Lab 2: Exploration by visualization: the streaming movies dataset

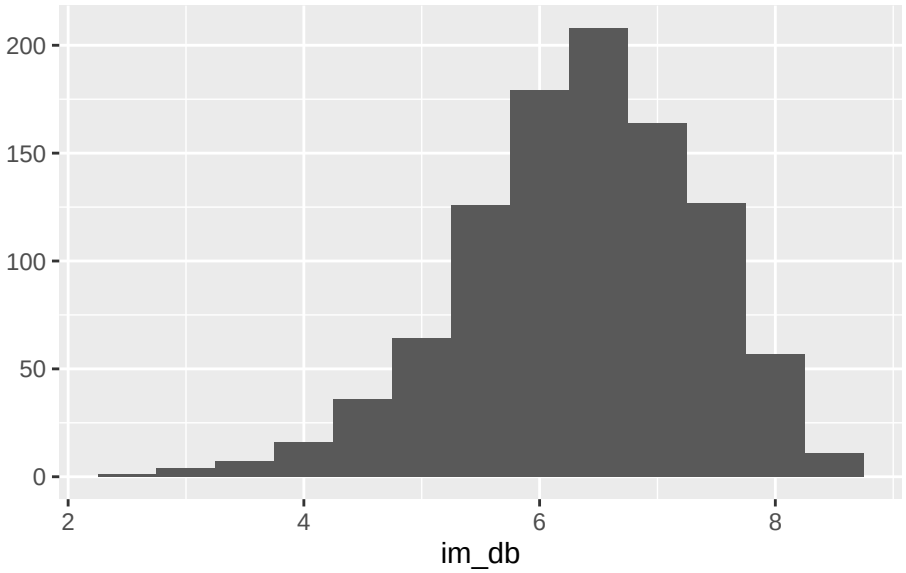
ji won mok

2022-02-15

Visualization by example

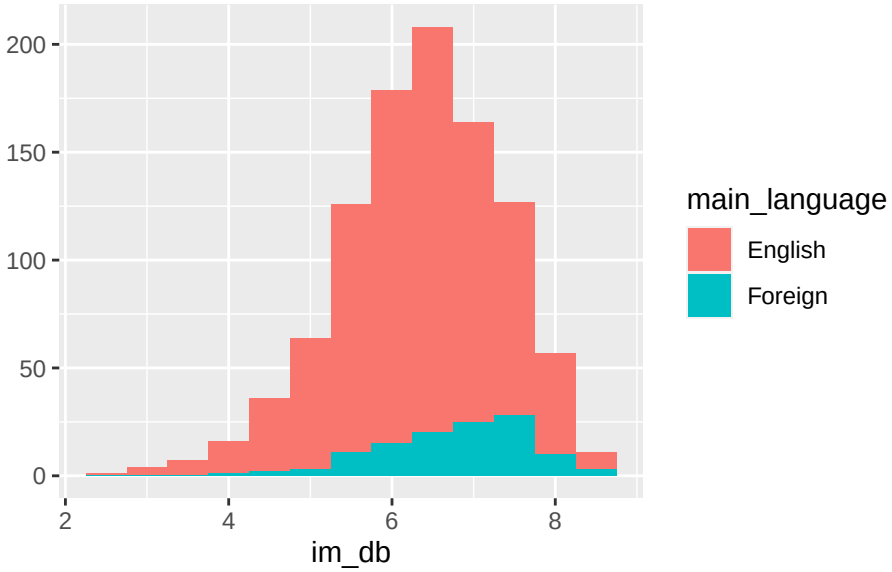
Exercise 1

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



Exercise 2

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



1) what did adding "fill = main_language" do? -> It fills the histogram depending on the categories, English and Foreign.
2) What language are most of the movies in this database? -> English and Foreign

Exercise 3

English and Foreign were skewed to the right.

1) Upon your visual inspection, does there appear to be a tangible difference in the average IMDB rating for these two distributions?

-> Yes, there is.

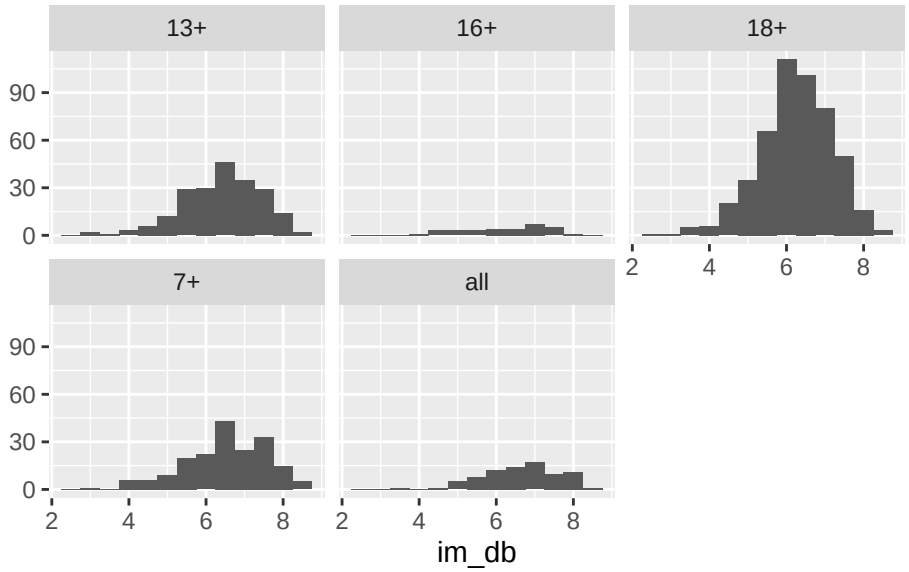
2) Based on your experience of watching movies in recent years, is this a result that you would have expected to see? (Explain why or why not.)

->

Yes. English-based films have been the movies that I have mostly watched.

Exercise 4

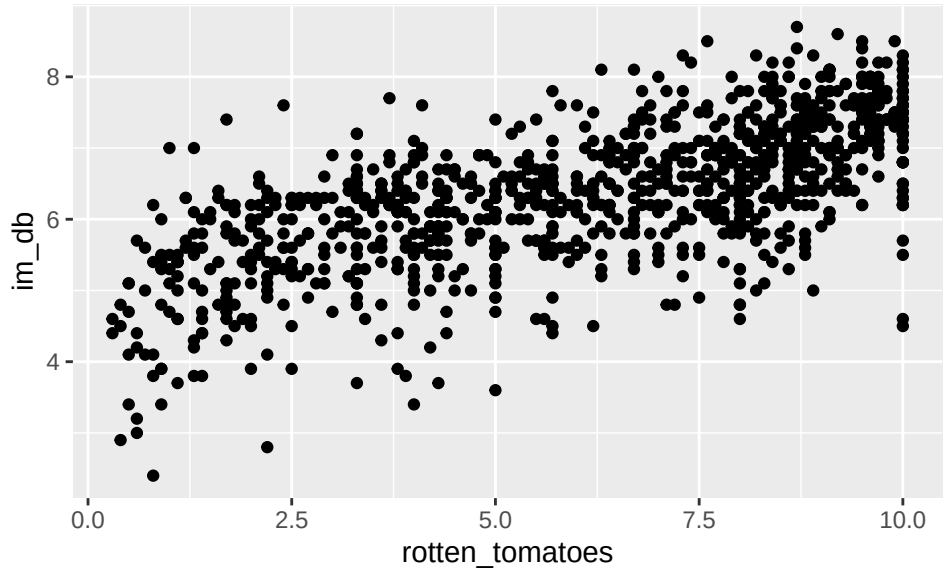
```
qplot(
  x = im_db,
  binwidth = 0.5,
  facets = ~ age,
  data = streaming
  )
```



1) How many facets are there? 5
2) What does each faceted sup-plot represent? sub-plot (Hint: what is the variable we are faceting over?)
3) Which facet's distribution contains the most movies? 18+

Exercise 5

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



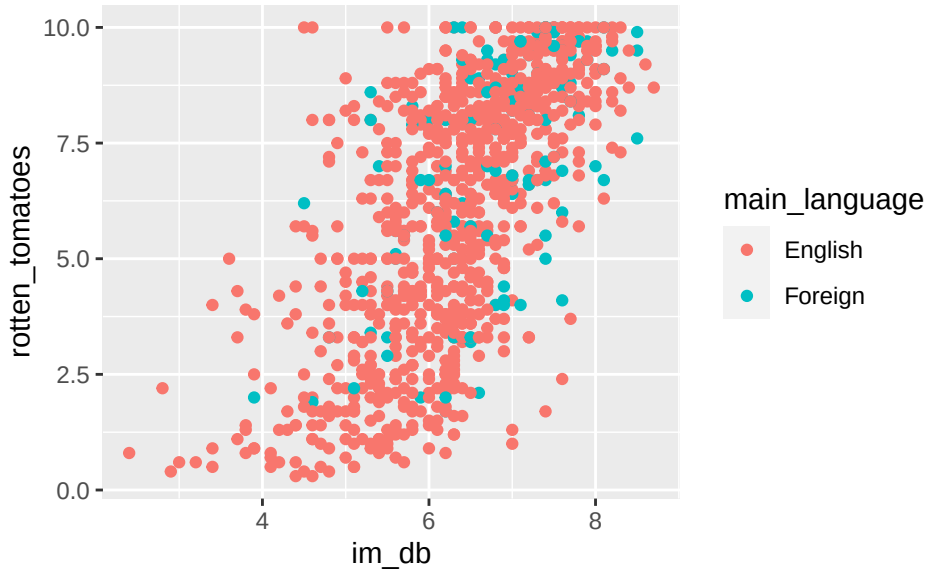
-> rotten_tomatoes and im_db have the positive relationship.

Exercise 6

1) What does this plot tell you about the relationship between the ratings and the main_language variable? (I.e. does the relationship between ratings on IMDB and Rotten Tomatoes look different for English and foreign language movies?)

```
qplot(
  y = rotten_tomatoes,
  x = im_db,
  binwidth = 0.5,
  color = main_language,
  data = streaming
  )
```

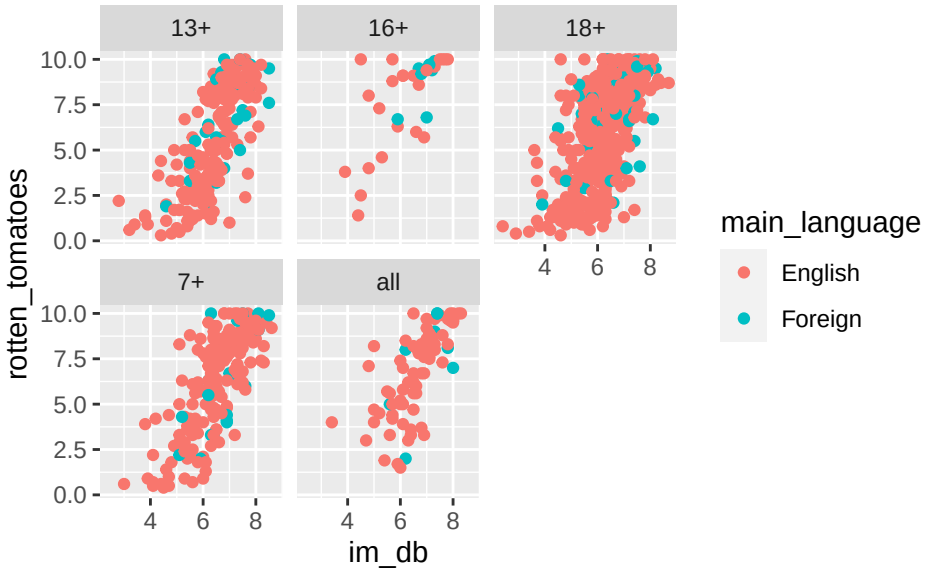
```
## Warning: Ignoring unknown parameters: binwidth
```



Exercise 7

```
qplot(
  y = rotten_tomatoes,
  x = im_db,
  binwidth = 0.5,
  color = main_language,
  data = streaming,
  facets = ~ age
  )
```

```
## Warning: Ignoring unknown parameters: binwidth
```



1. Is the information presented here any different from the information in Exercise 5?

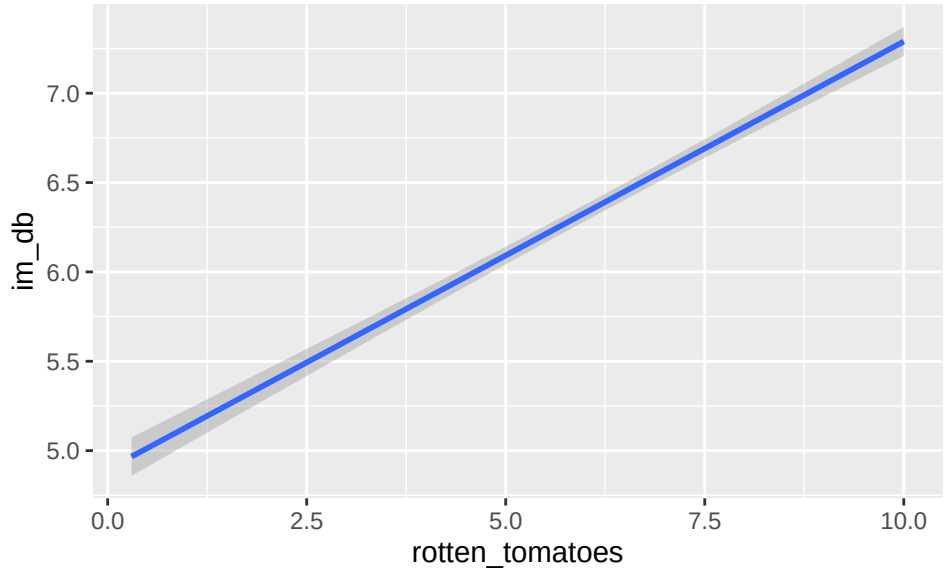
-> It clearly shows that the use of lanauge between english and foreign are significantly different.

(I.e. do any age categories show a different relationship between their ratings on IMDB vs Rotten Tomatoes?)

Exercise 8

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

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



1) Does it follow the trends (if any) you've previously described in the data?

-> Yes. im_db and rotten_tomatoes have positive relationship.

Exercise 9

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

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

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