Youtube Trending

AN EXPLORATORY D3.JS TOOL FOR WHAT WE WATCH

Authors:

Saul Woolf *(saulvhwoolf)*
Jon Venne *(jonvenne)*
Drew Ciccarelli *(dcicc4)*

Website:

https://saulvhwoolf.github.io/DataVisFinal/

Github Repository:

https://github.com/saulvhwoolf/DataVisFinal/

Overview and Motivation

We all watch YouTube, and often find ourselves browsing the trending page. Some days the video presented are hilarious or fascinating, and some days they are just not good. We wanted to make a tool to explore YouTube video data in order to further an understanding of the trending algorithm. Trending videos are the top 200 hundred videos chosen by the YouTube algorithm as the "most popular" on any given day. Our goal was to create a tool that allows people to explore YouTube data, and determine what tends to make a video trend.

Related Work

Drew did a data-visualization-oriented IQP in Venice and created a system that allowed users to construct their own visualizations and export code to embed them in their own websites. sandbox.veniceprojectcenter.org and dataviz.veniceprojectcenter.org/visuals are the tool and display sites respectively. The following is a screenshot of the bubble chart tool.
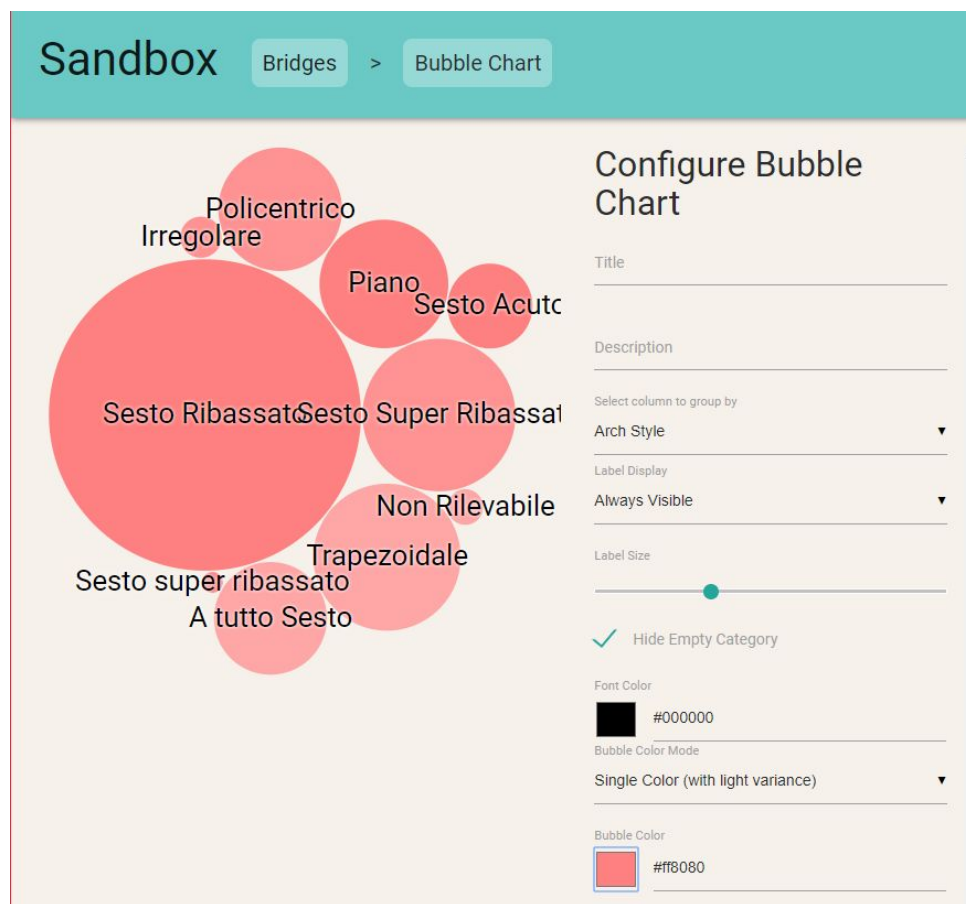
Questions

How trending videos differ from country to country?

We were very interested to see how this one would play out, but upon a deeper investigation of our data, it became evident that trending data did not differ from country to country greatly. As such, this question was dropped.

What correlates most to if a video is trending (likes, views, etc.)?

This was the main goal of our tool. A deep exploration of the data with our tool should enable a user to determine these correlations. We believe the main contributing factor is views. Likes, dislikes and comments don't seem to have as significant an impact.

How do downvotes affect trending videos?

Some videos were massively disliked and still remained on trending, so it seems that as long as a video brings in lots of views it can reach the trending page (i.e. logan paul)

What categories of videos are most often trending?

It was pretty clear that Music is the most popular category on youtube. That being said, it had less likes and comments per view than the other popular categories.

What tags are found most often in trending videos?

The tags "funny", "comedy", and "humor" were 3 of the top 4 tags, and 7 of the top 20 were comedy-related.
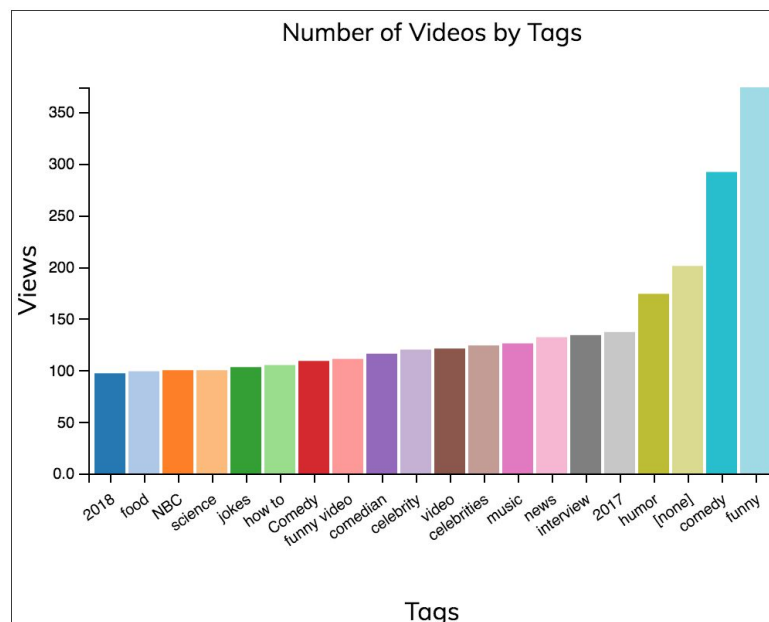
What effects do disabling likes and comments have on videos?

This did not seem to affect a video's ability to become trending, as controversial videos (i.e. the Roy Moore campaign statement) bring lots of views.

As the project progressed we dropped a few of the questions, but our tool enabled us to answer quite a few. We focused on looking at certain tags or categories and how they broke down, in terms of likes, dislikes, etc.

Data

We collect our data from https://www.kaggle.com/datasnaek/youtube-new/data . This database contains information on all trending videos between December 2017 and January 2018. It is also updated frequently, but we were satisfied with the 200 videos per day for the 2.5 months. The csv file contains information on likes, comments, dislikes, and whether these features were disabled. It also includes all categories and tags.

We performed some cleanup on the data. We converted the CSV file, which presented each video once for each day it was active, to a JSON file of video objects where one attribute lists all days where it was active and each day will have the likes, comments, and view counts for that video on that day. This allowed us to easily calculate the derived value of how many days a video was trending for as well get cumulative data for an individual video. Finally, the data originally used an number id system for each category, we replaced each category id with the appropriate category String. The file was converted with a python script which additionally calculated any missing values. For efficiency purposes, we derived a list of all tags with more than 10 occurrences in all trending videos, so we could aggregate data on tags more easily.

Exploratory Data Analysis

As our tool is primarily exploratory in nature, we started by making a Graph Creator tool to explore the data. This tool allowed string inputs to control the graph types, axis values, and filters. This tool was used to make all of the graphs in our project, as it was modeled to be able to fit graphs modularly into passed in divs.

In our initial testing of these functions we found that it was best to only show a small numbers of videos for the bar graphs. We also decided that bar graphs are best left for summations, averages, and other aggregating functions as the sheer number of videos made it impractical to load (and display). We additionally limited the tags to those with 10 or more occurrences, with shrunk the tag data to ⅕ its original size. Ultimately, playing with these initial graphs made us curious of how individual videos compared which drove us to spend a lot of time developing our Video Explorer page.

Design Evolution

We choose to use scatter plots, bar graphs, line graphs, and table views. Each of the graphs is good for showing off a particular aspect of our data. Line graphs were great at showing individual videos' trends over time. Scatter plots were good for trends in comparing video variables (such as likes to views) for subsets of our data. Bar graphs were helpful for comparing the aggregations of data by certain categories or tags. The table format for selecting individual videos enabled us to dive deep and compare individual stats. The table enabled users to easily search for videos, or types of videos that they wished analyze and compare.

Implementation

Although the actual appearance of each of our visualization changed considerably over the course of the project, the current versions are easily identified in comparison to the old.

Home page

The first page of our website is the home page which contains information on our visuals, the dataset, and some examples of interesting graphs about the data, shown below.
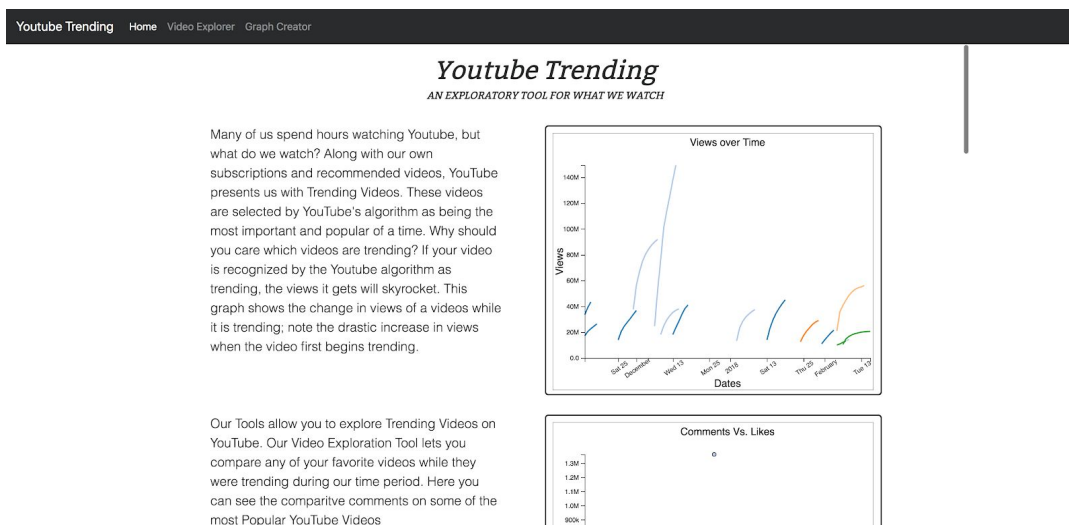

All of the graphs on this page are interactive, through the use tooltips and external links. When user hovers their mouse over a data point a box pops up indicating what that data point is. Clicking on a data point that represents an individual video will open that video in YouTube on another tab. This page was populated last as it contains information and tools from other pages and visualizations.

Video Explorer

The second page of our website enables users to look at specific video statistics. We initially planned this to be a fairly simple table view with checkboxes for selection.

	Video Name	Creator	Date	Category	Likes	Dislikes	Comments
●							
○							
○							
○							
○							
○							
○							
○							
●							
○							
○							

Selected videos were then supposed to appear in a graph like the model below. The axes on this graph were intended to be dropdowns, enabling the user to customize this visualization.
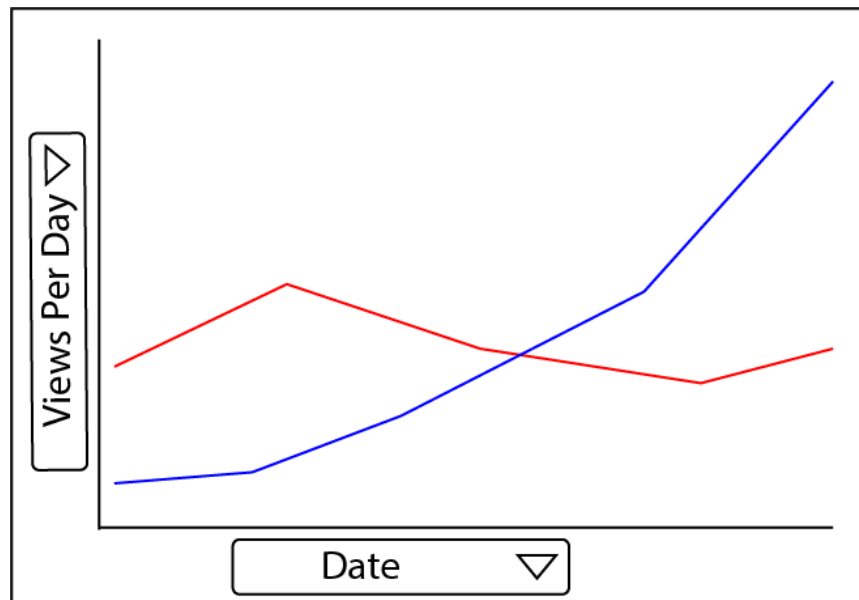
It became quickly became clear that this wouldn't work, as the initial list of videos was many thousands strong and displaying and manipulating the table in this state was unwieldy. As such, we turned to bootstrap tables for their pagination feature.
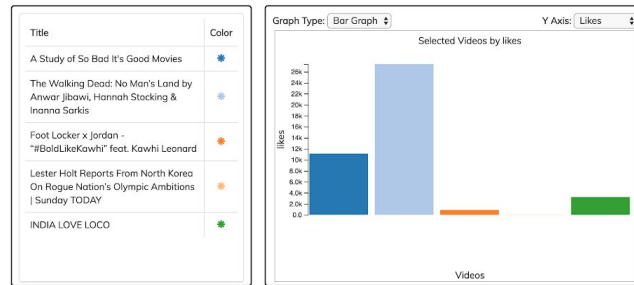
Video Explorer

Title	Color
A Study of So Bad It's Good Movies	✳
The Walking Dead: No Man's Land by Anwar Jibawi, Hannah Stocking & Inanna Sarkis	✳
Foot Locker x Jordan - "#BoldLikeKawhi" feat. Kawhi Leonard	✳
Lester Holt Reports From North Korea On Rogue Nation's Olympic Ambitions	Sunday TODAY	✳
INDIA LOVE LOCO	✳

Graph Type: Bar Graph ▼   Y Axis: Likes ▼

Selected Videos by likes

Table of Videos   Search [   ]   ▦▾

	Title	Channel	Category	Views	Likes	Selected
✚	2017 Champions Showdown: Day 3	Saint Louis Chess Club	Education	71089	460	👁
✚	Machine Gun Kelly, X Ambassadors & Bebe Rexha - Home (from Bright: The Album) [Music Video]	Atlantic Records	Music	3202879	110088	👁
✚	Meghan Markle Engaged To Prince Harry | The View	The View	Entertainment	502681	2699	👁
✚	Americans could see tax bill impact in early 2018	ABC News	News & Politics	285483	1754	👁
✚	The Making of Kendrick Lamar's LOVE. With Teddy Walton | Deconstructed	Genius	Music	390695	18203	👁
✚	venting online about a hard part in a video game	ProZD	Film & Animation	643737	31028	🚫👁
✚	Roy Moore & Jeff Sessions Cold Open - SNL	Saturday Night Live	Entertainment	2688797	19042	🚫👁
✚	5 Ice Cream Gadgets put to the Test	CrazyRussianHacker	Science & Technology	1251577	28951	🚫👁
✚	The Greatest Showman | Official Trailer 2 [HD] | 20th Century FOX	20th Century Fox	Film & Animation	2671756	12699	🚫👁
✚	Why the rise of the robots won't mean the end of work	Vox	News & Politics	635985	20721	🚫👁

Showing 1 to 10 of 4079 rows   10 ▲   rows per page   ‹ 1 2 3 4 5 … 408 ›

User can use the bottom table to sort, search, and select videos. Selected videos are inserted into the top-left table and assigned a color. One a video is added to the top left table, a graph is generated in the top-right according the the above dropdowns. To remove videos from the list, users simply have to click the same button used to select the videos. Rather than making the dropdowns on the axis labels themselves, we placed them above the graph. We felt that this placement would be more logical, and wouldn't take away from the meaning of the graph. The plus button on the left expands the row and shows an embedded copy of the YouTube video that row represents. Some videos disable this function. These videos instead have a link to the YouTube page itself.

Graph Creator

Our final page hosts the Graph Creator tool. The tool enables users to create their own graphs and explore the data however they want to. Below are our original sketches for this page and the final product.
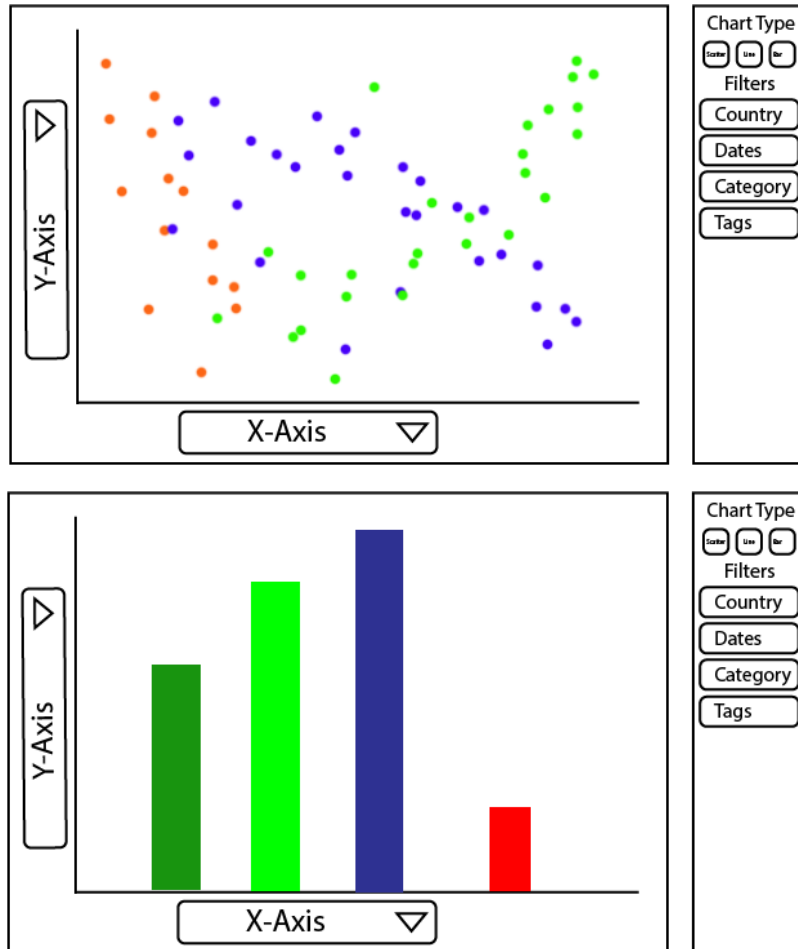

We also drafted a tool for creating the different types of graphs from the same interface, as pictured below.
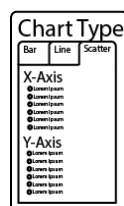

We abandoned the idea of axis labels on the graphs in favor of putting them in a sidebar, we moved the filters to the bottom and added titles to the graph. Each graph has tooltips and the line and scatter plots support clicking on them to bring you to the page on YouTube for the video. The bar graph does not have this function at it represents aggregations of videos rather than a single videos. Using the filter text boxes, (tags and categories) users can filter the data to

certain tags and categories by separating each by a space. The following is the final design of this page. One can select any of the graph types, and the div will be replaced with a customizable graph of that type.
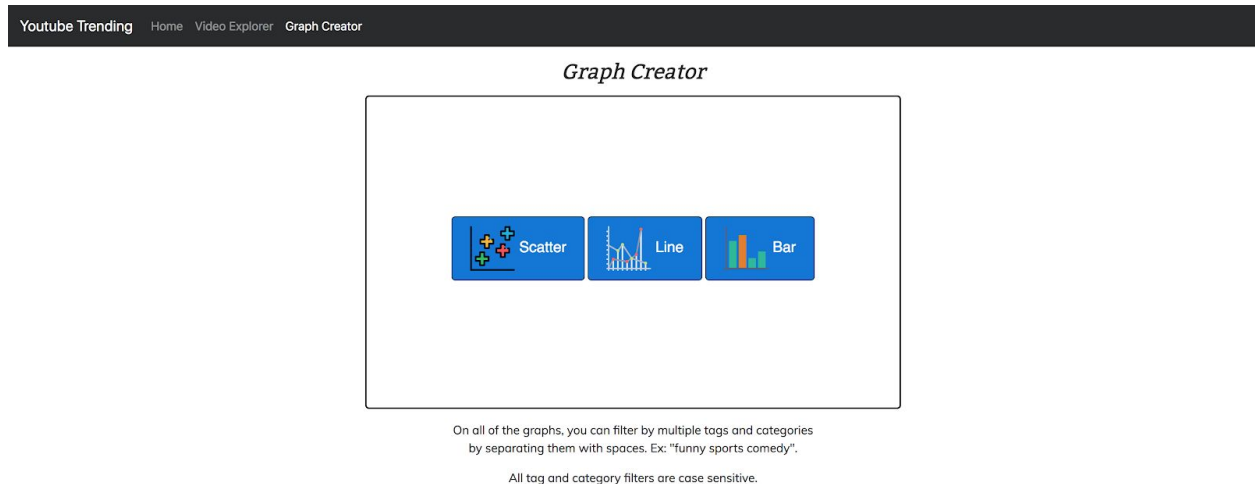
The graph type selection box then appears below the new graph, as show below.
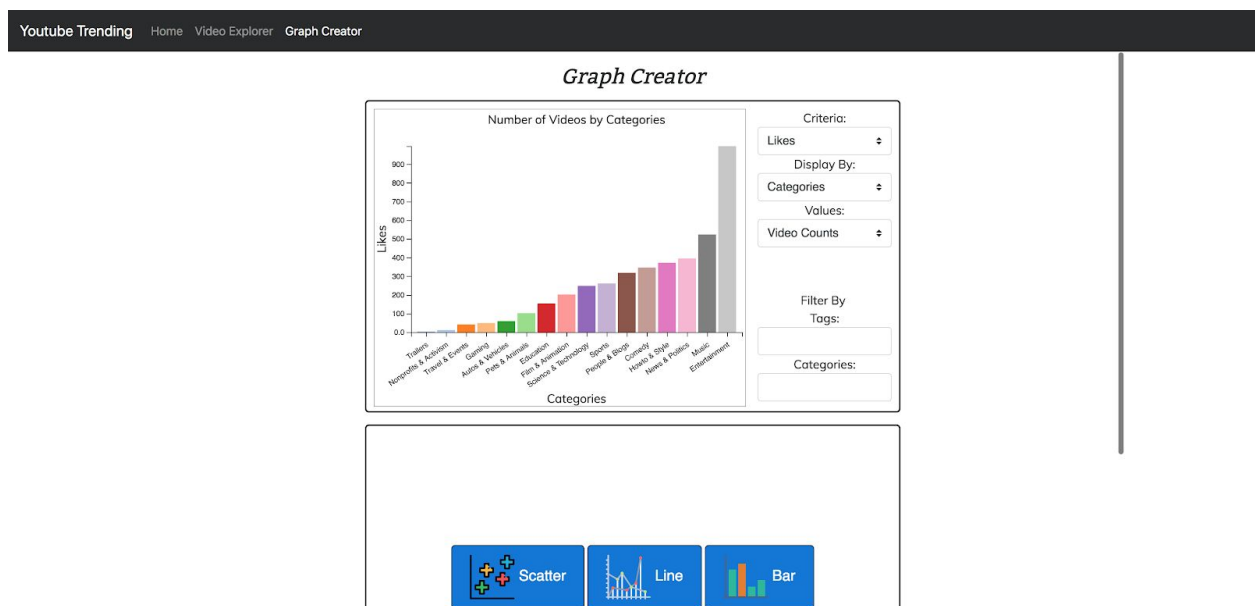
Evaluation

      We learned that videos across different countries share the same like, views, and comment counts, which makes an intuitive kind of sense. And while the respective positions on trending for each video may differ for different areas, that ranking was not included in our data set, so we removed the country filtering option from our tools as they wouldn't show anything new. We found that views correlate most with trending videos and while some trending videos have low ratings, they still trend due to sheer view count. We also found that dislikes and disabling comments/rating has no real impact on a video's trending status. Additionally, we found that Music is the most popular category in terms of Likes and Views following by Comedy and Entertainment. In a similar finding, we fond "Pop" (as a musical genre) was the most popular video tag in terms of Likes and Views followed again by "funny","comedy," and similar tags. However, this aggregated result combines the like and view counts for all videos that included those tags, so this makes sense given the high number of comedy and music related videos that appear in our trending data. It was also interesting to see that the 2017 YouTube rewind video gained massive popularity, which we felt was undeserved and only happened because of how hard YouTube advertised it. This made us think that YouTube, at times, likely artificially influence the selection of trending videos to suit themselves, however, we do not have concrete evidence to support this. Finally, we can see the Logan Paul controversy was hitting its peak around the time of this data collection as 5 of the top 12 most liked tags included his name, which is a display of the popularity and influence he had amassed to that point. Interestingly though, he does not appear in the top 20 tags sorted by views, indicating his viewership is very dedicated but not large enough to put him on par with influencers like Ed Sheeran, Kendrick Lamar, Casey Neistat, etc.

      Our visualization works well and overall achieves the intended purpose of exploring the structure of what makes it to trending these days. Some possible improvements include the following: auto complete for categories and tags, the ability to filter by dates, incorporating the YouTube API to allow users to pull current data and compare their own favorite videos, and to improve the load times of graphs (specifically the bar graph when sorting by tags). A further option could be to allow users to filter by videos with disabled comments and/or dislikes as the current functions allows for this, but they are not included in the filtering interface. Additionally, since the scatter plot can be so cramped, it could be good to include a feature where highlighting a region enables a zoomed in view of that area to help improve readability of the data. It may also be useful to include a "remove graph" option for the Create Graphs page and/or to enable persisting graphs so that if you go to the Home or Explore Videos pages, the Create Graphs pages doesn't reset.