

0.0.1 Question 2biv:

We analyzed the data but something seems odd. On a closer look, there are many negative values for `profit`. For example, the movie `102 Dalmatians` looks to have lost ~\$18M but it was a widely successful film! What may account for this issue?

```
In [46]: %%sql
SELECT *
FROM movie_budget
WHERE title = '102 Dalmatians'

* postgresql://jovyan@127.0.0.1:5432/imdb
  postgresql://jovyan@127.0.0.1:5432/postgres
1 rows affected.
```

```
Out[46]: [(85000000.0, 1635380, '102 Dalmatians')]
```

There might be a lot going on here but one might be that that children's films might have been more commonly watched on VHS or bluray disks back then which may not be taken into account here. It also might be that this was a film that was well received internationally rather than in the US which we have not taken account for.

0.0.2 Question 3c

What do you notice about the summary values generated in `earnings_summary`? Identify two properties about the histogram of the data.

Hint: Think in terms of about concepts from statistics like spread, modality, skew, etc. and how they may apply here.

I noticed that the mean is actually much higher than the median, and pretty close to the 75th percentile so the data is left skewed or skewed positively, and there might be some films that are making so much that they skew the data and these might be outliers. I also noticed that the range here is extremely wide, from around 30 to 750 million, so there must be many bins.

