

1 Notation & Definitions

In this section we introduce a mathematical description of the visualization pipeline where artist \mathcal{A} functions transform data \mathcal{E} to an intermediate representation in a prerendered display space \mathcal{H} .

$$\mathcal{A} : \mathcal{E} \rightarrow \mathcal{H} \quad (1)$$

We use fiber bundles[7, 18] to model data and graphics because they are products of a topological total space that encodes points and a topological base space that can encode the connectivity of those points:

- E is a locally trivial fiber bundle over K representing data space.
- H is a fiber bundle over S representing visual space
- K and S are a triangulizable topological space or a CW complex encoding the connectivity of points in E and H respectively

The fiber bundles mentioned in this work are assumed to be locally trivial[11, 20].

1.1 Data Fiber Bundles

As proposed by Butler [3, 4], we model data as a fiber bundle (E, K, π, F)

$$F \hookrightarrow E \xrightarrow{\pi} B \quad (2)$$

with topological total space E , base space K , fiber space F , and the map from total space to base space $\pi : E \rightarrow K$. Maps from K to E are called sections and select specific points in K . The global space of sections in E is $\Gamma(E)$.

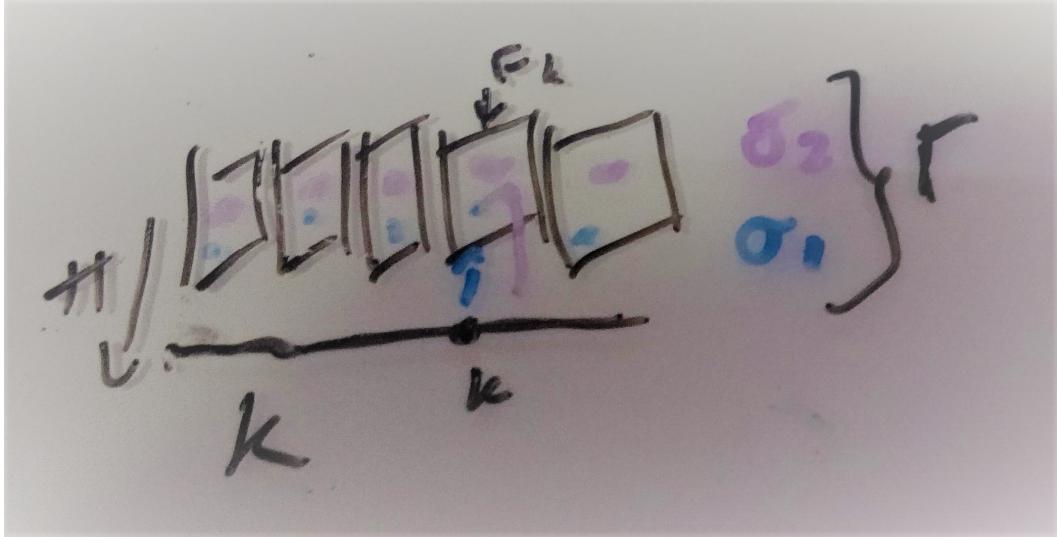


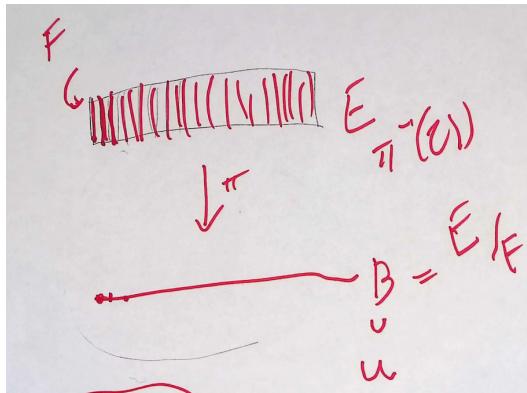
Figure 1: F is an $n \times m$ space of points. The section τ_1 returns the blue points, while τ_2 returns the purple points. $\Gamma(E)$ is the set of all sections, including τ_1 and τ_2

example The fibers F in figure ?? lie in some total space E and are points that lie in a 2D space. For example, this space can be temperature \times pressure or time \times precipitation. K is a line on the interval $[0,1]$ and indicates that the points lie on a 1D continuous space, but we derive whether the data is a distribution or a timeseries or a function from the fields in F . The sections τ_1 and τ_2 are distinct maps from k to the blue and purple points in E respectively.

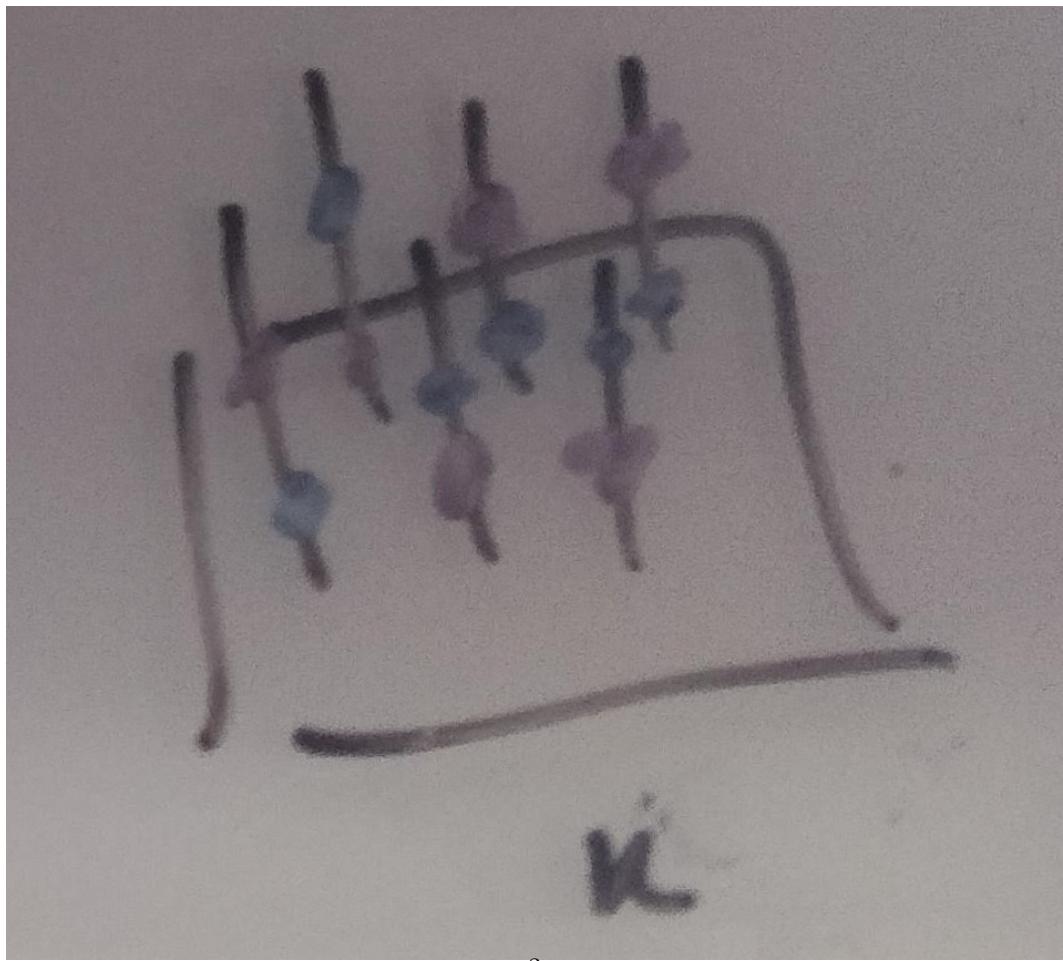
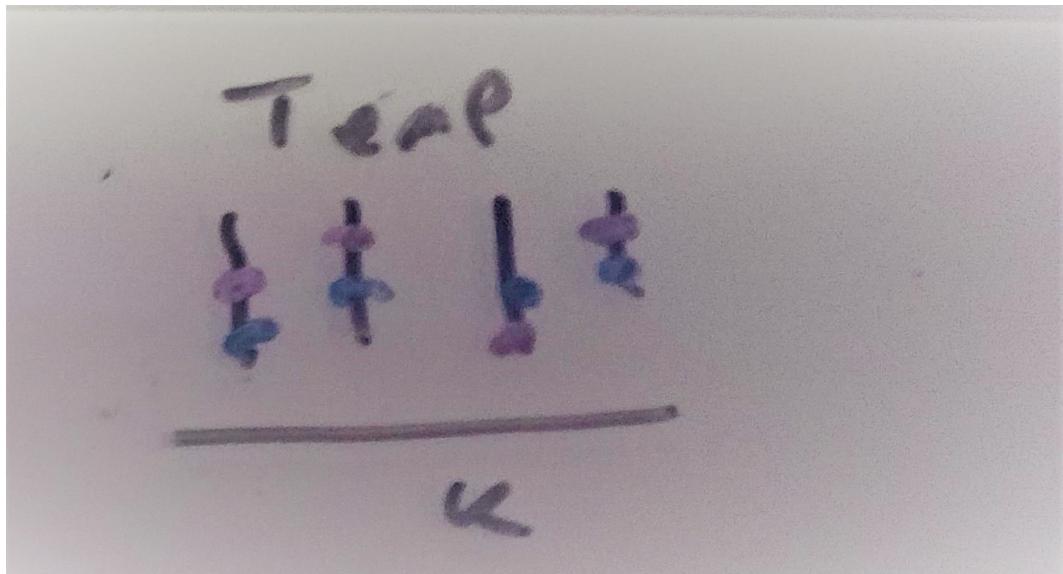
1.1.1 Base Space

Figure 2: The topological base space K encodes the connectivity of the data space, for example if the data is independent points or a map or on a sphere

Data can be discrete observations, timeseries, maps, fields [15] and K is a set of points k that can act as keys from a representative space, such as seen in figure ??, to the data values in E [15].



K and F are not intrinsic to E , rather they are choices in how E is subdivided[16]. In figure ?? we can divide a rectangular base space such that there is a short fiber and long base space or a long fiber and short base space. This is analogous to long and wide forms of the same table [23].



Example in figure ??, temperature is the only one data field in r but the K base spaces are different. subfig[1] is a timeseries, so the temperature in r at time t is dependent on the temperature in r_{t-1} and the temperature in r_{t+1} is dependent on r_t ; this connectivity is expressed as a one dimensional K where K is the number line. In the case of the map, every temperature in r is dependent on its nearest neighbors on the plane, and one way to express this is by encoding K as a plane. K does not know the time or latitude or longitude of the point as those are metadata variables describing the k rather than the value of k . The mapping $\tau : K \rightarrow E$ provides the binding between the key $k \in K$ and the value r in E [14].

1.1.2 Triangulation

The base space K is a representation of the connectivity of the data, specifically whether the points in E are discrete or sampled from a continuous space. The same dataset can be expressed with different K .

In our draft implementation of the data as fiber bundle model, we represent K as a simplicial complex.

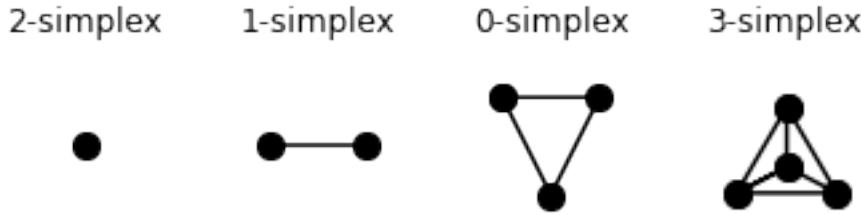


Figure 3: Simplices encode the connectivity of the data, from fully disconnected (0 simplex) observations to all observations are connected to at least 3 other observations. Higher order simplices are outside the scope of this paper.

K is a triangulizable topological space; one triangulization scheme is as a set composed of simplices, such as those shown in figure ??.

Example chopping up a torus maybe? talk about how that gets unpacked into triangles and then into vertices

1.1.3 Fiber Space

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\varphi} & U \times F \\ \pi \downarrow & \swarrow \text{proj}_U & \\ U & & \end{array} \quad (3)$$

such that $\varphi : \pi^{-1}(U) \rightarrow U \times F$ is a homeomorphism where π and proj_U both map to U and the fiber over k

$F_b = \pi^{-1}(b \in B)$ is homomorphic to the fiber F .

Definition A monoid[12] M is a set that is closed under an associative binary operator $*$ and has an identity element $e \in M$ such that $e * a = a * e = a$ for all $a \in M$. A left monoid action [1, 19] of M on a set X is a set \bullet with the properties:

closure $\bullet : M \times X \rightarrow X$,

associativity for all $m, t \in M$ and $x \in X$, $m \bullet (t \bullet x) = (m \bullet t) \bullet x$

identity for all $x \in X$, $e \in M$, $e \bullet x = x$

Example

1.1.4 Section

The section τ is the mapping from base space to total space $\tau : K \rightarrow E$

$$\begin{array}{ccc} F & \hookrightarrow & E \\ \pi \downarrow & \nearrow f & \\ B & & \end{array} \quad (4)$$

such that f is the right inverse of π

$$\pi(f(b)) = b \text{ for all } b \in B \quad (5)$$

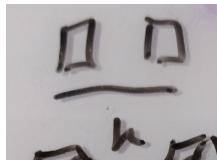
In a trivial fiber bundle, $E = K \times F$ [7, 18]:

$$f(b) = (b, g(b)) \quad (6)$$

where the domain of $g(b)$ is F_b and returns a point p in F_b . The space of all possible sections f of E is $\Gamma(E)$. All sections $f \in \Gamma(E)$ have the same fibers F and connectivity B .

Example For each field $c \in C$, the record function $r : C \rightarrow U_\sigma$ returns an object of type $\sigma(c) \in DT$. The set of all records $\Gamma(\sigma)$ is the set of all sections on U_σ . Spivak defines the τ mapping from an index of databases K to records $\Gamma(\sigma)$ as $\tau : K \rightarrow \Gamma(\sigma)$. This is equivalent to $\tau : k \rightarrow E$ since $F = \Gamma(\sigma)$ and F is the embedding in E on which the records r lie.

1.1.5 Example



The fiber in figure ?? is the space of possible temperature values in degrees celsius, such that $F = [temp_{min}, temp_{max}]$ and is named Temp. In figure ?? time is encoded as a second dimension. This means that the set of possible values F with $C = \{\text{Temp}, \text{Time}\}$:

$$F = [temp_{min}, temp_{max}] \times [time_{min}, time_{max}] \quad (7)$$

and the function τ that retrieves records from F is

$$\tau(k) = (k, (r : \text{Temp} \rightarrow temp, r : \text{Time} \rightarrow time)) \quad (8)$$

$$temp \in [temp_{min}, temp_{max}], time \in [time_{min}, time_{max}] \quad (9)$$

Since $\tau(k) = (k, r)$, *temp* is bound to a named data field and *sigma* binds *temp* to a temperature data type.

1.1.6 Sheaf and Stalk

As described in equation ??, there is a local space $U \subset B$ around every b . The inclusion map $\iota : U \rightarrow B$ can be pulled back such that $\iota^* E$ is the space of E restricted over U .

$$\begin{array}{ccc} \iota^* E & \xleftarrow{\iota^*} & E \\ \pi \downarrow \lrcorner & \lrcorner \iota^* f & \pi \downarrow \lrcorner \\ U & \xleftarrow{\iota} & B \end{array} \quad (10)$$

The localized section of fibers $\iota^* f : U \rightarrow \iota^* E$ is the sheaf $\mathcal{O}(E)$ with germ of $\xi^* f$ which is the localized section. The neighborhood of points the sheaf lies over is the stalk \mathcal{F}_b [20, 21] which is a Restricting the artist to the sheaf means the artist knows the data in F and also has access to derivatives of the data. This property is useful for some visual transformations.

1.2 Prerender Space

We define a graphic space H such that we do not have to assume the physical output space of the renderer. This means that the graphic in H can be output to a screen or 3D printed space or a dome.

We model the prerender space as a fiber bundle (H, S, π, D) . H is the predisplay space, with a fiber D dependent on the target display and a base space of S .

1.2.1 Base space

The underlying topology S of a graphic often needs more dimensions than the data topology K because of the specifications of the display space. For example, a line plot on a plane (such as a screen or a piece of paper) by necessity needs to also have a thickness so that it is visible, which maps back to a set of connected points in H . The topology of these connected points is therefore the region $s \subset S$ such that $\xi : S \rightarrow K$ is a deformation retraction [17]

$$\begin{array}{ccc} E & & H \\ \pi \downarrow & & \pi \downarrow \\ K & \xleftarrow{\xi} & S \end{array} \quad (11)$$

that goes from a region $s \in S_k$ to its associated point k , such that when $\xi(s) = k$, $\xi^*\tau(s) = \tau(k)$.

1.2.2 Fiber and Section

A section $\rho : S \rightarrow H$ is a mapping from a region s on a mathematical encoding of the image to a region xy on the screen that the renderer then maps to visual space as defined in D.

Example For a physical screen display, we can consider a predisplay space that is a trivial fiber bundle $H = \mathbb{R}^5 \times S$ such that ρ is

$$\rho(s) = [x(s), y(s), r(s), g(s), b(s)] \quad (12)$$

To draw an image, a region, H is inverse mapped into a region $s \in S$ where

$$s = \rho_{XY}^{-1}(xy) \quad (13)$$

such that the rest of the fields in \mathbb{R}^7 are then integrated over s to yield the remaining fields:

$$r = \iint_s \rho_R(s) ds^2 \quad (14)$$

$$g = \iint_s \rho_G(s) ds^2 \quad (15)$$

$$b = \iint_s \rho_B(s) ds^2 \quad (16)$$

Here we assume a single opaque 2D image such that the z and *alpha* fields can be omitted. To support overplotting and transparency, we consider $D = \mathbb{R}^7$

1.2.3 Example

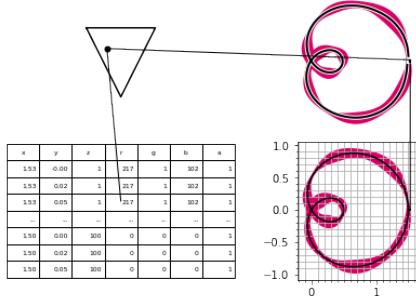


Figure 4

As illustrated in figure 4, words.

1.3 Artist

In this section we will define the artist as a mapping from a sheaf $\mathcal{O}(E)$ to $\mathcal{O}(H)$.

$$A : \mathcal{O}(E) \rightarrow \mathcal{O}(H) \quad (17)$$

The artist decomposes to mapping data to visual $\nu : E \rightarrow V$, then compositing V pulled back along ξ to ξ^*V to a visual mark in prerender space $Q : \xi^*V \rightarrow H$.

$$\begin{array}{ccccc} E & \xrightarrow{\nu} & V & \xleftarrow{\xi^*} & \xi^*V \xrightarrow{Q} H \\ & \searrow \pi & \downarrow \pi & \xi^* \pi \downarrow & \swarrow \pi \\ & & K & \xleftarrow{\xi} & S \end{array} \quad (18)$$

The pullback map ξ^* copies each value in V over k to s in corresponding S_k such that ξ^*V can have multiple values that map to one value in V .

The visual fiber bundle (V, K, π, P) has section $\mu : V \rightarrow K$ that resolves to a visual variable [2, 13] in fiber P . The visual transformer ν is a set of functions each targeting a different μ

$$\{\nu_0, \dots, \nu_n\} : \{\tau_0, \dots, \tau_n\} \mapsto \{\mu_0, \dots, \mu_n\} \quad (19)$$

where μ_i are the visual parameters in the assembly function $Q(\mu_0, \dots, \mu_n)(s) = \rho(s)$.

1.3.1 Example: Matplotlib Visual Fiber

For example, for Matplotlib [8], some of the possible types in P are:

ν_i	μ_i	$\text{codomain}(\nu_i)$
position	x, y, z, theta, r	\mathbb{R}
size	linewidth, markersize	\mathbb{R}^+
shape	markerstyle	$\{f_0, \dots, f_n\}$
color	color, facecolor, markerfacecolor, edgecolor	\mathbb{R}^4
texture	hatch	\mathbb{N}^{10}
	linestyle	$\{f_0, \dots, f_n\} \times (\mathbb{R}, \mathbb{R}^{+n, n \% 2 = 0})$

1.3.2 Visual Channels

$\nu : E \rightarrow V$ is an equivariant map such that there is a homomorphism from left monoid actions on E_i to left monoid actions on V_i where i identifies a field in the fiber. E_i and V_i each contain a set of values as defined in F and P respectively. A validly constructed ν is

one where the diagram

$$\begin{array}{ccc} E_i & \xrightarrow{\nu_i} & V_i \\ m_e \downarrow & & \downarrow m_v \\ E_i & \xrightarrow{\nu_i} & V_i \end{array} \quad (20)$$

commutes such that $\nu_i(m_e(E_i)) = m_v(\nu_i(E_i))$.

Example: Partial Order To preserve ordering of elements in E_i , ν must be a monotonic function such that given $e_1, e_2 \in E_i$

$$\text{if } e_1 \leq e_2 \text{ then } \nu(e_1) \leq \nu(e_2) \quad (21)$$

Example: Translation fairly certain I lost the thread here According to Stevens, interval data is a set with general linear group actions [10, 22]. Position is a visual variable that can support translation [2, 9, 13].

$$\nu(x + c) = \nu(x) + \nu(c) \quad (22)$$

Example: Invalid ν Given a transform $t(x) = x + 2$, we construct a ν that always takes data to .5:

$$\begin{array}{ccc} E_1 & \xrightarrow{\lambda:e \mapsto .5} & V_i \\ 2e \downarrow & & \downarrow 2v \\ E_1 & \xrightarrow{\lambda} & V_1 \end{array} \quad (23)$$

This ν is invalid because the graph does not commute for t :

$$\nu(t(e)) \stackrel{?}{=} t(\nu(e)) \quad (24)$$

$$.5 \stackrel{?}{=} t(.5) \quad (25)$$

$$.5 \neq 2 * .5 \quad (26)$$

To construct a valid ν , the diagram must commute for all monoid actions on the sets E_i, V_i .

1.3.3 Assembling Marks

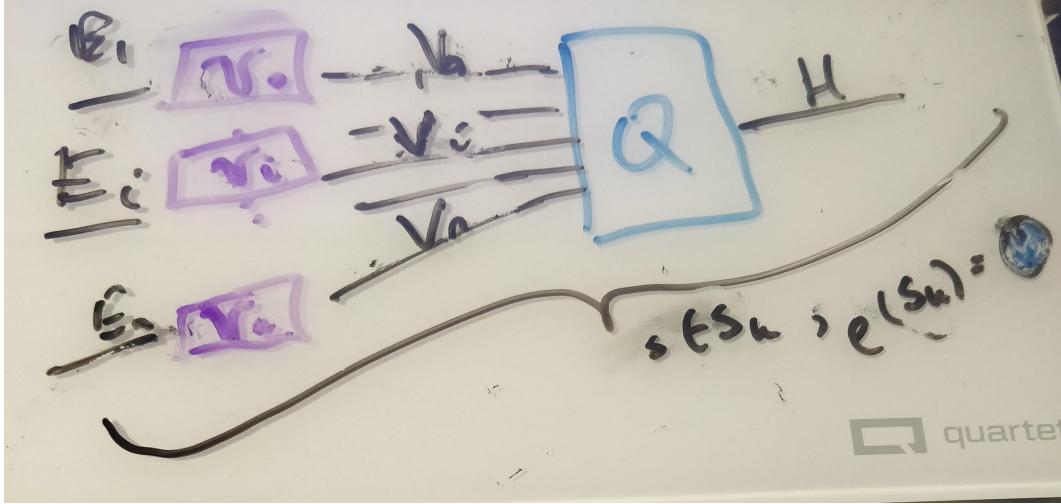


Figure 5: The ν functions convert data E to visual V . Q assembles the different types of visual parameters V_i into a graphic in H . $Q \circ \mu(\xi^{-1}J)$ forms a visual mark by applying Q to a region mapped to connected components $J \subset K$.

As shown in figure ??, Q takes the individual fields in V as input and outputs a single piece of a graphic on H . As with ν , the constraint on Q is that for every monoid actions on the input in V there is a corresponding monoid action on the output in H .

Proposition If $\forall g \in M$ and $\forall x_1, x_2 \in \Gamma(V)$ then $Q(x_1) = Q(x_2)$ implies $Q(g \circ x_1) = Q(g \circ x_2)$; therefore we can define a group action on $Q(\Gamma(V)) = Y$ as $g \circ y = y'$ where $y' = Q(g \circ x)$ with $x \in f^{-1}(y)$

To output a mark [2, 5], Q is called with all the regions s that map back to a set of connected components $J \subset K$:

$$J = \{j \in K \text{ s. t. } \exists \gamma \text{ s.t. } \gamma(0) = k \text{ and } \gamma(1) = j\} \quad (27)$$

where the path[6] γ from k to j is a continuous function from the interval $[0,1]$.

We define the mark as the graphic generated by $Q(S_j)$:

$$H \xrightleftharpoons[\rho(S_j)]{} S_j \xrightleftharpoons[\xi^{-1}(J)]{} J_k \quad (28)$$

where the set $j \subset J$ is the set of marks in the graphic.

1.3.4 Visual Idioms: Equivalence class of artists

in $O(E)$ of the same type, they output the same type of prerender $O(H)$:

Natural transformation + composition is partial ordering? Back and forth is equivalent