# Lab 2: Exploration by visualization: the streaming movies dataset

Fill in your full name

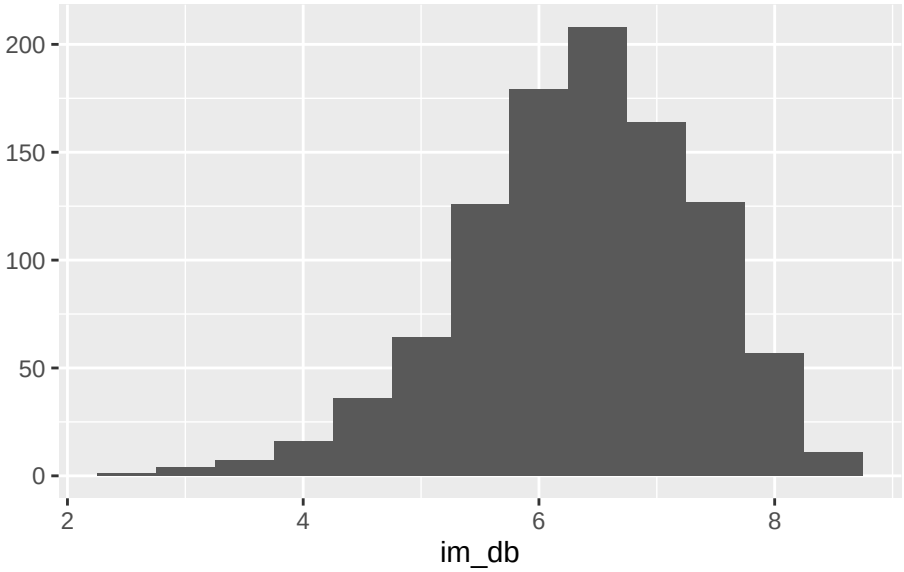
2022-06-29

---

## Visualization by example

### Exercise 1

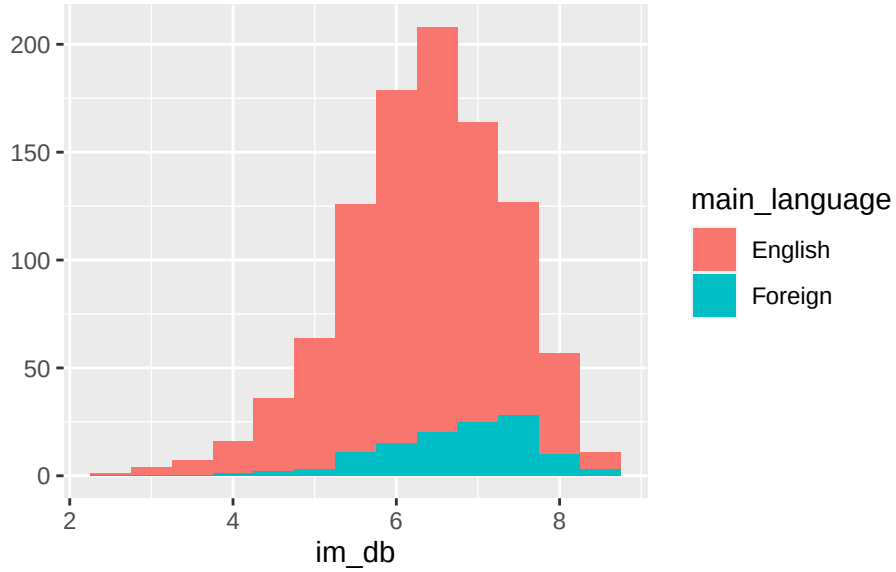
```
View(streaming)
qplot(x=im_db, binwidth=0.5, data=streaming)
```




- Question1 The View function shows whole data on Console window. The histogram with qplot function showed in plots window. The histogram shows the shape of distribution and density of im_db.

### Exercise 2

```
qplot(x=im_db, binwidth=0.5, fill=main_language, data=streaming)
```



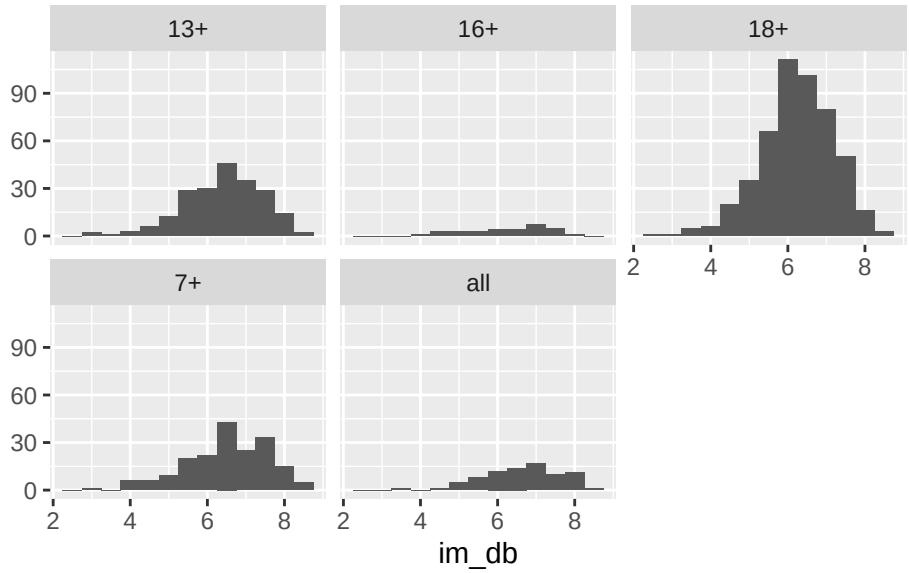

- Question1 It fill out the bar graph with color.
- Question2 Most of movies made by english.

**Exercise 3**

- Question1 English and Foreign IMDB shows left-skewed distribution, and centered around 6 to 8. The average between English and Foregin film have a substantial difference.
- Question2 Yes. It is hard to foray into American movie market as a Foreign film. So, only who has perfect stories, or great creativity foreign films can success and can be created by IMDB. So only the famous one can translated and rated. Plus, for the translating, it costs more than English movie. However, English movies have low obstacles to entry and rated, and even it is small films it is easy to rated. So the quality of movie is not guaranteed.

**Exercise 4**

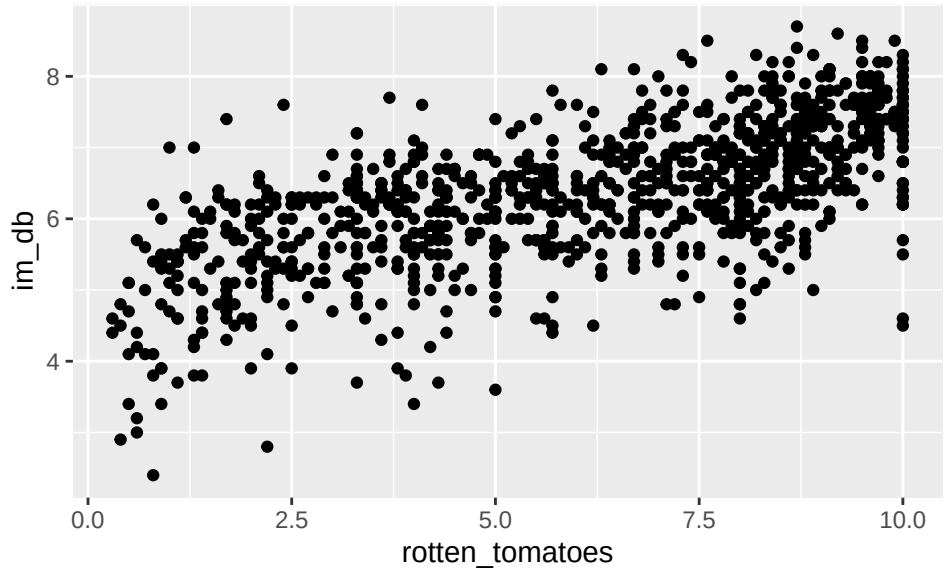
```
qplot(
  x = im_db,
  binwidth = 0.5,
  facets = ~ age,
  data = streaming)
```




- Question1 There are 5 facets.
- Question2 The facets represent IMDB rating depends on age(film rating).
- Question3 R rated movie distribution facet contains most movies.

**Exercise 5**

```
qplot(x=rotten_tomatoes, y=im_db, data=streaming)
```




- Question 1 When Rotten_tomatoes going up, The IMDB also rated high(positive, linear relationship).

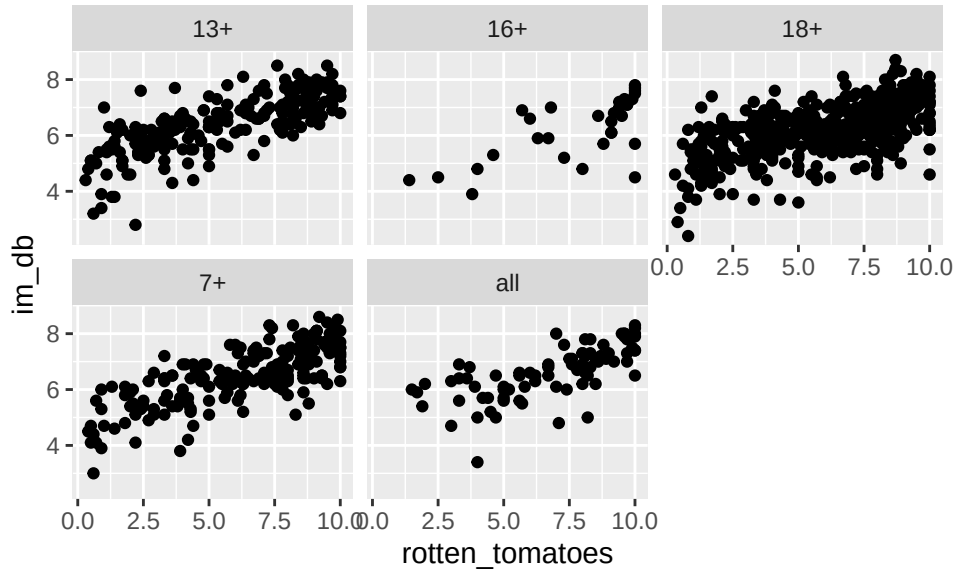
**Exercise 6**

```
qplot(x=rotten_tomatoes, y=im_db, data=streaming, color=main_language)
```


- Question 1 English movie scattered all over the rates, but Foreign films dense in right, upper quadrant.

**Exercise 7**

```
qplot(x=rotten_tomatoes, y=im_db, data=streaming, facets= ~age)
```

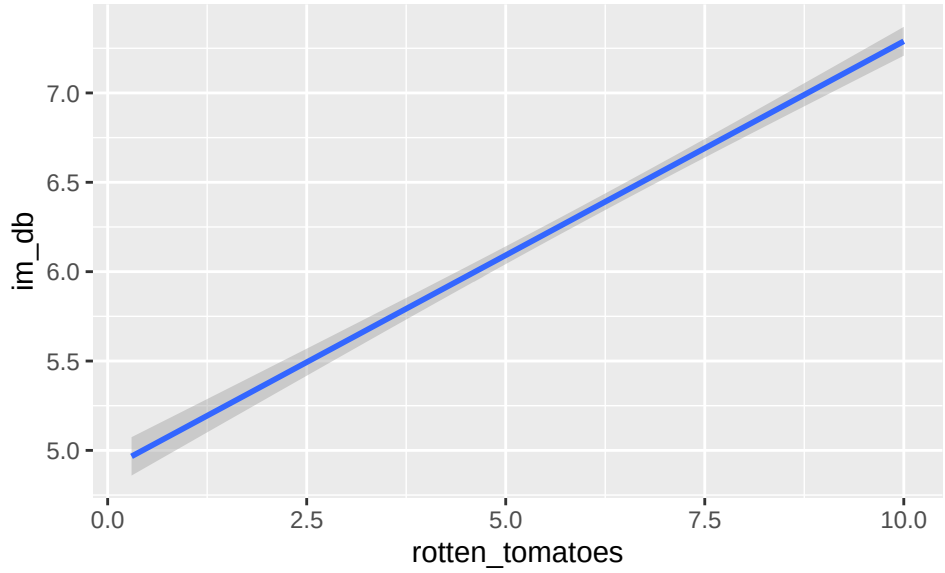


- Question 1 It shows same data, but divided by ages(film ratings).

**Exercise 8**

```
qplot(
  x = rotten_tomatoes,
  y = im_db,
  geom = "smooth",
  method = "lm",
  data = streaming
  )
```

```
## `geom_smooth()` using formula 'y ~ x'
```




- Question 1 It follows same linear trends.

**Exercise 9**

```
qplot(
  x = rotten_tomatoes,
  y = im_db,
  geom = c("smooth","point"),
  method = "lm",
  data = streaming
  )
```

```
## Warning: Ignoring unknown parameters: method
```

```
## `geom_smooth()` using formula 'y ~ x'
```