

MS 491

Activity #7

Group 7

Sai Yashverdhan 19110027
Gaurav Sharma 19110181
Anushka Niti 19110040
Deepak Patel 1911010



```
String code table for variable 4
1 = 'Metamorphosis'
2 = 'Quest'
3 = 'Monster Force'
4 = 'Rescue'
5 = 'Pursuit'
6 = 'The Riddle'
7 = 'Comedy'
8 = 'Sacrifice'
9 = 'Transformation'
10 = 'Maturation'
11 = 'Love'
12 = 'pursuit'
13 = 'Journey and Return'
14 = 'Underdog'
15 = 'Rivalry'
16 = 'Revenge'
17 = 'discovery'
18 = 'Discovery'
19 = 'Escape'
20 = 'sacrifice'
21 = 'Wretched Excess'
22 = 'Temptation'
23 = 'quest'
24 = 'love'
25 = 'rescue'
26 = 'fish out of water'
27 = 'rivalry'
28 = 'escape'
29 = 'maturation'
30 = 'underdog'
31 = 'journey & return'
32 = 'Fish Out of Water'
33 = 'comedy'
34 = 'monster force'
35 = 'riddle'
36 = 'tragedy'
37 = 'temptation'
38 = 'transformation'
39 = 'Tragedy'
40 = 'Rags to Riches'
41 = 'Rags To Riches'
42 = 'Fish Out Of Water'
```

```
String code table for variable 1
1 = 'Sony'
2 = 'Paramount'
3 = 'Disney'
4 = 'Warner Bros'
5 = 'Universal'
6 = 'Fox'
7 = 'Independent'
8 = 'Weinstein'
9 = 'Lionsgate'
10 = 'MGM'
11 = 'Mediaplex'
12 = 'Summit'
13 = 'Relativity Media'
14 = 'Buena Vista'
15 = 'Overture'
16 = 'Liberty Starz'
17 = 'New Line'
18 = 'Focus'
19 = 'Miramax'
20 = 'Crest'
21 = 'CBS'
22 = 'Highlight Communications'
23 = 'Vertigo'
24 = 'Aardman Animations'
25 = 'Pixar'
26 = 'Morgan Creek'
27 = 'Reliance'
28 = 'Spyglass Entertainment'
29 = 'DreamWorks'
30 = 'Village Roadshow'
```

Code for Lead Studio

```
31 = 'Regency Enterprises'
32 = 'Happy Madison'
33 = 'Virgin'
34 = 'Columbia'
35 = 'Legendary Pictures'
36 = 'Entertainment One'
37 = 'Magnolia'
38 = 'Eros'
39 = 'Roadside Attractions'
40 = 'Atlas Distribution'
41 = 'UTV'
42 = 'Fox Searchlight'
43 = 'Yash Raj'
44 = 'Open Road'
45 = 'Cohen Media'
46 = 'ARC Entertainment'
47 = 'IFC'
48 = 'LD Entertainment'
49 = 'Rocky Mountain'
50 = 'TriStar'
51 = 'Music Box Films'
52 = 'Samuel Goldwyn'
53 = 'FilmDistrict'
54 = 'Screen Gems'
55 = 'Oscilloscope'
56 = 'A24'
57 = 'Millenium Entertainment'
58 = 'Cinedigm Entertainment'
59 = 'Radius-TWC'
```

Code for Lead Studio
(cntd.)

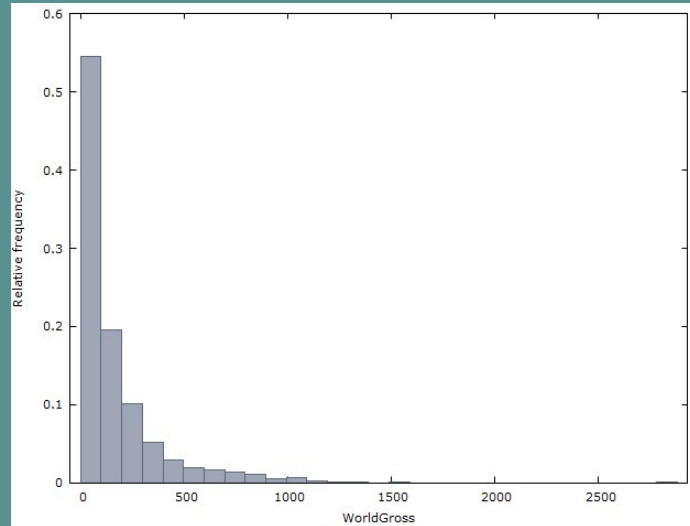
```
String code table
1 = 'Action'
2 = 'Animation'
3 = 'Adventure'
4 = 'Thriller'
5 = 'Comedy'
6 = 'Musical'
7 = 'Drama'
8 = 'Biography'
9 = 'Horror'
10 = 'Romance'
11 = 'Fantasy'
12 = 'Documentary'
13 = 'Crime'
14 = 'Mystery'
```

Code for Genre

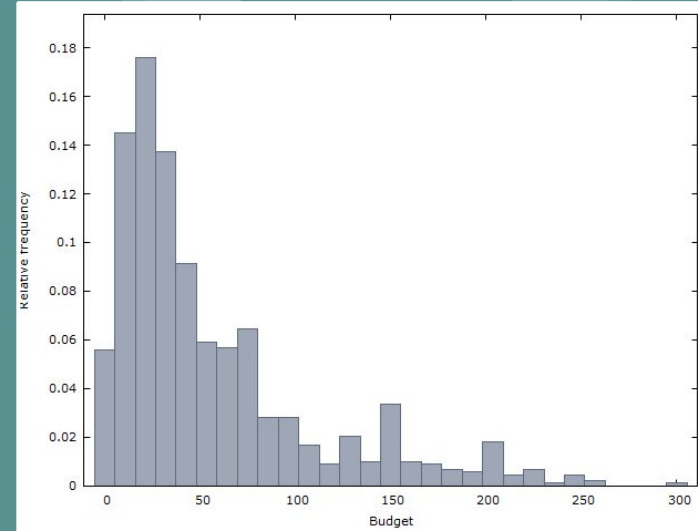
Code for Story

Year wise full
summary statistics

Year	World Gross				Profitability			
	Mean	Median	Standard Deviation	Coefficient of Variation	Mean	Median	Standard Deviation	Coefficient of Variation
2007	188.35	99.97	210.69	1.11	355.15	277.9	364.22	1.02
2008	136.34	69.4	167.81	1.23	379.68	244.05	632.37	1.66
2009	166.71	77.52	290.58	1.74	298.76	218	258.55	0.86
2010	166.5	85.3	204.3	1.22	394.55	248.25	671.45	1.7
2011	166.85	96.05	215.94	1.29	373.51	253.64	650.84	1.74
2012	233.52	131.52	274.28	1.16	567.02	305.43	1119.1	1.97
2013	151.73	71.63	214.68	1.41	366.37	258.78	447.89	1.23



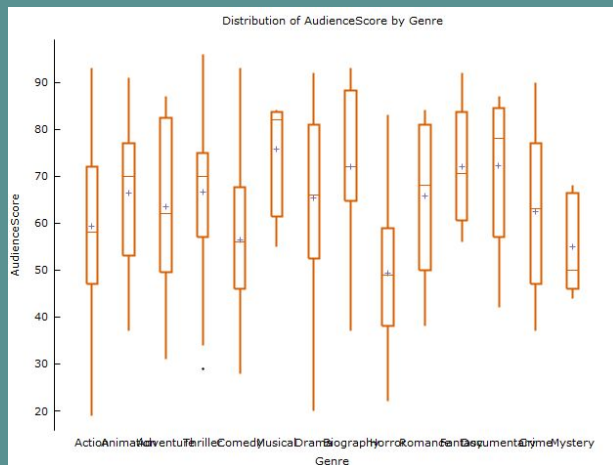
Frequency distribution for world gross



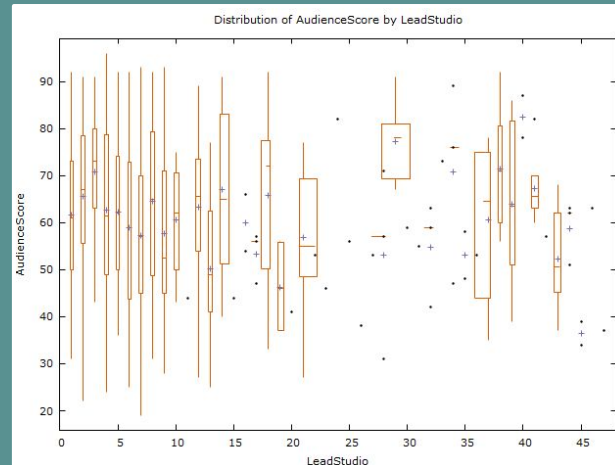
Frequency distribution for budget

	Budget				Opening weekend			
Genre	Mean	Median	Standard Deviation	Coefficient of Variation	Mean	Median	Standard Deviation	Coefficient of Variation
Action	85.86	66.25	64.026	0.7457	32.444	20.9	32.189	0.99215
Animation	108.81	100	59.26	0.5446	35.569	34.08	25.639	0.72083
Adventure	80.35	60	58.648	0.72991	26.086	14	31.721	1.216
Thriller	58.034	48	42.915	0.73947	24.254	14.81	28.847	1.1894
Comedy	36.896	30	26.66	0.72256	15.524	12.22	13.568	0.87397
Musical	53.5	62.5	28.85	0.53926	18.285	18.17	8.0312	0.43922
Drama	29.103	24.5	22.943	0.78835	10.758	7.33	16.357	1.5205
Biography	35.071	24.5	27.565	0.78595	11.204	6.58	11.044	0.98573
Horror	21.725	20	13.321	0.61316	16.83	14.1	10.972	0.65192
Romance	36.85	34	25.417	0.68974	20.901	12.37	31.255	1.4954
Fantasy	74.45	68.25	61.783	0.82985	39.525	15.7	64.157	1.6232
Documentary	19	9	23.95	1.2606	12.134	3.41	18.651	1.5371
Crime	33.32	30	20.895	0.6271	15.733	16.3	8.4353	0.53616
Mystery	47.8	20	58.73	1.2287	15.58	11.6	17.628	1.1314

Genre wise Full
summary statistics



Box plot (audience score vs genre)



Box plot (audience score vs lead studio)

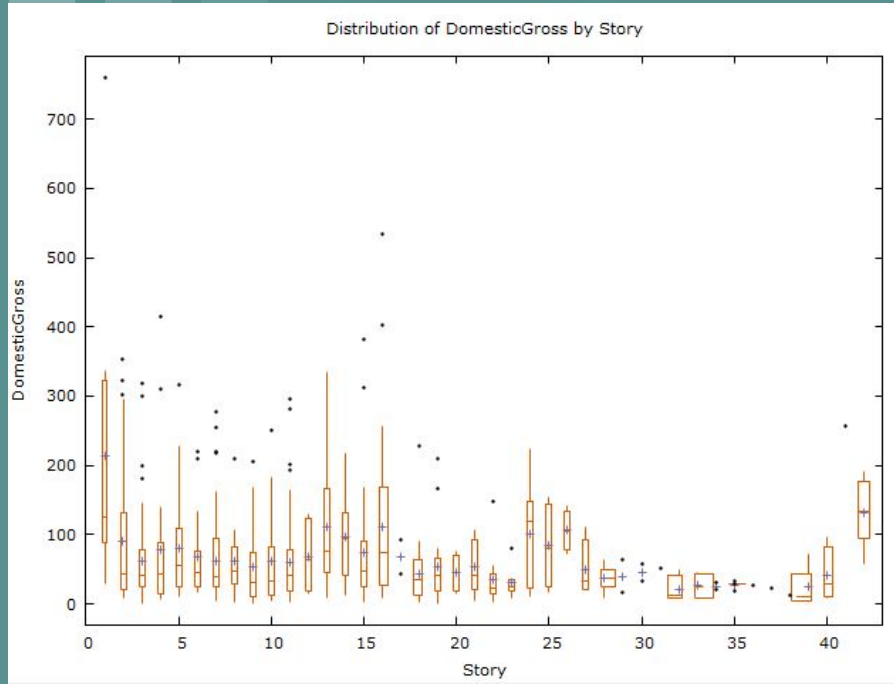
Year wise summary statistics

	Theaters Open Week					Opening Weekend				
Year	Mean	Median	Standard Deviation	Minimum	Maximum	Mean	Median	Standard Deviation	Minimum	Maximum
2007	2754	2785	896.9	605	4362	25.42	14.5	27.08	1.6	151.1
2008	2664	2714	902.6	9	4366	19.72	13.51	21.34	0.09	158.4
2009	2597	2734	946.9	4	4325	21.74	14.9	22.53	0.13	142.8
2010	2683	2877	1023	2	4468	21.84	15.8	22.99	0.032	128.1
2011	2721	2979	1006	4	4375	20.75	13.07	24.71	0.05	169.2
2012	1903	2433	1531	1	4404	18.42	8.43	30.21	0.03	207.4
2013	2345	2695	1277	1	4253	18.77	8.78	27.13	0.01	174.1

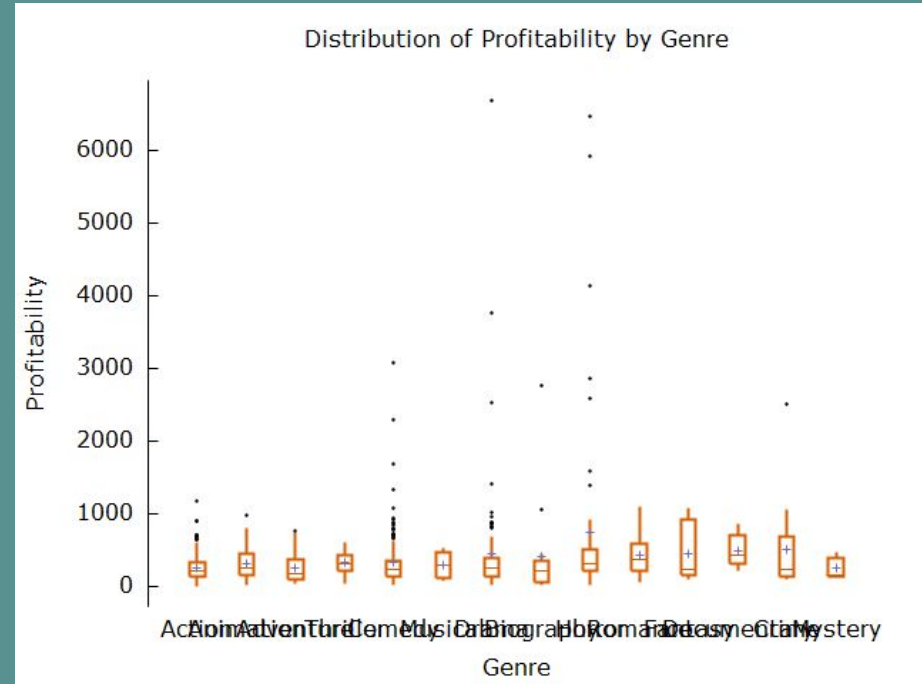
	Rotten Tomatoes					Audience Score				
Genre	Mean	Median	Standard Deviation	Minimum	Maximum	Mean	Median	Standard Deviation	Minimum	Maximum
Action	44.75	41	23.85	4	94	59.3	58	17.1	19	93
Animation	62.14	66	27.4	4	99	66.29	70	14.94	37	91
Adventure	54.27	53	28.32	9	96	63.47	62	16.99	31	87
Thriller	60.91	69	26.29	14	94	66.54	70	15.75	29	96
Comedy	44.65	43	26.93	1	97	56.39	56	14.39	28	93
Musical	61.5	63.5	32.36	27	92	75.75	82	13.89	55	84
Drama	57.16	56	25.35	5	97	65.41	66	17.93	20	92
Biography	72.71	68	14.41	50	96	72	72	15.08	37	93
Horror	37.39	26	27.34	0	92	49.27	49	14.09	22	83
Romance	50.68	49	25.48	8	93	65.68	68	14.17	38	84
Fantasy	62.33	68	26.33	27	96	72	70.5	13.16	56	92
Documentary	63.4	70	31.94	10	93	72.2	78	17.75	42	87
Crime	50.93	51	28.48	10	94	62.53	63	16.88	37	90
Mystery	42	36	29.81	12	83	55	50	10.77	44	68

Genre
wise
summary
statistics

Box Plot of Domestic Gross v/s Story



Box Plot of Profitability v/s Genre



	Mean	Median	S.D.	Min	Max
DomesticGross	68.16	40.41	80.41	0.06000	760.5
ForeignGross	101.2	46.66	156.0	0.0000	2021
WorldGross	169.0	88.18	227.7	0.1000	2782
Profitability	384.6	254.8	631.7	2.300	10176

Top 5

	Profitability			
Story	Mean	Median	Standard Deviation	Coefficient of Variation
Wretched excess	742.85	304.72	1021.4	1.375
Monster Force	645.06	273.6	1269.9	1.9686
Maturation	591.87	364.76	775.28	1.3099
Escape	482.15	367.16	343.88	0.71322
Underdog	438.94	272.06	573.59	1.3068

Story wise summary statistics of Profitability(top 5 by mean)

Bottom 5

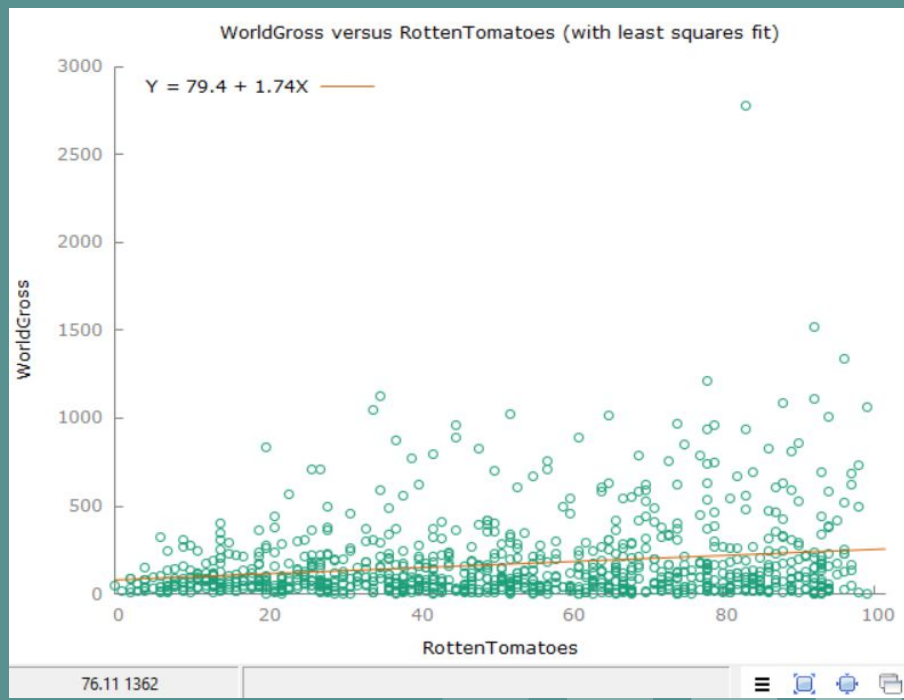
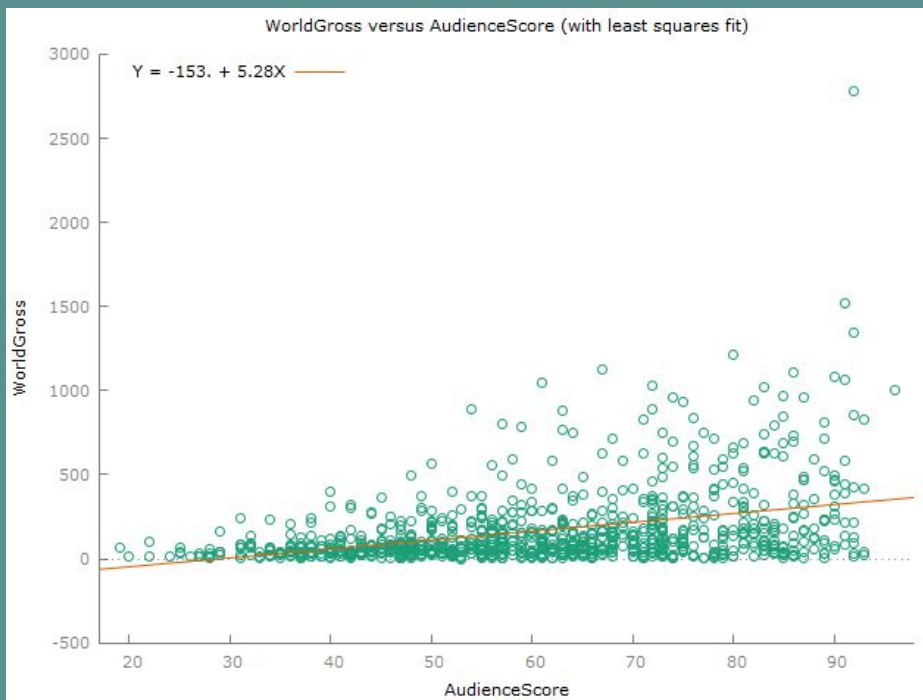
	Profitability			
Story	Mean	Median	Standard Deviation	Coefficient of Variation
Fish out of water	166.42	155.16	94.881	0.57013
Rescue	188.84	174.08	116.03	0.6144
Transformation	256.2	256.18	160.33	0.62579
Discovery	265.19	238.69	214.84	0.81014
Revenge	268.52	248.73	171.62	0.63914

Story wise summary statistics of Profitability(bottom 5 by mean)

Regression Analysis (Plots)

```
corr(AudienceScore, WorldGross) = 0.38377096  
Under the null hypothesis of no correlation:  
t(904) = 12.4955, with two-tailed p-value 0.0000
```

```
corr(RottenTomatoes, WorldGross) = 0.20415456  
Under the null hypothesis of no correlation:  
t(910) = 6.29107, with two-tailed p-value 0.0000
```



Regression Analysis (World Gross Vs Audience Score)

Model 4: OLS, using observations 1-970 (n = 906)

Missing or incomplete observations dropped: 64

Dependent variable: WorldGross

	coefficient	std. error	t-ratio	p-value	

const	-153.270	26.8116	-5.717	1.48e-08	***
AudienceScore	5.27925	0.422492	12.50	3.64e-033	***
Mean dependent var	170.0987	S.D. dependent var	228.3854		
Sum squared resid	40252374	S.E. of regression	211.0141		
R-squared	0.147280	Adjusted R-squared	0.146337		
F(1, 904)	156.1372	P-value(F)	3.64e-33		
Log-likelihood	-6133.401	Akaike criterion	12270.80		
Schwarz criterion	12280.42	Hannan-Quinn	12274.48		

Regression Analysis

File Edit Tests Save Graphs Analysis LaTeX



Model 7: OLS, using observations 1-970 (n = 891)

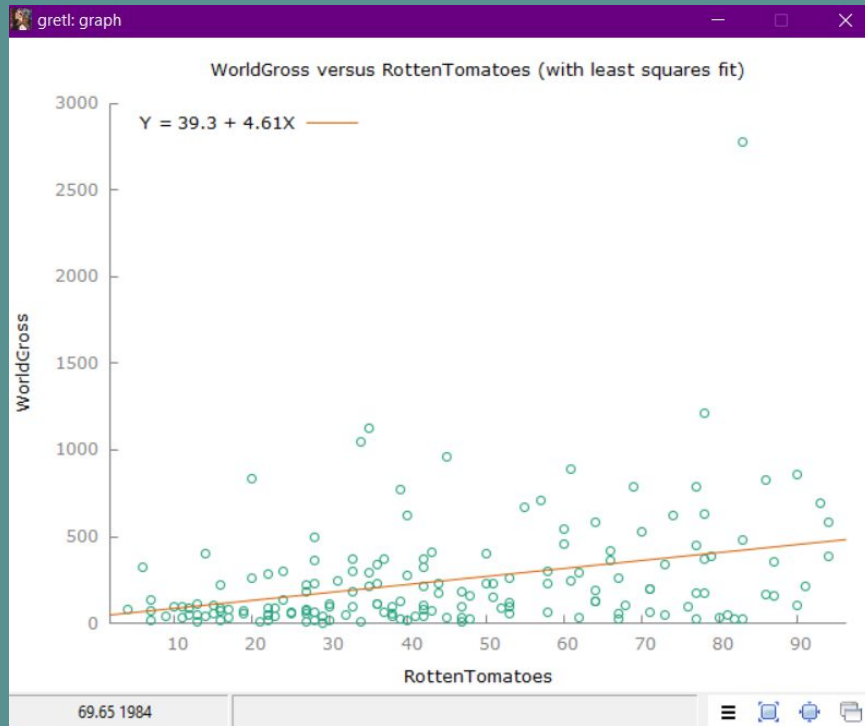
Missing or incomplete observations dropped: 79

Dependent variable: WorldGross

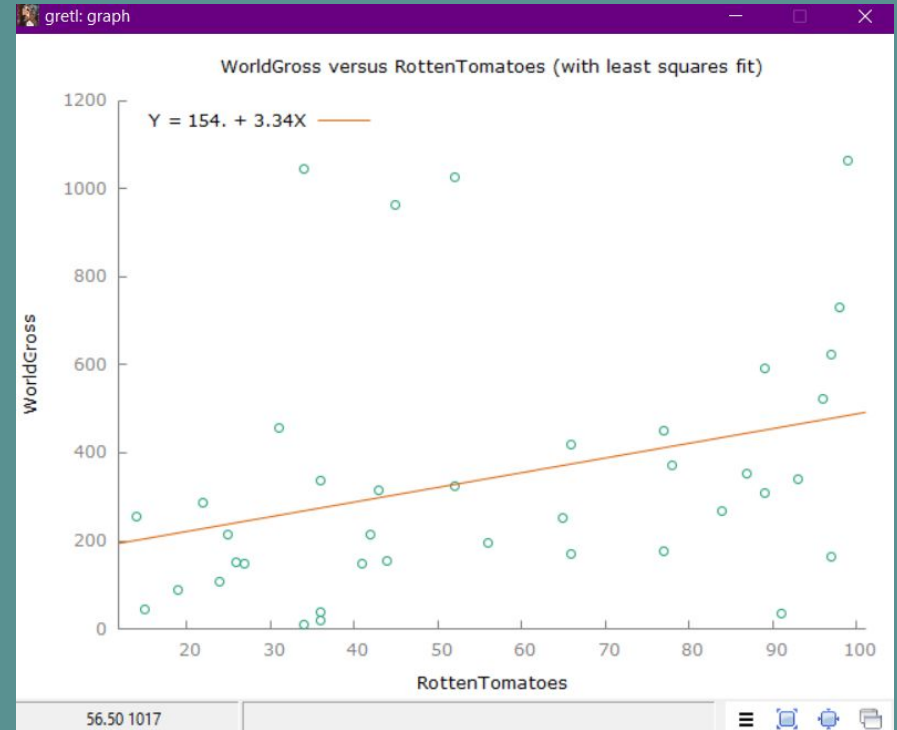
	coefficient	std. error	t-ratio	p-value
const	-3869.62	4999.18	-0.7741	0.4391
Budget	2.94515	0.0923085	31.91	2.18e-149 ***
AudienceScore	3.25401	0.299673	10.86	7.09e-026 ***
Year	1.82894	2.48737	0.7353	0.4624
Mean dependent var	172.2920	S.D. dependent var	229.5740	
Sum squared resid	18592279	S.E. of regression	144.7786	
R-squared	0.603633	Adjusted R-squared	0.602293	
F(3, 887)	450.2756	P-value(F)	1.1e-177	
Log-likelihood	-5695.178	Akaike criterion	11398.36	
Schwarz criterion	11417.53	Hannan-Quinn	11405.68	

Excluding the constant, p-value was highest for variable 15 (Year)

Regression Analysis (Restricted Dataset)



```
corr(RottenTomatoes, WorldGross) = 0.34480179  
Under the null hypothesis of no correlation:  
t(163) = 4.68973, with two-tailed p-value 0.0000
```



```
corr(RottenTomatoes, WorldGross) = 0.32639141  
Under the null hypothesis of no correlation:  
t(37) = 2.10039, with two-tailed p-value 0.0426
```



Thank You