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 $\Gamma(H)$:

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

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

- 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 the data in E
- $\tau : K \to E$ 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, 9].

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:

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

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, 9]

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

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 τ is the mapping $\tau : K \to E$

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

such that it is the right inverse of π

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

In a locally trivial fiber bundle, $E = K \times F$ [5, 9]:

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

where the domain of $g(k)$ is F_k and returns a data point r. The space of all possible sections τ 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 $\tau_i \neq \tau_j$.

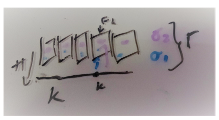
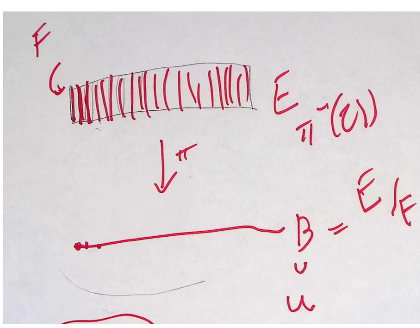


Figure 1: write up some words here

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 continous in one dimension.

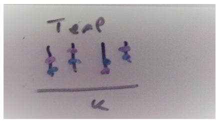
1.1.1 Base Space K

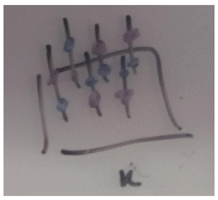


K is the quotient space of E, meaning it is the set of equivalence classes of elements r in E defined via the map $\pi : E \to K$ that sends each $r \in E$ to its equivalence class in $[r] \in K$ [8].

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 $\mathcal{T}$

$$\mathcal{T}_k = \{U \subseteqq K : \{r \in E : [r] \in U\} \in \mathcal{T}_E\} \tag{8}$$





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 [7].

1.1.2 Fiber Space F

We use Spivak's formalization of data base schemas as the basis of our fiber space F [10]. He defines the type specification

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

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

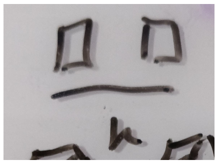
$$\sigma : C \rightarrow DT \tag{10}$$

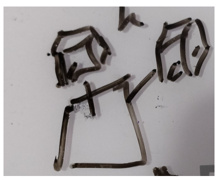
where C is the finite set of names of columuns, which we generalize to data fields in E. The set of all values restricted to the datatypes in DT is U_σ

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

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 $\pi_\sigma : U_\sigma \rightarrow C$ 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 $\tau : K \to \Gamma(\sigma)$. 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.





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

$$F = [temp_{min}, temp_{max}] \times [time_{min}, time_{max}] \tag{12}$$

and the function τ that retrieves records from F is

$$\tau(k) = (k, (r : \mathrm{Temp} \to temp, r : \mathrm{Time} \to time)) temp \in [temp_{min}, temp_{max}], time \in [time_{min}, time_{max}]) \tag{13}$$

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 and Stalk

As described in equation 4, there is a local space $U \subset K$ around every k. The inclusion map $\iota : U \to K$ can be pulled back such that $\iota^* E$ is the space of E restricted over U.

$$\begin{tikzcd}
\iota^* E \arrow[r, hook] \arrow[d] & E \arrow[d, bend right] \\
U \arrow[u, dashed, bend right, "\iota^*\tau"'] \arrow[r, hook, "\iota"] & K \arrow[u, dashed, bend right, "\tau"']
\end{tikzcd} \tag{14}$$

The localized section of fibers $\iota^*\tau : U \to \iota^* E$ is the sheaf $O(E)$. The neighborhood of points the sheaf lies over is the stalk $\mathscr{F}_k$ [11]

$$\iota^{-1}\mathscr{F}(\{k\}) = \varinjlim_{k \subseteq U} \mathscr{F}(U) = \varinjlim_{k \in U} = \mathscr{F}_k \tag{15}$$

which through ι gets the data in E at and near to k.

1.2 Prerender Space

Every point $k \in K$ maps to a space $S_k \in S$, 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 §:

$$\begin{tikzcd} D \arrow[r, hook] & H \arrow[d, shift right, "\pi"'] \\ & S \arrow[u, shift right, "\rho"'] \end{tikzcd} \tag{16}$$

where $\rho : H \to S$ is mapping from a mathematical encoding of the image to a region xy on the screen that the renderer then maps to pixel space.

For a physical screen display, the fiber D encodes position and color of the visual elements as $\mathbb{R}^7 = \{X, Y, Z, R, G, B, A\}$, where the patch xy is a region in H that we query into S to get values that we integrate into an $\{r, g, b\}$ value. 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 H is inverse mapped into a region $s \in S$ where s is the inverse mapping from pixel to region on $s = \rho_X^{-1} Y(xy)$ such that the rest of the fields in $\mathbb{R}^7$ are then integrated over s to yield the remaining fields in p

$$R(p) = \oint_s \rho_R(s) ds^2 \tag{17}$$

$$G(p) = \oint_s \rho_G(s) ds^2 \tag{18}$$

$$B(p) = \oint_s \rho_B(s) ds^2 \tag{19}$$

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

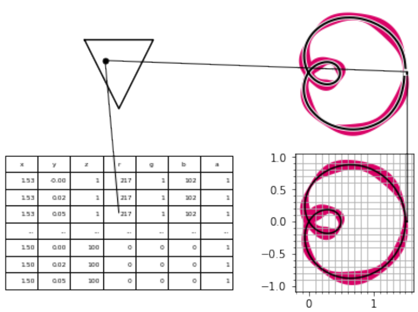


Figure 2

As illustrated in figure 2, words.

1.3 Artist

The artist is a mapping from the sheaf $O(E)$ which a dataset to a pre-render space $O(H)$ with display target properties D and visualization structure S.

$$A : \Gamma(E) \rightarrow \Gamma(H) \tag{20}$$

where A is composed of an asthethic mapping $\Xi = Q \circ \nu$ that builds the prerender render function $rho : S \rightarrow H$ functions described in section 1.2. builds ρ functions on S by compositing $\nu_{param} : c \mapsto \rho_{d \in D}$ functions that map data in some field c to a subset of visual variables such as position or color[1, 7] that are identified as parameters to the artist function.

$$\rho(u, v) = Q(\nu_param_0(\tau_{c_0}), \ldots \nu_param_n(\tau_{c_n}))(u, v) \tag{21}$$

Q composites ν functions because multiple parameters can target the same field in D, for example ν_{ypos} and $\nu_{linewidth}$ together determine the y position of a line drawn on screen such that $Q : \nu_{ypos}, \nu_{linewidth} \rightarrow \rho_Y$. The artist is

$$\begin{tikzcd}
\iota^* E \arrow[dr, "\Xi"] & H \arrow[d, bend right, "\rho^{-1}"'] \\
U \arrow[u, "\tau"] & S \arrow[l, "\xi"'] \arrow[u, bend right, "\rho"']
\end{tikzcd} \tag{22}$$

1.3.1 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].

$$H \underset{\rho(\xi^{-1}(J))}{\overset{}{\leftrightarrows}} S \underset{\xi^{-1}(J)}{\overset{\xi(s)}{\leftrightarrows}} J_k = \{j \in K | \exists \Gamma \text{ s.t. } \Gamma(0) = k \text{ and } \Gamma(1) = j\} \tag{23}$$

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

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.

1.3.2 Channels

Visual channels are mappings from data into visual space, for example distinguishing categories by coloring them differently [1, 6]. Channels are an example of $\nu : c \mapsto d \in D$, which are the mappings from data $r \in \Gamma(\tau)$ to some part of the prerender fiber D. the ν functions can target only one field in D, such as X, or multiple such as R, G, B.

Can in theory approximate hatching/dashing/etc can be approximated w/ functions and neighborhood of k.

1.3.3 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 $\Gamma(S)$:

$$
\begin{array}{lc}
A_{\tau_2}: & \Gamma(E) \longrightarrow \Gamma(H) \\
\downarrow\uparrow & \\
A_{\tau_1}: & \Gamma(E) \longrightarrow \Gamma(H)
\end{array}
\tag{24}
$$