# Analysis of Twitch streamer success

class: STP 494 by Jonathan Franco

## Twitch Data Set

- **source:** Kaggle Twitch Data Set
- The data set consists of the top 1000 streamers on twitch and their statistics barring actual income
- **Question:** Does stream time (*as in time that a streamer is online streaming*) have a significant effect on the amount of watch time that a streamer gets?


Figure 1: twitch logo

## Question Methodology

- The reason this question is important is because it could show that if streamers put in the time streaming a lot and consistently then they could grow.
- If this is not the case, what other feature is effecting the growth and size of channels on the top 1000?
- Is it *pure luck* or a *meritocracy* for those on Twitch?

## Top 10 chart

Table 1: Top 10 Twitch Streamers and Stream Time

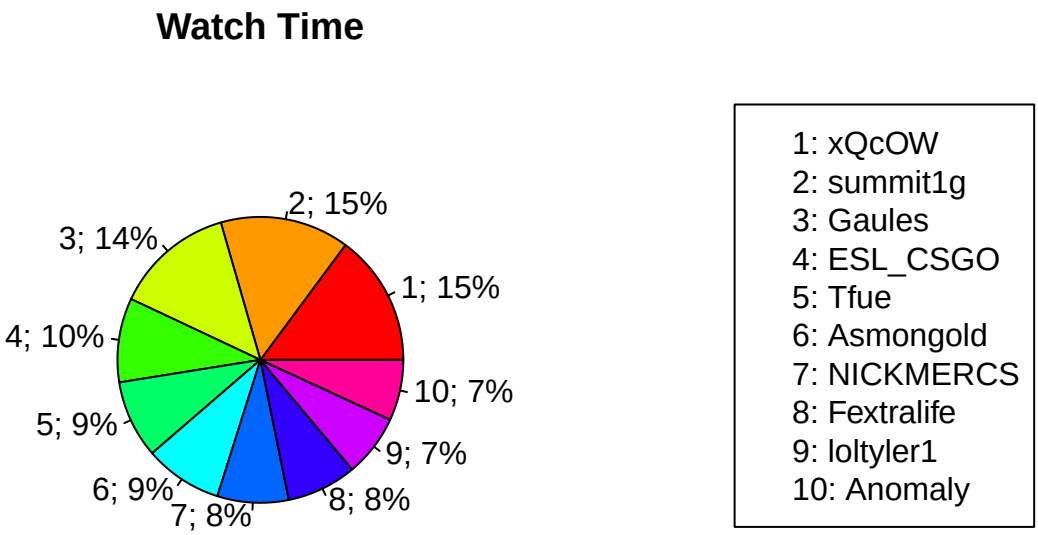
| Channel | Watch time(Minutes) | Stream time(minutes) |
|---|---|---|
| xQcOW | 6196161750 | 215250 |
| summit1g | 6091677300 | 211845 |
| Gaules | 5644590915 | 515280 |
| ESL_CSGO | 3970318140 | 517740 |
| Tfue | 3671000070 | 123660 |

## Top 10 (cont.)

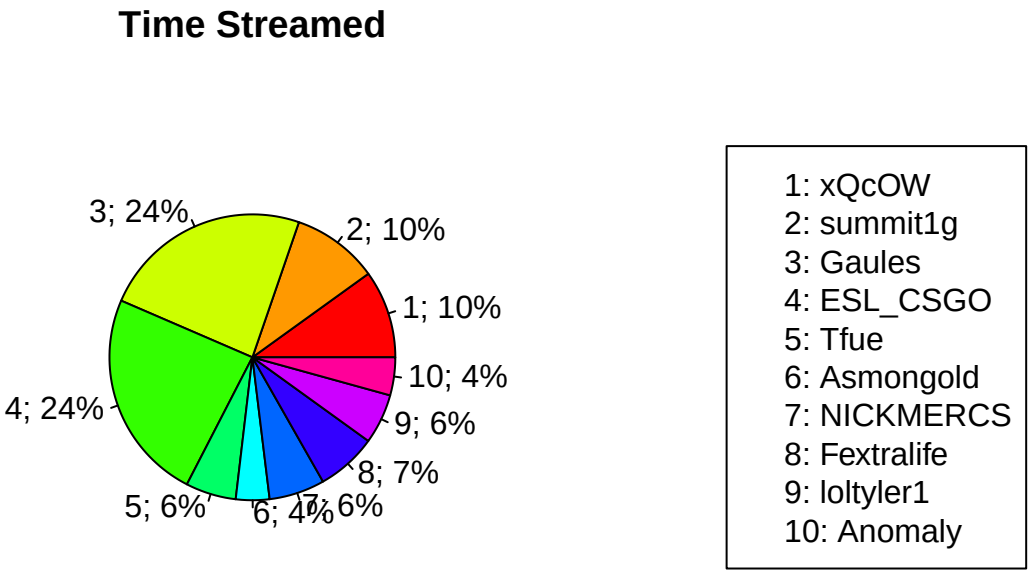
Table 2: Top 10 Twitch Streamers and Stream Time

| Channel | Watch time(Minutes) | Stream time(minutes) |
|---|---|---|
| Asmongold | 3668799075 | 82260 |
| NICKMERCS | 3360675195 | 136275 |
| Fextralife | 3301867485 | 147885 |
| loltyler1 | 2928356940 | 122490 |
| Anomaly | 2865429915 | 92880 |

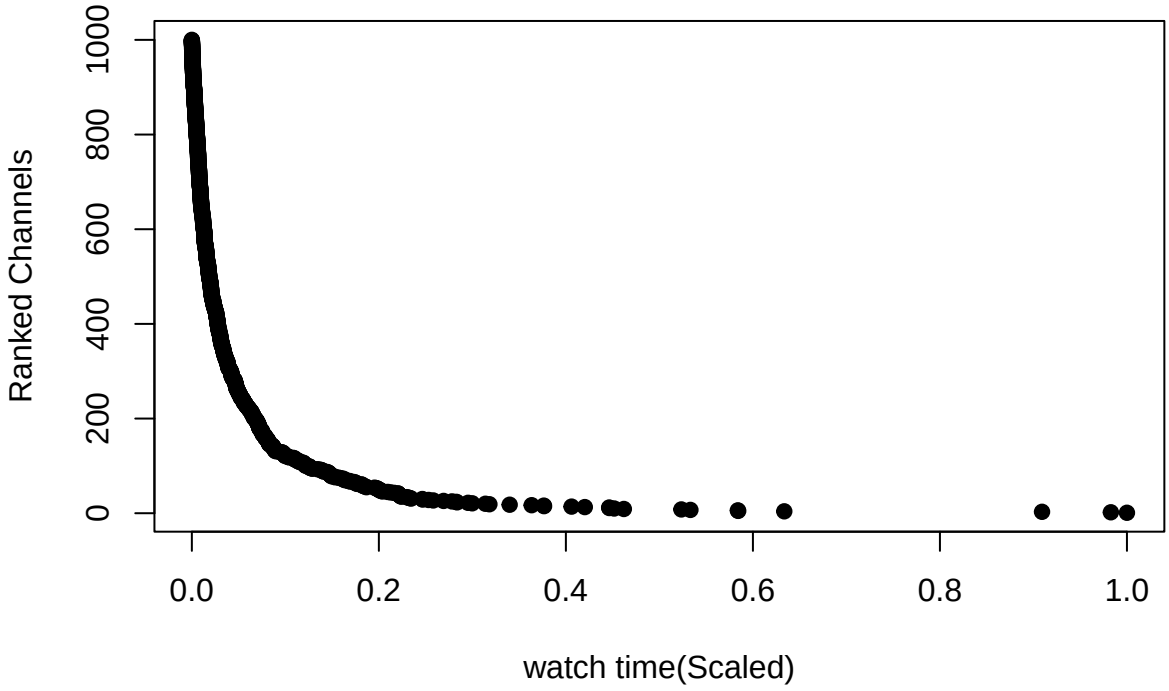
## Top 10 watch time



**Top 10 time streamed**



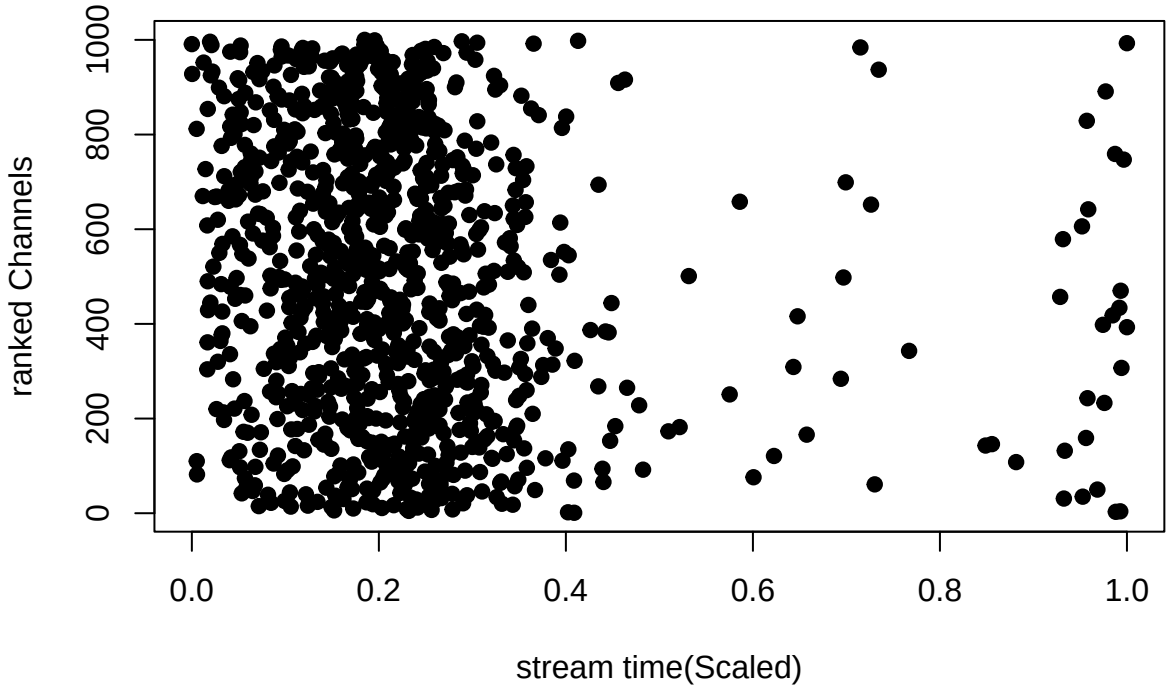
### Distribution of watch time among top 1k




### Differences in the top 1k

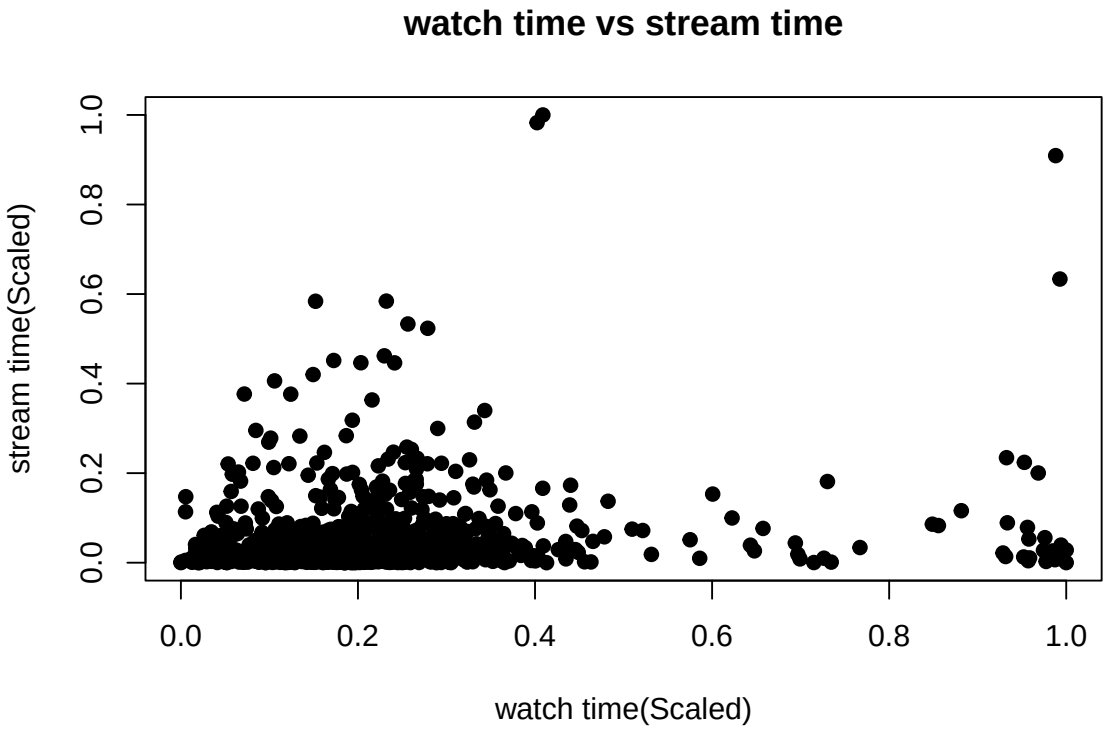
- The previous chart showed the major differences between watch time of the top 1k channels, but to put it in to better perspective allow us to look at #1 and #1000
- xQcOW (#1) has 51 times the amount of watch time that Remx (#1000) has
- But, he only has 2 times the stream time

### Distribution of stream time among top 1k




### Differences in the top 1k (cont)

- This new chart of stream times shows that there is a *bit* of a trend towards larger streamers streaming for longer but it's not very significant and as you can see below its not a good metric for predicting success




### top 5 ratios

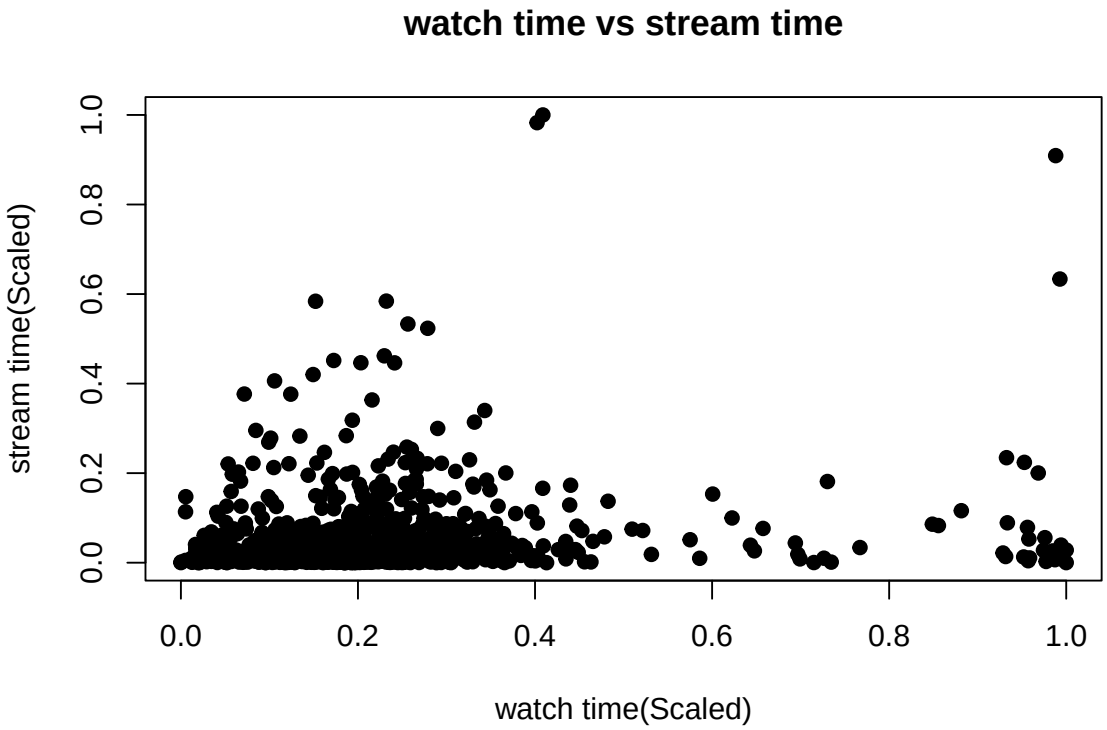
- In this column of watch time vs stream time its clear that the ratio varies wildly without any extra visualizations

Table 3: Top 10 Twitch Streamers w/ ratio (watch time/stream time)

| Channel | ratio |
|---|---|
| xQcOW | 28785.885 |
| summit1g | 28755.351 |
| Gaules | 10954.415 |
| ESL_CSGO | 7668.556 |
| Tfue | 29686.237 |

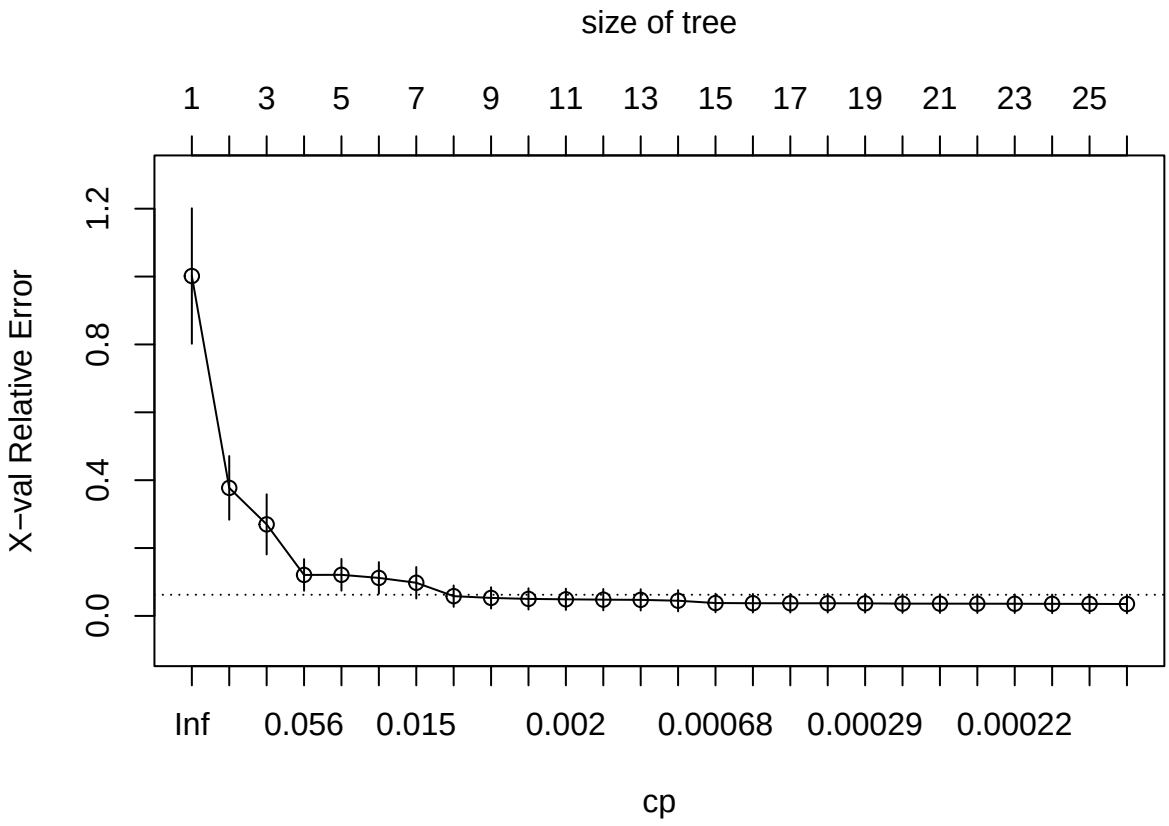
### predictions

- This new chart of stream times shows that there is a *bit* of a trend towards larger streamers streaming for longer but it's not very significant and as you can see below its not a good metric for predicting success

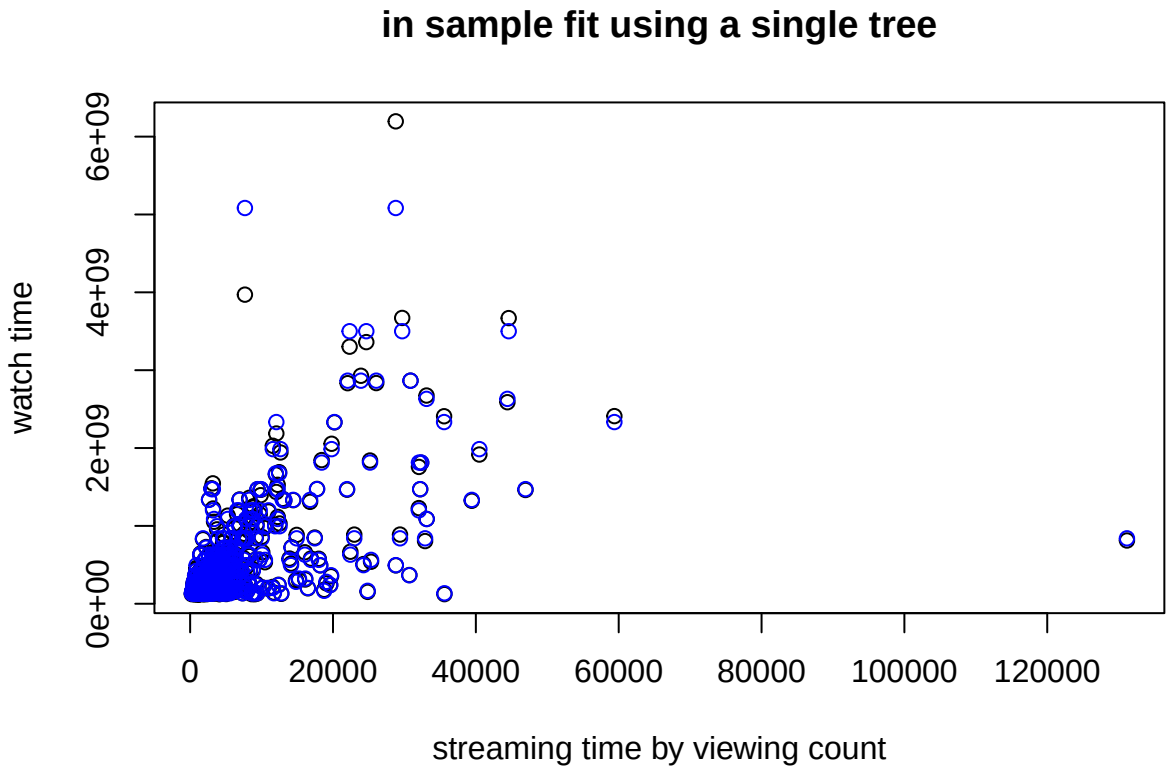


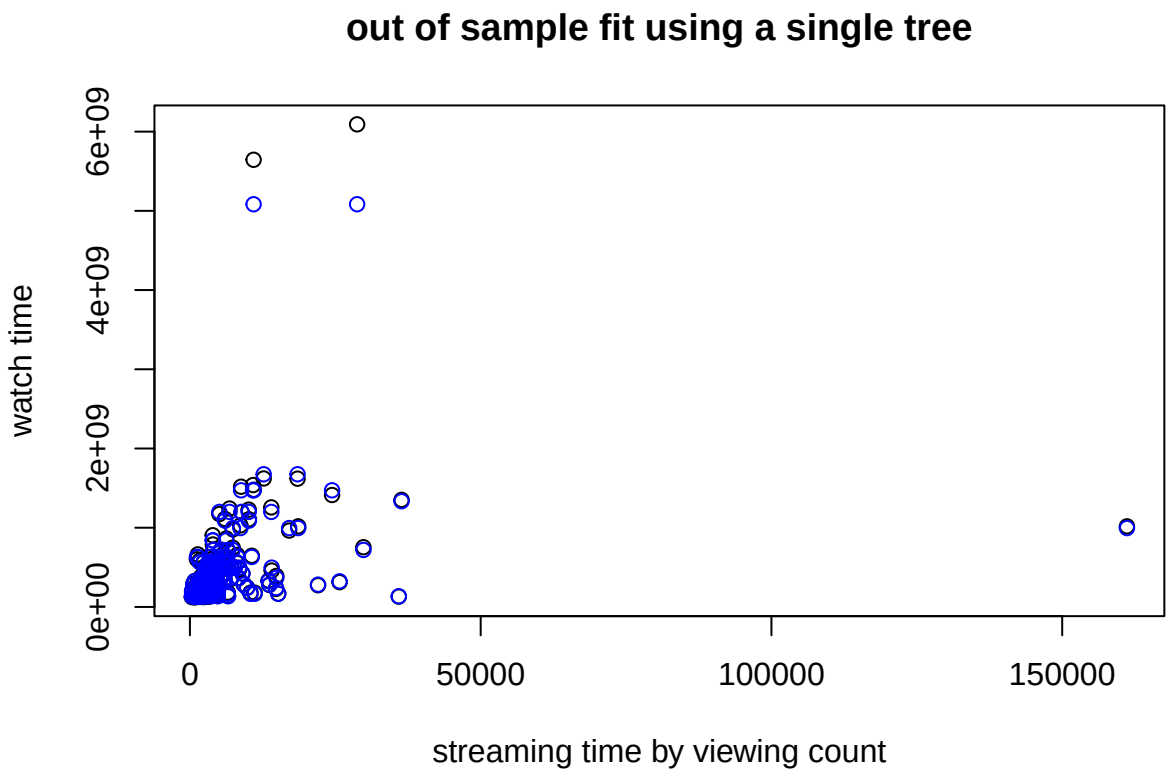

**Using Lasso, trees, and boosting to find any sort of correlation between variables**

find what makes a good streamer using all possible data

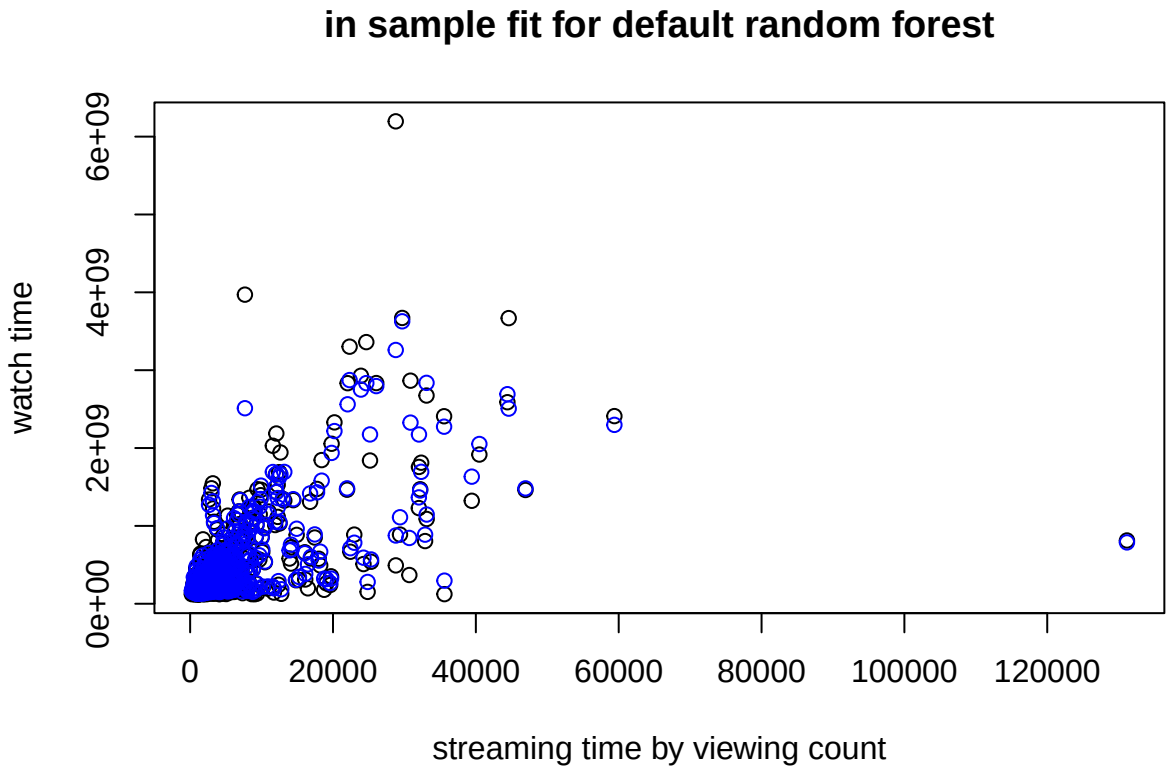


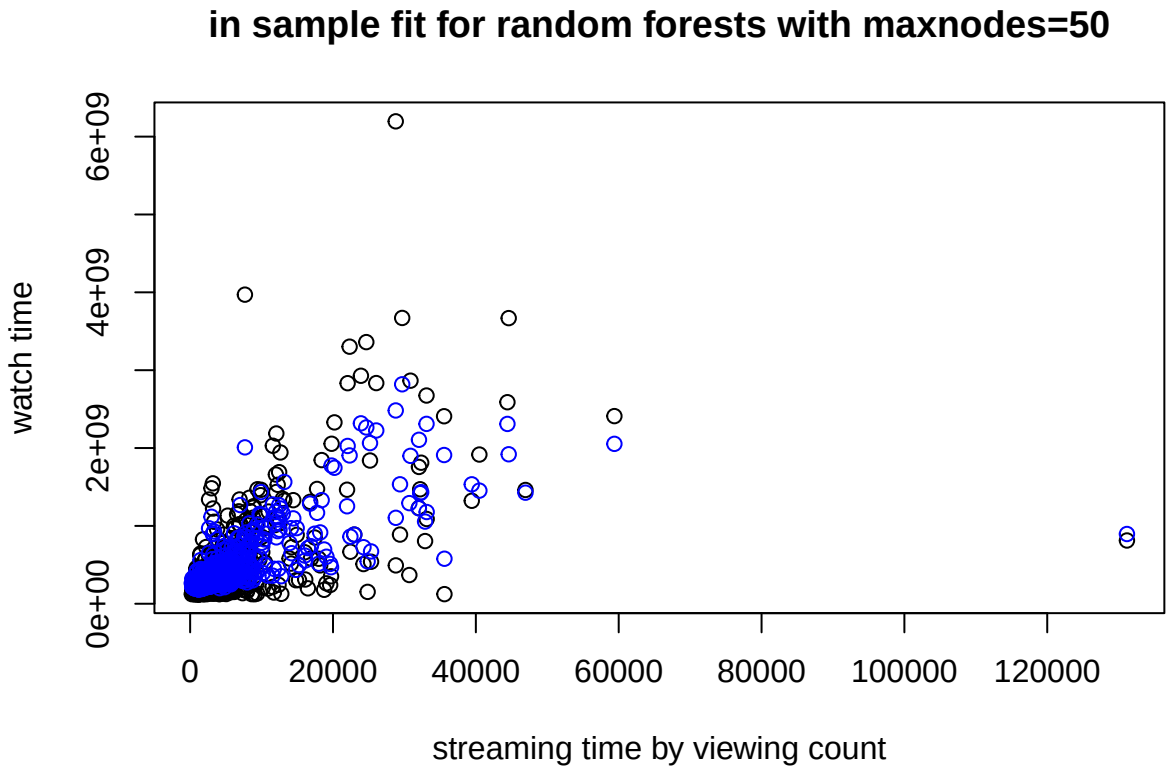

```
## size of best tree:  26
```

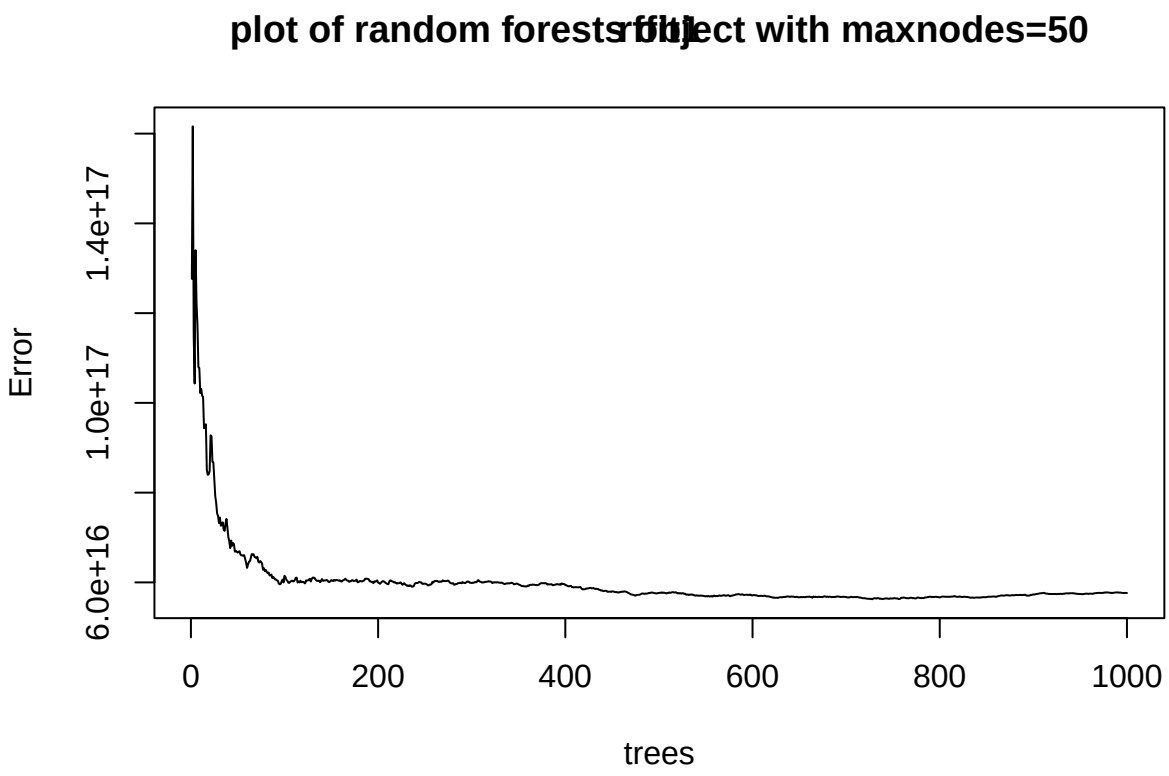

in sample fit using a single tree
watch time
0e+00
2e+09
4e+09
6e+09
0
20000
40000
60000
80000
100000
120000
streaming time by viewing count




```
## test rmse for tree:  75019246
```

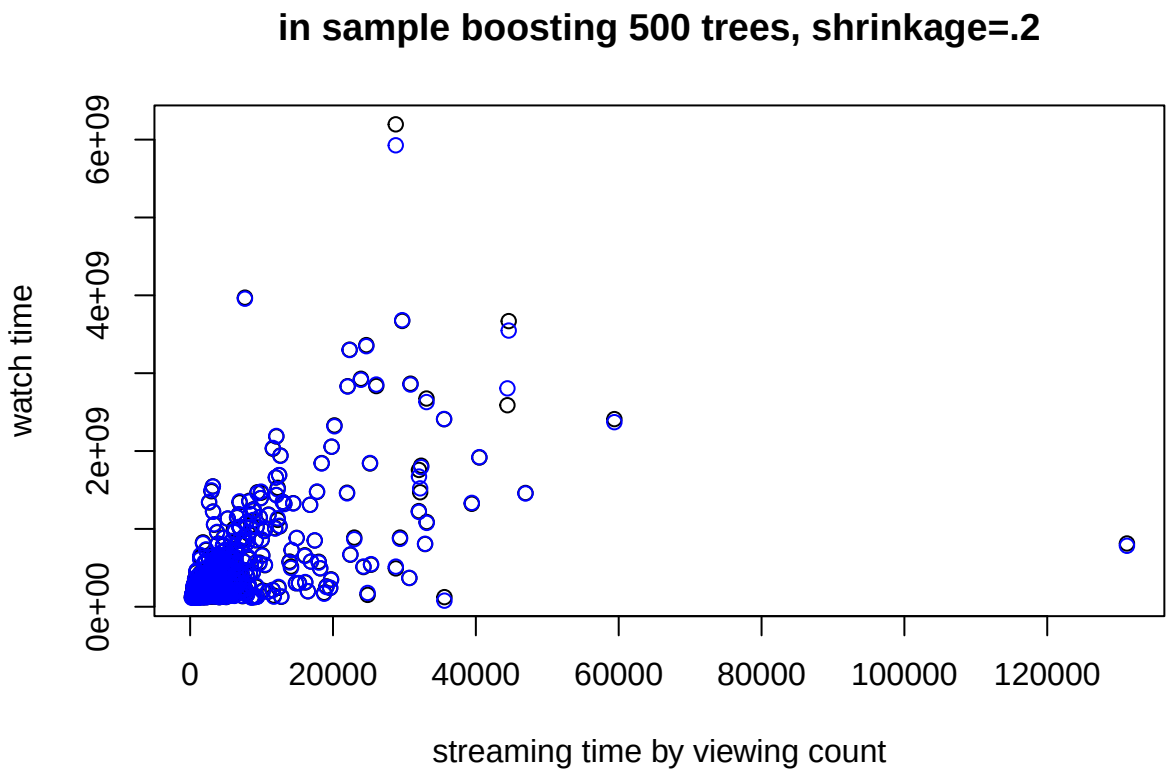

in sample fit for default random forest
watch time
0e+00
2e+09
4e+09
6e+09
0
20000
40000
60000
80000
100000
120000
streaming time by viewing count

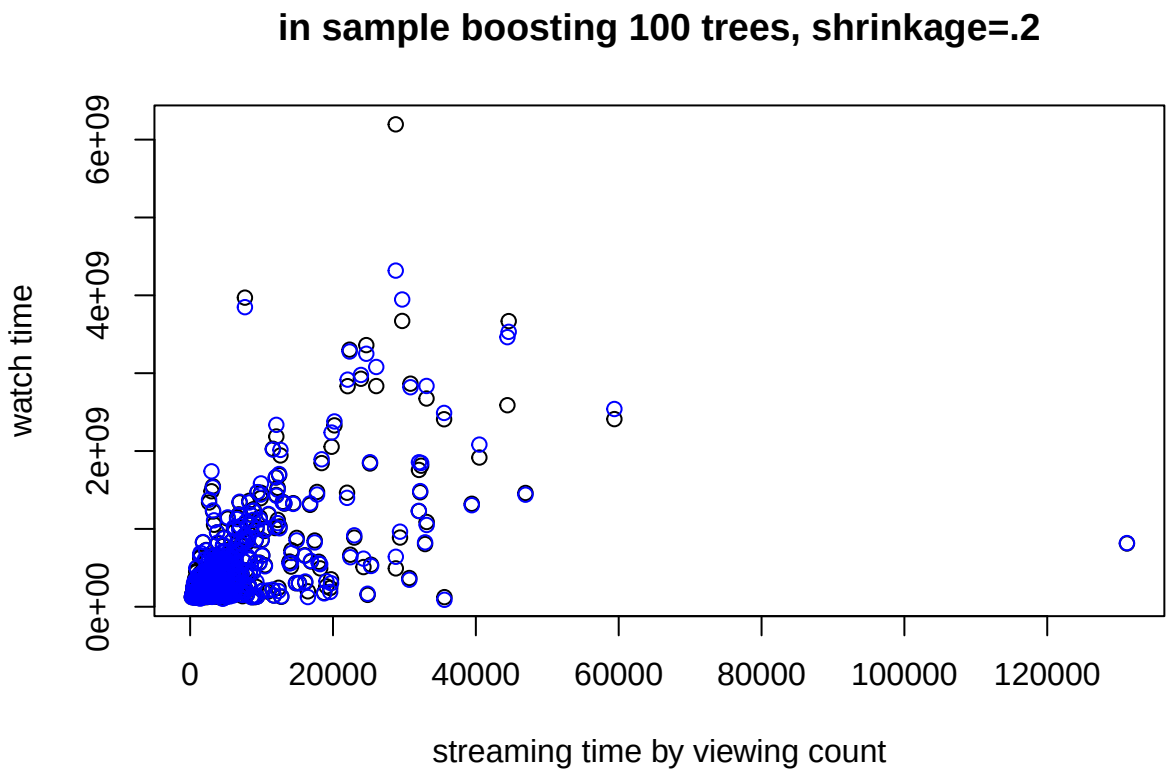

in sample fit for random forests with maxnodes=50
watch time
0e+00
2e+09
4e+09
6e+09
0
20000
40000
60000
80000
100000
120000
streaming time by viewing count




```
## rmse for random forests:  310801289
```

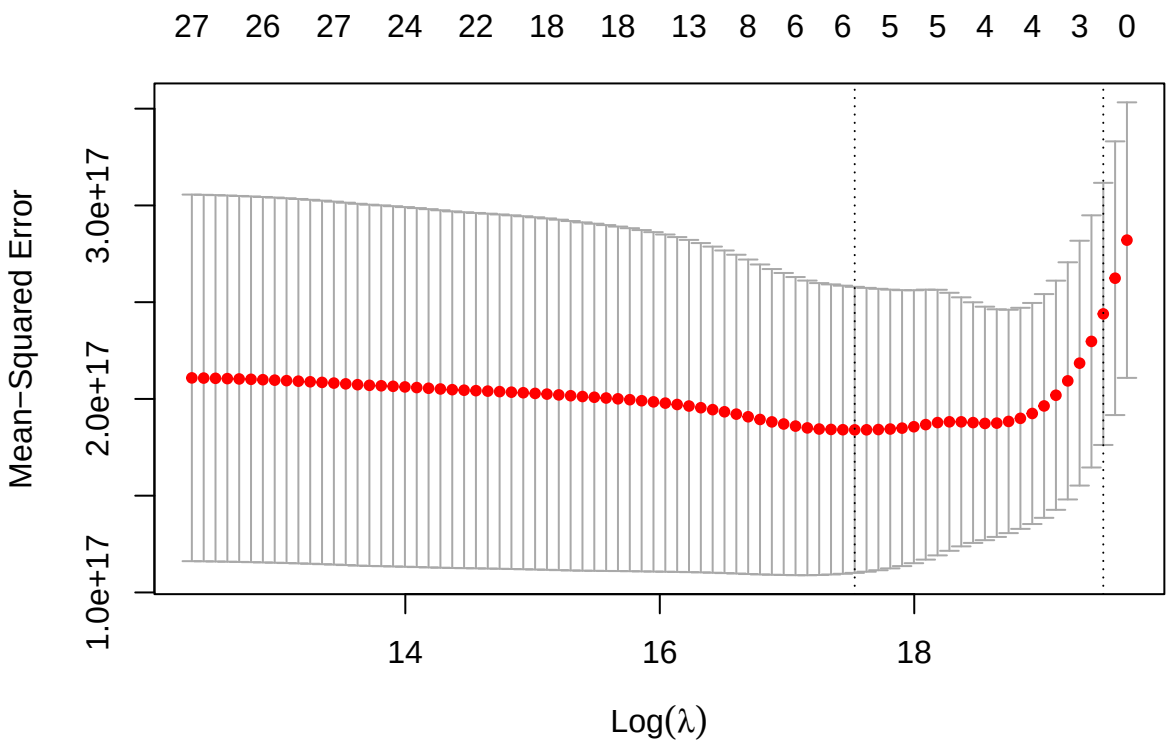
```
## [1] "Trees has led to nothing particularly interesting, and has shown little in terms of what makes a
```


in sample boosting 500 trees, shrinkage=.2
watch time
0e+00
2e+09
4e+09
6e+09
0
20000
40000
60000
80000
100000
120000
streaming time by viewing count



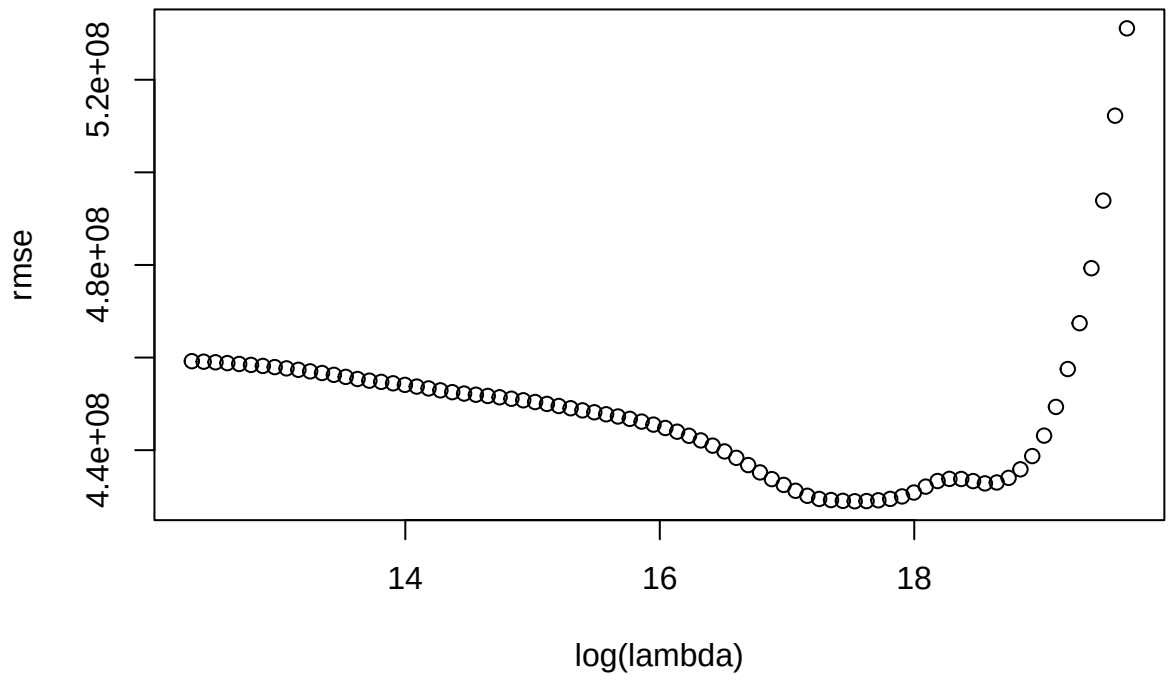

```
## boosting rmse with mileage:  183098734
```

```
## [1] "Boosting has also failed to show anything of significance other than its looking to be more luck
```

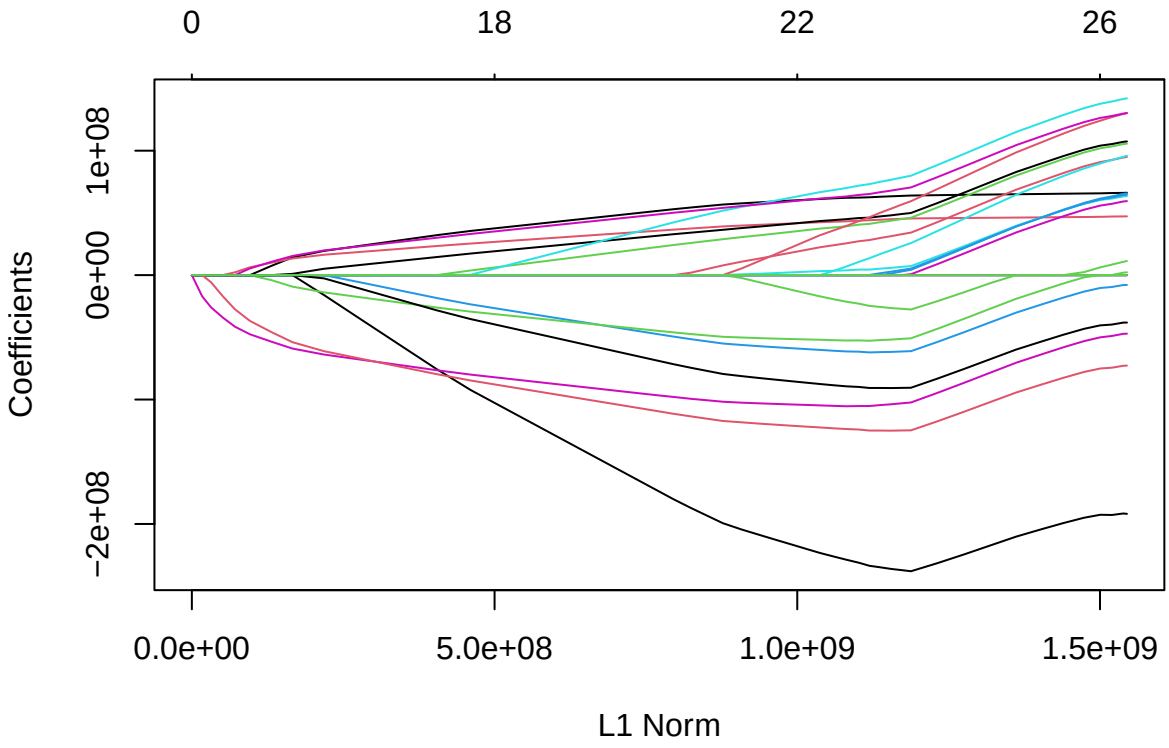



```
## lambda min:  41117244
```

```
## lambda 1se:  290074029
```




```
## min rmse:  428968507
```




```
## out of sample rmse:  485218843
```

```
## [1] "Terrible Rmse, which shows that atleast my data can not show what makes a streamer"
```

### In conclusion

- There is no significant correlation between watch time and time streamed
- Does that mean that the success of streamers is completely up to luck?
- No, but it could. The watch time seems to be a function of avg viewers and time streamed and that combination of the two varies from channel to channel
- each channel is different some are independent streamers that may stream a little but garner a large audience others are large organizations that can afford to stream all the time
- I think that if anything a more important question to ask would be if watch time should be considered a metric of success when something like income generated could be more tangible and involves amount of ads streamers run as well the income generated from stream time and large audiences
- so, if I could get data for income next time I would have to pull add frequency as well as these watch times to see if they correlate
- The machine learning aspect also showed no results other than I either need to pull more and better data or that it really is all luck when it comes to success on Twitches platform