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 $\Gamma(H)$:

$$A : \Gamma(E) \to \Gamma(H) \tag{1}$$

$$A : \sigma \to \rho \tag{2}$$

- 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 simplacial complex of triangles encoding the connectivity of the visualization of $\Gamma(E)$
- $\sigma : K \to E$ is the data being visualized
- $\rho : S \to H$ 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 toplological 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, π, F) consisting of topological spaces E, K, F and the map from total space to base space:

$$\begin{tikzcd} F \arrow[r, hook] & E \arrow[d, dashed, "\pi"', bend right] \\ & K \arrow[u, bend right] \end{tikzcd} \tag{3}$$

where there is a bijection from F to every fiber F_k over point $k \in K$ in E and the function $\pi : E \to K$ 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]

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, dashed, "\varphi"] \arrow[d, dashed, "\pi"] & U \times F \arrow[dl, "\mathrm{proj}_U"] \\ U & \end{tikzcd} \tag{4}$$

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.

The section σ is the mapping $\sigma : K \to E$

$$\begin{tikzcd} F \arrow[r, hook] & E \arrow[d, shift right, "\pi"'] \\ & K \arrow[u, shift right, "\sigma"'] \end{tikzcd} \tag{5}$$

such that it is the right inverse of π

$$\pi(\sigma(k)) = k \text{ for all } k \in K \tag{6}$$

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

$$\sigma(k) = (k, g(k)) \tag{7}$$

where the domain of $g(k)$ is F_k. The space of sections over $U \subset K$ is called a sheaf and the space of all possible sections σ of E is $\Gamma(E)$. All datasets $\sigma \in \Gamma(E)$ have the same variables F and connectivity K but can have different values such that $\sigma_i \neq \sigma_j$.

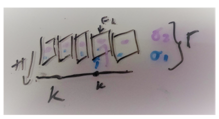
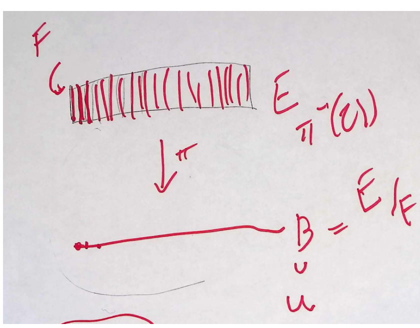


Figure 1: write up some words here

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 describes the connectivity of the points in E; in figure 1 the connectivity of the timeseries is encoded in the line representation of K.

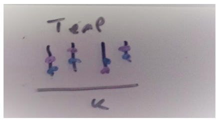
1.1.1 Base Space K

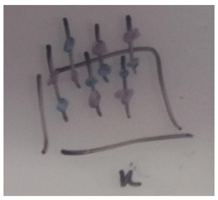


K is the quotient space of E, meaning it is the set of equivalence classes of elements p in of E defined via the map $\pi : E \to K$ that sends each point $p \in E$ to its equivalence class in $[p] \in K$ [7]. As shown in figure **??**, the fibers F divide E into smaller spaces consisisting of F and an open set neighborhood around F. This subdivision is projected down to the toplology τ_k

$$\tau_K = \{U \subseteq K : \{p \in E : [p] \in U\} \in \tau_E\} \tag{8}$$

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





We use K to encode the connectivity of the points p. In figure **??**, there is only one data field in p, temperature, but the points p are connected differently. In a timeseries, the temperature p at time t is dependent on the value at p_{t-1} and p_{t+1} is dependent on the value in p_t; this connectivity is expressed as a one dimensional K where K is the number line. In the case of the map, every point p 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 - those are metadata variables potentially encoded in p because they are ways of describing the connectivity rather than the connectivity itself. The mapping $sigma : K \to E$ provides the binding between the key on K and the value p in E [6].

1.1.2 Fiber Space F

Spivak models the fiber space F as schema and the data as sheafs (localized σ functions) on the schema [9]. He defines the type specification

$$\pi : U \to DT \tag{9}$$

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

$$\sigma : C \to DT \tag{10}$$

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

$$\begin{tikzcd} U_\sigma \arrow[r] \arrow[d, "\pi_\sigma"] & U \arrow[d, "\pi"] \\ C \arrow[r, "\sigma"] & DT \end{tikzcd} \tag{11}$$

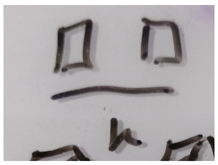
The pullback $U_\sigma \coloneqq \sigma^{-1}(U)$ restricts U to the datatypes of the fields in C such that U_σ is the fiber product $U \times_{DT} C$, and the pullback $\pi_\sigma : U_\sigma \to C$ specifies the domain bundle U_σ

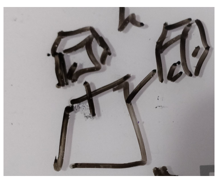
over C induced by σ. This domain bundle backs the fiber F in the data total space E

$$F = \prod_{i \in I} U_{\sigma_i} = \tag{12}$$

where F is the cartesian product of all sets in the disjoint union U_σ

The record function is the sigma function





The fibers F are a topological space embedded in E on which lie the set of all possible values. For example, if F is the interval $[0, 1]$, then $g(k)$ from equation **??** returns a single measurement x in the interval F:

$$g(k) = x, \text{ where } 0 \leq x \leq 1 \tag{13}$$

The fiber in figure **??** is the space of possible temperature values in $\circ$ celsius, ranging from [start, end], similar to the interval F in equation 13. F can be any number of dimensions, for example in figure **??** time is encoded as a second dimension. Given:

- interval of all possible temperature values $[T_{min}, T_{max}]$
- interval of all possible time values $[t_{min}, t_{max}]$

then F is the cross product $F = [T_{min}, T_{max}] \times [t_{min}, t_{max}]$, and $g(k)$ listed in equation **??** is:

$$g(k) = (x_0, x_1) \text{ where } x_0 \in [T_{min}, T_{max}], x_1 \in [t_{min}, t_{max}] \tag{14}$$

When E is trivial, then we can decompose E so that each E

1.1.3 Subset

$\Gamma(E)$ is the space of all points in F returned by σ; 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

$$\begin{tikzcd}
\iota^* E \arrow[r, hook] \arrow[d] & E \arrow[d, bend right] \\
U \arrow[u, dashed, bend right, "\iota^*\sigma"'] \arrow[r, hook, "\iota"] & K \arrow[u, dashed, bend right, "\sigma"']
\end{tikzcd} \tag{15}$$

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

1.2 Prerender Space

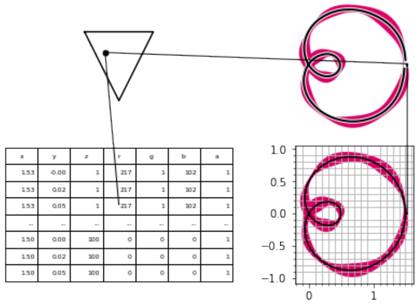


Figure 2

A physical display space can be thought of sets of $\mathbb{R}^7$ tuples, where

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

and 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 of $\mathbb{R}^7$ and a base space of §:

$$\begin{tikzcd} \mathbb{R}^7 \arrow[r, hook] & H \arrow[d, bend right, "\pi"'] \\ & S \arrow[u, bend right, "\rho"'] \end{tikzcd} \tag{17}$$

In the case of 2D screens, the predisplay space is a trivial fiber bundle $H = \mathbb{R}^7 \times S$. As illustrated in figure 2, a region on the screen defined by the corners (x_1, y_1) and (x_2, y_2) maps into a region on a 2-simplex in S defined by (α_1, β_1) and (α_2, β_2). The function on the simplex f returns the (R, G, B, A) value for that (α, β) pair. For a region,

$$\rho(S) = \int_{\alpha_1}^{\alpha_2} \int_{\beta_1}^{\beta_2} \int_{z_1}^{z_2} R, G, B, A$$

where the R,G,B,A values are derived from the how the data values are mapped to visual characteristics. The z component of the mapping to $\mathbb{R}^7$ is moved to the integration because this is a trivial space representing a 2D screen; ρ varies depending on H.

1.3 Artist

$$A : \Gamma(E) \to \Gamma(H) \tag{18}$$

1.3.1 Screen to Data

$$\begin{tikzcd} E \arrow[d, bend right, "\pi"'] & H \arrow[d, bend right, "\pi"'] \\ K \arrow[u, bend right, "\sigma"'] & S \arrow[u, bend right, "\rho"'] \arrow[l, "\xi"'] \end{tikzcd} \tag{19}$$

The pullback ξ on $S \to K$ means that the values in E can be directly mapped to a simplex in S, which means there's a mapping from screen space back to the values.

$$
\begin{tikzcd}
\xi E \arrow[rr, leftrightarrow, "\tau"] \arrow[dr, "\xi\sigma"'] & & H \arrow[dl] \\
& S \arrow[ul, bend right] &
\end{tikzcd}
\tag{20}
$$

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].

$$
\begin{tikzcd}
H \arrow[r, shift left] & S \arrow[l, shift left, "\rho(\xi^{-1}(J))"] \arrow[r, shift left, "\xi(s)"] & J_k \arrow[l, shift left, "\xi^{-1}(J)"]
\end{tikzcd}
= \{j \in K | \exists \Gamma \text{ s.t. } \Gamma(0) = k \text{ and } \Gamma(1) = j\}
\tag{21}
$$

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 got back to the display space H from the simplacial 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: Equivalance class of artists

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

$$
\begin{tikzcd}
A_{\tau_2}: \arrow[d, shift right] & \Gamma(E) \arrow[r] & \Gamma(H) \\
A_{\tau_1}: \arrow[u, shift right] & \Gamma(E) \arrow[r] & \Gamma(H)
\end{tikzcd}
\tag{22}
$$