Christopher Dillard
MIS 505
8/18/2023

Final Project – "YouTube Success Requires High-Quality Niche Content Creation"

My story for this dataset is based on the need for high quality content in a popular category to reach massive success on YouTube using the following dataset:

https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023

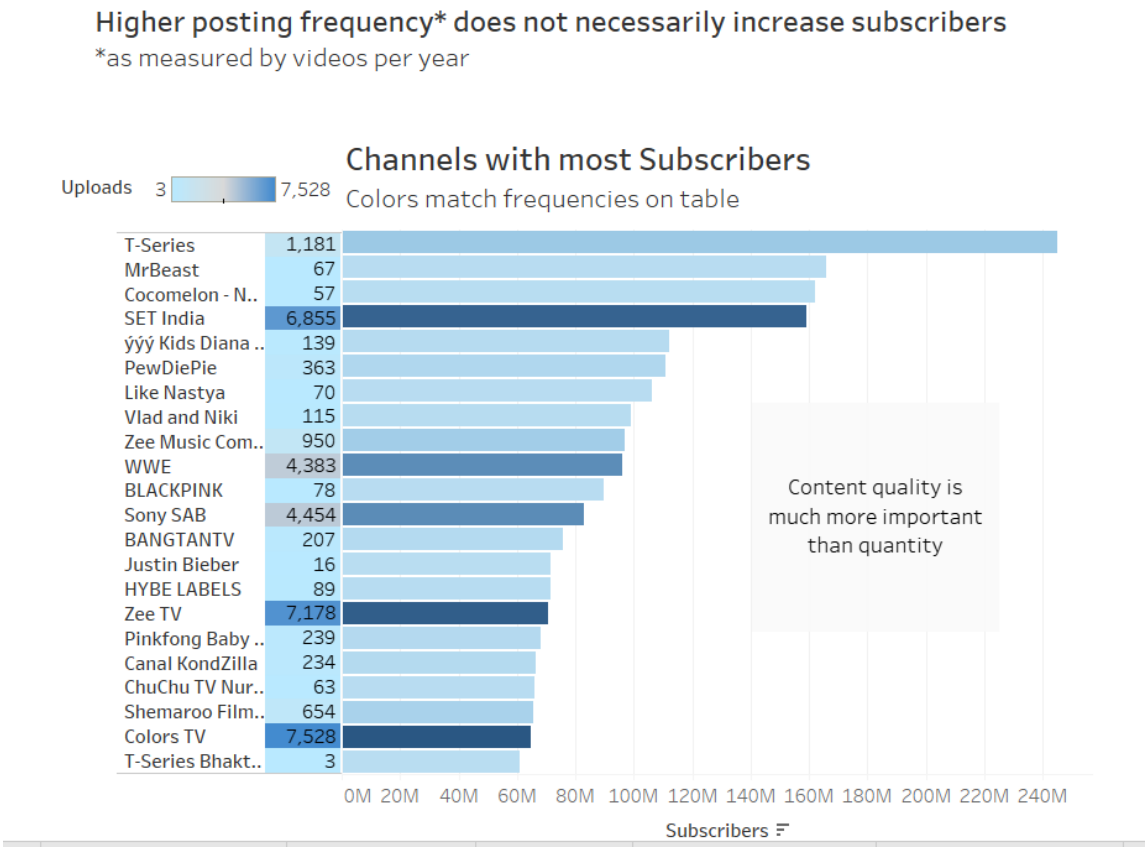
I found the dataset that I used for this project on Kaggle. It contains 995 of the most popular YouTube channels in terms of their subscriber count, among several other interesting variables. This dataset lends itself to an insightful analysis of the factors determining channels success on YouTube and which categories are most popular on the video sharing platform. This dataset is, however, not perfect and does contain some missing values, so I decided to only analyze the variables that would lead to a clear conclusion and left out some of the demographic information that contained things like the unemployment rate of the country of origin, the gross tertiary education enrollment percentage, and the urban population. When looking at the year each account was created, I identified one outlier which was listed as having been created in 1970, which of course predates the creation of YouTube by 35 years. This was YouTube's official account which I felt was not useful for analysis, so I excluded this channel from my research.

I created a calculated field for the age of the channel which I used to gain insight into how the quantity of videos posted per year (as a ratio between total videos and channel age) could be a determining factor in the success for channels and whether the frequency of posting for the most successful channels differs across categories. In other words, I intended to measure whether the most successful bloggers were frequent posters and whether the popularity of accounts for popular music artists whose accounts might just contain music videos could be less contingent on this factor since they may not have relied on YouTube to rise to fame. Unfortunately, there are a few instances where values of zero (null values) in certain places do not make sense like the count of videos uploaded being equal to zero despite the channel registering millions of views and currently having videos on the site upon checking. This led me to ultimately filter out any uploads per year to be at least 1 when creating the table, which yielded 3 videos per year as the smallest.

Visualization designs and key takeaways:

Christopher Dillard
MIS 505

8/18/2023

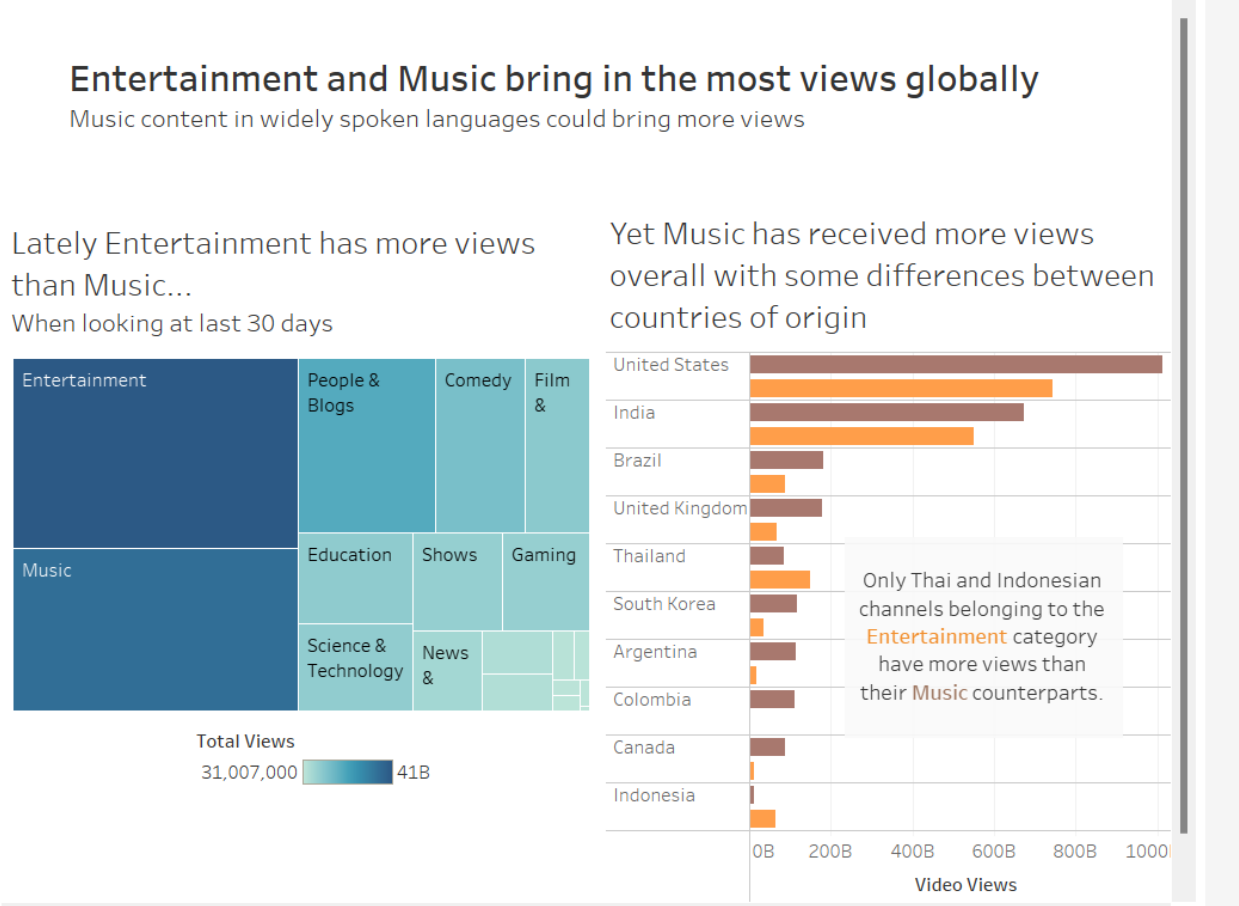


With this first story point I wanted to draw attention to how the number of videos a channel uploads does not necessarily lead to more success in terms of subscriber count. At the same time, I wanted to set the tone for the rest of the story by introducing the idea that content success depends heavily on its quality despite this being a difficult factor to quantify.

To create this visualization, I combined a table with a bar graph and then eliminated some of the headers to make sure that the bars aligned with the cells in the table. Here I encoded the posts-per-year in color and subscriber count with the length of the bars. I did need to make the color of the table slightly different to accommodate the text, but I didn't feel like doing so detracted from the overall look and interpretability of the visualization.

Christopher Dillard
MIS 505

8/18/2023

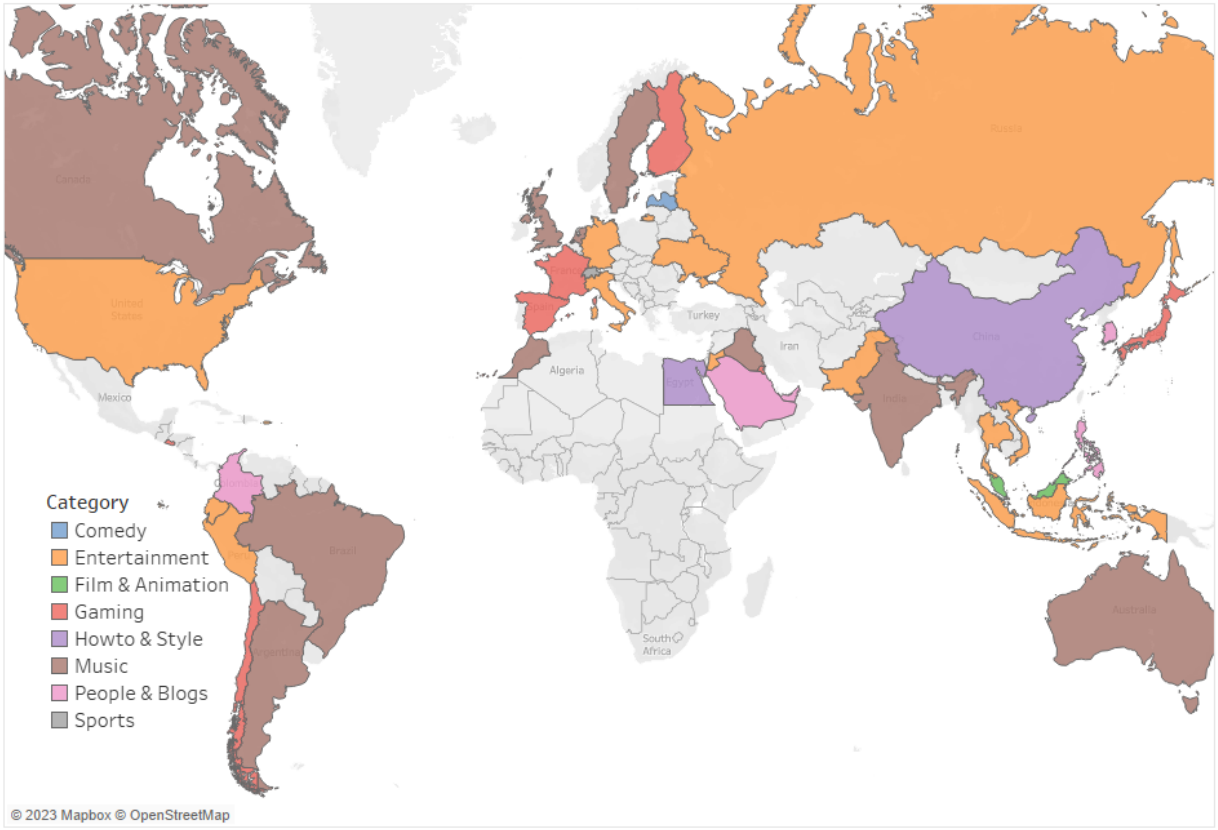


I found it helpful to visualize what the most successful categories were and wanted to use up the space on this story point wisely. Here I opted for a tree map to show just how much more views content belonging to the Entertainment and Music categories got than the other ones in the last 30 days. I wanted to compare this to overall totals in some way by also shedding light on the differences that arise between view counts when looking at a channel's country of origin. This bar graph displays the top ten countries of origin when looking at the combined view count sums for both categories. Music got more views than Entertainment, but why? I suggested in the top subtitle that language might be a key factor here when looking at widely spoken languages that the content is being produced in and how that could lead youtubers from countries like Colombia and Canada to reach broader audiences in larger markets given the international presence of English, French, and Spanish. As for South Korea, K-Pop is a popular genre which could explain how they ranked so high on the graph. In the annotation I pointed out the two exceptions to the trend which goes to show results vary across markets.

Christopher Dillard
MIS 505

8/18/2023

What type of content do youtubers with the most suscribers from each country make?

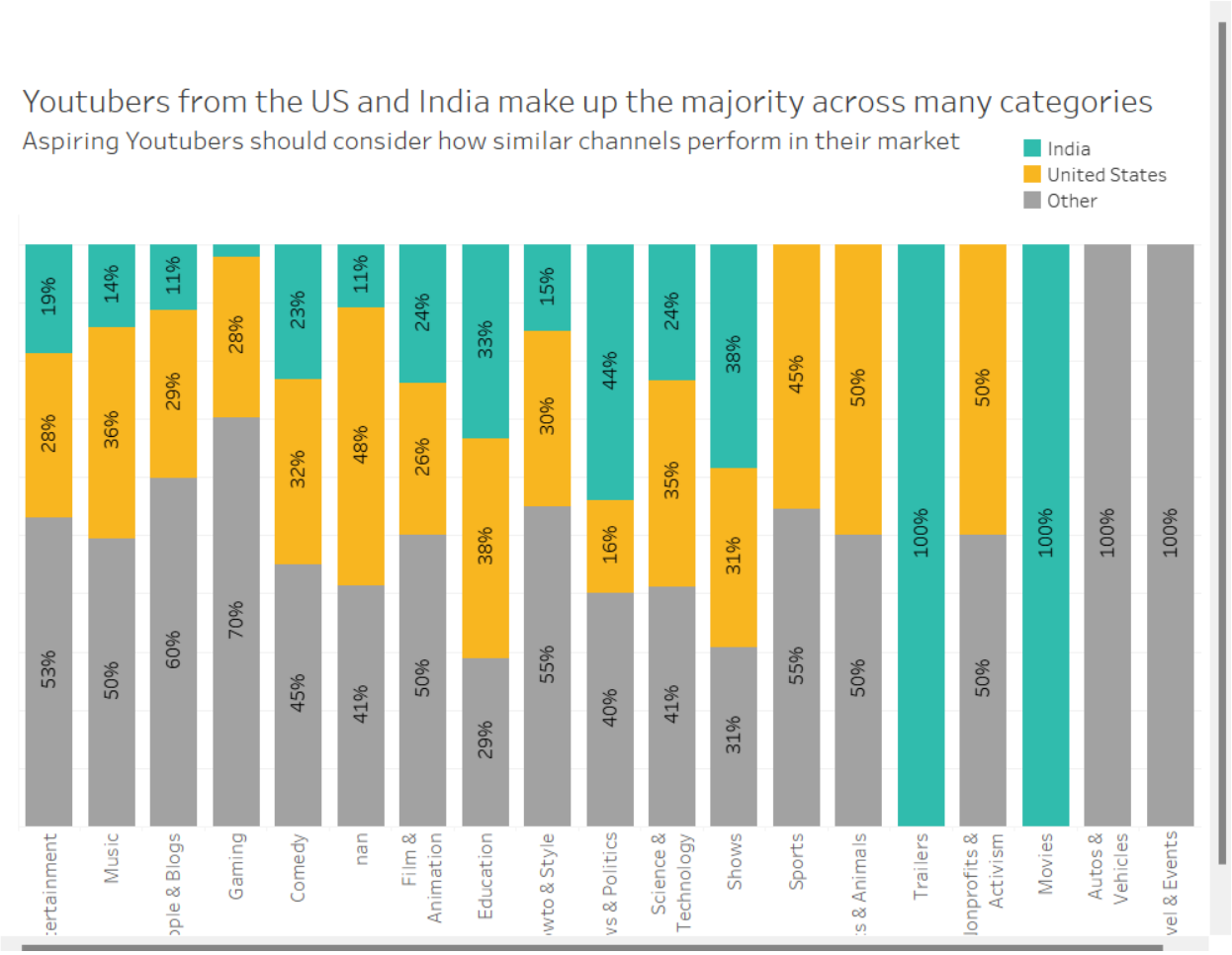


This map requires the audience to really stop and study the colors to draw any conclusions which is why I left the title as an open-ended question. Some trends that I picked out is how the most successful channel is from Japan is dedicated to gaming, which I thought was interesting considering that Nintendo, Sega, and Sony also hail from this country and are major names in the video game industry. Music and entertainment appear in many countries as the most successful category which was displayed in the previous story point's tree map.

7 categories were not the most successful in any of the countries here, being: News and Politics, Education, Science and Technology, Autos and Vehicles, Pets and Animals, Travel and Events, and Non-Profit and Activism.

Christopher Dillard
MIS 505

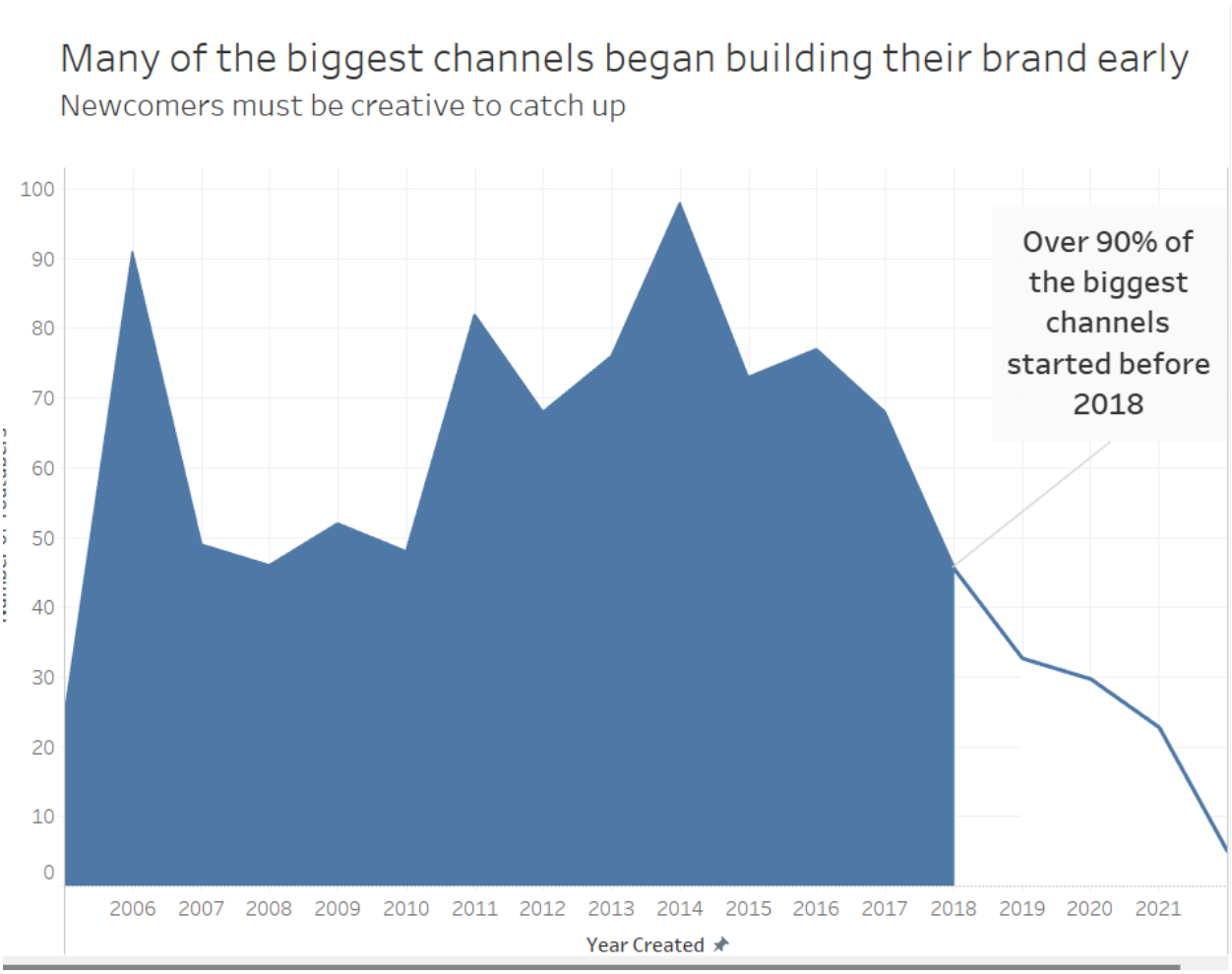
8/18/2023



By highlighting India and the US here, I wanted to draw attention to how these two nationalities are the most popular among the most successful youtubers, even though they may not be present in every category. This further alludes to the fact that content should be tailored towards a specific target audience by capitalizing on certain biases that could exist in how people perceive content from different countries by studying what works best. Given the popularity of Bollywood films it was no surprise that Indian youtubers were so successful in the movie and trailers categories. However, it is noteworthy that Indian youtubers made up 100% of the channels in these categories in the dataset. This was the easiest graph to make in the entire project, but it does present the issue of legibility in terms of the orientation of the text, which unfortunately I could not find a better alternative given the length of the category names. So, I opted to emphasize the overall trend as opposed to focusing on specific categories here.

Christopher Dillard
MIS 505

8/18/2023



I wanted to end the story with a more compelling graph to really sell the idea that youtubers who want to reach high success need to be extremely creative when competing against channels that have already had a large head start on the platform. To get the data behind the annotation, I just did a simple count of how many times each year appeared in the "Created Year" dimension to then divide by the total. This came out to roughly 94% for 2005-2017 so I just rounded down for a cleaner number. To create the graph, I then needed to create a group for these years and assign the rest to another group. Since no one year can belong to two groups at the same time, I needed to find a way to overcome how the color wasn't filling in continuously in the graph by making the groups 2005-2018 so that the shading stopped on the 2018 mark to include the entire year of 2017 when I color filled the area under the line. I then colored the other group white, which I think led to an easy-to-interpret graph.