We chose to visualize a Youtube dataset containing the video ID, uploader name, age, category, length, views, rate, ratings, comments and related IDs. The datasets are in tsv files and each one contains data on at least tens of thousands of videos and up to 749361 videos. Video ID is an 11-digit unique string; uploader is a string of the video uploader’s username; age is an integer number of days between the date when the video was uploaded and Feb. 15, 2007; category is a string of the video category chosen by the uploader; length is an integer number of the video length; views is an integer number of the views; rate is a float number of the video rate; ratings is an integer number of the ratings; comments is an integer number of the comments; related IDs contains up to 20 strings of the related video IDs.

You can ask many interesting questions. One example are questions that ask how certain attributes correlate with each other. How does ratings compare with number of views? What interesting data can we get when we compare the related videos with each other? For example, what are the average number of views and ratings of the related videos and the original video? Which category has the most comments per video?

We plan to use Processing to visualize the data because it gives us more control over what we want to create. While D3 is useful for creating set visualizations of data, Processing allows us to control what we want to make and how we want to make it including the details like color, size, and animation. One idea we want to implement is to have each individual video as an object and then when you click on that object, it shows the 20 related videos. You can then click on a related video to see more information about that video.

We plan to represent each category as a person object with the size of the person being the total length of all the videos in that category. We have many ideas for the person object. As of right now, we were thinking about using color and faces to represent the rating and popularity of the video. The number of comments would represent the color of the person. Red would represent high popularity and blue would represent no popularity. If a person has a smiley face, then the video has a high rating; otherwise if the smiley face is a frowny face, then the video has a low rating. The user can navigate by category, which can be presented as a person object that contains the aggregate totals of each attribute.

We are thinking of having a space on the left side of the screen dedicated to filtering data and displaying relevant statistics on them. For example, the user could filter by category, number of views, etc; and the space would display the average ratings, and views of the subset. We could save how the user filters the data and replay it. Each video will be a dot, but if the user zooms in, the dot becomes a person object. Each related video would be connected to the original.

One partner will do most of the data processing and displaying detailed overview of the attributes of each video. The other partner will create the interactions for the users to display the main visualization and the person objects. Both will work on the animations.

