Assignment 1 Parts 1-2

January 24, 2020

1 Part 1

1.0.1 Intro

In this replication experiment, our goal is to quantify abstraction. Fine art has always been one of the least quantifiable studies. Judgement of art has no objective measure. Therefore, to call a work "abstract" or "futurist" or "experimental" is inherently a risky venture, as someone else (including, but not limited to, the artist) could interpret the work entirely differently. Since the quantifiability of art was never 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 mind is the one interpreting a work and deeming it abstract. Can we replicate this assessment using computers, and if so, can be begin to approach the accuracy of the general layman assessing art?

1.0.2 The Plan

The goal of quantifying art, in this study, will be to label works in attempt to match to predetermined, subjetive categorizations. As domain experts, we recognize Piet Mondrian's journey from realism to abstraction. We can then gauge the success of our model's determination of abstraction through its ability to correctly follow Mondrian's journey, as understood by art historians. This will be in conjunction with our 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.

2 Part 2

2.0.1 Generating the Data

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 wa-

termarks 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 our 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.

2.0.2 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.

[ ]: