



Beauty is in the bid of the beholder: An empirical basis for style



William N. Goetzmann^a, Peter W. Jones^b, Mauro Maggioni^c, Johan Walden^{d,*}

^a Yale School of Management, United States

^b Yale University, Department of Mathematics, United States

^c Duke University, Department of Mathematics, United States

^d UC Berkeley, Haas School of Business, United States

ARTICLE INFO

Article history:

Received 29 April 2016

Accepted 10 May 2016

Available online 30 May 2016

Keywords:

Art market

Auctions

Art movement

Art style

Dimensionality reduction

Laplacian eigenmap

ABSTRACT

We develop a method for classification of works of art based on their price dynamics. The method is in the same spirit as factor models commonly used within financial economics. Factor models assume that price dynamics of assets are related to underlying fundamental characteristics. We assume that such characteristics exist for works of art, and that they are associated with what we intuitively think of as style. We use a clustering algorithm to group artists that represent similar styles. This algorithm is specifically well-suited for situations where statistical distributions are far from normal – a description we believe fits well with markets for art. We test the method empirically on a ten-year sample of price data for paintings by 58 artists. Even with this limited data set, we clearly identify five groups and show that these are related to a standard classification of style.

© 2016 University of Venice. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Style has long been the basis for classification in the history of art, however a precise definition of style, and a consistent basis for grouping works of art into styles has eluded scholarship. Style as applied to works of art is necessarily based upon not only a large variety of visual and material attributes of an object, but also upon the manner in which these attributes are executed and assembled. If engineers are only now beginning to develop optical recognition tools that can consistently identify the same face in front of a camera, imagine how long it will take to use optical data to distinguish say, a Renaissance from a Baroque painting, or to understand the subtleties of visual allusion and allegory. Given that works of art from the 20th Century were often, by their very nature, crafted to challenge stereotyping or easy classification, it is hard to imagine that style recognition will ever be meaningfully automated. It could happen, just as we have seen Gary Kasparov's chess prowess equaled by an IBM computer. However the limited dimensionality and the clarity of the rules of chess make it more susceptible to analysis. Aesthetic development has often taken place by breaking rules and replacing them. As Morse Peckham suggests, the most influential works of art are those that are initially perceived as chaotic. Art, in Peckham's view,

* Corresponding author at: 545 Student Services Building #1900, CA 94720-1900. Tel.: +1 510 643 0547.

E-mail address: walden@haas.berkeley.edu (J. Walden).

... serves to break up orientations, to weaken and frustrate the tyrannous drive to order, to prepare the individual to observe what the orientation tells him is irrelevant, but what may every be relevant.¹

A difficult challenge indeed, for a tool trained on patterns to detect regular, logical structure.

Research by economist David Galenson has suggested an economic basis for the identification and the analysis of quality in works of art (Galenson, 2002). Rather than using the visual characteristics of works of art as the grounds for evaluating which works are the most important, Galenson uses auction prices. In effect, he projects the vast, complex dimensionality of the visual and physical and historical characteristics of works of art down to two dimensions, price and time. This allows hypothesis testing about which works of art, for example, are perceived as most important, and the identification of a few basic career trajectories of successful artists.

In this paper, we propose adopting this economic approach to the problem of identification of style. Rather than constructing an automaton and feeding it the world's books on the history of art and training it ultimately to synthesize visual and physical input into stylistic classification, we propose to use auction prices. The use of price data has the potential to employ the cognitive and aesthetic capacities of the world's art auction market participants in order to define aesthetic styles. By relying upon the bidding behavior of market participants, under some simplifying assumptions, we can interpret the world's auction markets as a continually active and constantly changing market for opinions about the relative value and associations of works of art. The proposal in this paper is to use auction data in an econometric model to uncover "associations" among the works of artists, and these associations one could label "style." While we cannot claim that this approach uncovers style as an art historian might understand it, we *would* claim that it can, when executed with the proper econometrics, approximate the idea of style as collectors and dealers might understand it. Since collectors and dealers necessarily rely upon art historical scholarship and interpretation as the basis for value, this approach may then, at least secondarily, reflect expert opinion.

An important limitation of using market information to define style is that it displaces the specialists – the connoisseur and the art historian – from the definition of style, and replaces them with the customers for works of art, few of whom are likely to be trained in art evaluation. In asset pricing, where object values are common values, this problem is addressed through the processes of arbitrage in expectations. The Arbitrage Pricing Theory (Ross, 1976) for example shows that the existence of a single risk-neutral investor with unconstrained borrowing capacity – the expert – can enforce efficient pricing. This person can drive the value of two economically equivalent assets to the same price by bidding up the price of shares of the undervalued asset and shorting the shares of the over-valued asset until these values align themselves according to the "Law of One Price." A tricky aspect of this theory is that two such financial titans cannot agree to permanently disagree on the economic equivalence of the assets, or else they would furiously and infinitely bid their views without driving the prices towards some common value. In a sense you would get two equally powerful invisible hands arm wrestling with no resolution. Thus, even in the world of asset pricing models there must be some agreement on style. In the example, of course, at least one of the financial titans must be wrong, for, in an economic framework of objective values are equivalent to expected sums of future discounted cash flows. In the world of art, this need not be the case. There is no ultimate economic value for objects apart from the tastes of those with the money to indulge them.

Curiously, scholarship in financial economics has found the term "style" useful in the analysis of asset values and investor behavior. Barberis and Schelifer (2003), Chan et al. (2002), Brown and Goetzmann (1997), Sharpe (1992) among many others use the term style to describe common strategies of investors and/or common characteristics of investment securities. In doing so, they implicitly assume that broad market perceptions, behavior and perhaps even tastes are important in the world of assets. Indeed, Kumar and Lee (2006) and Kumar (2002), explore the pricing implications of style investing and find empirical evidence that common perceptions – not just the perceptions of the well-capitalized arbitrageur – may affect prices. Thus, an interesting implication of asset pricing theory and empirical evidence is that subjective style "factors" exist in the investment world and that they matter.

The suggestion that financial economics has begun to profitably borrow from the concepts of art historical scholarship may be a slim motivation for art historical scholarship to borrow the tools of modern finance. After all, in a private values market, in which personal taste, not expected sums of future discounted cash flows, drive market prices, the law of one price does not necessarily hold. Theoretically, tastes for works of art could be orthogonal. There might be no agreement on what is good or bad art, or what objects are meaningful substitutes for each other in the collector's imagination. In fact, we know this is not the case. A by-product of the construction of art price indices (Anderson, 1974; Goetzmann, 1993; Mei and Moses, 2002; Renneboog and Spaenjers, 2013 and Korteweg et al., 2015) is a measure of the variance in price changes explained by the common factor – the art index. The index explains a lot; typically half of the price change in a painting's purchase then re-sale over long holding periods can be attributed to broad market movements. Even if some of this can be explained by the shifting economic fortunes of collectors (Goetzmann and Spiegel, 1995; Ait-Sahalia et al., 2004; Lovo and Spaenjers, 2014), there remains a common component of value to works of art that reflects the degree to which a set of auctioned objects are regarded as substitutes for each other in collector utility functions. Formalizing this result, and expressing the idea of style in terms that will eventually allow estimation, will require some algebraic notation.

The next section develops a more formal framework for defining a relationship between style and prices, and for estimating styles from observed prices. In the third section, we use this framework to empirically analyze art styles. In a limited

¹ Peckham (1967), *Man's Rage for Chaos: Biology, Behavior and the Arts*, Patrick Wilkinson, editor, Maiseonneuve Press, University Park Maryland, reprint of the 1967 edition.

data set of auction prices over ten years, we clearly identify five groups of artists. These groups fit quite well with a standard classification of art style. The fourth section concludes.

2. Defining style

There are two steps in developing a method for empirically estimating styles from observed prices. The first step is to assume a model of the relationship between styles and prices: Consider an hedonic valuation model of a work of art j for collector i at time t . We assume a functional, f , that is conditioned upon the stochastic state of the world, where ω is the wealth of the collector and γ is the percentage of that wealth the collector wishes to invest in art. This model generalizes the model in Goetzmann and Spiegel (1995). The characteristics of the artworks; the artist [a], the size [s], the date [d] and other characteristics [x] all figure into the price, P_{ijt} , the collector is willing to pay at that time, and are represented by a vector X . We simplify the valuation to a linear model $X_j \beta_{it}$ for which the β_{it} represent factor loadings at time t and the X_j represent perceived factors that are associated with the definition of style, for example, landscape subject matter, pointilist technique and so forth. These are scaled by the value of the investor's art investment at a given point in time $\omega_{it}\gamma_{it}$:

$$P_{ijt} = f_{it}(a_j, s_j, d_j, x_j) = [X_j \beta_{it}] \omega_{it} \gamma_{it}. \quad (1)$$

In an auction, the bidder with the highest valuation at time t obtains the work of art. If all bidders had the same wealth and the same preference for art, then the determinant of the winning bid would rely solely on the characteristics X_j and the factor loadings β_{it} . If everyone had the same aesthetic tastes as well as the same wallet, then the highest priced item would be determined solely by the characteristics X_j . If these common sets of tastes evolved through time, prices of objects would change according to their characteristics.

So how does this models help us with style? Style can be thought of as classifying work of art, $j = 1, \dots, N$, into $K < N$ styles, according to the characteristics, X_j . Notice that although we are actually interested in X_j , P_{ijt} can make this classification problem easier, since the biggest challenge to classification of works of art using characteristics is the misspecification problem – we may mismeasure or leave out key variables, or not be able to correctly capture them in our functional form. If we only had X_j we would be back to the problem of programming the subtleties of the Baroque into a computer – with no way to check our work except by asking a human expert! In a setting in which tastes evolve, however, the dynamics of object prices can help differentiate styles. We observe $P_{l,j,t} = X_j \beta_{it} \omega_{it} \gamma_{it}$. In this case, the capital l subscript indicates the winning bidder. For example, as factor loadings for some works of art decrease, their relative prices will decline. *Without* dynamics, the classification is infeasible unless an econometrician is willing to specify X and estimate a classic hedonic regression. *With* dynamics, it is not necessary to separately estimate X_j and β_{it} .

This leads us to the second step in empirically classifying styles from observed prices, which is to use an algorithm to group together artworks with similar price dynamics. Input to such a classification algorithm will be price changes of artworks through time; output will be a number of *clusters* – a parsimonious set of groups with artists in the same clusters representing the same style. Brown and Goetzmann (1997) exploited this idea to estimate styles of mutual fund investment managers, applying a clustering algorithm to the time-series of returns of funds. Each observation consisted of the time-series of monthly returns of a single fund. The algorithm groups funds so as to minimize the within-group sums of squared residuals from the group center. This is a so-called *K-means algorithm*. It requires *ex ante* specification of the number of groups – as long as $K < N$ there will be squared residuals to minimize. More general algorithms allow for hierarchical clustering – the estimation of a graphic tree expressing the distance among subgroups of observations. They relieve the investigator from the burden of guessing the number of appropriate styles. In recent years, new algorithms have been developed that have attractive robustness characteristics. Traditional methods often depend on the “noise” in the model being close to normal (Gaussian) and relationships being linear. The new methods also work well when “noise” is far from normal and relationships are nonlinear – a reasonable assumption on the relationship between art styles and prices. We shall use a specific such algorithm, the *Laplacian eigenmap method*, and we will see that it performs better than standard methods when applied to our art data set.

3. Style classification – an example

In order to exploit the dynamics of taste to classify works of art into styles, it is necessary to observe a time-series of prices for each individual work of art. Alas, individual works of art do not sell every day – the illiquidity would seem to be a hopeless barrier. One approach, which we will follow, is to specify subgroups – that is, to lump the works of a single artist into one category and then apply a classification algorithm at the level of the artist, rather than the artwork. We do this for a specific data set provided by ArtNet.²

² <http://www.artnet.com>

3.1. Data

ArtNet identified the 100 most widely represented artists in their database of auction prices over the period 1984–1995, according to how many works of art by the artist were sold in the time period. The raw data consisted of 115,812 reported paintings and other works of art.

For each artist, we constructed an index of the median sales price of the artist's works each year. The median is better than the mean, of course, because it eliminates the effects of extreme sales prices. Bought in prices were not recorded as transactions. This is an admittedly crude measure of the price dynamics of the artist's works. Heterogeneity of quality alone is likely to cause fluctuations in the median price – fluctuations in X_j due to changing works of art will be misconstrued as changes in β_{it} .

Many artists did not have any reported sales for 1984–1985, so these years were not included in the sample. Furthermore, some artists were dropped from the sample because they did not have at least one painting sold in each of the remaining years. This left us with 10 years of data for 58 painters and altogether 20,700 observations. The bulk of the remaining artists are late 19th century and early 20th century painters, represented by giants like Picasso, Renoir and van Gogh. “Younger” artists, like Warhol and Lichtenstein are also well represented, whereas only five artists in the sample were born before 1840 – the oldest being Jean Baptiste Corot (born 1796).

The prices varied a lot between different types of art. We therefore identified three subtypes: *Paintings*, *Watercolors* and *Drawings* that were treated separately. Yearly returns (defined as relative increase in median price) were calculated for each subtype. For each year, the return was defined as the average of the returns of the three subtypes. For a specific subtype, if observations were lacking for any of the two years used to calculate return, this subtype was excluded.

In Table 1, we show summary data for the 58 remaining artists, i.e., total number of paintings sold and average price over the entire time-period for the three subtypes. We see that there are large variations between artists within a specific type, e.g., paintings, from a low average median price of \$22,000 for Alexander Calder (Artist 27), to a high of \$1,903,000 for Vincent van Gogh (Artist 4). Moreover, the average median price for paintings is typically, but not always, significantly higher than for the other two types – a notable exception is Henri Matisse (Artist 6) whose median price for paintings were \$540,000, compared with \$1,333,000 for watercolors.

3.2. The clustering algorithms

Formally, a clustering algorithm applied to a set of points aims at dividing the set into K clusters. Here, K can either be endogenously determined or exogenously specified. Points within a cluster are “similar,” whereas points in different clusters are “different”. In this paper, the aim is to cluster artists with similar styles by observing price dynamics of sold works of art.

There are a large number of traditional clustering algorithms, including parametric approaches, like Gaussian mixture methods and K -means methods, and non-parametric approaches, like hierarchical clustering algorithms. For a survey paper, see Jain et al. (1999). We will use two such traditional algorithms, the K -means algorithm and a hierarchical tree algorithm. Inspired by the interest in the learning community about nonlinear methods for dimensionality reduction and data mining (Tenenbaum et al., 2000; Donoho and Grimes, 2003; Roweis and Saul, 2000; Belkin and Niyogi, 2003; Coifman and Lafon, 2006), we also decided to use a third algorithm, based on this research.

Various constructions have been proposed that seek to nonlinearly embed a set of data points in a lower dimensional space while minimizing distortion. These algorithms can be directly applied to clustering problems, as solutions in lower-dimensions offer advantages in terms of stability, interpretability and speed. The algorithm focus on local estimations of properties of the data, and have different ways of incorporating these local structures into a global structure. We will use the Laplacian eigenmap method, as a representative for these new types of methods (Belkin and Niyogi, 2003). By focusing on local similarities, the Laplacian eigenmap algorithm has the potential to outperform Gaussian probabilistic models, when probability distributions are highly non-normal.

We give a brief description of the three algorithms that we shall use: the first algorithm is the *K-means algorithm* (also used in Brown and Goetzmann, 1997). It works as follows: Given a set of N points, X , with a distance d (inversely related to a similarity measure), and a fixed integer K , it returns a partition of X into K subsets S_1, \dots, S_K . The partition is constructed by finding the K “best” centroids, and then by assigning each point of X to the closest centroid. The minimization problem solved by K -means is

$$\arg \min_{\{S_1, \dots, S_K\}} \sum_{i=1}^K \left(\sum_{x \in S_i} d(x, x_i) \right)$$

over all partitions $\{S_1, \dots, S_K\}$ of X , where x_i is the centroid of S_i . The expression in the round bracket is a measure of dissimilarity of each cluster, hence K -means tries to minimize the sum of these measures, over all possible K -partitions of X . Common choices for d , are the squared Euclidean distance, or one minus the cosine of the angle between points viewed as vectors.

The second algorithm is a *hierarchical clustering algorithm*. It works as follows: given a set of N points, X , with a distance d (inversely related to a similarity measure), it returns a (usually binary) tree. The root of the tree is the whole set X , the leaves are the single elements of X . The other nodes are sets that are union of the sets associated to the children nodes. Each level

Table 1

Summary data for 58 artists in sample. For each type of painting (Painting, Watercolor and Drawing), the total number of works sold and the average price over the 10 year period is shown.

Artist #	Artist name	Paintings		Watercolors		Drawings	
		#	\bar{P} , [\$1000]	#	\bar{P} , [\$1000]	#	\bar{P} , [\$1000]
1	Picasso, Pablo	460	528	123	155	712	38
2	Renoir, Pierre Auguste	588	239	19	63	118	113
3	Monet, Claude	198	1698	0		21	162
4	Gogh, Vincent van	46	1903	5	511	37	245
5	Chagall, Marc	238	572	263	145	179	19
6	Matisse, Henri	109	540	13	1333	331	51
7	Degas, Edgar	51	393	5	83	329	195
8	Miro, Joan	188	262	135	83	148	31
9	Pissarro, Camille Jacob	198	557	89	140	198	6
10	Cezanne, Paul	64	1553	54	206	35	31
11	Leger, Fernand	220	280	272	36	126	21
12	Dubuffet, Jean	310	201	85	47	223	14
13	Gauguin, Paul	59	666	14	157	48	18
14	Vlaminck, Maurice de	574	98	127	26	118	7
15	Modigliani, Amedeo	58	1058	7	107	109	31
16	Bonnard, Pierre	209	293	21	55	92	5
17	Braque, George	138	280	35	26	36	61
18	Warhol, Andy	475	56	66	7	139	4
19	Kandinsky, Wassily	62	515	55	192	32	48
20	Utrillo, Maurice	473	132	147	56	30	15
21	Klee, Paul	58	298	157	159	111	30
22	Toulouse-Lautrec, Henri de	55	355	10	678	120	291
23	Sisley, Alfred	105	678	0		17	42
24	Dufy, Raoul	266	118	324	38	250	4
25	Dongen, Kees van	208	191	71	35	38	10
26	Foujita, Tsuguharu	166	150	76	44	190	18
27	Calder, Alexander	62	22	401	8	90	7
28	Bacon, Francis	42	1369	0		1	8
29	Buffet, Bernard	467	72	48	30	71	10
30	Magritte, Rene	108	394	68	118	99	10
31	Lichtenstein, Roy	134	185	1	244	25	51
32	Rouault, Georges	171	155	83	42	11	29
33	Laurencin, Marie	248	126	184	34	140	8
34	Signac, Paul	70	493	264	16	40	11
35	Fantin-Latour, Henri	182	160	0		38	2
36	Chirico, Giorgio de	228	110	34	44	68	10
37	Dali, Salvador	94	222	97	38	203	10
38	Gris, Juan	59	513	12	94	50	28
39	Boudin, Eugene Louis	389	80	74	13	66	9
40	Stella, Frank	389	80	74	13	66	9
41	Nolde, Emil	35	594	231	74	16	15
42	Ernst, Max	173	129	16	104	58	17
43	Derain, Andre	214	28	50	14	109	2
44	Francis, Sam	246	131	78	53	0	
45	Redon, Odilon	59	215	10	50	73	198
46	Corot, Jean Baptiste Camille	247	83	0		42	4
47	Kisling, Moise	357	78	14	9	14	5
48	Morandi, Giorgio	76	357	11	48	45	11
49	Marquet, Albert	223	98	47	9	73	1
50	Kline, Franz	71	212	9	14	39	16
51	Jawlenskij, Alexej von	170	114	15	26	12	13
52	Kirchner, Ernst Ludwig	31	405	62	34	191	7
53	Poliakoff, Serge	163	104	102	25	3	5
54	Martin, Henri	250	68	3	0	9	3
55	Le Sidaner, Henri Eugene Augustin	223	57	4	6	21	11
56	Delvaux, Paul	30	689	28	37	69	21
57	Basquiat, Jean-Michel	240	44	9	8	81	9
58	Nicholson, Ben	128	74	25	31	59	11
Total		10,917		4158		5625	

of the tree simulates the structure of X by means of a certain partition into clusters. Hierarchical clustering algorithms can work either bottom-up, by agglomerating points and clusters of points into new, coarser level nodes, or top-down, by partitioning X recursively until singletons are reached. In the agglomerative approach, the key ingredient is the similarity measure between two clusters. This is computed between any pair of clusters, at each level, in order to decide which two clusters are the most similar. These will be joined into a new cluster. Common measures of similarity between clusters

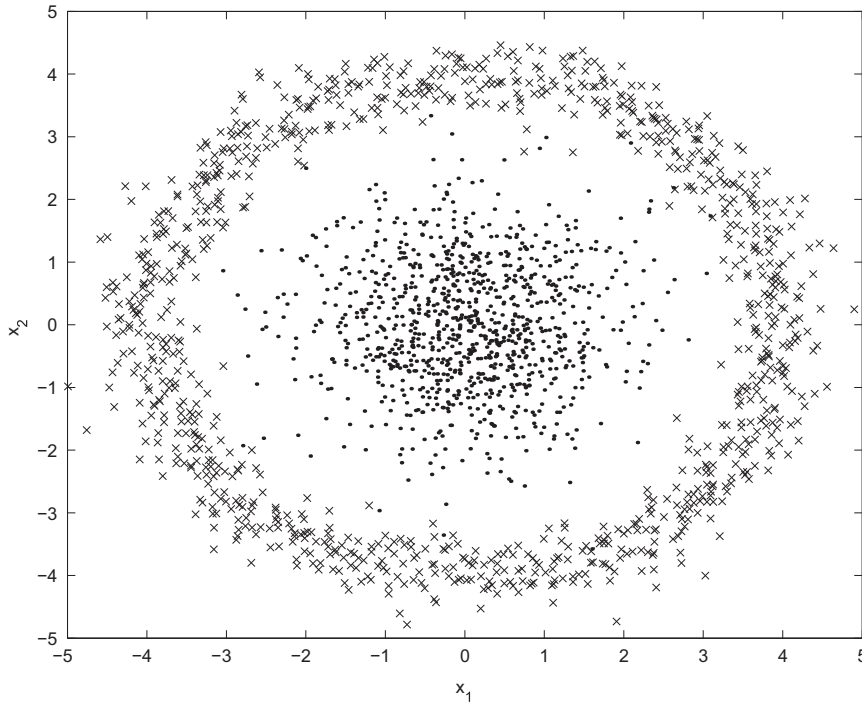


Fig. 1. Set of points with highly non-Gaussian distribution. There are 2000 points that form two separated groups (“ring” and “blob”).

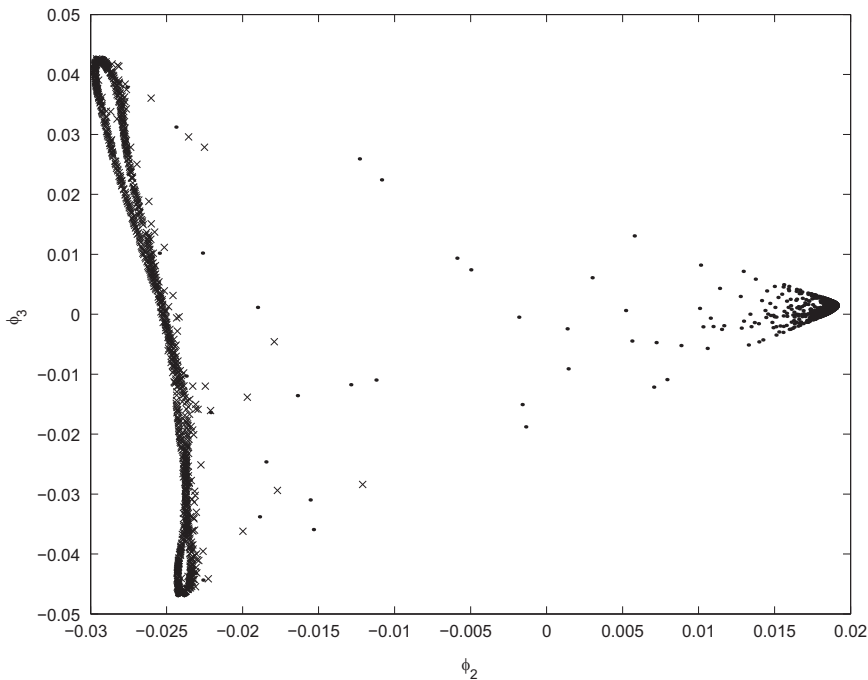


Fig. 2. Decomposition of points with Laplacian eigenmap method. A separation is achieved, with “ring” points forming a line (left part of figure), and “blob” points forming a cone (right part of picture).

include the following:

$$\begin{aligned} \text{Nearest distance: } & \min_{x_i \in X_i, x_j \in X_j} d(x_i, x_j), \\ \text{Average distance: } & \text{ave}_{x_i \in X_i, x_j \in X_j} d(x_i, x_j), \\ \text{Complete distance: } & \max_{x_i \in X_i, x_j \in X_j} d(x_i, x_j). \end{aligned}$$

The third algorithm is the *Laplacian eigenmap algorithm*. It rests upon more complicated mathematics than the previous two algorithms. We give a detailed description of the algorithm in [Appendix A](#). Here, we show how it works with an example: we are given a set of N points, X , each point being a vector of M numbers. This could, e.g., be the value over M consecutive years for N paintings. We wish to divide these into K clusters. For simplicity, let us assume that we have $N=2000$ vectors of length $M=2$, as shown in [Fig. 1](#). It is trivially clear to the eye that there are two distinct clusters of points in [Fig. 1](#). However, as the mean and covariance terms between points in the two clusters are zero, it is also clear that any method based on a normal distribution of points will fail to separate the two clusters, even if means, variances and covariances are known! Another way of stating this, is that there is simply no way to separate the two clusters with a straight line, which is what standard Gaussian methods do. A similar argument also applies to the K -means algorithm. A method for separating the two sets needs to rely on more local measures than what is provided by means and variances. Phrased differently, we need a method that uses the fact that *points that are close to neighbors of a specific point, are close to that specific point too*, whereas not “bothering” too much of the distance of points that are far apart.

The Laplacian eigenmap method starts with a symmetric similarity matrix $Q \in \mathbb{R}^{M \times M}$, where Q_{ij} defines the similarity of points i and j . Thus, a large value of Q_{ij} implies that point i and j are similar. The similarity measure must be accurate for points that are close, but it does not have to be very accurate for points that are far apart as long as it is small – these points will typically not end up in the same cluster anyway. The measure is thus defined in a way such that a *focus on locality* is introduced. We can now view each point as an element in \mathbb{R}^M . In the example, we have embedded each point in \mathbb{R}^2 in \mathbb{R}^{2000} ! The key here is that even though we now have a much larger space to work with, our embedding is a low-dimensional manifold in (ideally, two-dimensional). If we make an eigenvector decomposition of this mapping, the two eigenvectors corresponding to the second and third largest eigenvalues will provide a good representation of this manifold (the first eigenvector does not provide any information at all – it is a constant). The eigenvector decomposition is a *nonlinear map*.

Finally, by looking at the eigenvector representation of a specific point, we can determine which cluster it belongs to. We do this for the points in [Fig. 1](#). The resulting plot of the coefficients for the second and third eigenvectors are shown in [Fig. 2](#). We see that there is a perfect separation between the two sets, and that they can easily be separated by a straight line. In general, clustering algorithms can be expected to be more efficient when working in the image of the points under the Laplacian eigenmaps. A heuristic argument for why this works, and how the similarity mapping should be chosen, is given in [Belkin and Niyogi \(2003\)](#). Now, although there are many other traditional clustering algorithms that would also succeed in separating these two clusters, there are more complicated distributions where the Laplacian eigenmap is superior, e.g., when the two sets are “spirals”, see [Coifman and Lafon \(2006\)](#).

3.3. Results

3.3.1. K-means algorithm

We run K -means clustering with Euclidean distance. After trying different numbers, we arrive at specifying 5 clusters (remember that the number of clusters must be exogenously given to the K -means algorithm). The resulting clusters are shown in [Table 2](#). The resulting clusters are unbalanced – the smallest cluster containing 3 artists and the largest containing 22. Ideally, we would want the first number – the average distance between artists within the same cluster – to be as small as possible, whereas we would want the nondiagonal elements of columns two to six – the minimal distance between artists in two different clusters – to be as large as possible. We see that, except for cluster K1, this is by no means the case for the other clusters. Especially, cluster K5 has an average distance of 40.5 between elements within cluster K5, whereas the

Table 2

Clusters identified with K -means algorithm, K1–K5. The clusters are unbalanced, and it is difficult to visualize how separated they are. The artist associated with respective number is given in [Table 1](#).

Distance	Within	K1	K2	K3	K4	K5	Artists
K1	2.3	–	3.3	6.5	3.9	9.1	2,11,15,18,21,25,26,27,31,32,33, 35,36,39,40,42,43,44,45,50,51,58
K2	8.9	3.3	–	5.7	3.0	7.4	1,6,10,17,30,37,38,41,57
K3	17.2	6.5	5.7	–	5.4	8.4	4,22,48
K4	6.6	3.9	3.0	5.4	–	5.6	5,8,9,12,13,14,16,19,20,24,28, 34,46,47,49,52,53,54,55
K5	40.5	9.1	7.4	8.4	5.6	–	3,7,23,29,56

minimum distance to each of the other clusters is less than 10. This is a clear indicator that this cluster contains scattered elements that did not fit into the other clusters. This suggests that the K-means algorithm is not very well suited for our dataset.

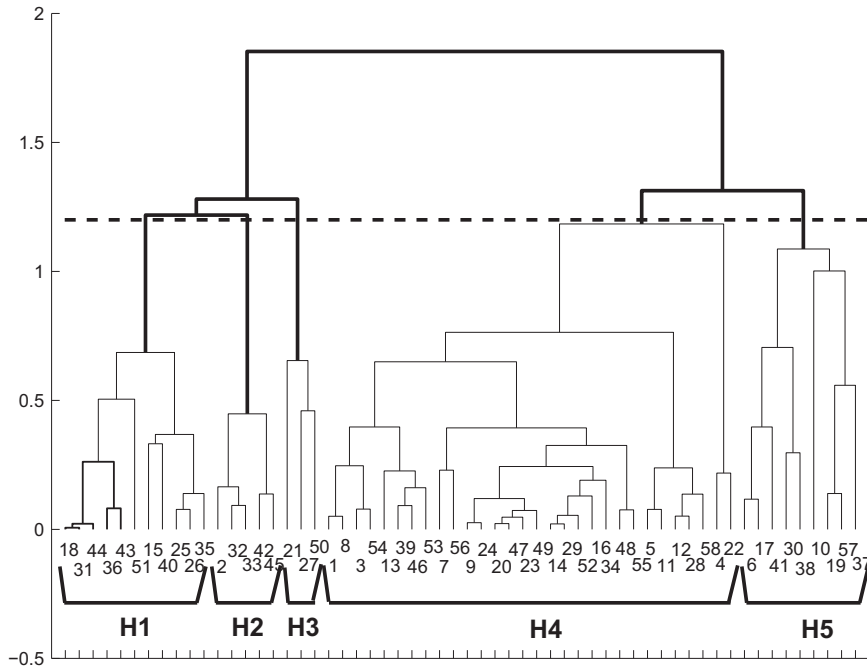


Fig. 3. Tree built by hierarchical clustering algorithm with cosine distance and complete similarity measure. The horizontal line represents the similarity level at which there are 5 clusters: H1–H5, shown in bottom part of figure. The clusters are unbalanced. The artist associated with respective number is given in Table 1.

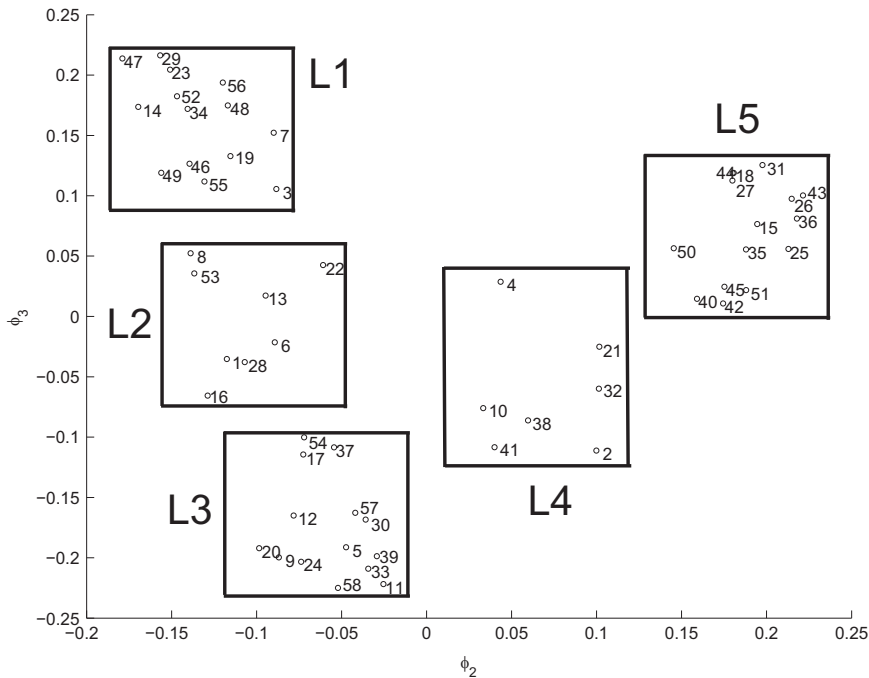


Fig. 4. Clusters identified with Laplacian eigenmap algorithm. The clusters are fairly well balanced. The artist associated with respective number is given in Table 1.

3.3.2. Hierarchical tree algorithm

We next run a hierarchical clustering algorithm. As discussed, there are two degrees of freedom when running this algorithm: the choice of similarity function between points, and the choice of similarity between clusters. We tried several combinations: most led to highly unbalanced trees, with one or two very large clusters, and the rest of the clusters consisting of a single artist. The choice that gave the best result was using the angle (cosine) distance function, and the complete distance among clusters at different levels in the hierarchy. The results with this choice are shown in Fig. 3. The vertical axis describes the distance between different branches in the tree. To separate, e.g., 5 clusters that minimize the distance between elements within a cluster, we should start from the top, and proceed downwards until the tree is split into 5 branches. This level is represented by the horizontal line in Fig. 3. We see that, even with this parameter choice, such a procedure produces even more unbalanced clusters than the K -means algorithm, with cluster H3 containing 3 elements, and cluster H4 containing 30! Thus, hierarchical tree algorithms also seem quite poorly suited for our data.

3.3.3. Laplacian eigenmap algorithm

Finally, we run the Laplacian eigenmap method. The results are shown in Fig. 4. We see that five clearly separated clusters can be identified. These clusters are also identified when we run a K -means algorithm on the transformed two-dimensional dataset. The clusters are relatively well balanced, ranging in size between 7 and 15 artists. Also, we note that the representation in Fig. 4 offers a nice two-dimensional visualization of the dataset (Coifman and Lafon, 2006).

We now have three suggestions for how artists should be clustered. In some cases they agree: For example, the impressionists Claude Monet and Edgar Degas (# 3 and 7 respectively) are put in the same cluster by all three methods. In other cases they do not: For example, the Laplacian eigenmap method groups Vincent van Gogh (# 4) together with Paul Klee (# 21, expressionist) and Pierre August Renoir (# 2, impressionist). The hierarchical tree algorithm on the other hand chooses to put these three artists in separate clusters, rather focusing on the connection between Renoir and Max Ernst (# 42), and between Klee and Alexander Calder (# 27). Finally, the K -means algorithm agrees with the Laplacian eigenmap method that Renoir and Klee should be in the same cluster, but puts Van Gogh in a small cluster together with Henri de Toulouse Lautrec (# 22) and Giorgio Morandi (# 48). Now, the situation is analogous to having three art connoisseurs in the same room, and having as many opinions on what constitutes style. However, the point of our exercise is to try to objectively measure what constitutes style, or at least market substitutes, *from observed price data*. We should therefore be able to pick one of the three methods, *solely based on the price dynamics* and then compare if it relates to what we think of as style.

We would like an objective way of measuring which method gives the best clusters for this dataset. The different degrees of balancedness of the clusters gives us some indication, but we would also like an “objective” measure of how “similar” artists within a cluster are, and how “different” artists in different clusters are. Right now, we are comparing apples and pears, as the three methods have different measures of what constitutes a cluster. Unfortunately, it is almost a *folk theorem* that for any method, one can define a measure of success and a data distribution, such that the method is optimal (for a semi-serious article, see Laloudouana and Tarare, 2003). For example, we compared the methods with their own “measures of success,” i.e., comparing the Laplacian eigenmap clusters with the K -means clusters, *with the K -means algorithm's measure of similarity*, etc. Not surprisingly, each method outperforms the others when its own measure of success is used. We have performed, but do not report the details of, these tests.

A more objective, frequently used, measure is the N -Cut criterion, Shi and Malik (2000). It measures how close points within a cluster are, compared with points in different clusters, by summing distances within clusters, and dividing over total distances, in a way such that a scale-free parameter is achieved. The measure has some optimality implications. Furthermore, it is not immediately related to any of our algorithms. In this sense, the N -Cut measure is like calling for an outsider to judge which of three paintings has the highest quality, instead of letting one of the three painters decide. For our data, the points represent price returns, and a high N -Cut measure means that returns for artists within a cluster are highly correlated, whereas returns for artists in different clusters have low correlation. It therefore fits well with the factor model described in Section 2. We compare the N -Cut-measure for the three algorithms in Table 3. Here, a higher number means that the algorithm is more successful. We see that the K -means algorithm and the Laplacian eigenmap algorithm outperform the hierarchical tree algorithm, and that even though it is a close call, the Laplacian eigenmap algorithm has a higher score than the K -means algorithm. Thus, in addition to the balancedness, we have a quantitative indication that the Laplacian eigenmap algorithm is to prefer. Finally, we prefer the visualization provided by the Laplacian eigenmap algorithm compared with the distance matrices provided by the K -means algorithm. We therefore focus on the clusters identified by the Laplacian eigenmap method.

The aim with this paper is, of course, not to find the ultimate clustering algorithm – rather we wish to show that there is a connection between style as an economist and as an art historian might define it. In the next section, we analyze the

Table 3

Comparing the three algorithms with the N -Cut criterion of success. The Laplacian eigenmap method has highest score, closely followed by the K -means algorithm, both outperforming the hierarchical tree algorithm.

Algorithm	K -means	Hierarchical tree	Laplacian eigenmap
N -Cut (%)	10.2	6.3	10.9

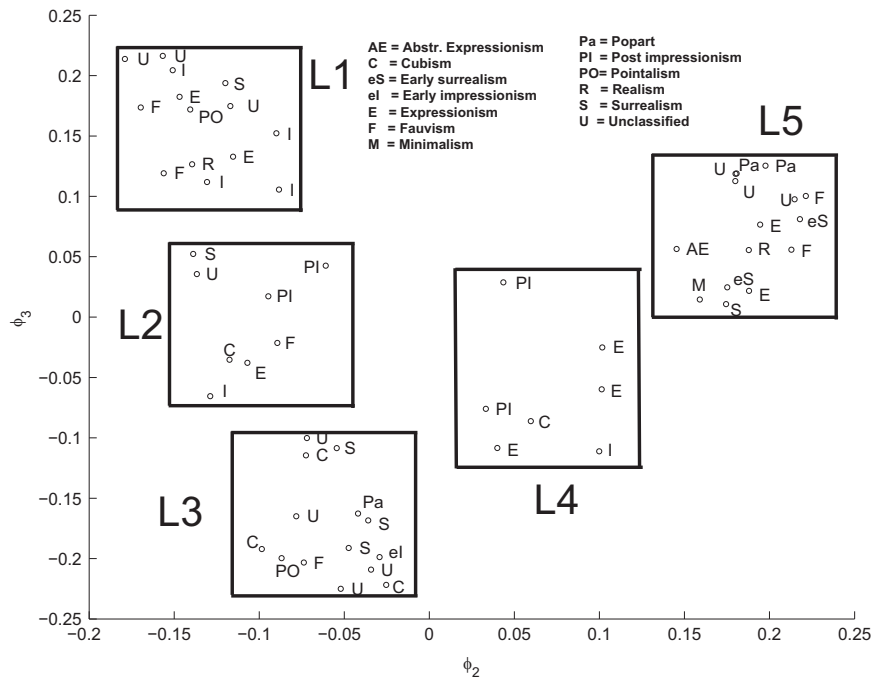


Fig. 5. Style for artists. Hypotheses: Impressionists are over-represented in cluster L1, Post-impressionists are overrepresented in clusters L2 and L4, Expressionists dominate cluster L4, Surrealists are well represented in cluster L3, as are Popart painters in cluster L5.

clusters, L1–L5, identified by the Laplacian eigenmap method, and how they relate to style. For completeness, we also give a brief summary of the (weaker) results for the clusters identified by the *K*-means algorithm (K1–K5) and by the hierarchical clustering algorithm (H1–H5).

3.4. Relationship with style

What do the clusters in Fig. 4 represent? To answer this question, we first study the price dynamics for the different clusters, then we compare the clusters with “style” as it might be classified by an Art historian. There is not much relationship between *price level* and clusters. For example, average price for each artist (shown in Fig. 6 in Appendix B, where size of price is represented by the area of each ring), seems fairly evenly distributed across clusters. *Returns* for different artists on the other hand (shown in Fig. 7 in Appendix B) have a clear relationship with the clusters – from large returns for cluster L1, and then decreasing, with the lowest returns for cluster L5. Thus, average returns partly explain the different clusters.

However, this is just one dimension – let us now see if the clusters do in any sense represent how the art historian might classify style. We naturally have to choose a very simple classification: one comparison is to look at the year of birth for respective artist and see if it is related to our clusters: We would expect artists that were active in the same time-period to be more similar in style. This comparison is shown in Fig. 8 in Appendix B. It seems like clusters L1 and L4 on average contains “older” artists, whereas cluster L5 contains “younger” artists. We test whether these hypotheses bear any statistical significance. We count the number of artists in each group that are older and younger than the median age of the sample respectively. We use a χ^2 -test to see whether these groups have a random distribution. The hypothesis is rejected at the 5% level (Table 5b), and thus the clusters do not seem to have a random age distribution. We also perform a two-sample t-test to see whether the artists in clusters L1 and L4 are significantly older than the rest of the sample. Finally, we test whether the artists in cluster L5 are significantly younger than the rest of the sample. These results are shown in Table 5a. We can conclude that the artists in cluster L5 are younger than the rest of the sample at the 5% significance level. We do not get 5% significance for either of the separate hypotheses that clusters L1 and L4 are older than the rest of the sample, although the hypotheses that artists in clusters L1 and L4 are older than the rest of the sample is supported at the 1% level (not shown in table). Thus, we have at least an indirect indicator that the clusters are related to style.

A direct test can be performed by comparing the clusters with a style classification. We use *Artcyclopedia*³ as a primary source to associate one style with each artist. When no clear classification was given by *Artcyclopedia*, we used *The-Artists.org*⁴ as a secondary source. Artists that were not classified by either of these sources were defined as “Unclassified”. The

³ <http://www.artcyclopedia.com>

⁴ <http://www.the-artists.org>

Table 4

(a) Number of artists in each cluster that are born before and after median year of birth. (b) Number of artists in each cluster that represent different styles.

(a) Cluster	L1	L2	L3	L4	L5
Older than median	4	4	10	1	10
Younger than median	10	4	4	6	5
Total	14	8	14	7	15
(b) Cluster	L1	L2	L3	L4	L5
Style					
Impressionism	4	1	0	1	0
Post-impressionism	0	2	0	2	0
Expressionism	2	1	0	3	2
Surrealism	1	1	3	0	1
Popart	0	0	1	0	2
Other	7	3	10	1	10
Total	14	8	14	7	15

Table 5Statistical tests of style distribution of identified clusters: (a) two-sample *t*-test for the year of birth of artists in different clusters. (b) χ^2 -test for the relationship between age, styles and different clusters.

(a) Hypothesis	χ^2 / <i>t</i> -statistic	Supported with <i>p</i> -value	
Cluster L1 born earlier than rest of sample	– 1.64	0.10	
Cluster L4 born earlier than rest of sample	– 1.50	0.14	
Cluster L5 born later than rest of sample	1.97	< 0.05*	
(b) Hypothesis	χ^2 -statistic	Degrees of freedom	Supported with <i>p</i> -value
Age-distribution is not random	10.38	4	< 0.05*
Style-distribution is not random			< 0.007**
Cluster L1 more Impressionists	6.61	1	< 0.025*
Cluster L2 more Post-impressionists	5.82	1	< 0.025*
Cluster L3 more Surrealists	2.44	1	< 0.2
Cluster L4 more Expressionists	5.66	1	< 0.025*
Cluster L4 more Post-impressionists	4.74	1	< 0.05*
Cluster L5 more Popart	2.75	1	< 0.1

* Corresponds to a significance level of 5%.

** Corresponds to a significance level of 1%.

resulting (admittedly very rough) classification is shown in Fig. 5. The results are mixed, but there seems to be some structure:

- Impressionists are mainly in cluster L1.
- Cluster L2 is small, but contains two out of four Post-impressionists, as does cluster L4.
- Cluster L3 is mixed, but has a high representation of Surrealists.
- Expressionists dominate cluster L4.
- Two out of three Popart painters are in cluster L5.

We use an exact contingency test⁵ with the null hypothesis that our identified structures are random. In doing this, look at the following styles: *Impressionism*, *Post-impressionism*, *Expressionism*, *Surrealism* and *Popart*. Artists that do not belong to any of these are grouped into *Other*. A frequency table is shown in Table 4b. The hypotheses and the significance level that the respective alternative hypothesis can be rejected at are shown in Table 5b. We see that we can reject that the clusters are random at the 1% level. The hypotheses on Impressionists, Post-impressionists and Expressionists are all supported at the 5% level or better (i.e., the alternative hypothesis is rejected). The hypotheses on Surrealists being over-represented in cluster L3, and Popart painters in cluster L5 are not significant.

⁵ Rather than a χ^2 -test, as the frequency table contains several zeros.

We also perform the same tests with the clusters identified by the K -means algorithm. The results are somewhat weaker – summarized in what follows: the test does not reject that the clusters are random with respect to *style* at the 5% level, although it is close ($p=0.052$). Neither does it, reject randomness with respect to *age* ($\chi^2 = 3.6$ with 4 degrees of freedom). Furthermore, no cluster has a t -statistic higher than 1.15 when testing whether age distribution of the specific cluster differs from the rest. For styles within individual clusters, the two small clusters (K3 and K5) contains statistically significant overrepresentation of Impressionists and Post-Impressionists respectively – similar to what was found by the Laplacian eigenmap method. However, no relations (not even weak) are identified for Expressionists, Popart and Surrealism, with highest χ^2 statistics of about 1.2. Thus, it seems like the K -means algorithm does a decent job in separating out Impressionists and Post-Impressionists, in the two small groups K3 and K5, but fails to separate styles among the rest.

Finally, we test the clusters identified by the hierarchical algorithm. Here, we are not even close to reject randomness of either age or style. The p -value for randomness of style is $p=0.30$, and for randomness of age, $\chi^2 = 1.6$ with 4 degrees of freedom. Thus, there does not seem to be any relationship between style and the clusters identified by the hierarchical tree algorithm.

4. Conclusions

We have empirically verified a connection between price dynamics and style of works of art, using a clustering algorithm. The Laplacian eigenmap algorithm identified five clusters (L1–L5) and we could reject that these clusters were random with respect to style and age of artist. Furthermore, we found support for:

1. Impressionists being overrepresented in cluster L1.
2. Post-impressionists being overrepresented in clusters L2 and L4.
3. Expressionists being overrepresented in cluster L4.

Our results should, of course, not be over-stated. We realize that we are using *post hoc* hypotheses that were identified from the data. A Bonferroni correction of the p -values for the individual hypotheses (in Table 5), to take this into account, would destroy statistical significance, except for the overall hypotheses that the clusters are not random over age and style. Moreover, the clusters are noisy and the correspondence with what we call style is far from one-to-one. Our interpretation is therefore that there is a relationship between price dynamics of artwork and style, and that a richer data-set would permit us to further explore and refine this relationship.

Acknowledgment

We thank ArtNet, Inc. for the use of historical data for research.

Appendix A. Description of Laplacian eigenmap method

We are given a set of points and their pairwise distances representing similarities (the smaller the distance, the more similar two points are). These distances are assumed to be rather accurate when they are small (i.e. the two objects being compared are very similar) but they may become unreliable when they are big (i.e. the two objects being compared are quite not similar). This is a rather common situation in many applications.

We look for a representation of the set of points in as low a dimension as possible, while preserving the important, reliable distances (the small ones), and allowing distortion for less-important distances (the large ones). Geometric harmonics (Coifman and Lafon, 2006) and the related eigenmaps are non-linear maps that try to preserve local distances while in general distorting non-local distances. Moreover, these maps have the property of pulling clusters apart, automatically detecting possible good cuts for separating different clusters. Geometric Harmonics have a scaling parameter, which allows for looking at different level of details in the data set.

We want to look at the geometric harmonics on a data set $X \subset \mathbb{R}^N$, each point being, in our particular case, one of the time series described above. We get the Laplacian eigenmap when we choose kernel

$$K(x, y) = e^{-\|x-y\|^2/\delta^2} \quad (2)$$

for some choice of δ and normalized as follows: let

$$D(x) \stackrel{\text{def}}{=} \sum_{y_i} K(x, y_i), \quad (3)$$

and consider the (normalized) Laplacian operator on the data set given by

$$L = D^{-1}(D - K). \quad (4)$$

We compute the eigenvalues and eigenvectors of this operator:

$$L\varphi_i = \lambda_i\varphi_i, \tag{5}$$

for $i = 1, 2, \dots$, where the λ_i 's are ordered in nondecreasing order. It is always the case that $\lambda_1 = 0$ (since the operator is identity minus averaging). The first few eigenvectors are particularly interesting. For any fixed k we can consider the map

$$\Phi_k: X \rightarrow \mathbb{R}^k, \text{ defined by } \Phi_k(X) = (\varphi_2(x), \dots, \varphi_{k+1}(x)). \tag{6}$$

This map minimizes the distortion as defined by

$$N - \text{tr}(\varphi K \varphi^T) = \sum_{i=2}^{k+1} (1 - \lambda_i) = \frac{1}{2} \sum_{i=2}^{k+1} \sum_{x,y} (\varphi_i(x) - \varphi_i(y))^2 K(x,y), \tag{7}$$

among all “projection maps” (AKA^T with A orthogonal and $AKA^T = I$). Since the kernel $K(x,y)$ is larger for x,y close, the emphasis in the distortion of Φ_k is in keeping close points close.

Observe that the matrix $P = D^{-1}K$ is a Markov matrix, whose (i,j) -entry defines the transition probability of jumping from point x_i to point x_j , thus defining a random walk on X , which we think of as a heat diffusion. Also, observe that the eigenvectors of P are exactly $\{\varphi_i\}_i$ and that the eigenvalues of P are just $1 - \lambda_i$. Hence the eigenvectors $\{\varphi_i\}_i$ are also the eigenfunctions of the heat operator on X . Now observe that two clusters in X would by definition be weakly linked together, where weakly is measured with respect to the strength of the connections between points inside each cluster. Then the heat diffusion will be slow along the links connecting different clusters, compared to the speed of diffusion inside the clusters. We expect the second eigenfunction φ_2 of the heat operator to have its 0-level set in the middle of these weak connections, and be of different sign on different clusters. Along these lines, it can for example be proved that the second eigenfunction φ_2 arises from the relaxation of a (NP-hard) clustering problem (Shi and Malik, 2000). In the example in Fig. 1, the eigenfunction φ_2 is negative in the core cluster and positive in the annulus around it, as one can clearly see in Fig. 2, and its 0 level set is a good cut in the graph determined by the set of points.

Appendix B. Clusters vs styles – additional figures from Section 3.4

See Figs. 6–8.

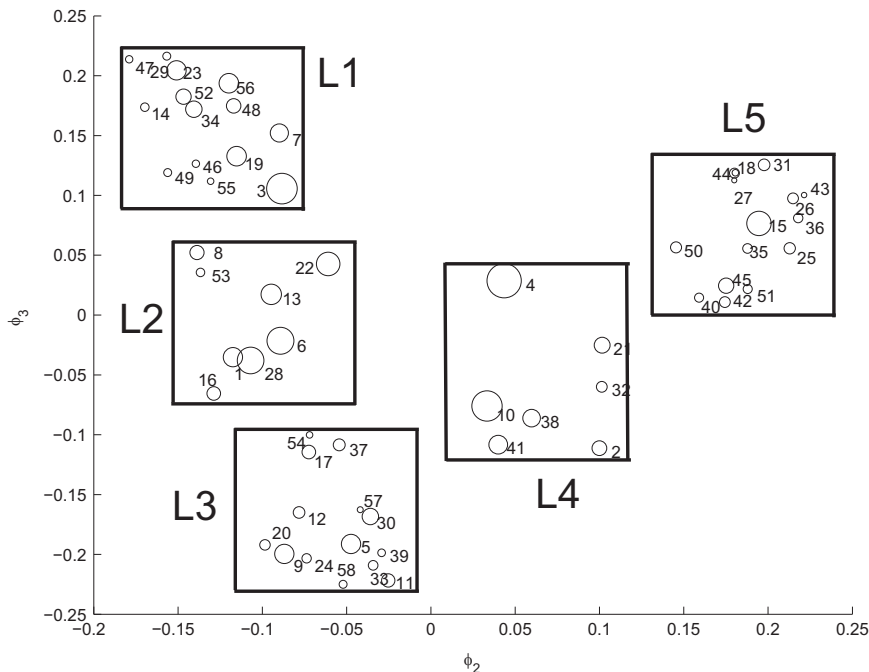


Fig. 6. Average prices over ten year period. There is no clear relationship between prices and clusters. The artist associated with respective number is given in Table 1.

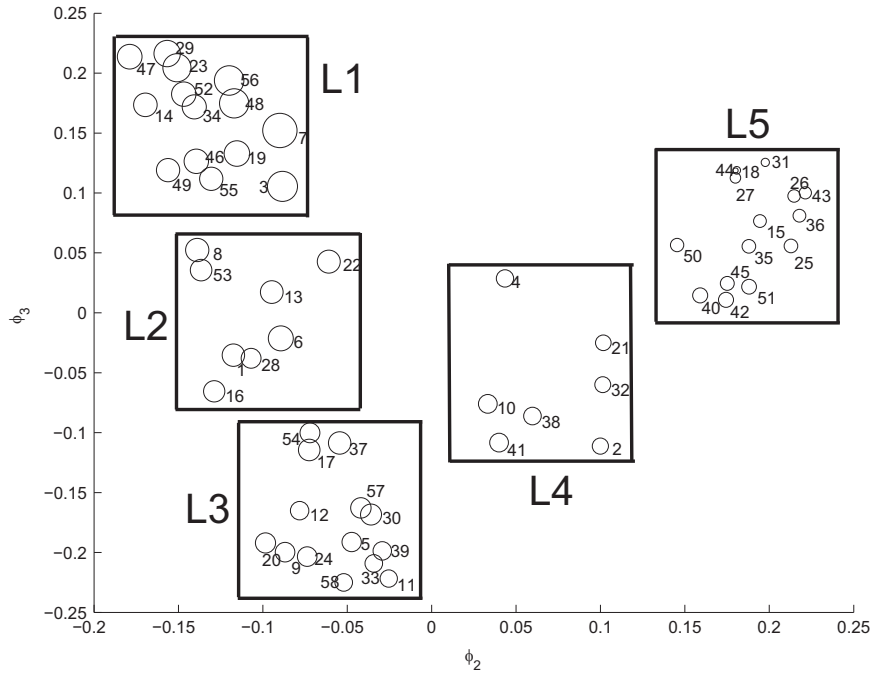


Fig. 7. Average returns over ten year period. There is a clear relationship between prices and clusters. Basically, the average return increases in the second eigenvector ϕ_2 . The artist associated with respective number is given in Table 1.

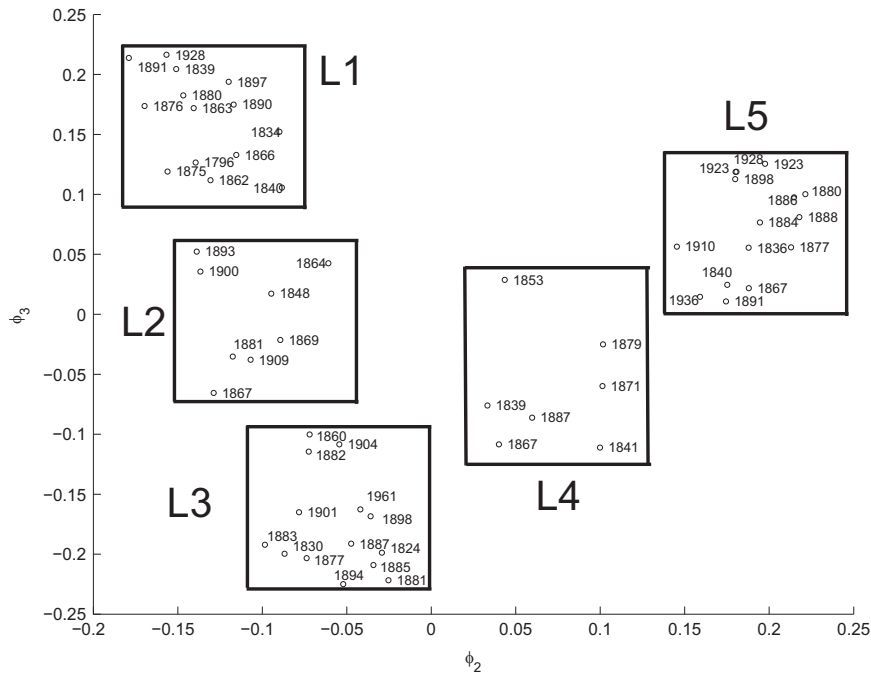


Fig. 8. Year of birth for artists: Clusters L1 and L4 contain “Old artists”, cluster L5 contains “young artists”. Average birthyear: Cluster L1 – 1867, Cluster L2 – 1878, Cluster L3 – 1883, Cluster L4 – 1864, Cluster L5 – 1891.

References

Ait-Sahalia, Y., Parker, J.A., Yogo, M., 2004. Luxury goods and the equity premium. *J. Finance* 59, 2959–3004.
 Anderson, R.C., 1974. Paintings as an investment. *Econ. Inq.* 12, 13–26.
 Barberis, N., Schelifer, A., 2003. Style investing. *J. Financ. Econ.* 68, 161–199.
 Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 1376–1393.

- Brown, S.J., Goetzmann, W.N., 1997. Mutual fund styles. *J. Financ. Econ.* 43, 373–399.
- Chan, L.K.C., Chen, H., Lakonishok, J., 2002. On mutual fund investment styles. *Rev. Financ. Stud.* 15, 1407–1437.
- Coifman, R.R., Lafon, S., 2006. Geometric harmonics. *Appl. Comput. Harmon. Anal.* 21, 31–52.
- Donoho, D.L., Grimes, C., 2003. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci.* 100, 5591–5596.
- Galenson, D.V., 2002. *Painting Outside the Lines: Patters in Creativity in Modern Art*. Harvard University Press, Cambridge.
- Goetzmann, W., 1993. Accounting for taste: an analysis of art returns over three centuries. *Am. Econ. Rev.* 83, 1370–1376.
- Goetzmann, W., Spiegel, M., 1995. Price value components and the winner's curse in an art market. *Eur. Econ. Rev.* 39, 549–555.
- Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. *ACM Comput. Surv.* 31, 264–323.
- Korteweg, A., Kräussl, R., Verwijmeren, P., 2015. Does it pay to invest in art? A selection-corrected returns perspective. *Rev. Financ. Stud.*, forthcoming, <http://dx.doi.org/10.1093/rfs/hhv062>.
- Kumar, A., 2002. Style switching and stock returns. Working paper, University of Notre Dame.
- Kumar, A., Lee, C.M.C., 2006. Retail investor sentiment and return comovement. *J. Finance* 61, 2451–2486.
- Laloudouana, D., Tarare, M.D., 2003. Data set selection. Conference paper, NIPS 2003.
- Lovo, S., Spaenjers, C., 2014. Unique durable assets. Working paper, HEC Paris.
- Mei, J., Moses, M., 2002. Art as an investment and the underperformance of masterpieces. *Am. Econ. Rev.* 92, 1656–1668.
- Peckham, M., 1967. In: Patrick Wilkinson (Ed.), *Man's Rage for Chaos: Biology, Behavior and the Arts*, Maisonneuve Press, University Park Maryland, Reprint of the 1967 edition.
- Renneboog, L., Spaenjers, C., 2013. Buying beauty: on prices and returns in the art market. *Manag. Sci.* 59, 36–53.
- Ross, S., 1976. The arbitrage theory of capital asset pricing. *J. Econ. Theory* 13, 341–360.
- Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326.
- Sharpe, W.F., 1992. Asset allocation: management style and performance measurement. *J. Portf. Manag.* 18, 7–19.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 888–905.
- Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290 (5500), 2319–2323.