Trending_YouTube_Video_Exploration

YuzheHuang

Dec. 14 2019

Contents

Key Takeaways

- 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.
- Deep Dive to the Algorithm of Trending Videos: Views, voting rate and comment rate will affect the going-viral days, and have different influence mechanism within different countries, especially U.S. and UK.

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 algorithm of the YouTube trending charts.

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

Data Processing

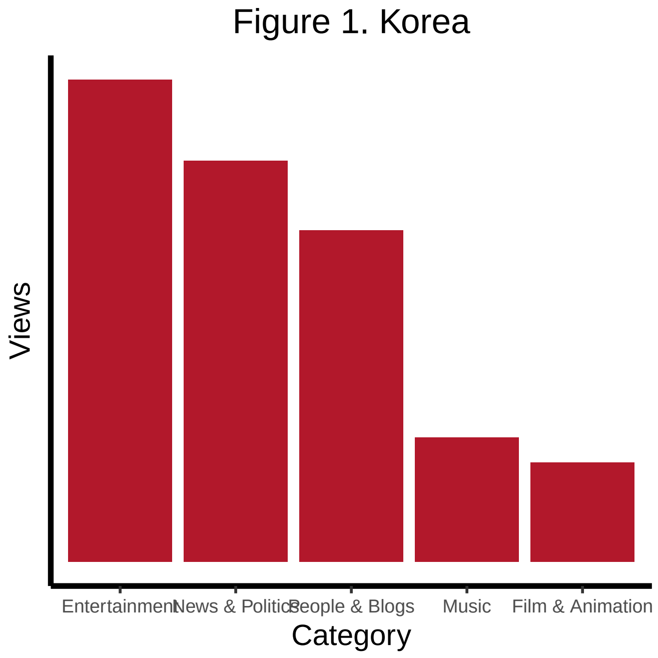
- Drop null values and corrected time format;
- Pair the category ids with their names for further analysis.

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).



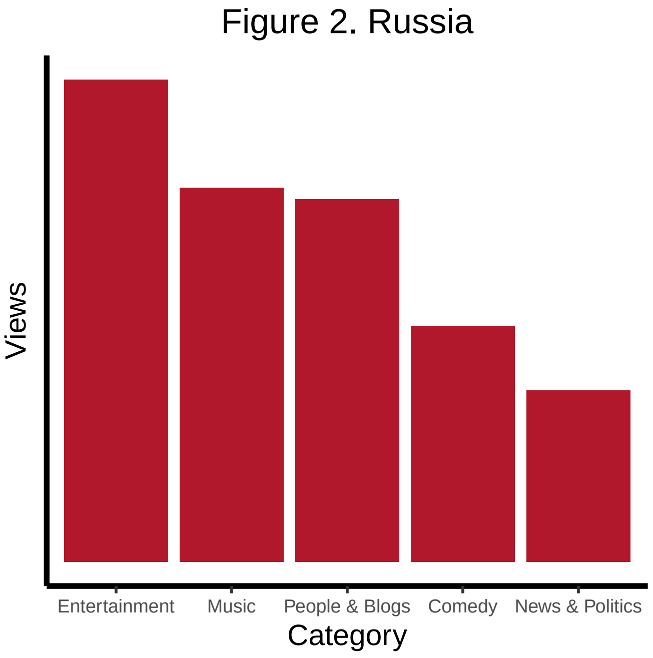


Figure 3. Germany

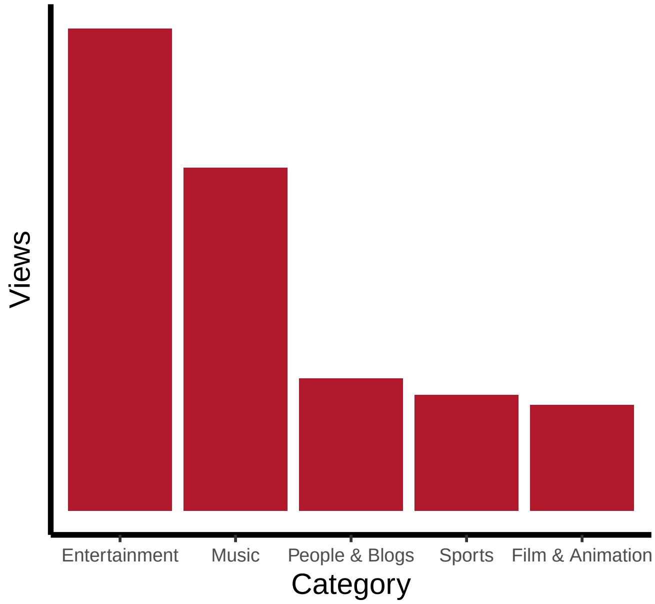


Figure 4. Mexico

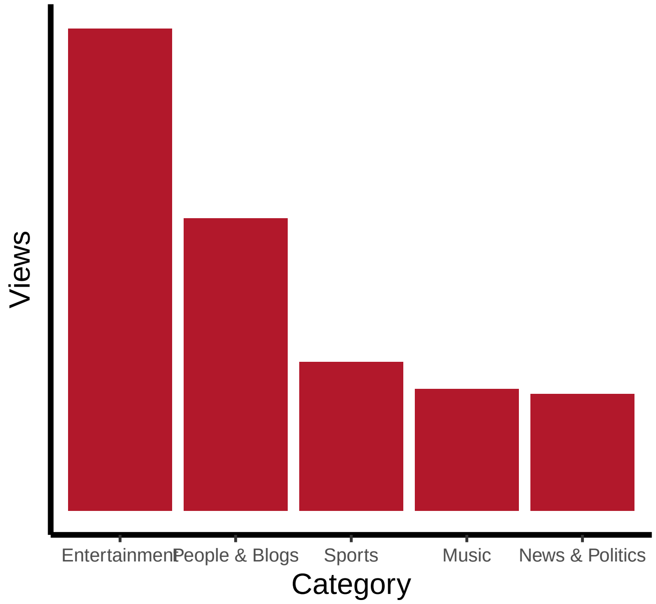


Figure 5. Japan

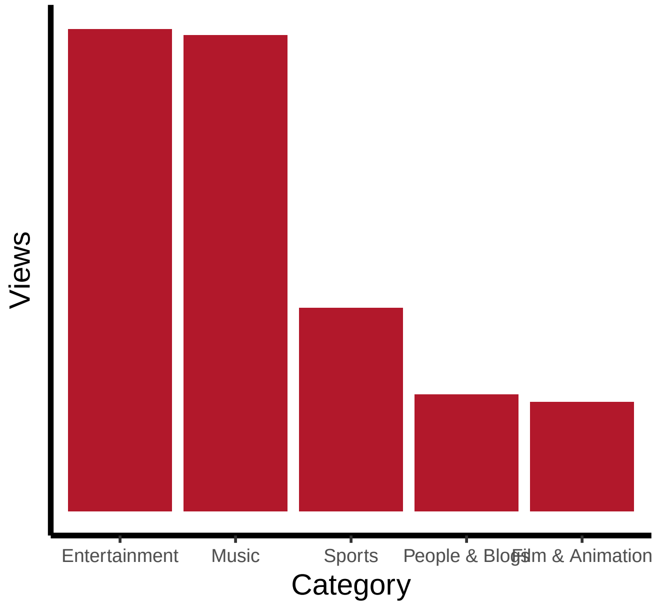


Figure 6. Canada

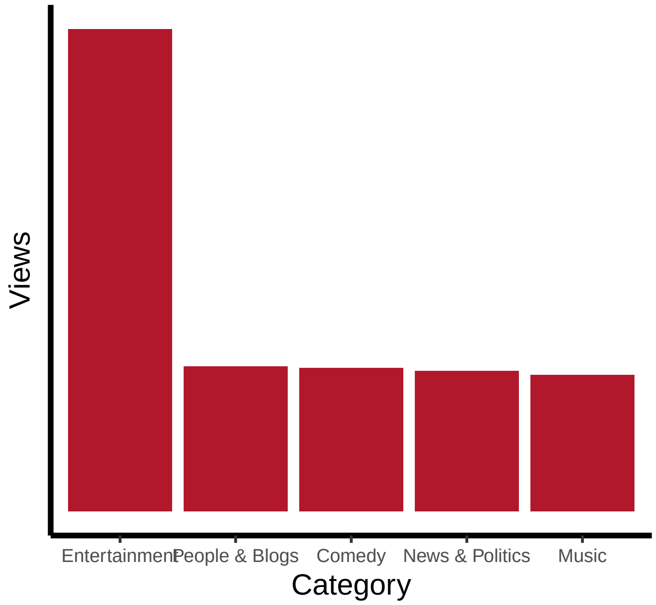


Figure 7. United Kingdom

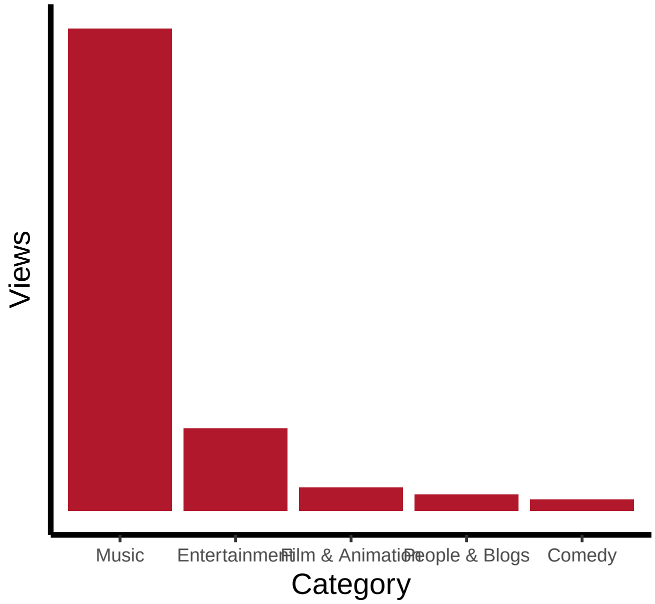


Figure 8. United States

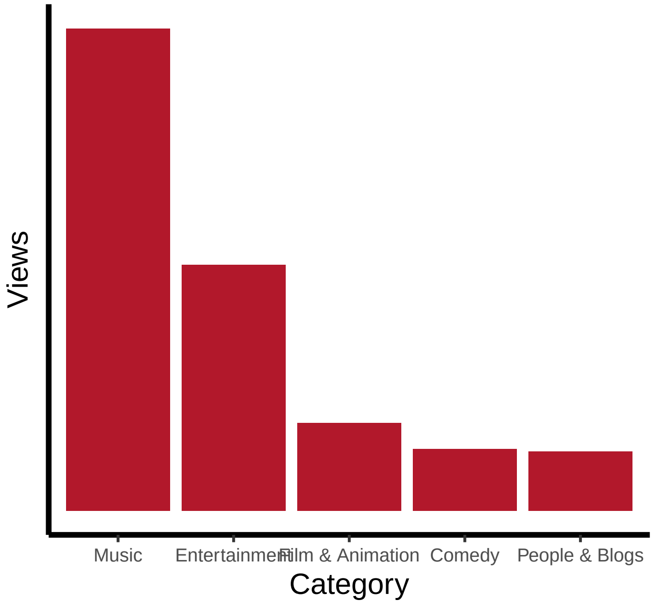


Figure 9. India

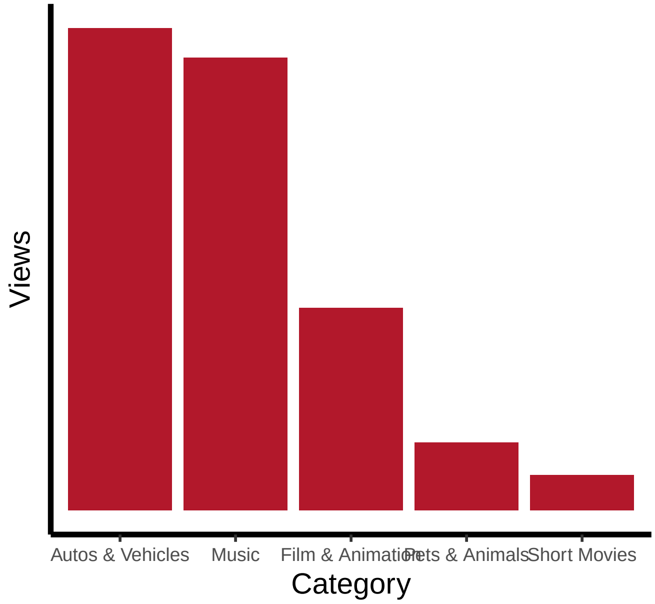
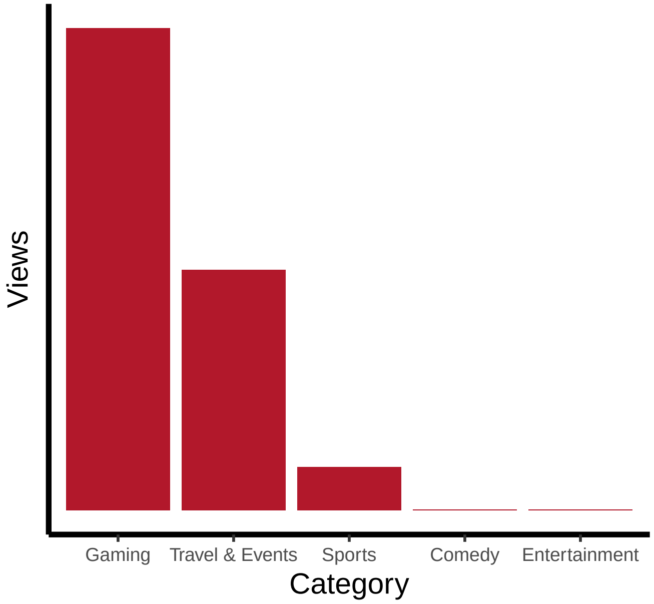


Figure 10. France



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.

(Nearly 200 years ago, Henry Wadsworth Longfellow asserted “Music is the universal Language of mankind.”)

Figure 11. Korea

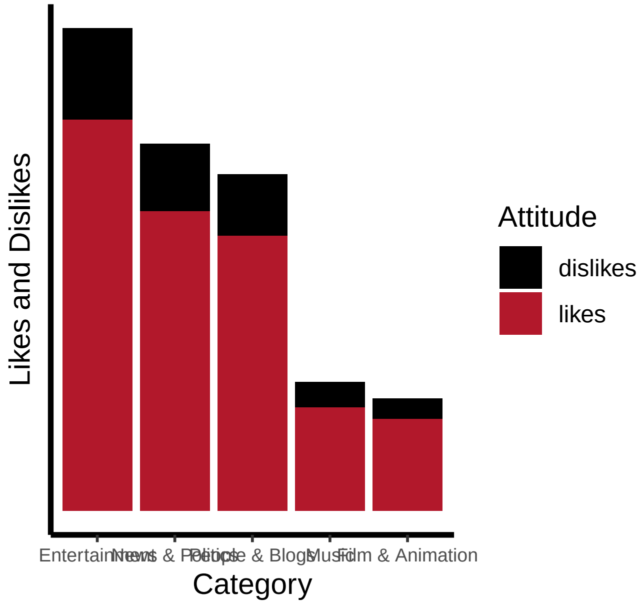


Figure 12. Mexico

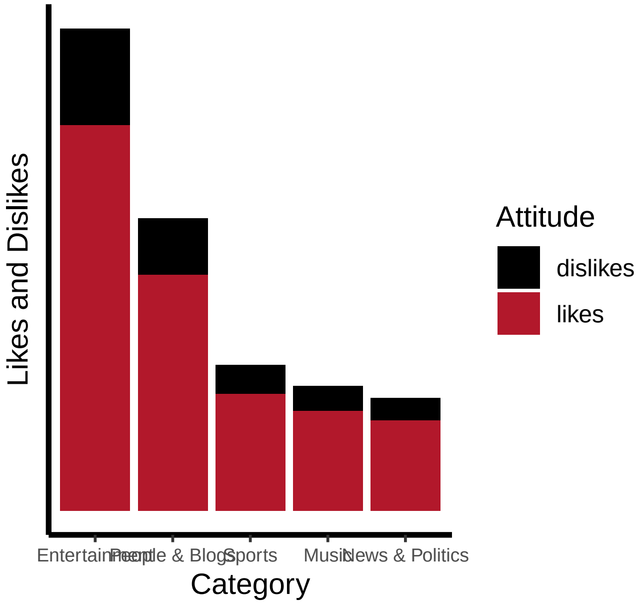


Figure 13. Canada

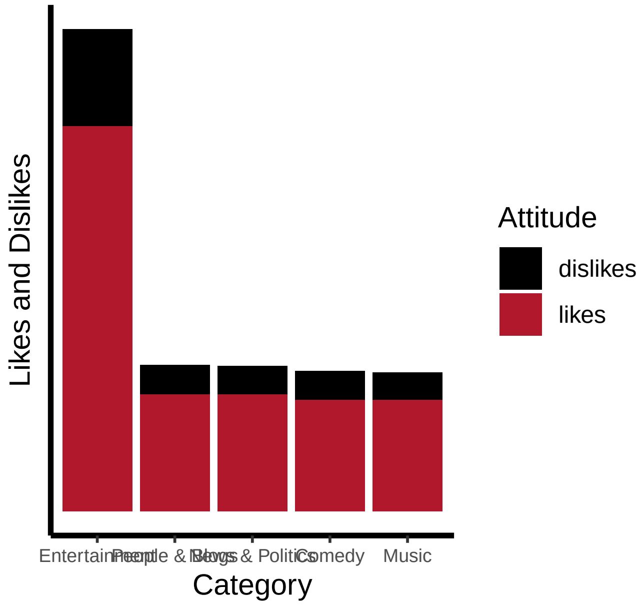


Figure 14. United States

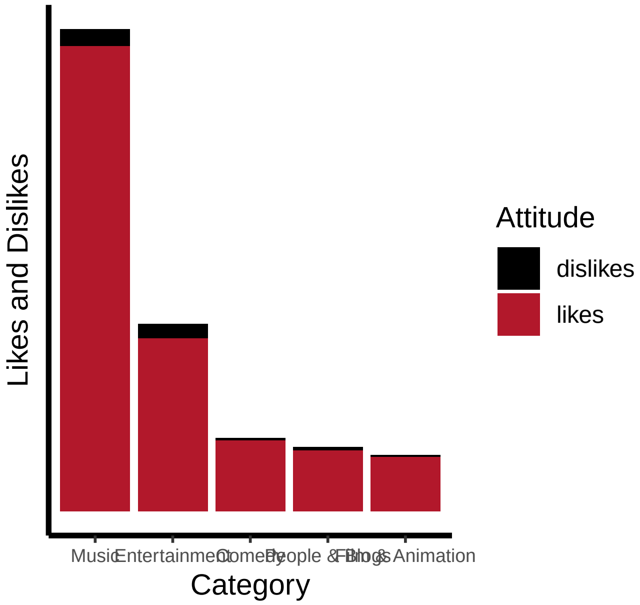


Figure 15. Russia

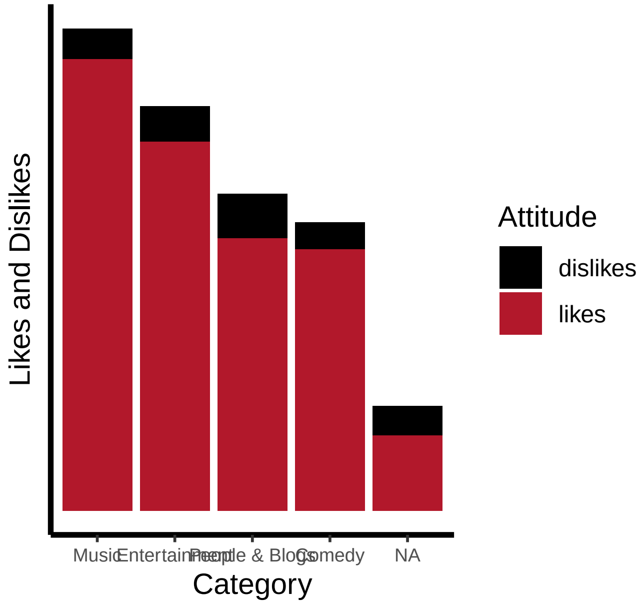


Figure 16. Germany

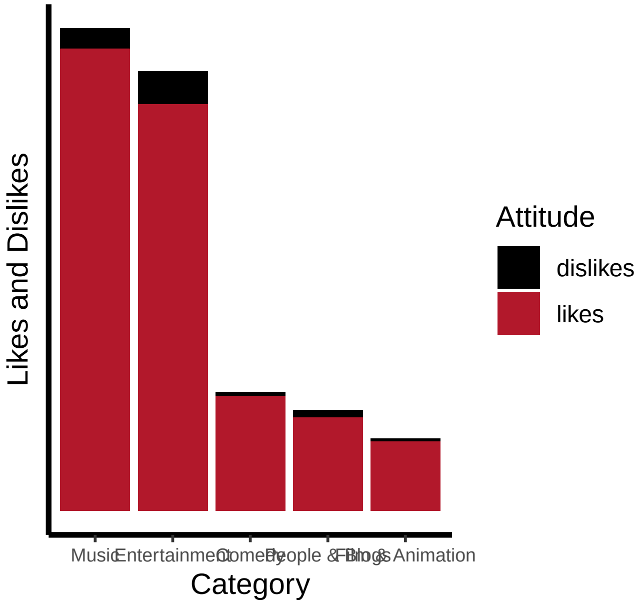


Figure 17. Japan

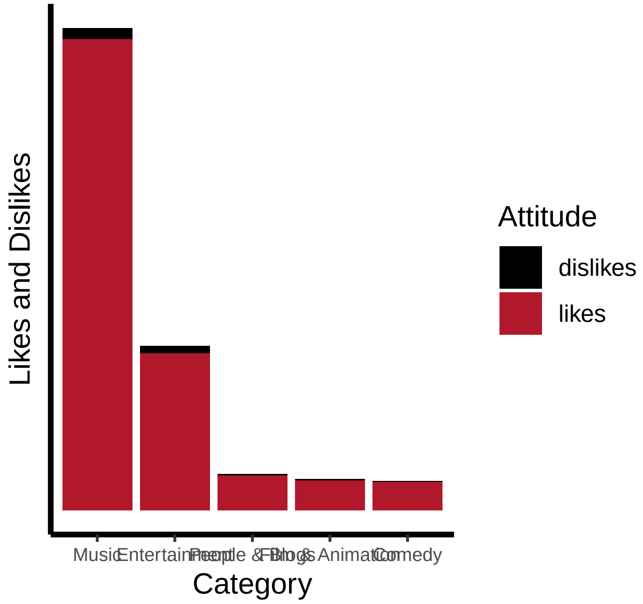


Figure 18. United Kindom

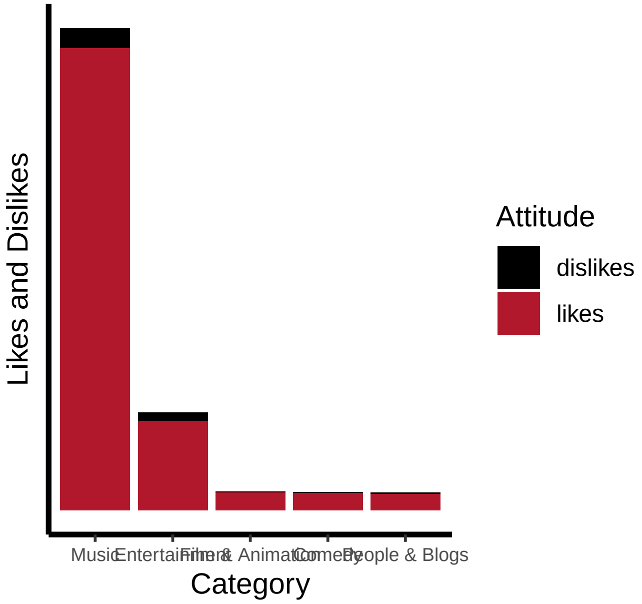


Figure 19. India

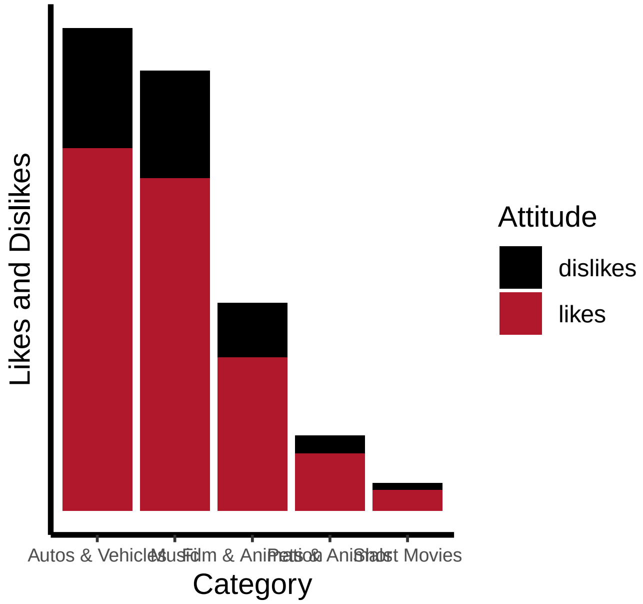
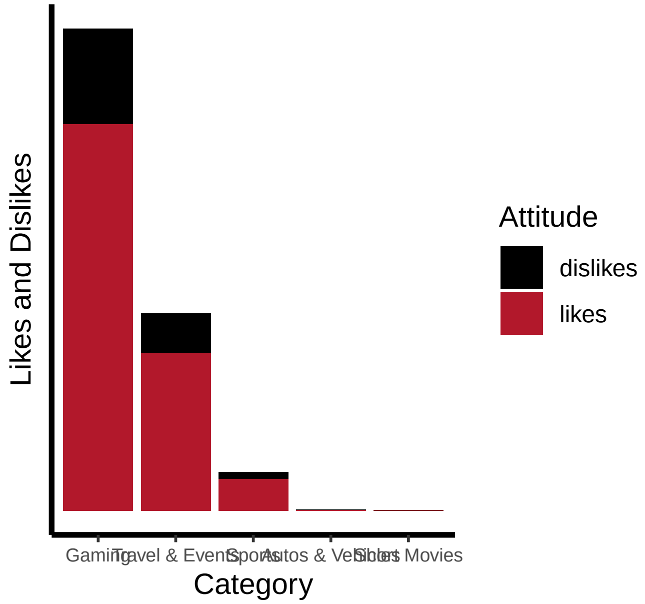


Figure 20. France



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

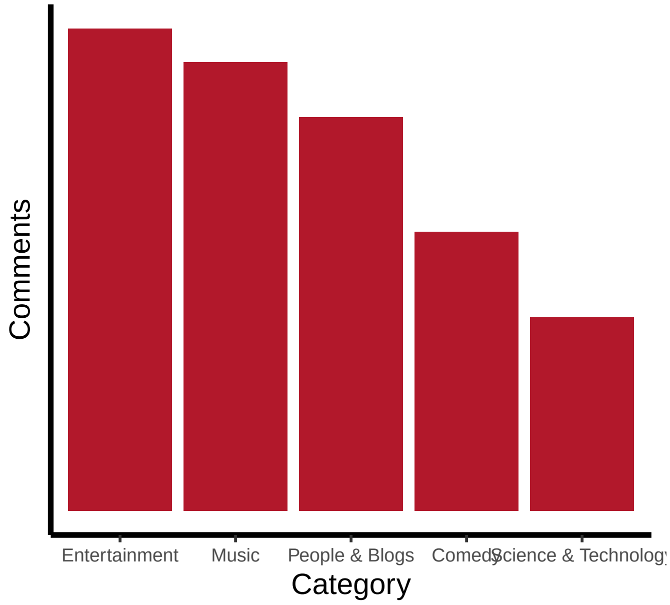


Figure 22. Germany

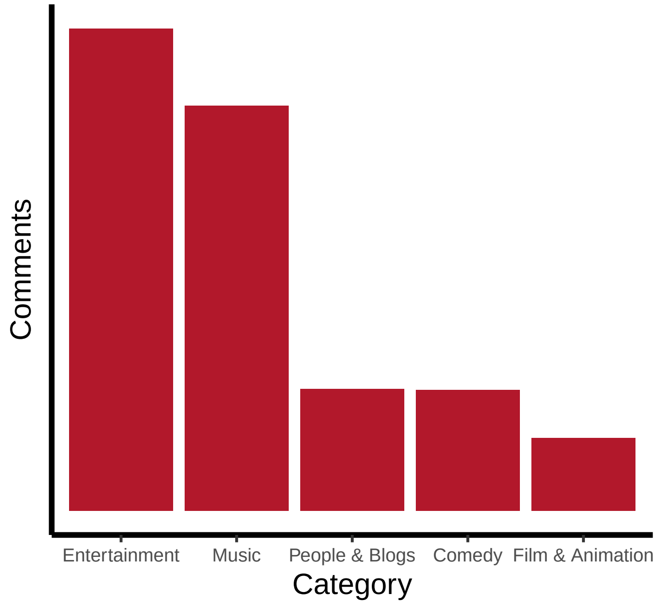


Figure 23. Mexico

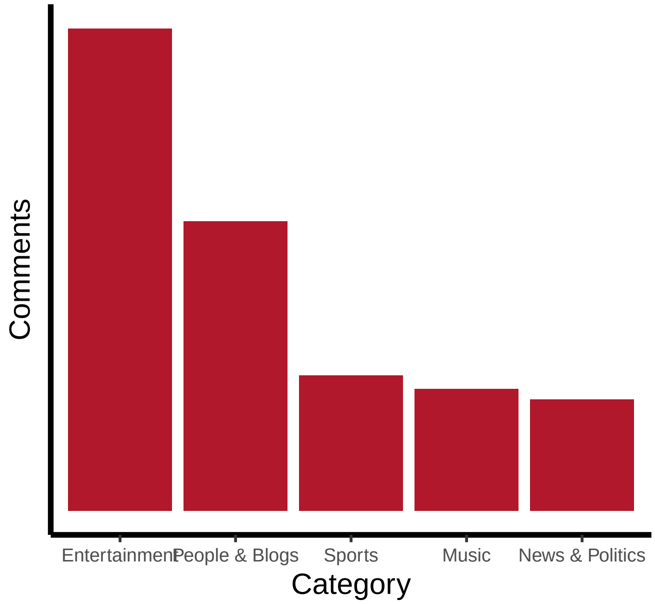


Figure 24. Canada

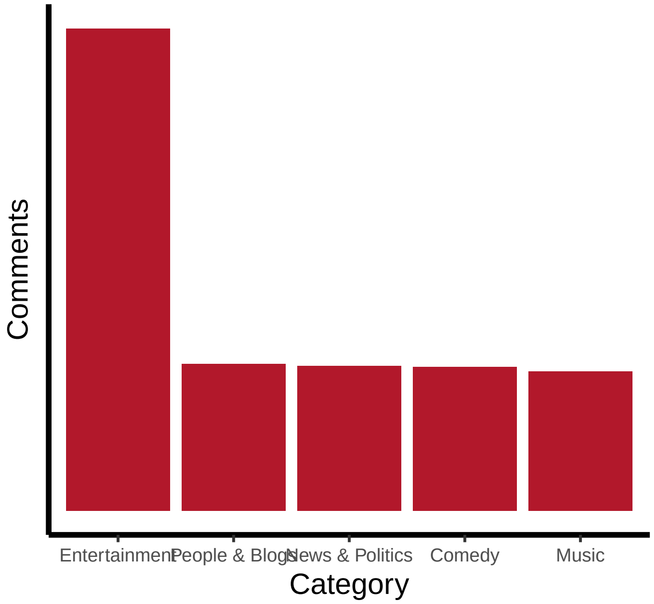


Figure 25. United States

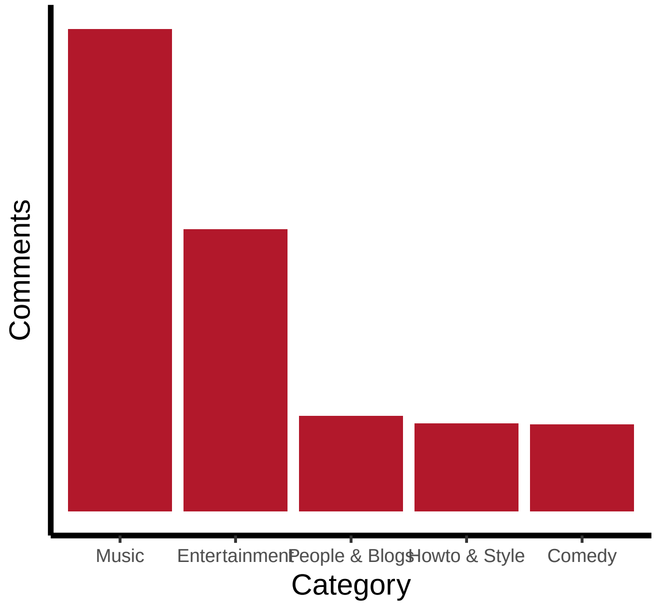


Figure 26. United Kingdom

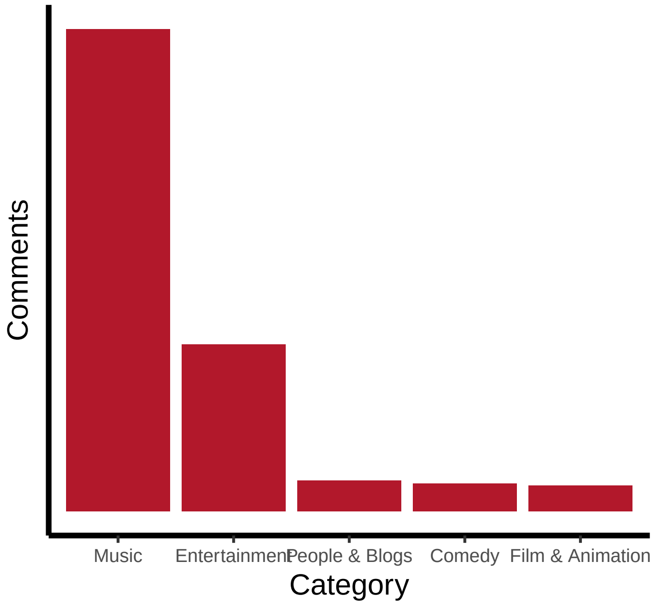


Figure 27. Japan

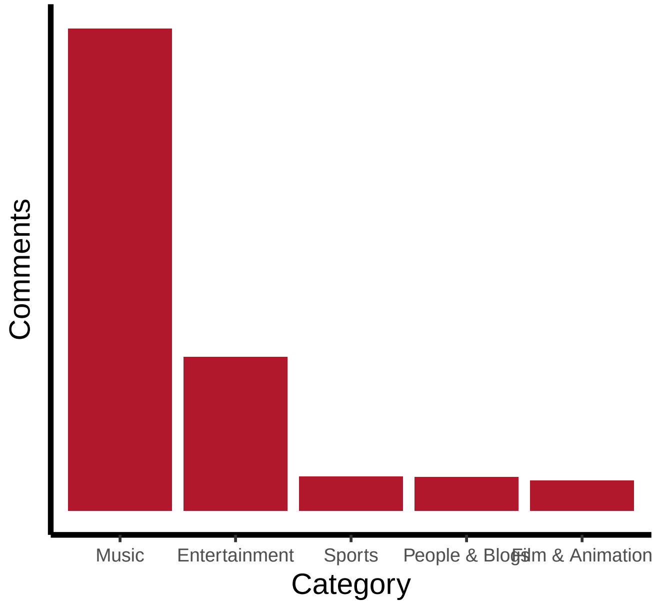


Figure 28. Korea

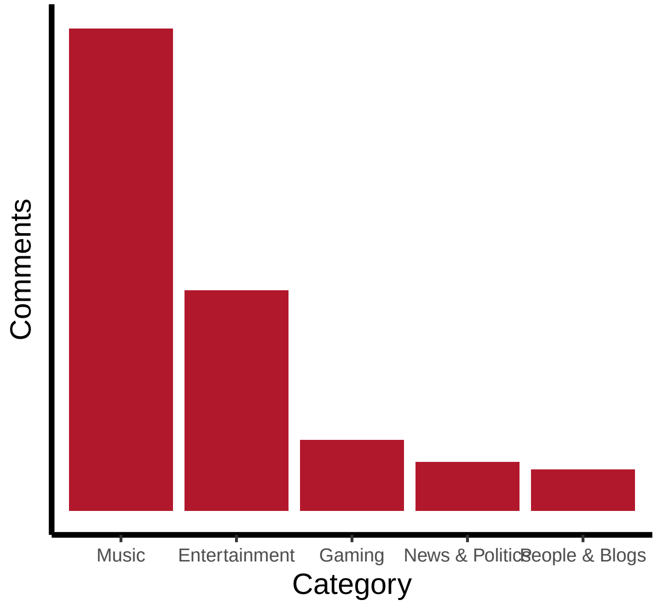


Figure 29. France

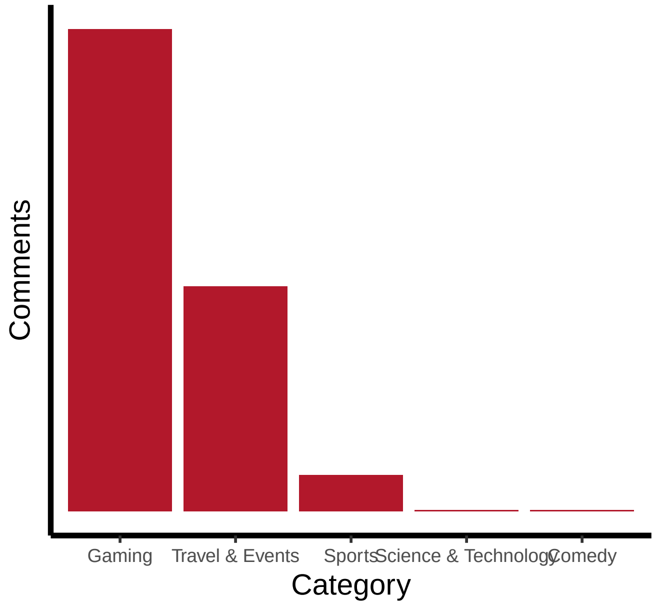
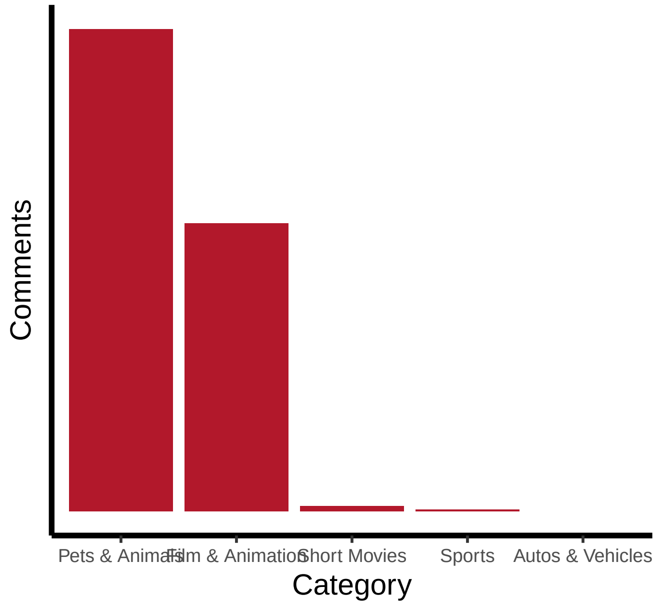


Figure 30. India



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

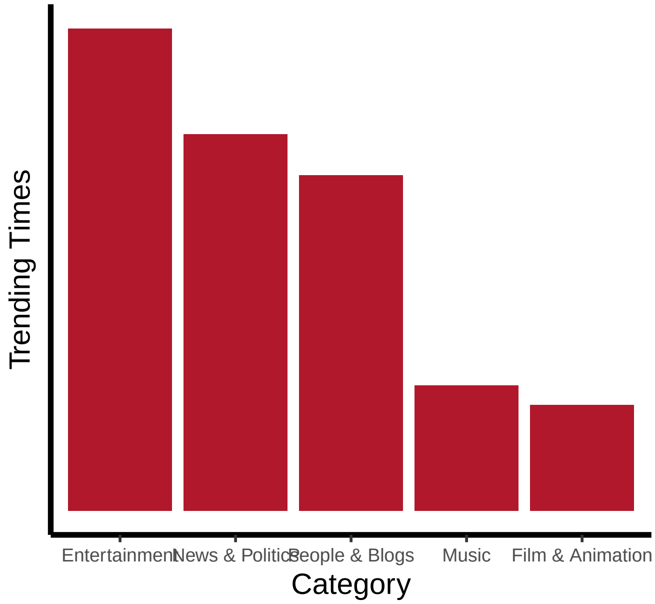


Figure 32. Germany

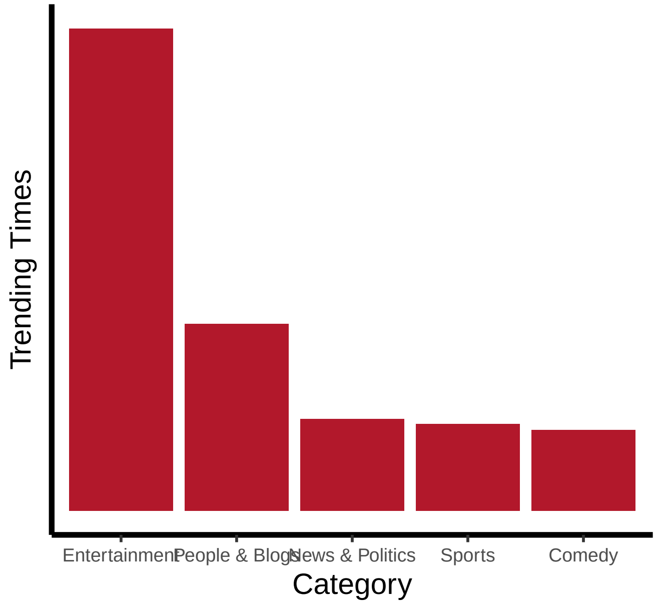


Figure 33. Mexico

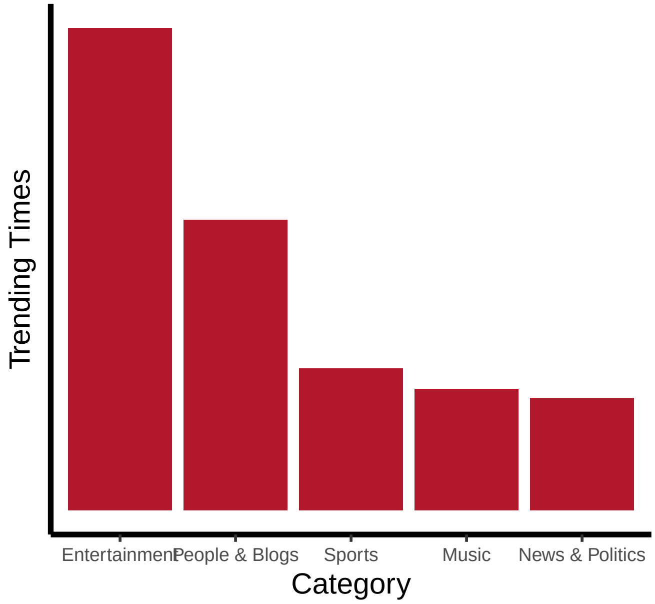


Figure 34. Japan

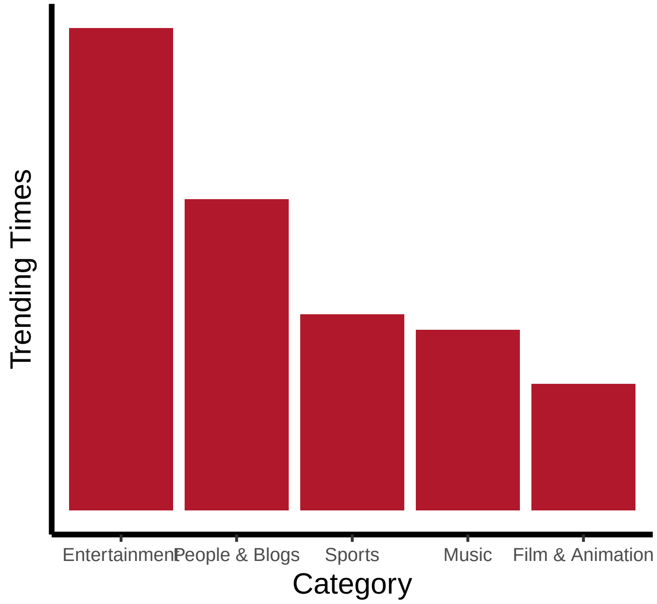


Figure 35. Canada

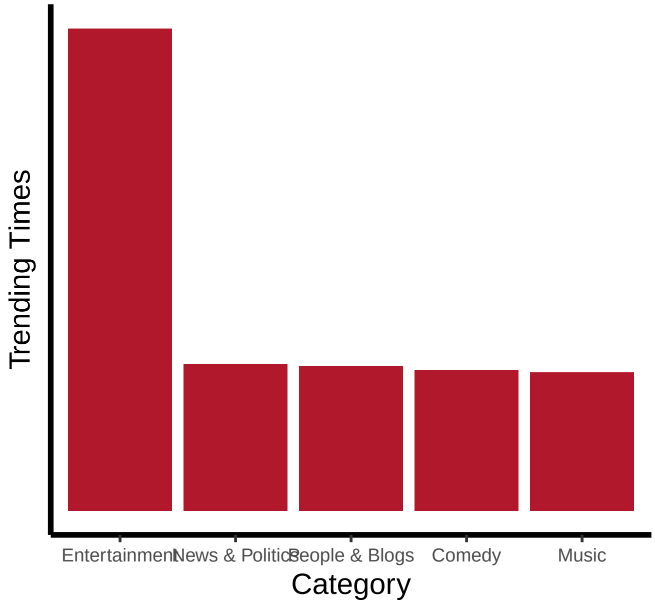


Figure 36. United States

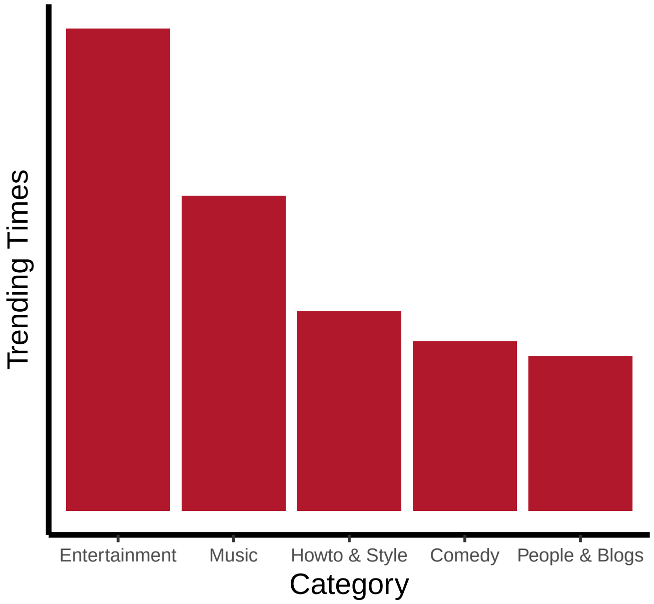


Figure 37. United Kingdom

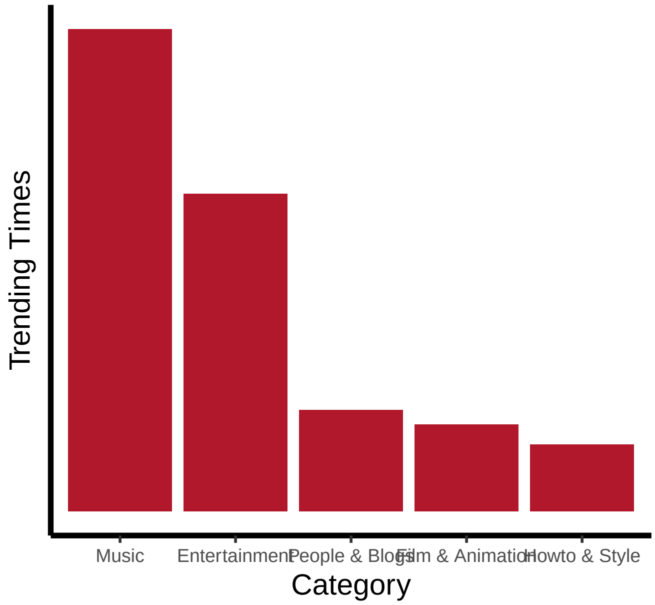
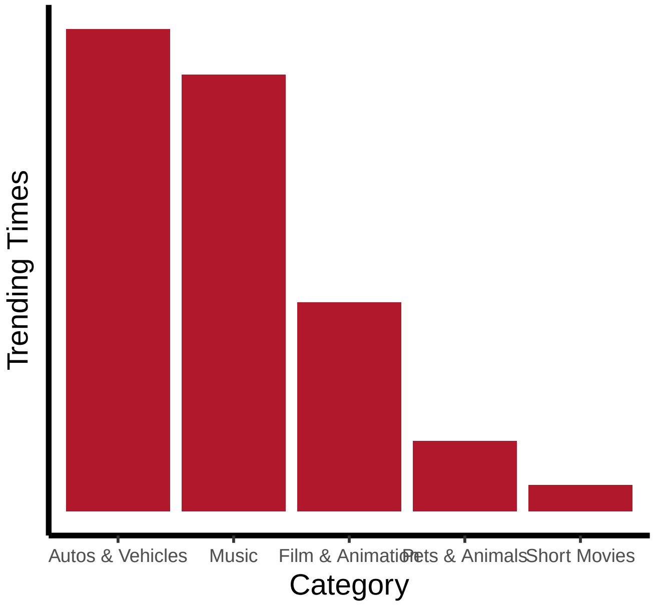


Figure 38. India



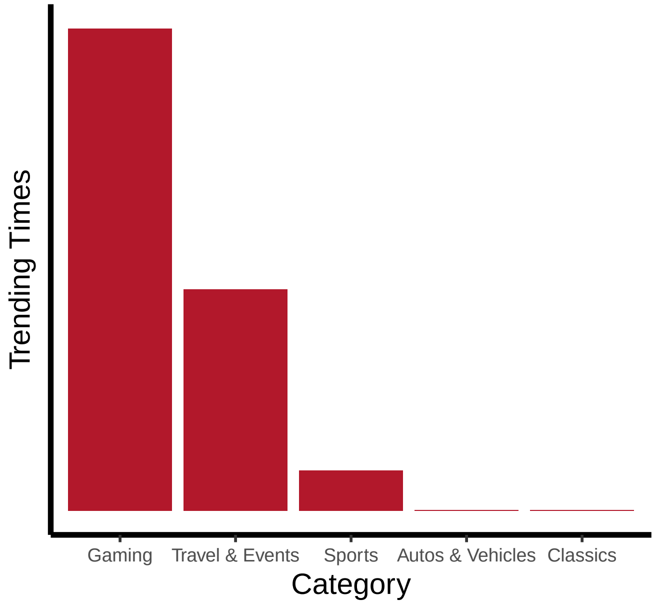


Figure 39. France

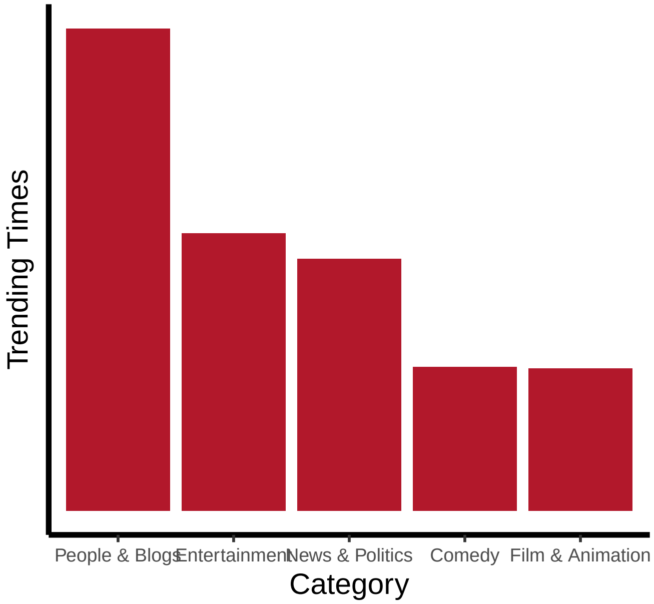


Figure 40. Russia

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):
  - Voting rate of a video=(likes+dislikes)/views X 100%;
  - Comment rate of a video=(# of comments)/views X 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

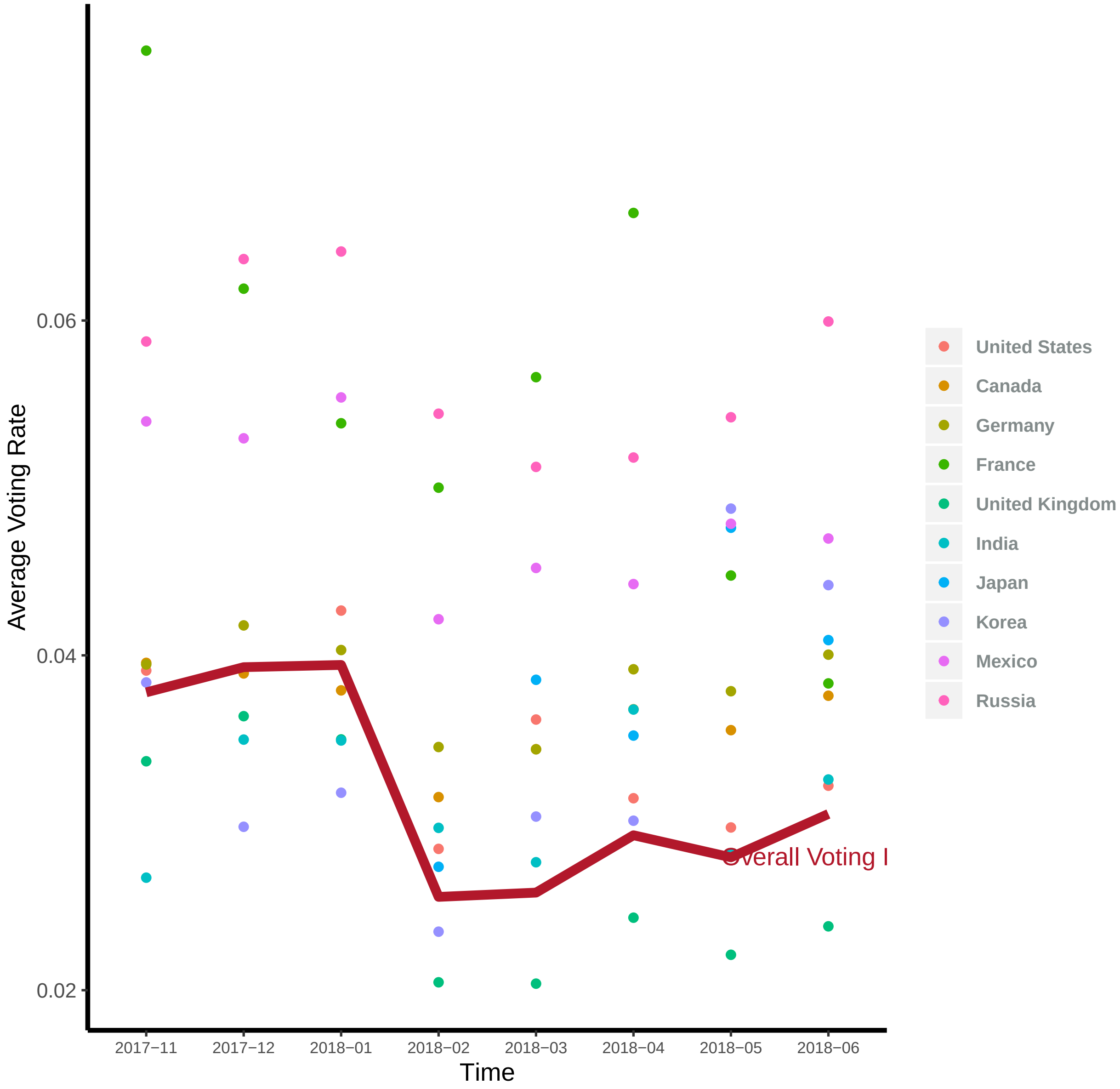
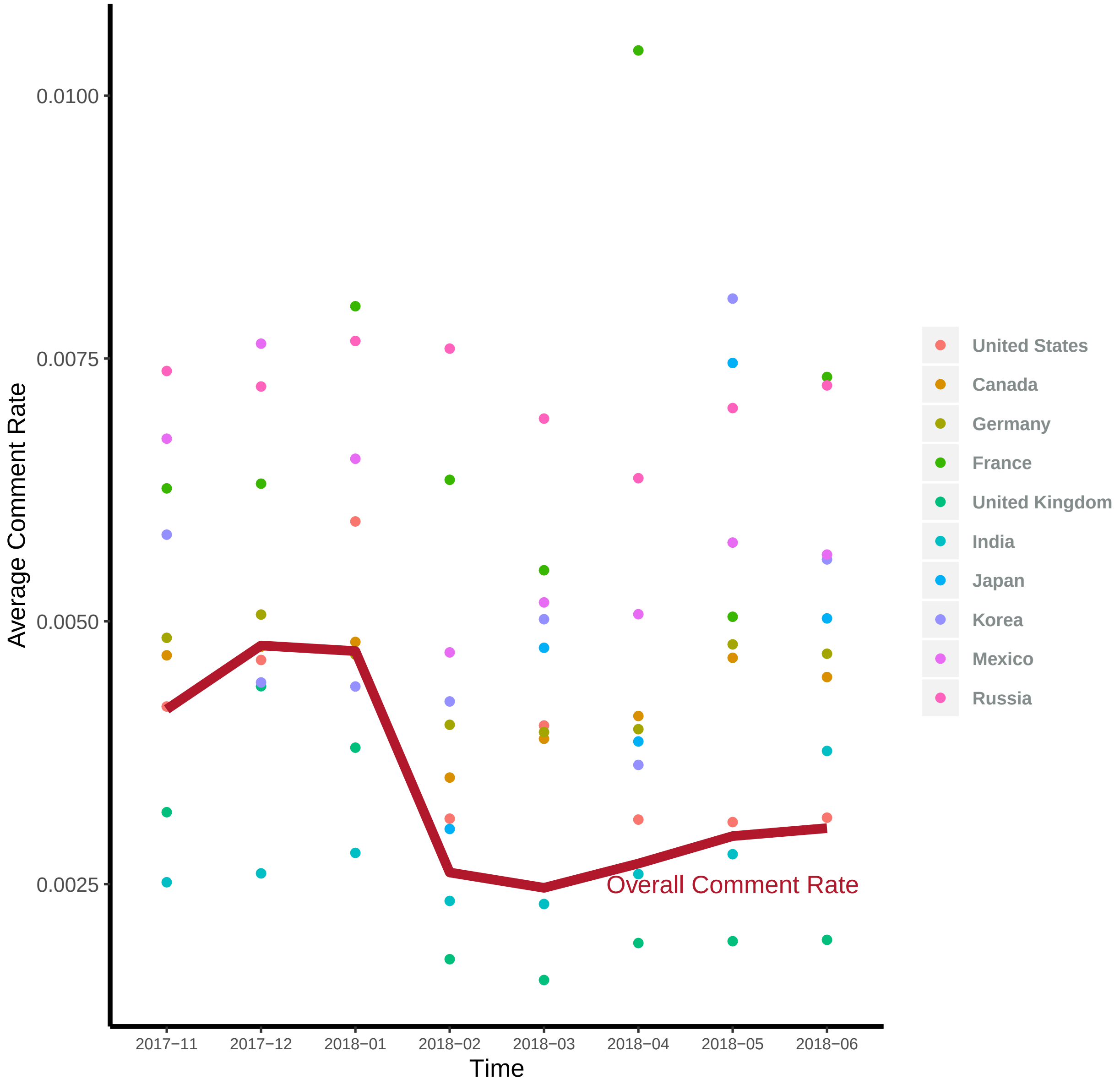


Figure 42. Average Comment Rate



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 two key time intervals:
  - 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.

- The average trending days for each category explain on average how long a video can keep attracting audience. The longer time interval is, the larger number of target customers 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).
  - Average trending days: It showed similar trend as the average going-viral time (Figure 51-60).
- By category
  - Average going viral time: Music videos take the longest time to go viral on YouTube. This may 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.
  - Average trending days: Shows and movies are outstanding among all categories.

Figure 41. Japan

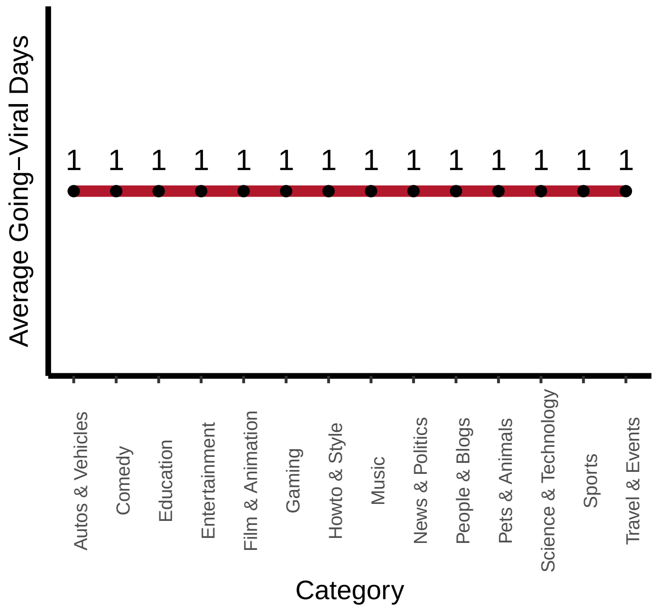


Figure 42. Russia

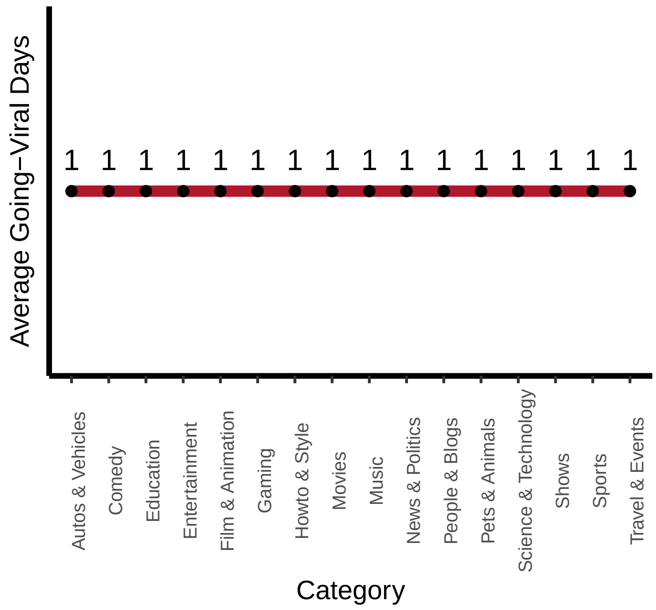


Figure 43. France

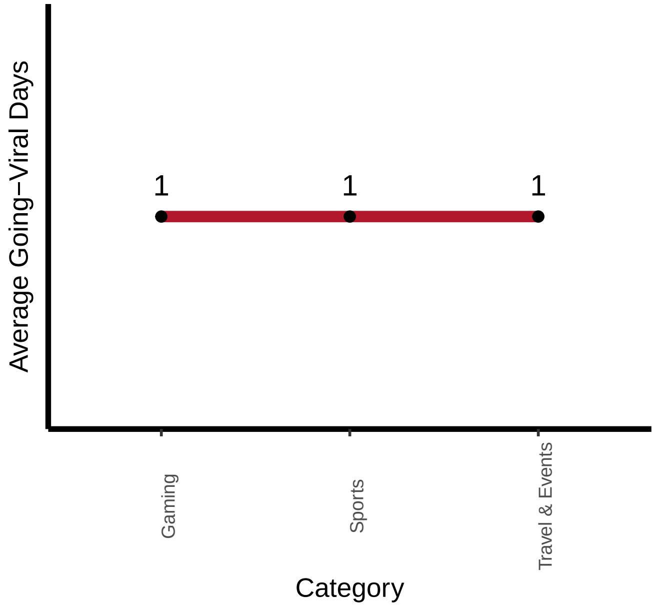


Figure 44. Mexico

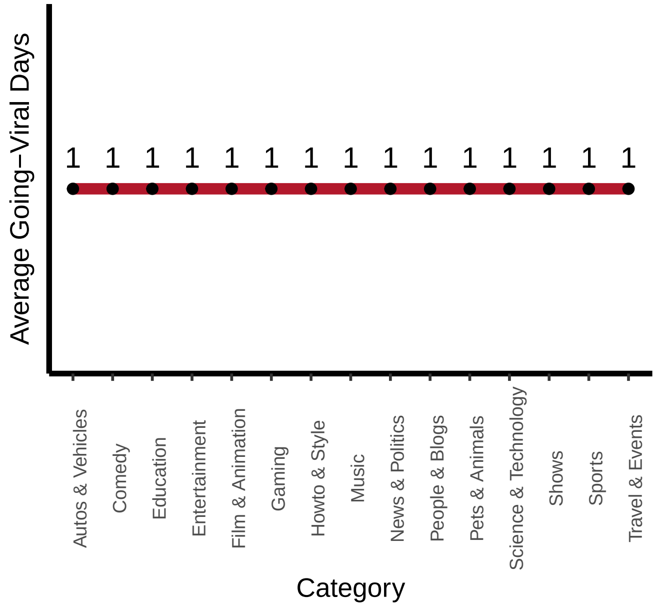


Figure 45. Germany

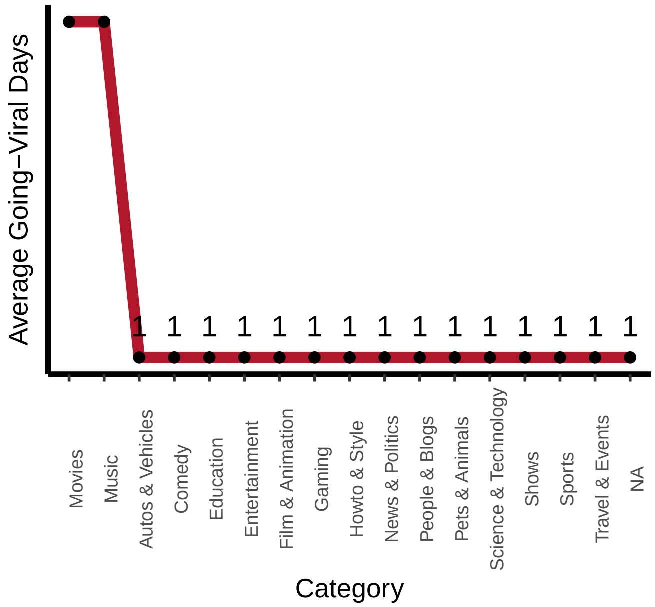


Figure 46. Canada

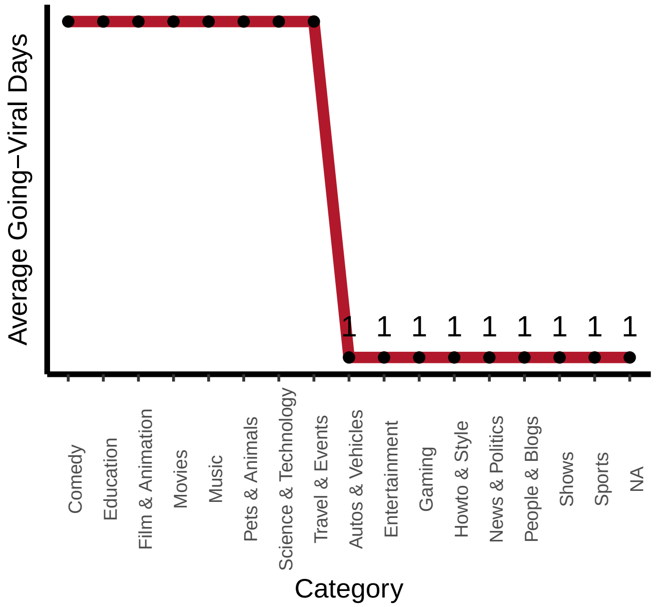


Figure 47. Korea

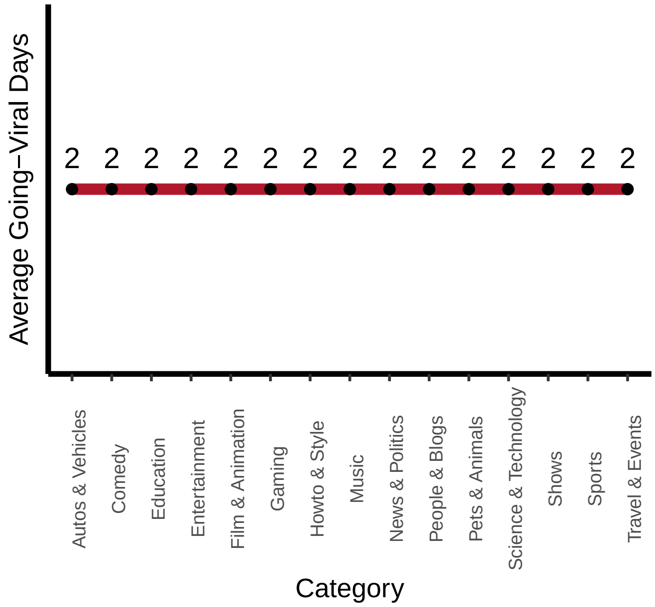


Figure 48. India

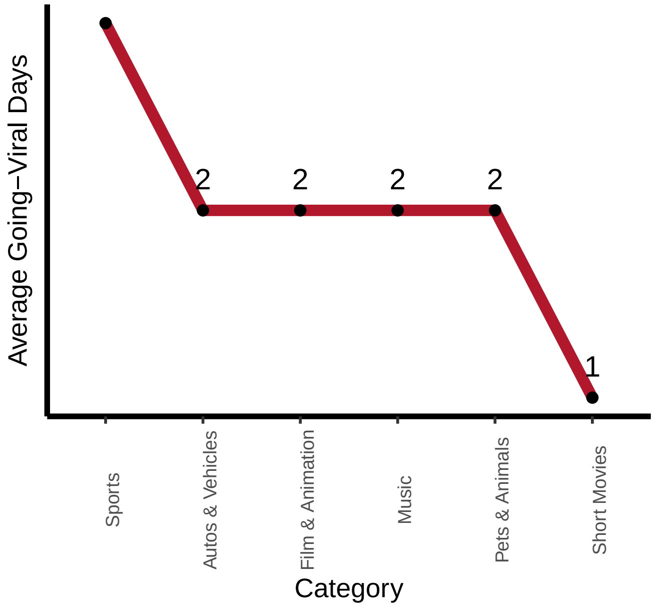


Figure 49. United States

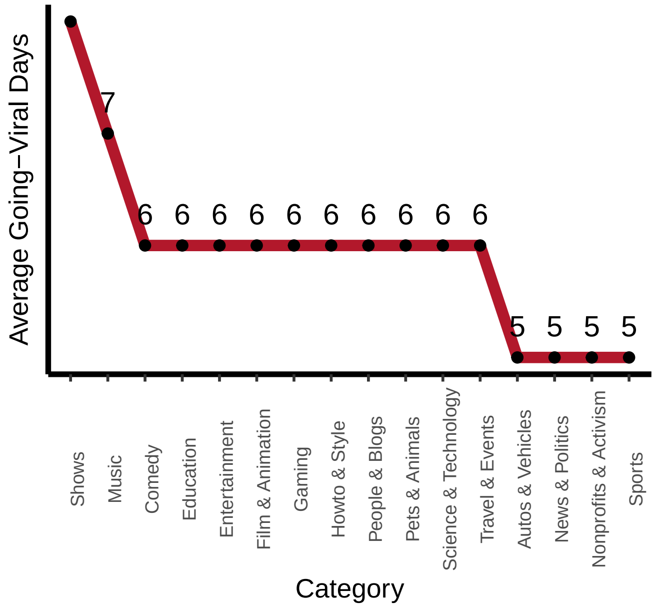
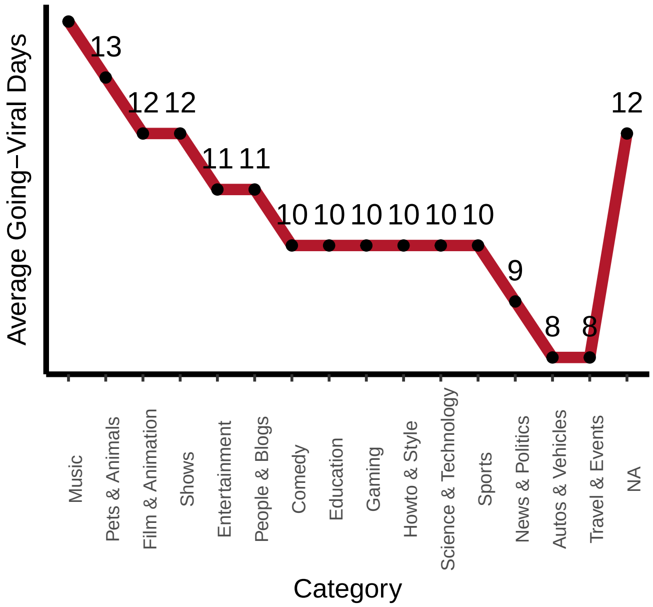


Figure 50. United Kingdom



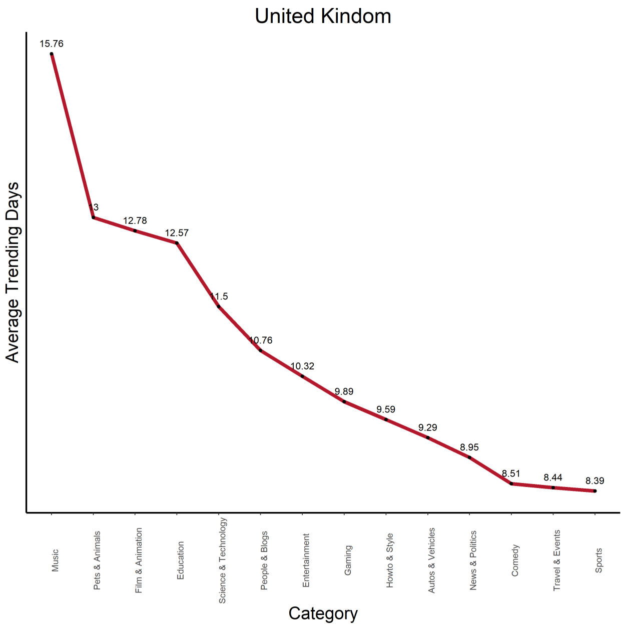
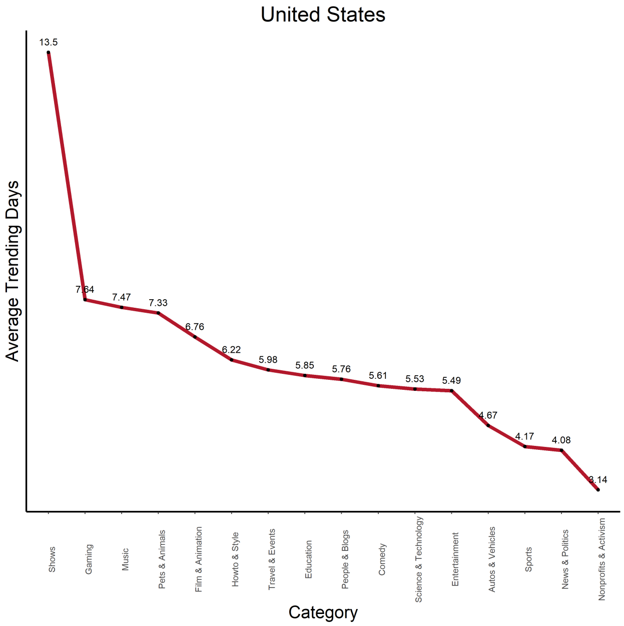
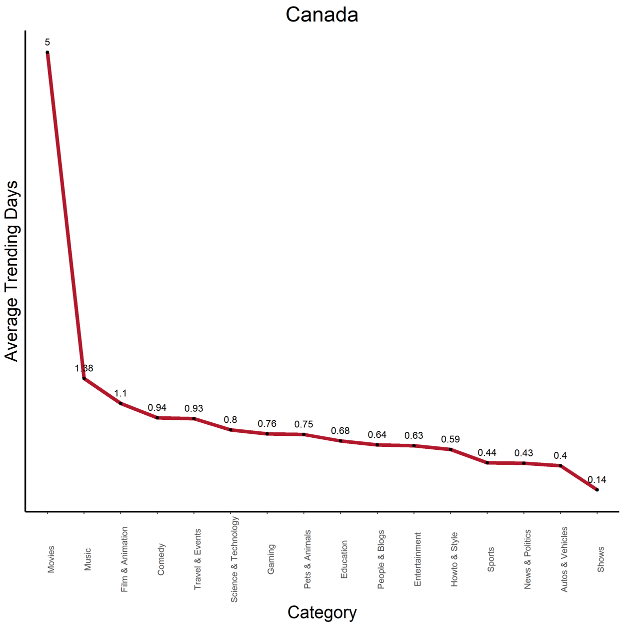
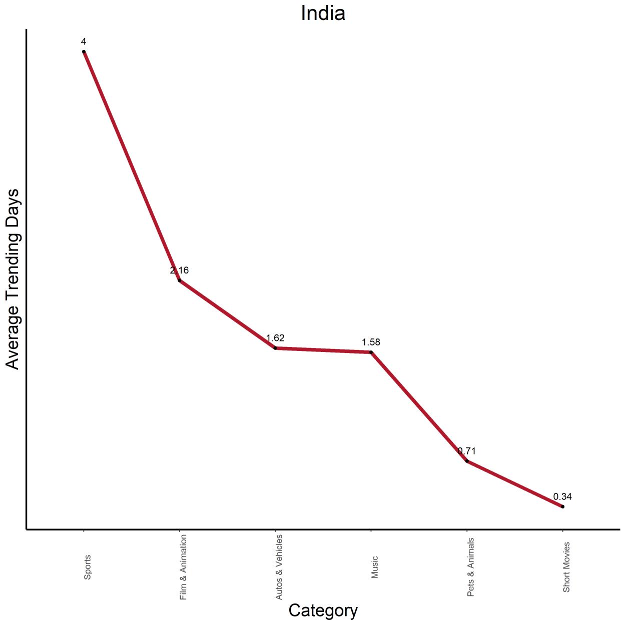


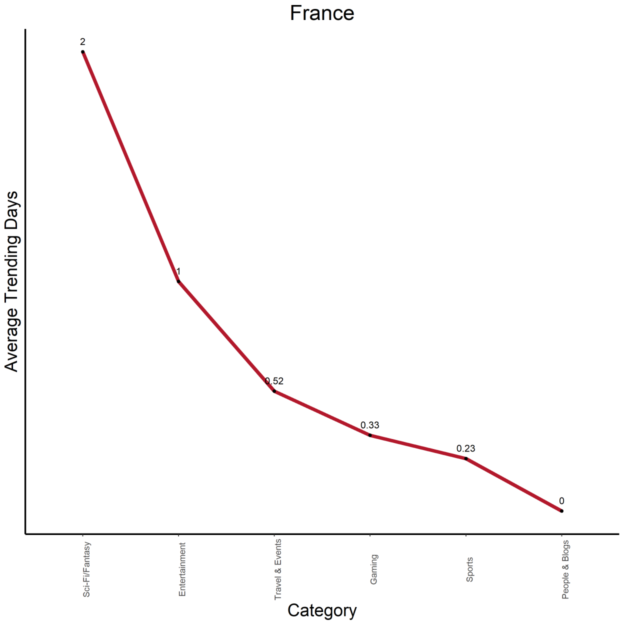
Figure 1: Figure 51-60

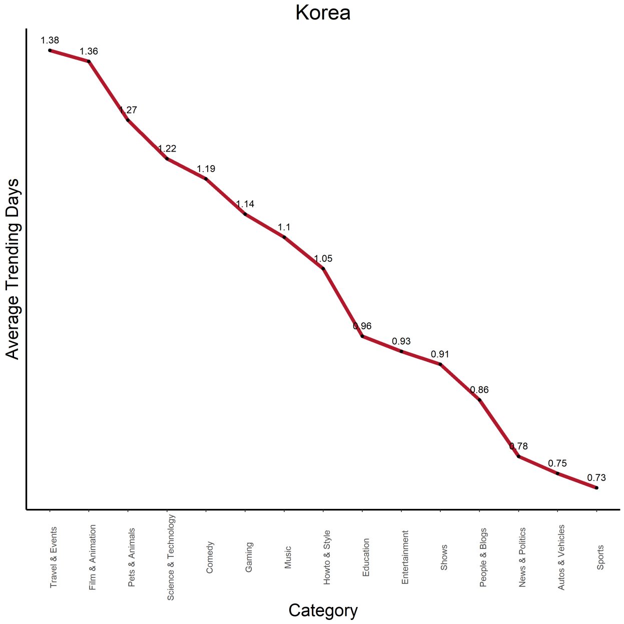
- Average Trending Days are shown below:

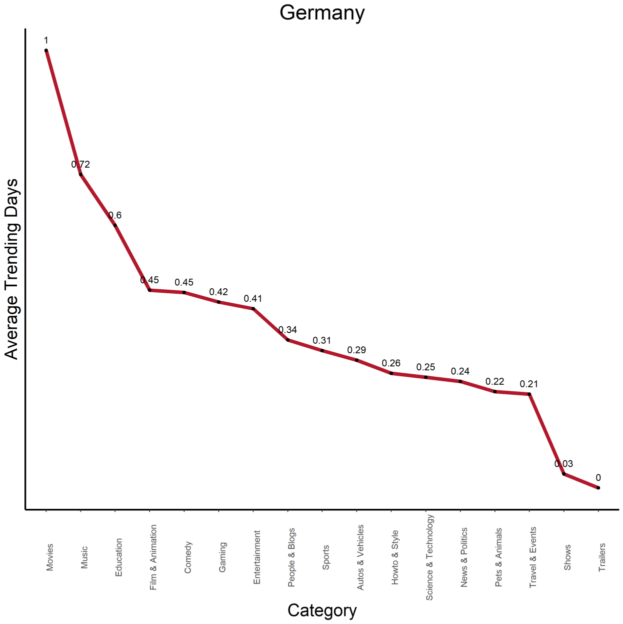


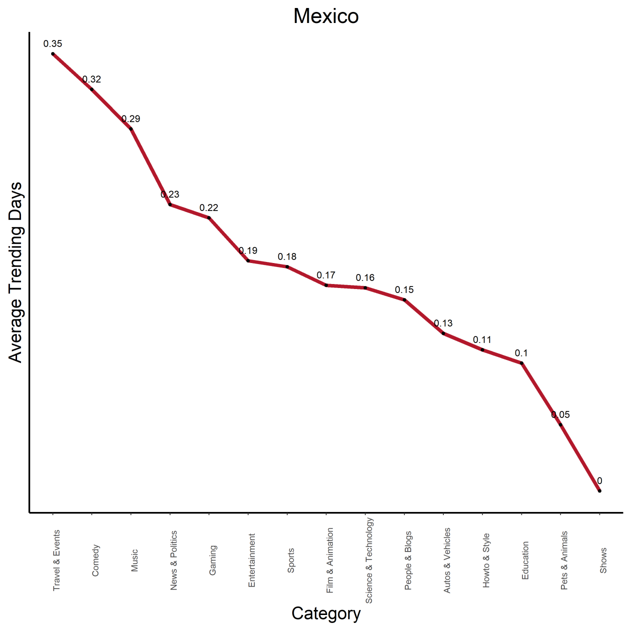


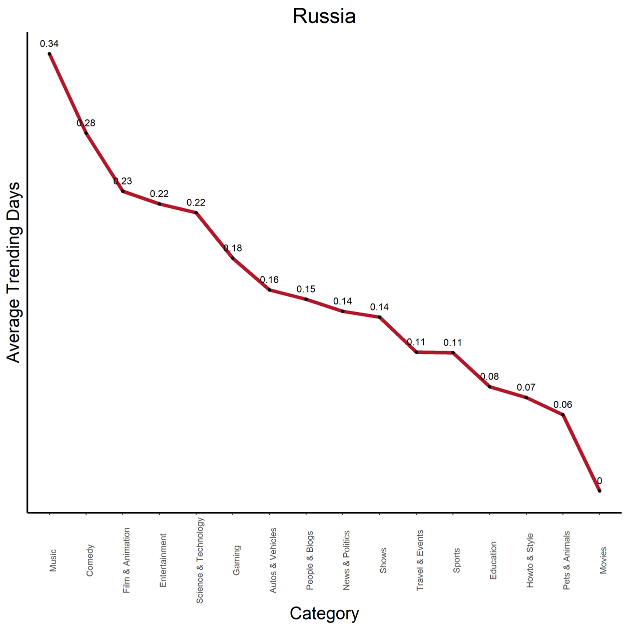

India
Average Trending Days
4
2.16
1.62
1.58
0.71
0.34
Sports
Film & Animation
Autos & Vehicles
Music
Pets & Animals
Short Movies
Category

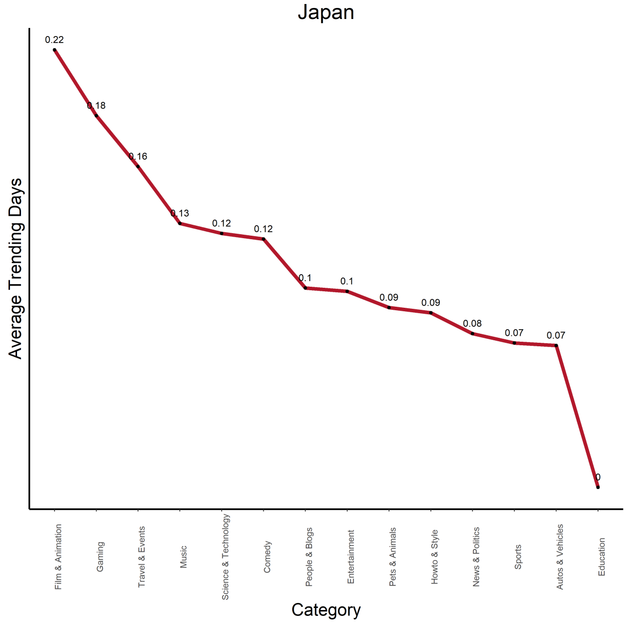

France
Average Trending Days
2
1
0.52
0.33
0.23
0
Sci-Fi/Fantasy
Entertainment
Travel & Events
Gaming
Sports
People & Blogs
Category


Korea
Average Trending Days
1.38
1.36
1.27
1.22
1.19
1.14
1.1
1.05
0.96
0.93
0.91
0.86
0.78
0.75
0.73
Travel & Events
Film & Animation
Pets & Animals
Science & Technology
Comedy
Gaming
Music
Howto & Style
Education
Entertainment
Shows
People & Blogs
News & Politics
Autos & Vehicles
Sports
Category


Germany
Average Trending Days
1
0.72
0.6
0.45
0.45
0.42
0.41
0.34
0.31
0.29
0.26
0.25
0.24
0.22
0.21
0.03
0
Movies
Music
Education
Film & Animation
Comedy
Gaming
Entertainment
People & Blogs
Sports
Autos & Vehicles
Howto & Style
Science & Technology
News & Politics
Pets & Animals
Travel & Events
Shows
Trailers
Category







4. Category Selection

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

Category_name	Views	VotingRate	CommentRat	GoingViralTime	Total_Score
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

Findings

- In the U.S., **How to & Style, Gaming and Comedy** are among Top 3 popular categories based on evaluation matrix, which partly different from the previous EDA, i.e. music, entertainment and gaming videos'influence continues.
- **How to & Style and Comedy** become new hitmakers and thus may be the next potential categories for advertisers.

Modelling

Deep dive to the algorithm of trending chart

Questions of interest

- How the number of views, likes, dislikes and comments influence the going-viral time?
- Is the influence vary across different countries?

Method

- Multilevel linear model with varying intercepts and slopes
- Predictors: Views (normalized because of scaling issues), Comment Rate (cv), Voting Rate (av)
- Response: Going-viral Days

Result and Discussion

- Basically, the higher voting rate, comment rate and views would lead to shorter going-viral days.
- This influence varies across countries. Especially in U.S. and UK, the comment rate and voting rate play more important roles compared with other countries.
- In UK, since the intercept is largest, chances are are that there exist other more significant factors affecting the going-viral days.

```
## Loading required package: Matrix
```

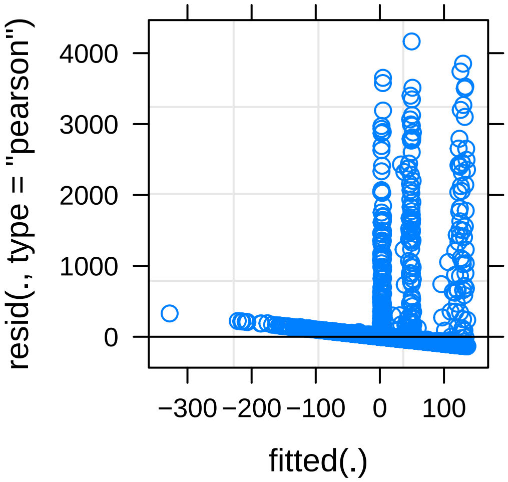
```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## boundary (singular) fit: see ?isSingular
```

	(Intercept)	av	cv	view_ct
us	51.227547	-544.335586	-689.234000	-2.2686074
ca	5.064221	-44.584420	-60.814239	-0.1982068
de	2.077616	-9.171318	-8.250338	0.0031592
fr	4.977902	-34.674242	-35.013804	-0.0137654
gb	135.164295	-1369.302398	-1645.735363	-4.4176161
ind	1.454879	-2.960247	-2.349352	0.0186521
jp	1.702033	-11.640025	-20.677526	-0.1108470
kr	2.895657	-14.744144	-12.163082	0.0297100
mx	2.766321	-15.008470	-12.998389	0.0051266
ru	3.754060	-21.578342	-22.155093	0.0327222

** Model checking ** As is shown in the residual plot, there is no clear pattern.


resid(., type = "pearson")
fitted(.)