Process Book

1. Team Info

- **Team Name:** Team Like and Subscribe
- **Team Members:**
    - Ewuresi Amoo-Adare (eamooadare3@gatceh.edu),
    - Carlos Robinson(crobinson304@gatech.edu),
    - Joni Isbell (joniisbell@gatech.edu),
    - Sean Mulloy (smulloy3@gatech.edu)
- **Team Leader:** Joni Isbell

2. Team Agreement

- **Meeting Times:** Every Friday @ 3:30 pm - 4:30 pm EST
- **Communication Method**: Microsoft Teams and Text Messaging
- **Collaboration:** Shared github repo for source control self assign weekly tasks on Trello, progress updates on Teams

3. Project Info

- **Title**

  Meta definition and redefinition: Changes in the construction and popularity of youtube videos over time

- **Abstract:**

  We will analyze Youtube video popularity over time. Due to both hard factors, such as platform policy changes, copyright enforcement, and implementation of features like youtube shorts, as well soft factors, such as ‘algorithm folklore,’

trends, and wider changes in the attention economy, the performance and viability of videos changes drastically year over year. How do global events affect video topics? Do US presentational elections create a spike in political content? Do we see this same spike globally? Social media greatly impacts our world. Content produced on sites like Youtube can tell the story of what is important to us.

Such is common in similar platforms, which in turn visibly alter the trends of Youtube and its content. So how exactly does the rise and popularity of the competitor social media app Tiktok affect the content style and production of content on Youtube. Will Youtube shorts become more popular than actual long form content videos? Will the style of Monetization change on Youtube as well?

- **Potential Data Sources**
    - YouTube Videos Data for ML and Trend Analysis https://www.kaggle.com/datasets/cyberevil545/youtube-videos-data-for-ml-and-trend-analysis,
    - Direct Youtube API access
    - Other SNS platforms API

Map

Data Collection & Project Plan

1. Who is the audience?
    a. Youtubers interested in SEO and growth
    b. Advertisers interested in what their ads tend to run against

    c. Viewers? Parents?

2. Audience in detail
    a. Youtubers, advertisers, and views will all be familiar with Youtube. As these are active users, we won't have to explain some basic features of the platform, such as likes, trending views, and Youtube Shorts. It is also likely that our audience is familiar with other social media platforms and internet websites such Instagram, Tiktok, and Google. Youtube is a visual heavy site, so we expect most viewers to have decent visualization literacy. We expect even higher literacy for Youtubers and Advertisers as they have to look at insight visualizations for their job. Because of the high level of visualization literacy and familiarity with the topic, we will present the information in a straightforward and detailed way.
3. Questions
    a. Correlation in Youtube Shorts Popularity and when Tiktok shut down (moving traffic) ~ Carlos
    b. How important is engagement - SEAN :)
        i. Dislike & negative engagement - what is the impact on performance?
        ii. Positive engagement - is it as valuable as we think?
        iii. Does the content of engagement matter (Kids leaving the comment "G" vs adults writing essays in comments)
    c. How important is content
        i. Video length and view count correlation ~ JONI
        ii. Length of titles ~ JONI
        iii. Topic ~ JONI

iv. Genre ~ JONI

v. Hashtags ~ Carlos

d. Overall popularity of Youtube shorts

e. Popularity of other social media sites and features Youtube adopted. ~ Carlos

i. Vine, Snapchat, Tiktok, Twitch

ii. Features like stories, hashtag, continuous scroll, paid subscriptions,

f. Popularity of Youtube videos over 10 minutes. ~ Ewuresi

g. Popularity of Youtube videos under 10 minutes~ Ewuresi

h. Popular categories of Youtube videos over 10 minutes. ~ Ewuresi

i. Popular categories of Youtube videos under 10 minutes . ~ Ewuresi

4. Data Types (categorical, ordinal, or quantitative)

a. Categorical

i. Genre

ii. Creator

iii. Title

iv. Tags

b. Ordinal ~ Carlos

i. Channel Rank

ii. Video Rank eg. Trending

iii. Hype Feature

c. Quantatative

i. Likes & Dislikes

ii. View Count

iii. Comment count

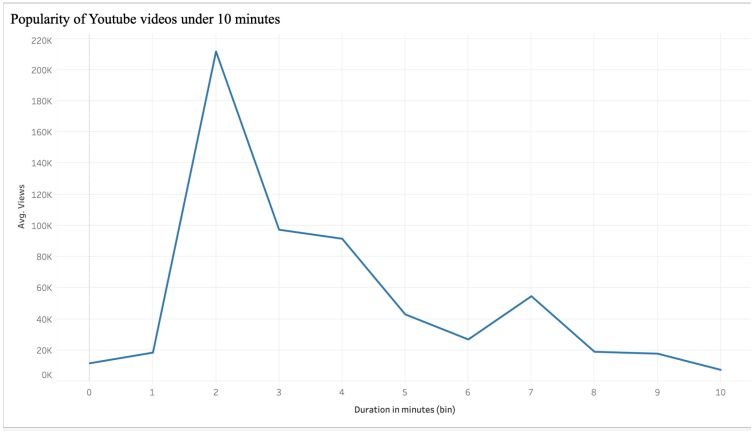
5. Cleanup Expectation

   a. Youtube, Google n-gram, and Instagram API is widely available. There are already datasets available on sites like Kaggle that will make the majority of the data already clean. However, if we find ourselves unable to answer a substantial question, we will use Youtube API directly, the processing workload will be a lot more significant. Will we try to not run into this case and prioritize pre-made datasets when possible.

Data Exploration

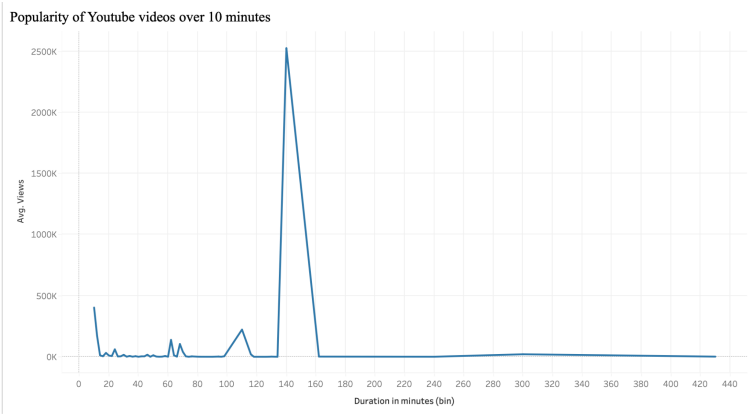
Ewuresi's Visuals and Insights

1. Popularity of Youtube videos under 10 minutes



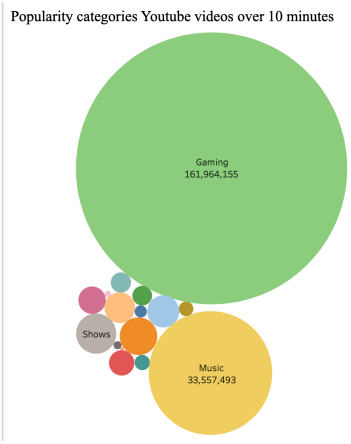
   - Spikes around the 2 minute mark
   - Declined and spikes a little bit around the 7 minute mark

2. Popularity of Youtube videos over 10 minutes



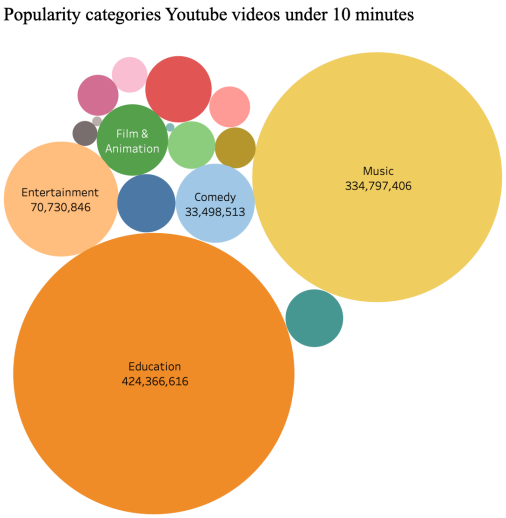
- Spikes around the 2 and half hour mark
- Videos over 3 hours get less attention

3. Popular categories of youtube videos over 10 minutes



- Gaming most popular category likely do to streams
- Music is second most popular probably because of music videos

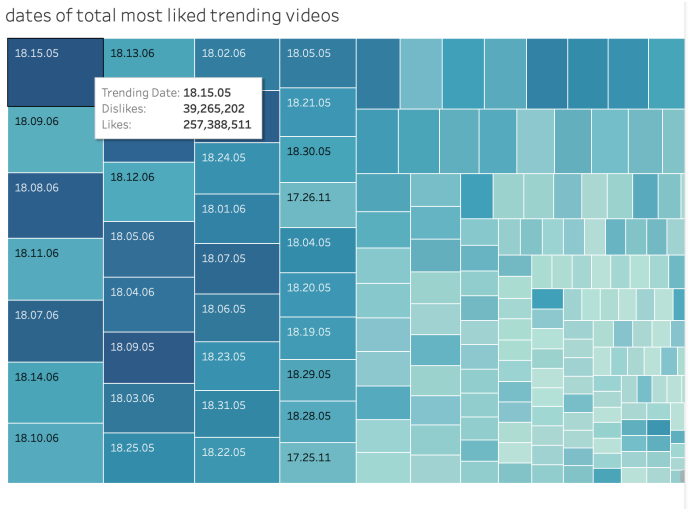
4. Popular categories of youtube videos under 10 minutes



- Education and music popular categories

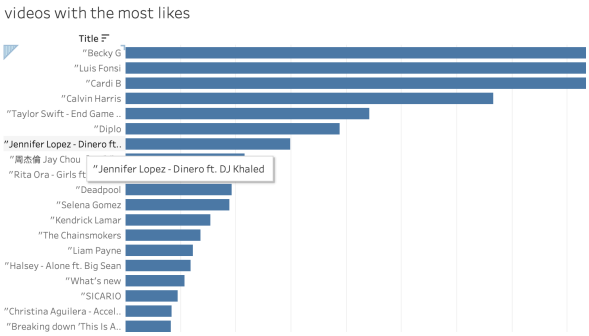
Joni's (Jacqueline) Visuals and Insights

5. Dates of total most liked trending videos (darker the circle, the most dislikes)



- The number of likes weakly correlates with the number of dislikes.

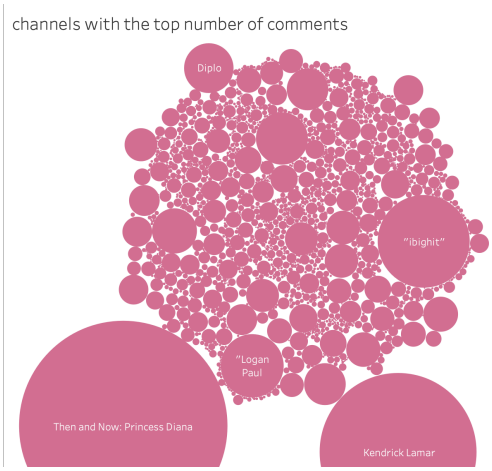
6. Videos with the most likes



- The majority of the topped liked videos are Music Videos

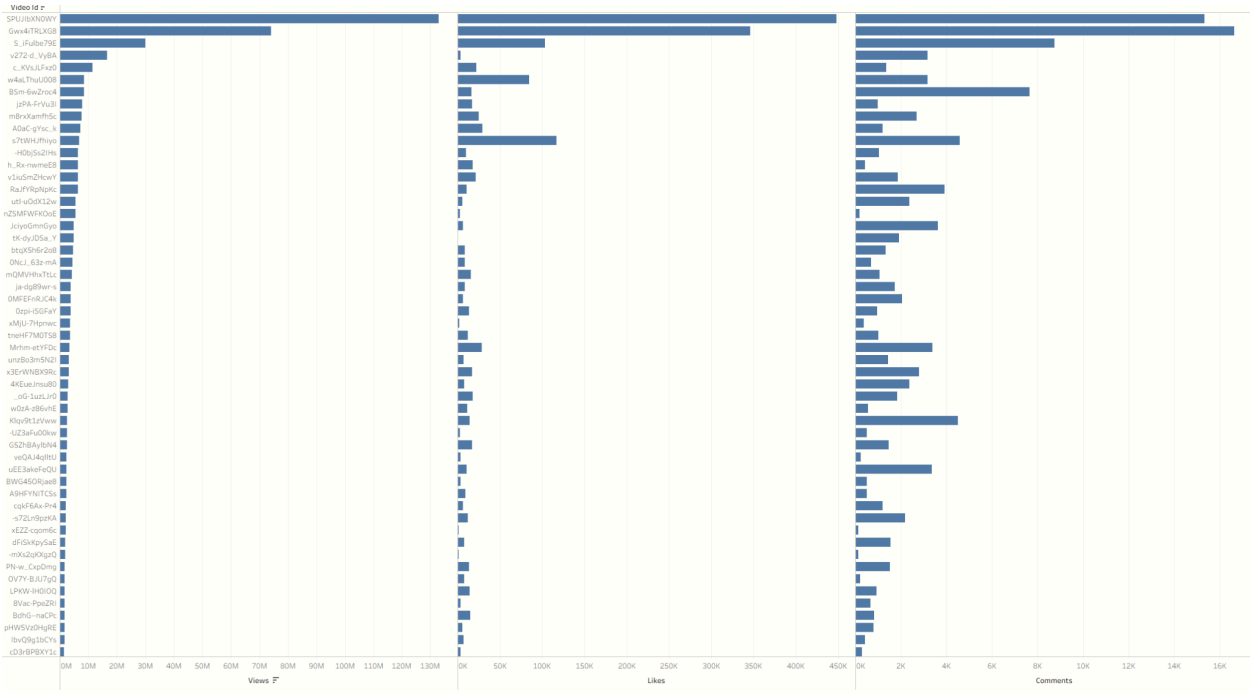
- Should consider combining graphs by insight ex. Likes, comments, etc.

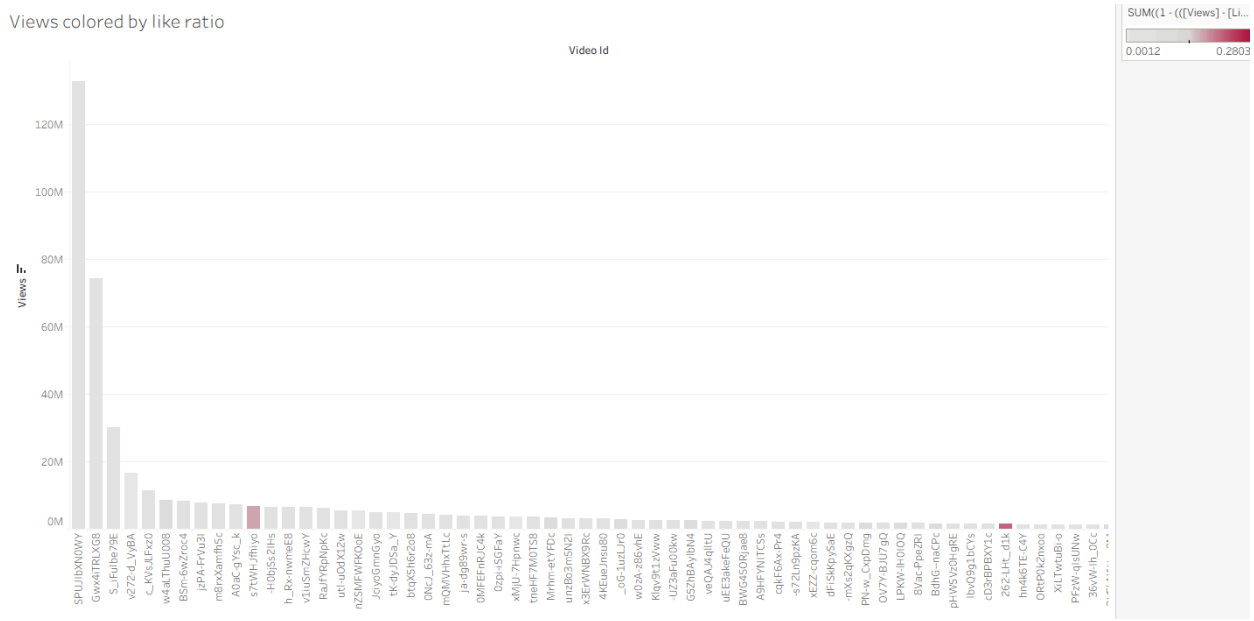
7. Channels with the top number of comments



-Should parse dataset to make more in depth visualization. For example, hashtags were not automatically separated in the dataset, if we can make a python script to do this, we could know the most popular tags for trending videos.

Sean's Exploration & Insights

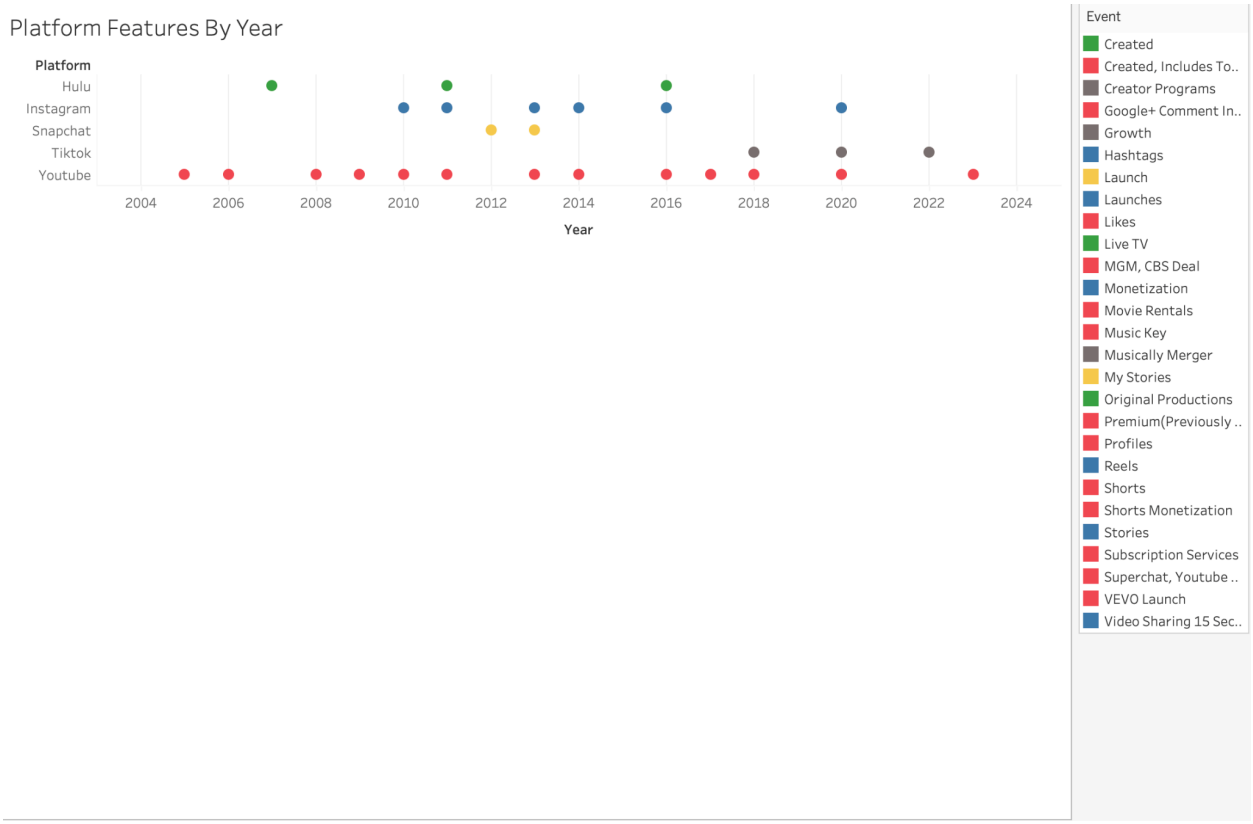




1. At the higher end, the ratio of comments to views tends to lower, as expected from videos with repeatable viewing (i.e. music videos, children's videos), but for more moderate viewership, higher comment ratios are typically correlated with higher viewership.

2. Excluding the higher tier, like quantity is a metric of lesser importance than comment quantity.
3. Like ratio stays relatively standard through the data, and widely beloved (e.g. high like ratio) videos do not tend to perform significantly better, and are relatively rare in the data.

Carlos's Exploration and Insights



Insights regarding the above graph

- Youtube adopted various features from their competitors over the years in response to trends such as Snapchat's stories, TikTok's "for you page" styled content delivery and creator program, and providing Television in response to Hulu.
- Looking at the features added by dates, you get a feel for how youtube was trying to compete. After Hulu launched, Youtube made deals with several entertainment companies like MGM and Lionsgate to offer TV services on their platform. Shorts implemented monetization features quickly after Tiktok began offering it in 2022.

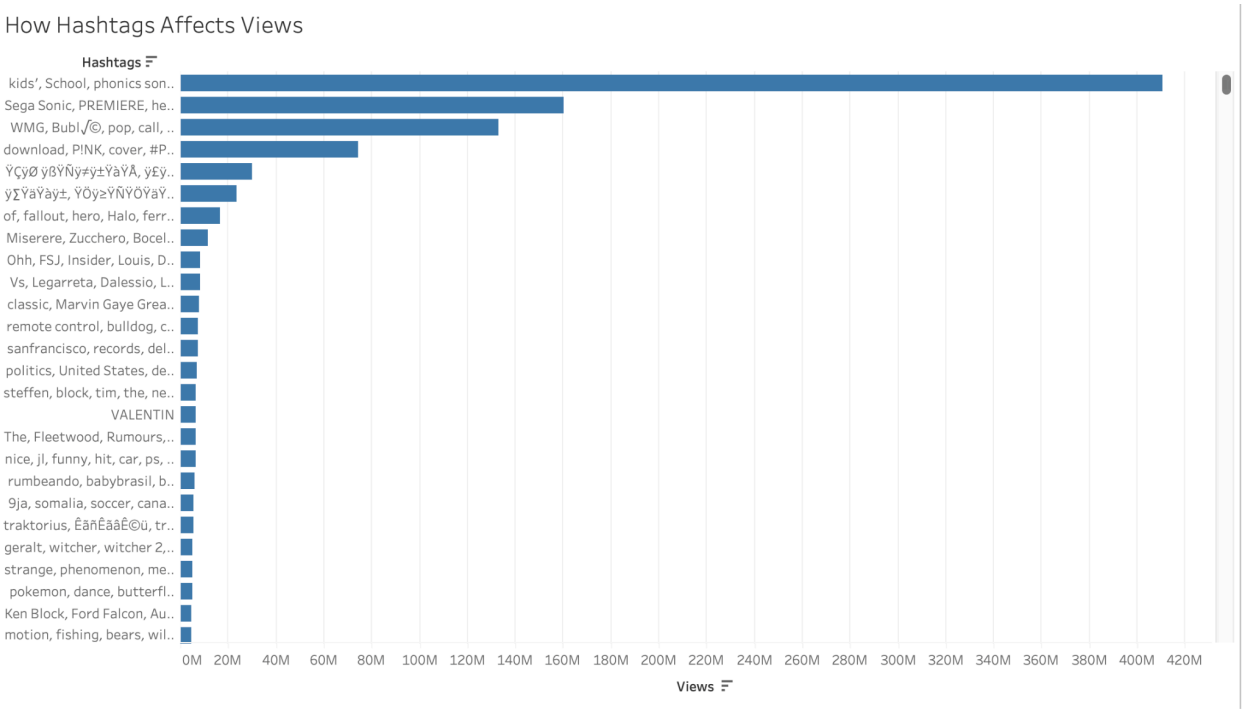
- No direct data showcasing the impact of the TikTok Ban on platforms such as Youtube Shorts. Apps like Redone and Lemon 8 such a surge in popularity following the announcement and short ban of the app.

Regarding Ordinal Data

- Youtube implemented a feature allowing viewers to hype a video published on YouTube up to 7 days after it's uploaded. They then are ranked on country-specific leaderboards to boost visibility.
- Youtube doesn't officially provide channel ranks, but instead channels are ranked externally using their publicly available subscriber count.
- While recently shuttered, Youtube's trending page selected videos based on view counts, region, topics, and relative views to other videos by the same uploader.

Other

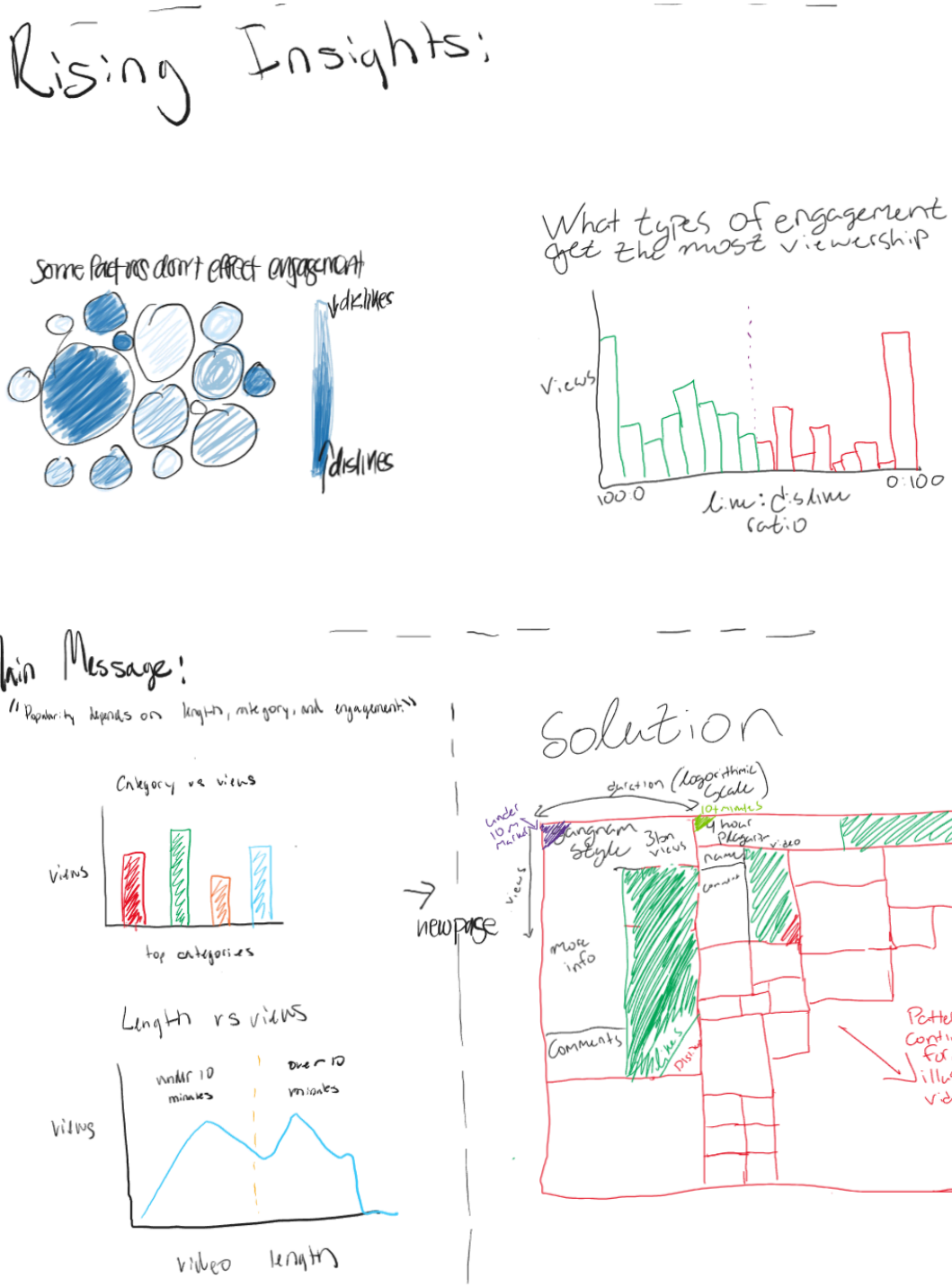
- For youtube shorts, hashtags can potentially boost a shorts visibility, but not necessarily. Youtube advises creators to put them in on a need by basis.
  Source: https://www.youtube.com/watch?v=n3jsYK_-aRU&t=316s
- The graph below shows an idea of this, many videos include a ton of hashtags with varying amounts of views. So it's hard to say the impact that an individual hashtag has on a video's visibility.



Storyboard

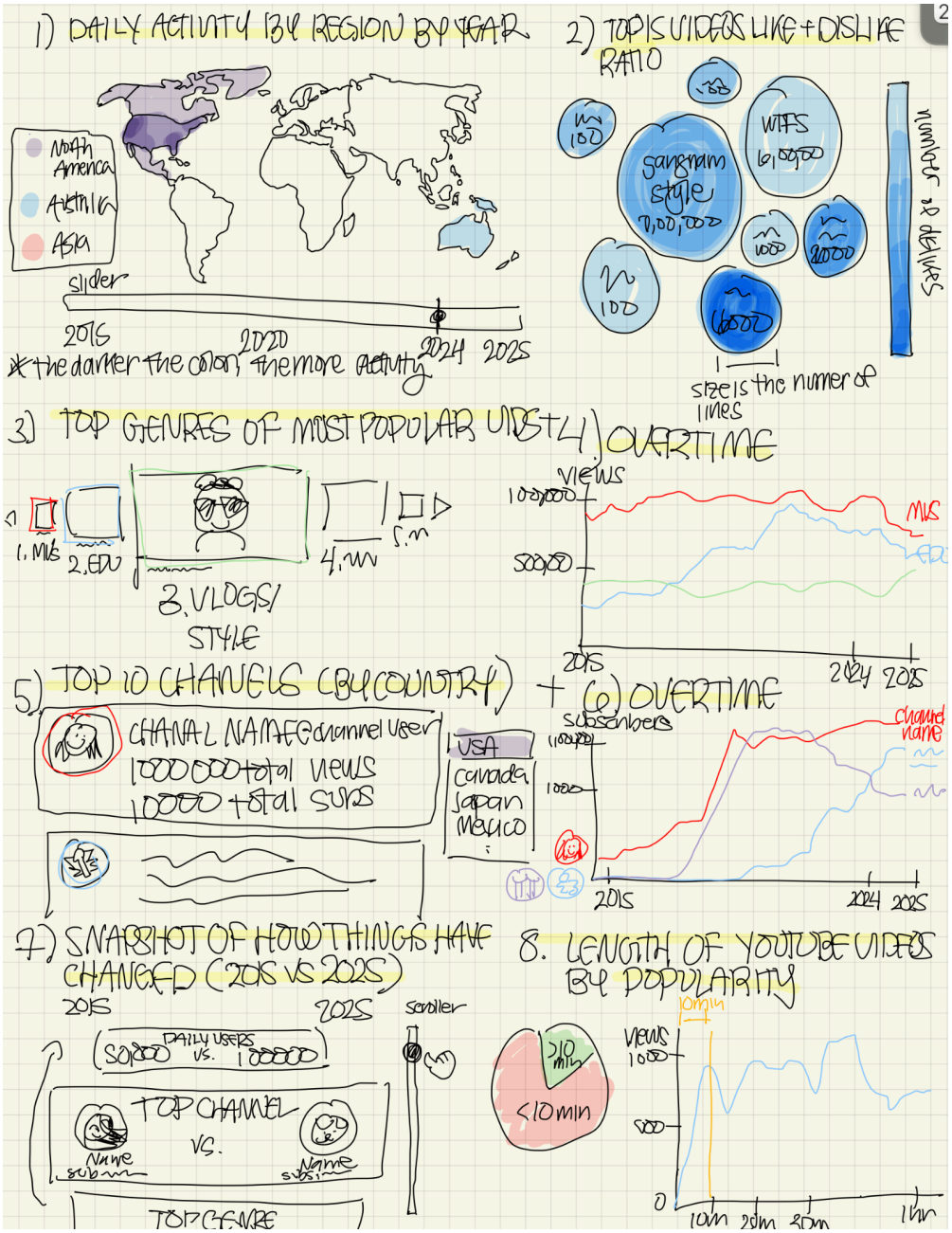
Main Message

From the data visualizations we made as a team, we decided that our main insight is what factors are in play to create the most popular youtube video. Meaning that the popularity of the video is not random, but depends on key factors such as video length, category, and audience engagement. Our team thought this thesis reflected all of our visualizations, as all of our questions were pointing to the factors to produce engagement.


Rising Insights:
Some features don't affect engagement
likes
dislikes
What types of engagement get the most viewership
views
100:0
0:100
like:dislike ratio
Main Message:
"Popularity depends on length, category, and engagement"
Category vs views
views
top categories
Length vs views
under 10 minutes
over 10 minutes
views
video length
new page
Solution
duration (logarithmic scale)
10+ minutes
Gangnam Style
3bn views
4 hour Pleasure video
name
comments
views
more info
Comments
likes
Pattern continues for most illustrative videos

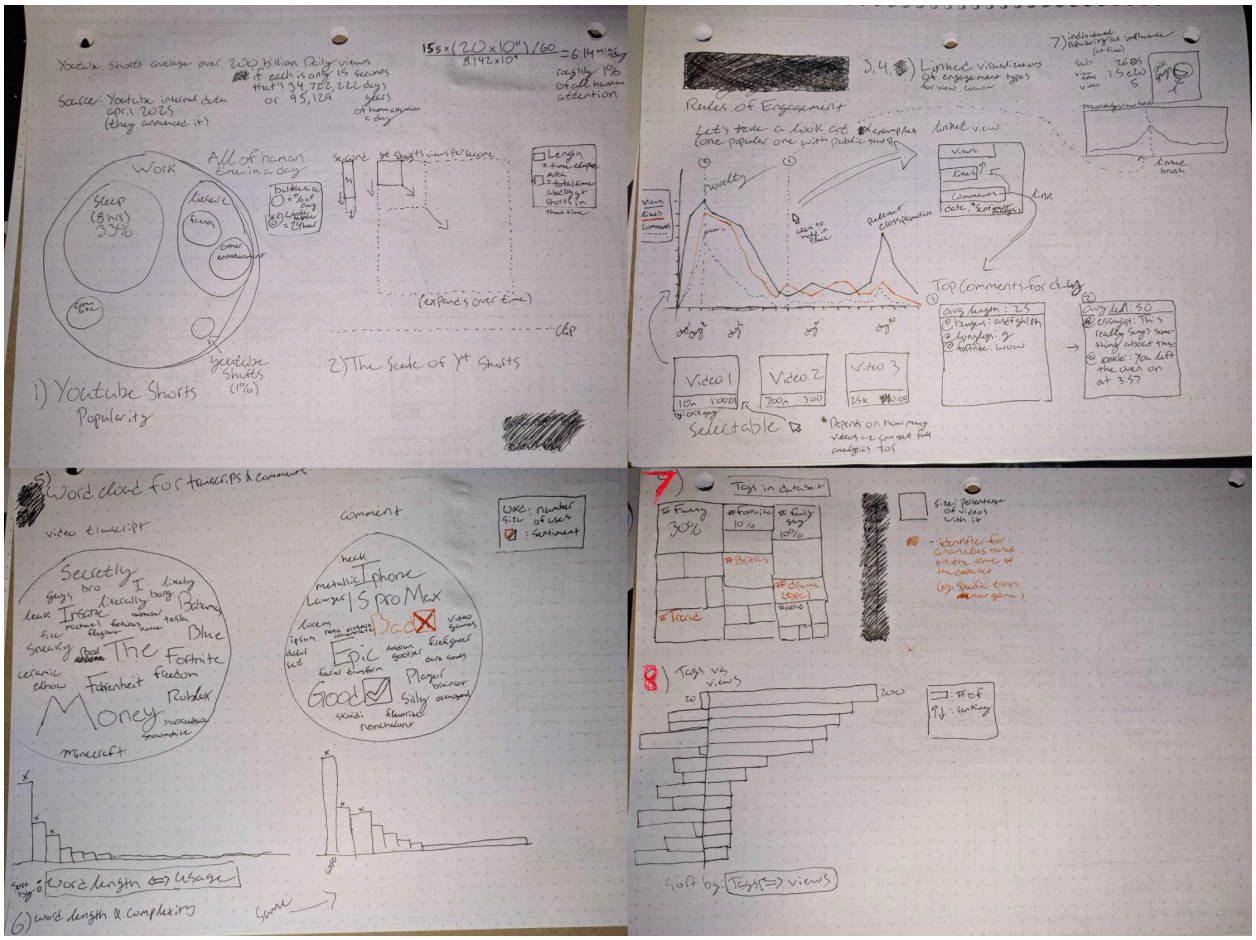
Sketches

Joni's Sketches



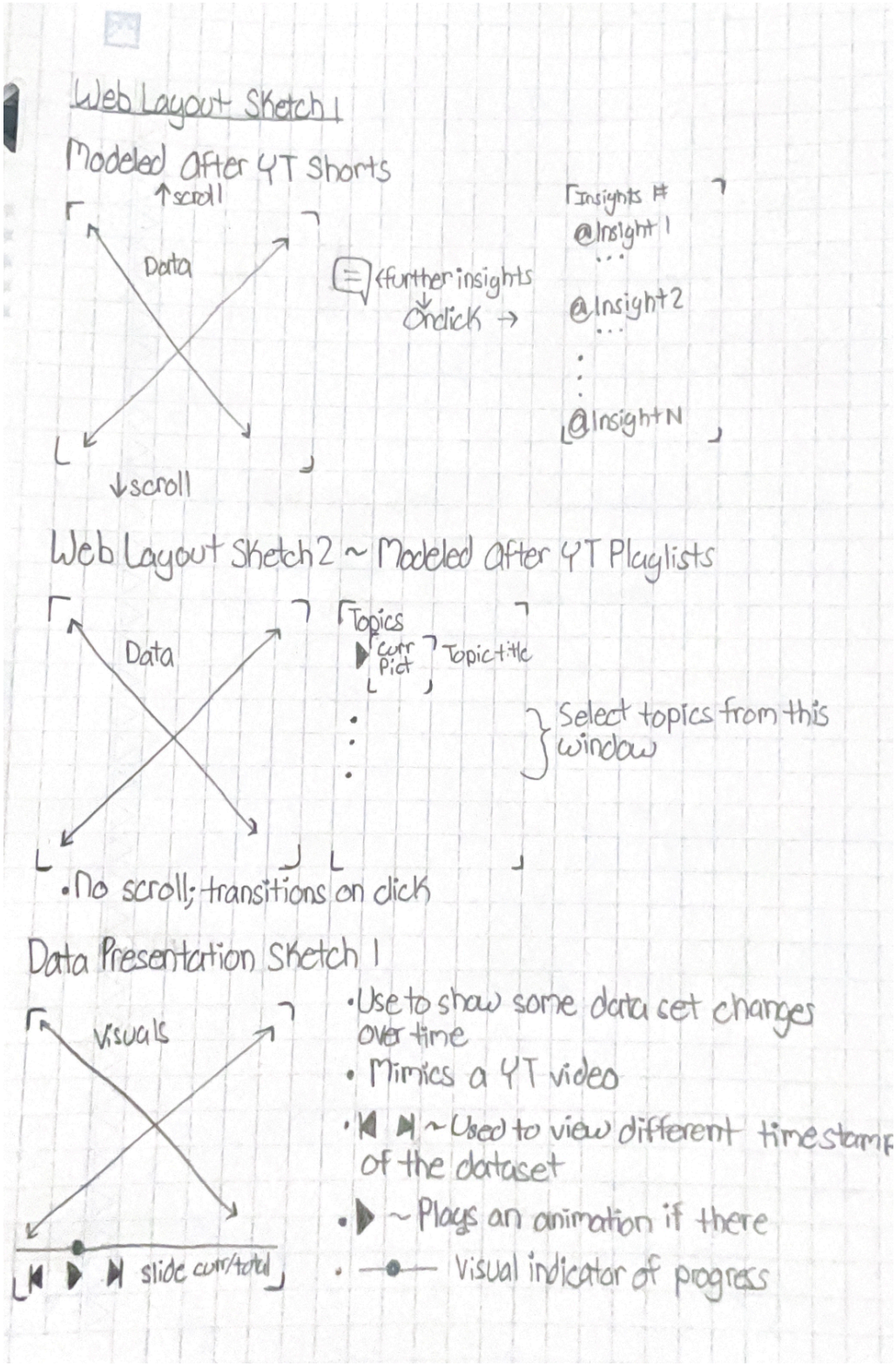
J1, J2… J8, in same order

Sean's Sketches

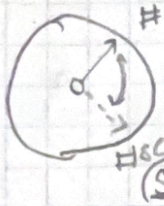


S1, S2, S3… as labeled (S(n))

Carlos' Sketches



- Sketch of Views Vs Length ~ Based Off Stopwatch App



- Clicking start moves the hand on the clock the distance of a video length
- Should maybe always start at 1 to make clear the difference

Length 1
⋮
Length n

~The bottom half is populated by most popular videos by views and length

- Sketch of Trending Topics VS View Count in an Area

Map of States or Countries

State/country

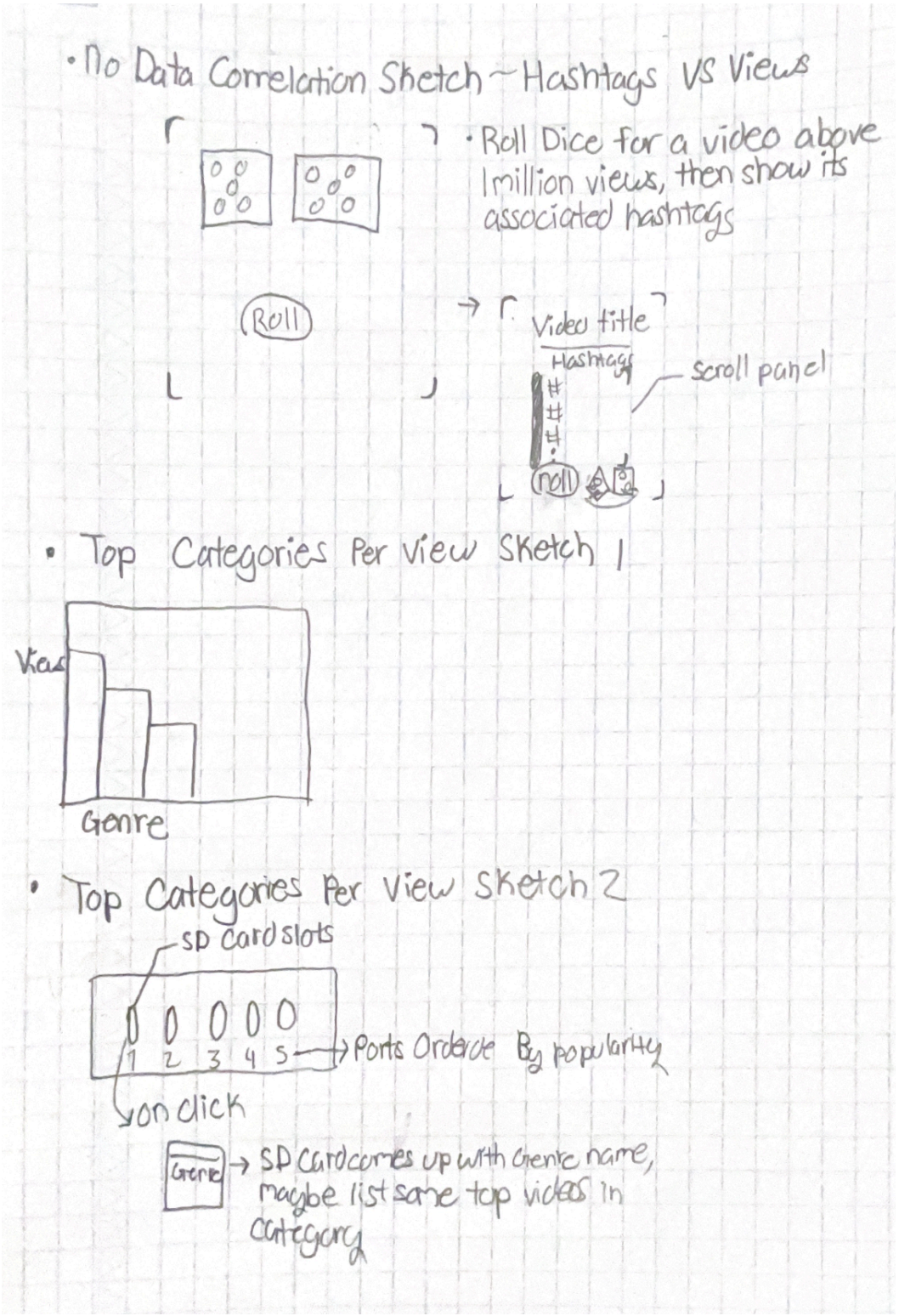
onclick

1. Pop. Video
2. Pop. Video
3. Pop. Video

} Videos ranked on popularity

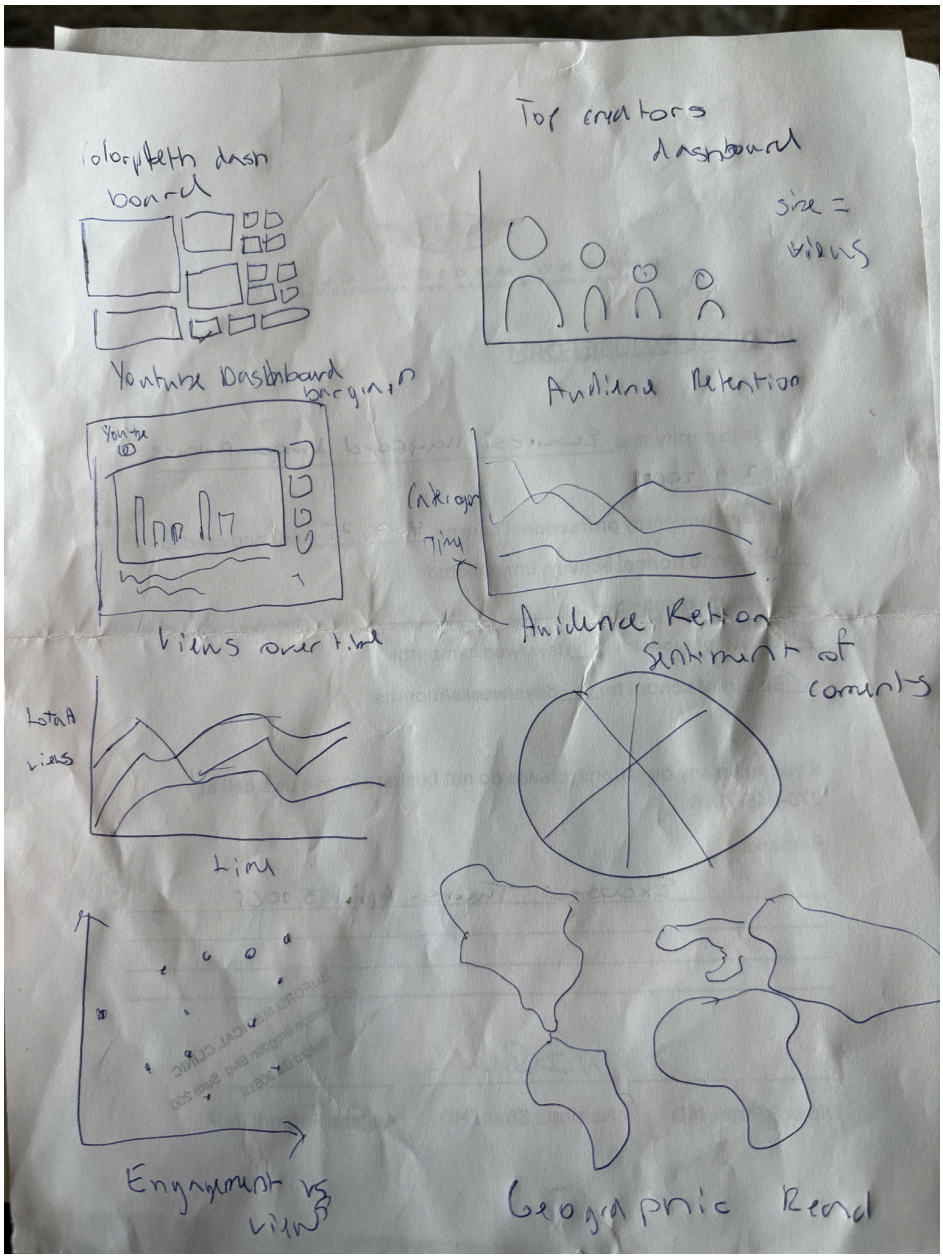
1) Topic
2) Topic
3) Topic

} Top trending topics for some time frame



The sketch ID's are C1-8 in order of appearance. Starting with Weblayout Sketch 1 and ending with Top Categories Per View Sketch 2.

Ewuresi's Sketches



Sketch IDs are A1-8 from left top column down and then the right top corner down.

Decide

SKETCH ID	MESSAGE ID	AUTHOR	VOTES
J1, C6, A8	1, 12	JI, CR, EA	*$
J2	2	JI	+
J3	3	JI	CR
J4	3	JI	
J5	4	JI	*$CR
J6	4	JI	
J7	5	JI	
J8	6	JI	
C1	7	CR	
C2	7	CR	
C3	8	CR	
S1	10	SM	*
S2	10	SM	*CR
C5	11	CR	*$
S7	11		
S8	11		
C4	6	CR	+CR
C7	3	CR	

C8	**3**	**CR**	**+$**
S3	**12**	**SM**	**+$**
S4	**12**	**SM**	
S5	**13**	**SM**	
S6	**13**	**SM**	
A1	**7**	**EA**	
A2	**7**	**EA**	**+CR**
A8	**1**	**EA**	
A3,A4	**11**	**EA**	
A5,A7,A6	**13**	**EA**	

Message IDs:

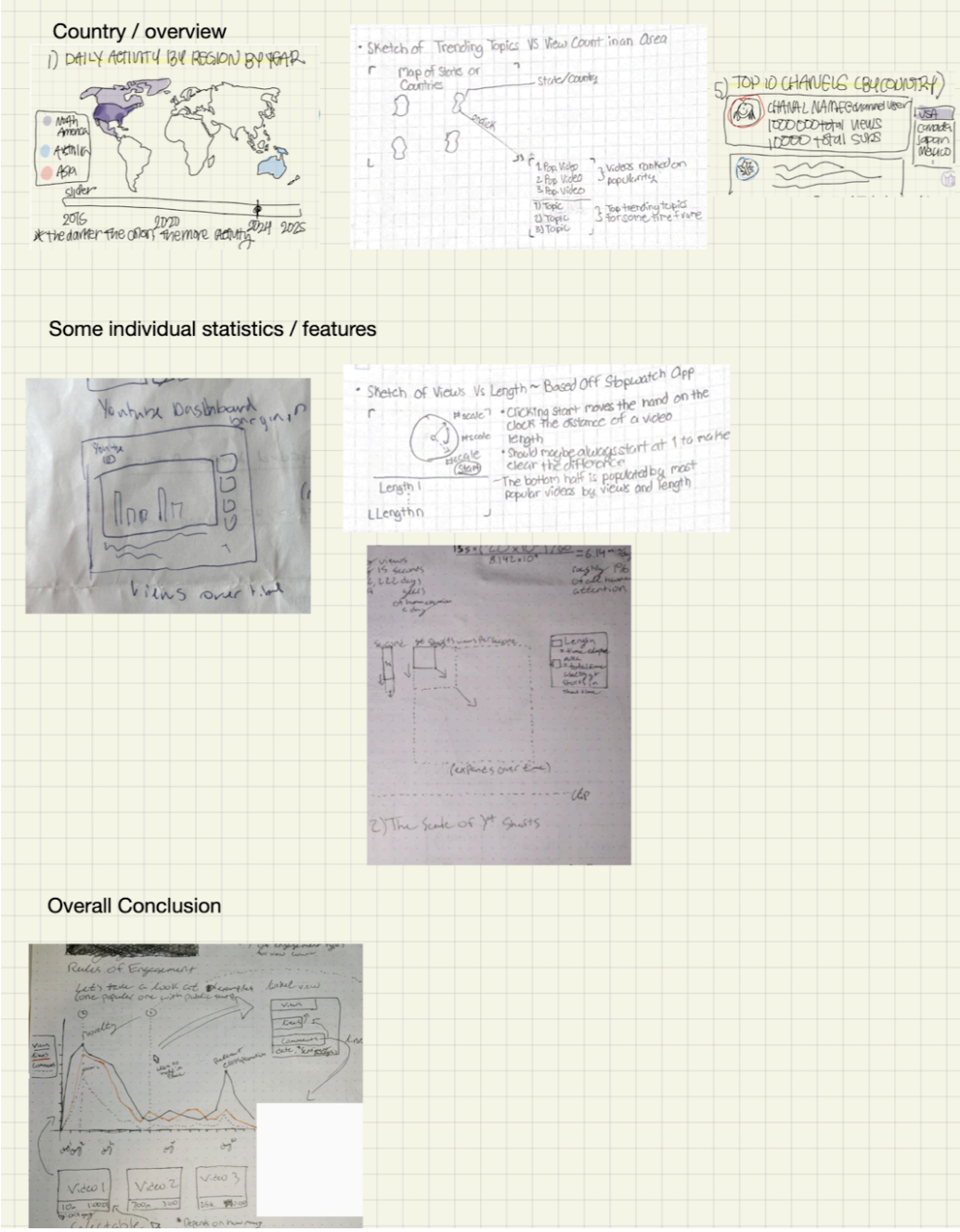
1. **Daily activity by region by year**
2. **Top 15 videos like + dislike ratio**
3. **Top genres of most popular vids**
4. **Top 10 channels by country**
5. **Snapshot of 2015 vs 2025 analytics**
6. **Length of youtube videos by popularity**
7. **Web layout**
8. **Way to present data over time**
9. **Overall popularity of Youtube shorts**
10. **Most popular videos in a region VS topics for a time frame**

11. **Hashtags Vs Views, Display of no correlation/random**

12. **How Engagement correlates with likes**

13. **General Vernacular of youtube with sentiment**

CHOSEN SKETCHES



Decisions and Rationale

Our decisions were based on what might impart the scale of the data with specific visualizers for cases to try and give an overall illustration of the data. The hook of the story is impact, so showing the area and amount of youtube being watched was important. Then, we needed to show what's important on a larger scale with wider trends, while also having smaller case studies which are able to show more. We wanted to give a variety of approaches to the data, and these illustrations were the ones that best fit that goal.

Prototype V1 & V2

Student design segments and TODO

Sean Mulloy

- Implemented animated and interactive shorts visualization
- Filled out main structure and styling, TODO: make it better
- TODO: add visualization and interaction for conclusion (rules of engagement)

Carlos Robinson

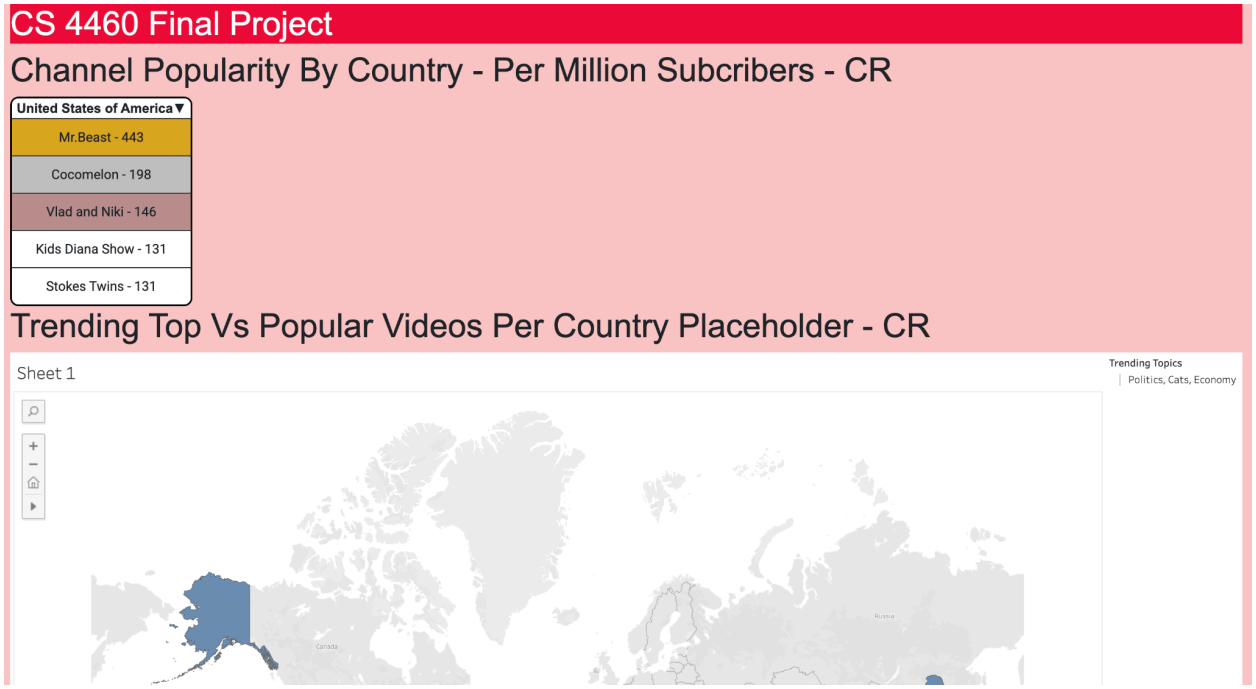
- Channel Popularity By Country Per Million Subscribers D3 Draft
- Popular Videos Vs Trending Topics Global Tableau Draft
- TODO: Change Tableau Draft to D3 implementation with geoData.

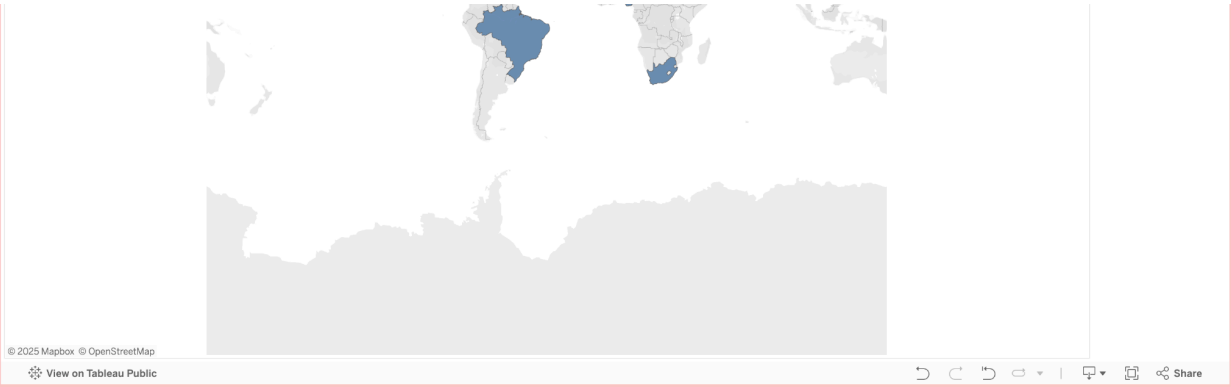
- TODO: Get more complete dataset for popular videos/genre in a given region using views
- TODO: Further fill out dataset for top 5 most subscribed youtubers in every country.

Ewuresi Amoo-Adare

- Implemented basic pie chart placeholder D3 visualization for views vs length with dummy data.
- TODO: finish up D3 visualization for pie chart
- TODO: start and complete d3 youtube dashboard/ bar chart visualization.

Screenshots of Webpage V1





Current Tableau Implementation, ToDO: Change to D3 w/ Complete Data

Views vs Length - EA

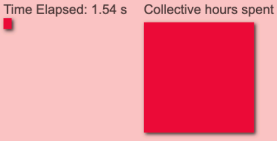
TODO: Under here there will be a d3 visualization pie chart comparing view count of video to its length.

The Scale of Time spent on Youtube - SM

TODO: Make this text description make sense

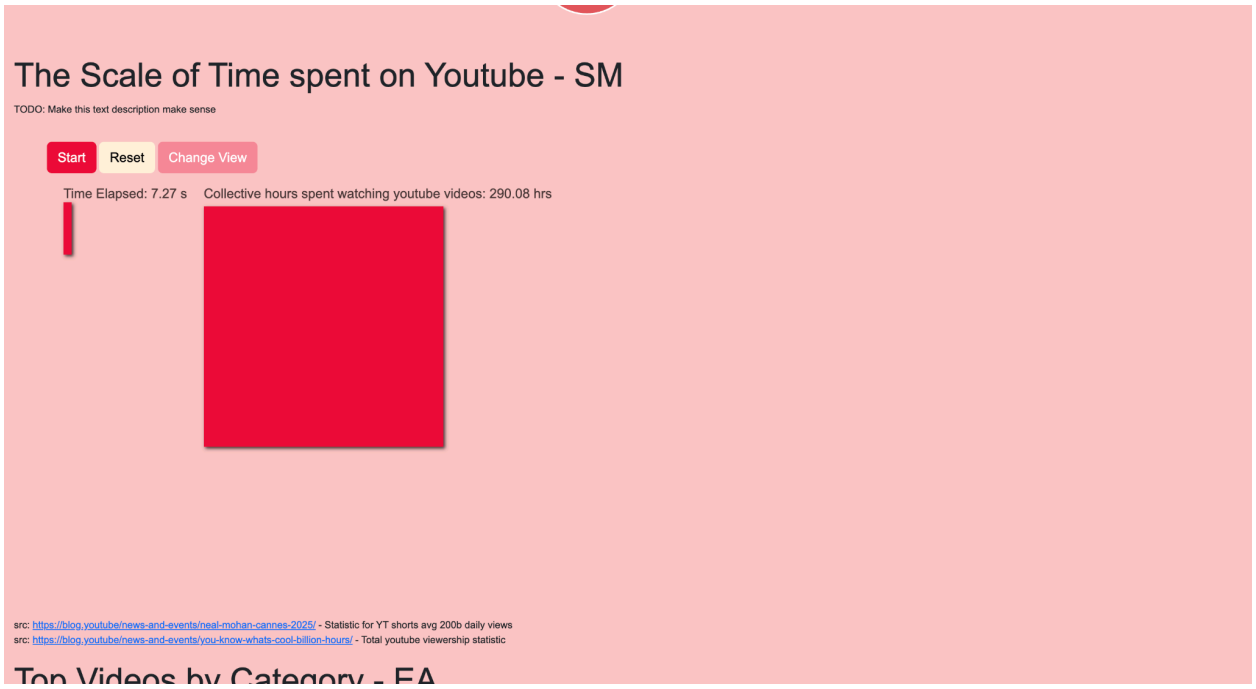
Start Reset Change View

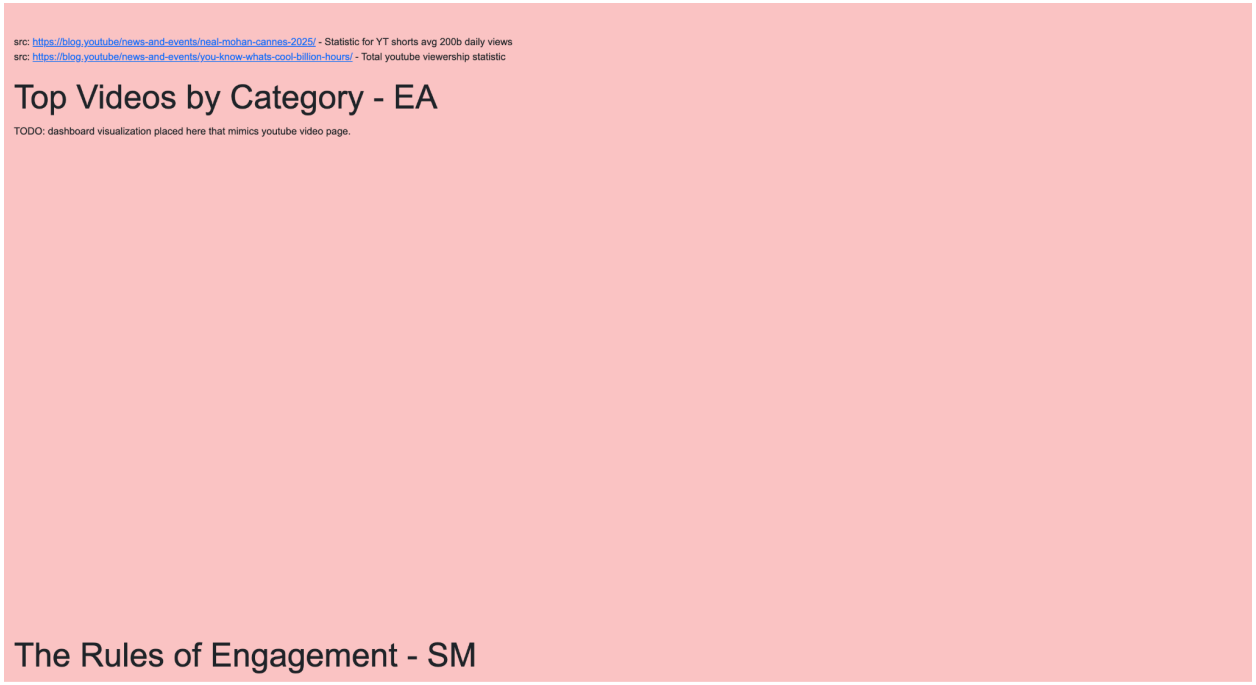
Time Elapsed: 1.54 s Collective hours spent watching youtube videos: 133.51 hrs



src: https://blog.youtube/news-and-events/neal-mohan-cannes-2025/ - Statistic for YT shorts avg 200b daily views
src: https://blog.youtube/news-and-events/you-know-whats-cool-billion-hours/ - Total youtube viewership statistic

Top Videos by Category - EA





Students Working on Prototype V2

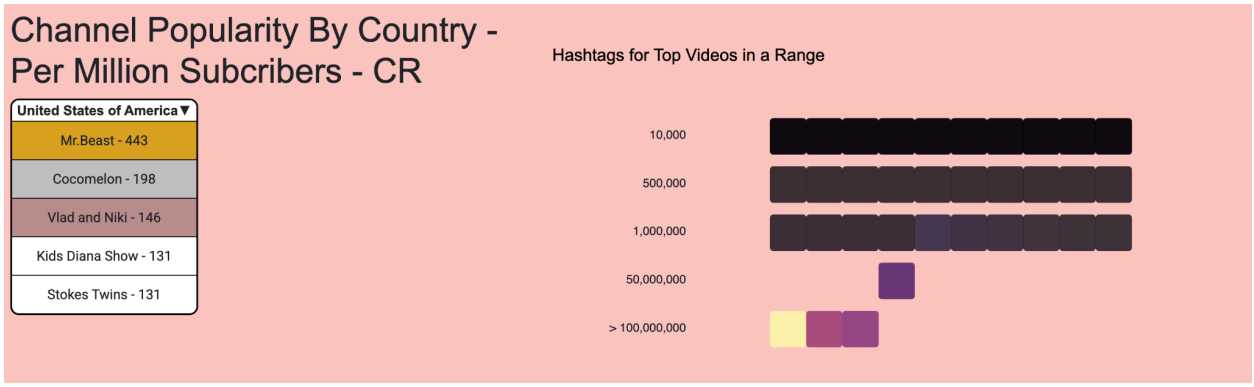
- Ewuresi Amoo-Adare, Carlos Robinson, Sean Mulloy

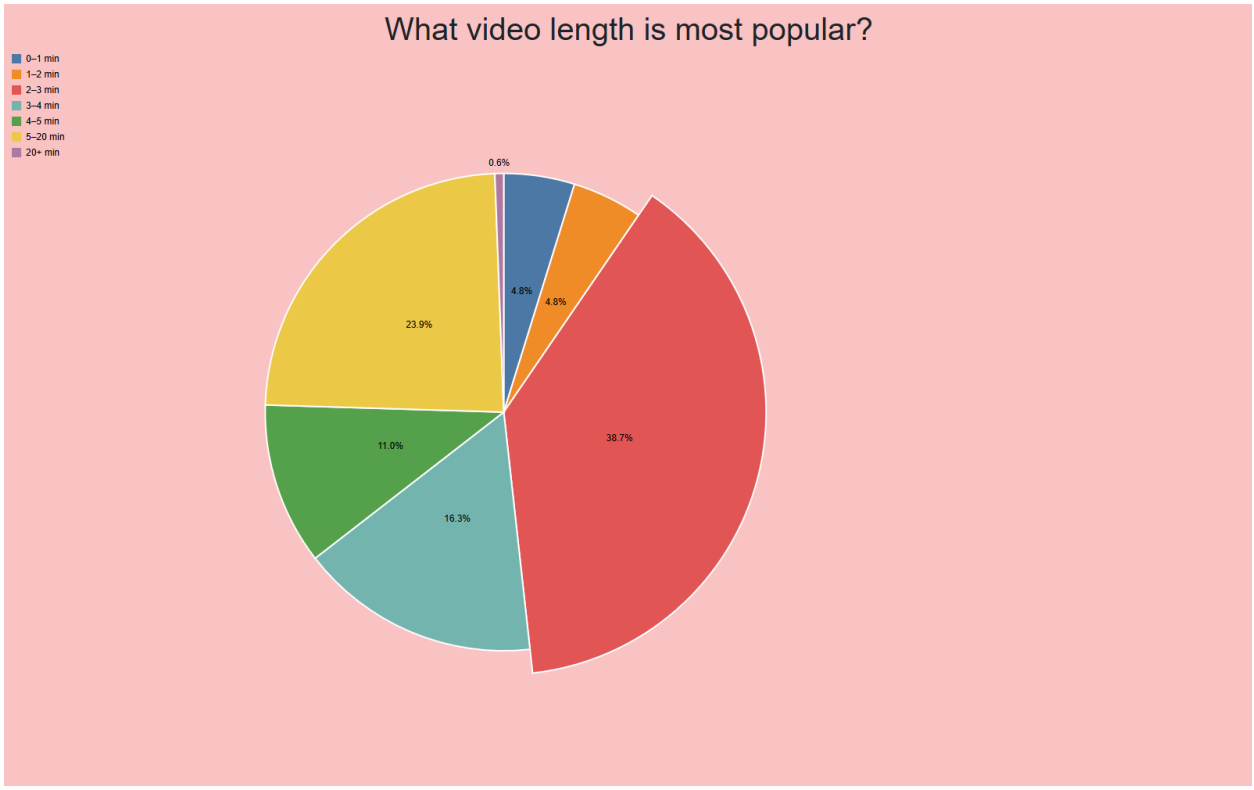
Prototype V3 & V4

Students Working on Prototype V3 & V4

- Ewuresi Amoo-Adare
    - Implemented Video Length and Video Dashboard vis
    -
- Carlos Robinson
    - Implemented Channel Popularity and Country vis
    -
- Sean Mulloy
    - Implemented Shorts Vis & Rules of Engagement Vis
    - Some behind the scenes refactoring
    - Dealt with all the merge conflicts
    - TODO: Fix a few bugs with vis 3 & 5 and restyle page

Screenshots of Implementation




What video length is most popular?
0–1 min
1–2 min
2–3 min
3–4 min
4–5 min
5–20 min
20+ min
0.6%
4.8%
4.8%
23.9%
38.7%
11.0%
16.3%

The Scale of Time spent on Youtube

Youtube announced that Youtube Shorts were reaching 200 billion daily views in 2025.
In 2017, they had announced that, in total, people were watching a billion hours of Youtube each day.
Lets put that into perspective...
(made by Sean)

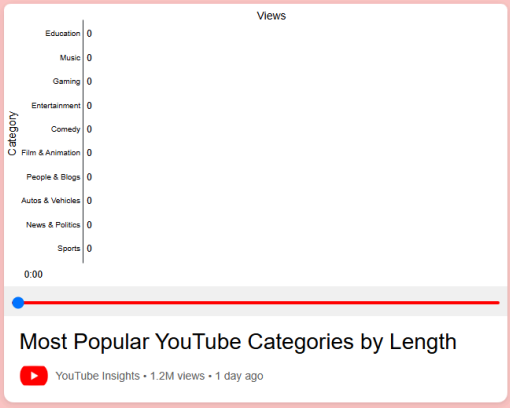
Start | Reset | Change View

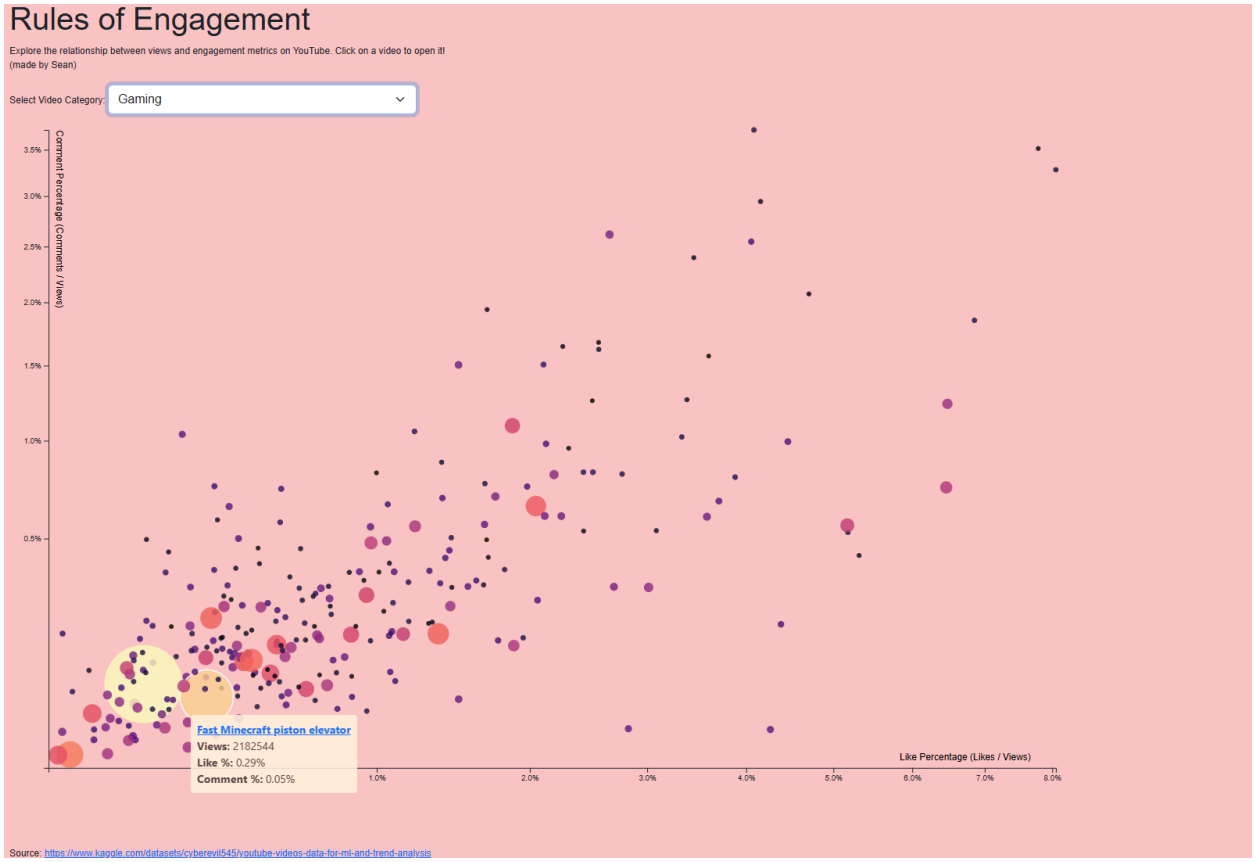
Time Elapsed: 28.09 s Collective hours spent watching youtube shorts: 520.51 hrs

src: https://blog.youtube/news-and-events/neal-mohan-cannes-2025/ - Statistic for YT shorts avg 200b daily views
src: https://blog.youtube/news-and-events/you-know-whats-cool-billion-hours/ - Total youtube viewership statistic

src: https://blog.youtube/news-and-events/neal-mohan-cannes-2025/ - Statistic for YT shorts avg 200b daily views
src: https://blog.youtube/news-and-events/you-know-whats-cool-billion-hours/ - Total youtube viewership statistic

the yt dashboard one - Ewuresi





Test / Evaluation

Tester Name: [Tester 1] [Tester 2]

Tester Email: [Redacted] [Redacted]

Unfortunately, we did not get the name or email of either tester

General Observations from the think-aloud study:

Thesis was generally unclear

The general color palette was too harsh for tester 1, Tester 2 did not seem to mind. Individual vis colors were not of note.

Some elements were unclear in their interactivity for tester 1, tester 2 had no issues

There were a few minor bugs that neither tester audibly noticed but slightly misrepresented the data.

What does the tester like about your data story?

Testers enjoyed the interactivity (specifically the elements that take them offsite)

Testers were curious about the data and spent a while exploring it

Tester 2 was impressed by the scale of the data

The general premise was relatable, as both testers were familiar with youtube

What improvements does the tester point out?

Shorts vis should have a pause button

The dimensions of the youtube dashboard vis should be clarified

Engagement vis should have prominent channels and axes more prominent

Tester 2 wanted the heatmap to have clickability like the final visualization (note: this would distract people early on so we should not add it)

Overall, both desired more context

Was the intended key message clear to the tester? Why or why not?

For both testers, the thesis was overall unclear and there was a clear desire for more context. While the visualizations were able to give insight, both asked questions like “Huh, that’s interesting. Why is that? What’s the deal with that?” that could be better served by text in and around the visualization giving more related information about the trends in the data.

Did the tester get your next steps or call to action? Why or why not?

Both testers appeared to correctly follow the next steps (tester 2 in particular seemed pretty taken by it) of “See for yourself!” after the final visualization allows the users to go off-site and actually see many of the videos for themselves.

Changes to be Made

- [x] ~~Make stronger narrative throughline~~
- [x] ~~Redesign page to be more appealing and followable~~
    - [x] ~~Add further text to explain insights visible in visualizations~~
        - [x] ~~Vis 1~~
        - [x] ~~Vis 2~~
        - [x] ~~Vis 3~~
        - [x] ~~Vis 4~~
        - [x] ~~Vis 5~~
        - [x] ~~Vis 6~~
- [x] ~~Vis 1 (Shorts vis)~~
    - [x] ~~Start on render~~
    - [x] ~~Remove clipping [Could not be done without occluding other visualizations or interfering with their color encoding]~~
    - [x] ~~Make encoding more clear (eg. clarify what the "size is meant to mean")~~
- [x] ~~Change vis 2 (Top Channels by Country) to be a map CR~~
- [x] ~~Clarify vis 3 (Heatmap)~~
    - [x] ~~Convert to weighted graph~~
    - [x] ~~Changed to network graph CR~~
- [x] ~~Convert vis 5 into a linked view~~
    - [x] ~~Add a tree map showing proportions of the whole taken up by each category~~

- ☑ ~~Start Vis at a value greater than 0:00 (Whatever looks best)~~

☑ ~~Make vis 6 more of a clear end point~~

- ☑ ~~Add links explaining parasociality and apathy to better explain engagement relationships~~
- ☑ ~~Make it more clear the bubbles are clickable~~
- ☐ [Stretch Goal] Make a linear regression

Final Product

Algorithm Folklore

What makes for a viral Youtube video?

The Scale of Time spent on Youtube

Youtube announced that Youtube Shorts were reaching 200 billion daily views in 2025.
Assuming the average viewership of a Youtube Short is 15 seconds, that would mean 3 trillion seconds
or approximately 0.5% of all time spent by humanity in a day is spent watching youtube shorts.
Lets put that into perspective...

Restart

Time Elapsed: 2.35 s

One Second

Collective hours spent watching youtube shorts: 22665.90 hrs

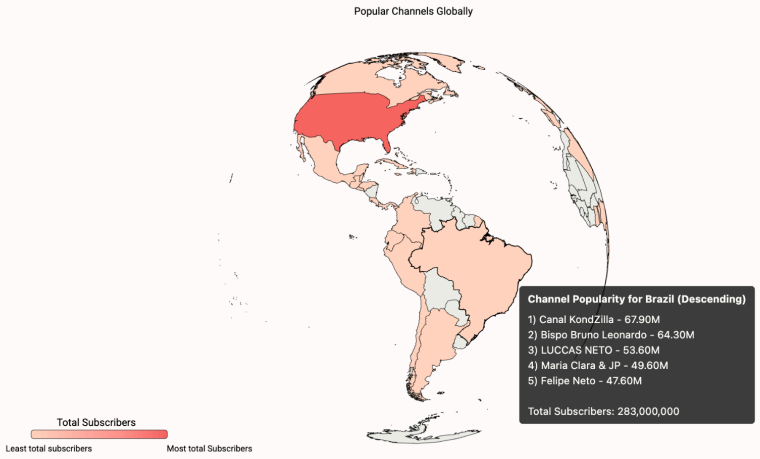
src: https://blog.youtube/news-and-events/neal-mohan-cannes-2025/ - Statistic for YT shorts avg 200b daily views
src: https://blog.youtube/news-and-events/you-know-whats-cool-billion-hours/ - Total youtube viewership statistic

Popular Channels Across the World

Considering that so much time is spent on youtube, it would be worth considering who is spending that time and on which channels.

Language barriers, cultural preferences, and regional trends all factor into popularity within a given region.
Virality and viewership is thereby a relatively local phenomenon.
Given that, demographic context is important for the popularity and virality of any Youtube video.

For exceptions, countries with poor or restricted internet access (e.g. Antarctica and North Korea) are excluded in grey.
The second most populous nation in the world, China, is also suspiciously underrepresented due to The Great Firewall's blocking of most google services, including Youtube.



Hashtag Associations

Tags and tagging are often touted as an exploit to gain popularity, but this is a rather superstitious piece of folklore from times long past. In the modern era, tags that affect outcome tend to be more clear identifiers of content as to create a valid dataset for the almighty Youtube Algorithm.

Due to this, tags are often utilized haphazardly, but ones of relevance are still able to create a clear picture of the Youtube landscape.

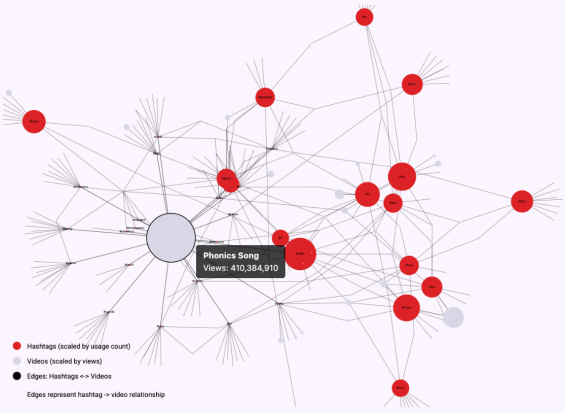
The original data set includes 110,235 unique hashtags and 17,588 videos.

This visualization filters out 20 hashtags and the 10 most popular videos associated with each hashtag.

Here, we can identify what tags are relevant for success, but success does not simply come to those who tag incorrectly.

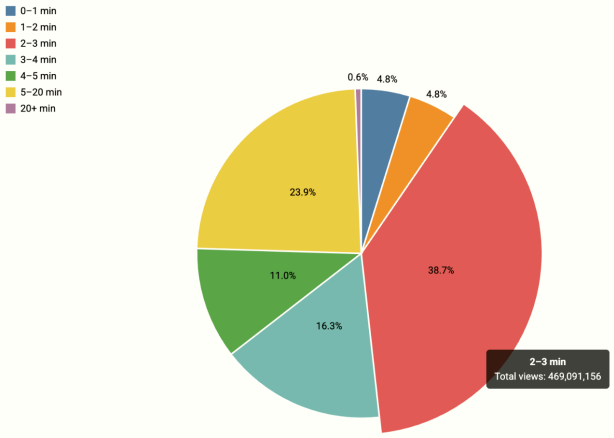
Select Network View

Top 20 Hashtags By Frequency



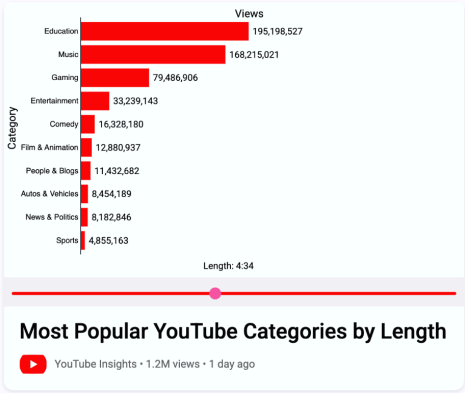
What video length is most popular?

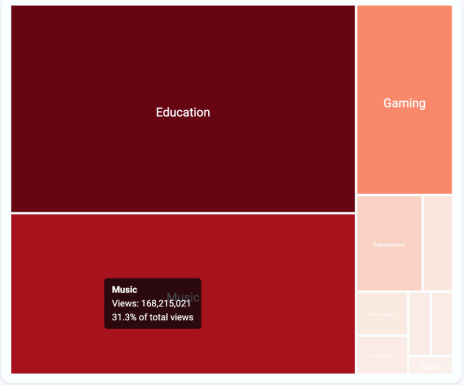
Youtube has a history of restrictions and preferences for certain video lengths.
For instance: The maximum length of a video was changed from 15 minutes to 12 hours in 2010.
In the modern lore, the optimal time is expected to be around 8-10 minutes, specifically for increased advertizing opportunities. However, much of the current landscape has been shifting shorter and shorter, especially in the wake of youtube shorts.



Length & Category

We can expect that different categories and types of videos will vary in length.
For instance, a fun fact about cornbread might be around the same length as the music video "Hips Don't Lie" by Shakira, but a CS50 lecture might run for the entirety of the Nier: Automata OST. Specific categories may have their preferences and hold a dominance in certain ranges of length.





Most Popular YouTube Categories by Length

YouTube Insights • 1.2M views • 1 day ago

Rules of Engagement

How important is engagement on youtube? It is common for career Youtube creators to focus on engagement metrics and interaction reminders. However, at a large scale, it is not strictly correlated with viewership.

Due to factors such as parasociality, niche interests, and algorithmic recommendations, videos can be directed to smaller but more engaged viewerbases. Larger audiences are less likely to be highly engaged, and a more varied audience will tend to include more apathetic viewers.

Click on a video to open it in a new tab!

Select Video Category: News & Politics