

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 $\Gamma(H)$:

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

- 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 points in E
- H is a fiber bundle over S representing visual space
- S is a simplicial complex encoding the visualization

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, 12].

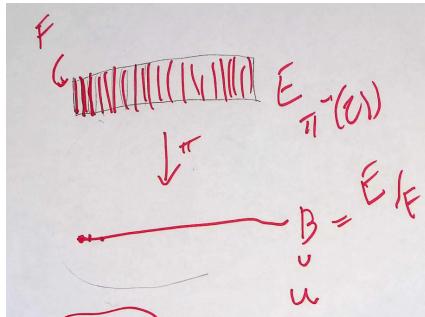
1.1 Fiber Bundles

We provide a brief description of fiber bundles because we model data, visual transformations, and a prerendered visual graphic as fiber bundles. A fiber bundle is a structure (E, B, π, F) consisting of topological total space E , base space B , fiber space F and the map from total space to base space:

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

where there is a bijection from F to every fiber F_b over point $b \in B$ in E and the function $\pi : E \rightarrow B$ is the map into the B quotient space of E . By defintion of a fiber bundle, π is always a mapping from total space to base space, independent of the points $p \in E$, and therefore we call this mapping π for all the fiber bundles in the model.

1.1.1 Base space



B is the quotient space of E , meaning it is the set of equivalence classes of elements p in E defined via the map $\pi : E \rightarrow B$ that sends each $p \in E$ to its equivalence class in $[p] \in B$ [10].

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}

$$\mathcal{T}_b = \{U \subseteq B : \{p \in E : [p] \in U\} \in \mathcal{T}_E\} \quad (3)$$

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

1.1.2 Fiber

As shown in equation!??, every point $b \in B$ has a local open set neighborhood U [5, 12]

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

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 .

1.1.3 Section

The section f is the mapping from base space to total space $f : B \rightarrow E$

$$\begin{array}{ccc} F & \hookrightarrow & E \\ \pi \downarrow & \nearrow f & \\ B & & \end{array} \quad (5)$$

such that f is the right inverse of π

$$\pi(f(b)) = b \text{ for all } b \in B \quad (6)$$

In a locally trivial fiber bundle, $E = B \times F$ [5, 12]:

$$f(b) = (b, g(b)) \quad (7)$$

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 .

1.1.4 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 .

$$\begin{array}{ccc} \iota^* E & \xleftarrow{\iota^*} & E \\ \pi \downarrow & \nearrow \iota^* f & \pi \downarrow \\ U & \xleftarrow{\iota} & B \end{array} \quad (8)$$

The localized section of fibers $\iota^* f : U \rightarrow \iota^* E$ is the sheaf $\mathcal{O}(E)$ with germ of $\xi^* f$. The neighborhood of points the sheaf lies over is the stalk \mathcal{F}_b [13, 15]

$$\iota^{-1} \mathcal{F}(\{b\}) = \varinjlim_{b \subseteq U} \mathcal{F}(U) = \varinjlim_{b \in U} = \mathcal{F}_b \quad (9)$$

which through ι gets the data in E at and near to b . 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 Data Model

As proposed by Butler [2, 3], we model data as a fiber bundle (E, K, π, F) with $\pi : E \rightarrow K$ where K which can be thought of as a set of keys k . A section $\tau : K \rightarrow E$ is an instance of the data that lies in E and is discussed in section 1.2.4.

1.2.1 Example

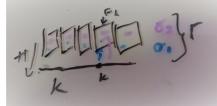


Figure 1: write up some words here

As illustrated by figure ??, 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.

1.2.2 Base Space

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.

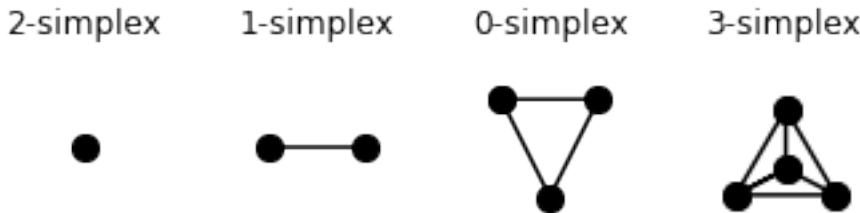


Figure 2: Simplices encode the connectivity of the data, from fully disconnected (0 simplex) observations to all observations are connected to at least 3 other observations. Higher order simplices are outside the scope of this paper.

One way to represent the topological space K is as a set composed of simplices, such as those shown in figure ???. Simplices are a way of encoding the connectivity of each observation ($\sigma(k)$) to another:

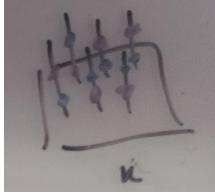
0-simplex discrete observations (inventory records)

1-simplex 1D continuos data (timeseries)

2-simplex 2D continuos data (map)

3-simplex 3D continuos data (video)

1.2.3 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 [9].

1.2.4 Fiber and Sections

We use Spivak's formalization of data base schemas as the basis of our fiber space F [14]. He defines the type specification

$$\pi : U \rightarrow DT \quad (10)$$

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, σ) of type π , where π is the universe of all types, such that

$$\sigma : C \rightarrow DT \quad (11)$$

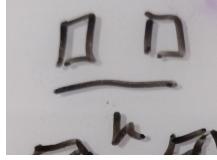
where C is the finite set of names of columuns, which we generalize to data fields in E . The set of all values restricted to the datatypes in DT is U_σ

$$\begin{array}{ccc}
 U_\sigma & \longrightarrow & U \\
 \pi_\sigma \downarrow & & \downarrow \pi \\
 C & \xrightarrow{\sigma} & DT
 \end{array} \tag{12}$$

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

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.

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

$$F = [temp_{min}, temp_{max}] \times [time_{min}, time_{max}] \tag{13}$$

and the function τ that retrieves records from F is

$$\tau(k) = (k, (r : \text{Temp} \rightarrow temp, r : \text{Time} \rightarrow time)) \tag{14}$$

$$temp \in [temp_{min}, temp_{max}], time \in [time_{min}, time_{max}] \tag{15}$$

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

1.3 Prerender Space

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

1.3.1 Base space

K can be considered a subspace of the screen base space S such that $\xi : S \rightarrow K$ is a deformation retraction [11]

$$\begin{array}{ccc} E & & H \\ \pi \downarrow & & \pi \downarrow \\ K & \xleftarrow{\xi} & S \end{array} \quad (16)$$

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

1.3.2 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. For a physical screen display, the predisplay space is a trivial fiber bundle $H = \mathbb{R}^7 \times S$ such that ρ is

$$\rho(s) = \{x, y, z, r, g, b, a\} \quad (17)$$

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

$$s = \rho_{XY}^{-1}(xy) \quad (18)$$

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

$$r = \iint_s \rho_R(s) ds^2 \quad (19)$$

$$g = \iint_s \rho_G(s) ds^2 \quad (20)$$

$$b = \iint_s \rho_B(s) ds^2 \quad (21)$$

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

1.3.3 Example

As illustrated in figure 3, words.

1.4 Artist

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

$$A : \mathcal{O}(E) \rightarrow \mathcal{O}(H) \quad (22)$$

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 $Q : \xi^*V \rightarrow H$.

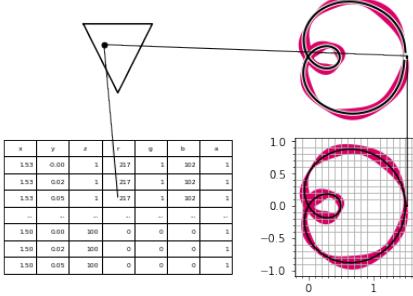


Figure 3

$$\begin{array}{ccc}
 E & \xrightarrow{\nu} & V \\
 & \searrow \pi & \downarrow \pi \\
 & K & S
 \end{array}
 \quad
 \begin{array}{ccc}
 \xi^* V & \xrightarrow{Q} & H \\
 \xi^* \pi \downarrow & & \swarrow \pi \\
 S & \xleftarrow{\xi} & K
 \end{array}
 \quad (23)$$

The visual fiber bundle (V, K, π, P) has section $\mu : V \rightarrow K$ that resolves to a visual variable [1, 8] in fiber P . The visual transformer ν is a set of functions each targeting a different μ

$$\nu : \{\nu_0, \dots, \nu_n\} \rightarrow \{\mu_0, \dots, \mu_n\} \quad (24)$$

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

1.4.1 Example: Matplotlib Visual Fiber

For example, for Matplotlib [6], 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, linecolor, facecolor, markerfacecolor, edgecolor	\mathbb{R}^4
texture	hatch	\mathbb{N}^{10}
	linestyle	$(\mathbb{R}, \mathbb{R}^{+n, n \% 2 = 0})$

1.4.2 Visual Channels

$\nu : \tau \mapsto \mu$ is an equivariant map such that there is a homomorphism from monoid actions on E_i to monoid actions on V_i

$$\begin{array}{ccc} E_i & \xrightarrow{\nu_i} & V_i \\ f \downarrow & & \downarrow f \\ E_i & \xrightarrow{\nu_i} & V_i \end{array} \quad (25)$$

where the diagram commutes such that $\nu_i \circ f(\tau_i) = f \circ \nu_i(\tau_i)$. A monoid is a set M with a closed associative binary operation (\cdot) and an identity element e .

Example: Translation When τ_i of E_1 returns interval data, Steven's defines the group actions on the data as the general linear group [7, 16]; therefore to construct a valid ν , we choose a target variable in P that is

$$\begin{aligned} \nu_i(\tau_i) &= \tau + 2 \\ \lambda e &: .2 \end{aligned}$$

Example: Invalid ν shifts in either space need to be mirrored in the other space, this is not valid here's why:

$$\begin{array}{ccc} E_1 & \xrightarrow{\lambda:e \mapsto .5} & V_1 \\ e+2 \downarrow & & \downarrow h+2 \\ E_2 & \xrightarrow{\lambda} & V_2 \end{array} \quad (26)$$

Example: Partial Order ν induces a monoid homomorphism to allow for partial ordering. For example if E_1 is the set '

1.4.3 Constructing Marks

The visual fiber bundle V gets pulled back over S via ξ such that

$$\begin{array}{ccc} \xi^*V & \xrightarrow{Q} & H \\ & \searrow \xi^*\pi & \downarrow \pi \\ & S & \end{array} \quad (27)$$

the function $Q : \xi^*V \rightarrow H$ composites points in V into \mathbb{R}^d tuples. The section μ is pulled over s

$$\begin{array}{ccc} \xi^*V & \xrightarrow{Q} & H \\ & \nwarrow \xi^*\mu & \uparrow \rho \\ & S & \end{array} \quad (28)$$

such that the composition Q of μ is equivalent to the render $\rho(s) = Q \circ \mu(s)$. (In practice $\mu \cong \xi^*\mu$) Q is the constructor for the mark, one Q per artist, $A = \text{sum of artists } Q$ as there's default line, point, etc that's stretched and tweaked and whatever else by the μ visual variable + structure preserving, Q must have map that commutes such that a change

in μ carries over somehow to ρ , functorial in its arguments - commutative diagram - into rho

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 are the point, line, and area marks [1, 4]. For each region s in the display space H , the mark it belongs to can be found by mapping s back to S then taking $\xi(s)$ back to a point on $k \in K$ which lies on the connected component $J \subset K$.

$$H \xrightleftharpoons[\rho(\xi^{-1}(J))]{\quad} S \xrightleftharpoons[\xi^{-1}(J)]{\quad} J_k = \{j \in K \mid \exists \Gamma \text{ s.t. } \Gamma(0) = k \text{ and } \Gamma(1) = j\} \quad (29)$$

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 .

1.4.4 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