Youtube Trending: What do we watch?

Saul Woolf (saulvhwoolf)

Jon Venne (jonvenne)

Drew Ciccarelli (dcicc4)

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

**Background and Motivation**

Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

We all watch YouTube, and are interesting in understanding YouTube’s trending algorithm. Trending videos are the top 200 hundred videos chosen by the YouTube algorithm as the “most popular” for a given time period.

**Project Objectives**

Our visualizations will be primarily exploratory and so will focus on allowing users to explore data on various trending videos over two months.

We would like to look into:

How trending videos differ from country to country?

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

How do down votes affect trending videos?

What categories of videos are most often trending?

What tags are found most often in trending videos?

What effects do disabling likes and comments have on videos?

The benefits of our project include:

1. Helping people understand the YouTube Algorithm

2. Helping YouTubers to create content more likely to be shown as trending

3. Showing the effects of disabling likes and comments on videos

4. Letting people compare their own favorite YouTube videos to trending ones.

**Data**

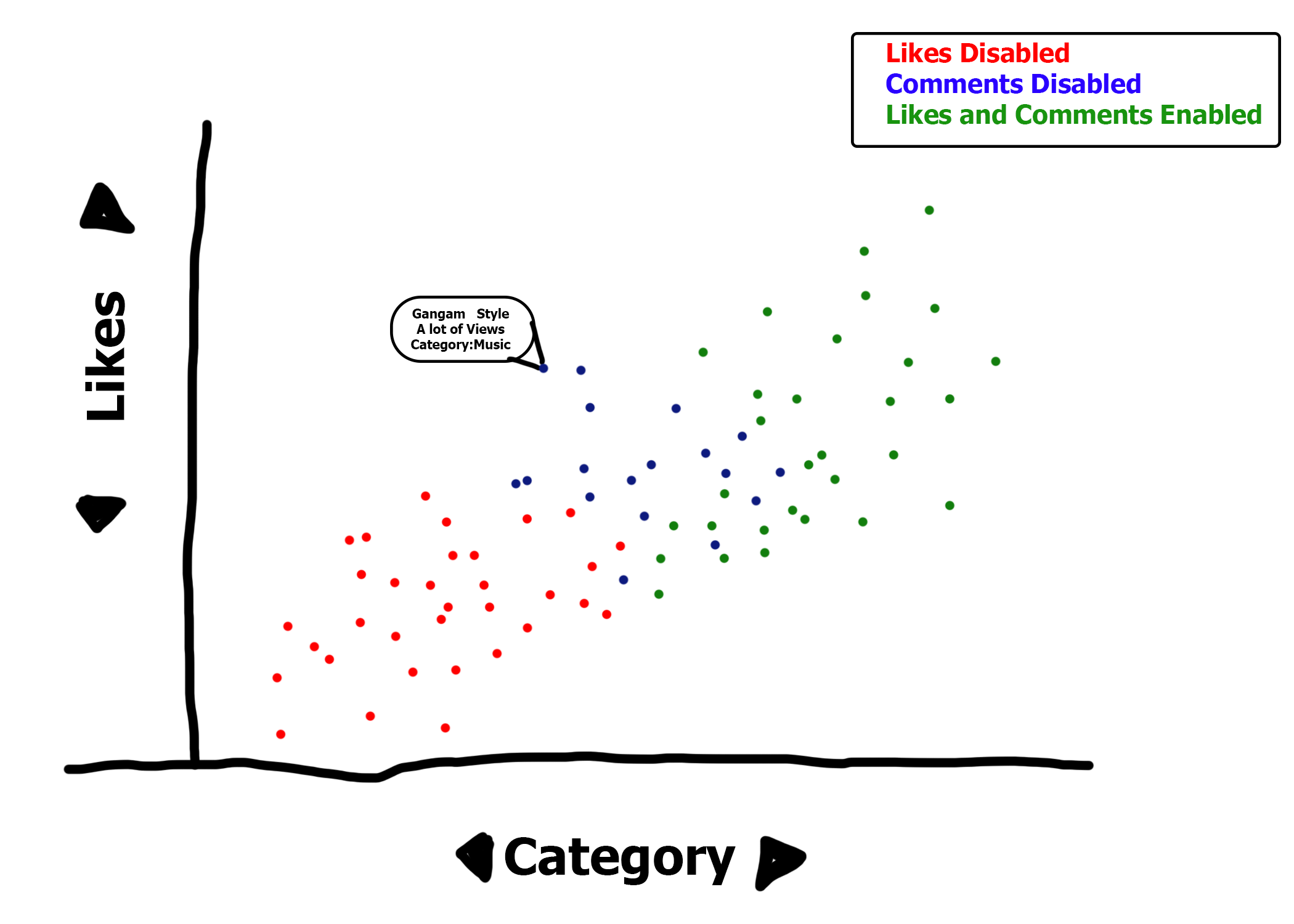
We primarily collect our data from <https://www.kaggle.com/datasnaek/youtube-new/data> . This database contains information on all trending videos in five different countries from between December 2017 and January 2018. It contains information on likes, comments, dislikes, and whether these features were disabled. We also hope to allow users to import data from any YouTube video using a Google API.

**Data Processing**

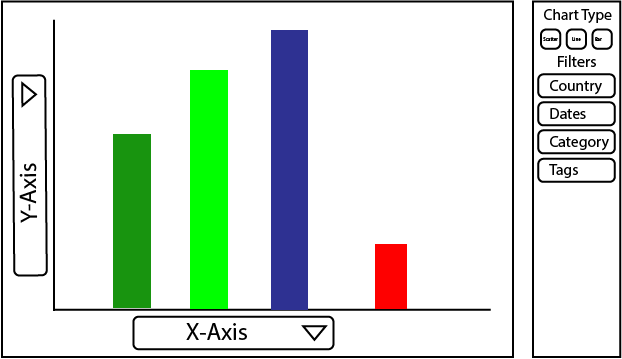
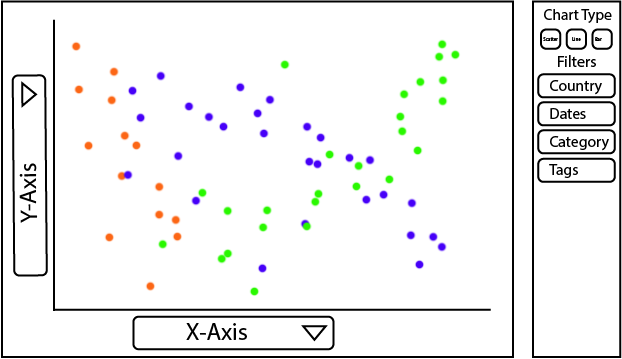
We will perform some cleanup of our data. We hope to convert the CSV file which currently displays a video once for each day it was active, to a JSON file of videos where one attribute of the video will be a list of days where it was active and each day will have the likes, comments, and view counts for that video on that day. This will allow 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 currently uses an number id system for each category, we will replace each id with the appropriate category String. We will use a python script to convert the file and calculate the missing value. Additionally, we will be supplementing the current data set by including extremely popular viral videos (such as Gangnam Style, Despacito, etc) from the past. This will help ground what the user is seeing by giving them a “known” point of reference by which they can compare the rest of the videos to.

**Visualization Design**

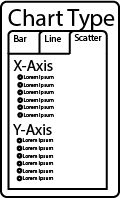
How will you display your data? Provide some general ideas that you have for the visualization design. Develop three alternative prototype designs for your visualization. Create one final design that incorporates the best of your three designs. Describe your designs and justify your choices of visual encodings. We recommend you use the [Five Design Sheet Methodology](http://fds.design/).

Drawn Graph:

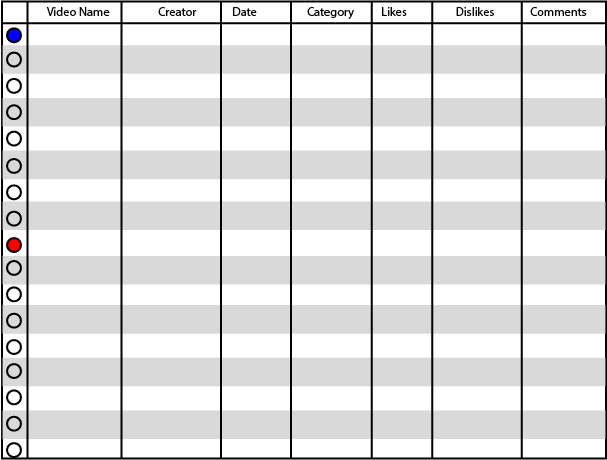
Simple Customizable Graphs:

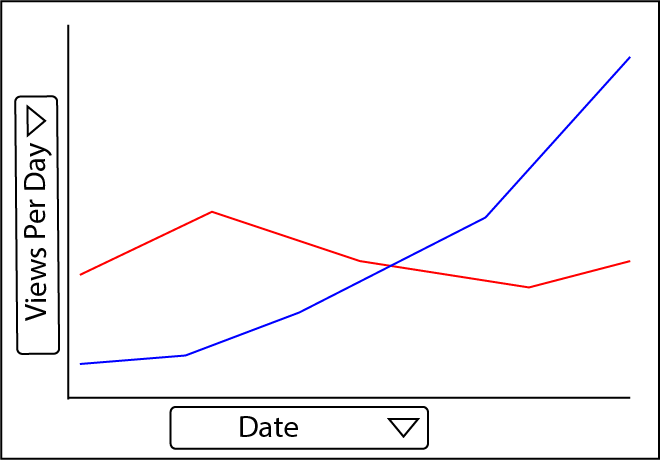


Graph Type Editor:



Song Selection Graphs:





* Scatterplot of videos on a particular day. This would act as one our headliner visualizations. (The big one that most people use most frequently for learning about the data set)
  + By likes, views.
  + With tooltip and link to video
* There would additionally be a set of sub visualizations that are more customizable by the user that could be a bar chart, scatter plot, line graph, etc.
* Exploratory Visualization
  + Give the users the interactive ability to view whatever data they’re interested in and compare various views to discover new info about how trending works
* Supplementary Visualizations
  + Top 3 trending On certain date
  + Top Trending global vs by country
  + On a specific day, Explore the trending page
* Comparison of videos to top trending of all time (Gangnam Style)(Despacito)
  + Scatter for multiple videos or bar for single
* Simple Graphs (likely bar charts)
  + Time on trending
  + Number of trending videos per channel
  + Top Likes/Dislikes/like-dislike-ratio/comments
  + Country color fill comparison to show country viewership data

**Must-Have Features**

The primary feature of our site will be a set of exploratory visualizations. They will consist of the following sub features

* Users can filter the data displayed by:
  + Country
  + Category
  + Trending day(s)
  + Video Tags
* Users can put the following on the X-axis
  + Days Trending
  + Views
  + Likes/Dislikes
  + Comment Count
  + Country
  + Category
* Users can put the following on the Y-axis
  + Days Trending
  + Views
  + Likes/Dislikes
  + Comment Count
* Users can display data as
  + Scatter Plot (discrete Y-axis)
  + Bar Graph (summation Y-axis)
  + Line Graph (summation Y-axis, when X-axis is time only)

Another feature will allow users to explore some premade visuals on data we found interesting, such as channels with many trending videos, high dislike counts, and non-consecutive trending dates.

Finally, users will be able to choose individual videos and make visualizations comparing them. The user will see visualizations on these videos displayed as a line graph. These will show likes, views, or comment counts over time while they were trending.

All of our visualizations displaying discrete values will support tools tips allowing users to hover over a dot to see which video it is, and give them the option to view the video on YouTube.

**Optional Features**

* allow users to add their own videos from YouTube and compare them to others using our visualization tool. We may also add more filter options for the user such as thresholds for likes and comment counts.
* Enable Animated Visualization Transitions for the dynamics graphics
* Collapsable Filtering Settings next to each user customizable graph
* Enable viewers to see maps of each country where the percentage of viewers is represented by shade of color (color possible customizable)
* Zoom Out Feature that displays multiple visualizations in a new pop up view side by side and allows the user to click and jump to any one of those visualizations on the actual page
* A complete scatterplot showing all data points linked to a table that displays the raw data and is searchable/sortable by Name, Genre, Country, etc then highlights the currently selected video on the overall plot to help give the user a sense of relative video performance
* Make the page translatable to multiple languages
* Tooltips displaying the video Thumbnail in addition to the other information
* A page that gives users the option to try and predict which video had a higher certain characteristic (Views, Likes, Comments, etc) Possibly in a “compete with your friends” kinds of format. This not only helps engage the user, but helps test whether they can recognize the trends in what gets you to trending.

**Project Schedule**

|  |  |
| --- | --- |
| Milestone to Have Completed | Date |
| Clean Data Set | Feb 16 |
| Establish Page Layout/Views/What Content goes on which pages | Feb 17 |
| Have a functioning main scatter plot | Feb 18 |
| Have Functioning Customizable Plots | Feb 20 |
| Have full prototype of the final product usable | Feb 22 |
| Receive input from friends on usability and what can be improved | Feb 23 |
| Implement Changes Outlined by Test Users | Feb 24 |
| Attempt to have around half of the optional features implemented | Feb 25 |
| Scramble to make everything that doesn’t work work | Feb 27 |
| Functioning Website | Feb 28 |