Youtube Rewind

Data visualisation final report

Date: 08-06-2020

Carlos de Bourbon	s2153874
Elsi Muller	s1917668
Eva Lahuis	s1851985
Aine van den Aakster	s2113570
Iven van Horck	s2007665

In this report we explain briefly where our data comes from.

User characterisation

Visualisation 1: Correlation between comments and likes on Youtube per category
Made by: Elsi Muller
The data for this visualisation was gathered from:
https://www.kaggle.com/datasnaek/youtube-new which shows the daily trending videos from 2006 to 2018. With every video it includes, the likes/dislikes, amount of comments, used tags, the publish/trending date, the country and to which category it belongs. The amount of data was very irregular in this data set. From the year 2006 to late 2017, there are nearly no data points. For this reason the time period of november 2017 to june 2018 was used, because of the higher reliability. Some sub-categories were removed due to the few number of data points.

Visualisation 2: Amount of youtube users per country and Percentage of the population of youtube users per country
Made by: Elsi Muller
The data for this visualisation was gathered from:
https://www.worldatlas.com/articles/which-countries-watch-the-most-youtube.html and https://databank.worldbank.org/data/download/POP.pdf
On the first link, a list of the amount of youtube users there are per country in 2018 was found. This has been extracted and used to create a radial bar graph. The second link shows the population per country in 2018. This has been used to create a percentage of the population that views youtube, by dividing the amount of users by the size of the population.

Visualisation 3: Top 3% versus bottom 97%
Made by: Carlos de Bourbon
The data for this visualisation was gathered from:
Bärtl, M. (2018). YouTube channels, uploads and views: A statistical analysis of the past 10 years. *Convergence: The International Journal of Research into New Media Technologies*, *24*(1), 16–32. https://doi.org/10.1177/1354856517736979

The top 3% of channels (by views) and bottom 97% are compared in views and uploads.

Comparisons to other platforms

Visualisation 1: Timeline of of existing and defunct video sharing platforms
Made by: Iven van Horck
The data for this visualisation was gathered from:
https://en.wikipedia.org/wiki/List_of_video_hosting_services where a table visualised the popularity of the website in alexa ranking, the name of the website, and the languages used on that website.   For defunct websites it also showed startyear and endyear. To find the start and end date for each of the websites the wikipedia page for that website was used, if no starting month was mentioned the month of january was used for that website..

Visualisation 2: Comparison of popularity of the website, date of creation, and language used
Made by: Iven van Horck
The data for this visualisation was gathered from:
Same source as visualisation 1.

Visualisation 3: Top 10 music artists on Youtube versus Spotify
Made by: Aine van den Aakster
The data for this visualisation was gathered from:
https://socialblade.com/youtube/top/100/mostsubscribed
https://en.wikipedia.org/wiki/List_of_most-streamed_artists_on_Spotify
The top 10 most subscribed music artists on YouTube were taken from a database of the top 100 most subscribed channels. From this, every channel in the category 'music' and 'entertainment' were checked (googled) to see if they were a music artist and if so, added to the list. The top 10 most followed music artists on Spotify were taken from a very up to date list on Wikipedia. Year of international breakthrough was searched for online (Wikipedia page of every artist). This data was directly injected into the visualisation (in Adobe Illustrator), as it was not made with any graphing/data processing software.

Popular content

Visualisation 1: Number of views per category
Made by: Elsi Muller
The data for this visualisation was gathered from:
https://www.kaggle.com/datasnaek/youtube-new (the same as visualisation 1 in User characterisation).

Visualisation 2: Word clouds of popular tags by category
Made by: Carlos de Bourbon
Source : https://www.kaggle.com/datasnaek/youtube-new

From the kaggle database, a database containing the top trending movies in Russia, Mexico, South Korea, Japan and India respectively from 2006 - 2018. A visualization of the most occurring tags are shown, in the form of a word cloud, in which the size of the word gives in indication of its occurrence. The visualization is limited to the top 5 most viewed categories.

Visualisation 3: Sponsors behind most subscribed channels
Made by: Carlos de Bourbon
Source : www.youtube.com,
https://www.businessinsider.nl/most-popular-youtubers-with-most-subscribers-2018-2-2/

Sponsorship data was manually gathered from YouTube and added to data of the top 23. Sponsor companies were categorized, to allow comparison to the channel content.

Rules and Regulations

<u>Visualisation 1</u>: Youtube disruptions by country vs. Christianity - Pct. Adherents
Made by: Eva Lahuis
Sources: In this source the disruptions measured by Youtube can be found. This is what Youtube is saying about the disruptions: "We add events to the list after we detect or discover significant drops in our traffic graphs and obtain related information from government sources, news outlets or ISPs." .https://transparencyreport.google.com/traffic/overview?hl=en_GB In this source data about information about religious adherence worldwide since 1945. https://correlatesofwar.org/data-sets/world-religion-data For this vizualisation tha data out of 2010 is used, this is the most recent data, so this is the best data to relate to Youtube.

<u>Visualisation 2:</u> Youtube product traffic per country against Youtube COVID-19 measures during November 2019 and June 2020
Mabe by: Eva Lahuis
Source: In this source the product traffic per country to and from Youtube around the world is found:https://transparencyreport.google.com/traffic/overview?hl=en_GB In this next source the measures taken by Youtube during COVID-19 can be found: https://support.google.com/youtube/answer/9777243?p=covid19_updates&visit_id=637264243678297513-1469679444&rd=1 This visualization displays the changes in Youtube product traffic per country and the timing of the measures taken by Youtube during the COVID-19 pandemic.

<u>Visualisation 3:</u> Number of flags assigned by users against removed videos per category
Made by: Eva Lahuis
Sources: YouTube's Terms of Service prohibit the posting of videos which violate copyrights or depict pornography, illegal acts, gratuitous violence, or hate speech (this can be found here: https://www.youtube.com/intl/nl/about/policies/#community-guidelines). User-posted videos that violate such terms may be removed and replaced with a message stating: "This video is no longer available because its content violated YouTube's Terms of Service". Users can assign flags when they think a video violates such terms. In this source the data about these flags and removals can be found: https://transparencyreport.google.com/youtube-policy/removals?hl=en From this visualization you can conclude that the users and Youtube do not always think the same about violating the terms, does this mean the users want Youtube to be more strict?

History

<u>Visualisation 1:</u> Growth and achievements
Made by: Aine van den Aakster
Source:
https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/
https://www.officetimeline.com/blog/youtube-history-timeline

This graph shows the amount of hours uploaded to YouTube every minute. The data was placed on a timeline using Tableau (so that the spacing between the months was proportional) and big milestones were highlighted, from the timeline source.

Visualisation 2: Timeline of interesting channels
Made by: Aine van den Aakster
Source: www.youtube.com
Manually gathered dates of channels in three categories. Data was directly injected into visualisation (in Adobe Illustrator).

Visualisation 3: History of popularity of topics on youtube
Made by: Iven van Horck
Source:
Bärtl, M. (2018). YouTube channels, uploads and views: A statistical analysis of the past 10 years. *Convergence: The International Journal of Research into New Media Technologies*, *24*(1), 16–32. https://doi.org/10.1177/1354856517736979

Table 6. They visualised this with an increasing area graph, in our visualization the increasing of the graph is removed to an it is instead a bar graph, this is because the interesting part of the data is the differences in interest in various subjects.