ROTTEN TOMATOES WEB SCRAPING

By Marissa Joy

WHY?

- Uncertainty of what people want

- Do critic or audience score give a better prediction of how much profit a movie will make?

DATA

- Title
- Box Office Profits
- Ranking
- Audience Rating
- Audience Review Count
- Critic Rating
- Critic Review Count
- Genre
- Runtime
- Rating

REGRESSION SUMMARY

```
-157410844   -33534468    -5131019    21508171   554294564

Coefficients:
                       Estimate Std. Error t value            Pr(>|t|)
(Intercept)            97626079    7469141  13.071 < 0.0000000000000002 ***
ranks                  -8740478    1897627  -4.606    0.000004477746069 ***
a_count                15526813    1542897  10.063 < 0.0000000000000002 ***
a_score                 7723462    2053115   3.762             0.000176 ***
c_count                32778693    1930473  16.980 < 0.0000000000000002 ***
c_score               -20326920    2426247  -8.378 < 0.0000000000000002 ***
genre.Art             -53822511    5673612  -9.486 < 0.0000000000000002 ***
genre.Comedy          -29884230    4693618  -6.367    0.000000000260745 ***
genre.Documentary     -48205304    6631031  -7.270    0.000000000000597 ***
genre.Drama           -42981349    4475680  -9.603 < 0.0000000000000002 ***
genre.Kids             10729656   23142501   0.464             0.642982
mp_rating.NR          -12624081    9162421  -1.378             0.168482
mp_rating.PG            5721879    8363903   0.684             0.494015
mp_rating.PG.13         -266383    8224971  -0.032             0.974168
mp_rating.R           -34272100    7802484  -4.392    0.000012051285977 ***
runtime                10653181    1707399   6.239    0.000000000580697 ***
monthJan              -24199355    7001045  -3.457             0.000564 ***
monthOct              -20802734    5054601  -4.116    0.000040870243890 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59820000 on 1400 degrees of freedom
Multiple R-squared:  0.4701,    Adjusted R-squared:  0.4636
F-statistic: 73.05 on 17 and 1400 DF,  p-value: < 0.00000000000000022
```
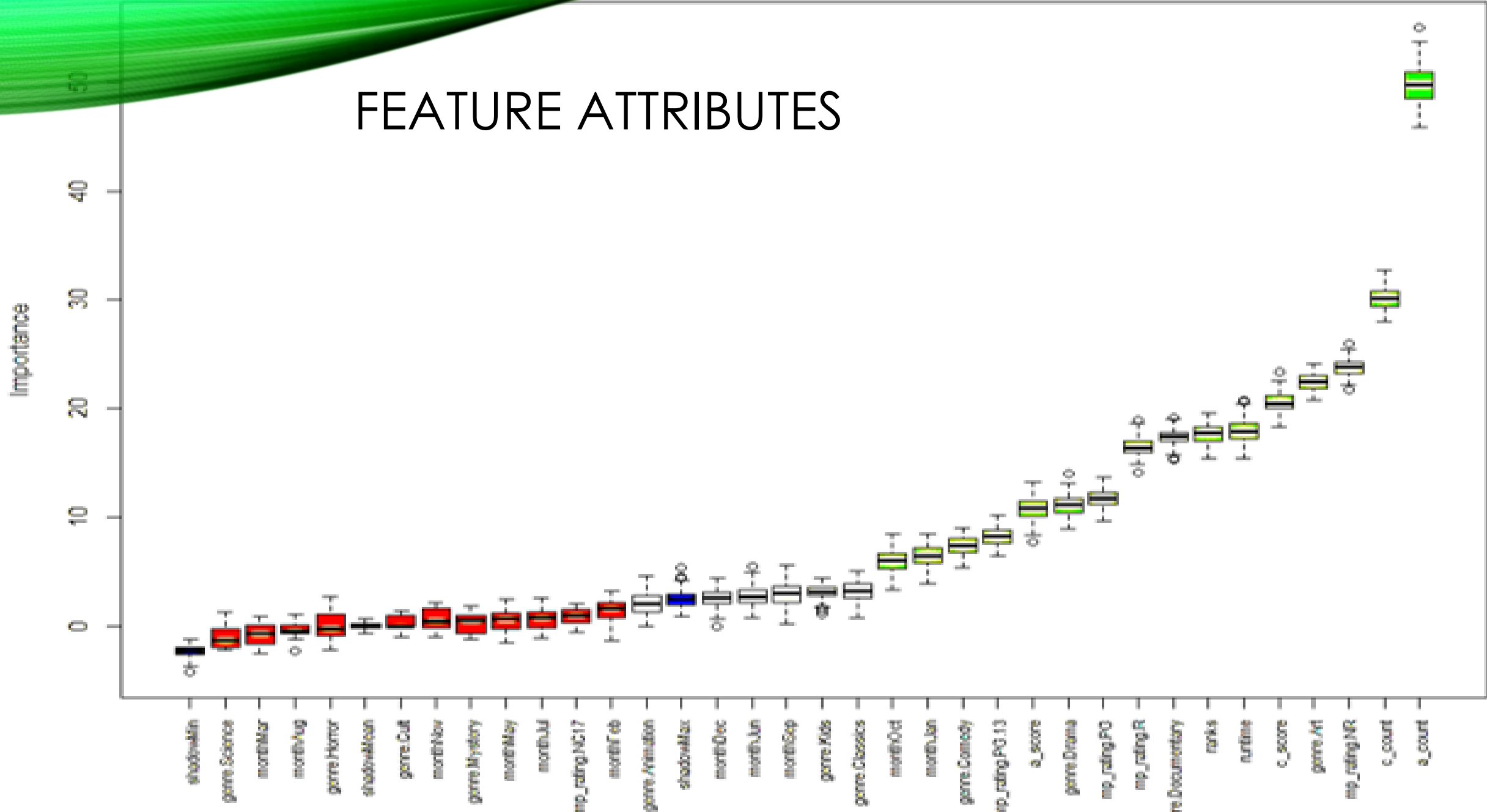
FEATURE ATTRIBUTES

MOST IMPORTANT

Linear Model Important Attributtes