

No Swayze, No Wayze

Comprehensive Write-Up

www.ahubers.com/no-swayze-no-wayze

Brian Hixson-Simeral, Alex Hubers, Kent Schlorff, Dawit Tsigie

The first question that sparked this project is, “Can a man ever really know Patrick Swayze? Can his glory ever truly be captured or analyzed?” This project came about because, deep down, we all want to be a little bit more like this legend of a man. Of course, there are things that will never be known about Patrick Swayze, but this cannot stop us from trying to capture the essence of the man. Since we knew that it would be impossible to recreate his spirit, we decided to catalogue and analyze a more physical aspect: his color.

The analysis of Patrick Swayze’s color holds many possible gains. In this analysis we may be able to determine how the film industry, or at least the films that Swayze has been a part of, have changed their palette and tone over the years. By comparing the colors of each of his films with their gross earnings or ratings we may be able to find a combination of colors that tends to make a movie more successful.

As an initial strategy, we started by breaking the project up into four parts. Dawit undertook the web scraping and image collection of Swayze as he has had experience working with search API’s used to gather the images. Kent was given the job of writing the algorithm to convert each of these images into their three main colors. He had already worked with OpenCv before the project, which made him the ideal man for the job. Alex had the job of managing the data and inserting it into a MySQL database, as it gave him the chance to finish up early and assist with any other part of the project. Brian was assigned to working through most of the data

analysis since he is fairly familiar with statistics and math, but less experienced with the computer science side of things. Each task was disjoint enough so that all could be developed in tandem. Analysis obviously required the most attention, so we all were able to funnel our efforts that direction upon completion of our own task.

In the end, our method had changed slightly from what we originally intended. Instead of using the Google search API, we used Bing, as it was less restrictive and allowed us to gather all the data necessary. We switched from our initial plan of turning images into histograms to instead using a method called K-means clustering to find the three most prominent colors, pull out the hexadecimal code, and transfer the data to our database. The database works exactly as expected, though it was significantly harder to set up than we thought it would be. For analysis, we used a color histogram to compare color frequency with certain search terms (this was done using the Shiny library in R). We created a stream graph to do some more advanced analysis, such as the gross earnings of a movie versus the main colors it used as well as how the colors of Swayze's films have changed over the years. These two demos, our color frequency histogram and stream graph, can be viewed via our website listed above, under "data visualization" and "stream graph", respectively. In addition, we constructed a 3d scatter plot of our swayze colors, which can be subsetting via search query. This is also viewable under "data visualization". This plots each RGB color within an RGB cube. We found this option to optimize data clarity whilst also being comprehensive.

Our analysis gave us some great insight into Swayze's color palette. For one, we now know that mixing several thousand colors together usually results in an average color of olive drab. We also know that Swayze's movies contain lots of blacks, browns, and reds - in fact,

pretty much every relevant keyword for Swayze contained more reds and oranges than any other colors by a very wide margin. What implications does this have? Certainly, that Mr. Swayze's movies contain a lot of browns and oranges - that is, lots of dirt, brown hair, and white people. Peach flesh tones were also pretty common in the 1970s and 1980s cinematography - the era where most of Swayze's movies were concentrated.

As an aside, our data certainly shows that 1. further cleaning is necessary, and 2. we should develop an alternative to solely hue for color analysis. By visiting our main page, you'll see that not exactly every image is Swayze. Naturally, collecting 500-1,000 images per search query will lead to quite a few not-Swayzes. Our group proposes, in particular, to use Haar classification to learn Swayze's face and recognize it as a filtering tool. Done effectively, this will certainly increase our accuracy. In regards to color analysis, our intention in using hue was the simplicity in a single dimension for color. However, it seems this will not suffice. Our proposed solution is to develop a binning algorithm that more accurately labels each color. For example, a predominantly black color will be recognized as such, and not considered "red" if that's where its hue lies. This will project the three color dimensions into a single bin variable.

We also wanted to look at Swayze's top grossing films, and see if there was a correlation between color and gross. Ultimately, it was hard to tell, but it should be noted that some of Swayze's lowest grossing films, *Waking Up in Reno* and *Dirty Dancing: Havana Nights*, feature some of the lowest amounts of red hue of all of his films. It should also be mentioned that *Waking Up in Reno* is a rom-com, and *Havana Nights* is an awful sequel. Ultimately, Mr. Swayze's most popular movies were action flicks (or, in the case of the original *Dirty Dancing*, just an acceptably decent movie); just like Arnold Schwarzenegger, Swayze was smart when he

stuck to what he knew. We invite you to explore our data for yourself using our visualizations, as we could not hold all this Swayze to ourselves.