

# 1 Notation & Definitions

In this section we introduce a mathematical description of the visualization pipeline where artist  $A$  functions transform data of type  $\Gamma(E)$  to an intermediate representation in prerendered display space of type  $\Gamma(H)$ :

$$A : \Gamma(E) \rightarrow \Gamma(H) \quad (1)$$

$$A : \tau \rightarrow \rho \quad (2)$$

- $A$  is the function that converts an instance of data  $\Gamma(E)$  to an instance of a visual representation  $\Gamma(H)$
- $E$  is a locally trivial fiber bundle over  $K$  representing data space.
- $K$  is a triangulizable space encoding the connectivity of the observations in the data.
- $H$  is a fiber bundle over  $S$  representing visual space
- $S$  is a simplicial complex of triangles encoding the connectivity of the visualization of  $\Gamma(E)$
- $\tau : K \rightarrow E$  is the data being visualized
- $\rho : S \rightarrow H$  is the render map

When  $E$  is a trivial fiber bundle  $E = F \times K$ , it can be assumed that all fibers  $F_k$  over  $k \in K$  are equal. Fiber bundles are product spaces of topological spaces, which are a set of points with a set of neighborhoods for each point[5, 8].

## 1.1 Data Model

We use a fiber bundle model to represent the data, as proposed by Butler [2, 3]. A fiber bundle is a structure  $(E, K, \pi, F)$  consisting of topological spaces  $E, K, F$  and the map from total space to base space:

$$\begin{array}{ccc} F & \hookrightarrow & E \\ & \pi \downarrow & \nearrow \\ & K & \end{array} \quad (3)$$

where there is a bijection from  $F$  to every fiber  $F_k$  over point  $k \in K$  in  $E$  and the function  $\pi : E \rightarrow K$  is the map into the  $K$  quotient space of  $E$ . Every point in the base space  $k \in K$  has a local open set neighborhood  $U$  [5, 8]

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\varphi} & U \times F \\ \pi \downarrow & \swarrow \text{proj}_U & \\ U & & \end{array} \quad (4)$$

such that  $\varphi : \pi^{-1}(U) \rightarrow U \times F$  is a homeomorphism where  $\pi$  and  $\text{proj}_U$  both map to  $U$  and the fiber over  $k$   $F_k = \pi^{-1}(k \in K)$  is homomorphic to the fiber  $F$ .

The section  $\tau$  is the mapping  $\tau : K \rightarrow E$

$$\begin{array}{ccc} F & \xhookrightarrow{\quad} & E \\ \pi \swarrow & \nearrow \tau & \\ K & & \end{array} \quad (5)$$

such that it is the right inverse of  $\pi$

$$\pi(\tau(k)) = k \text{ for all } k \in K \quad (6)$$

In a locally trivial fiber bundle,  $E = K \times F$  [5, 8]:

$$\tau(k) = (k, g(k)) \quad (7)$$

where the domain of  $g(k)$  is  $F_k$  and returns a data point  $r$ . The space of all possible sections  $\tau$  of  $E$  is  $\Gamma(E)$ . All datasets  $\tau \in \Gamma(E)$  have the same variables  $F$  and connectivity  $K$  but can have different values such that  $\tau_i \neq \tau_j$ .

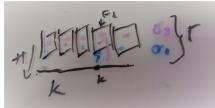
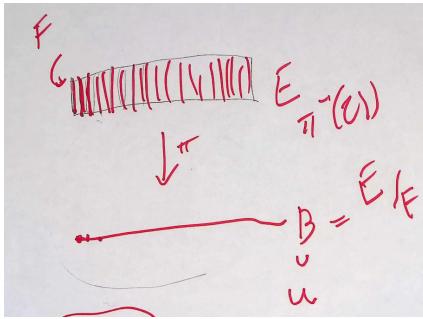


Figure 1: write up some words here

As illustrated by figure 1, the vertical lines  $F$  are the range of possible temperature values embedded in the total space  $E$ . The base space  $K$  of the fiber bundle is a line because the data points  $r$  in  $E$  are on a space that is continuous in one dimension.

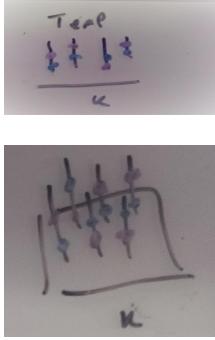
### 1.1.1 Base Space $K$



$K$  is the quotient space of  $E$ , meaning it is the set of equivalence classes of elements  $r$  in  $E$  defined via the map  $\pi : E \rightarrow K$  that sends each  $r \in E$  to its equivalence class in  $[r] \in K$  [7].

As shown in figure ??, the fibers  $F$  divide  $E$  into smaller spaces consisting of  $F$  and an open set neighborhood around  $F$ . This subdivision is projected down to the topology  $\mathcal{T}$

$$\mathcal{T}_k = \{U \subseteq K : \{r \in E : [r] \in U\} \in \mathcal{T}_E\} \quad (8)$$



where  $[r] \in U$  is the point  $k \in K$  with an open set surrounding it that has an open preimage in  $E$  under the surjective map  $\pi : r \rightarrow [r]$ .

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  $E$  [6].

### 1.1.2 Fiber Space $F$

We use Spivak's formalization of data base schemas as the basis of our fiber space  $F$  [9]. He defines the type specification

$$\pi : U \rightarrow DT \quad (9)$$

where  $DT$  is the set of data types (as identified by their names) and  $U$  is the disjoint set of all possible objects  $x$  of all types in  $DT$ . This means that for each type  $T \in DT$ , the preimage  $\pi^{-1}(T) \subset U$  is the domain of  $T$ , and  $x \in \pi^{-1}(T) \subset U$  is an object of type  $T$ . Spivak then defines a schema  $(C, \sigma)$  of type  $\pi$ , where  $\pi$  is the universe of all types, such that

$$\sigma : C \rightarrow DT \quad (10)$$

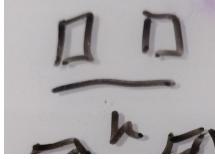
where  $C$  is the finite set of names of columns, which we generalize to data fields in  $E$ . The set of all values restricted to the datatypes in  $DT$  is  $U_\sigma$

$$\begin{array}{ccc} U_\sigma & \longrightarrow & U \\ \downarrow \pi_\sigma & & \downarrow \pi \\ C & \xrightarrow{\sigma} & DT \end{array} \quad (11)$$

The pullback  $U_\sigma := \sigma^{-1}(U)$  restricts  $U$  to the datatypes of the fields in  $C$  such that  $U_\sigma$  is the fiber product  $U \times_{DT} C$ , and the pullback  $\pi_\sigma : U_\sigma \rightarrow C$  specifies the domain bundle  $U_\sigma$  over  $C$  induced by  $\sigma$ . The fiber  $F$  is the cartesian product of all sets in the disjoint union  $U_\sigma$ .

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  $U_\sigma$ . Spivak defines the  $\tau$  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.



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

$$F = [temp_{min}, temp_{max}] \times [time_{min}, time_{max}] \quad (12)$$

and the function  $\tau$  that retrieves records from  $F$  is

$$\tau(k) = (k, (r : \text{Temp} \rightarrow temp, r : \text{Time} \rightarrow time)) \quad \begin{aligned} temp &\in [temp_{min}, temp_{max}], \\ time &\in [time_{min}, time_{max}] \end{aligned} \quad (13)$$

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

### 1.1.3 Sheaf: Subset

$\Gamma(E)$  is the space of all points in  $F$  returned by  $\sigma$ ; therefore the points being visualized in a streaming or animation example can be considered a subset that lives on base space  $U \subset K$  with the same fiber  $F$

$$\begin{array}{ccc} \iota^* E & \longrightarrow & E \\ \downarrow \iota^* \sigma & & \downarrow \sigma \\ U & \xhookrightarrow{\iota} & K \end{array} \quad (14)$$

where  $\iota^* E$  and  $\iota^* \sigma$  are  $E$  and  $\sigma$  restricted to points  $k \in U \subset K$ .

## 1.2 Prerender Space

The sets correspond to the sections on §, which is the topology of the output of the artist  $A$ . The space  $H$  is a total space representing the predisplay space, with a fiber dependent on the render space and a base space of §:

$$\begin{array}{ccc} D & \longrightarrow & H \\ \pi / & \nearrow \rho & \\ S & & \end{array} \quad (15)$$

where  $\rho : H \rightarrow S$  is mapping from a mathematical encoding of the image to pixels on the screen.

For a physical screen display, the fiber  $D$  encodes position and color of the visual elements as  $\mathbb{R}^7$ :

$$\mathbb{R}^7 = \{X, Y, Z, R, G, B, A\} \quad (16)$$

where the pixel  $p \in H$  has the fields  $\{x, y, z, r, g, b\}$ . In the case of 2D screens, the predisplay space is a trivial fiber bundle  $H = \mathbb{R}^7 \times S$ .

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

$$s = (\rho_x^{-1}(X(p)), \rho_y^{-1}(Y(p))) \quad (17)$$

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

$$R(p) = \oint_s \rho_R(s) ds^2 \quad (18)$$

$$G(p) = \oint_s \rho G(s) ds^2 \quad (19)$$

$$B(p) = \oint_s \rho B(s) ds^2 \quad (20)$$

Here we assume a single opaque 2D image such that the  $z$  and *alpha* fields can be omitted.

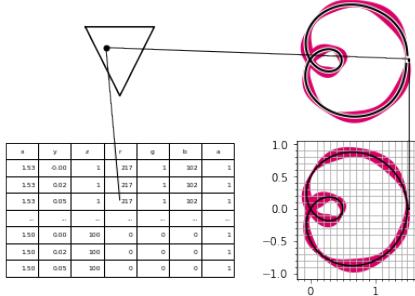


Figure 2

As illustrated in figure 2, ....

### 1.3 Artist

The artist is a mapping from the  $\Gamma(E)$  space of all datasets with a shared  $F$  and  $K$  to a pre-render space  $H$  with display target properties  $D$  and visualization structure  $S$ .

$$A : \Gamma(E) \rightarrow \Gamma(H) \quad (21)$$

### 1.3.1 Screen to Data

$$\begin{array}{ccc} E & & H \\ \downarrow \pi \sigma & & \downarrow \pi \rho \\ K \xleftarrow{\xi} S & & \end{array} \quad (22)$$

The pullback  $\xi$  on  $S \rightarrow K$  means that  $r \in E$  can be mapped to from a simplex in  $S$

$$\begin{array}{ccc} \xi E & \xleftarrow{\tau} & H \\ \downarrow \xi \sigma & \curvearrowright & \downarrow \rho^{-1} \\ & S & \end{array} \quad (23)$$

### 1.3.2 Marks

Bertin describes a location on the plane as the signifying characteristic of a point, measurable length as the signifying characteristic of a line, and measurable size as the signifying characteristic of an area and that in display (pixel) space these are marks [1, 4].

$$H \xleftarrow[\rho(\xi^{-1}(J))]{\xi(s)} S \xleftarrow[\xi^{-1}(J)]{} J_k = \{j \in K \mid \exists \Gamma \text{ s.t. } \Gamma(0) = k \text{ and } \Gamma(1) = j\} \quad (24)$$

Each point  $s$  in the display space  $H$ , the mark it belongs to can be found by mapping  $s$  back to  $K$  via the lookup on  $S$  described in section 1.2 then taking  $\xi(s)$  back to a point on  $k \in K$  which lies on the connected component  $J \subset K$ . To get back to the display space  $H$  from the simplicial complex  $J$  of the signifier implanted in the mark, the inverse image of  $J \in S, \xi^{-1}(J)$  is pushed back to  $S$ , and then  $\rho(\xi^{-1}(J))$  maps it into  $R^7$ .

### 1.3.3 Channels

Acts on different parts of  $F$ , types means measurement groups , can be broken out so Tau can preserves the measurement type properties (group scales)

Tau is fully flexible and can do whatever; knows about fiber & neighborhood of fiber. Can in theory approximate hatching/dashing/etc can be approximated w/ functions and neighborhood of k.

### 1.3.4 Visual Idioms: Equivalence class of artists

Two artists are equivalent when given data containers  $\Gamma(E)$  of the same type, they output the same type of prerender  $\Gamma(S)$ :

$$\begin{array}{ccc} A_{\tau_2} : & \Gamma(E) & \longrightarrow \Gamma(H) \\ \downarrow & & \\ A_{\tau_1} : & \Gamma(E) & \longrightarrow \Gamma(H) \end{array} \quad (25)$$