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 $\mathcal{H}$.

$$\mathcal{A} : \mathcal{E} \to \mathcal{H} \tag{1}$$

We use fiber bundles[7, 20] 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[12, 22].

1.1 Total Space $\mathcal{E}$

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

$$F \hookrightarrow E \xrightarrow{\pi} K \tag{2}$$

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

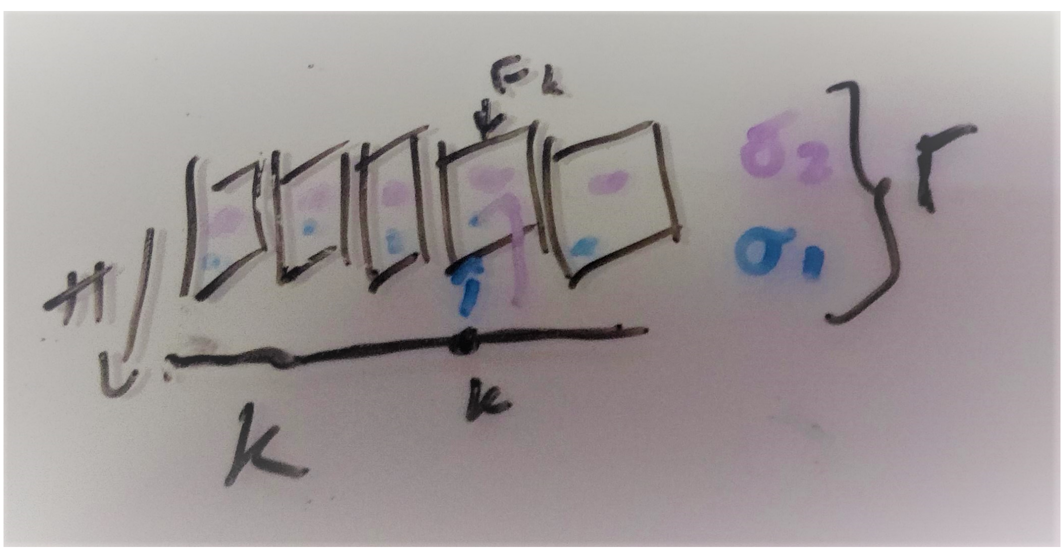


Figure 1: F is an $n \times m$ space of points. The section τ_1 returns the blue points, while τ_2 returns the purple points. $\Gamma(E)$ is the set of all sections, including τ_1 and τ_2

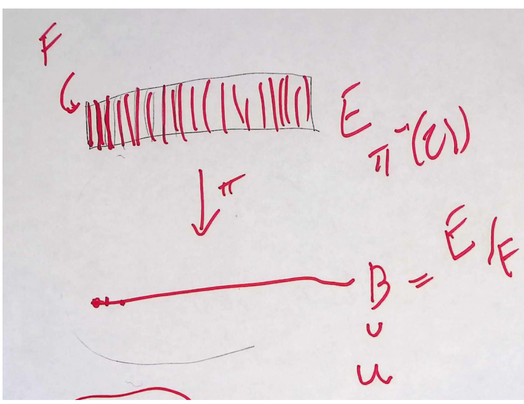
.

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 continous 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.

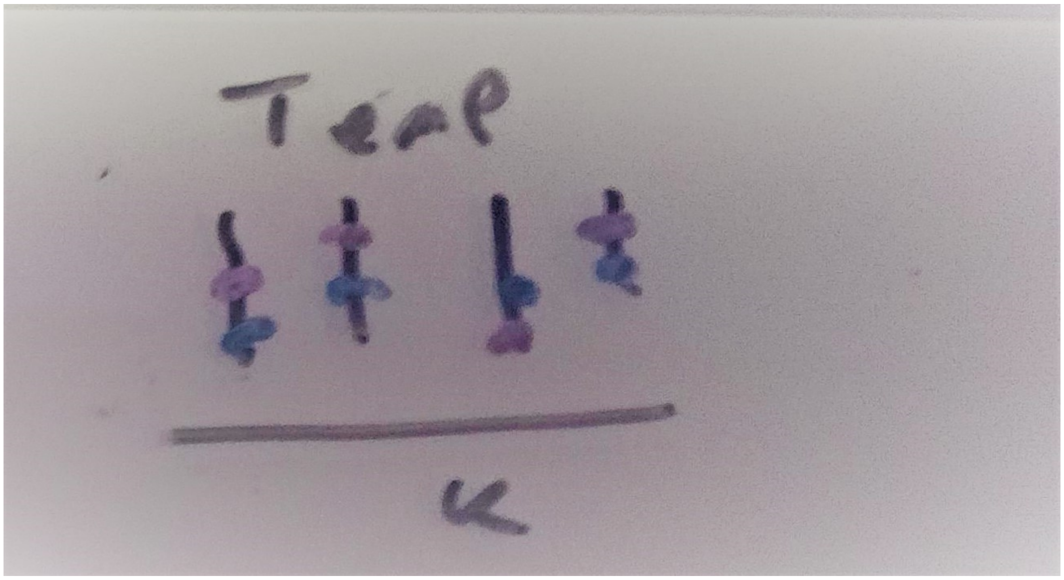
1.1.1 Base Space $\mathcal{K}$

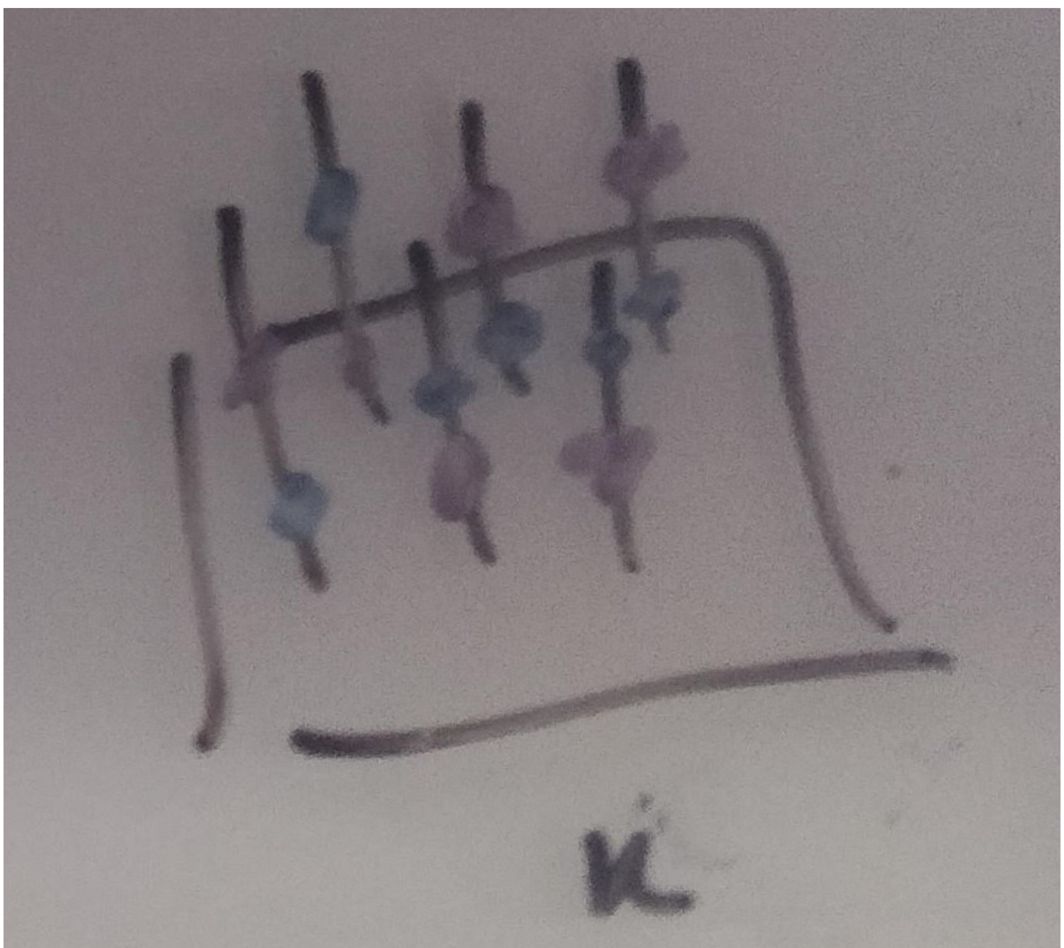
Figure 2: The topological base space K encodes the connectivity of the data space, for example if the data is independent points or a map or on a sphere

Data can be discrete observations, timeseries, maps, fields [16] 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 [16].



K and F are not intrinsic to E, rather they are choices in how E is subdivided[18]. 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 [25].


Temp
K


K

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 $\tau : K \to E$ provides the binding between the key $k \in K$ and the value r in E [15].

Triangulization 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 continous 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 simplacial complex.

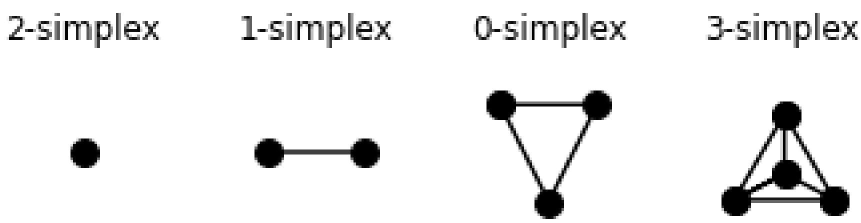


Figure 3: Simplices encode the connectivity of the data, from fully disconnected (0 simplex) observations to all observations are connected to at least 3 other observations. Higher order simplicies are outside the scope of this paper.

K is a triangularizable topological space; one triangularization 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

1.1.2 Fiber Space

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, "\varphi"] \arrow[d, "\pi"'] & U \times F \arrow[dl, "\mathrm{proj}_U"] \\ U & \end{tikzcd} \tag{3}$$

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

$F_k = \pi^{-1}(k \in K)$ is homomorphic to the fiber F.

Definition A monoid[13] 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, 21] of M is a set X with an action $\bullet$ with the properties:

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

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

1.1.3 Section

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

$$\begin{tikzcd} F \arrow[r, hook] & E \arrow[d, "\pi"'] \\ & B \arrow[u, bend right, "f"'] \end{tikzcd} \tag{4}$$

such that f is the right inverse of π

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

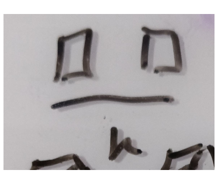
In a trivial fiber bundle, $E = K \times F$ [7, 20]:

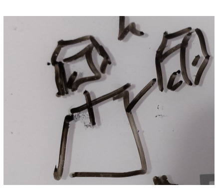
$$f(b) = (b, g(b)) \tag{6}$$

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 $r : C \to U_\sigma$ 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 $\tau : K \to \Gamma(\sigma)$. This is equivalent to $\tau : k \to E$ since $F = \Gamma(\sigma)$ and F is the embedding in E on which the records r lie.

1.1.4 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 $C = \{\text{Temp}, \text{Time}\}$:

$$F = [temp_{min}, temp_{max}] \times [time_{min}, time_{max}] \tag{7}$$

and the function τ that retrieves records from F is

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

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

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

1.1.5 Sheaf and Stalk

Often a graphic may need to be updated with live data or support zooming in on a segement of the dataset; to support working with a subset of data, we can use the sheaf $\mathcal{O}(E)$:

$$\begin{tikzcd} \iota^* E \arrow[d, shift right, "\pi"'] & E \arrow[l, "\iota^*"'] \arrow[d, shift right, "\pi"'] \\ U \arrow[u, shift right, bend right, "\iota^*\tau"'] & K \arrow[l, "\iota"'] \arrow[u, shift right, bend right, "\tau"'] \end{tikzcd} \tag{10}$$

As shown in equation 3, there is a local space $U \subset K$ around every k. The inclusion map $\iota : U \rightarrow K$ is pulled back such that $\iota^* E$ is the space of E restricted over K. The localized section of fibers $\iota^* \tau : U \rightarrow \iota^* E$ is the sheaf with a germ of $\xi^* \tau$. The neighborhood of points k_i surrounding the point k the sheaf lies over is the stalk $\mathcal{F}_b$ [22, 23].

The jet bundle $\mathcal{J}$ [9, 17] is a type of sheaf that allows for writing differential equations on sections of fiber bundles; this information is required for some visual characteristics, such as line thickness.

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 $\xi : S \to K$ is a deformation retraction [19]

$$\begin{tikzcd} E \arrow[d, "\pi"'] & H \arrow[d, "\pi"'] \\ K & S \arrow[l, "\xi"'] \end{tikzcd} \tag{11}$$

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 $\rho : S \to H$ 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

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

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

$$s = \rho^{-1}_{XY}(xy) \tag{13}$$

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

$$r = \iint\limits_s \rho_R(s) ds^2 \tag{14}$$

$$g = \iint\limits_s \rho_G(s) ds^2 \tag{15}$$

$$b = \iint\limits_s \rho_B(s) ds^2 \tag{16}$$

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

1.2.3 Example

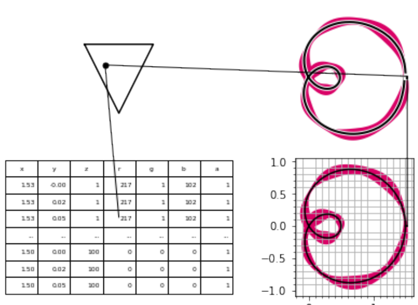


Figure 4

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 $\mathcal{O}(H)$.

$$A : \mathcal{O}(E) \to \mathcal{O}(H) \tag{17}$$

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

$$\begin{tikzcd} E \arrow[r, "\nu"] \arrow[dr, "\pi"'] & V \arrow[d, "\pi"] & \xi^*V \arrow[l, "\xi^*"'] \arrow[r, "Q"] \arrow[d, "\xi^*\pi"'] & H \arrow[dl, "\pi"] \\ & K & S \arrow[l, "\xi"'] & \end{tikzcd} \tag{18}$$

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 $\mu : V \to K$ that resolves to a visual variable [2, 14] in fiber P. The visual transformer ν is a set of functions each targeting a different μ

$$\{\nu_0, \ldots, \nu_n\} : \{\tau_0, \ldots, \tau_n\} \mapsto \{\mu_0, \ldots, \mu_n\} \tag{19}$$

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

1.3.1 Example: Matplotlib Visual Fiber

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

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

1.3.2 Visual Channels

$\nu : E \to V$ 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

$$\begin{tikzcd} E_i \arrow[r, "\nu_i"] \arrow[d, "m_e"'] & V_i \arrow[d, "m_v"] \\ E_i \arrow[r, "\nu_i"] & V_i \end{tikzcd} \tag{20}$$

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 $e_1, e_2 \in E_i$

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

Example: Translation fairly certain I lost the thread here According to Stevens, interval data is a set with general linear group actions [11, 24]. Position is a visual variable that can support translation [2, 10, 14].

$$\nu(x + c) = \nu(x) + \nu(c) \tag{22}$$

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

$$\begin{tikzcd} E_1 \arrow[r, "\lambda : e \mapsto .5"] \arrow[d, "2e"'] & V_i \arrow[d, "2v"] \\ E_1 \arrow[r, "\lambda"] & V_1 \end{tikzcd} \tag{23}$$

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

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

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

$$.5 \neq 2 * .5 \tag{26}$$

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

1.3.3 Assembling Marks

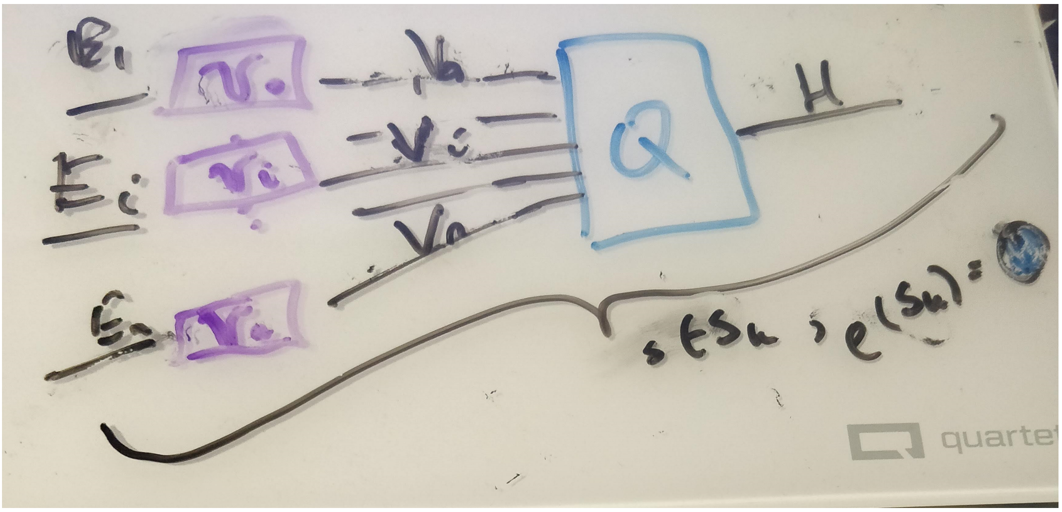


Figure 5: The ν functions convert data E to visual V. Q assembles the different types of visual parameters V_i into a graphic in H. $Q \circ \mu(\xi^{-1}J)$ forms a visual mark by applying Q to a region mapped to connected components $J \subset K$.

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 $J \subset K$:

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

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

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

$$H \underset{\rho(S_j)}{\rightleftarrows} S_j \underset{\xi^{-1}(J)}{\overset{\xi(s)}{\rightleftarrows}} J_k \tag{28}$$

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

1.3.4 Visual Idioms: Equivalance class of artists

Given $O(E)$ of the same type that output to the same type of graphic $O(H)$, the

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