



# ANIME DATA MINING REPORT

**GROUP 13**

- Jue Gu
- Xuwei Zhang

# About the data:

# 01



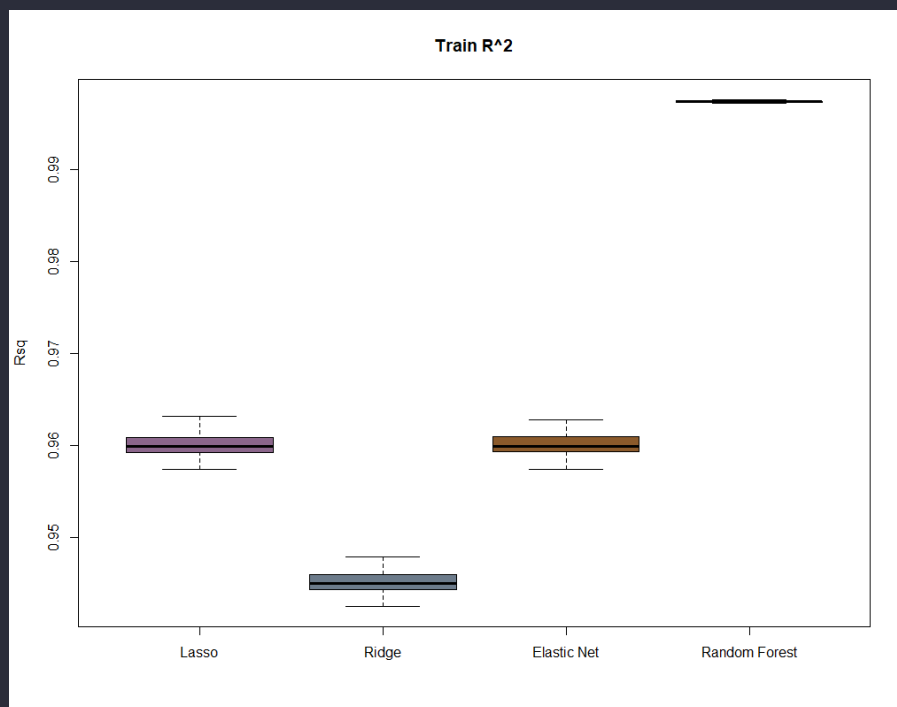
- Data Description
  - Kaggle
  - Information of every anime.
  - Score, Genre, English Name, Japanese Name, Type, Studio, etc.\*
- Data Cleaning
  - Removed columns
    - English Name, Japanese Name, Anime ID
  - Removed rows with missing value
  - Converted categorical columns into dummy values
    - Studio, Genre, Type, etc.
- Finalized Data
  - Dimension: 10,639 x 105
  - Response Variable: Score
  - Predictors: Episodes, Ranked, Popularity. etc\*\*

\* Full Column List included in the appendix

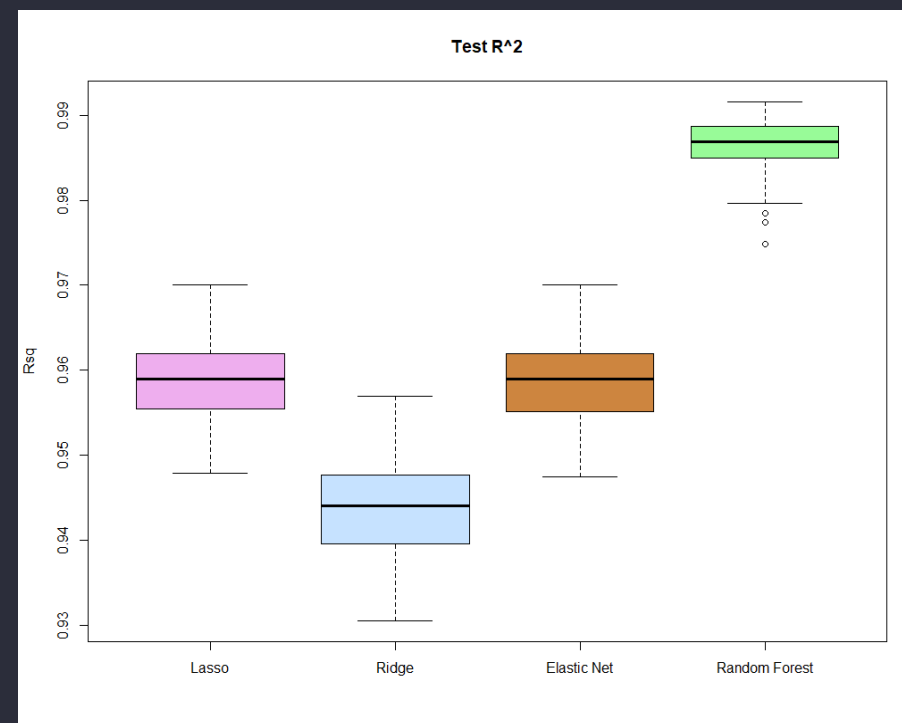
\*\* Full list of predictors included in the appendix

# R<sup>2</sup>-Boxplots

02



Train

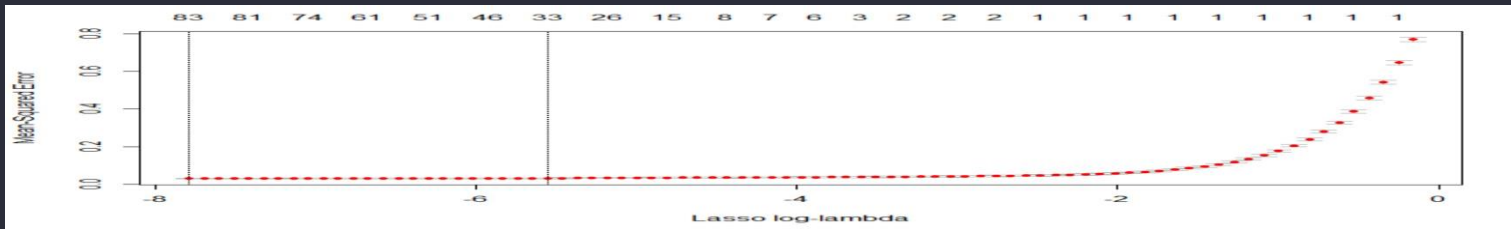


Test

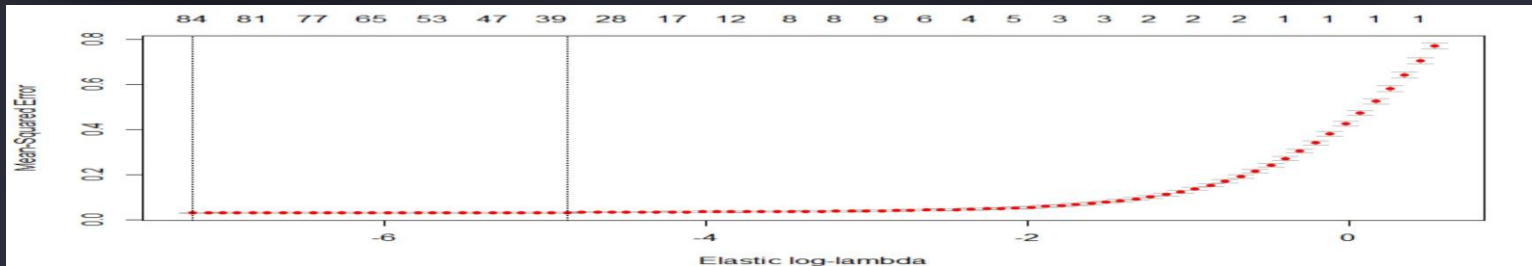
10-fold cv LASSO/RIDGE/EN

# Cross Validation

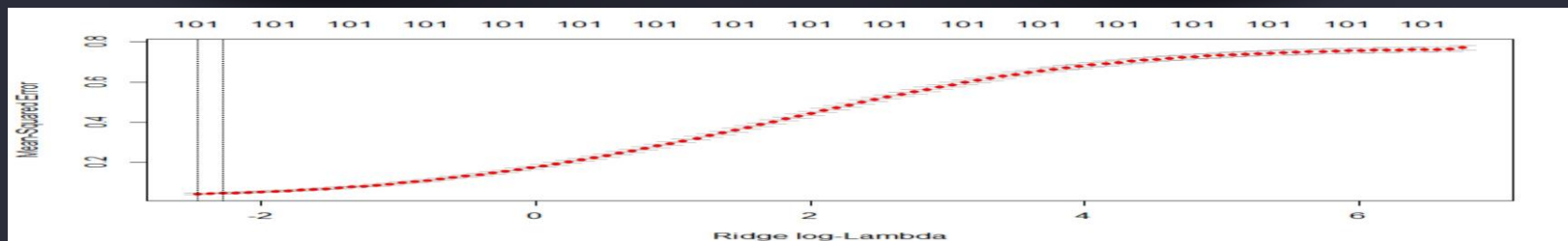
03



LASSO 10- FOLD CV



ELASTIC-NET 10-FOLD CV



RIDGE 10-FOLD CV

# 04

10-FOLD CV FOR LASSO /ELASTIC-NET/RIDGE

## Time

LASSO

#1

10-fold-lasso: 11.98 sec elapsed

ELASTIC-NET

#2

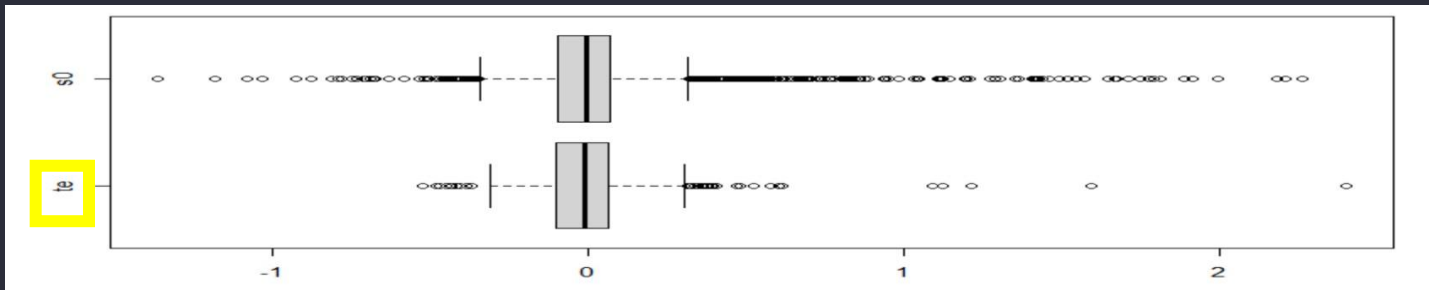
10-fold-en: 13.98 sec elapsed

RIDGE

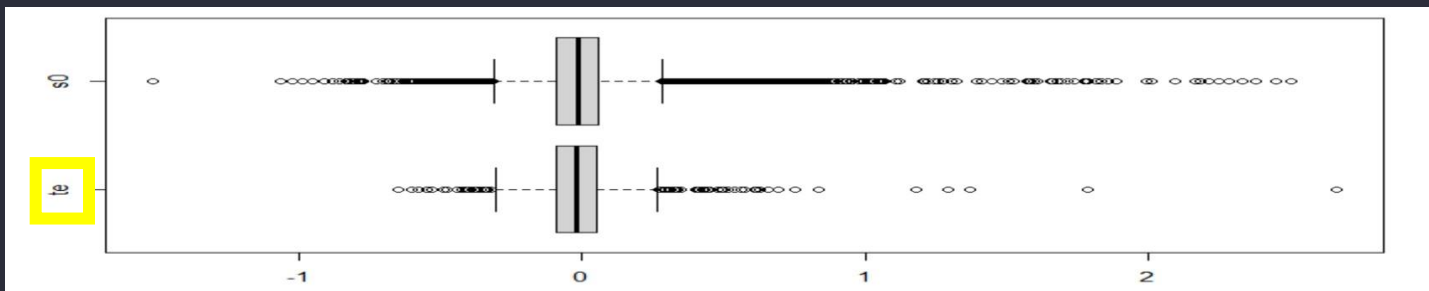
#3

10-fold-ridge: 15.1 sec elapsed

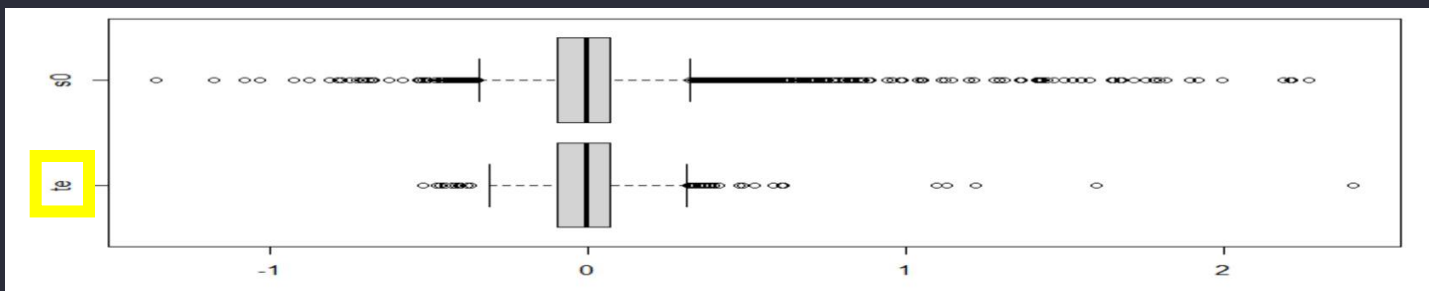
# LASSO/RIDGE/ELASTIC-NET RESIDUALS



LASSO



RIDGE



ELASTIC-NET

# Trade-off

# 06

## Between Time and Predictive Performance

90% test R-square  
interval based on  
100 samples



LASSO

5%	95%
0.9357316	0.9532638

Ridge

5%	95%
0.9357316	0.9532638

Elastic-Net

5%	95%
0.9517095	0.9665393

Random Forest

5%	95%
0.9811175	0.9906774

Time

15.27 sec

12.89 sec

12.47 sec

4.08 mins (245.23 sec)



# ABOUT Coefficients

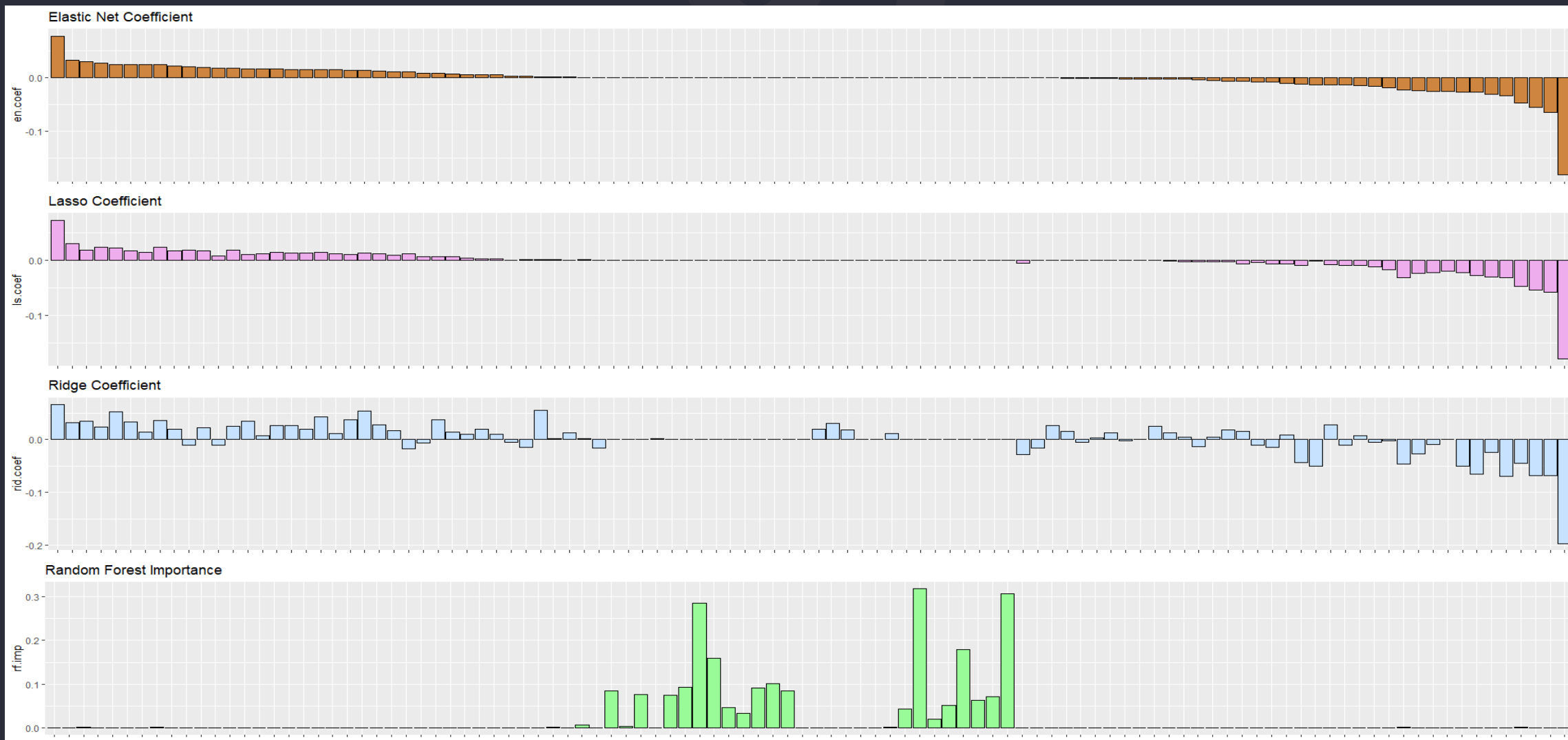
## Feature Importance





# ABOUT Coefficients

Feature Importance



# Feature Importance Rank

LASSO/RIDGE/ELASTIC-NET/RANDOM FOREST

Feature Importance

## EN

Rank	Predictor	EN Coefficient
1	g_samurai	0.077358
2	g_parody	0.032592
3	Studio_Top25per	0.029202
4	g_sci_Fi	0.026434
5	g_sports	0.024753
6	g_cars	0.024233
7	Studio_25per_50per	0.024081
8	rating_G	0.023659
9	g_game	0.020919
10	g_ecchi	0.019929
11	g_fantasy	0.018204
12	Studio_50per_75per	0.017628
13	sce_unknown	0.017552
14	g_police	0.016456
15	g_shounenAi	0.015945

## LASSO

Rank	Predictor	Lasso Coefficient
1	g_samurai	0.072108
2	g_parody	0.029530
3	g_sci_Fi	0.023663
4	rating_G	0.023237
5	g_sports	0.021782
6	g_ecchi	0.018452
7	sce_unknown	0.017728
8	Studio_Top25per	0.017663
9	g_cars	0.017361
10	g_game	0.017328
11	g_fantasy	0.016837
12	Studio_25per_50per	0.014377
13	g_drama	0.013810
14	type_special	0.013737
15	g_emons	0.013162

## RID

Rank	Predictor	Ridge Coefficient
1	g_samurai	0.065985
2	type_music	0.055247
3	g_historical	0.053306
4	g_sports	0.052462
5	g_drama	0.043280
6	g_mystery	0.037415
7	g_sliceofLife	0.037066
8	rating_G	0.035586
9	Studio_Top25per	0.034904
10	g_police	0.034844
11	g_cars	0.033256
12	g_parody	0.031647
13	g_shounen	0.030508
14	Studio_Sunrise	0.027858
15	g_kids	0.027610

## RF

Rank	Predictor	Random Forest Imp
1	Score_9	0.322098
2	Score_8	0.316886
3	Ranked	0.288404
4	Score_7	0.208636
5	Score_10	0.135813
6	Completed	0.123617
7	Members	0.101886
8	Popularity	0.097503
9	Score_6	0.093084
10	Score_3	0.089240
11	Score_2	0.086249
12	Score_5	0.070848
13	Score_4	0.064446
14	Score_1	0.063728
15	Favorites	0.055157

What we learned from....

# INSIGHTS

01

**Intercept**

02

**Trade-off**

03

**Important Predictors**



**THANK YOU !**



# Appendix

# \*Full Column List

1. MAL\_ID: MyAnimeList ID of the anime. (e.g. 1)
2. Name: full name of the anime. (e.g. Cowboy Bebop)
3. Score: average score of the anime given from all users in MyAnimeList database. (e.g. 8.78)
4. Genres: comma separated list of genres for this anime. (e.g. Action, Adventure, Comedy, Drama, Sci-Fi, Space)
5. English name: full name in english of the anime. (e.g. Cowboy Bebop)
6. Japanese name: full name in japanses of the anime. (e.g. カウボーイビバップ)
7. Type: TV, movie, OVA, etc. (e.g. TV)
8. Episodes': number of chapters. (e.g. 26)
9. Aired: broadcast date. (e.g. Apr 3, 1998 to Apr 24, 1999)
10. Premiered: season premiere. (e.g. Spring 1998)
11. Producers: comma separated list of produducers (e.g. Bandai Visual)
12. Licensors: comma separated list of licensors (e.g. Funimation, Bandai Entertainment)
13. Studios: comma separated list of studios (e.g. Sunrise)
14. Source: Manga, Light novel, Book, etc. (e.g. Original)
15. Duration: duration of the anime per episode (e.g. 24 min. per ep.)
16. Rating: age rate (e.g. R - 17+ (violence & profanity))
17. Ranked: position based in the score. (e.g. 28)
18. Popularity: position based in the the number of users who have added the anime to their list. (e.g. 39)
19. Members: number of community members that are in this anime's "group". (e.g. 1251960)
20. Favorites: number of users who have the anime as "favorites". (e.g. 61,971)
21. Watching: number of users who are watching the anime. (e.g. 105808)
22. Completed: number of users who have complete the anime. (e.g. 718161)
23. On-Hold: number of users who have the anime on Hold. (e.g. 71513)
24. Dropped: number of users who have dropped the anime. (e.g. 26678)
25. Plan to Watch': number of users who plan to watch the anime. (e.g. 329800)
26. Score-10': number of users who scored 10. (e.g. 229170)
27. Score-9': number of users who scored 9. (e.g. 182126)
28. Score-8': number of users who scored 8. (e.g. 131625)
29. Score-7': number of users who scored 7. (e.g. 62330)
30. Score-6': number of users who scored 6. (e.g. 20688)
31. Score-5': number of users who scored 5. (e.g. 8904)
32. Score-4': number of users who scored 4. (e.g. 3184)
33. Score-3': number of users who scored 3. (e.g. 1357)
34. Score-2': number of users who scored 2. (e.g. 741)
35. Score-1': number of users who scored 1. (e.g. 1580)



# Full List of Response Variable and Predictor

Response Variable:

```
Score
Min.    :1.85
1st Qu.:6.03
Median  :6.60
Mean    :6.59
3rd Qu.:7.22
Max.    :9.19
```

Predictor:

[1] "score"	"Episodes"	"Ranked"
[4] "Popularity"	"Members"	"Favorites"
[7] "Watching"	"Completed"	"On_Hold"
[10] "Dropped"	"Plan_to_watch"	"Score_10"
[13] "score_9"	"score_8"	"score_7"
[16] "score_6"	"score_5"	"score_4"
[19] "score_3"	"score_2"	"score_1"
[22] "Premiere_Year"	"Duration_Minute"	"Studio_Unknown"
[25] "Studio_Sunrise"	"Studio_Madhouse"	"Studio_Toei_Animation"
[28] "Studio_J.C.Staff"	"Studio_Studio_Pierrot"	"Studio_Studio_Deen"
[31] "Studio_Production_I.G"	"Studio_OLM"	"Studio_TMS_Entertainment"
[34] "Studio_A_1_Pictures"	"Studio_Top25per"	"Studio_25per_50per"
[37] "Studio_50per_75per"	"Studio_75per_100per"	"sce_manga"
[40] "sce_game"	"sce_lightnovel"	"sce_novel"
[43] "sce_original"	"sce_unknown"	"sce_visualnovel"
[46] "sce_webmanga"	"sce_other"	"g_adventure"
[49] "g_cars"	"g_comedy"	"g_dementia"
[52] "g_emons"	"g_drama"	"g_ecchi"
[55] "g_fantasy"	"g_game"	"g_harem"
[58] "g_hentai"	"g_historical"	"g_horror"
[61] "g_josei"	"g_kids"	"g_magic"
[64] "g_martial_Arts"	"g_mecha"	"g_military"
[67] "g_music"	"g_mystery"	"g_parody"
[70] "g_police"	"g_psychological"	"g_romance"
[73] "g_samurai"	"g_school"	"g_sci_Fi"
[76] "g_seinen"	"g_shoujo"	"g_shoujoai"
[79] "g_shounen"	"g_shounenAi"	"g_sliceofLife"
[82] "g_space"	"g_sports"	"g_superPower"
[85] "g_supernatural"	"g_thriller"	"g_vampire"
[88] "g_yaoi"	"g_action"	"g_demons"
[91] "g_sciFi"	"g_unknown"	"g_misc"
[94] "rating_G"	"rating_PG"	"rating_PG13"
[97] "rating_R"	"rating_R_Plus"	"rating_Rx"
[100] "type_movie"	"type_music"	"type_ONA"
[103] "type_OVA"	"type_special"	"type_TV"