**Final Project Prototype**

**Project Description**

*YT* is a data visualization video project about *YouTube* culture. Through this project, I want to explore its identity and discover how and why this platform resists and adapts to the changes and trends of the digital culture over the past 18 years. My initial idea was to display popular and niche *YouTube* videos of the past 18 years that are data-generated on a webpage. The data was supposed to be *YouTube* videos and comments from various channels from 2005 to 2023. I wanted to use pre-existing *YouTube* datasets that cover the years between 2005 (i.e., *YouTube*’s first year of existence) and 2023 to create a randomly generated list of videos. The page would generate new videos after a delay. The usual thumbnail icons would display videos instead of their usual images. Generated comments will be displayed. If the video does not interest the user, they could generate new videos by clicking the “skip” button. The skipping button would be used to create apophenia, which is the perception of patterns within random data. If it is pressed, the comments will start floating around on the webpage. Each time the user clicks on this button, more floating comments will be generated, and even generated videos will start floating around on the webpage. Comments and images will blend to create an abstract representation of the platform. The web page design was supposed to be like *YouTube*’s video layout. I wanted to use video and web development for this project because it represents *YouTube*’s essence. *Node.js,* *React.js*, *npm* packages like *node-ytdl-core* or *react-youtube, HTML* and *CSS* were the programming methods planned in the initial idea because I am new to data collection and its way of collecting it through programming, I wanted to use something that I was more familiar with. The goal was to loop sequences of generated *YouTube* videos on a webpage, create a skip interactive button and generate comments and videos on the page.

After a discussion with Sabine, the data visualization of the project changed. The design will look more like Mary Flanagan’s *pile of secrets*. *pile of secrets* is a series that displays video game footage on LCD screens. Flanagan collected video game footage from 1980 to 2011 to answer the question: “What makes a game a game?” The artist gathered common elements in video games into her database. So, instead of displaying videos on a webpage, the *YouTube* data will be displayed on multiple screens in a room. Each screen in this installation will be associated with a specific year. The videos will loop on the screens.

At first, I wanted to display the videos by looping them one after the other. But while pulling the data from their datasets, I thought of testing layering the videos like Jason Salavon’s *All the Ways (The Simpsons)*. Thisseries gathers all the frames of every episode of *The Simpsons*’ 26 seasons into a single episode. I wanted to layer all the videos like in Salavon’s project. However, while doing the video editing, I noticed that the various lengths of the videos created some gaps in the layering at the end of the dataset video. Instead, I balanced the number of layers on a time frame of approximately five minutes, so the current result of the videos is closer to the aesthetic found in another inspirational data project, *Particle*. *Particle* is a processed urban imagery that fluctuates between recognizable landscapes and abstract data-like patterns combined with dense sound textures. It deconstructs visual and audible urban landscapes. The images and sounds have been broken into fragments and then reconfigured.

**Artistic Process**

My goal in this prototype was to select specific years, find datasets that give results fitting to these years, and download their associated videos. I decided to use a five-year gap, resulting in four collections. The years selected were 2005, 2010, 2015, and 2020. Each year, a total of ten videos are collected. All datasets used for this prototype come from the platform *Kaggle*. For the year 2005, I used the dataset *Youtube Oldest Videos (2005) Dataset*, which is a dataset gathering videos released during the first year of the platform. For 2010, I used the *All PewDiePie Videos* and *Youtube videos having more than 1 Billion views* datasets. *All PewDiePie Videos* is a dataset gathering all the videos released by the famous YouTuber PewDiePie. *Youtube videos having more than 1 Billion views* is a dataset gathering all the videos that reached over one billion views on *YouTube*. For 2015, I used the *Top 14 Ever Most Viewed YouTube Videos* and *Youtube videos having more than 1 Billion views* datasets. The *Top 14 Ever Most Viewed YouTube Videos* is a dataset gathering the 14 most-ever watched videos on YouTube. For 2020, I used the *Youtube videos having more than 1 Billion views, Cyberpunk 2077 YouTube Reception* and *Data of YouTube Videos*. *Cyberpunk 2077 YouTube Reception* is a dataset that collected all the reactions related to the controversial release of the video game *Cyberpunk 2077* in 2020. *Data of YouTube Videos* is a dataset collecting videos over the year 2020. For the code, I have reused what I have built up for the first assignment of this course to apply to this project. After collecting the data, I downloaded the collected videos and created a data video data visualization through video editing. I am currently satisfied with the resulting data visualization.

The steps missing in the processing are creating more yearly videos (i.e., the best would be to create all the years from 2005 to 2023), the number of screens to use to display the videos, and how to configure the space with the screens, and find the equipment for the installation.

**Images/Screenshots**

Detailed images/diagrams of the overall system (i.e. how data flows between the various components)

**Code**

The code was used to collect the data but not to visualize it; the datasets served me to discover the most influential or popular videos over the selected years to download them and create my data visualization. I created a code prototype through the first assignment to collect video data I reused for the final project prototype. However, different functions were used depending on the dataset format or what I was looking for within a dataset. I have used the *.aggregate()*, *.sort()*, *.find()*, *.toArray()* and *.limit()* functions. For the 2005 dataset, I looked for videos released in 2005 that now have views over 32,500,000. For 2010, I created two *js* files, one for each dataset used for that year. The first one was to pull data from videos released in 2010 with over one billion views. I used the find, sort, and limit functions with this dataset. The second one served to pull PewDiePie videos released in 2010 (i.e., the year his channel was created). I used the aggregation function for this dataset. For 2015, two *js* files were created. The first is pulling the list of the most viewed videos on *YouTube*. The second one is pulling videos released in 2015 with over a billion views. Three datasets have been pulled for the year 2020. The first one is pulling a list of videos released in 2020 that have more than a billion views. The second pulls a list of the most viewed videos in 2020. The third one pulls a list of the most popular videos related to the release of the video game *CyberPunk 2077* in 2020. The results from the code are the videos that I’ve used for the data visualization. The only exception to the rule is the music videos of Justin Bieber from the 2010 video because only his title, *Baby,* was part of the results. Still, due to the phenomenon he brought in 2010, I decided to put his most viewed videos in 2010 in the data visualization.

**Challenges & Issues**

The most challenging part of the project is finding the right datasets or the right way to pull the data from them to get the results needed for the project. Most datasets I have found so far do not have a column within their collections telling just the year of video creation. It often provides the very specific date and hour the video was published. This accuracy is not helpful to pull a group of videos because the code requires a specific date to pull its data. Otherwise, I receive no results in the terminal. Suppose I can’t find a dataset describing the year of publication. In that case, I must pull data by the highest number of views or find datasets of a subject that happened specifically during the selected year.

Regarding data visualization, I’ll need to find a way to make the video layers opaque without making them hide each other. For now, I feel the opacity is still low and makes the videos a bit too dark to watch, but it is very hard to intensify their opacity without hiding the other layers. I tested some colour editing, but so far, I barely see any improvement by doing that.

Another challenge was downloading good-quality videos for the data. *YouTube* restricted the downloads of videos by adding their paid subscription to *YouTube Premium*. For the sake of the project, I decided to get a free month's trial. However, I got fooled by the platform; the downloading function is not allowing you to download the videos on your local computer, it stays on the platform and can only be watched offline. I tried to find an alternative to download the videos, and I found *TarTube*, a GUI front-end for *youtube-dl* and *yt-dlp*. This app allows you to download videos with the quality of your choice. This GUI was very useful, and I could download the videos I needed for my data collection and video editing.

**Bibliography**

“D-Fuse - About.” Accessed September 30, 2023. http://www.dfuse.com/about.html.

“Jason Salavon.” Accessed September 30, 2023. http://salavon.com/page/biography/.

“Jason Salavon | All the Ways (The Simpsons).” Accessed September 30, 2023. http://salavon.com/work/all-the-ways-video/.

*JASON SALAVON - ALL THE WAYS Feb 25 - Apr 9, 2016 @ Mark Moore Gallery*, 2016. https://www.youtube.com/watch?v=ja1c1HbQdKQ.

Mary Flanagan. “About - Mary Flanagan -.” Accessed September 30, 2023. https://maryflanagan.com/about/.

“MOST LIKED COMMENTS ON YOUTUBE.” Accessed September 30, 2023. https://www.kaggle.com/datasets/nipunarora8/most-liked-comments-on-youtube.

npm. “React-Youtube,” November 22, 2022. https://www.npmjs.com/package/react-youtube.

npm. “Ytdl-Core,” July 14, 2023. https://www.npmjs.com/package/ytdl-core.

“Particle.” Accessed September 30, 2023. http://www.dfuse.com/particle.html.

Statista. “Biggest Social Media Platforms 2023.” Accessed September 30, 2023. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Statista. “Global Top Websites by Monthly Visits 2022.” Accessed September 30, 2023. https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/.

user, mf default. “[Pile of Secrets].” *Mary Flanagan* (blog), July 3, 2011. https://maryflanagan.com/pile-of-secrets/.

“YouTube Age Demographics [Updated Aug 2023] | Oberlo.” Accessed September 30, 2023. https://www.oberlo.com/statistics/youtube-age-demographics.

“Youtube Videos Dataset (~3400 Videos).” Accessed September 30, 2023. https://www.kaggle.com/datasets/rajatrc1705/youtube-videos-dataset.