Analysis of movies data

School: SJTU Name: Wuhao ID: No.67

1. What data do we have?

- title: Title of the movie
- year: Year of release
- budget: Total budget (if known) in US dollars
- length: Length in minutes
- rating: Average IMDB user rating
- votes: Number of IMDB users who rated this movie
- r1-10: Multiplying by ten gives percentile (to nearest 10%) of users who rated this movie a 1
- mpaa: MPAA rating
- action, animation, comedy, drama, documentary, romance, short: Binary variables representing if movie was classified as belonging to that genre

2. Which data do I focus on?

2.1 My plan

In my opinion, the budget and the rating would change with time passing by. Therefore, I focus on the relationship with rating, budget and year. The results are shown on figure1 and figure2.

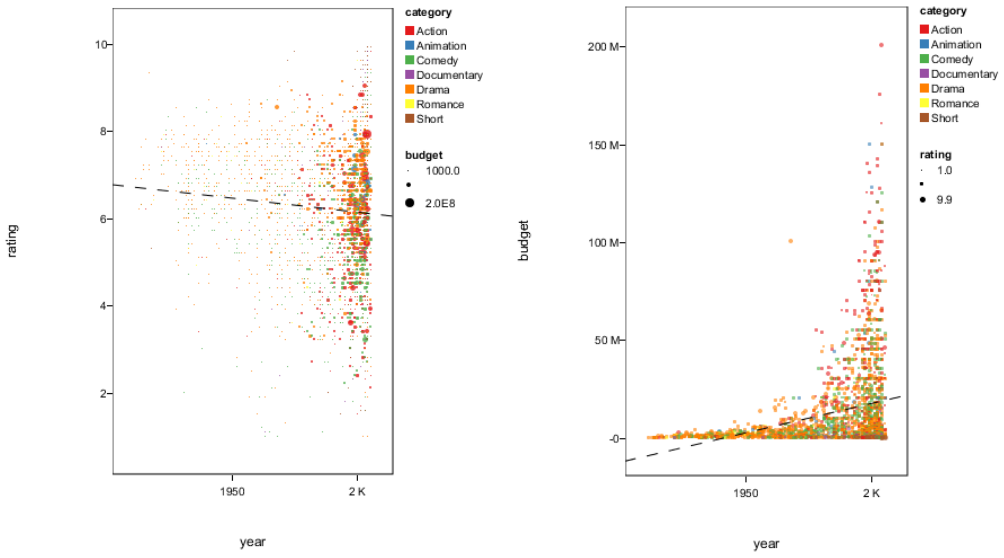

Figure 1

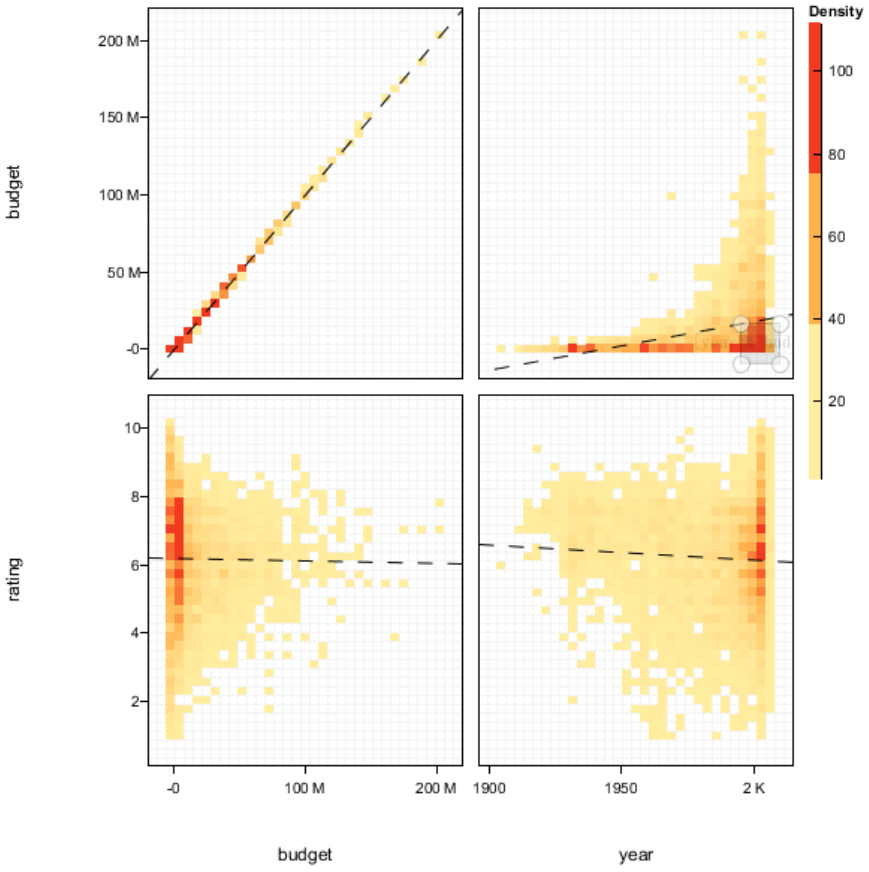

Figure 2

2.2 The details about figures

In the figure1, xAxis stands for year and yAxis stands for rating or budget. The size of circles represents it's budget or rating, the bigger, the higher budget or rating.

In the figure2, xAxis stands for budget and year each and yAxis stands for budget and rating each. The color bar represents the density of every square, the redder, the more crowded.

2.3 The result I find

In the figure1, I can find that the proportion of high-raking movies becomes lower and lower but the proportion of high-budget movies becomes more and more, which are mainly about action and drama as time goes by like "The lord of the kings"(9.0/$94,000,000) or "Spider man2"(7.9/$200,000,000).

And it shows that not all high-budget movies will get high raking like "Batman&Robin"(3.6/$110,000,000)

In the figure2, the second picture can clearly show the budget becomes higher in the recent years as well. And the third picture shows that the budget doesn't have so obvious relation with rating.

3. The wrong data!

I find an unusual point when I focus on the length of short as shown on figure3.

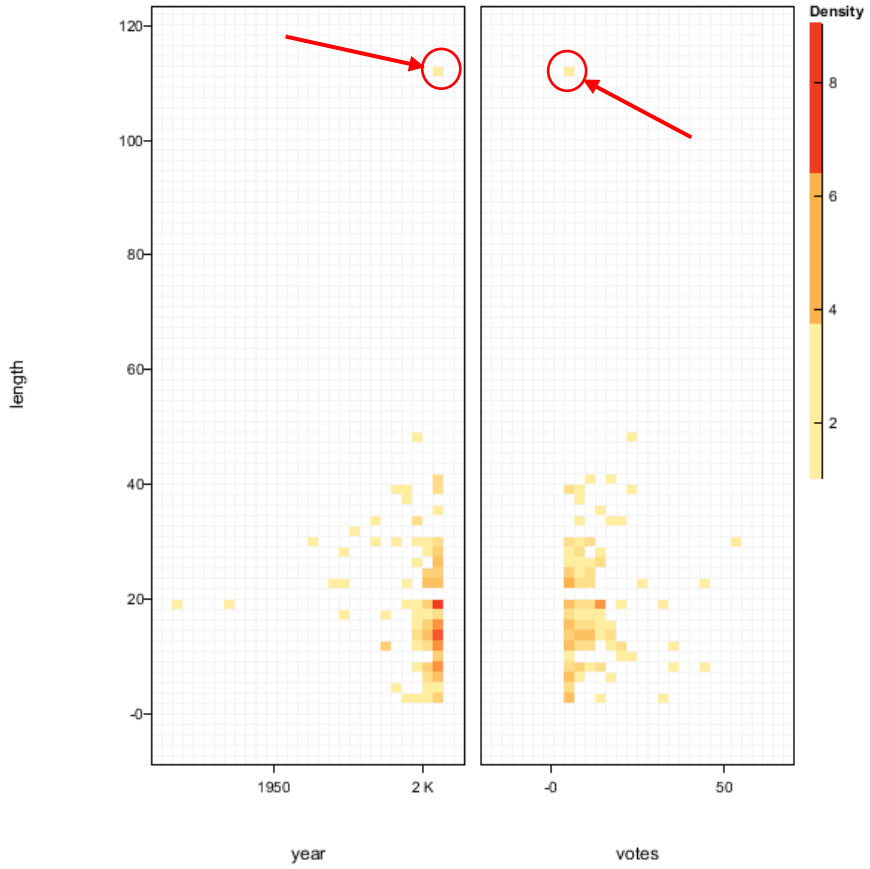

Figure 3

I search for this movie, entitled "inamorata", and released in 2009. Therefore, I found the length of this short should be 7mins rather than 112 minutes as shown in figure4!!!

Figure 4