

YouTube_Trending_Videos_Exploration_615_Final_Project

YuzheHuang

12/11/2019

Introduction

YouTube is one of the largest video hosting websites in the world, which has profound influences on the society in all aspects. Therefore, analyzing of YouTube's dataset become significant for advertisers and investors to analyze the social trends and make a prediction for future strategic business planning.

Therefore, the key objectives of the report are as follows:

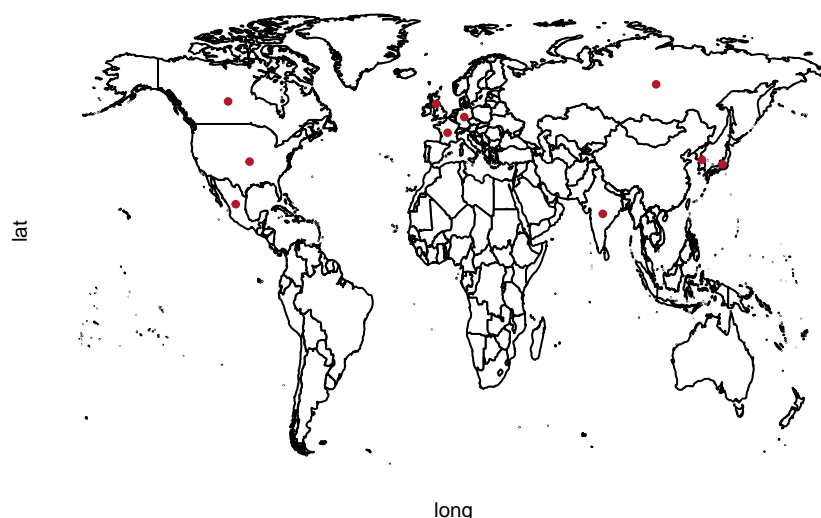
- Find the view patterns of YouTube videos by region;
- Select the category that have relatively higher business potential by constructing evaluation matrix;
- Explore the popular words used in the YouTube trending videos.

Data Description

Datasets

- The dataset includes daily trending YouTube videos of 10 countries: United States, United Kingdom, Germany, Canada, France, Japan, Korea, Mexico, Russia and India.
- Trending time interval: 2017.11-2018.6.
- The variables used contain: video id, trending date, category id, category name, publish time, views, likes, dislikes, comment count.
- Source: <https://www.kaggle.com/datasnaek/youtube-new#header>

Distribution of Countries in Dataset



Data Processing

- Drop null values and corrected time format;
- Pair the category ids with their names for further analysis.

Exploratory Data Analysis

1. View pattern

1.1 Top 5 Most Viewed YouTube Categories (by country)

- Entertainment videos are most viewed in six out of ten countries (Figure 1-6), followed by music videos dominated both in UK and U.S. (Figure 7,8).
- Indian viewers like auto and vehicle videos best (Figure 9), while French prefer game streaming (Figure 10).

Figure 1. Korea

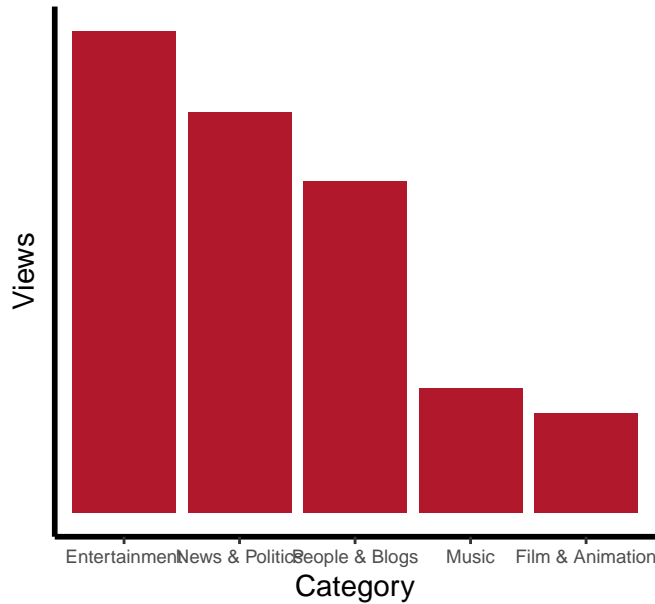


Figure 2. Russia

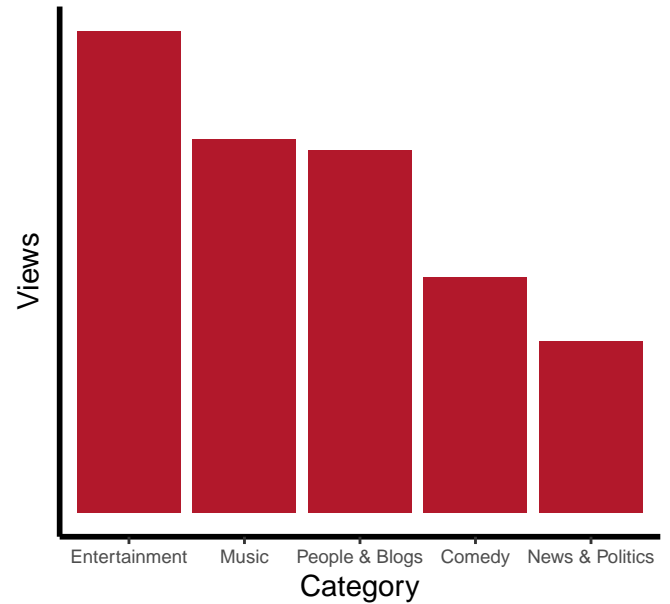


Figure 3. Germany

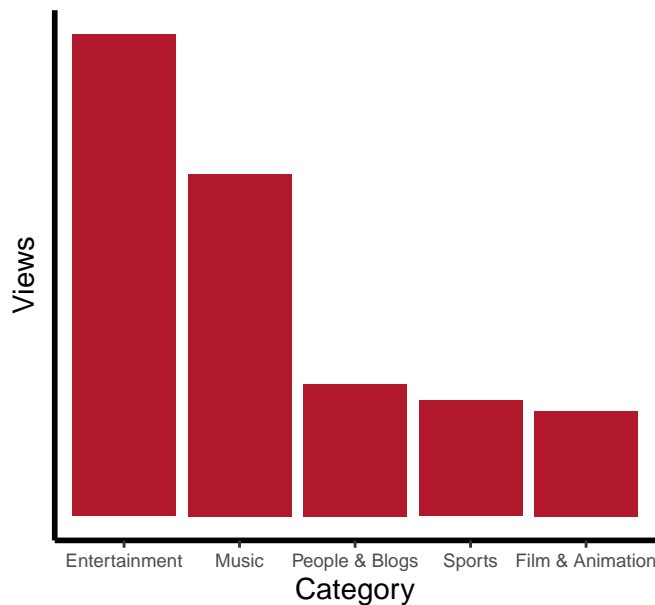


Figure 4. Mexico

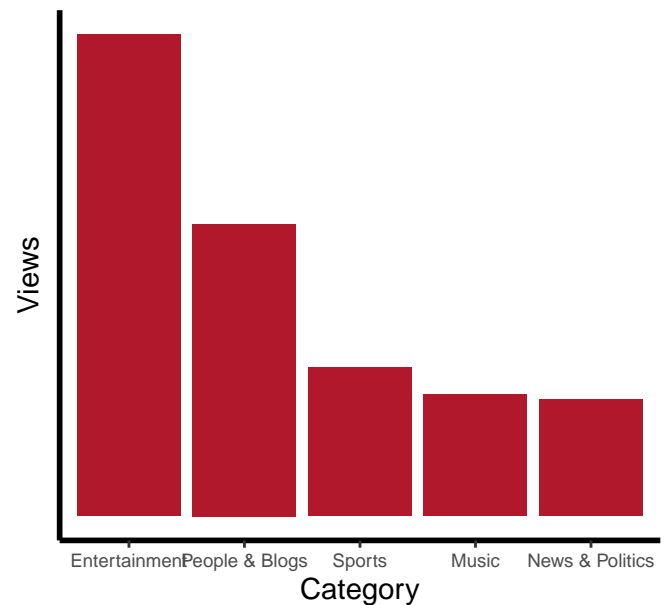


Figure 5. Japan

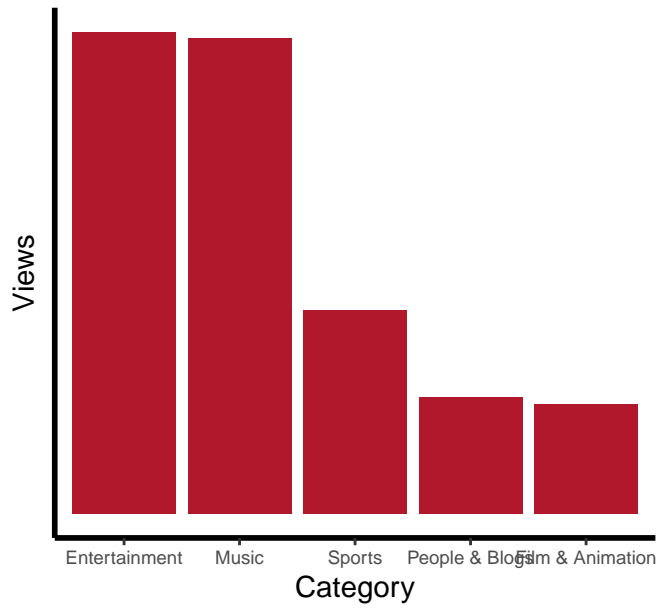


Figure 6. Canada

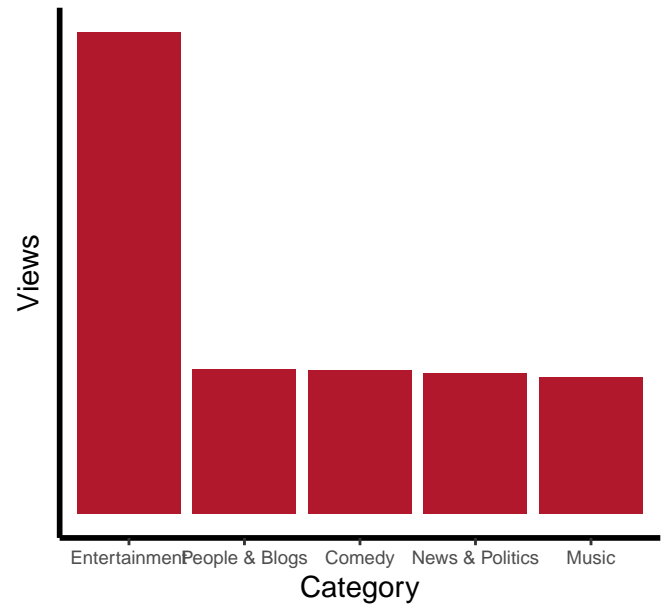


Figure 7. United Kingdom

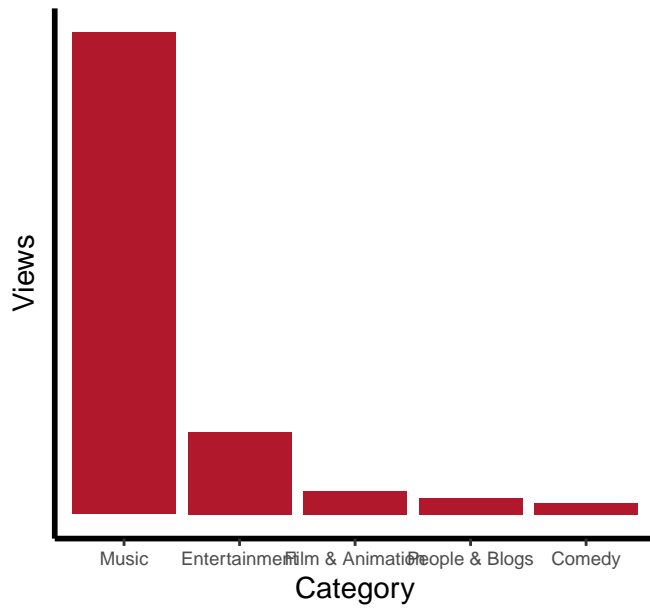


Figure 8. United States

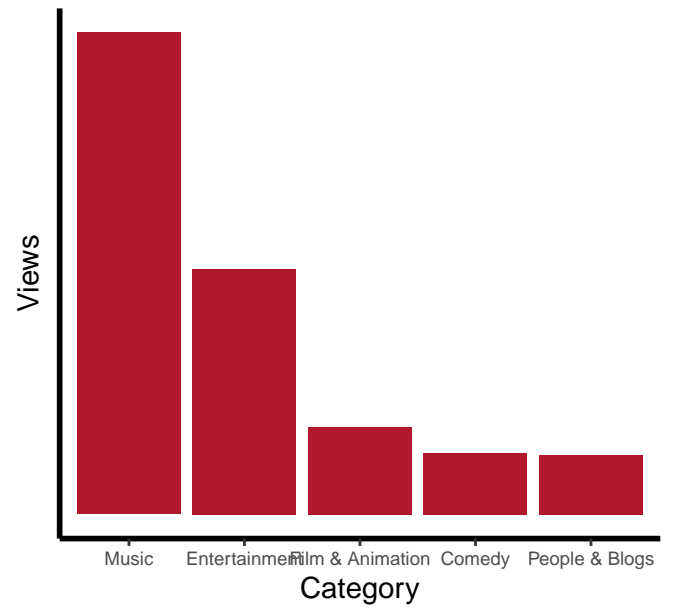


Figure 9. India

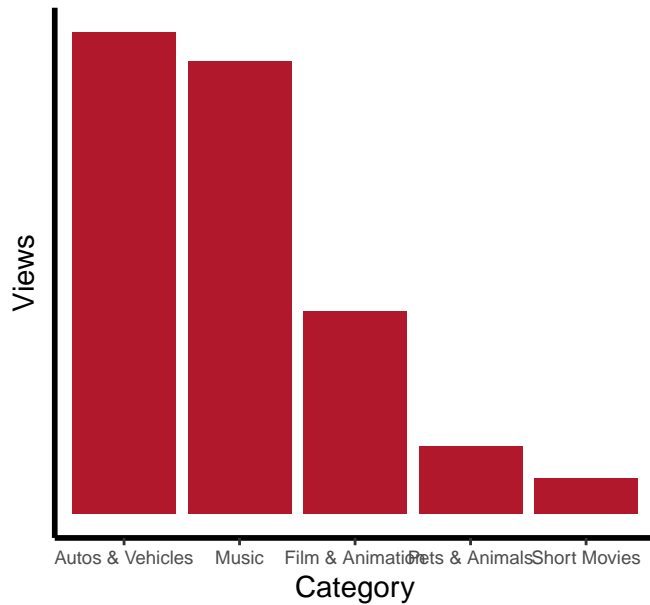
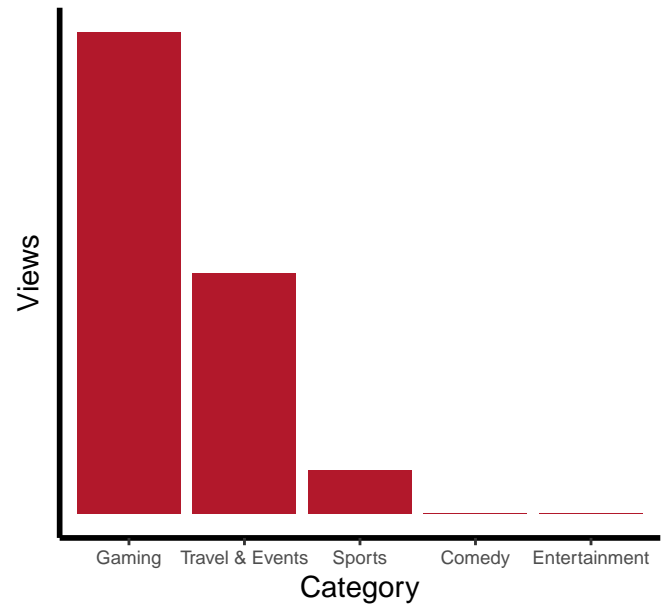


Figure 10. France



1. View pattern

1.2 Top 5 Most Rated YouTube Categories (by country)

- In terms of likes and dislikes, similarly, videos regarding to entertainment (Figure 11-13), music (Figure 14-17), gaming and auto (Figure 18-20) gain most popularity.
- The percentage of dislikes in music videos is generally lower than that in other three Top 1 genres, which indicates that music videos are more acceptable than others and thus a “safe” choice for advertiser.

Figure 11. Korea

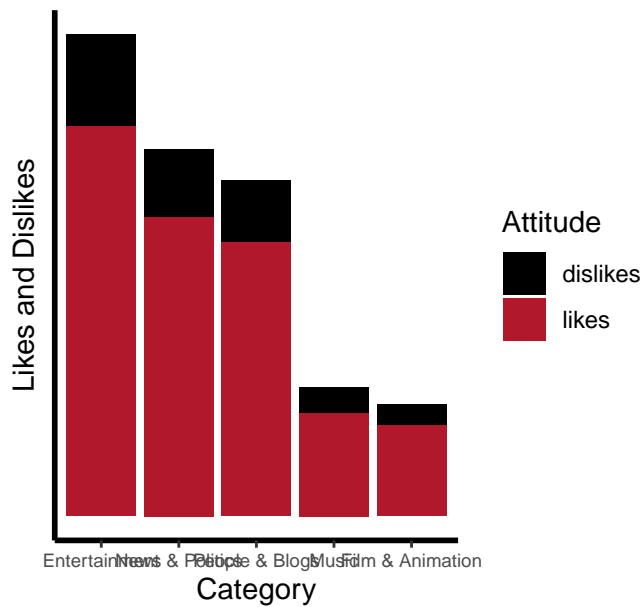


Figure 12. Mexico

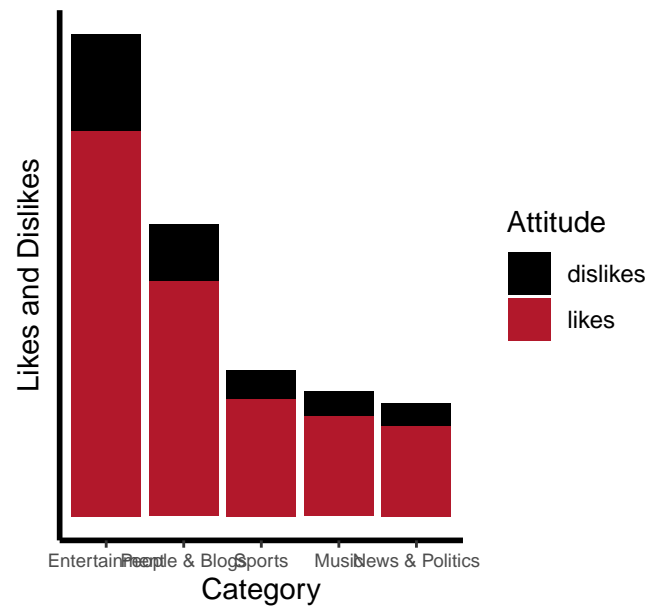


Figure 13. Canada

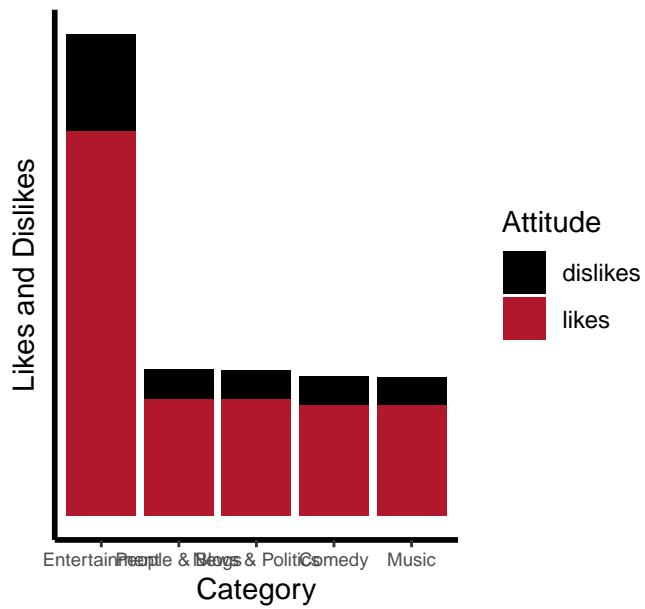


Figure 14. United States

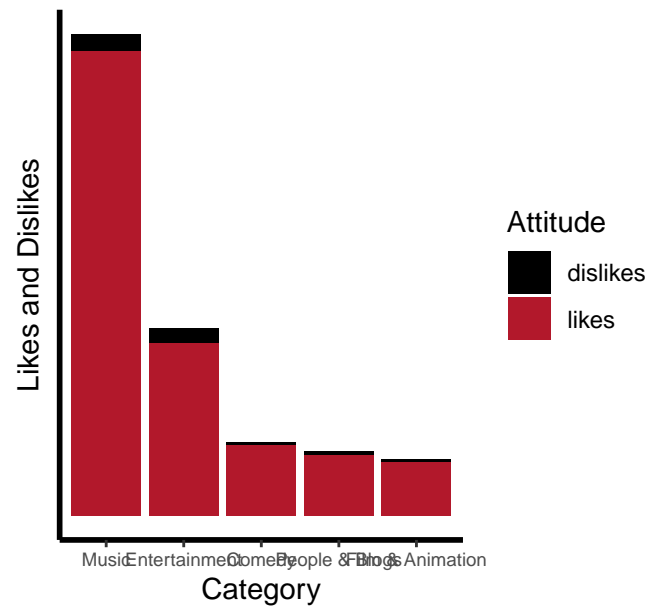


Figure 15. Russia

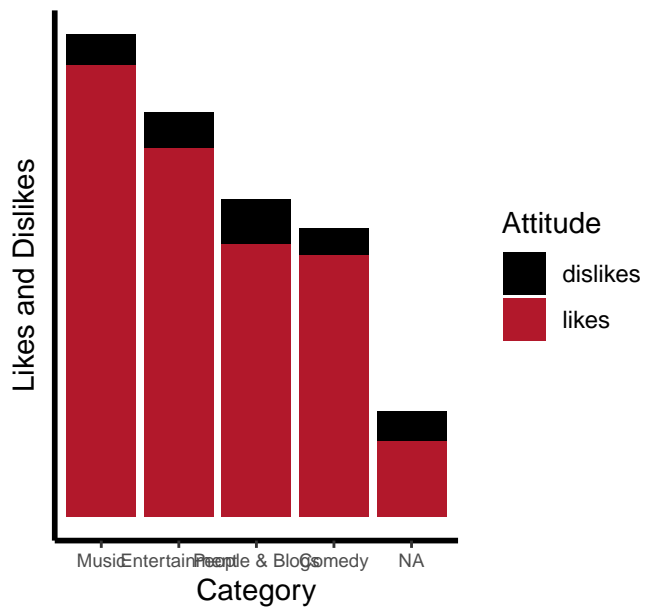


Figure 16. Germany

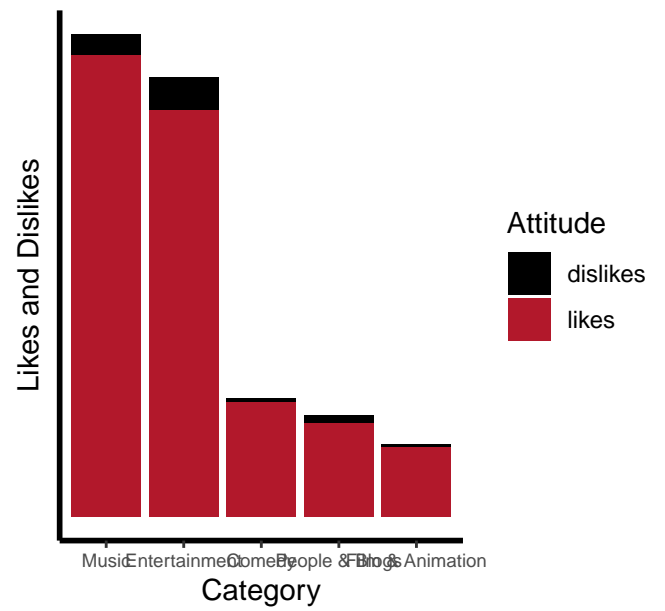


Figure 17. Japan

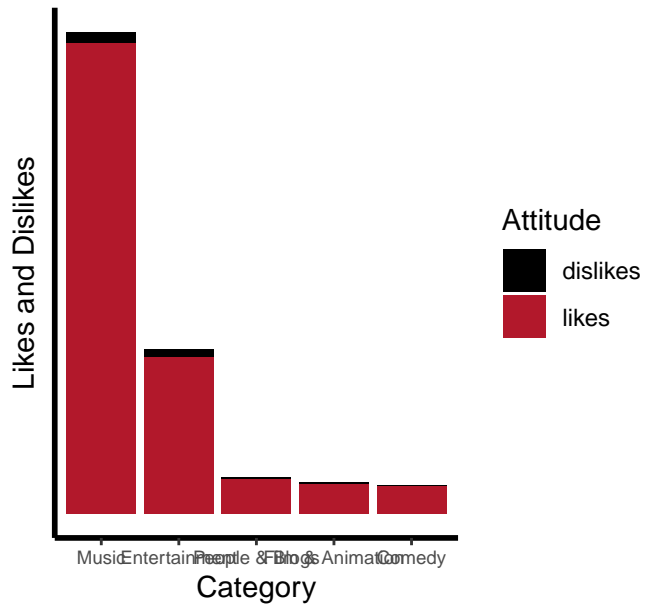


Figure 18. United Kindom

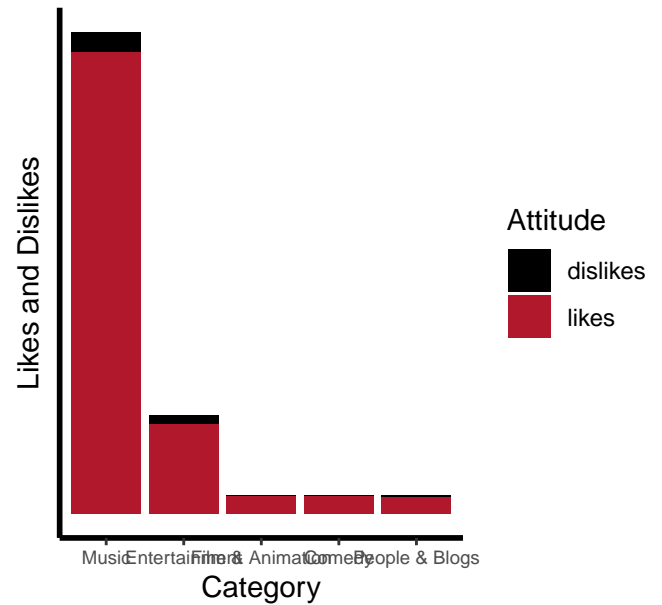


Figure 19. India

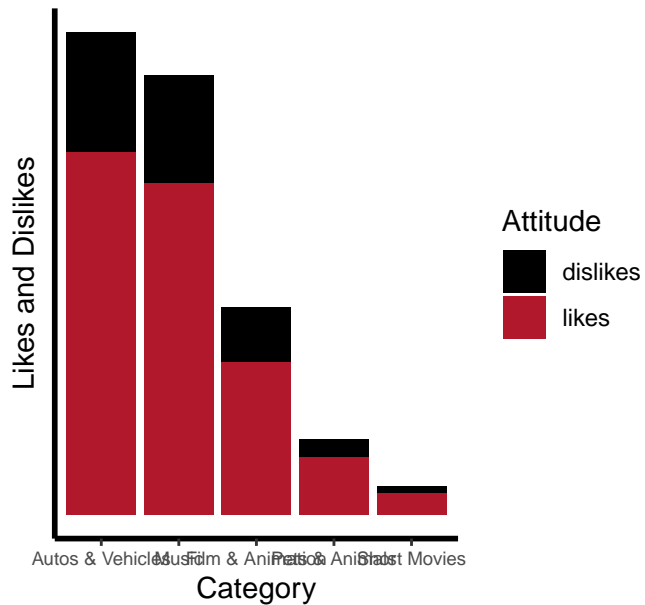
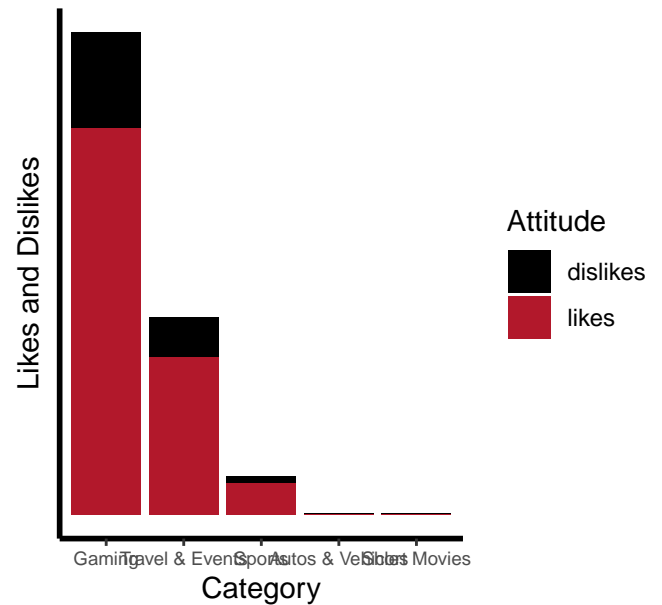


Figure 20. France



1. View pattern

1.3 Top 5 Most Commented YouTube Categories (by country)

- Entertainment (Figure 21-24), music (Figure 25-28), gaming and auto videos'(Figure 29-30) influence continues. . .

Figure 21. Russia

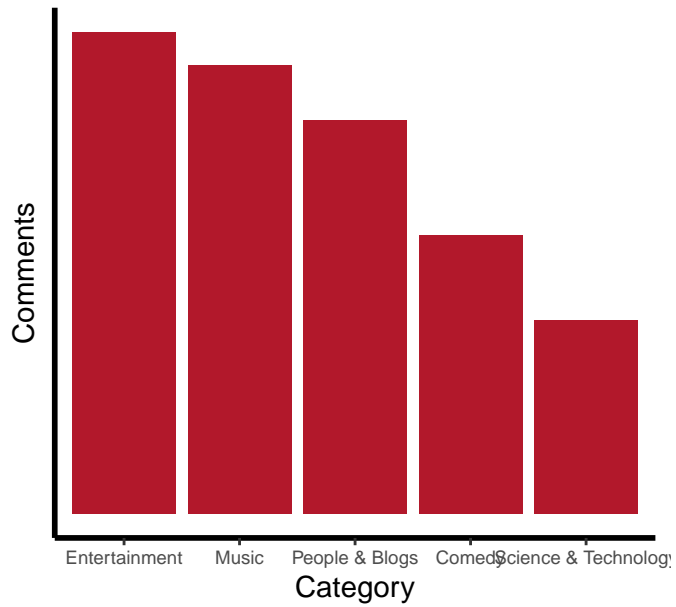


Figure 22. Germany

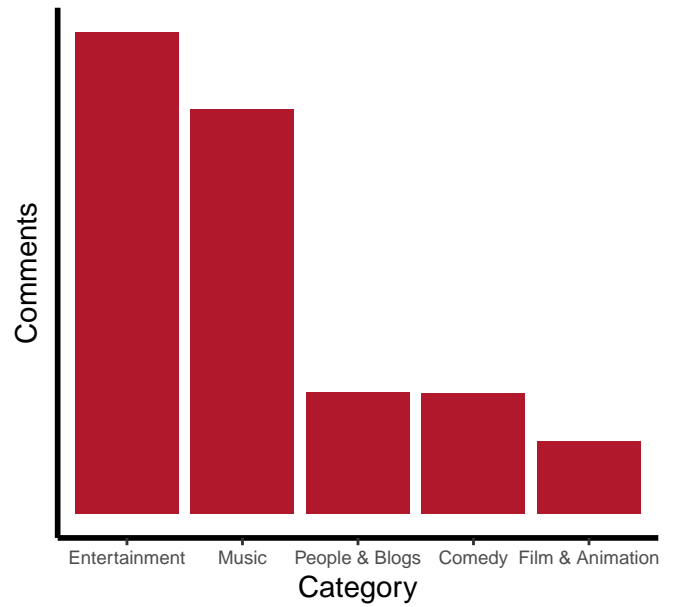


Figure 23. Mexico

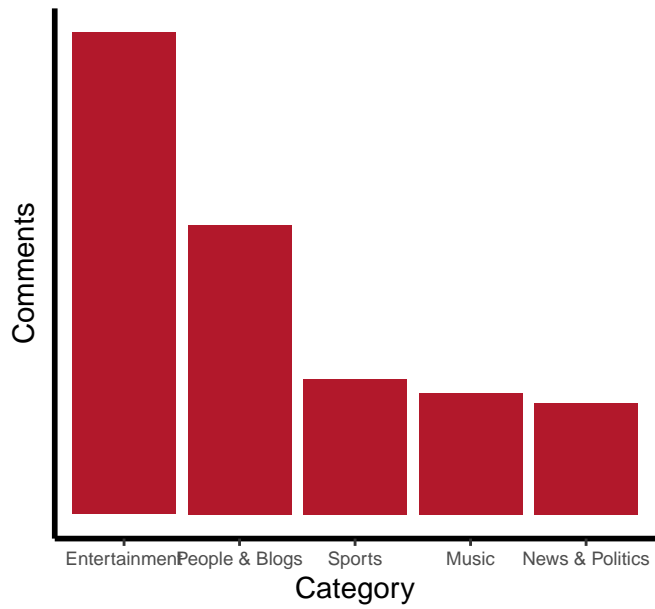


Figure 24. Canada

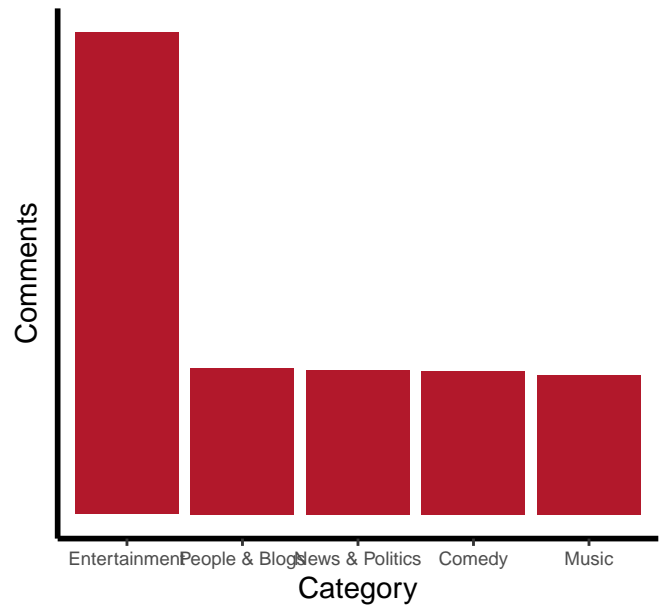


Figure 25. United States

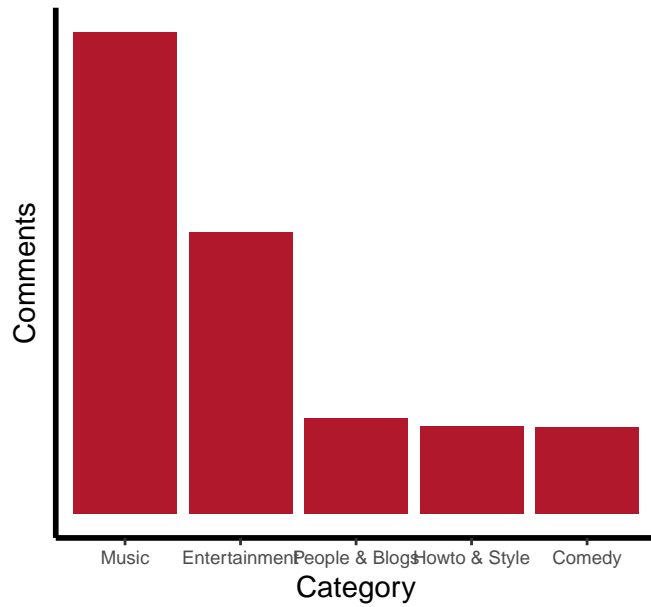


Figure 26. United Kingdom

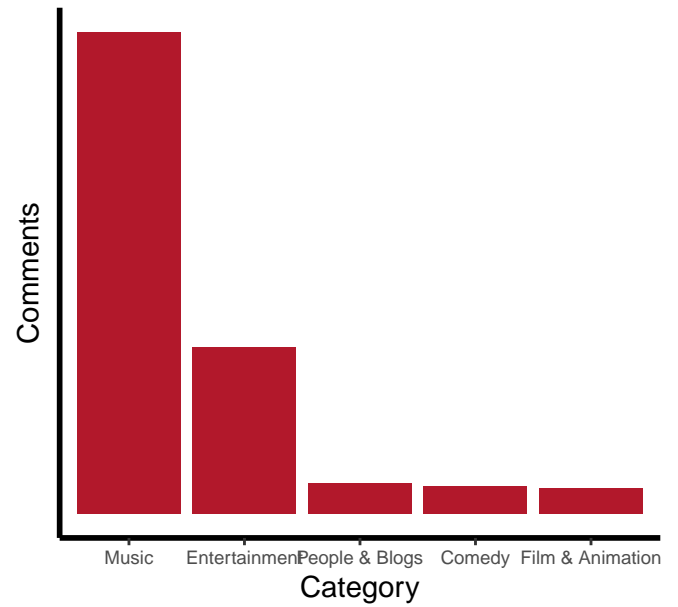


Figure 27. Japan

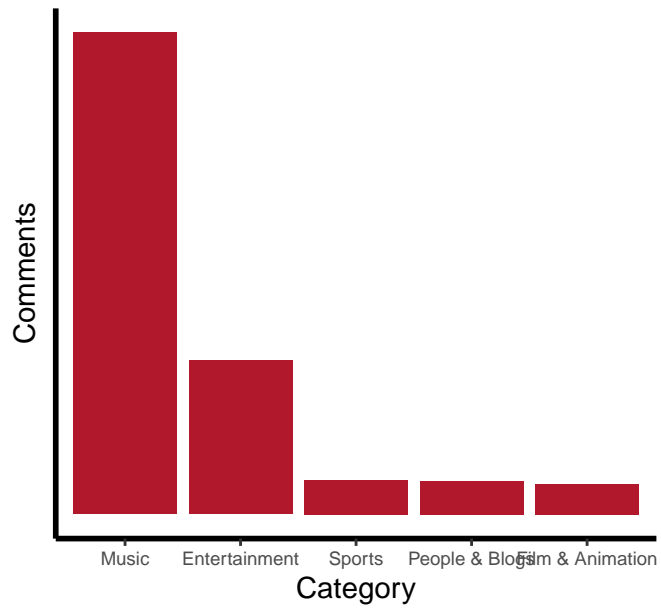


Figure 28. Korea

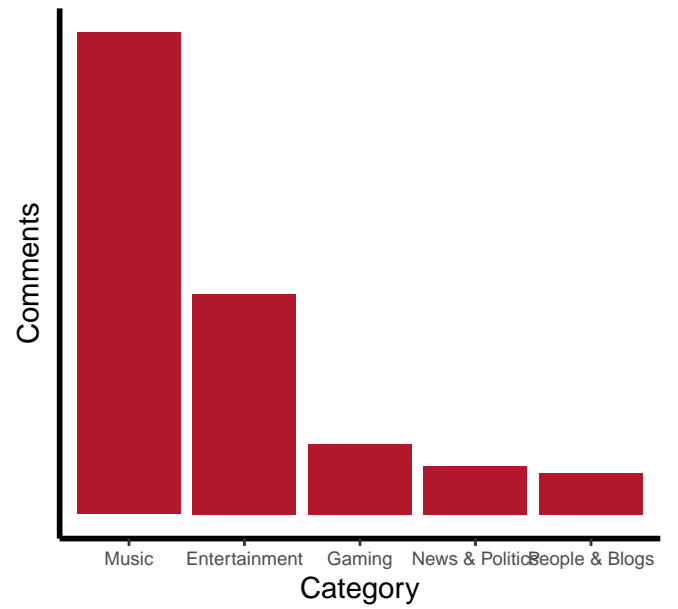


Figure 29. France

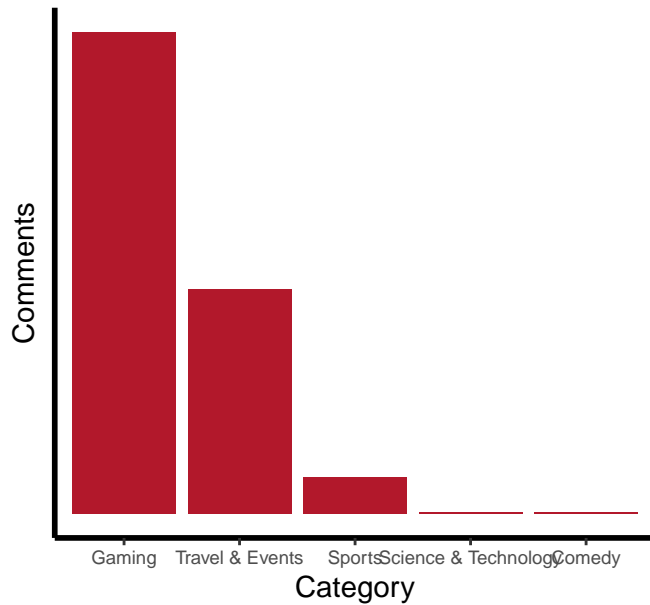
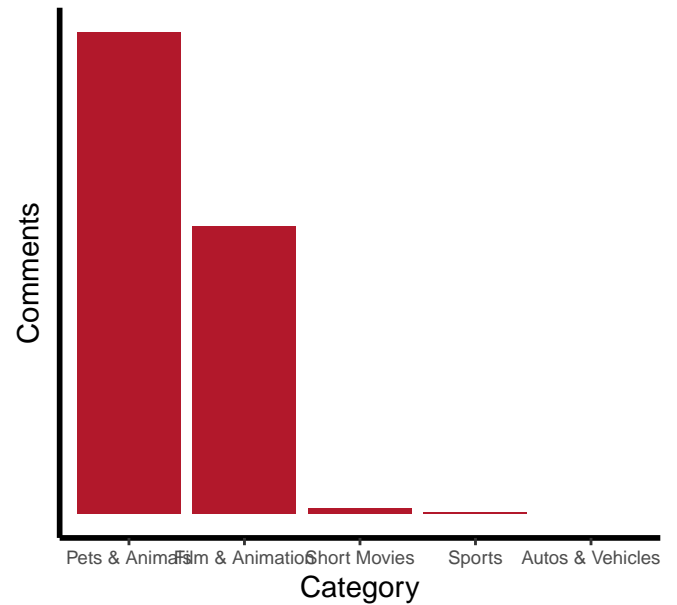


Figure 30. India



1. View pattern

1.4 Top 5 Most Popular YouTube Categories (by country)

- Finally, according to the times that videos went on charts, the four categories (Entertainment, Music, Auto and Gaming, see Figure 31-39) still dominate, which indicates that the views, likes, dislikes and number of comments are correlated with whether a video goes on chart.
- Besides, people and blog (Figure 40) videos gain ground in Russia.

Figure 31. Korea

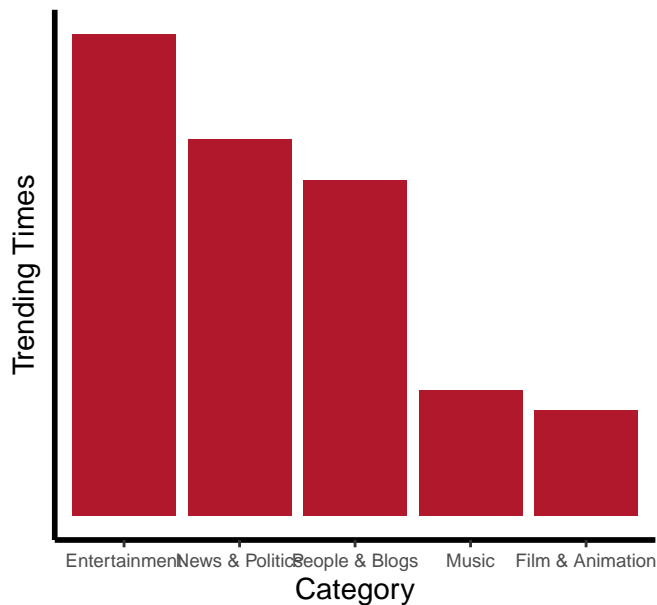


Figure 32. Germany

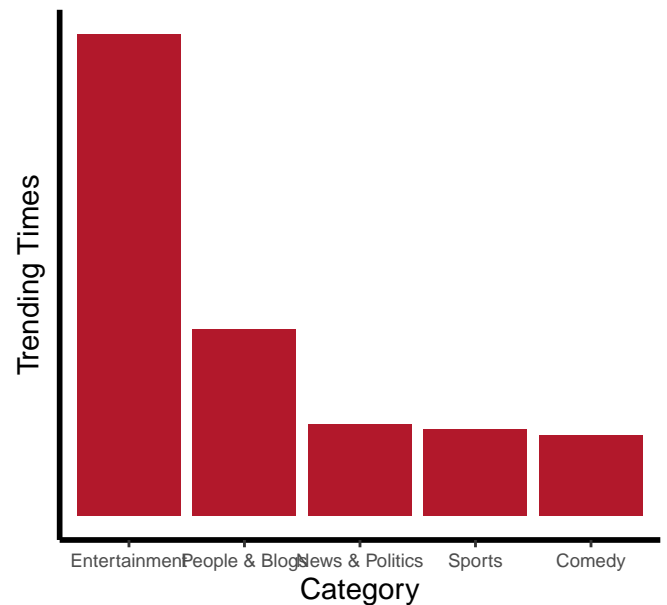


Figure 33. Mexico

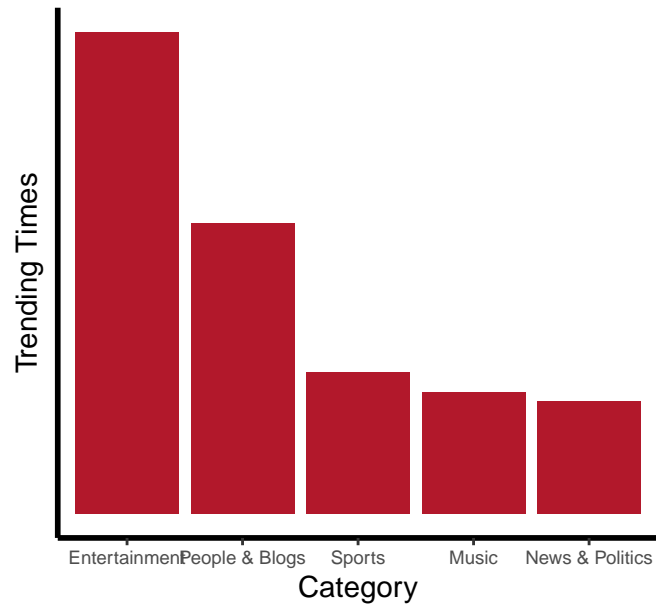


Figure 34. Japan

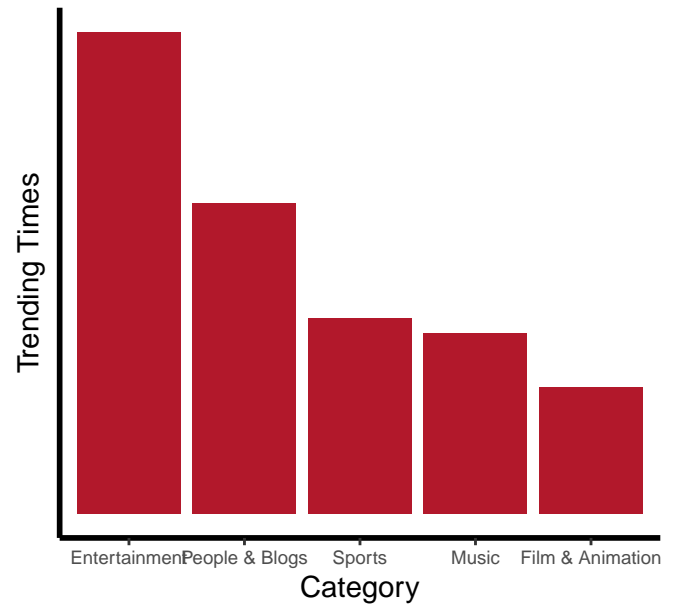


Figure 35. Canada

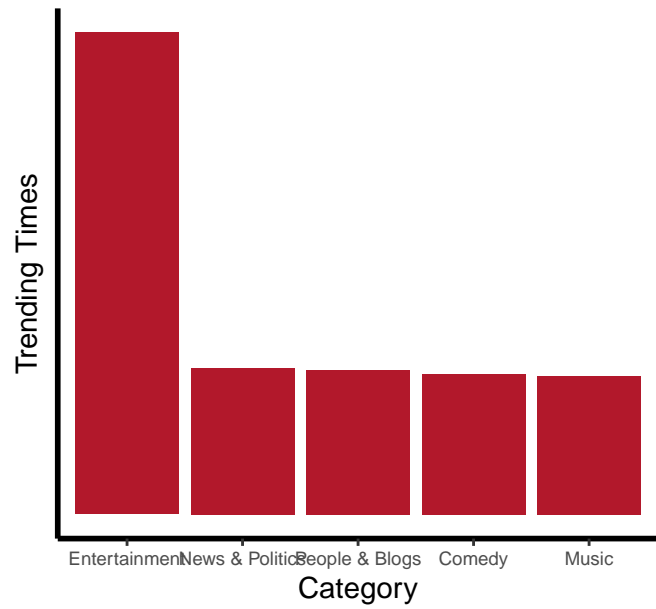


Figure 36. United States

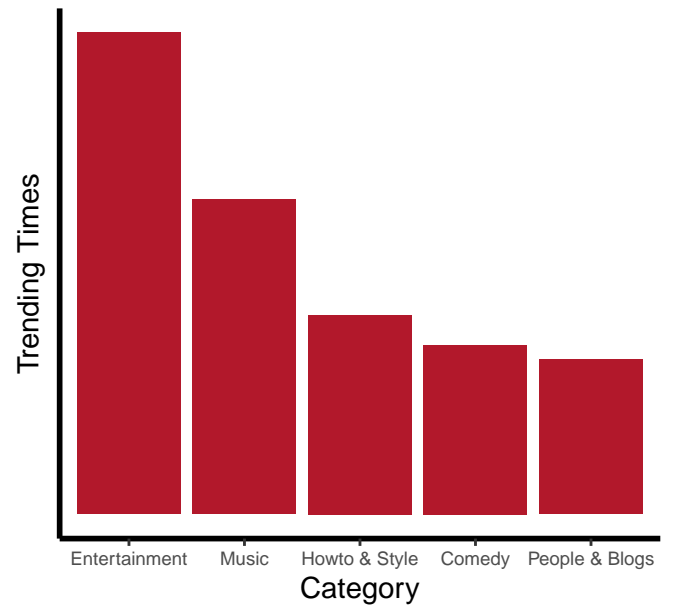


Figure 37. United Kingdom

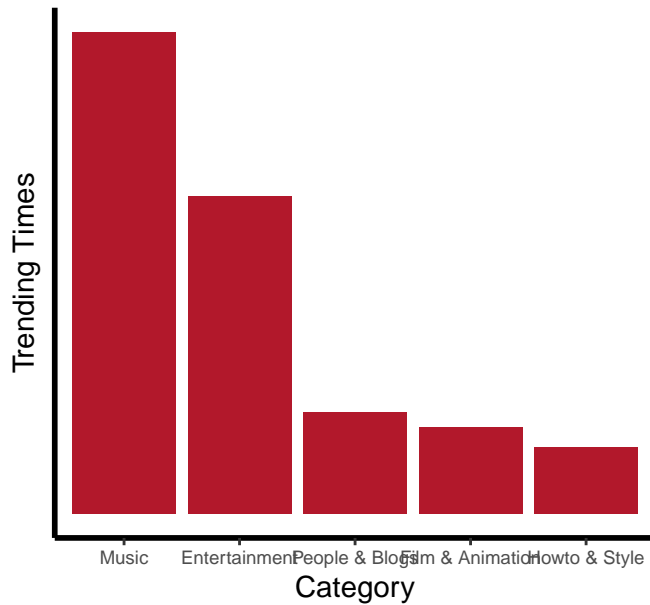


Figure 38. India

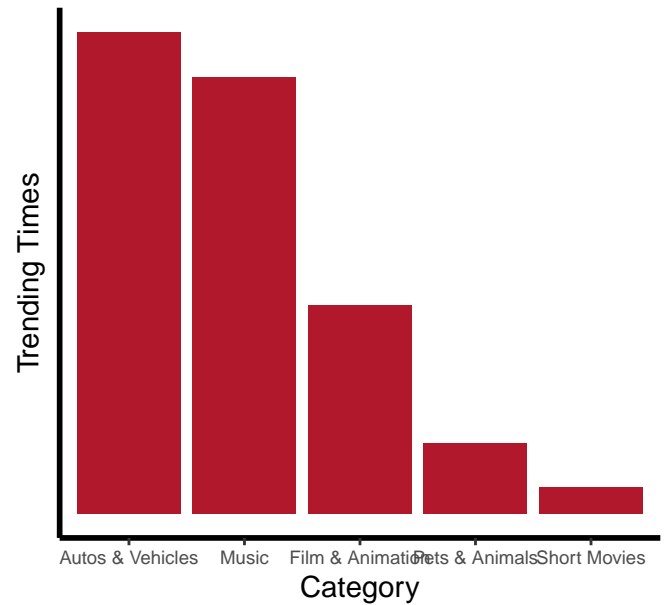


Figure 39. France

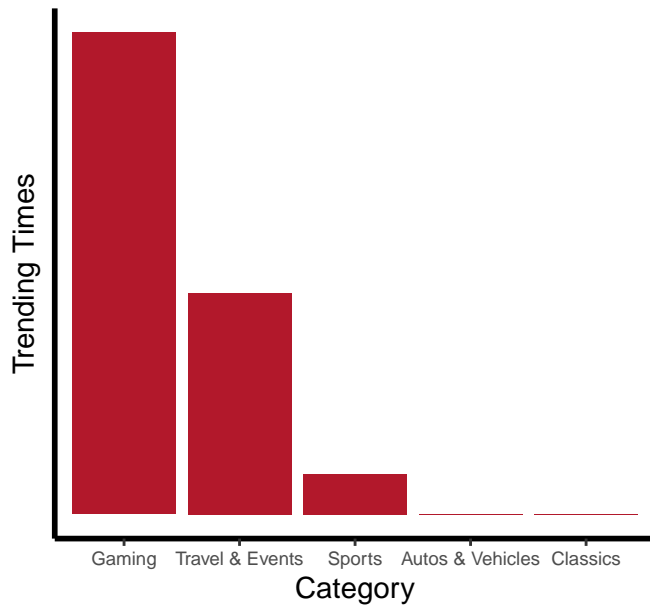
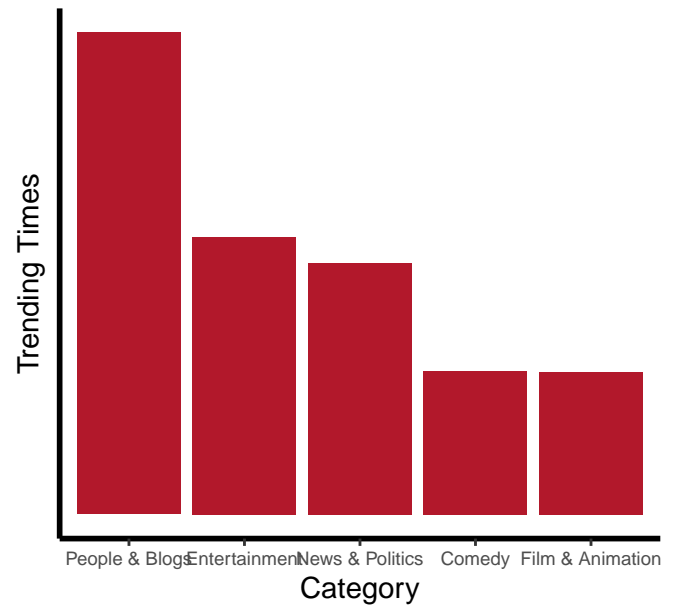


Figure 40. Russia



2. User participation

- User participation is broke down into two parts:
 - showing their attitude to the videos, i.e. clicking “likes and dislikes”;
 - leaving comments on videos.
- The average voting rate, 3% of ten countries is 10 times of the average comment rate,0.3% (red lines in the figures below):
 - $\text{Voting rate of a video} = (\text{likes} + \text{dislikes}) / \text{views} \times 100\%$;
 - $\text{Comment rate of a video} = (\# \text{ of comments}) / \text{views} \times 100\%$.
- The British show the least willingness to rate and comment on videos, while Russians are enthusiastic about doing so, which could be partly explained by culture difference.

Figure 41. Average Voting Rate

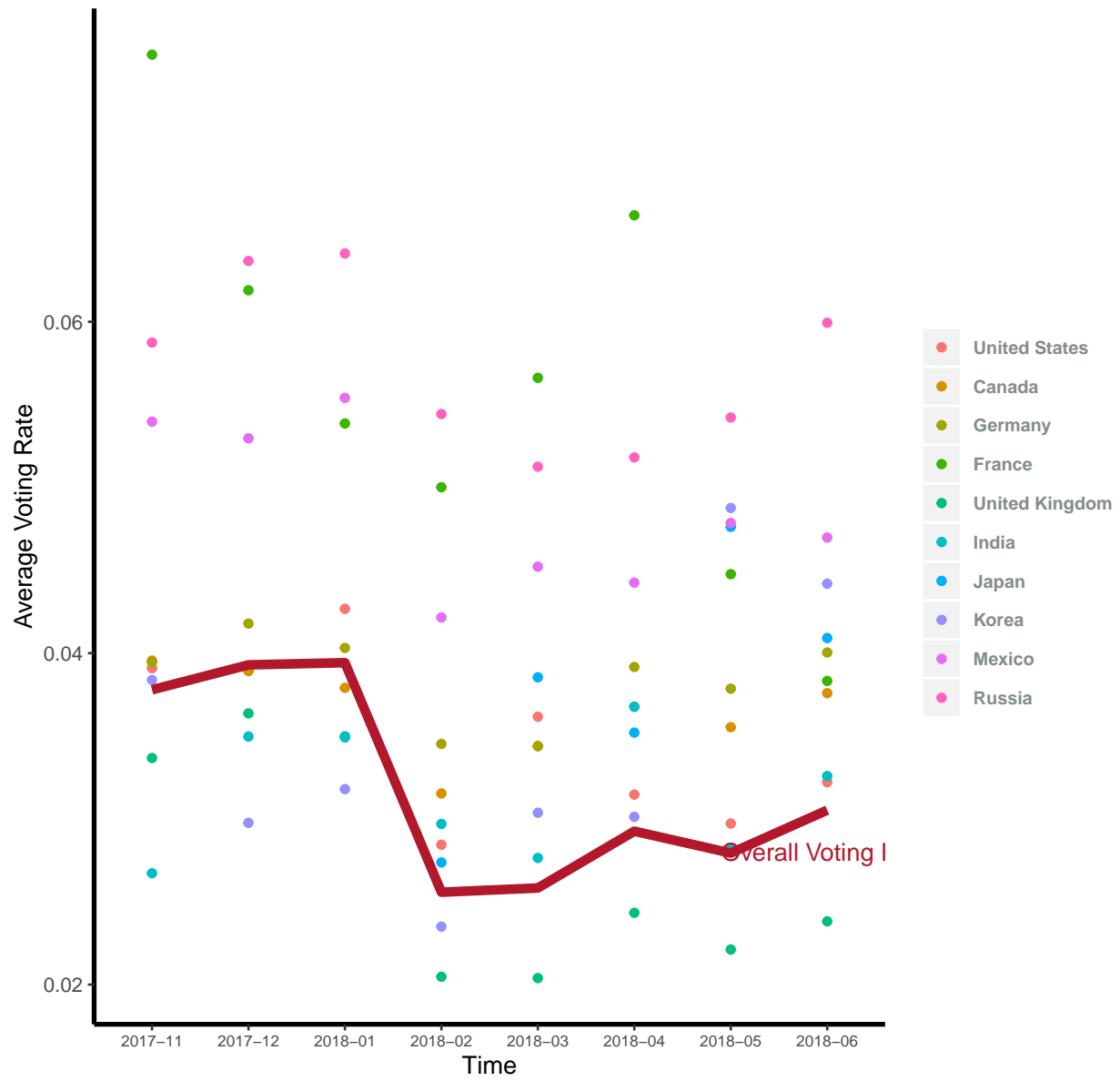
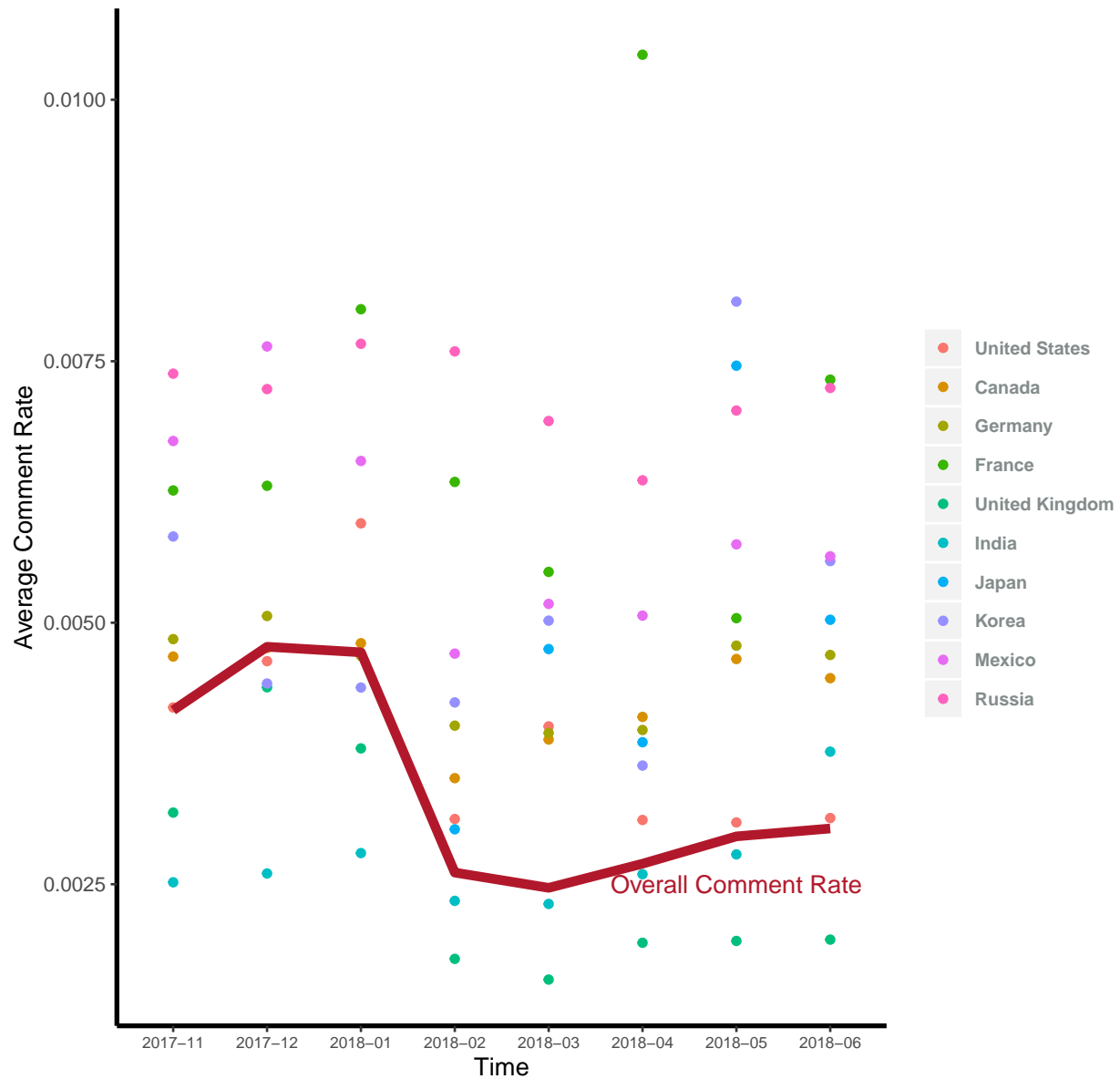


Figure 42. Average Comment Rate



3. Trending Lifecycle

“Trending lifecycle is from the moment the video is published to the last trending date on the charts” **Metrics explanation**

- I focus on: Average going-viral time=Initial trending date-publish date
 - Average trending days=Final trending date-Initial trending date
- The average going-viral time for each category describes on average how fast a video can show up on the trending charts. The longer time interval is, the larger the time cost will be.

Insights from visuals

- By country
 - Average going-viral time: In most countries, it took **1-3 days** for a video to go viral (Figure 41-48). However, in U.S and UK, it took **1-2 weeks** (Figure 49-50).

- By category
 - Average going viral time: Music videos take the longest time to go viral on YouTube. This may be explained by the fact that:
 - * Music videos normally lack breath-taking storylines, so they take time to attract audience;
 - * There exist other more focused music video platforms, like Vimeo and TikTok, making audience switching to them.

Figure 41. Japan

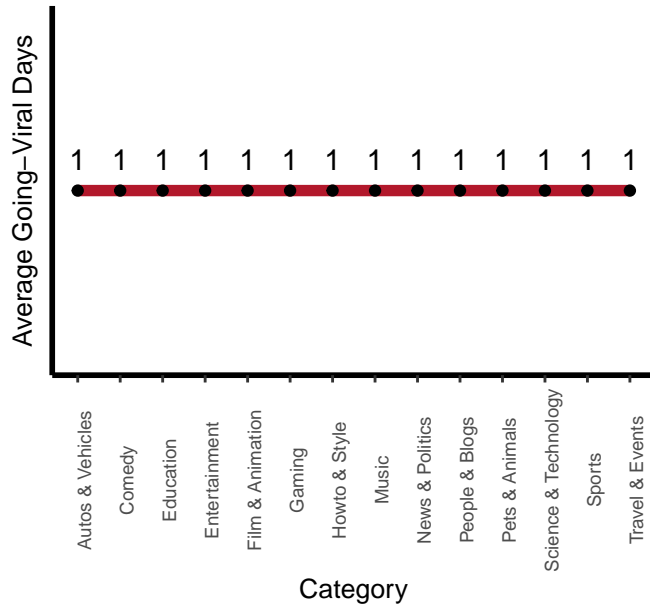


Figure 42. Russia

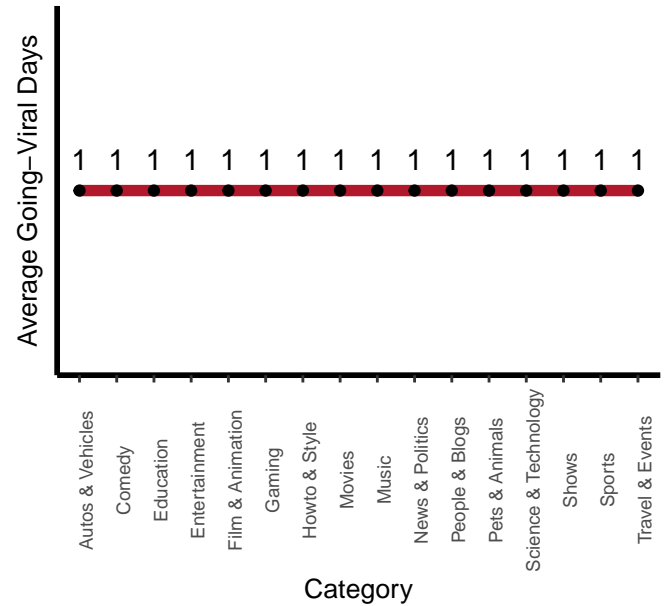


Figure 43. France

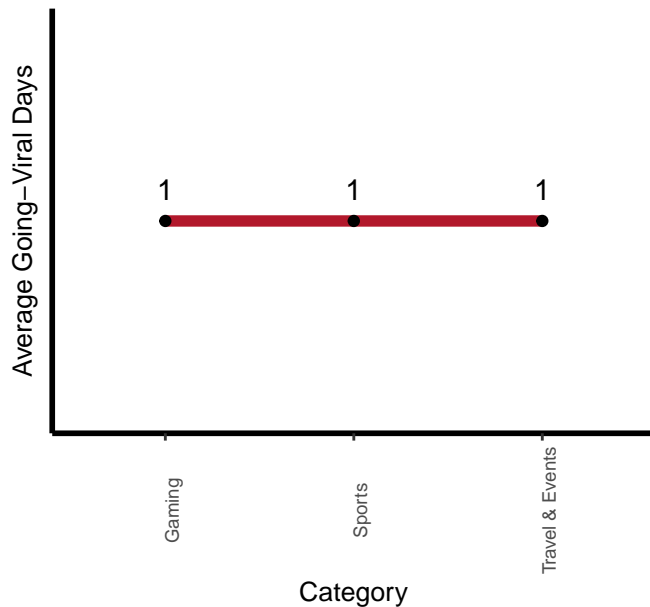


Figure 44. Mexico

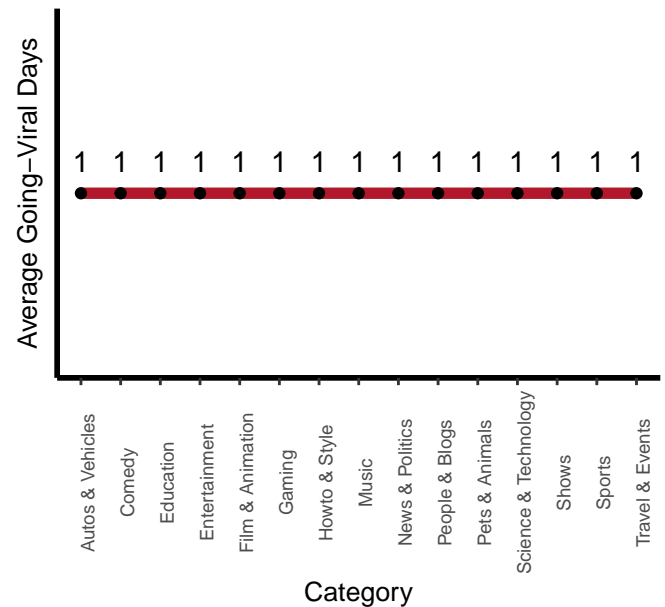


Figure 45. Germany

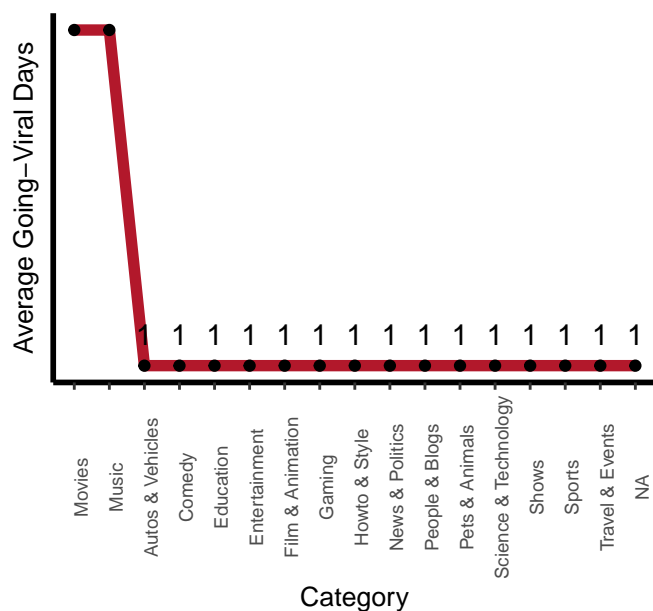


Figure 46. Canada

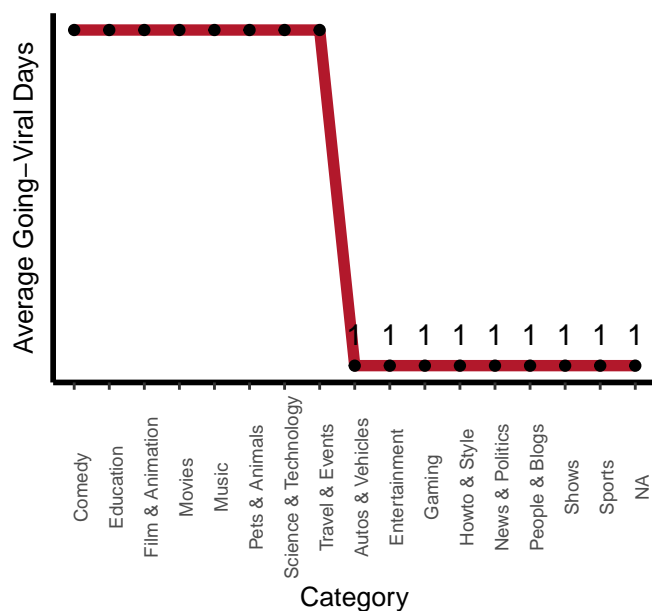


Figure 47. Korea

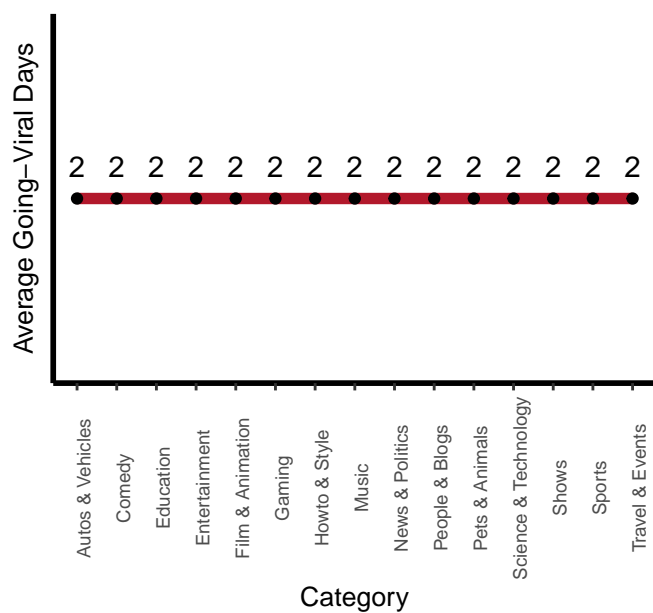


Figure 48. India

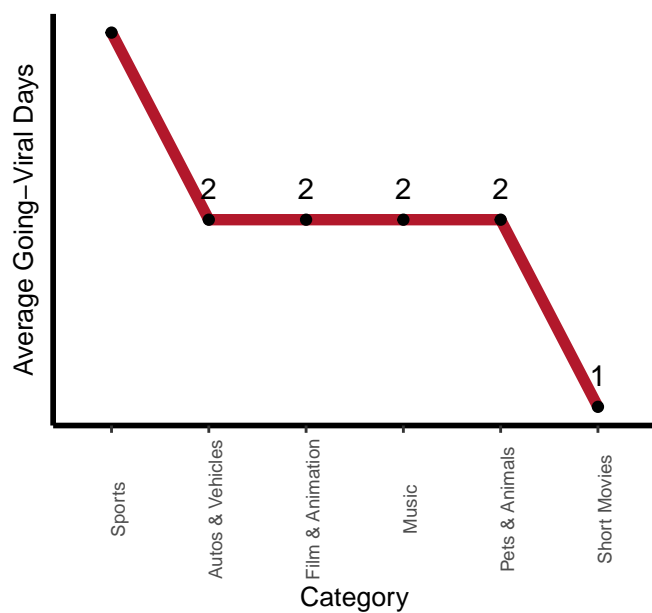


Figure 49. United States

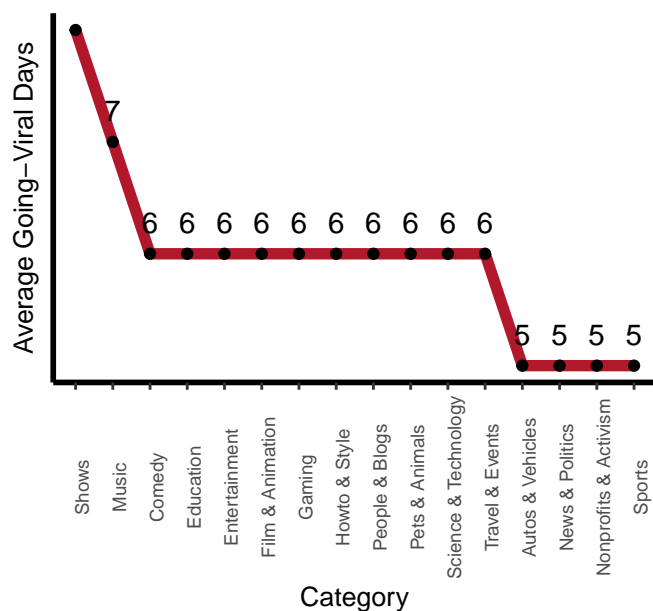
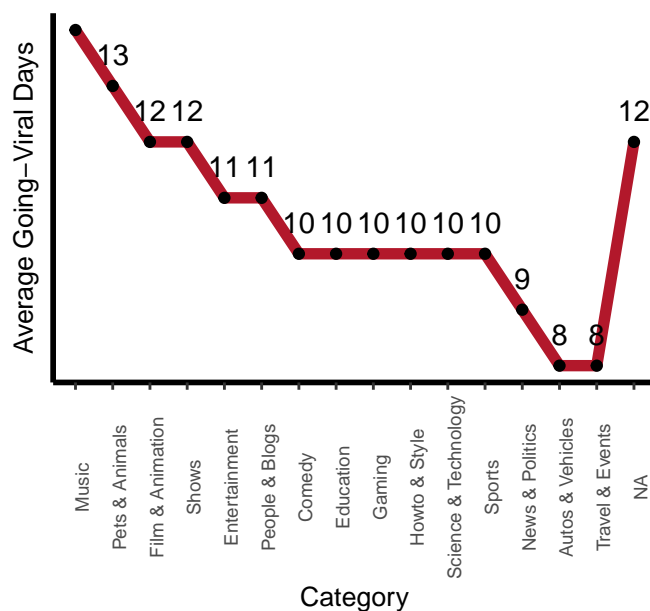


Figure 50. United Kingdom



4. Category Selection

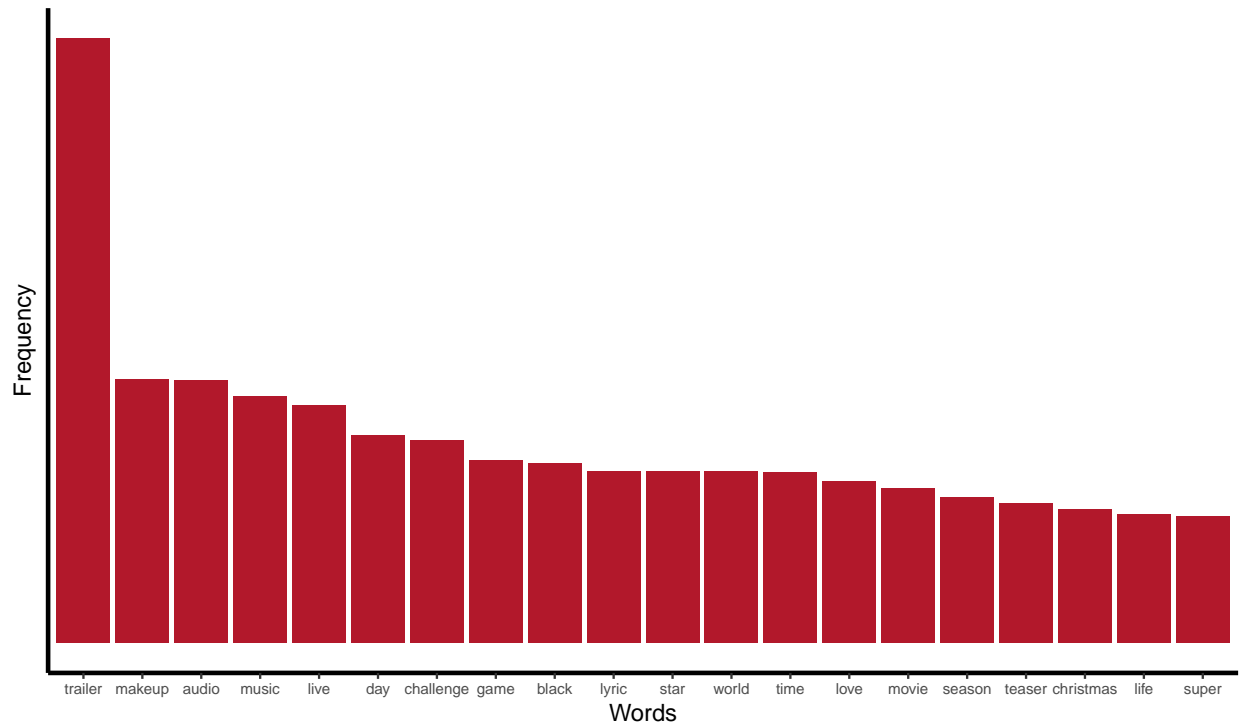
According to previous analyses, different Top 5 categories are generated under different criteria (we can see this more clearly in the **shiny app**, where I put four criteria of a country in one screen), therefore, a comprehensive evaluation is needed.

Method

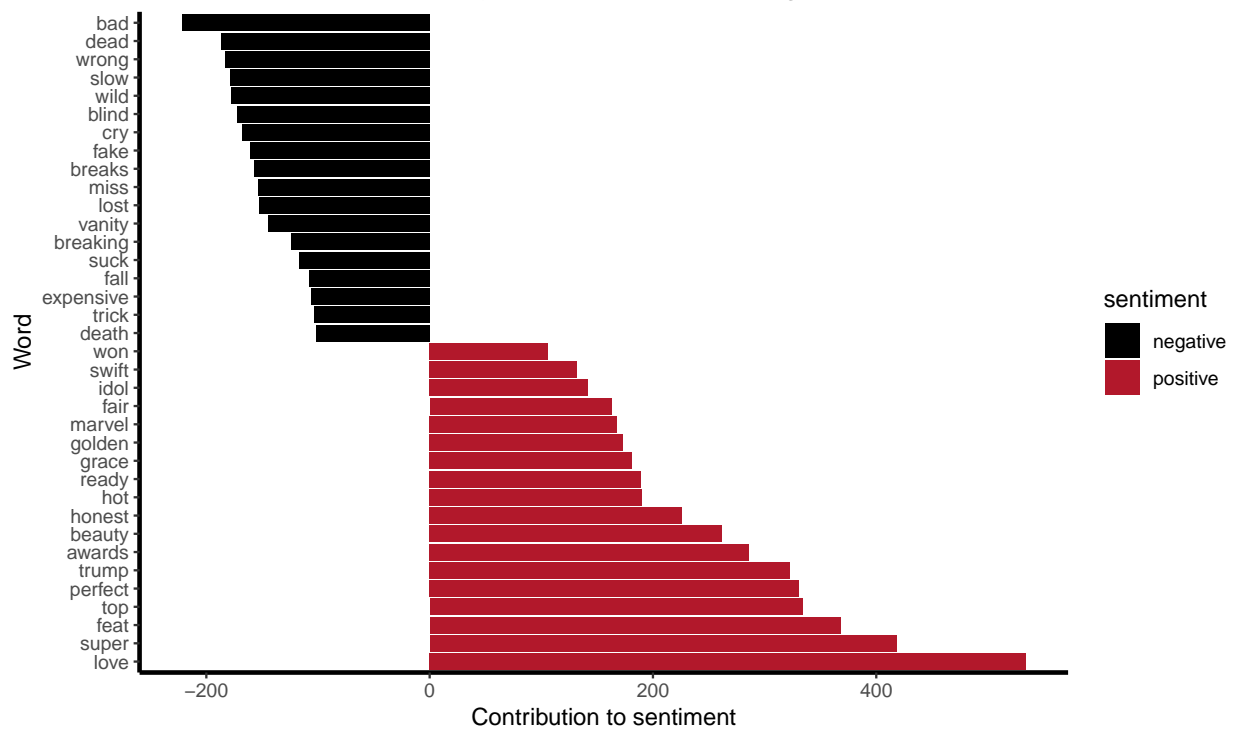
- In the evaluation matrix, I took view pattern, user participation and trending lifecycle into account and pick the variables below.
- Used the reverse ranking of each attribute as its score for categories, and then summed up all the scores to get a final score of each category to select most valuable category to do further analysis.

Category_name	Views	VotingRate	CommentRat	GoingViralTime	Total_Score
Howto & Style	14.0	13	14	7.5	48.5
Nonprofits & Activism	1.5	16	16	14.5	48.0
Comedy	13.0	15	10	7.5	45.5
People & Blogs	12.0	12	13	7.5	44.5
Entertainment	16.0	9	9	7.5	41.5
Education	7.0	14	12	7.5	40.5
News & Politics	11.0	3	11	14.5	39.5
Gaming	5.0	11	15	7.5	38.5
Music	15.0	10	6	2.0	33.0
Science & Technology	10.0	7	7	7.5	31.5
Sports	8.0	5	4	14.5	31.5
Pets & Animals	6.0	8	8	7.5	29.5
Film & Animation	9.0	6	3	7.5	25.5
Autos & Vehicles	3.0	1	1	14.5	19.5
Travel & Events	4.0	2	5	7.5	18.5
Shows	1.5	4	2	1.0	8.5

Top 20 Popular Words of YouTube Trending Videos' Titles

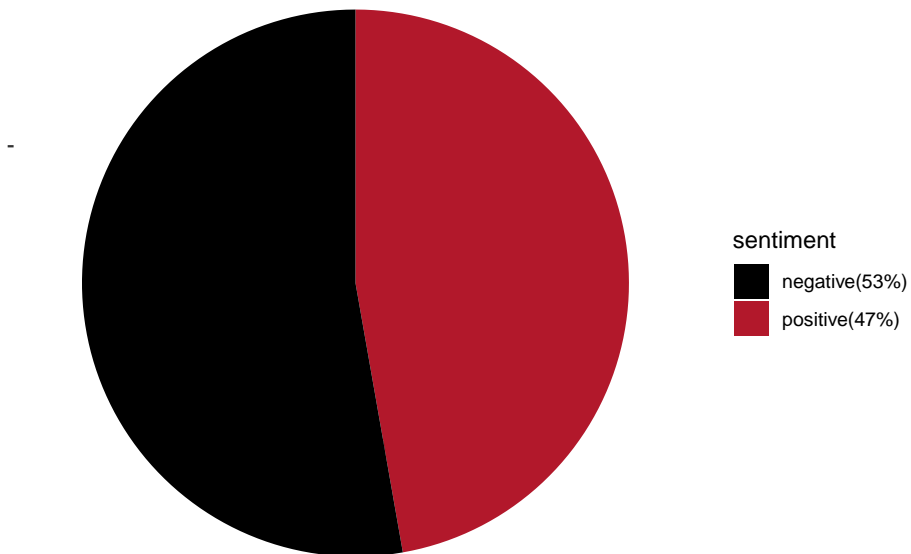


Sentimental Analysis for YouTube Trending Video Title



Scale for 'fill' is already present. Adding another scale for 'fill',
which will replace the existing scale.

Sentimental Analysis for YouTube Trending Video Title (%)



Conclusion

- View Pattern: Entertainment, music, gaming and auto gain most popularity in 10 countries.
- User Participation: The average voting rate, 3% of ten countries is 10 times of the average comment rate, 0.3%.
- Trending Lifecycle: The time interval of both going-viral and keeping trending ranges from 1 day to 2 weeks.
- Category Selection: Entertainment, music, film and animation have the most business potential based on overall performance.
- Text Analysis of YouTube Trending Videos' Titles
 - Top 3 popular words: “official”, “trailer”, “makeup”;
 - Negative sentiment > Positive sentiment.