INFO 247 Final Project

YouTube Conspiracy Videos Exploration

Grace Chung, Chloe Lee, Lynn Yeom

Project Goals

Inspired by Professor Farid's research on YouTube's recommendation of conspiracy videos, our project team wanted to create a visualization platform that raises awareness on the false information presented in YouTube. Our goal focused on exploring various visualization tools learned in class (illustrator, tableau, and d3) and combining with our expertise in UX design and data science to create visualization that provides social impact. We have three target audiences for our tool - YouTube users, YouTube, and policy makers with the following goals:

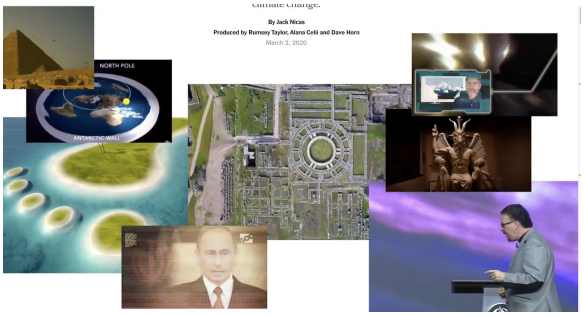
- For YouTube users: As a viewer, users should be aware and further be able to raise flags and start conversations about conspiratorial content in YouTube. Moreover, users should also be vigilant about the videos that are being recommended to them via YouTube.
- For YouTube: YouTube has the capability to expand on the research initiated by the researchers form UC Berkeley by incorporating user specific data into further analysis. With the results, YouTube needs to better develop algorithms that discourages recommendation of conspiratorial content, and be more transparent about the process.
- For policymakers: Social media platforms should not be held solely accountable for eradicating conspiratorial content. Policymakers should further collaborate with the platforms on monitoring conspiratorial content information and further set legislative boundaries to protect users from exposure to false information.

Related Work

1. [A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos](), by Marc Faddoul, Guillaume Chaslot, and Hany Farid

   This is the original study that inspired this visualization project. The authors collect information on more than 8 million+ recommendations to 200K+ videos made by YouTube's watch-next algorithm over the span of 15 months to analyze conspiratorial videos that are often recommended by YouTube. Our project uses data collected and created by these researchers, and we use their study as a stepping stone to explore how information visualization can help make further impact.

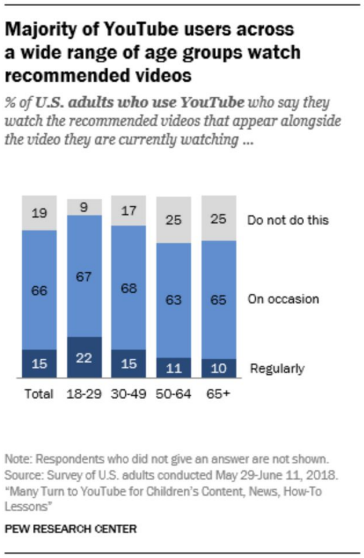
2. [Can YouTube Quiet Its Conspiracy Theorists?](), by Jack Nicas, produced by Rumsey Taylor, Alana Celii and Dave Horn (New York Times)



   This New York Times article was written based on the study (related work #1) by Faddoul et al. (2020). It gives a summary of the study, and discusses some feedback on the study from other sources, including Mr. Farshad Shadloo, a YouTube spokesman. Like our project, this article also focuses on providing more visualizations related to the study. Compared to our project, this article's visualizations and content puts a heavier emphasis on the timeline, whereas we use a timeline to introduce the issue, but create a visualization that is more topic- and content-centric.

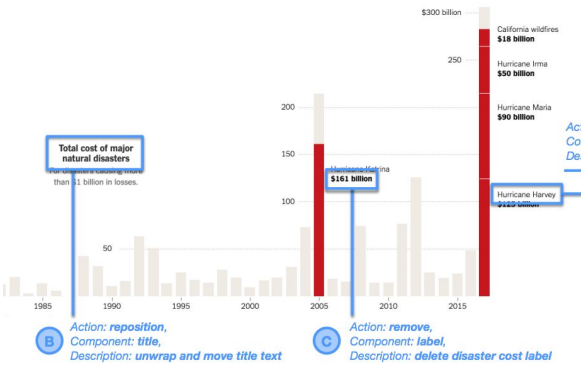
3. [Many Turn to YouTube for Children's Content, News, How-To Lessons](), by Aaron Smith, Skye Toor and Patrick van Kessel

   This study from the Pew Research Center shows that many users rely on YouTube not just for entertainment content, but also to "help[ing] them understand events that are happening in the world", and that many adult users rely on YouTube for content for children. This survey also shows how influential YouTube's recommendation algorithm can be, as about 80% of the YouTube users say they at least occasionally watch what the platform suggests in their "up next" recommendations.
   This study is relevant to our work, because it illustrates how impactful these algorithms can be to influence users who use these platforms to gain more insight into the



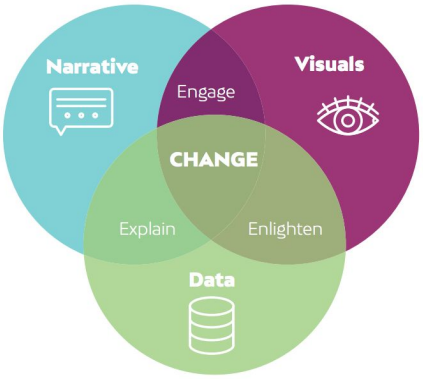
world/current events. This highlights how dangerous these algorithms can be when they continue to recommend false videos or conspiratorial videos.

4. Techniques for Flexible Responsive Visualization Design, by Jane Hoffswell, Wilmot Li, and ZhichengLiu



Techniques for Flexible Responsive Visualization Design gave us ideas and principles towards how to think about interactive visualization in relation to the html setting that we present our final project. The paper also defines five different components(reposition, remove, resize, modify, add) in interactive visualization that helps users intuitively understand the visualization. Combining the five components and other interactive skills that we learned from class, we were able to identify key interactive elements for each visualization, and further make decisions on what interactions to keep or leave out to maintain simplicity. Moreover, the paper let us think about presentation of data visualization in different screen sizes. This not only led us to decide the layout of our website in one centered column to enable the visualizations to stay as coherent as possible in different screen sizes, but also take into account where the interactive part of the visualization should be placed.

5. Data Visualization Guide by NetQuest by NetQuest



This pdf gave us guidance to putting all the visualizations together in one setting. In the pdf, the paper introduces scrollytelling, which is a type of technique where it aims to tell a story as users scroll through a graphic. Taking this concept, we decided to tell our story of conspiratorial content and YouTube progressively as users scroll down the page – revealing the overview first and drilling down to the details in later pages, ending with overall takeaways.

6. [Which Chart or Graph is Right For You?](#) By Maila Hardin, Daniel Hom, Ross Perez, and Lori Williams

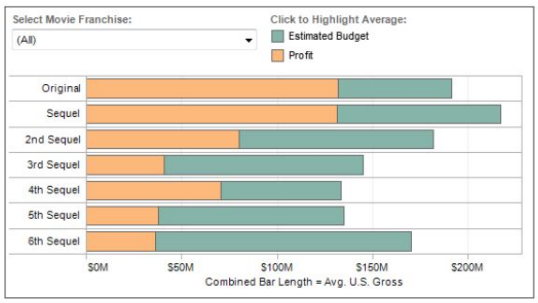
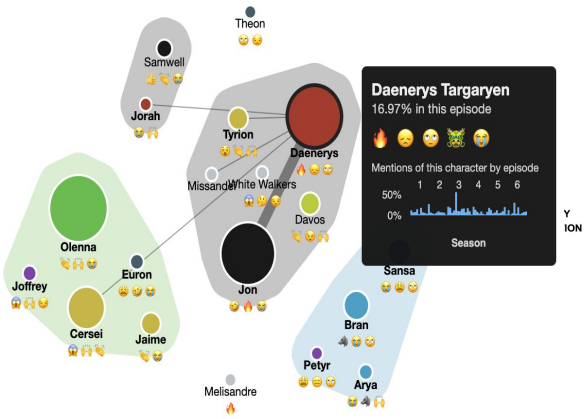
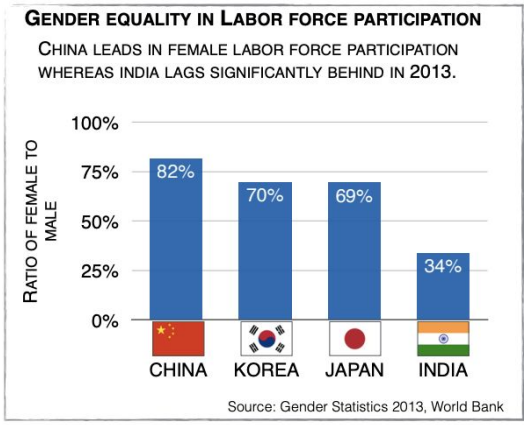


Figure 1: Tell stories with bar charts

This guideline helped us further define in detail the type of chart or graph fit for our data – especially the visualization for 'Conspiracy Videos by Scores and Topics'. Following the different types of visualization identified in the guideline, we explored multiple versions for 'Conspiracy Videos by Scores and Topics', including versions with stacked bar chart, and pie chart. When the visualization was finalized, we further finessed the looks by following the guidelines' points for consideration such as combining charts.

7. [Beyond Memorability: Visualization Recognition and Recall](#) by Michelle A. Borkin et al.





DATA & MESSAGE REDUNDANCY

The study emphasizes how the quality of visualizations not only depends on how memorable they are but also their ability to have users recognize and recall after viewing them. We wanted to ensure that our visualizations included both qualities and incorporated into our usability test and counted the number of times participants asked questions during completing tasks and the number of times participants had to go back to visualization when they were completing the knowledge test. These measures allowed us to evaluate how memorable and recallable our visualizations were. In addition, one of the examples that the authors use is by redundant encoding of data to emphasize the message (image on the right). We incorporated this quality to our tableau visualization (visualization 3) where we included additional quantitative (additional label) and qualitative (topic icon) to emphasize the visualization.

8. How every #GameOfThrones episode has been discussed on Twitter by Krist Wongsuphasat

   This visualization was presented as an example during one of the guest lecturers, Krist Wongsuphasat, on display of Game of Thrones character mentions on twitter. When each character node is hovered over, a tooltip is displayed highlighting the percentage of mentions. The nodes change based on the appearance of the character in the selected season and episode and the size of the node differs based on the percentage of mentions. This interaction became an inspiration to our final visualization on display of YouTube videos based on topic and top tags. We have also adopted the size of the nodes to be proportional to the frequency of tags and allowed users to hover over each video thumbnail to see additional information along with ability to click the thumbnails to display actual videos on YouTube website.

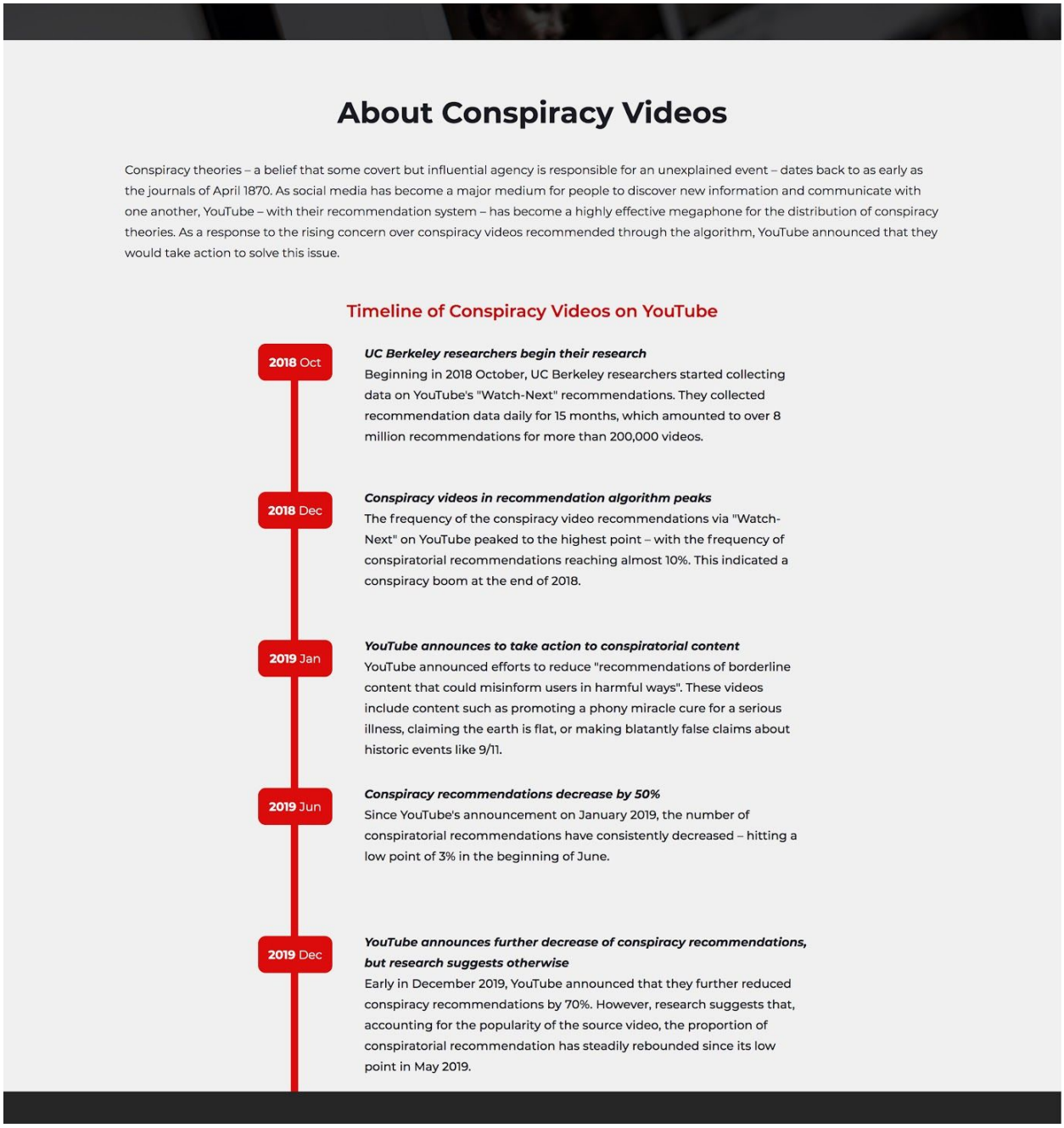
9. Eye Tracking in Retrospective Think-Aloud Usability Testing: Is There Added Value?
   By Ratma Elbabout, Obead Alhadreit, Pam Mayhew

   One challenge we came across during our project is getting substantial feedback from users during user testing. Given the circumstances of COVID-19, it was impossible for us to conduct a face-to-face usability study and further identify places where improvements could be made in detail. This paper enabled us to identify two skills – concurrent think-aloud (CTA), and retrospective think-aloud (RTA) – when conducting usability study via Zoom. CTA enabled us to identify the sections that are less intuitive to users. While conducting a knowledge test after going through the website, we asked users to conduct RTA which let us identify the types of information communicated and further come up with visualization techniques such as iconography to help users remember and understand the data communicated.

Visualization Description

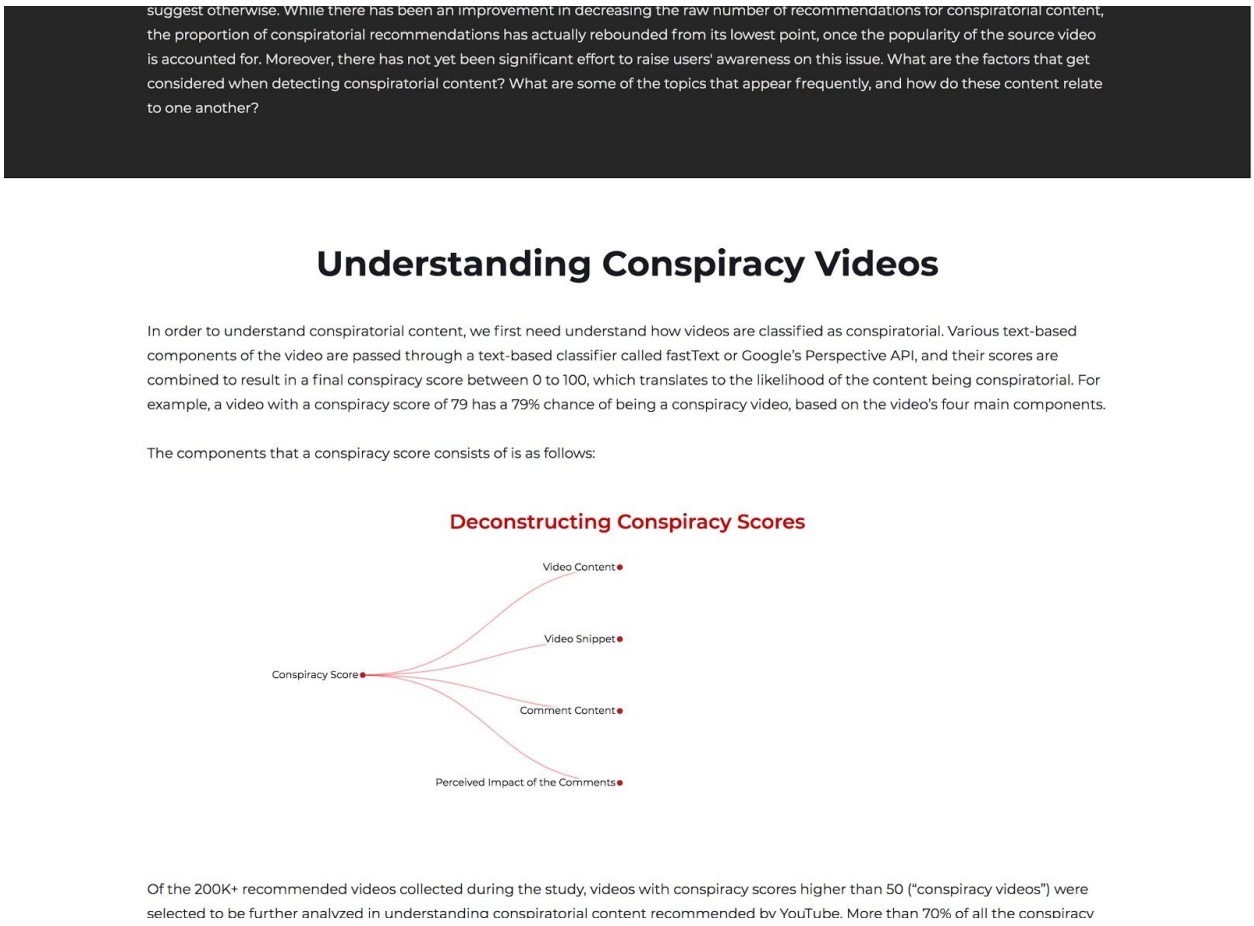
Introduction

The website starts with a catchphrase on conspiratorial content and YouTube. Short, impactful sentence with a visual was chosen to work not only as an opening that introduces the main topic, but also in an interesting way to encourage the users to scroll further down on the page and get to know more about the topic.



Timeline

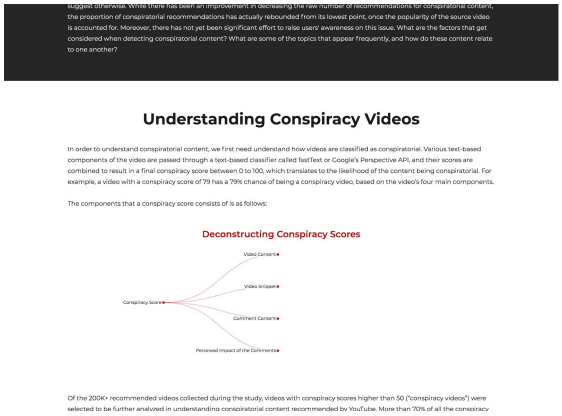
The timeline illustrates five important events related to conspiratorial content and YouTube. By providing an overview, we aim to provide users with an idea of how conspiratorial content has affected YouTube, and YouTube's response to the issue at hand.



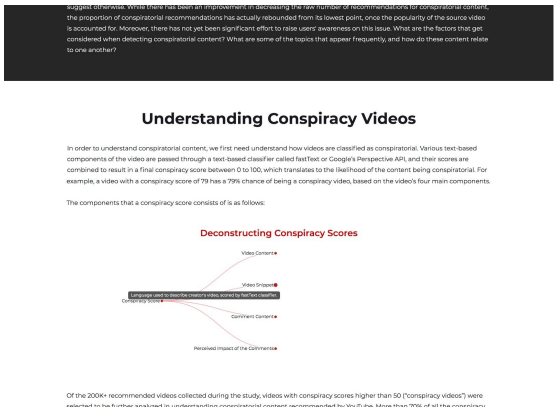
Understanding Conspiracy Scores

A video's conspiracy score (or conspiracy likelihood) - calculated by the researchers of the original study - is a key metric heavily used in both the original research and our project, as it measures how conspiratorial the video is. However, the measure is not as easy to understand or (especially) explain to a wider audience because it uses a combination of NLP and logistic classification techniques on various different components of the video.

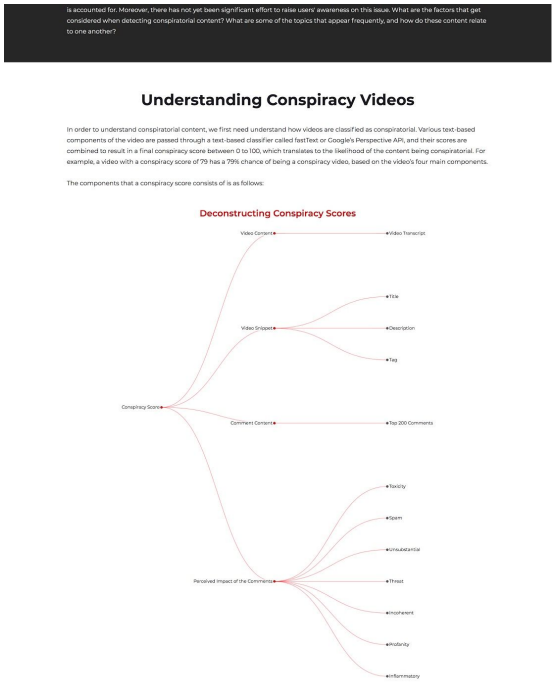
This visualization tries to show how the overall conspiracy score is constructed as simply as possible, with an emphasis on showing what components go into calculating a final likelihood score for the video. We decided that a tree diagram would be effective in illustrating the different components that come together to create an overall score.



In the default state, we show the four main text-based components that make up the overall conspiracy score. The red circles (which grows bigger when the user's mouse hovers over it) indicate that the nodes can be expanded).

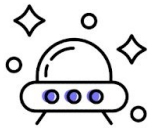


When the user hovers over the component node text, a grey textbox appears and gives a brief explanation of the component and how it is scored (and it disappears once the mouse moves away).



Once the tree is fully expanded, it displays the full list of components that are considered when calculating a video's conspiracy score. For example, a video snippet is a concatenation of the video's title, description, and tag, and the comment impact score is scored based on 7 different properties (toxicity, spam, threat, etc.).
Compared to the earlier design, we focused on making the visualization itself more intuitive and illustrative, while adding some of the deeper explanation to the text portion that introduces the visualization.

Conspiracy Topics

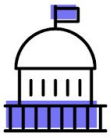


Alternative Science & History

Redefinition of the mainstream historical narrative of human civilization and development. This content uses scientific language, without the corresponding methodology, often to reach a conclusion that supports a fringe ideology less well served by facts.

Prophecies & Online Cults

Explanations of world events as prophetic, such as claims that the world is coming to an end or that natural catastrophes and political events are religious realizations. Many videos from this category intertwine religious discourse based on scripture interpretations with conspiratorial claims.



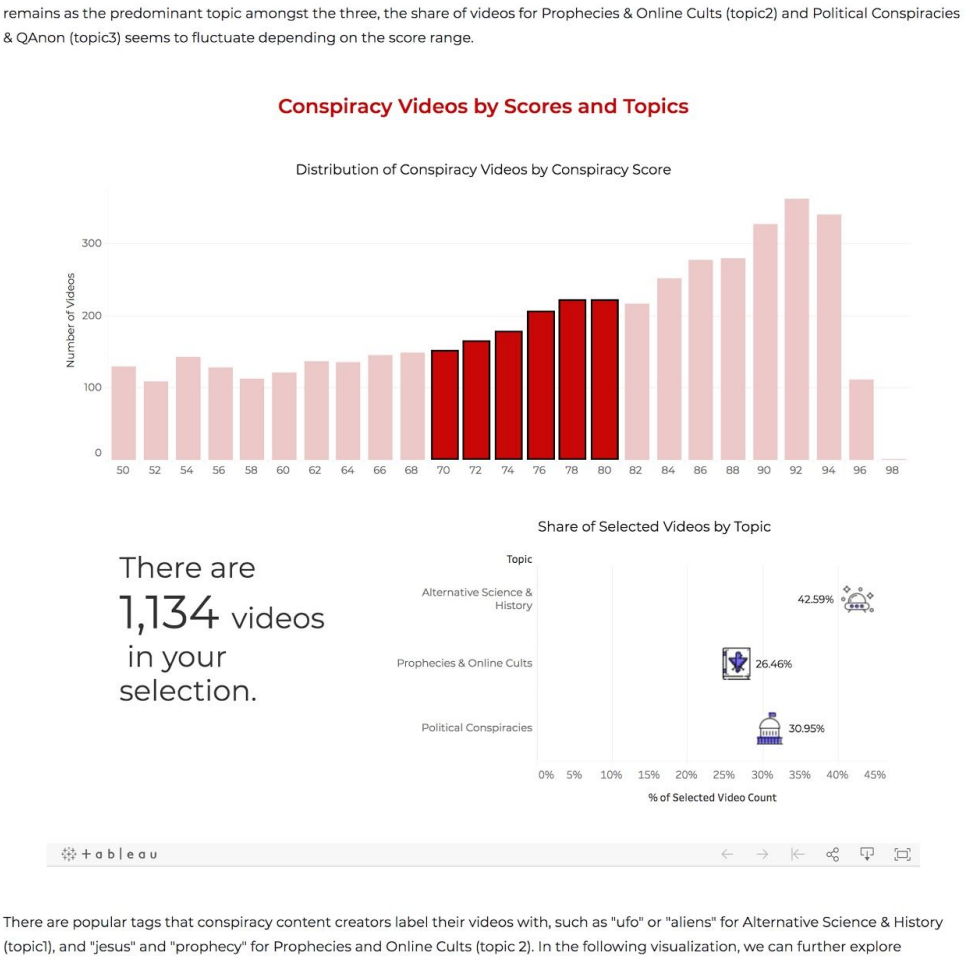
Political Conspiracies & QAnon

The third main topic is comprised of political conspiracies, the most popular of which is QAnon, a conspiracy based on aseries of ciphered revelations made on the 4chan anonymous message board by a user claiming to have access to classifiedU.S. government secrets.

One of the main drivers of data visualization for this website is the topics of conspiratorial content that YouTube videos fall into. The topics were predefined from the UC Berkeley research, where the researchers used a topic modeling technique called non-negative matrix factorization(NMF). Considering the high complexity of conspiratorial content that users are introduced with, we decided to work with the top three major topics : ‘Alternative Science and History’, ‘Prophecies and Online Cults’, and ‘Political Conspiracies and QAnon’.

This portion shows a brief explanation about the three major topics, and we used iconography to make a more visual impact that helps users remember what the topics are. This section also sets the stage for the topic-centered exploration that we lay out for the rest of the project.

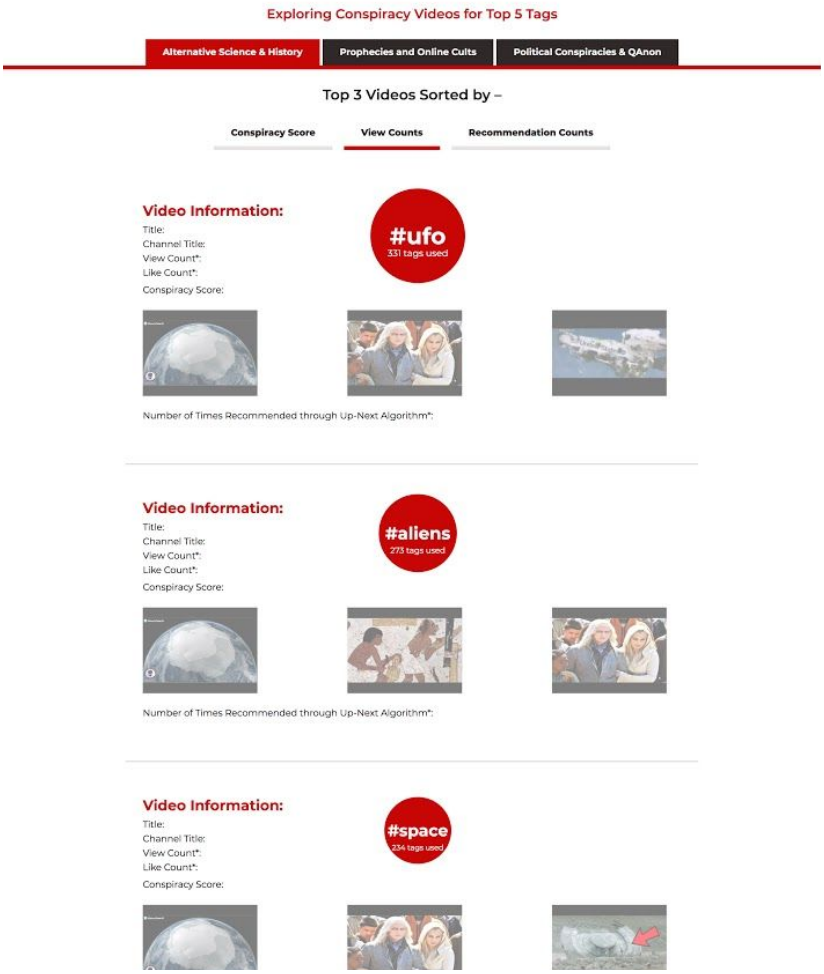
Distribution of Conspiracy Videos per Topic and Conspiracy Score



Through this interactive dashboard, we let users broadly explore the landscape of conspiratory videos. This visualization shows the overall distribution of conspiracy videos by their conspiracy score (top panel), and users can select (brush) score segments to see how the topics are distributed in the selected score range (bottom-right panel) and how many videos fall into their selection (bottom-left panel). For the bottom right panel, we used pictogram markers that are consistent with the icons from the previous section.

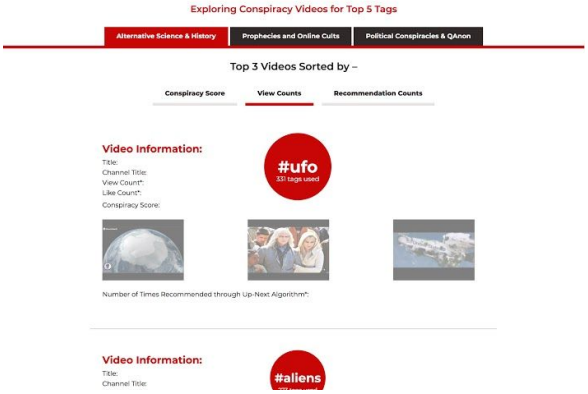
We intentionally set the default view to already have a selection, because we wanted to make sure users notice that they can interact with the visualization. If the selection is cleared, the visualization gives information on the whole sample (4806 videos).

Exploring Conspiracy Videos for Top 5 Tags

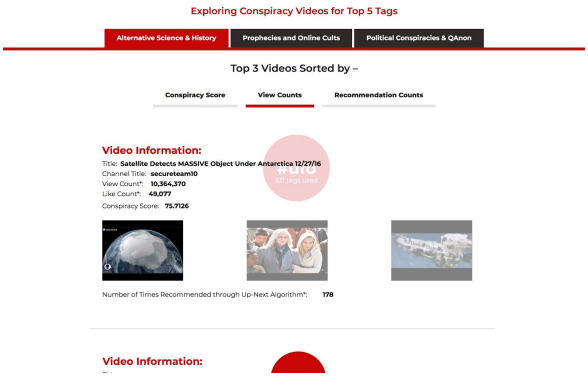


Through this interactive visualization, we allow users to explore conspiracy videos in three topics based on their top five tags. Users can look up each topic by clicking through the first set of tabs - 1) Alternative Science & History, 2) Prophecies and Online Cults, 3) Political Conspiracies and QAnon.. For example, for the "Political Conspiracies and QAnon'' tab, the top five tags are the following: q anon (or qanon), trump, economic collapse, economy, and gold.

In each tag, we presented three thumbnails to show snippets of what kind of conspiracy videos are in this topic on YouTube currently. Users can also sort the videos of each tag by conspiracy score, view counts, or recommendation counts by selecting the second set of tabs. For example, by selecting "Recommendation Counts", the thumbnails will display three videos of each tag that have the highest recommendation counts (the frequency at which the video has been recommended).

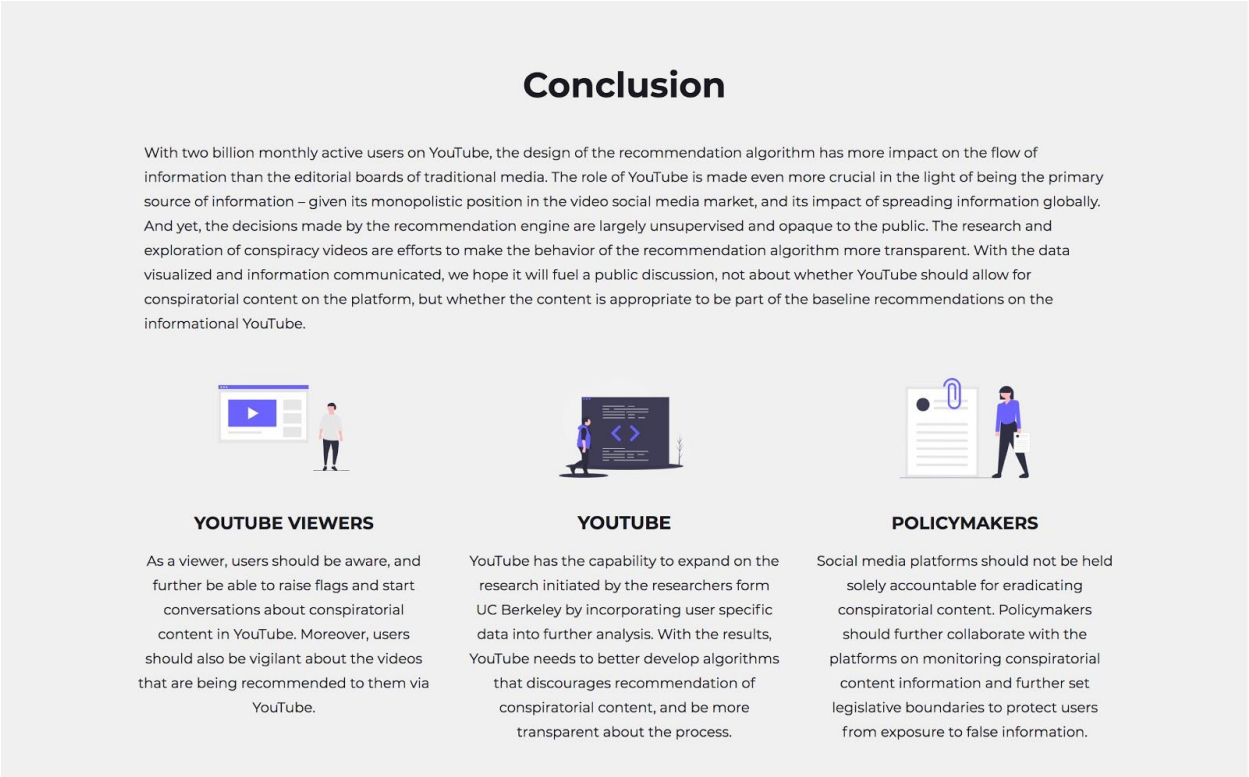


At the default state, users can see three thumbnails for all the tags and view information at a high level.



When hovered over to a thumbnail, then the tag becomes more transparent while the thumbnail image gets more opaque and video information for the thumbnail is shown in bold text in the video information section. The information contains title, channel title, view count, like count, conspiracy score, and the number of times the video was recommended through YouTube's up-next algorithm. Users can also click the thumbnail image to view the actual video on YouTube (clicking on the thumbnail redirects the users to the actual YouTube page).

Conclusion



This section helps tie all our visualization together and make the final concluding remarks. We also add more specific guidelines/recommendations for our different target audiences. To relieve cognitive overload, we decided to make the target audience recommendation into three columns and identify each of them with pictograms that resemble each target audience.

Approach

Data

We received a dataset of 6752 videos that received a conspiracy score greater than 0.50 from Professor Farid and researcher Marc Faddoul. During our initial exploratory data analysis, we found out that around 70% of the data (4608 out of 6752 videos) fell into one of three large conspiratory topics: 1. Alternative Science and History, 2. Prophecies and Online Cults, and 3. Political Conspiracies and QAnon. We decided to focus on these three topics to simplify our visualizations and show the most relevant topics to our audience.

The components of the dataset are the following:

- Index (string)
    - Unique identifier for YouTube video URL
    - *Example: CqfMDX74ks0*
- Title (string)
    - Title of the video
- Channel_title (string)
    - Title of the channel
- Reco_count (integer)
    - How many times researchers detected the video each month in the recommendation emulation
- View_count (integer)
    - Number of views in the video at the time of collection (query_date)
- Comment_count (integer)
    - Number of comments in the video at the time of collection (query_date)
- Like_count (integer)
    - Number of likes at the time of collection (query_date)
- Topic_number (integer)
    - Topic number that the video is categorized to
- Tags (list or string)
    - Tags of the video
    - *Example: [david icke, Julian Assange, WikiLeaks, JFK], "Dark Journalist|UFO|Mystery Schools"*
- Description
    - Description wrote by content creator about the video
- Conspiracy_likelihood
    - Conspiracy score calculated using inputs such as video content, video snippet, comment content, and perceived impact of the comments
    - This is essentially the calculated probability (likelihood) of each video being a conspiracy video
- Query_date
    - Date the data was queried

Tools

1. **Overall Webpage**
   Website building and hosting was done through Webflow. Using basic tools provided by Webflow, further customized code was embedded to display data visualization and make timeline visualization.
2. **Exploratory Data Analysis (EDA)**
   Importing the original .json file to pandas dataframe with Python, we have performed EDA and data cleaning. We have also performed text analysis on tags to find out which tags were the most frequently used among the three topics and sorted the videos by conspiracy scores, like counts, and view counts (descending order). Then each data frame created was converted into json files to be used in visualization 4.
   - Tools used: python, jupyter notebook
3. **Visualizations**
   a. **Visualization 1 - Timeline**
      In order to give an overview of the conspiratorial content on YouTube and its relationship to research done at UC Berkeley, we first identified five crucial dates related to the topic. After identifying the dates Illustrator was used to brainstorm different styles of timeline design. Considering the complexity of the information, the designs were simplified and focused on delivering the information in different levels(dates, titles, and deeper information). After designs were finalized, we further moved to HTML & CSS where the timeline was hardcoded to the website.
      - Tools used: Illustrator, HTML & CSS

   b. **Visualization 2 - Collapsible Tree**
      In order to give users a better understanding of the components that go into calculating an overall conspiracy score (likelihood that the video is a conspiracy video), we used D3 to create a collapsible tree. Initially (before the usability test), our tree only consisted of the 4 main components and most of the explanation was enclosed in a textbox. But after receiving feedback about the explanation being hard to digest and the visualization not fully utilizing the interactivity of a collapsible tree, we added another layer to the tree to make the score composition easier and more intuitive to understand and added a hover-text to give a brief explanation of how each of the components are scored.
      - Tools used: D3

   c. **Visualization 3 - Interactive Dashboard**
      After a little bit of further data cleaning using Excel, we used Tableau to create an interactive dashboard that helps users get a better overview of the conspiracy videos in terms of their conspiracy scores and video topic. We created a dashboard that combines 3 different sheets, but linked the sheets using the dashboard actions functionality (filtering), so that all components will reflect information about the same selected target (brushing and linking).

- Tools used: Excel, Tableau, Illustrator(topic iconography)

d. Visualization 4 - Interactive YouTube Video Display per Topic and Tag

Using the json files created based on topic and sort-by categories, we created an interactive D3 display of thumbnails of the videos and their information. We had faced unknown challenges from D3 which gave us limitations on how many items could be hovered/clicked before it froze the interaction. To address this issue and to aid the aesthetics of information presented, we incorporated webflow's multi-tab property in the framework to separate each block of visualization based on sorting into different tabs.

- Tools used: D3, webflow

4. Usability Test

We have conducted usability tests on three participants that matched our target audience. Our interviews were held and recorded through Zoom. Participants were asked to complete tasks, complete questionnaires, and provide feedback. All the contents during the interview were recorded.

- Tools used: Zoom, google sheets(raw data input, data analysis), google docs (raw data input), google forms (knowledge quiz, survey)

Discussion of Results from Usability Test

Method

For our usability test, we selected three users in the bay area that had similar characteristics as our target audiences. Our team members rotated following roles during our usability test: time keeper, note taker, participant facilitator during the interviews. We wrote a script to read during the session with our test participant (available in Appendix) that the participant facilitator read out. All contents of the interview were recorded. Participants performed the following tasks:

1) Walk through visualizations and completed four scenarios
2) Complete a knowledge test after using our system
3) Complete a multiple choice survey questionnaire
4) Provide feedback in speech after using the system about each visualizations and overall impressions

The participants performed four task scenarios for the system that are the core functionalities and visualizations of our system. The participants were asked to think aloud while performing the tasks and the notetaker wrote down the feedback. When the participant performed each task, the time keeper kept track of how much time it took the participant to complete the task. After completing all the task scenarios, participants were asked to complete a knowledge test through google forms (available in Appendix). The knowledge test consisted of five questions that tested participants' gained knowledge after interacting with the visualization on our platform. The participants were allowed to go back to the visualization to answer questions but each time going back to the visualization was noted. After completion of the test, participants completed a survey through google forms that had likert style questions asking for feedback after using the interface (also available in Appendix). Finally, participants provided feedback orally on what worked and did not work as they used the system, and the notetaker wrote it down.

Result Analysis

Results of the quantitative measures

1. *Time taken to complete each task overall*
   In order to measure exactly how long it takes for participants to complete the task on each of our visualizations and also to allow participants to freely ask questions during their exploration of our visualizations, we decided to start measuring time after participants took time to explore texts and visualizations related to the task. Below is a table of the time (in seconds) each participant took to complete a task. The time may seem short, but it is because participants already had some time to explore the visualizations.

	User 1	**User2**	**User3**	**Average**
Task1	10 seconds	20 seconds	8 seconds	12.67 seconds
Task2	5 seconds	7 seconds	5 seconds	8 seconds

Task3	5 seconds	3 seconds	7 seconds	5 seconds
Task4	3 seconds	3 seconds	4 seconds	3.33 seconds

2. *Number of clarification questions asked*
   During the time when each participant was looking at our visualizations, we counted the number of times they asked questions. This measure helped us understand how clear information presented was to the users.
   - User 1: 4 (more content-based questions)
   - User 2: 7 (more about the design and functionality)
   - User 3: 11 (more content-based questions)

3. *Knowledge test accuracy overall & number of times participants referred back to visualization to answer knowledge test overall*
   After each participant looked through the visualizations and completed the tasks, the participants completed a knowledge test which asks about key contents from the visualizations. The participants were allowed to go back to the visualization but the number of times they were going back per question was recorded. These measures allowed us to measure how recognizable and memorable information presented in the visualizations were.
   - Question 1: When did YouTube announce to take actions ending conspiracy scores? When did the YouTube conspiracy recommendation content reach its lowest?
   - Question 2: What are the components used to calculate an overall conspiracy score of YouTube videos?
   - Question 3: What is the conspiracy score range that has the most number of conspiracy videos?
   - Question 4: What is the most frequently used tag for conspiracy videos about alternative science and history?

	User 1	**User2**	**User3**	**Average**
Question 1	100%	0%, had to go back	100%	67%
Question 2	100%	0%, had to go back	100%, had to go back	67%
Question 3	100%, had to go back	100%	100%	100%
Question 4	100%	100%	100%	100%

4. *Answers to Likert scale survey questions (average score)*
   The participants answered the following survey questions after they looked at our visualizations. The questions below were presented as multiple choice likert style questions.
   - Question 1: How aware were you about conspiracy videos before taking this usability test?
   - Question 2: Was the information communicated through TIMELINE data visualization clear to understand?
   - Question 3: Was the information communicated through CONSPIRACY SCORE COMPONENTS data visualization clear to understand?
   - Question 4: Was the information communicated through SHARE OF VIDEOS BY CONSPIRACY SCORE data visualization clear to understand?
   - Question 5: Was the information communicated through EXPLORATION OF CONSPIRACY VIDEOS BY TOPIC data visualization clear to understand?
   - Question 6: Compared to before interacting with this visualization, I am ____ about being able to identify conspiracy videos. (much less confident - much more confident)

	User 1	**User2**	**User3**	**Average**
Question 1	4	1	5	4.33
Question 2	5	5	4	4.67
Question 3	5	2	3	3.33
Question 4	4	5	4	4.33
Question 5	5	5	4	4.67
Question 6	4	4	4	4 - more confident

Results of qualitative measures

- Overall Impression
  - Users rated the website to be generally intuitive and easy to understand. The introductory and conclusion section of the site seemed to drive the storytelling element fluently.
  - User3 expressed that it would be better to have a clearer conclusion to targeted users.
  - Two of the three users expressed cognitive overload while going through the first few parts of the 'Study on Conspiracy Videos' section.
- Comments on individual visualizations
  - Visualization 1 (timeline)
    - Users found the layout very clear and appreciated the bold fonted headlines

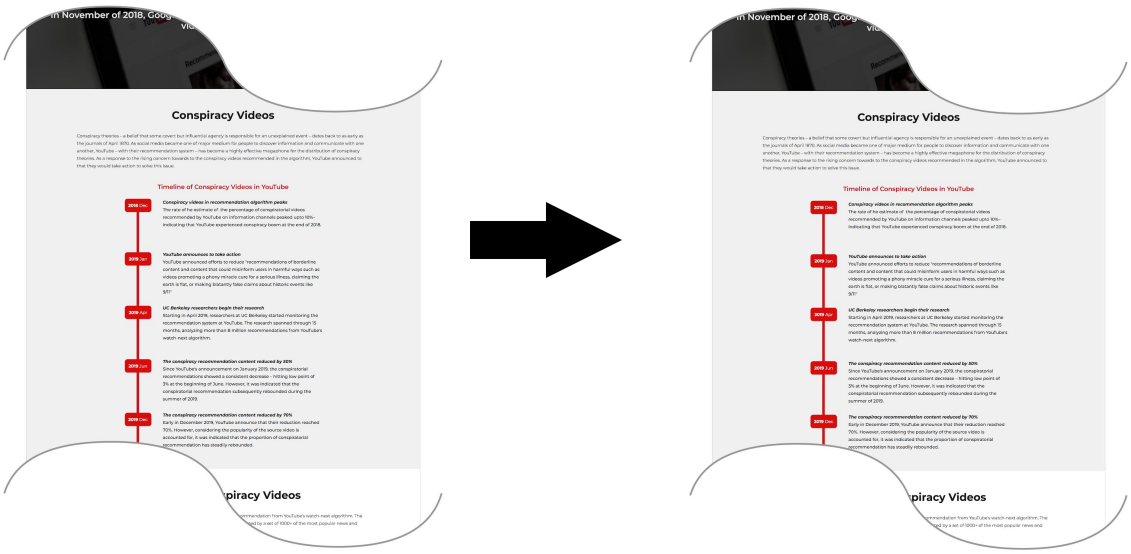
        - There was a comment about how the timeline is text heavy.
        - There was a suggestion that it might be helpful to space out the time bullets by the actual time gap.
        - There was also a suggestion that it might be helpful to make the timeline horizontal
    - Visualization 2 (collapsible tree diagram)
        - Users found that the tree was text heavy, and that it takes a lot of concentration to read in all the information.
        - There was a comment on how the texts seem cluttered.
        - There was some confusion about what the visualization was explaining about the conspiracy score.
        - There was a suggestion that it might be useful to put the explanation text in bullet point format for legibility.
        - There was positive feedback about how the structure of the tree makes the relationship between the left and right nodes clear.
        - One of the users pointed out that they would like more interactivity from the visualization
    - Visualization 3 (stacked bar chart)
        - Some of the users were confused about the toggles (score buckets), and didn't immediately find how to interact with the visualization.
        - One of the users suggested that it would be helpful to distinguish between the text on the left and the information from the bar chart, and to make the color legend more visible.
        - One of the users didn't find the use of the toggle tool right away, but figured it out by themselves and commented that they find the functionality useful.
    - Visualization 4 (exploration tool)
        - Two out of three of the users found a bug with the hovering tooltip (the thumbnails sometimes start flickering for the bottom two tags when the users hover over the node).
        - One of the users asked if they can click on the thumbnails.
        - One of the users noticed that the tag nodes had different sizes, proportional to how many times the tag was used in the topic, which is consistent with our intention.
        - Users expressed interest in finding out more about the videos that were shown
        - Two out of three of the users said that it might be interesting to see more tags than the three we have in our visualization.

Feedback Incorporation

- **Visualization 1**

Incorporating the feedback that the timeline is text heavy, we edited the text and title of each timeblocks to be more concise and clear than the previous version. There was a feedback that it was hard to identify the length of time between the dates. While visually expressing the length of time between dates seemed desirable, we decided to continue with the current design and keep the timeline simple – putting more focus on communicating detailed content related to conspiratorial videos and simplicity of design.

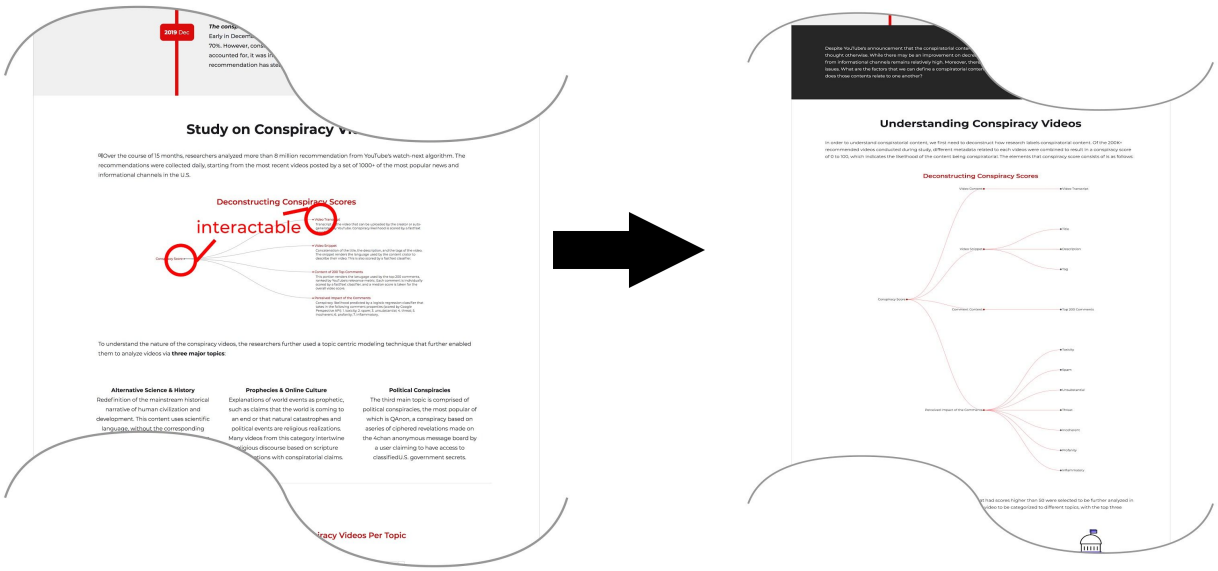
Before: *After:*



- **Visualization 2**

  Initially, we created a collapsible tree that only had 2 layers (1 parent node, 4 child nodes) with a textbox describing each component. However, after receiving feedback from our usability test that mainly said the textboxes contained too much information and was hard to read, we added another layer to the tree to make the score composition easier and more intuitive to understand and added a hover-text to give a brief explanation of how each of the components are scored.

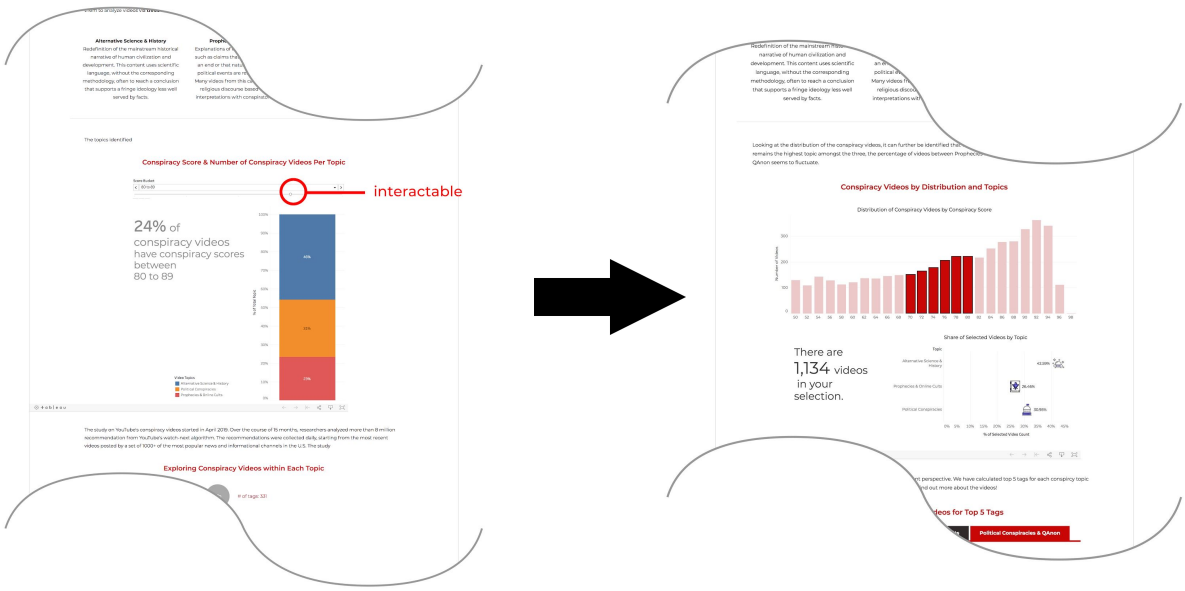
Before: *After:*



- **Visualization 3**

  After receiving feedback from usability study participants and the professor, we made a drastic redo of the visualization to more effectively achieve our goal which is to show a big overarching view of our data and then allow for more granular exploration based on the three topics and the conspiracy scores. We first created a histogram to show the overall picture first (distribution by conspiracy score) and emphasized the interactiveness of the visualization. Once users brush sections of the histogram that they would like to explore, the linked dot chart and text reflects information about that selected subsection of videos (linking). For the dot chart, we replaced the dots with iconography (consistent with website images) to enhance memorability of the diagram.

  Before: *After:*



- **Visualization 4**

  After receiving feedback from the usability study, we made various changes to the visualization to enhance the interactiveness and provide broader facets of information on conspiratorial videos. Instead of users hovering only to find video thumbnails and its title, we allowed users to hover each of the videos to find more detailed information such as view counts and recommendation counts and made thumbnail images clickable - clicking on the thumbnail redirects users to the original videos on YouTube. Additionally, we added additional filters to sort videos not only by conspiracy score but also by view count and recommendation counts. As a result, the visualization became a lot more interactive and interesting and provided more information about the contents.

Before:

After:

Visualization Link

[Link to our visualization]

Allocation of Project Tasks

The table below shows a breakdown of the task allocated to each team member, and the percentages are an estimate of the members' contribution to each task. However, these numbers are not fully representative of the overall work that every member put into the project. A significant portion of the work involved a lot of design and idea brainstorming, which we always did together. We believe that all team members put in an equal amount of effort into the project.

Project Component	Task	Grace	Chloe	Lynn
Data Preparation	Data Sourcing	33%	33%	33%
	Data Preprocessing	50%	50%	0%
	EDA	33%	33%	33%
Visualizations	1. Timeline	0%	0%	100%
	2. Collapsible Tree on Conspiracy Score	0%	100%	0%
	3. Stacked Bar Chart on Share of Videos by Conspiracy Score	0%	100%	0%
	4. Interactive Diagram on Topics and Tags	100%	0%	0%
Design	Website Text Writeup	10%	20%	70%
	Website Layout Design	0%	0%	100%
User testing and Others	Final Presentation	33%	33%	33%
	User Testing	33%	33%	33%
	Report Writing	33%	33%	33%

Appendix

Code Reference

Visualization 2:

Adapted code from:
https://observablehq.com/@d3/collapsible-tree (collapsible tree)
http://bl.ocks.org/PandaRider/84b9c096e37334c2a4fd945a66b8ccf9 (hovering textbox)

Visualization4:
Adapted code from:
https://observablehq.com/d/dbd32de2a71baffa (previous D3 assignment)
https://stackoverflow.com/questions/37722130/d3-js-opening-a-new-tab-by-dbclick-event (opening new tab by click)
https://fabiofranchino.com/blog/how-to-load-image-in-svg-with-d3js/ (uploading image files)

Usability Test Questions

- **Knowledge Test**
  - Q1. When did YouTube announce to take actions ending conspiracy scores? When did the YouTube conspiracy recommendation content reach its lowest?
    A: 2019 Jan, 2019 Dec
  - Q2. What are the components used to calculate an overall conspiracy score of YouTube videos? (select all that apply)
    a. Number of comments and likes
    b. Perceived Impact of Comments
    c. Content of top comments
    d. The title, description, and tags of the video
    e. Number of views
    f. Video Transcript
  - Q3. What is the conspiracy score range that has the most number of conspiracy videos?
    a. 50-59
    b. 60-69
    c. 70-79
    d. 80-89
    e. 90-99
  - Q4. What is the most frequently used tag for conspiracy videos about alternative science and history?
    A: UFO

- **Survey Questionnaire**
    - Q1. How aware were you about conspiracy videos before taking this usability test? (Not aware – Very aware)
    - Q2. Was the information communicated through data visualization clear to understand? (Not clear – Very clear)
        i. Visualization 1
        ii. Visualization 2
        iii. Visualization 3
        iv. Visualization 4
    - Compared to before interacting with this visualization, I am _____ about being able to identify conspiracy videos. (Much less confident - Very confident)

Usability Test Script

Demo script:
Interviewer: Hi [Participant Name]! It's nice to meet you. Thank you for making the time to speak with me today about understanding Youtube's Conspiracy Video. [Introductions to other team members]

Before we begin, I'm going to give you a brief overview of the test and how it will work. During this session, we'll be assessing the usability of different data visualizations within our Explaining Conspiracy Video within Youtube project. I'll give you four tasks to complete, and I might ask questions as we go along. Before I tell you the task, I'll give you a bit of context behind it, such as why you might be doing it and what you hope to achieve. After completing the tasks, you will be given a brief knowledge check about the visualizations and survey asking for your feedback on your experience interacting with our platform.

It's really important to know that we are only testing the visualization, not you. You can't do or say anything wrong here. Please feel free to let me know at any time if there's something you like, dislike, if you're confused, etc.

Also, I'd like you to "think aloud" as much as possible. By that, I mean that I'd like you to speak your thoughts as often as you can. For example, you may be looking at a page, suddenly see something you didn't see before and want to click on it. In that case, saying something like "this caught my eye so I'm going to see what it is" would be very useful.

If at any point you have questions, please don't hesitate to ask. We will also be recording this session and it may be shared with the class instructors. Do we have your permission to record?

Do you have any other questions?

Ok, Let's get started.

Begin Recording

Warm Up

Before we dive into our visualization, could you tell me a little bit about your previous experience with Youtube – how often you use it and the things you watch?

Have you heard about conspiracy videos within Youtube/have you ever come across one of them?

Begin Screen Sharing

Task Completion

Now I'm going to give you four tasks as you look through each of the visualizations. Each task will correspond to each of the visualization sections we have on our website. For each task, I'll ask you to go over the visualization first, thinking out loud. After you are done exploring the visualization, we will ask you a few questions regarding the visualization in order to identify the clarity of the visualization. .

Task 1
The first task is to help us understand if the timeline of Youtube Conspiracy Video is intuitive and easy to understand. Can you first look at the timeline and think out loud about the information that you notice?
(user goes through the timeline)
Now that you've looked at the timeline, can you describe when the conspiracy video in youtube peaked, and when youtube decided to address this issue?
(time keeper tracks time it takes the participant to complete the task)

Task 2
The second task is to ensure that users can clearly understand how conspiracy scores from our analysis are calculated. Can you take a look at the tree diagram on the conspiracy score and think out loud about the information that you notice?
(user goes through the diagram)
Now that you've looked at the diagram, can you describe which factors are considered in calculating how conspiratory each video is?
(time keeper tracks time it takes the participant to complete the task)

Task 3
The third task is to help us understand if the bar chart is easy to understand the proportion of videos among top three topics depending on the conspiracy score for users. Can you take a look at the stacked bar chart and think out loud about the information that you notice?
(user goes through the diagram)

Now that you've looked at the diagram, can you find out which topic among the three has the highest percentage of conspiracy videos out of all videos that received a score ranging from 80 to 89?
(time keeper tracks time it takes the participant to complete the task)

Task 4
The last task is to help us understand if our diagram thumbnails of top three conspiratory tags are easy to interact with and the information provided from the diagram is easy to understand. Can you first look at the diagram and think out loud about the information that you notice?
(user goes through the timeline)
Now that you've looked at the diagram, can you tell me which tag appears the most in the "Alternative Science and History" topic and describe the top videos under that tag?
(time keeper tracks time it takes the participant to complete the task)

Great, we're done with the task portion of the interview. Now we will move on to the knowledge test portion. Here, you will be given a quick knowledge test that asks for few of the components in the visualization. You are welcome to go back to the visualization and look for answers, but please tell us whenever you have to go back to the visualizations. I just sent you the link on the chat. Please let us know at any point in time when you have any questions or comments.
(User completes the test)

Next, we will move on to the survey portion of our interview. You will be asked a few multiple choice questions about your general opinions. I just sent you the link on the chat. Please let us know at any point in time when you have any questions or comments.
(User completes the test)

Finally, we would like to know your feedback and opinions on our visualizations.

Is there any portion that you liked or disliked?

Any comments on the timeline visualization?

Any comments on the conspiracy score tree diagram?

Any comments on the stacked bar chart on the percentage of conspiracy videos per topic?

Any comments on the interaction diagram?

What was your overall impression while going through the visualizations?

Great, we're done with the test for today.

Thanks again for your time.