Replication and Analysis of Efforts to Quantitatively Validate Artistic Judgement

March 16, 2020

Will Bates
DSC 180A

1 Introduction

In this replication experiment, my goal is to quantify the concept of abstraction. Fine art is generally considered one of the least quantifiable fields. This conclusion is formed from the idea that judgement of art has no objective measure. Therefore, to call a work "abstract" or "futurist" or "experimental" is inherently as risky a venture as calling a work "art" in the first place. Someone else (including, but not limited to, the artist) could interpret the work entirely differently. Since few elements of artistic style are quantifiability defined, you might both be right.

That said, art, like everything else, is made of numbers. Every pixel is an RGB value and can be converted to a numeric form. Every painting is a 2-dimensional array of color tuples. Yet, the original point still stands. Even if we quantify every individual pixel, the human brain is the one interpreting a work and categorizing it into a genre that only truly exists as a hypothetical collection of works they've seen before that share some subconscious pattern. These categorizations are combined by sources considered authoritative and labeled - by museums, informational websites, etc.

Can we prove or disprove the categorization of a human viewer numerically, and if so, what are the implications on the world of art? How will classification of artistic genres, movements, and eras change when granted a more rigid, quantitative definition? And if it can't conclusively classify artistic style, what attribute are humans given to work with that the computer isn't?

1.1 Historical Context

As domain researchers, we recognize Piet Mondrian's journey from realism to abstraction. We can then see the success of my model's determination of abstraction through its ability to correctly follow Mondrian's career path, as understood by art historians. This will be in conjunction with my own subjective analyses of each painting's complexity.

In addition to essential subjectivity, the fact is that Mondrian certainly did not become abstract suddenly, nor in a perfect linear fashion. In addition, there are difficulties as far as assessing the data. The image recreations range widely in quality to the point where a painting could be labeled abstract merely because it is of low quality. Historically, however, Mondrian has been a person of interest in quantitative analysis of art due to his vastness of work and relatively clearly defined transformation of style.

The studies being replicated and assessed are:
Quantifying Abstraction in Art: Mondrian by Jason Bailey, for Artnome.com
Mondrian v. Rothko: Footprints and Evolution in Style Space by Lev Manovich

1.2 Data Generation

The paintings involved in the study come from the Dutch Institute of Art History (RDK). The Institute contains an apparently complete list of Mondrian's works from 1879-1944. The site watermarks any images before being downloaded, though most images seem to be perfectly high-res in their web-available states. I'm not sure if this would cause any legal issues, so I would have to investigate that further. For the meantime, though, it seems like a fairly complete collection of useful data.

Each observation will refer to an individual painting. The attributes of this row will be the work's title, it's date (or date range), and the image itself. Most of my analysis will be performed on the image itself, so we shouldn't need to add more noise and size to the data by adding more unnecessary attributes.

1.3 Data Drawbacks

The images are saved in URLs, though using image processing packages these will be easy to turn into actual images and RGB arrays. Unfortunately, this data processing method does not yield distinct artwork IDs. They don't tend to intrinsically exist between various sources. Also, Mondrian's titles were not always the most unique and he seemed to place very little importance on them, so repeats seem inevitable. This makes it difficult to combine this data with any other metadata or lost paintings, except through the year of creation.

2 Data Reliability, Inclusion, and Cleaning

As with any cultural analysis, some onus falls upon the analyst to determine how best to define the population to be studied. This will be based on, among other factors: access to data, categorization, and quality of data. In this case, over the course of data cleaning I reduced the total number of pieces from 1482 to 785 through this process of assessing the quality and relevance of the data. This is definitely a big difference, but the rest of the data still exists if it should become useful for future analysis and I wanted to make sure to not be measuring distinct objects under the same scale - for example, should a painting and a chair both be judged by their resolutions? Probably not.

Speaking of resolution, this turned out to be a significant factor in thinning out the data. It spoke both to the identification of unrelated media as well as the quality of the images. I plotted the resolutions and developed percentiles, noting specifically where major drop-offs occur. I viewed some of those images to confirm my suspicions, that they were generally too low quality or of a different medium than which we are studying. For these reasons, I chose to remove those observations for this analysis. I set the threshold at the mean minus 2 times the standard deviation, which via plotting histograms resulted in a firm deletion of the outlier class while only removing 21 paintings.

At certain points, I also had to remove observations with bad IDs and missing images. I chose to do my best in interpreting the dates of the images, though some will absolutely be questionable. Some ranges on the catalog exceeed forty years, which I could only trust due to a lack of any other information. I created distinct start and end date columns if selecting either one, or both, turns out not to be the best interpretation of image date.

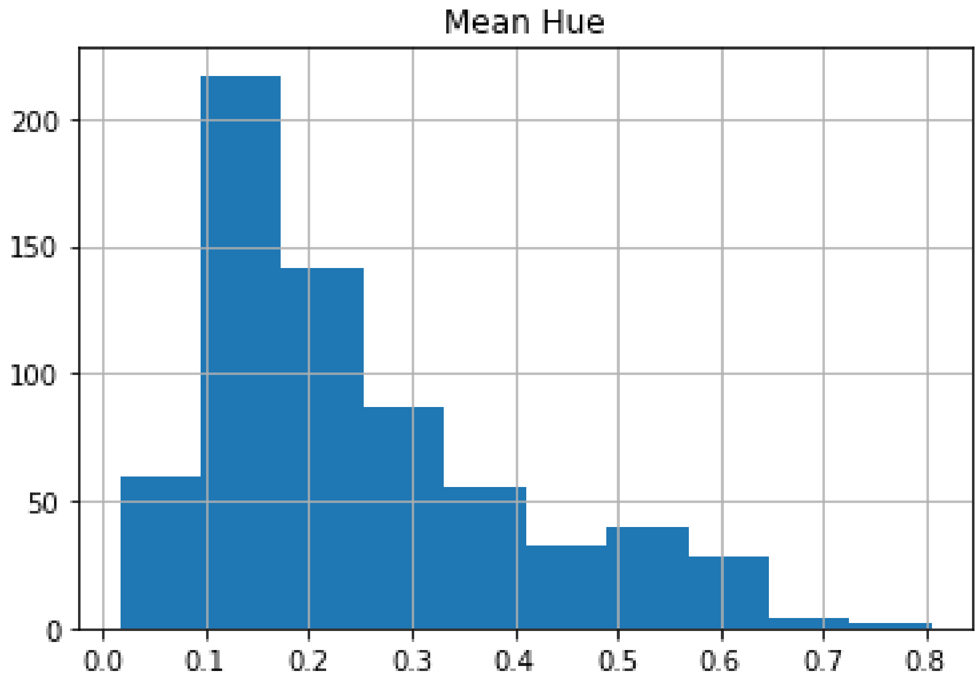
In addition, I was able to scrape the "category" of each piece. This considered paintings, drawings, illustrations, and other forms of art. For the time being, I will only be analyzing paintings. In order to also assess other forms of media in the same study, I would have to alternately weight features by category, which, while certainly possible, further complicates the study. Further, I would question whether non-painting works would be expected to match the development of abstraction foretold by art historians in regard to Mondrian's paintings.

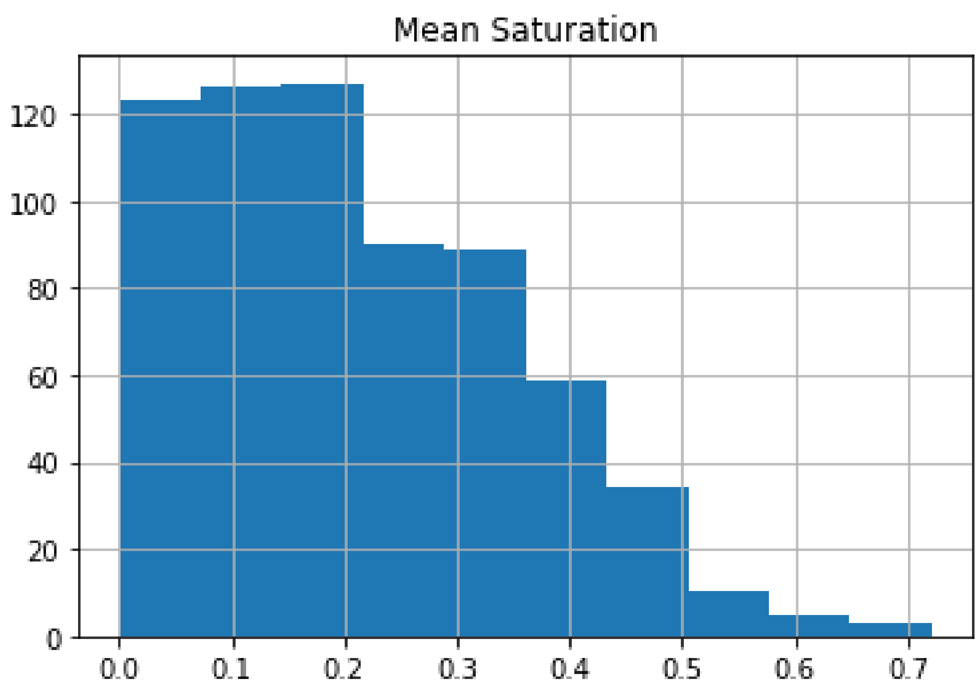
3 Exploratory and Historic Analysis

3.1 Manovich's Hue-Saturation-Value Features

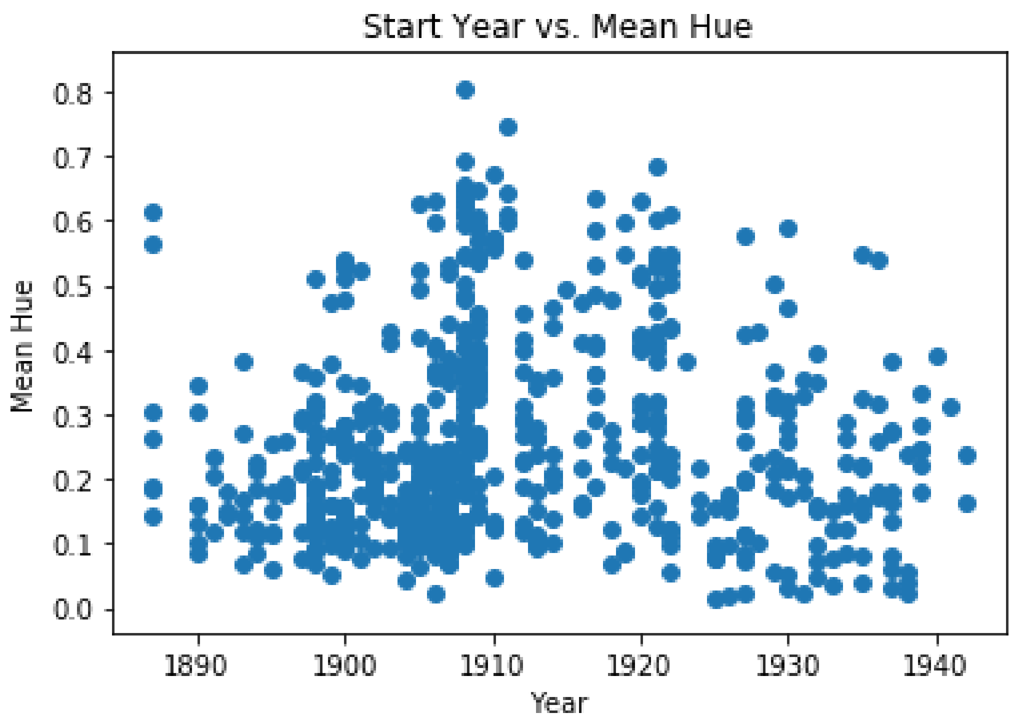
To validate my choices of inclusion and exclusion of works, I developed univariate and bivariate analyses of the features generated. Lev Manovich's features were hue, saturation and value: combined, these three characteristics can fully describe any image pixel. His technique used Principal Component Analysis on these image features in an attempt to cluster Mondrian's works. I tested bivariate analysis specifically on the start year of the work vs image features including mean hue and mean saturation. Correlations seemed to appear on the scatter plots of start date vs mean value, which showed a relatively strong positive correlation. Below are a few of the visual summaries of the exploratory analysis.

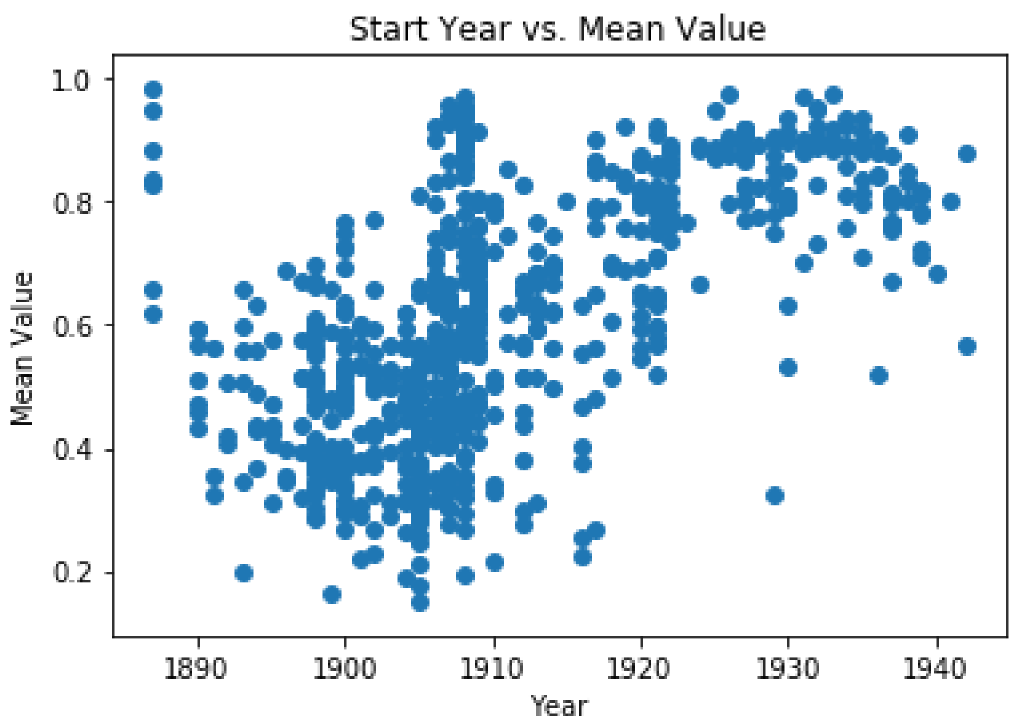
3.1.1 Frequency Histograms of Visual Features





3.1.2 Scatterplots of Visual Features Year of Composition





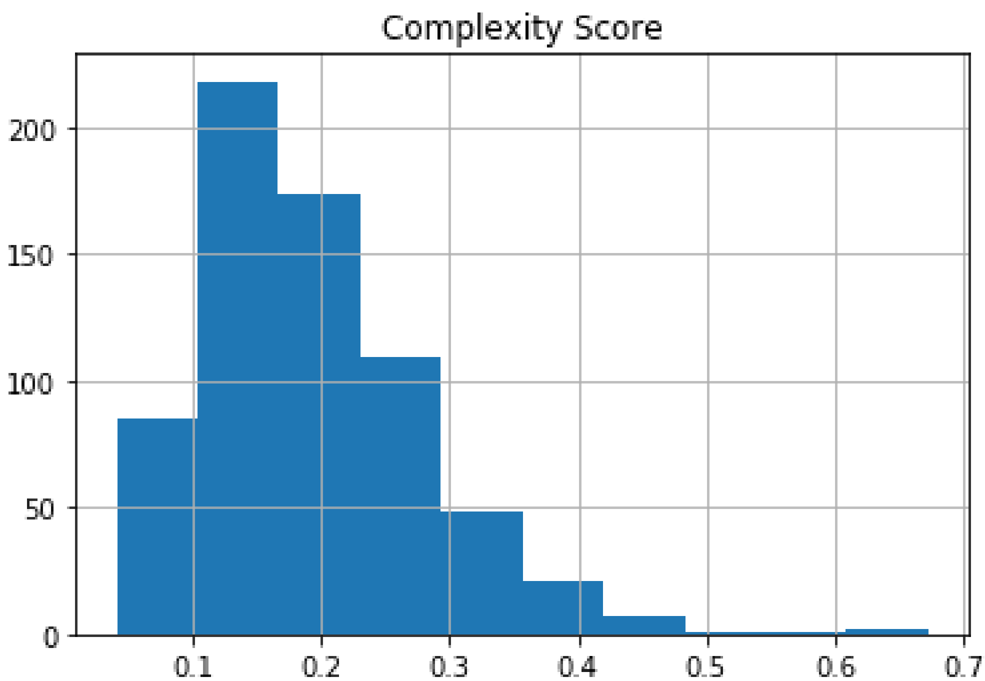
3.2 Bailey's Complexity Score

Bailey chose instead to select three computed features he theorized would be relevant to abstraction and averaged them to calculate an overall "complexity score" ("complexity" to stand as the inverse of "abstraction"). These three features are as follows:

- Color Score: Bailey's Color Score was the total number of distinct RGB values used in the image. In theory, the more colors used, the more complex a painting.
- Variance Score: This referred to the average row-wise variance in the value of the image. The idea being that complex paintings will change more substantially across the horizontal plane.
- Edge Score: The total number of edges, or contours, detected in an image.

The average of these normalized quantities became the "complexity score" of the image. Bailey hypothesized that these three features could capture Mondrian's evolution toward abstraction. In a typical Mondrian abstract work, it was likely to see large patches of single colors, which would correspond to a low color score. In addition, an abstract work might have less activity or dynamism across the landscape, suggesting a low variance score. Lastly, Bailey theorized that an abstract work would have fewer complicated mechanisms and objects, yielding a low edge score. Together, he hoped, abstraction in general could be granted a quantitative definition.

I plotted the distribution of Bailey's Complexity Score for reference.



3.2.1 Where Complexity Falls Short

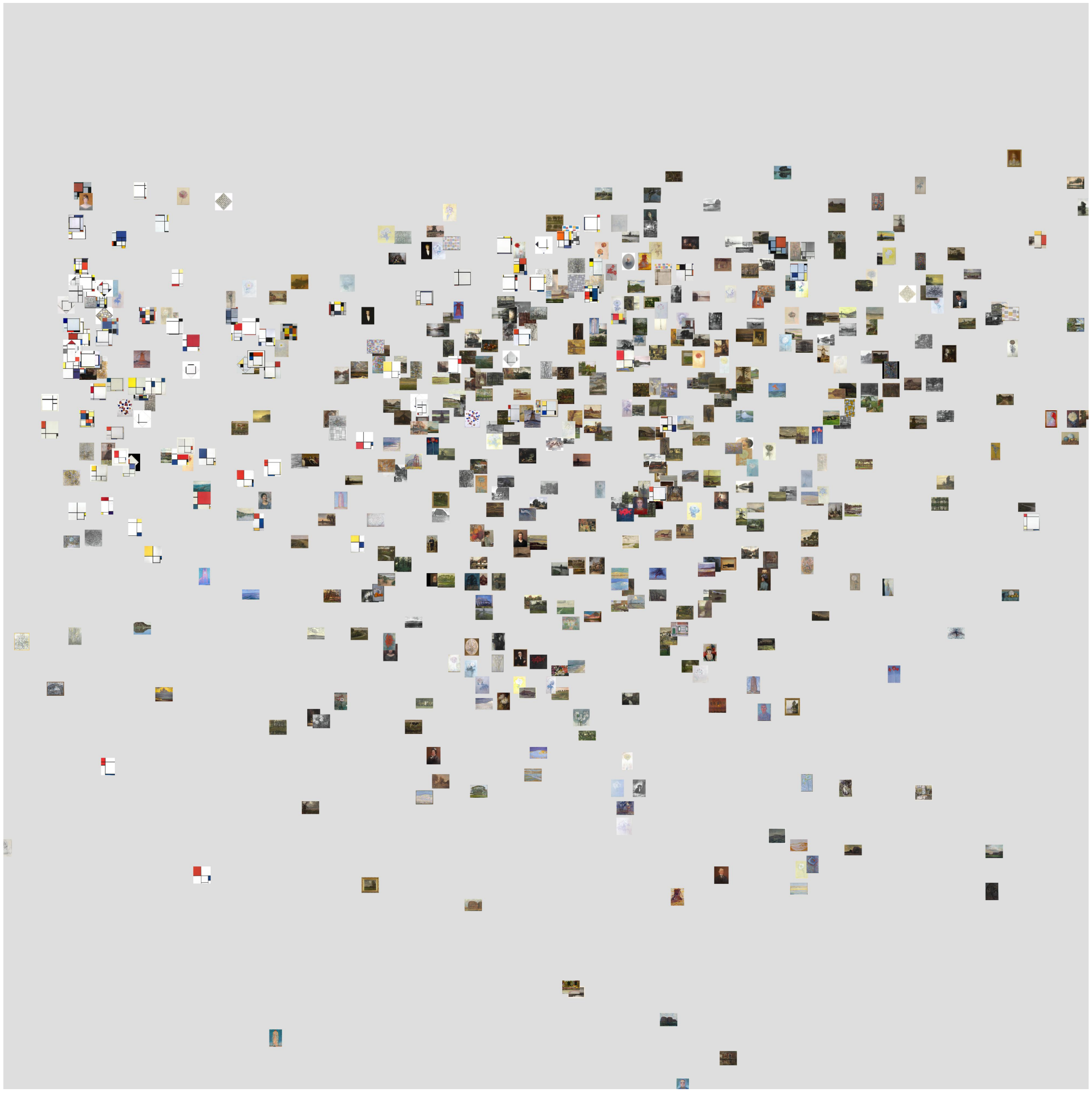
Bailey's vision of complexity contains some valuable baselines. The combination of features related to color, shape, and brightness makes a lot of sense, as it allows the computer to see the picture as more of a whole image, like a human does. However, Bailey's computation is naive and failing to attempt to weight the features using an intelligent classifier is an avoidable pitfall. Further, the choice to computer per-row variance in the variance score, as opposed to per-column or otherwise, seems to be a mostly arbitrary decision and might disregard relevant patterns. Lastly, though this may have been intentional, Bailey's color score considers every distinct RGB value equally, without regard for distances between colors.

4 Alterations and Improvements

4.1 Adding Features to PCA Regressor

Though Lev Manovich did employ PCA in an effort to isolate relevant aspects of his features, his input dataset was unnecessarily limited. His features: mean hue, mean value, and mean saturation, appropriately captured the average values of the pixels, but say nothing more about any variation or changing throughout the image. By adding in Bailey's features, as well as hue *variance*, value *variance* (as in an image-wide "variance score"), and saturation *variance*, the PCA became much more effective at locating relevant vectors.

This more comprehensive set of features fared better in a linear regression using two PCA vectors with the composition start date as the label. It yielded an R-Squared value of 0.055, compared to 0.017 with Manovich's original features. This projection shows clusters appear when the two PCA vectors are plotted on the x- and y-axes.

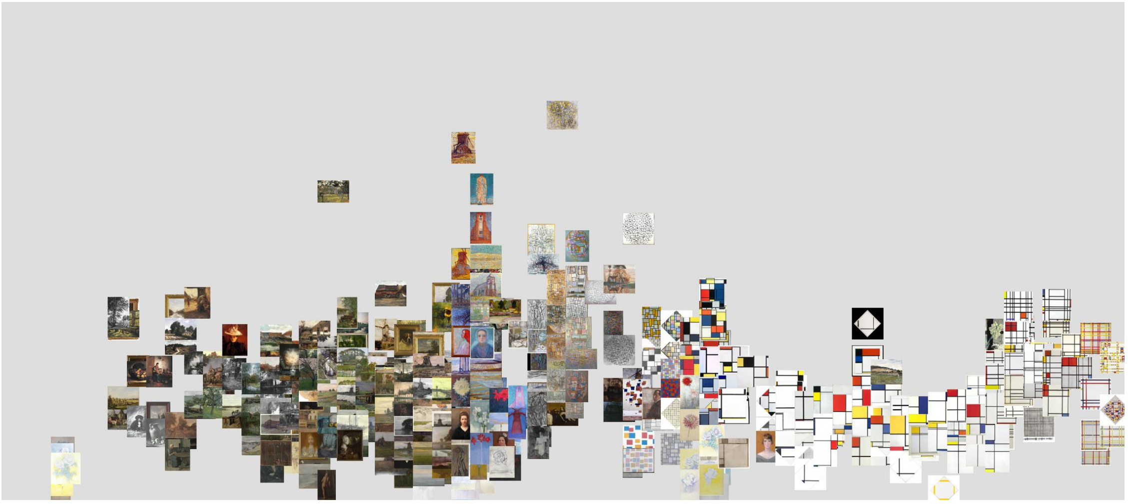


PCA 1 is on the x-axis, PCA 2 is on the y-axis

4.2 Advanced Complexity

Jason Bailey's analysis highlights a more subjective approach to determining attributes of abstraction. Upon analysis of his thought process and the results it generated, I believe he was on the right track. I chose to patch up a couple points in which I felt there were holes in his analysis to see if I could do better. Primarily, I took a closer look at his Variance Score. Instead of looking solely at a per-row value variance, I wanted to try to detect other patterns of drastic changes in value.

I generated two additional features: the Column Variance Score and the Square Variance Score. The Column Variance Score, intuitively, took the average per-column variance of the image. The Square Variance Score browsed the image by each 2x2 pixel square and took the variance of those four pixels. This measure was meant to detect value changes in horizontal, vertical, and diagonal directions simultaneously. These two features were combined with the Row Variance Score to create a more comprehensive Variance Score to be plugged into Bailey's complexity computation. The result is shown below:



Complexity is shown on the y-axis, Year is shown on the x-axis

5 Conclusion

In my update of Manovich's model, distinctions in paintings become clear. The upper left of the scale, based on the eye test, accurately identifies many of Mondrian's abstract works. Darker works seem to appear more commonly on the right side, while the bottom part seems to include more colorful paintings. The model captures abstraction, but only a certain "type" of abstraction, so to speak. The top left captures bright, white-backed paintings with very few colors. It misses, however, distinctly abstract works that lack these qualities.

For the other model, we see the complexity scale display a peak in complexity variation in the middle of Mondrian's career. At the beginning and end, he maintained low complexity scores. This suggests that my definition of abstraction correctly identified abstract works, but also falsely included earlier works, before his abstract period. These may have garnered low complexity scores due to low-quality images or high-simplicity paintings, despite not being abstract in the understood sense.

Ultimately, art is still largely in the eye of the beholder. Over the course of the data generation, inclusion, and cleaning processes, I felt that much of the results were coming from my decision making as a viewer. Should a drawing be on the same scale as a painting? What about an oil painting vs. a watercolor? With advanced neural networks and greater ability to obtain usable data, we may come closer to defining artistic qualities. But for now, while Bailey and Manovich both make efforts, neither their original works nor my alterations on them can fully capture what the human mind perceives.