

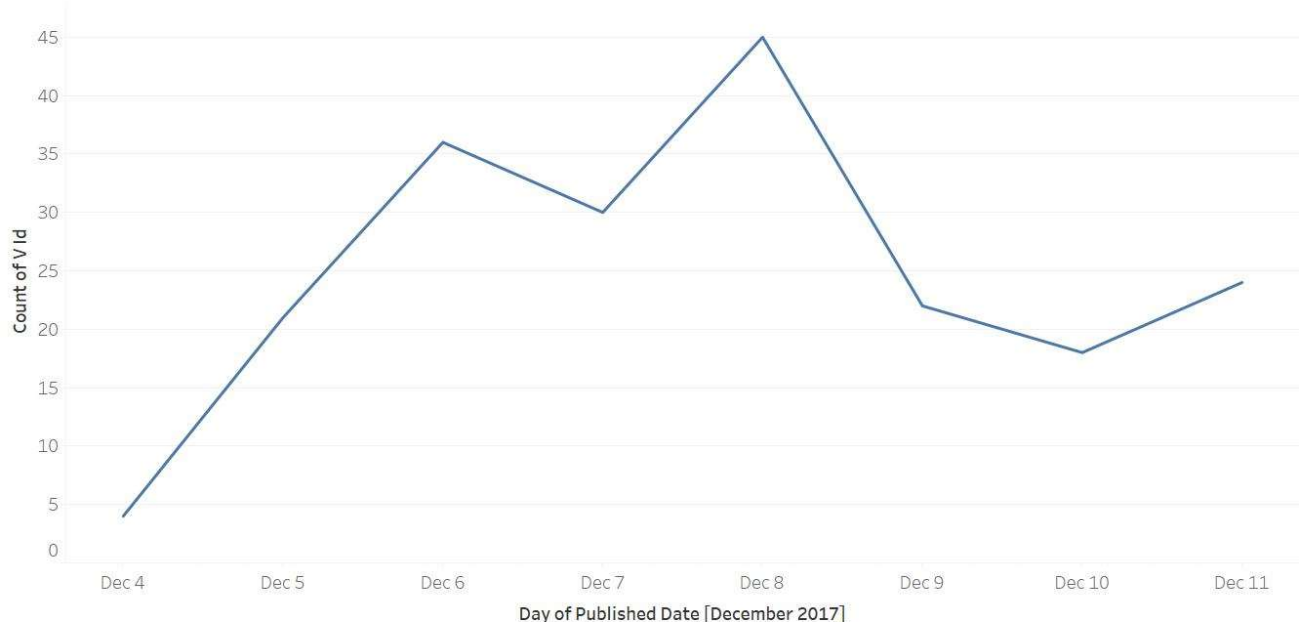
Notebook 3

In this project we utilized the YouTube Application Programming Interface, or API. Many web services provide API's to the public to use for developing and programming applications. We chose to use YouTube. YouTube is one of the biggest and most popular website for video sharing and is one of Google's web services. Odds are if you are a person who has been on the internet, you have been to or at least heard of YouTube.

Trending videos are videos that have a high popularity in a very short amount of time. YouTube selects videos that are the most popular at a given time and defines these as trending videos. We were curious about the relationships among data from these trending videos. Using our application that uses YouTube's API, we gathered data on the top two hundred most popular videos. The data in this project represents the top 200 videos from 6:30 pm on Monday December 11, 2017.

From the data we gathered we created four different tables. The first table contained, for each top 200 videos, the video ID, video title, video description, channel ID, channel title, and the date and time the video was published. From this table we were wanted see how long the videos had been posted for. Trending videos have usually not been posted for very long, however it takes time for a video to gain popularity. Using the video ID's and the date they were posted we were able to create the graph below. This graph shows the timeline of the posting of these popular videos. All of them were posted within the past week. The highest number of videos posted on one day is 45 videos, this was on December 8th; three days ago. The graph follows a rough bell curve, meaning that videos posted more than 4 days ago have lost some popularity and videos posted within the past 2 days have not reached a high popularity yet. This follows what we expected as we thought that it would take a few days for a video to get very popular but that the popularity does not last for very long.

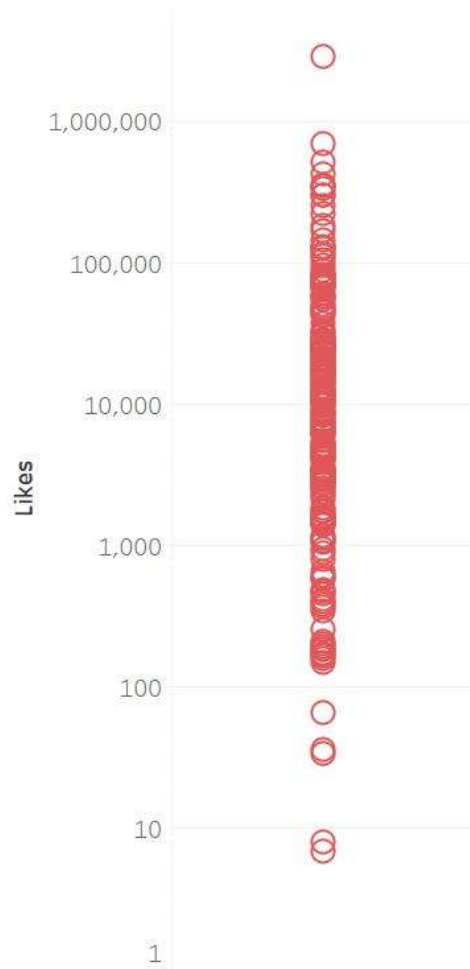
Published Videos by Date



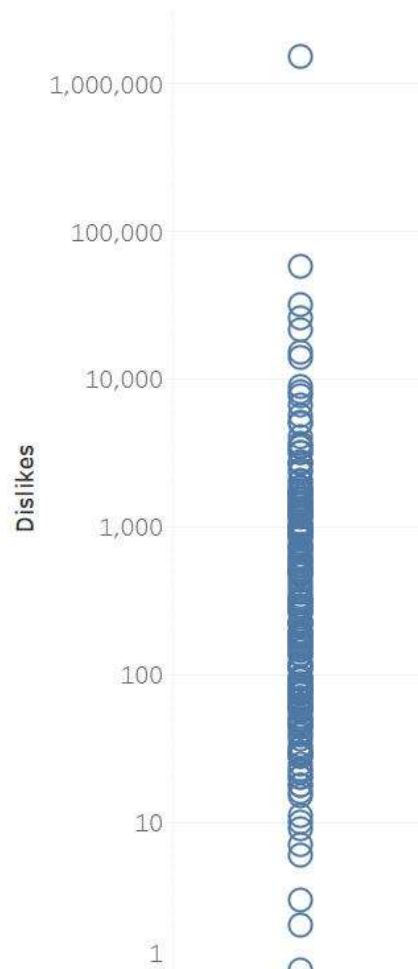
The trend of count of V Id for Published Date Day.

After viewing the trend among the publishing date of videos, we decided to examine the relationship between likes, dislikes, and views of a video. We suspected that the number of likes and dislikes would be drastically lower than the number of views, since not many people actually rate a video. From our second data table, containing the statistics of the videos, we created the graphs below. These graphs show the number of likes, dislikes, and views for each video. They are displayed on a logarithmic scale because the extremely large range of data points made viewing the data on a regular scale practically impossible to read. The data points on the graph that displays the number of likes ranges from 7 to 2,885,874, the number of dislikes ranges from 1 to 1,521,082, and the number of views ranges from 1,150 to 122,194,703. There was not much to see or learn from this relationship.

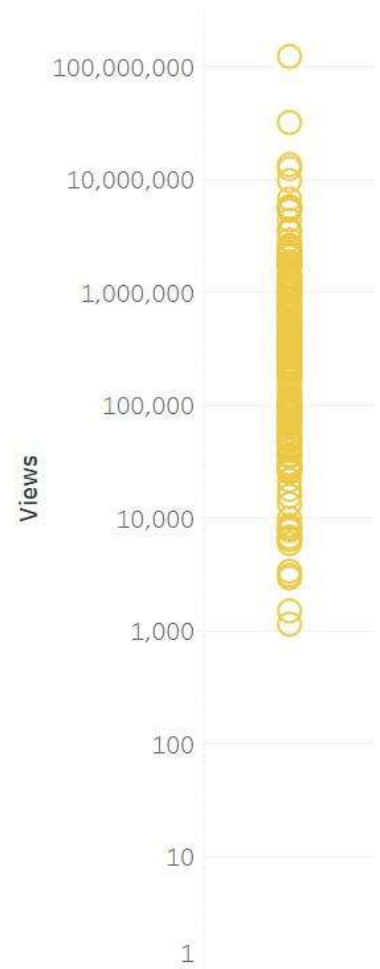
Likes



Dislikes

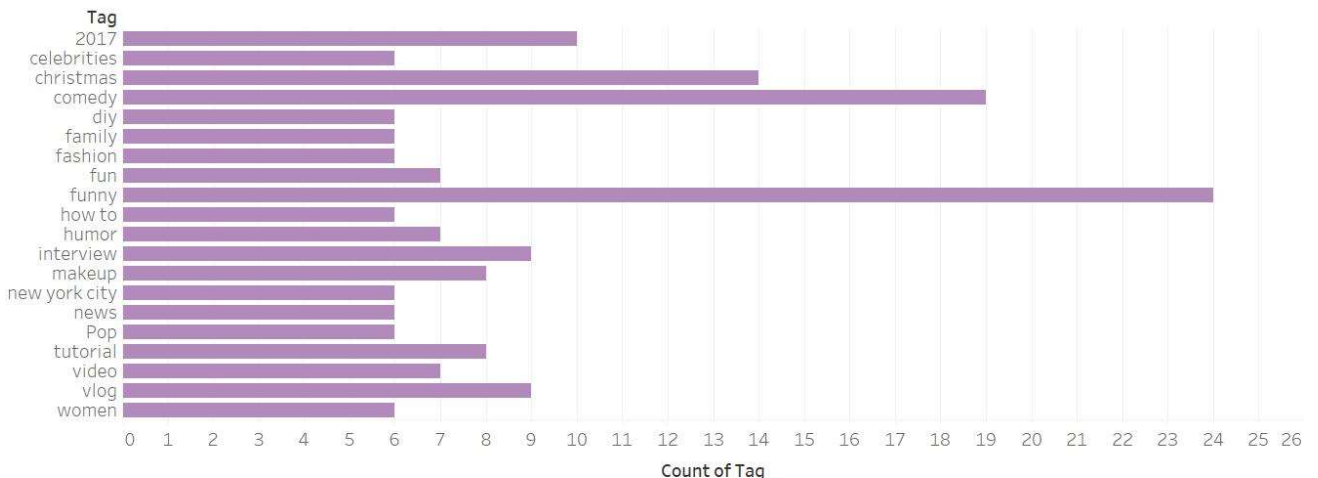


Views



While exploring the YouTube API, we noticed that videos often have many tags associated with them. This got us wondering if there were common themes or tags across popular videos. We wanted to know if certain tags were used more than others or if popular videos are tagged similarly. To examine this relationship we created a table that contained every tag associated with every video ID. Even though there were only two hundred videos, the tags table had 4,123 rows. This means that on average, each video had about 20 tags associated with it. Below is a graph that shows some of the most popular tags. We filtered these results based on what tags were used 6 or more times. These resulting tags are pretty expected. The tag "funny" is by far the most popular tag, followed by "comedy", which makes perfect sense as YouTube is often used to share funny videos and these funny videos are often very popular. "2017" and "christmas" were also very popular tags, which given the time of year makes sense that popular videos would be tagged with these attributes. Some of the other popular tags represent common YouTube video themes, like tutorial/how-to videos, news/interviews, and vlogs.

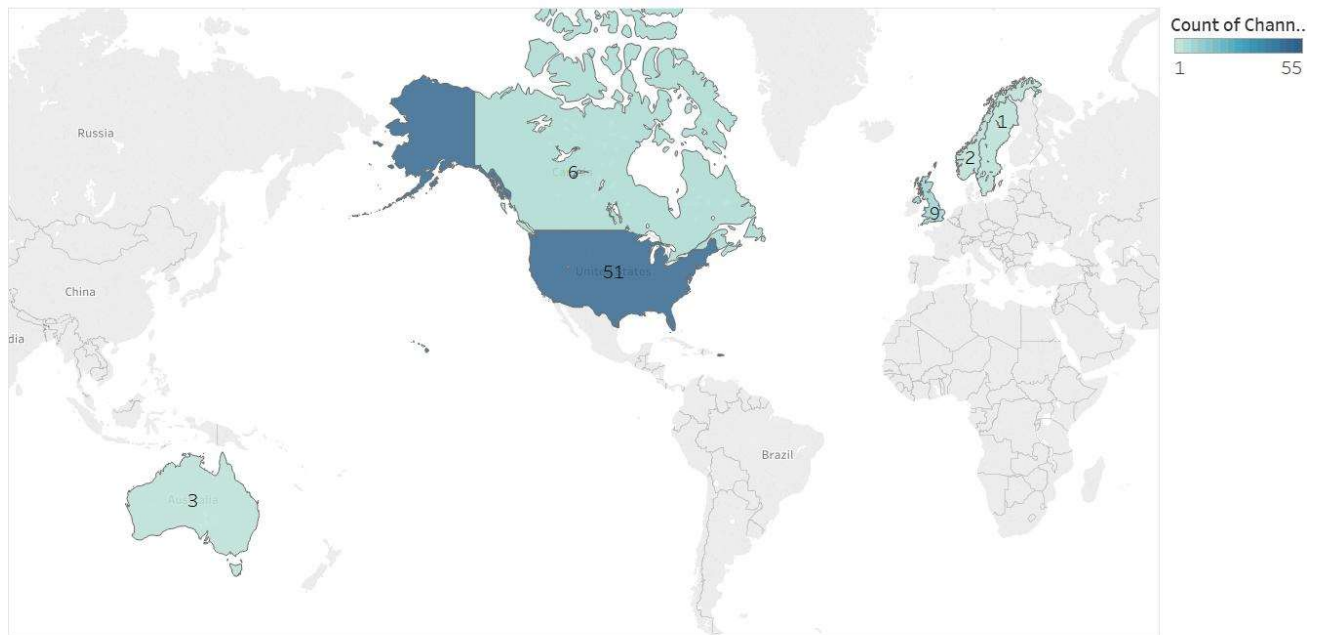
Popular Tags



Count of Tag for each Tag. The view is filtered on count of Tag, which includes values greater than or equal to 6.

Additionally, during our exploration of the API, we saw that we could gather the country of a specific channel. This made us curious about the country of origin for the trending videos. Unfortunately, YouTube's v3 API does not provide the country of a video, but it does provide the country of the channel a video belongs to. We used the channel ID of each video to create a list of all of the channels that have posted one of our trending videos. Then we found the country code for each of these channels. We generated the map below, which shows how many of the channels in our dataset originate from each country. Unfortunately, not all of the channels provide the country code, so this graphic is only representing a little less than half of the channels. From the channels that did provide their country code, however, we can see that the US has by far the most channels, at 51. Great Britain has the next highest number of channels, at only 9. While this data is not extremely descriptive, the fact that the US has the most channels is unsurprising.

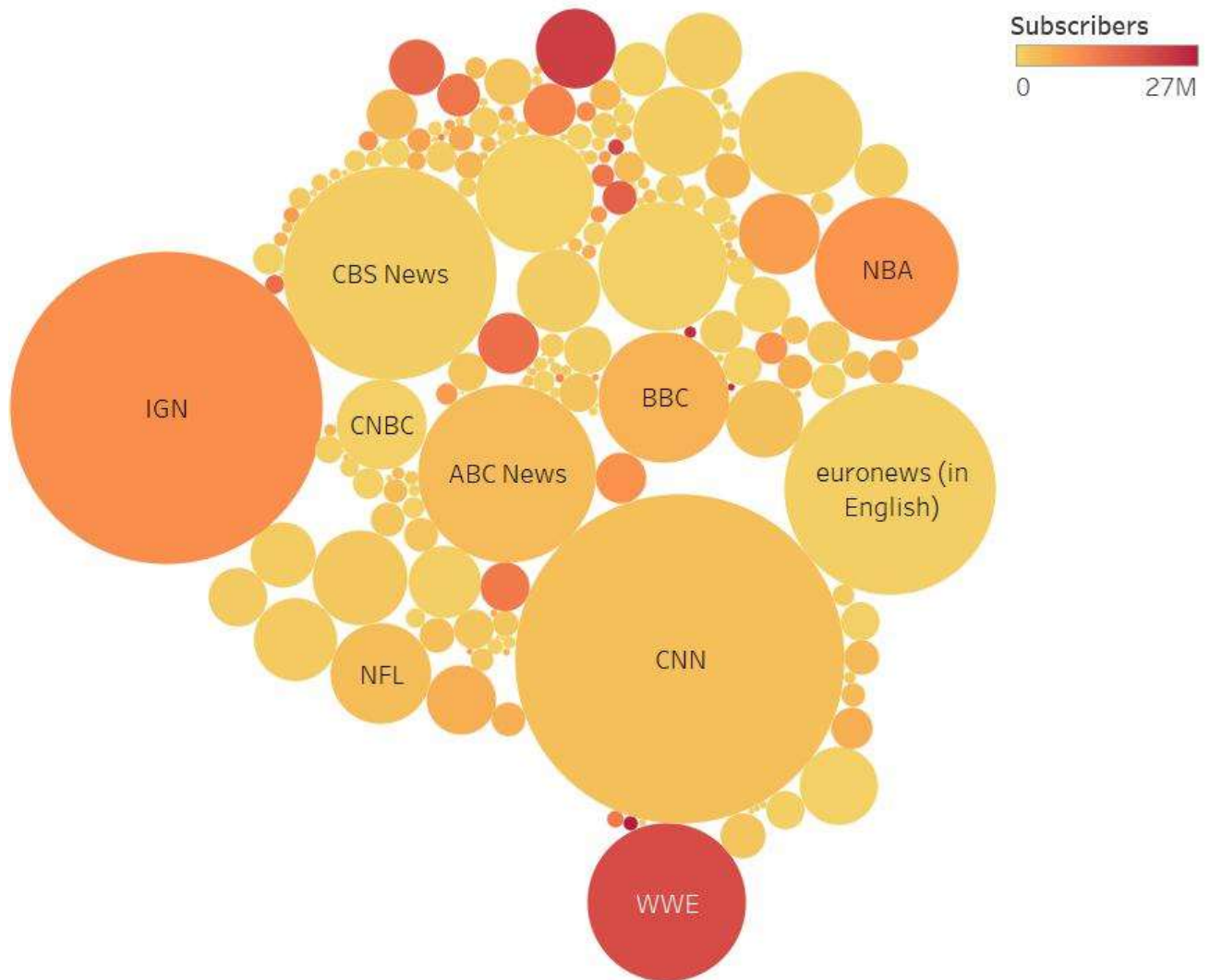
channels by country



Map based on Longitude (generated) and Latitude (generated). Color shows count of Channel Id. Details are shown for Country.

Finally, we examined the relationship between the number of subscribers a channel has and the number of videos that channel has posted. We expected that channels that have a lot of subscribers would also have a lot of videos they have posted. Using the tables we created with information on each channel that posted one of our trending videos, we created the data visualization below. In this image, the size of the circle is representative of the number of videos that channel has posted; the larger the circle, the more videos. The coloring of the circles is representative of the number of subscribers that channel has; the darker/more red the color, the more subscribers. In general, the channels with the most videos are news related channels, so it makes sense that they would have a large number of videos as news outlets produce and post videos daily, as well as a decently large number of subscribers. There are several channels that have relatively few posted videos but a very high number of subscribers; like EminemVEVO, which has only 69 videos but 24,657,063 subscribers. In some ways this followed what we expected but also included several outlier channels that, while not having as many posted videos, probably post really popular material, giving them more subscribers.

Channel Subscribers and Number of Videos



Channel Title. Color shows sum of Subscribers. Size shows sum of Num Videos. The marks are labeled by Channel Title.

In conclusion, we found many interesting data relationships surrounding trending videos, however, we also failed to find some in places where we thought they might be. In general, we learned a lot about the sample of trending videos, like how long a video stays popular, or what kind of videos are the ones that become popular. Through this project we also learned how to and gained first hand experience using a web service API and applying class material to real world data. With the combination of these two practices, we were able to achieve a clearer understanding of what composes trending videos each day and the information that can be learned from them.