# Activity Breakdown

Lynda Negron, Jordan Dominguez, and Brandon Tineo

This project analyzes and breakdowns YouTube activity by major city. Within our visualization, we plot the biggest cities in the world on a map, showing the amount of videos uploaded within that city by major category. We also show how popular those videos are, and are thus able which cities provide the most popular videos of each YouTube category.

# YouTube Activity Breakdown

# PROCESS BOOK

# INITIAL STEPS

Project proposal, initial sketches, first design studio

Since its creation in 2005, Youtube has grown quite a bit in terms of popularity. Currently, Youtube claims that roughly 60 hours of new videos are uploaded to the site every single minute and boasts 800 million unique users a month. Given that the Youtube interface can be accessed in 61 different languages, it comes at no surprise that three quarters of the material uploaded to Youtube every single day comes from outside the United States.

However, culture within the United States and culture across different parts of the world is extremely different and thus we can expect that way in which users use Youtube is different from one place to another. For example, while one country might value humorous videos a lot, another country might value sports videos more. Value in this case can be measured in a variety of ways (i.e view counts, rating…). Taking all of this into consideration is what initially sparked out interest in this project. Further motivation comes from the fact that three of us are very active users of Youtube and are well aware of the plethora of information that can be accessed from the Youtube API.

Finally, we think this project is extremely interesting because of how widely applicable our visualizations will be. Up to date, it might seem surprising, but there are no effective visualizations of the way in which Youtube users use the website, or what they find appealing. Our visualization can be useful to the every day Youtube user, as well as to companies/organizations looking to effectively target information to certain groups of people.

# Background and Motivation

- We want to visualize how Youtube use differs around the world, including which parts of the world have the most viewers and what genre of videos are most popular.

# Initial Objectives

Chloropleth map to visualize amount of videos uploaded total videos of selected category in every country

# First Sketches

These sketches would provide alternative data regarding regarding the information. One will be a bar chart that, on a mouseover of the country, would show the number of videos in each category. Another, on the side, would change when the map data changed, showing which countries had the most amount of videos or watch counts per videos within that category.

# First Sketches

- We soon found out that the YouTube API (at least, the updated version of it) would only provide which *city* these videos were uploaded in rather than the country.

# Problems

# CHANGE OF DIRECTION

- Instead of creating a chloropleth map, we decided to plot the major cities of the world on a world map and chloropleth color & base radius on amount of videos uploaded per city based on category.

# New Data?

# Design Studio Feedback

- The feedback that we received during the design studio was extremely helpful. Some of the concerns raised by the group we discussed our project with were:
  - Cities chosen:
    - Cities chosen (top 100) might lead to overly concentrated areas and overly empty areas
    - Dots for each city need to be carefully sized as to not get too messy
    - Map more cities than just the top 100
  - Categories chosen:
    - Good idea to structure our visualization around categories
    - Category information most likely will appeal to advertisers and thus providing information on the demographics of these cities and the categories these cities enjoy most is key.
  - Visualization:
    - Probably best to not color our city points chloroplethically because points are too small.
    - because the points weren't
- New ideas we got from the design studio included
  - Adding both a side visualization and a tooltip.
  - Showing information that supplements our understanding to the data we would show on the map.

# DATA COLLECTION

Data scraping, cleaning, aggregation, API calls methods etc.

- Our first question was which cities we wanted to look at YouTube information from.
- We decided to do some research on the top 100 cities in terms of population and scrape their longitude and latitude using JavaScript.

**The Biggest cities of the World (TOP 1000 cities)**

| N | Name | Population | Altitude | Country | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 1 | Tokyo : | 31 480 498 | 20 m | JP | 35°41'6.0" N | 139°45'3.6" E |
| 2 | Beijing Shi : | 14 933 274 | 59 m | CN | 39°55'44.0" N | 116°23'16.8" E |
| 3 | Shanghai : | 14 608 512 | 6 m | CN | 31°0'18.0" N | 121°24'32.4" E |
| 4 | Buenos Aires : | 13 076 300 | 14 m | AR | 34°35'15.0" S | 58°40'21.0" W |
| 5 | Bombay : | 12 692 717 | 35 m | IN | 18°58'30.0" N | 72°49'32.9" E |
| 6 | Karachi : | 11 627 378 | 14 m | PK | 24°52'0.1" N | 67°2'60.0" E |
| 7 | Constantinople : | 11 581 707 | 157 m | TR | 41°1'7.0" N | 28°57'52.9" E |
| 8 | Mexico (México) : | 11 285 654 | 2 224 m | MX | 19°23'60.0" N | 99°2'60.0" W |
| 9 | Delhi : | 10 928 270 | 215 m | IN | 28°40'0.1" N | 77°13'0.1" E |
| 10 | Manila : | 10 443 877 | 19 m | PH | 14°36'15.1" N | 120°58'55.2" E |
| 11 | Moscow : | 10 381 288 | 151 m | RU | 55°45'7.9" N | 37°36'56.2" E |
| 12 | Seoul : | 10 323 448 | 46 m | KR | 37°33'59.0" N | 127°0'0.0" E |
| 13 | Sao Paulo (S?o Paulo) : | 10 021 437 | 631 m | BR | 23°31'59.9" S | 46°37'0.1" W |
| 14 | Istanbul : | 9 797 536 | 28 m | TR | 41°1'7.0" N | 28°57'52.9" E |
| 15 | Lagos : | 8 789 133 | 39 m | NG | 6°27'11.0" N | 3°23'45.0" E |
| 16 | Mexico : | 8 720 916 | 2 224 m | MX | 19°26'3.1" N | 99°8'19.0" W |

# City Data Collection

- Scraping the city data was only the first step in getting a usable city dataset to use in our project.
- In wrangling our city we had to make sure our latitudes and longitudes were formatted the correct way so that they matched the expected input of our map. This involved splitting the latitudes and longitudes returned by the scrape and returns integers that accounted for whether they were North, South, West or East.
- Our final step in wrangling our city data was to create a CSV file and a JSON file that could be easily accessed by other parts of our program.

# City Data Wrangling

# YouTube API – Exploration Phase

- We started by exploring a variety of the information that the YouTube API could return to us.
- As previously mentioned we found that the version of the API that would be most suitable for us to use (Version 2) would not allow us to collect information about countries.
- At this point however we had collected all of the city data we needed and so we were ready to look into the kinds API calls that would return city information

- We quickly realized that it would be best for us to select a few categories to work with.
- In selecting categories we visited the YouTube.com actual page and looked through which categories seemed to be most popular. It came to our attention that YouTube consistently placed the following categories at the top of their suggested categories: Comedy, Sports, Movies, Music, and Gaming.

# YouTube API – Selecting Categories

# YouTube API – API Calls + Data Wrangling

- One we had selected our categories we moved on to actually making our API calls and building an object. This step was particularly cumbersome given the asynchronous nature of the various YouTube API calls we were making.
- To solve this issue we proceeded to wrap the API calls within a function that would allow us to make sure every API call was completely done before moving on to the rest of the code.
- The YouTube API returns XML which required us to parse through the different tags, ids and attributes to gather the information we wanted.
- After exploring the <feed></feed> returned by each API call we proceeded to clean and merge our data into an object. Our object is currently sorted by categories since our visualization will allow user to change what they see based on what category they select.
- One final problem we encountered was handling the scope of our object.

# Implementing Visualization – Initial Steps

- We first started by plotting a map of the world by following the way we did this in the extra credit of the last homework. Given the amount of data we were plotting we wanted to add a little more functionality to our map than we did in the extra credit of HW4. The first of these added functionalities were geometric zoom (to allow users to zoom into particular parts of the map), and tooltips (to provide additional information about the countries/cities.
- We then implemented the drop down menu which would give users the ability to interact with the visualization.
- Finally, we plotted a dots on the map to represent the 100 cities we were currently focusing on.

# Full Visualization – First Look

- We then started to look into making the dots on the map sized and colored by the number of YouTube videos a particular city had in a particular category.
- In doing so we needed to access the object we created using our API calls. At this point we ran into some scope problems with the object we created earlier.
- Once we solved our scope problems we were ready to move on to displaying more data.

# Implementing Visualization – Next Steps

# FIRST MILESTONE

First implementation, feedback on progress

# First Milestone Questions

- After submitting our first milestone we met with our TF to discuss our project.
- Main Questions:
  - Would it be better to have a lot of visualizations on our page or give the user the ability to select what data would be visualized?
    - We wanted to display a lot of information, but make sure to 1. Not overload the user 2. Not make the page look cluttered.
  - Scaling of our points
    - Both linear and log scales seem to have their problems. The data is extremely sparse (ranging from 9 entries in some cities, to thousands

# First Milestone Feedback

- Our TF seemed very excited about our YouTube data based visualization and had some very insightful feedback:
  - Perhaps it would be best to have a couple key visualizations and then give the user the ability to modify the kind of data that was displayed in these key visualizations.
  - Although scaling our points logarithmically has it's issues, having a tooltip that displays information about the total number of entries for each city on mouse-over can resolve some of the ambiguity causes by the sparseness in the data.
  - At the time we met with our TF we had no yet started to stylize our visualization in terms of color and so one piece of feedback that he gave us was to make sure our final project used the potential of color.

# BUILDING KEY VISUALIZATIONS

Finalizing the data displayed by the map and building bar charts

# Data Decisions

- One of the very first things we decided to do after meeting with our TF was take a closer look at the data we had collected from the API. We noticed that the movies and gaming category had little data and thus decided it would be best not use them. In doing more research, we decided to include Entertainment and News as categories instead. This choice seemed very logical as both of these categories are ones that we would expect to be universal.
- The JavaScript object we created earlier had a lot of information: average ratings, view counts, number of favorites, and we could always go back to add even more information such as video duration. However, after meeting with our TF we came to the conclusion that it would best to first get down our most key visualizations for this project: the map and the bar chart.
- As shown in previous pages, we had gotten our map paths to display and we had plotted some points. The points we choice to start working with was total number of videos for any given city within any given category because doing so would give us a solid/ easy to work with foundation. As mentioned later on this process book, we would later go on to change this.

# Map and Bar Chart

- The first step toward having a functional map that displayed some meaningful information about these cities and these YouTube categories was sizing the radius of the points according to the city. As previously mentioned we were working with total number of videos per city per category and so we coded our drop down menu to do this.
- After this we decided to add a tooltip to the points to display the total number of videos that were being fed through the scale to create the point.
- Once we got this to work, we moved on to creating the bar chart. We decided that it would make sense to give the user the ability to see the bar chart for any given city when mousing-over the city.

We decided to build a very simple bar chart to begin. The chart displayed the information for the 5 different categories in the city. The bar chart was scaled to use percentages instead of the actual total number of videos in the city in the category. We thought this would be more meaningful for comparing a city against the entire world.

In line with this reasoning, we switched from a simple bar chart to a double bar chart to compare the city data to the world data. We colored the city data one color and the world data another color. We also experimented with putting the bar chart inside of a tooltip as to allow more space for other visualizations, but decided against this idea as it would cover too much of the map on mouseover.

# Initial Bar Chart Progress

We then decided that it would be a good idea to give the user the ability to compare the data of one city against the data of another city. This choice was in line with the main point of our entire project which was to give the user the ability to visualization the difference in YouTube difference across the world.

# Initial Bar Chart Progress

# Full Visualization - Second Look

# Full Visualization - Second Look

# STYLISTIC CHANGES

Coloring the Map and Adding Bootstrap CSS

# Initial Color Changes

- Once we got our key visualizations down (the map and bar chart) and gave the use the ability to visualization different kind of data, it was time to finally change the colors to display meaningful information as well. We experimented with a few different color schemes:

- We wanted vivid colors, but this version seemed somewhat boring so opted for not using it:

- After the initial color changes, we switched to giving ever category a unique color using an ordinal scale.
- We then colored each bar according to its appropriate color.
- Finally, one design choice we made, which turned out be great was to color the dots on the map according to the category that was most popular in the city. For example if a city had the most number of videos in the comedy genre (colored blue) then its dot on the map would be blue.
- At this point however, we still sized our dots by the total number of entries in the category.

# Initial Color Changes

- While changing the colors on our map we decided to start implementing some bootstrap to make our visualization page look a little more like the YouTube page. Using bootstrap we formatted the drop down menus differently and also the layout of the page in general.
- To add the bootstrap to our website we ran into a couple of challenges. First of all, to use the bootstrap we needed to use jQuery and learn that the way in which we updated our visualization when using a drop down menu was inherently different than the way we would update our visualization with Bootstrap tabs.

# Bootstrap.css

# Full Visualization – Third Look

# ADDITIONAL FEATURES

Embedding YouTube Videos

- Adding and formatting the bootstrap to work with our visualization took a surprising amount of time. However, we decided to split up the work a little more at this point and have one of our group members focus on continuing with the bootstrap, have another one work on adding some additional cool features, and have the third one work on using more data in our visualization.
- Embedding YouTube Videos seemed like a really meaningful addition to our visualization for a variety of reasons. First of all, being able to show actual videos makes the visualization appeal to a larger public. Furthermore, seeing a comedy video from a different part of the world, and knowing that it was rated as the funniest video tells the user a lot about the type of people from any given place.
- Once we figured out how to embed YouTube videos in general, we decided to give the user the ability to select how many videos they wanted to see.
- We put all of these embedded videos into a window with a scroll so at to not make our page look too messy.
- Once again we styled the text above the embedded videos and scroll tabs to look like those on YouTube.com

# Embedding YouTube Videos

# MOVING TO MORE DATA

Building our visualization project using Total Number of Videos, Average Rating, and Average Favorites

# Adding More Data

- Being able to display the total number of videos that any given city has for any even category seems like quite a bit of information, but we debated for a long time about what we wanted to do with the rest of our data. Indeed we had data on average number of ratings and average number of likes for every video and so we decided that visualizing this information through both our map and our bar chart would be meaningful.
- To do this we started with a drop down menu and then moved on to bootstrap created tabs.
- We then made sure to update our map, our bar chart, and the embedded videos being displayed to match the city and the category the user was interested in seeing.

# FINAL OVERVIEW

Evaluation

# Adding More Data

- The visualizations that we included in our project provide a lot of information about the data we collected.
- First of all, it seems that Latin America really likes and values music a lot. As can be seen through the map, most of Latin America has the music as it's top category.
- The type of videos that are considered sports throughout the world varies quite a bit. If we looked at the embedded YouTube videos for different parts of the world we can see that most cities consider videos like soccer or football to be sports videos. However, other cities categorize practical jokes or fights as sports.
- There is a lot of variety in terms of usage for YouTube across the world. Based on the number of videos uploaded it seems that YouTube is most popular in really urban areas and least popular in less urban areas

# Learning from Data

- The USA cities included in our visualization seem to enjoy news and entertainment the most. Surprisingly enough, our visualization seems to indicate that music is not as big in the USA as it is in other parts of the world such as Latin America.
- However, when it comes to sizing by average rating, it seems that music becomes extremely popular around the entire world.
- The differences that we see across the world in terms of plotted circles and their colors indicates that YouTube users think about total number of entries, the way in which they give like/favorites, and the way in which they rate videos differently.

# Learning from Data

# Future Work

- If we had more time we perhaps would have opted for scraping the next 100 cities and plotting them as using more cities would make our visualizations all the more meaningful.
- In addition, it seems that providing a visualization that is unique to the United States of America would also be valuable. We think this because in looking at the next 100 cities it seems that a lot of them were from the USA.
- Finally, if we had more time we would have liked to find a way to work with updating the data provided by our visualization in real time. To make this visualization we make many many API calls that take a really long time and thus this is a problem that could only be solved if we wrangled the data using big data methods.

# FINAL PRODUCT

Al Qahirah vs. Bangkok
Total Entries %
100
90
80
70
60
50
40
30
20
10
0
Al Qahirah Data
Bangkok Data
Comedy
Music
Entertainment
News
Sports

# Screencast:

https://www.youtube.com/watch?v=lFfXwZtgMAc&feature=youtu.be