1. Introduction

The document provided has data for 2 artists but the majority is related to one artist only (Taryn Southern). For the simplicity, I focus on the data analysis for one artist. However, each one of the plots I show can be expanded to show multiple artists.

First thing to notice in the data for one artists is that each row contains views and earnings per asset and country. The most important data points are therefore: asset, total views, total earnings and possibly breakdown of those metrics per country. Any visual representation should accentuate the main metrics, show as much relevant detail as possible and display the rest on demand. Although the document contains data for one month only, while brainstorming I kept in mind that one of the main goals is to track metrics over time. I wanted to show at least views, earnings and time in one visualization.

The first question I had was: How are the number of views and earnings related? Is there a simple way to estimate earnings if we know total number of views, and vise versa? In the extreme case when views and earnings are consistently correlated, there may be no need to show both metrics in the visualization, because if we see one value we will know the other one too. To answer this question, I simply divided total earnings by total views in the Excel sheet, for the assets whose earnings came from Net AdSense only (I could have done this taking into account YouTube-sold revenue too, but there was more data for AdSense). Since the results varied across assets, I also tried dividing number of Ad-Enabled views by earnings. After doing so, it was clear the correlation between earnings and views is not simple, and both metrics should be plotted together for now. Either we may need more information (and I need more time) or the visualization can help to understand the correlation between the two.

2. Ideas

This section describes several initial ideas that display data from different angles. One uses only one month data but it can be expanded to capture changes over time, the other one gives a more detailed breakdown of metrics over time, and the third one is a geographical representation.

2.1 Bubble chart view

One simple way to show how much was earned by each asset is to plot assets in a coordinate system, with views and earnings as axes. I implemented this simple solution using d3.js, to depict the idea and demonstrate the knowledge of the tool. Plot can be seen at http://mat.ucsb.edu/~ivana/miscvis/bubbles-stem.html, but for the reference I include the screen-shot here as well (Figure 1). Besides the axes, we can use size and color of circles (asset nodes) to convey additional information. For now (as an experiment), I counted number of countries that brought earnings for each asset and mapped it to the size of the circles. I will discuss this and other ideas at the end of the section.

2.1.1 Rough data analysis

As expected, general trend is that as the number of views grows, so do the earnings. However, we can quickly spot some inconsistencies. For example, “Shave Your Ballsack” has 3,374 views but has less earnings than many assets with equal or lower number of views. It was also seen in more countries than

“Guys They Just Wanna $%*# You” that had 5,523 views.

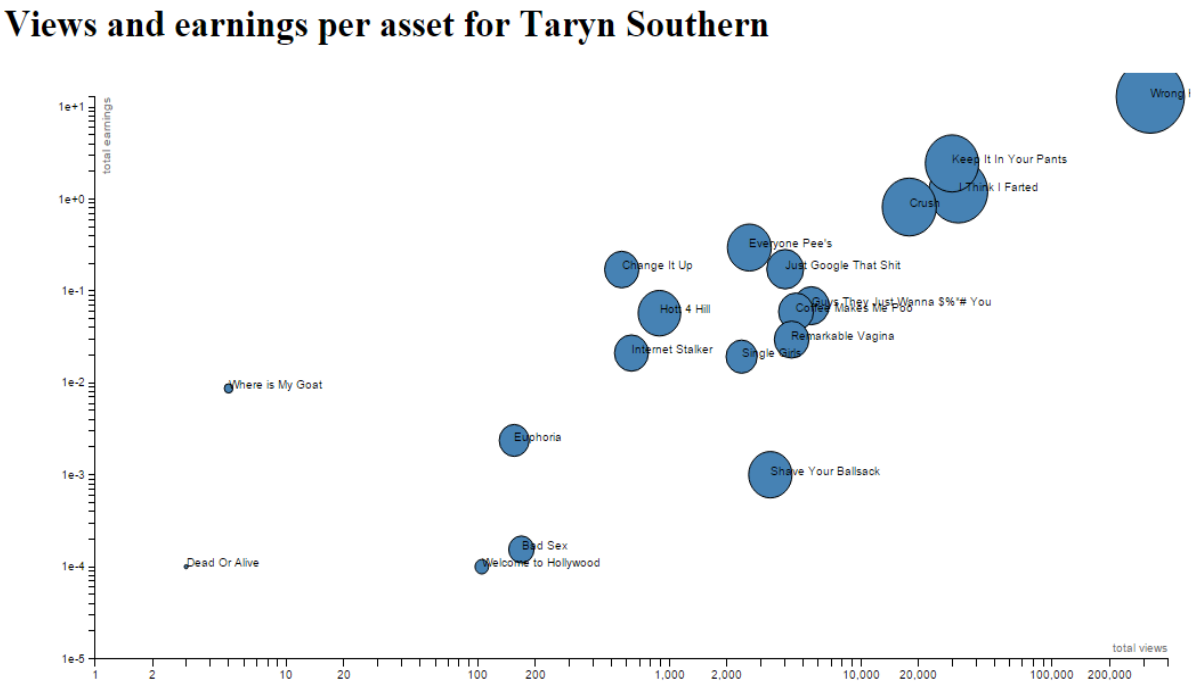

Figure 1: Bubble chart showing earnings and number of views per asset for one artist

Due to the time constraints, I skimmed though the data sheet looking for clues about why “Shave Your Ballsack” has lower earnings. Most of the earnings for all assets come from AdSense so the source itself may not have a great impact. Then I noticed that roughly one third of the views are from Taiwan, one third from Vietnam and the rest is spread across other countries. In comparison, “Just Google That Shit” (4,035 view total) had around 75% views from Taiwan and the rest is spread across other countries (only 6 from Vietnam). ““Guys They Just Wanna $%*# You” has similar distribution of views per country. Therefore, one explanation for lower earnings of ““Shave Your Ballsack” may be that large portion of views came from a country with a lower standard of living. Another consideration is where and how the viewers watched the videos – Ad-Enabled, Watch Page, Embedded or Channel Page. To my knowledge, it makes sense that we should look at the number of Ad-Enabled views to make conclusions about earnings. The above mentioned assets have the similar proportion of Ad-Enabled and total number of views so this metric is likely not sufficient to explain differences in earnings. Finally, please not that the discussion so far is based solely on skimming the data rather than doing detailed data analysis.

2.1.2 Visualization improvements

The plot discussed above can be improved in several ways. First, we can use pie-charts to show breakdown of views and earnings from different sources in place of asset nodes. An example of a pie-chart showing what portion of total views had ads and what portion of ad-enabled views led to revenue is shown in the Figure 2.

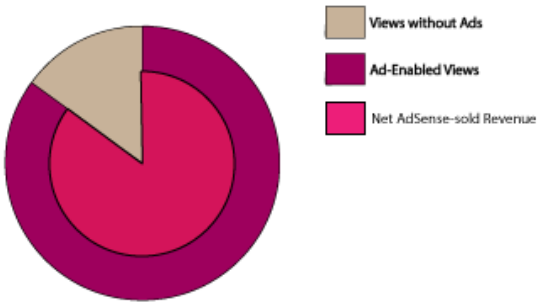

Figure 2: Pie-chart of views and ad-revenue for one asset

Second, any additional information about the asset can be displayed on demand (e.g. by clicking on an asset node) in a form of a table or another visualization. Third improvement is showing changes over time. This can be represented in a similar way as in the following example: http://romsson.github.io/dragit/example/nations.html

2.2 Bar chart views

2.2.1 Bar chart combined with line chart

If we want to show how views and earnings change over time, one option is to show views as bars and earnings as a line in the same chart. Following picture (Figure 3) is taken from the web as an example of this kind of a plot:

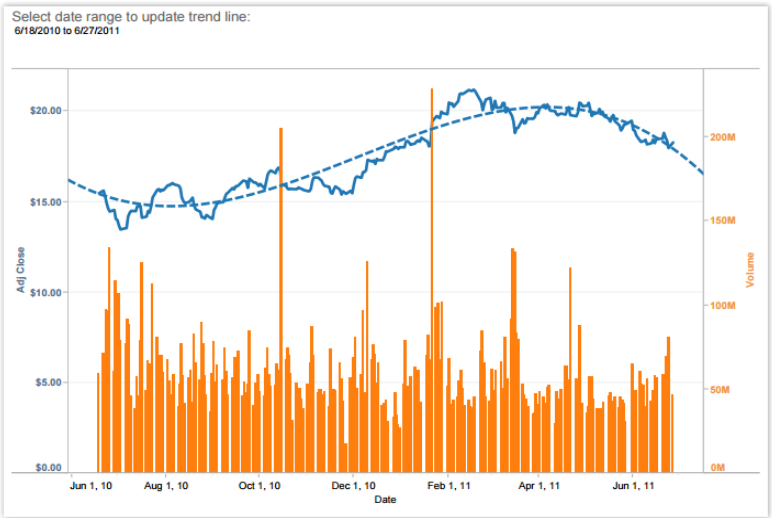

Figure 3: Example of a chart that plots time series of two types of data

Bottom (x) axis represents time, left would represent earnings and right views. Since we may also want to see the breakdown for both metrics, we could split bars to show breakdown of views per different sources (e.g. countries). But we cannot do the same in a simple way for earnings.

2.2.2 Customized bar chart

Figure 4 shows views and earnings for one artists in one customized bar chart. Grey area shows total earnings and views across all assets, and the separate bars within it (labeled as a1, a2… a7) show the same metrics for each asset individually (proportions are rough estimates). Each asset bar is split vertically to show earnings from different sources, and additionally it may even be split horizontally to show types of views. The latter could be done on demand because it may cause cognitive overload otherwise. From the given figure it can be seen that asset *a6* had the most views and brought in the most profit.

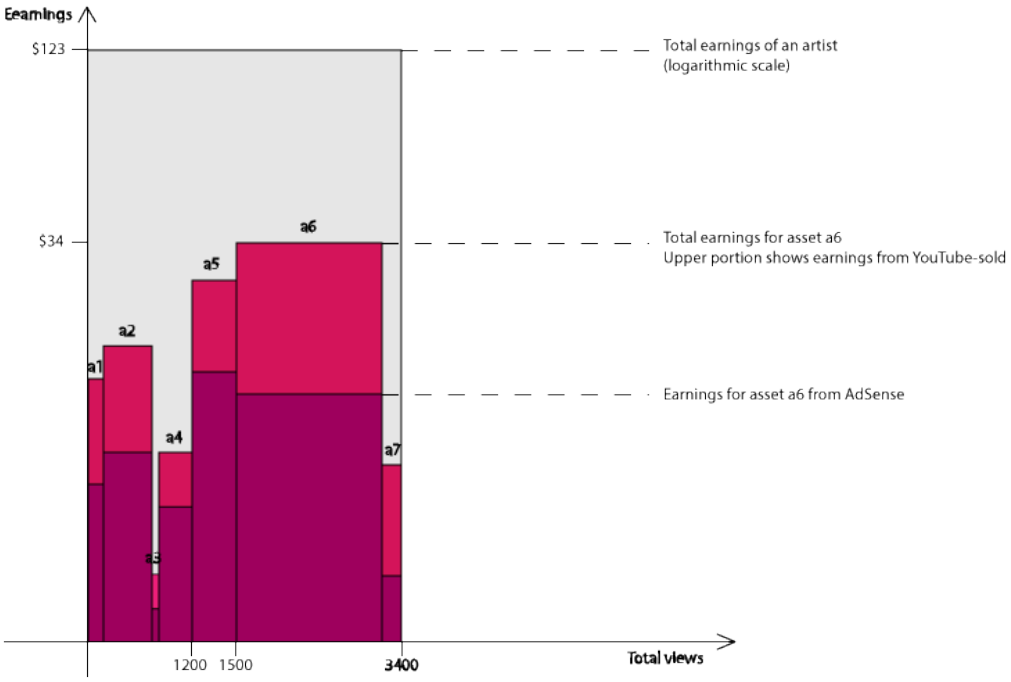

Figure 4: Customized bar representation of views and earnings for one artist. Bars a1 to a7 represent different assets of an artist.

Furthermore, we can show such bars for each month to observe the changes over time. Figure 5 shows an example.

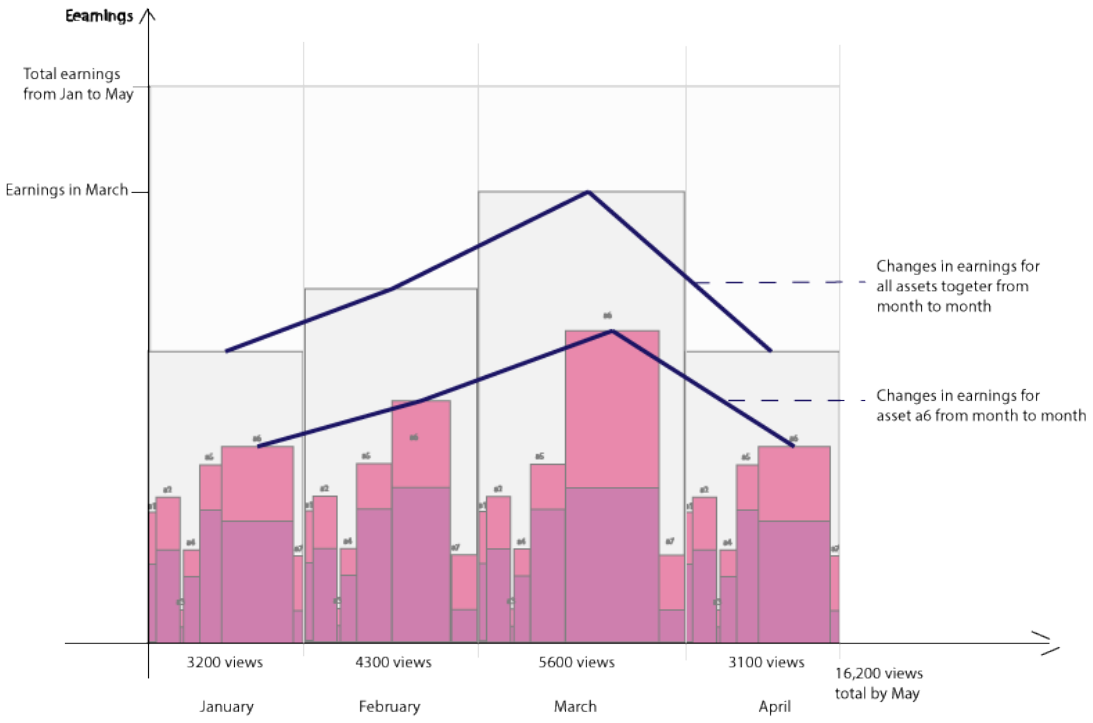

Figure 5: Monthly views and earnings in a bar chart

2.3 Geographical view

If we want to observe the metrics based on countries, we could plot data on a world map

Figure 6: Example of geographical data view

Figure 6 is taken from web as an example of this type of the visualization. Circles can be placed in the countries where videos were watched. Their size can correspond to the number views and earnings can be mapped to a color shade.

3. Data set expansions

Without specific knowledge of the availability of additional data nor extensive knowledge of the industry, I will mention what I think would be useful to know in order to make further conclusions about asset profitability.

- Demographics data: Knowing the audience (age, sex, education level, income etc) can help understand who the fans are and what fan groups should be reached. This is probably not readily available on YouTube.

- Types of ads: Some ads can bring more revenue than others. I am not sure if AdSense shows personalized ads to users, but in general observing the revenue based on ad types gives insight into the effectiveness, especially if we know something about the audience.

- Artist similarity data: If we know who the similar artists are to the one we are observing, based on style for example, we can look into how their earnings and views differ. For example, if one artists has views in one country, but the other one doesn't, the first one may profit from being introduced to audience in that country. We can get some insight into untapped market based on other relevant data too.