,pro$Percentage)
##to see the combination of two factors that produce a high Percentage in investment
movie.percentage2=lm(Percentage~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro)
anova(lm(Percentage~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro))
summary(movie.percentage2)
2.2.2
##to see the combination of three factors that produce a high Percentage in investment
movie.percentage3=lm(Percentage~Genre:MPAA.Rating:Source,data=pro)
anova(lm(Percentage~Genre:MPAA.Rating:Source,data=pro))
summary(movie.percentage3)
```

```

3.
## for G-rated
G=read.csv("D:\\Study\\MA331\\project\\final-G rate.csv", header=TRUE)
G
is.factor(G$Disney)
is.factor(G$Type)
G$Disney=factor((G$Disney), labels = c("no","yes"))
G$Type=factor((G$Type), labels = c("live action","cartoon"))

Gd=lm(GrossBudget~Disney, data=G)
Gt=lm(GrossBudget~Type, data=G)
Gdt=lm(GrossBudget~Disney+Type, data=G)
anova(Gd)
anova(Gt)
anova(Gdt)
summary(Gd)
summary(Gt)
summary(Gdt)
par(mfrow=c(1,1))
interaction.plot(G$Disney,G$Type,G$GrossBudget)

```

Output:

```

> pro =read.csv("D:\\Study\\MA331\\project\\NewDataFinal.csv",header=TRUE)
> pro

```

	Date	Name	Genre	M.P.A.A. Rating	Source	Worldwide.Gross	Budget	Gross.Budget	Percentage
1	11/2/2007	Bee Movie	Comedy	PG	Original	78621662	150000000	-71378338	-0.47585559
2	9/26/2008	Miracle at St. Anna	Drama	R	Based	9110458	45000000	-35889542	-0.79754538
3	6/5/2009	Land of the Lost	Comedy	PG-13	Based	68099914	100000000	-31900086	-0.31900086
4	10/13/2006	Alex Rider: Operation Stormbreaker	Action	PG	Based	9351567	40000000	-30648433	-0.76621083
5	5/2/1997	Warriors of Virtue	Action	PG	Original	6448817	35000000	-28551183	-0.81574809
6	6/7/1996	The Phantom	Action	PG	Based	17220599	45000000	-27779401	-0.61732002
7	10/10/2008	The Express	Drama	PG	Based	9808102	37500000	-27691898	-0.73845061
8	12/5/2008	Punisher: War Zone	Action	R	Sequel	8199130	35000000	-26800870	-0.76573914
9	5/9/2008	Speed Racer	Action	PG	Based	93394462	120000000	-26605538	-0.22171282
10	10/10/2008	City of Ember	Adventure	PG	Based	11817059	38000000	-26182941	-0.68902476
11	10/23/2009	Amelia	Drama	PG	Based	15526769	40000000	-24473231	-0.61183078
12	8/28/2009	Taking Woodstock	Drama	R	Based	9646830	29000000	-19353170	-0.66735069
13	11/2/2007	The Martian Child	Drama	PG	Based	9076823	27000000	-17923177	-0.66382137
14	10/3/2008	How to Lose Friends and Alienate People	Comedy	R	Based	12031443	28000000	-15968557	-
15	10/23/2009	Cirque du Freak: The Vampire's Assistant	Action	PG-13	Based	24289984	40000000	-15710016	-
16	5/23/2008	The Children of Huang Shi	Drama	R	Based	5527507	20000000	-14472493	-0.72362465
17	12/21/2007	Walk Hard: The Dewey Cox Story	Comedy	R	Original	20575121	35000000	-14424879	-0.41213940
18	11/16/2007	Love in the time of cholera	Drama	R	Based	31077418	45000000	-13922582	-0.30939071
19	6/5/2009	Away We Go	Comedy	R	Original	9467628	21000000	-11532372	-0.54916057
20	7/11/2008	Meet Dave	Comedy	PG	Original	50648806	60000000	-9351194	-0.15585323
21	11/21/2007	I'm Not There	Drama	R	Based	11498547	20000000	-8501453	-0.42507265
22	12/31/2008	Defiance	Drama	R	Original	42268745	50000000	-7731255	-0.15462510
23	8/20/2008	The Rocker	Comedy	PG-13	Original	8767338	15000000	-6232662	-0.41551080
24	6/6/2008	And When Did You Last See Your Father?	Drama	R	Based	2476491	8000000	-5523509	-
25	8/22/2008	Hamlet 2	Comedy	R	Original	4898285	9000000	-4101715	-0.45574611
26	9/19/2008	Igor	Comedy	PG	Original	26608350	30000000	-3391650	-0.11305500
27	6/19/2009	Year One	Comedy	R	Original	57637279	60000000	-2362721	-0.03937868
28	12/5/2008	Frost/Nixon	Drama	R	Based	26870381	29000000	-2129619	-0.07343514
29	1/23/2009	Inkheart	Adventure	PG	Based	58051454	60000000	-1948546	-0.03247577
30	3/27/2009	12 Rounds	Action	PG-13	Original	18175838	20000000	-1824162	-0.09120810
31	7/22/2007	Evan Almighty	Comedy	PG	Sequel	173219280	175000000	-1780720	-0.01017554
32	7/10/2009	I love you, Beth Cooper	Comedy	PG-13	Based	16447529	18000000	-1552471	-0.08624839
33	2/13/2009	Two Lovers	Drama	R	Original	11549034	12000000	-450966	-0.03758050
34	8/29/2008	College	Comedy	R	Original	5629618	6000000	-370382	-0.06173033
35	7/31/2009	Funny People	Comedy	R	Original	70935104	70000000	935104	0.01335863
36	10/3/2008	Rachel Getting Married	Drama	R	Original	13326280	12000000	1326280	0.11052333
37	9/5/2008	Bangkok Dangerous	Action	R	Original	46598133	45000000	1598133	0.03551407
38	10/17/2008	W.	Drama	PG-13	Based	28575778	25100000	3475778	0.13847721
39	12/29/2006	Miss Potter	Drama	PG	Based	35025861	30000000	5025861	0.16752870
40	9/18/2009	Love Happens	Drama	PG-13	Original	24051483	18000000	6051483	0.33619350
41	5/22/2009	Dance Flick	Comedy	PG-13	Original	31605080	25000000	6605080	0.26420320
42	8/22/2008	Death Race	Action	R	Based	72516819	65000000	7516819	0.11564337

43	9/9/2009	9 Adventure	PG-13 Original	38251490	30000000	8251490	0.27504967
44	2/22/2008	Be Kind Rewind	Comedy PG-13 Original	28505302	20000000	8505302	0.42526510
45	5/2/2008	Son of Rambow: A Home Movie	Comedy PG-13 Original	10573083	20000000	8573083	
	4.28654150						
46	8/18/2000	Godzilla 2000	Action PG Sequel	10037390	10000000	9037390	9.03739000
47	10/2/2009	The Invention of Lying	Comedy PG-13 Original	27894004	18500000	9394004	0.50778400
48	3/21/2008	Drillbit Taylor	Comedy PG-13 Original	49686263	40000000	9686263	0.24215658
49	12/1/2006	The Nativity Story	Drama PG Based	46432264	35000000	11432264	0.32663611
50	4/11/2008	The Visitor	Comedy PG-13 Original	16194905	40000000	12194905	3.04872625
51	10/24/2008	Pride and Glory	Drama R Original	43440721	30000000	13440721	0.44802403
52	9/28/2007	The Kingdom	Action R Original	86509602	72500000	14009602	0.19323589
53	8/24/2007	War	Action R Original	40666994	25000000	15666994	0.62667976
54	2/8/2008	In Bruges	Comedy R Original	30782621	15000000	15782621	1.05217473
55	8/6/2008	The Sisterhood of the Traveling Pants 2	Adventure PG-13 Sequel	44154645	27000000	17154645	
	0.63535722						
56	9/25/2009	Surrogates	Action PG-13 Based	102868790	80000000	22868790	0.28585987
57	8/29/2008	Babylon A.D.	Action PG-13 Based	70216497	45000000	25216497	0.56036660
58	1/16/2009	Notorious	Drama R Based	44437535	19000000	25437535	1.33881763
59	11/7/2008	The Boy in the Striped Pyjamas	Drama PG-13 Based	39830581	12500000	27330581	2.18644648
60	10/17/2008	The Secret Life of Bees	Drama PG-13 Based	39612166	11000000	28612166	2.60110600
61	11/26/2008	Milk	Drama R Based	50164027	20000000	30164027	1.50820135
62	12/12/2008	Doubt	Drama PG-13 Based	50923043	20000000	30923043	1.54615215
63	4/25/2008	Harold and Kumar: Escape from Guantanamo Bay	Comedy R Sequel	43231984	12000000	31231984	
	2.60266533						
64	2/29/2008	The Other Boleyn Girl	Drama PG-13 Based	72944278	40000000	32944278	0.82360695
65	9/26/2008	Fireproof	Drama PG Original	33451479	50000000	32951479	65.90295800
66	7/25/2008	The X-Files: I want to believe	Action PG-13 Based	68369434	35000000	33369434	0.95341240
67	9/12/2008	The Women	Comedy PG-13 Original	50042410	16000000	34042410	2.12765062
68	12/17/2008	The Wrestler	Drama R Original	43236603	60000000	37236603	6.20610050
69	2/8/2008	Fool's Gold	Adventure PG-13 Original	111231041	72500000	38731041	0.53422125
70	11/17/2007	August Rush	Drama PG Based	65627510	25000000	40627510	1.62510040
71	8/14/2009	The Time Traveler's Wife	Drama PG-13 Based	82351916	39000000	43351916	1.11158759
72	12/21/2007	Charlie Wilson's War	Drama R Based	118661095	75000000	43661095	0.58214793
73	8/22/2008	The House Bunny	Comedy PG-13 Original	70237389	25000000	45237389	1.80949556
74	3/6/2009	Watchmen	Action R Based	183831502	138000000	45831502	0.33211233
75	10/7/2005	Good Night and Good Luck	Drama PG Based	54601218	70000000	47601218	6.80017400
76	10/17/2008	Max Payne	Action PG-13 Based	85761789	35000000	50761789	1.45033683
77	1/25/2008	Meet the Spartans	Comedy PG-13 Original	84646831	30000000	54646831	1.82156103
78	6/26/2009	My Sister's Keeper	Drama PG-13 Based	83311181	27500000	55811181	2.02949749
79	4/4/2008	Nim's Island	Adventure PG Based	94081683	37000000	57081683	1.54274819
80	10/2/2009	Zombieland	Comedy R Original	83442128	23600000	59842128	2.53568339
81	3/23/2007	TMNT	Action PG Based	95009888	35000000	60009888	1.71456823
82	8/15/2008	Vicky Cristina Barcelona	Comedy PG-13 Original	77213577	16000000	61213577	3.82584856
83	2/6/2009	Coraline	Adventure PG Based	123280657	60000000	63280657	1.05467762
84	2/14/2008	The Spiderwick Chronicles	Adventure PG Based	162839667	92500000	70339667	0.76042883
85	12/8/2006	Blood Diamond	Action R Original	171377916	100000000	71377916	0.71377916
86	12/10/2008	The Reader	Drama R Based	106107610	33000000	73107610	2.21538212
87	8/6/2008	The Pineapple Express	Comedy R Original	100941380	26000000	74941380	2.88236077
88	7/11/2008	Hellboy 2: The Golden Army	Action PG-13 Sequel	160388063	82500000	77888063	0.94409773
89	10/9/2009	Couples Retreat	Comedy PG-13 Original	137911516	60000000	77911516	1.29852527
90	9/18/2009	Cloudy with a chance of meatballs	Comedy PG Based	179890972	100000000	79890972	
	0.79890972						
91	8/7/2009	Julie & Julia	Comedy PG-13 Based	121523567	40000000	81523567	2.03808918
92	11/21/2007	Hitman	Action R Original	99965792	17500000	82465792	4.71233097
93	11/26/2008	Four Christmases	Comedy PG-13 Original	163546040	80000000	83546040	1.04432550
94	1/9/2009	Bride Wars	Comedy PG Original	115150424	30000000	85150424	2.83834747
95	8/13/2009	Tropic Thunder	Comedy R Original	188163455	90000000	98163455	1.09070506
96	12/17/2007	Atonement	Drama R Based	128448676	30000000	98448676	3.28162253
97	7/1/2009	Public Enemies	Drama R Based	203317192	102500000	100817192	0.98358236
98	9/21/2007	Resident Evil: Extinction	Action R Sequel	146162920	45000000	101162920	2.24806489
99	2/16/2007	Bridge to Terabithia	Drama PG Based	136934139	25000000	111934139	4.47736556
100	12/19/2008	Seven Pounds	Drama PG-13 Original	166617328	54000000	112617328	2.08550607
101	6/13/2008	The Incredible Hulk	Adventure PG-13 Sequel	263349257	137500000	125849257	0.91526732
102	8/7/2009	G.I. Joe: The Rise of Cobra	Action PG-13 Based	301060793	175000000	126060793	0.72034739
103	11/26/2008	Australia	Drama PG-13 Original	207482792	78000000	129482792	1.66003579
104	12/20/2006	Rocky Balboa	Drama PG Sequel	155720088	24000000	131720088	5.48833700
105	2/14/2008	Step Up 2 the Streets	Drama PG-13 Sequel	150017783	17500000	132517783	7.57244474
106	2/14/2008	Jumper	Adventure PG-13 Based	222117068	82500000	139617068	1.69232810
107	1/16/2009	Paul Blart: Mall Cop	Comedy PG Original	180449670	26000000	154449670	5.94037192
108	11/21/2008	Bolt	Comedy PG Original	313953579	150000000	163953579	1.09302386

109	3/7/2008	10,000 B.C.	Adventure	PG-13	Original	269065678	105000000	164065678	1.56253027
110	6/15/2007	Fantastic Four: Rise of the Silver Surfer	Action	PG	Sequel	288215319	120000000	168215319	1.40179433
111	12/25/2008	The Curious Case Of Benjamin Button	Drama	PG-13	Based	329809326	160000000	169809326	1.06130829
112	5/21/2009	Terminator Salvation: The Future Begins	Action	PG-13	Sequel	371628539	200000000	171628539	0.85814269
113	12/19/2008	Yes Man	Comedy	PG-13	Based	225990976	500000000	175990976	3.51981952
114	7/24/2009	G-Force	Adventure	PG	Original	272614038	825000000	190114038	2.30441258
115	5/16/2008	The Chronicles of Narnia: Prince Caspian	Adventure	PG	Sequel	419490286	225000000	194490286	0.86440127
116	7/11/2008	Journey to the Center of the Earth	Adventure	PG	Based	240904370	450000000	195904370	4.35343044
117	8/1/2008	The Mummy: Tomb of the Dragon Emperor	Adventure	PG-13	Sequel	397912118	175000000	222912118	1.27378353
118	5/1/2009	X-Men Origins: Wolverine	Action	PG-13	Sequel	374762367	150000000	224762367	1.49841578
119	8/21/2009	Inglorious Basterds	Action	R	Original	307901514	700000000	237901514	3.39859306
120	5/8/2009	Star Trek	Adventure	PG-13	Sequel	384953778	140000000	244953778	1.74966984
121	5/22/2009	Night at the Museum: Battle of the Smithsonian	Comedy	PG	Sequel	414409338	150000000	264409338	1.76272892
122	11/14/2008	Quantum of Solace	Action	PG-13	Sequel	576368427	230000000	346368427	1.50594968
123	11/21/2008	Twilight	Drama	PG-13	Based	383815521	370000000	346815521	9.37339246
124	11/12/2008	Slumdog Millionaire	Drama	R	Original	363934575	140000000	349934575	24.99532679
125	5/30/2008	Sex and the City	Comedy	R	Based	415247258	575000000	357747258	6.22169144
126	3/9/2007	300	Action	R	Based	456068181	600000000	396068181	6.60113635
127	6/5/2009	The Hangover	Comedy	R	Original	462649453	350000000	427649453	12.21855580
128	11/7/2008	Madagascar: Escape 2 Africa	Comedy	PG	Sequel	599761756	150000000	449761756	2.99841171
129	5/29/2009	Up	Adventure	PG	Original	659979556	175000000	484979556	2.77131175
130	5/22/2008	Indiana Jones and the Kingdom of the Crystal Skull	Adventure	PG-13	Sequel	786558765	185000000	601558765	3.25166900
131	6/24/2009	Transformers: Revenge of the Fallen	Action	PG-13	Sequel	833229011	210000000	623229011	2.96775719
132	5/4/2007	Spider-Man 3	Adventure	PG-13	Sequel	890871626	258000000	632871626	2.45299080
133	7/15/2009	Harry Potter and the Half Blood Prince	Adventure	PG	Sequel	934431513	250000000	684431513	2.73772605
134	7/1/2009	Ice Age: Dawn of the Dinosaurs	Adventure	PG	Sequel	883661830	900000000	793661830	8.81846478

> 1.##the relationship between (Budget), (Gross.Budget), and (Profit.Percentage)

```
[1] 1
> 1.1
[1] 1.1
> is.factor(pro$Genre)
[1] TRUE
> is.factor(pro$MPAA.Rating)
[1] TRUE
> is.factor(pro$Source)
[1] TRUE
> is.factor(pro$Budget)
[1] FALSE
> is.factor(pro$Gross.Budget)
[1] FALSE
> is.factor(pro$Profit.Percentage)
[1] FALSE
> 1.2
[1] 1.2
> par(mfrow=c(1,2))
> boxplot(Budget)
> plot(Gross.Budget~Budget, data = pro, main="profit")
> plot(Percentage~Budget, data = pro, main="profit percentage")
> linear = lm(Gross.Budget~Budget, data = pro)
> linear2 = lm(Percentage~Budget, data = pro)
> summary(linear)
```

Call:

```
lm(formula = Gross.Budget ~ Budget, data = pro)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-311123949 -70650457 -16943390 22704999 652788800
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.436e+06 1.574e+07 -0.472 0.637
Budget      1.648e+00 1.877e-01  8.778 7.46e-15 ***
```

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

Residual standard error: 124600000 on 132 degrees of freedom
Multiple R-squared: 0.3686, Adjusted R-squared: 0.3638
F-statistic: 77.06 on 1 and 132 DF, p-value: 7.461e-15

> summary(linear2)

Call:

lm(formula = Percentage ~ Budget, data = pro)

Residuals:

```
   Min     1Q  Median     3Q    Max
-3.6006 -2.3825 -0.9972  0.4086 62.8958
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.014e+00 7.969e-01  3.782 0.000235 ***
Budget      -1.294e-08 9.502e-09 -1.362 0.175648
```

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

Residual standard error: 6.309 on 132 degrees of freedom
Multiple R-squared: 0.01385, Adjusted R-squared: 0.006379
F-statistic: 1.854 on 1 and 132 DF, p-value: 0.1756

> 2.##seperate to Gross.Budget and Profit.Percentage

[1] 2

> 2.1##focus on Profit(Gross.Budget)

[1] 2.1

> 2.1.1

Error: unexpected numeric constant in "2.1.1"

> par(mfrow=c(3,2))

> interaction.plot(pro\$Genre,pro\$MPAA.Rating,pro\$Gross.Budget)

> interaction.plot(pro\$MPAA.Rating,pro\$Genre,pro\$Gross.Budget)

> interaction.plot(pro\$Genre,pro\$Source,pro\$Gross.Budget)

> interaction.plot(pro\$Source,pro\$Genre,pro\$Gross.Budget)

> interaction.plot(pro\$MPAA.Rating,pro\$Source,pro\$Gross.Budget)

> interaction.plot(pro\$Source,pro\$MPAA.Rating,pro\$Gross.Budget)

> movie.profit2=lm(Gross.Budget~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro)

> anova(lm(Gross.Budget~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro))

Analysis of Variance Table

Response: Gross.Budget

```
      Df Sum Sq Mean Sq F value Pr(>F)
Genre:MPAA.Rating 10 6.5009e+17 6.5009e+16 4.1705 6.538e-05 ***
Genre:Source       8 7.3295e+17 9.1619e+16 5.8775 2.909e-06 ***
MPAA.Rating:Source  4 1.3481e+17 3.3702e+16 2.1621 0.07791 .
Residuals         111 1.7303e+18 1.5588e+16
```

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

> summary(movie.profit2)

Call:

lm(formula = Gross.Budget ~ Genre:MPAA.Rating + Genre:Source +
MPAA.Rating:Source, data = pro)

Residuals:

```
   Min     1Q  Median     3Q    Max
-314816055 -58459018 -16638195 31460334 368037325
```

Coefficients: (2 not defined because of singularities)

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  29133550 33304558  0.875 0.38359
```

```

GenreAction:MPAA.RatingPG -69626617 65783674 -1.058 0.29216
GenreAdventure:MPAA.RatingPG 64670027 58957872 1.097 0.27507
GenreComedy:MPAA.RatingPG 25045908 90729544 0.276 0.78302
GenreDrama:MPAA.RatingPG -16027748 52704251 -0.304 0.76162
GenreAction:MPAA.RatingPG-13 38506376 58306709 0.660 0.51036
GenreAdventure:MPAA.RatingPG-13 -93863052 91226065 -1.029 0.30576
GenreComedy:MPAA.RatingPG-13 53452051 66051703 0.809 0.42011
GenreDrama:MPAA.RatingPG-13 45740636 49381533 0.926 0.35632
GenreAction:MPAA.RatingR 111897457 67966649 1.646 0.10252
GenreAdventure:MPAA.RatingR NA NA NA NA
GenreComedy:MPAA.RatingR 101471350 70596697 1.437 0.15343
GenreDrama:MPAA.RatingR NA NA NA NA
GenreAction:SourceOriginal -4339531 86648529 -0.050 0.96015
GenreAdventure:SourceOriginal 214831940 85989450 2.498 0.01394 *
GenreComedy:SourceOriginal 896581 82400671 0.011 0.99134
GenreDrama:SourceOriginal 82925913 80776121 1.027 0.30683
GenreAction:SourceSequel 205734522 79376078 2.592 0.01083 *
GenreAdventure:SourceSequel 414882963 77538863 5.351 4.76e-07 ***
GenreComedy:SourceSequel 188179925 102417042 1.837 0.06883 .
GenreDrama:SourceSequel 97220322 102163549 0.952 0.34336
MPAA.RatingPG-13:SourceOriginal -60478849 80676001 -0.750 0.45505
MPAA.RatingR:SourceOriginal -64248867 83167786 -0.773 0.44145
MPAA.RatingPG-13:SourceSequel -18182761 83056760 -0.219 0.82711
MPAA.RatingR:SourceSequel -302240616 111250232 -2.717 0.00765 **

```

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

Residual standard error: 124900000 on 111 degrees of freedom
Multiple R-squared: 0.4673, Adjusted R-squared: 0.3617
F-statistic: 4.426 on 22 and 111 DF, p-value: 7.93e-08

```

> 2.1.2
Error: unexpected numeric constant in "2.1.2"
> movie.profit3=lm(Gross.Budget~Genre:MPAA.Rating:Source,data=pro)
> anova(lm(Gross.Budget~Genre:MPAA.Rating:Source,data=pro))
Analysis of Variance Table

```

```

Response: Gross.Budget
      Df Sum Sq Mean Sq F value Pr(>F)
Genre:MPAA.Rating:Source 30 1.6250e+18 5.4166e+16 3.4373 1.739e-06 ***
Residuals      103 1.6231e+18 1.5758e+16

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(movie.profit3)

```

Call:
lm(formula = Gross.Budget ~ Genre:MPAA.Rating:Source, data = pro)

```

```

Residuals:
      Min       1Q   Median       3Q      Max
-363037590 -57376126 -19666756 30001153 369056537

```

```

Coefficients: (6 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    31231984 125532693  0.249 0.804014
GenreAction:MPAA.RatingPG:SourceBased -37487855 140349818 -0.267 0.789924
GenreAdventure:MPAA.RatingPG:SourceBased 28513831 135590806  0.210 0.833854
GenreComedy:MPAA.RatingPG:SourceBased 48658988 177530037  0.274 0.784565
GenreDrama:MPAA.RatingPG:SourceBased -12915398 133147528 -0.097 0.922914
GenreAction:MPAA.RatingPG-13:SourceBased 9195897 135590806  0.068 0.946060
GenreAdventure:MPAA.RatingPG-13:SourceBased 108385084 177530037  0.611 0.542864
GenreComedy:MPAA.RatingPG-13:SourceBased 24783513 140349818  0.177 0.860182
GenreDrama:MPAA.RatingPG-13:SourceBased 50887326 132323077  0.385 0.701350
GenreAction:MPAA.RatingR:SourceBased 118573517 144952668  0.818 0.415237
GenreAdventure:MPAA.RatingR:SourceBased NA NA NA NA
GenreComedy:MPAA.RatingR:SourceBased 139657367 153745522  0.908 0.365805
GenreDrama:MPAA.RatingR:SourceBased -10320925 130271432 -0.079 0.937006
GenreAction:MPAA.RatingPG:SourceOriginal -59783167 177530037 -0.337 0.736991
GenreAdventure:MPAA.RatingPG:SourceOriginal 306314813 153745522  1.992 0.048979 *
GenreComedy:MPAA.RatingPG:SourceOriginal 22006765 135590806  0.162 0.871385

```

```

GenreDrama:MPAA.RatingPG:SourceOriginal 1719495 177530037 0.010 0.992291
GenreAction:MPAA.RatingPG-13:SourceOriginal -33056146 177530037 -0.186 0.852654
GenreAdventure:MPAA.RatingPG-13:SourceOriginal 39117419 144952668 0.270 0.787805
GenreComedy:MPAA.RatingPG-13:SourceOriginal -53235 130271432 -0.000409 0.999675
GenreDrama:MPAA.RatingPG-13:SourceOriginal 51485217 144952668 0.355 0.723176
GenreAction:MPAA.RatingR:SourceOriginal 39271341 135590806 0.290 0.772681
GenreAdventure:MPAA.RatingR:SourceOriginal NA NA NA NA
GenreComedy:MPAA.RatingR:SourceOriginal 27360932 131114622 0.209 0.835111
GenreDrama:MPAA.RatingR:SourceOriginal 34394009 135590806 0.254 0.800264
GenreAction:MPAA.RatingPG:SourceSequel 57394371 153745522 0.373 0.709687
GenreAdventure:MPAA.RatingPG:SourceSequel 526295892 144952668 3.631 0.000442 ***
GenreComedy:MPAA.RatingPG:SourceSequel 206231474 144952668 1.423 0.157830
GenreDrama:MPAA.RatingPG:SourceSequel 100488104 177530037 0.566 0.572602
GenreAction:MPAA.RatingPG-13:SourceSequel 257543297 137514175 1.873 0.063925 .
GenreAdventure:MPAA.RatingPG-13:SourceSequel 276318048 135590806 2.038 0.044125 *
GenreComedy:MPAA.RatingPG-13:SourceSequel NA NA NA NA
GenreDrama:MPAA.RatingPG-13:SourceSequel 101285799 177530037 0.571 0.569562
GenreAction:MPAA.RatingR:SourceSequel 5949041 153745522 0.039 0.969209
GenreAdventure:MPAA.RatingR:SourceSequel NA NA NA NA
GenreComedy:MPAA.RatingR:SourceSequel NA NA NA NA
GenreDrama:MPAA.RatingR:SourceSequel NA NA NA NA

```

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

Residual standard error: 125500000 on 103 degrees of freedom
Multiple R-squared: 0.5003, Adjusted R-squared: 0.3547
F-statistic: 3.437 on 30 and 103 DF, p-value: 1.739e-06

```

> 2.2##focus on (Profit.Percentage)
[1] 2.2
> 2.2.2
Error: unexpected numeric constant in "2.2.2"
> par(mfrow=c(3,2))
> interaction.plot(pro$Genre,pro$MPAA.Rating,pro$Percentage)
> interaction.plot(pro$MPAA.Rating,pro$Genre,pro$Percentage)
> interaction.plot(pro$Genre,pro$Source,pro$Percentage)
> interaction.plot(pro$Source,pro$Genre,pro$Percentage)
> interaction.plot(pro$MPAA.Rating,pro$Source,pro$Percentage)
> interaction.plot(pro$Source,pro$MPAA.Rating,pro$Percentage)
> movie.percentage2=lm(Percentage~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro)
> anova(lm(Percentage~Genre:MPAA.Rating+Genre:Source+MPAA.Rating:Source,data=pro))
Analysis of Variance Table

```

```

Response: Percentage
      Df Sum Sq Mean Sq F value Pr(>F)
Genre:MPAA.Rating 10 419.2 41.921 1.4459 0.169717
Genre:Source      8 791.3 98.907 3.4114 0.001528 **
MPAA.Rating:Source 4 899.8 224.943 7.7584 1.499e-05 ***
Residuals       111 3218.3 28.993

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(movie.percentage2)

```

Call:
lm(formula = Percentage ~ Genre:MPAA.Rating + Genre:Source +
    MPAA.Rating:Source, data = pro)

```

```

Residuals:
    Min     1Q   Median     3Q    Max
-11.2067 -1.5375 -0.1933  1.5969 35.0278

```

```

Coefficients: (2 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.4670    1.4363  -0.325 0.745698
GenreAction:MPAA.RatingPG -2.2282    2.8371  -0.785 0.433899
GenreAdventure:MPAA.RatingPG  0.7179    2.5427  0.282 0.778218
GenreComedy:MPAA.RatingPG -10.7279    3.9129  -2.742 0.007128 **
GenreDrama:MPAA.RatingPG   5.4359    2.2730  2.391 0.018461 *
GenreAction:MPAA.RatingPG-13  1.3742    2.5146  0.546 0.585841
GenreAdventure:MPAA.RatingPG-13  7.6438    3.9344  1.943 0.054569 .

```

```

GenreComedy:MPAA.RatingPG-13  3.9363  2.8486  1.382 0.169800
GenreDrama:MPAA.RatingPG-13   1.0000  2.1297  0.470 0.639603
GenreAction:MPAA.RatingR       5.8249  2.9312  1.987 0.049367 *
GenreAdventure:MPAA.RatingR    NA      NA      NA      NA
GenreComedy:MPAA.RatingR       4.9272  3.0447  1.618 0.108433
GenreDrama:MPAA.RatingR        NA      NA      NA      NA
GenreAction:SourceOriginal     13.0861 3.7369 3.502 0.000667 ***
GenreAdventure:SourceOriginal   7.5835 3.7085 2.045 0.043231 *
GenreComedy:SourceOriginal     14.9207 3.5537 4.199 5.44e-05 ***
GenreDrama:SourceOriginal      25.9063 3.4837 7.436 2.32e-11 ***
GenreAction:SourceSequel       7.7566 3.4233 2.266 0.025402 *
GenreAdventure:SourceSequel     2.1865 3.3441 0.654 0.514573
GenreComedy:SourceSequel       12.3671 4.4170 2.800 0.006031 **
GenreDrama:SourceSequel        7.1788 4.4061 1.629 0.106085
MPAA.RatingPG-13:SourceOriginal -17.5007 3.4794 -5.030 1.90e-06 ***
MPAA.RatingR:SourceOriginal    -18.1290 3.5868 -5.054 1.72e-06 ***
MPAA.RatingPG-13:SourceSequel  -6.7987 3.5820 -1.898 0.060294 .
MPAA.RatingR:SourceSequel     -12.9904 4.7980 -2.707 0.007852 **

```

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

Residual standard error: 5.385 on 111 degrees of freedom
Multiple R-squared: 0.396, Adjusted R-squared: 0.2763
F-statistic: 3.308 on 22 and 111 DF, p-value: 1.749e-05

> 2.2.2

Error: unexpected numeric constant in "2.2.2"

> movie.percentage3=lm(Percentage~Genre:MPAA.Rating:Source,data=pro)

> anova(lm(Percentage~Genre:MPAA.Rating:Source,data=pro))

Analysis of Variance Table

Response: Percentage

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Genre:MPAA.Rating:Source	30	4319.1	143.97	14.691	< 2.2e-16 ***
Residuals	103	1009.4	9.80		

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

> summary(movie.percentage3)

Call:

lm(formula = Percentage ~ Genre:MPAA.Rating:Source, data = pro)

Residuals:

Min	1Q	Median	3Q	Max
-5.4159	-1.2913	-0.3000	0.7361	19.7340

Coefficients: (6 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.6027	3.1305	0.831	0.408
GenreAction:MPAA.RatingPG:SourceBased	-2.5753	3.5000	-0.736	0.464
GenreAdventure:MPAA.RatingPG:SourceBased	-1.4377	3.3813	-0.425	0.672
GenreComedy:MPAA.RatingPG:SourceBased	-1.8038	4.4272	-0.407	0.685
GenreDrama:MPAA.RatingPG:SourceBased	-1.1798	3.3204	-0.355	0.723
GenreAction:MPAA.RatingPG-13:SourceBased	-2.0064	3.3813	-0.593	0.554
GenreAdventure:MPAA.RatingPG-13:SourceBased	-0.9103	4.4272	-0.206	0.837
GenreComedy:MPAA.RatingPG-13:SourceBased	-1.3145	3.5000	-0.376	0.708
GenreDrama:MPAA.RatingPG-13:SourceBased	-0.2836	3.2998	-0.086	0.932
GenreAction:MPAA.RatingR:SourceBased	-0.2530	3.6148	-0.070	0.944
GenreAdventure:MPAA.RatingR:SourceBased	NA	NA	NA	NA
GenreComedy:MPAA.RatingR:SourceBased	0.2230	3.8341	0.058	0.954
GenreDrama:MPAA.RatingR:SourceBased	-2.1240	3.2487	-0.654	0.515
GenreAction:MPAA.RatingPG:SourceOriginal	-3.4184	4.4272	-0.772	0.442
GenreAdventure:MPAA.RatingPG:SourceOriginal	-0.0648	3.8341	-0.017	0.987
GenreComedy:MPAA.RatingPG:SourceOriginal	-1.0815	3.3813	-0.320	0.750
GenreDrama:MPAA.RatingPG:SourceOriginal	63.3003	4.4272	14.298	<2e-16 ***
GenreAction:MPAA.RatingPG-13:SourceOriginal	-2.6939	4.4272	-0.608	0.544
GenreAdventure:MPAA.RatingPG-13:SourceOriginal	-1.8121	3.6148	-0.501	0.617
GenreComedy:MPAA.RatingPG-13:SourceOriginal	-1.0422	3.2487	-0.321	0.749
GenreDrama:MPAA.RatingPG-13:SourceOriginal	-1.2421	3.6148	-0.344	0.732
GenreAction:MPAA.RatingR:SourceOriginal	-0.9893	3.3813	-0.293	0.770


```

GenreAdventure:MPAA.RatingR:SourceOriginal      NA      NA      NA      NA
GenreComedy:MPAA.RatingR:SourceOriginal        -0.9413  3.2697 -0.288  0.774
GenreDrama:MPAA.RatingR:SourceOriginal         2.6586  3.3813  0.786  0.434
GenreAction:MPAA.RatingPG:SourceSequel        2.6169  3.8341  0.683  0.496
GenreAdventure:MPAA.RatingPG:SourceSequel      1.5375  3.6148  0.425  0.671
GenreComedy:MPAA.RatingPG:SourceSequel       -1.0190  3.6148 -0.282  0.779
GenreDrama:MPAA.RatingPG:SourceSequel         2.8857  4.4272  0.652  0.516
GenreAction:MPAA.RatingPG-13:SourceSequel     -1.0478  3.4293 -0.306  0.761
GenreAdventure:MPAA.RatingPG-13:SourceSequel  -0.8895  3.3813 -0.263  0.793
GenreComedy:MPAA.RatingPG-13:SourceSequel      NA      NA      NA      NA
GenreDrama:MPAA.RatingPG-13:SourceSequel      4.9698  4.4272  1.123  0.264
GenreAction:MPAA.RatingR:SourceSequel       -1.8615  3.8341 -0.486  0.628
GenreAdventure:MPAA.RatingR:SourceSequel      NA      NA      NA      NA
GenreComedy:MPAA.RatingR:SourceSequel      NA      NA      NA      NA
GenreDrama:MPAA.RatingR:SourceSequel      NA      NA      NA      NA

```

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

Residual standard error: 3.13 on 103 degrees of freedom
Multiple R-squared: 0.8106, Adjusted R-squared: 0.7554
F-statistic: 14.69 on 30 and 103 DF, p-value: < 2.2e-16

```

> 3.
[1] 3
> G=read.csv("D:\\Study\\MA331\\project\\final-G rate.csv", header=TRUE)
> G

```

Date	Name	Genre	Worldwide.Gross	Budget	MPAA.Rating	GrossBudget	Source	Disney	Type
1 8/15/2008	Fly Me To The Moon	Adventure	40087234	25000000	G	15087234	Original	0	1
2 7/18/2008	Space Chimps	Adventure	59517784	37000000	G	22517784	Original	0	1
3 3/14/2008	Horton Hears a Who	Adventure	297133947	85000000	G	212133947	Based	0	1
4 6/27/2008	WALL-E	Comedy	532936655	180000000	G	352936655	Original	1	1
5 6/29/2007	Ratatouille	Comedy	\$624,445,654	150000000	G	474445654	Original	1	1
6 8/25/2007	Mr. Bean's Holiday	Comedy	229,700,105	\$25,000,000	G	204700105	Sequel	0	0
7 12/15/2006	Charlotte's Web	Drama	\$143,985,708	\$82,500,000	G	61485708	Based	0	0
8 11/10/2004	Polar Express	Adventure	\$305,172,033	\$170,000,000	G	135172033	Based	0	1
9 6/9/2006	Cars	Comedy	\$461,982,881	\$70,000,000	G	391982881	Original	1	1
10 2/10/2006	Curious George	Comedy	\$70,114,174	\$50,000,000	G	20114174	Based	0	1
11 4/14/2006	The Wild Adventure		\$99,384,046	\$80,000,000	G	19384046	Original	1	1
12 11/4/2005	Chicken Little	Adventure	\$314,432,738	\$60,000,000	G	254432738	Original	1	1
13 6/22/2005	Herbie: Fully Loaded	Comedy	\$144,110,682	\$50,000,000	G	94110682	Sequel	1	1
14 10/7/2005	Wallace & Gromit: The Curse of the Were-Rabbit	Adventure	\$185,724,838	\$30,000,000	G	155724838	Original	0	1
15 3/18/2005	Ice Princess	Comedy	\$25,732,334	\$25,000,000	G	732334	Original	1	0
16 8/19/2005	Valiant	Adventure	\$61,746,888	\$35,000,000	G	26746888	Original	1	1
17 2/11/2005	Pooh's Heffalump Movie	Adventure	\$52,858,433	\$20,000,000	G	32858433	Based	1	1

```

> is.factor(G$Disney)
[1] FALSE
> is.factor(G$Type)
[1] FALSE
> G$Disney=factor((G$Disney), labels = c("no","yes"))
> G$Type=factor((G$Type), labels = c("live action","cartoon"))
> Gd=lm(GrossBudget~Disney, data=G)
> Gt=lm(GrossBudget~Type, data=G)
> Gdt=lm(GrossBudget~Disney+Type, data=G)
> anova(Gd)

```

Analysis of Variance Table

```

Response: GrossBudget
  Df Sum Sq Mean Sq F value Pr(>F)
Disney 1 2.6905e+16 2.6905e+16 1.2368 0.2836
Residuals 15 3.2631e+17 2.1754e+16
> anova(Gt)

```

Analysis of Variance Table

```

Response: GrossBudget
  Df Sum Sq Mean Sq F value Pr(>F)
Type 1 1.1666e+16 1.1666e+16 0.5123 0.4851
Residuals 15 3.4155e+17 2.2770e+16
> anova(Gdt)

```

Analysis of Variance Table

Response: GrossBudget

```
      Df Sum Sq Mean Sq F value Pr(>F)
Disney  1 2.6905e+16 2.6905e+16  1.1771 0.2963
Type    1 6.3213e+15 6.3213e+15  0.2766 0.6072
Residuals 14 3.1999e+17 2.2857e+16
> summary(Gd)
```

Call:

```
lm(formula = GrossBudget ~ Disney, data = G)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-182337701 -88959353 -41881270 101333127 291375619
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 103366978 52146631  1.982  0.0661 .
Disneyyes   79703057 71668689  1.112  0.2836
---
```

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

Residual standard error: 147500000 on 15 degrees of freedom
Multiple R-squared: 0.07617, Adjusted R-squared: 0.01458
F-statistic: 1.237 on 1 and 15 DF, p-value: 0.2836

> summary(Gt)

Call:

```
lm(formula = GrossBudget ~ Type, data = G)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-142601908 -130942254 -27487008 96743596 316756512
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 88972716 87120807  1.021  0.323
Typecartoon 68716426 96002454  0.716  0.485
```

Residual standard error: 150900000 on 15 degrees of freedom
Multiple R-squared: 0.03303, Adjusted R-squared: -0.03144
F-statistic: 0.5123 on 1 and 15 DF, p-value: 0.4851

> summary(Gdt)

Call:

```
lm(formula = GrossBudget ~ Disney + Type, data = G)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-169401609 -101139890 -3300831 95906823 285659999
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 64786539 90768807  0.714  0.487
Disneyyes   72558531 74707711  0.971  0.348
Typecartoon 51440586 97815354  0.526  0.607
```

Residual standard error: 151200000 on 14 degrees of freedom
Multiple R-squared: 0.09407, Adjusted R-squared: -0.03535
F-statistic: 0.7268 on 2 and 14 DF, p-value: 0.5008

> par(mfrow=c(1,1))

> interaction.plot(G\$Disney,G\$Type,G\$GrossBudget)

>