

# 1 Topological Artist Model

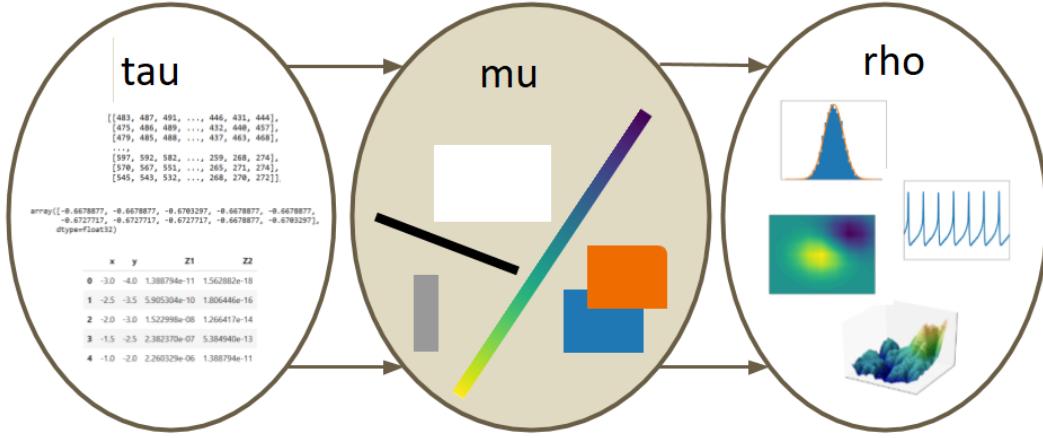


Figure 1: Visualization is equivariant maps between data and visual encoding of the variables and assembly of those encodings into a graphic. *not gonna name these bubbles tau, mu, rho, but might keep the same basic structure of different types of data and encodings*

*2 should this be in 3rd person passive?*

3 Visualization is generally thought of as structure preserving maps from data into graphics,  
4 and in this section we formally define that structure and how it is preserved via equivariant  
5 maps. We can then specify that a faithful visual mapping is structure preserving, and apply  
6 these constraints to visualizations we may want to develop or implement. We model the data,  
7 visual characteristic, and graphic stages of visualization, shown in figure 1, as topological  
8 structures that encapsulate types of variables and continuity; by doing so we can develop  
9 implementations that keep track of both in ways that let us distribute computation while  
10 still allowing assembly and dynamic update of the graphic.

11 We introduce a mathematical description of the visualization pipeline where artist  $\mathcal{A}$   
12 functions transform data space  $\mathcal{E}$  to an intermediate representation in a prerendered graphic  
13 space  $\mathcal{H}$ .

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

14 We first describe how we model data(1.1), graphics(1.2), and intermediate visual char-  
15 acteristics (1.3) as fiber bundles. We then discuss the equivariant maps between data and  
16 visual characteristics (1.3.2) and visual characteristics and graphics (1.3.3) that make up  
17 the artist.

## 18 1.1 Data Space $E$

19 We build on Butler's proposal of using fiber bundles as a common data representation  
20 format for visualization data[4, 5] because fiber bundles are mathematical structures that  
21 are flexible enough express all the types of data described in section ??.

22 We model data as the fiber bundle  $(E, K, \pi, F)$ , where  $E$   $F$  and  $K$  are topological  
23 spaces that encode

- <sup>24</sup>  $F$  the properties of the variables in the fiber (1.1.1)  
<sup>25</sup>  $K$  the continuity of the records in the base space (1.1.3)  
<sup>26</sup>  $\tau$  collections of records (1.1.4).

and  $E$  is the total space of data that  $F$  lives in. The bundle is the projection map  $\pi$

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

<sup>27</sup> that binds the variables  $F$  continuity  $K$ . The fiber bundles mentioned in this work  
<sup>28</sup> are assumed to be trivial[15, 24], unless otherwise specified, because the trivial bundle is  
<sup>29</sup>  $E = K \times F$  such that extra structure in the total space  $E$  falls out and discussion can be  
<sup>30</sup> focused on the fiber and base space.

### <sup>31</sup> 1.1.1 Variables: Fiber Space $F$

The fiber is a topological space that is the set of possible values of the data; the values themselves can be any dimension and type and have any continuity. We use Spivak's description of simplicial database schemas [25] as the basis of our fiber space because he binds the components of the fiber to variable names and types. Spivak constructs a set  $\mathbb{U}$  that is the disjoint union of all possible objects of types  $\{T_0, \dots, T_n\} \in \mathbf{DT}$ , where  $\mathbf{DT}$  are the data types of the variables in the dataset. He then defines the single variable set  $\mathbb{U}_\sigma$

$$\begin{array}{ccc} \mathbb{U}_\sigma & \longrightarrow & \mathbb{U} \\ \pi_\sigma \downarrow & & \downarrow \pi \\ C & \xrightarrow[\sigma]{} & \mathbf{DT} \end{array} \quad (3)$$

which is  $\mathbb{U}$  restricted to objects of type  $T$  bound to variable name  $c$ . Given  $\sigma$ , the fiber for a one variable dataset is

$$F = \mathbb{U}_{\sigma(c)} = \mathbb{U}_T \quad (4)$$

where  $\sigma$  is the schema binding variable name  $c$  to its datatype  $T$ . A dataset with multiple variables has a fiber that is the cartesian cross product of  $\mathbb{U}_\sigma$  applied to all the columns:

$$F = \mathbb{U}_{T_1} \times \dots \mathbb{U}_{T_i} \dots \times \mathbb{U}_{T_n} \quad (5)$$

which is equivalent to

$$F = F_0 \times \dots \times F_i \times \dots \times F_n \quad (6)$$

- <sup>32</sup> which allows us to decouple  $F$  into components  $F_i$ .

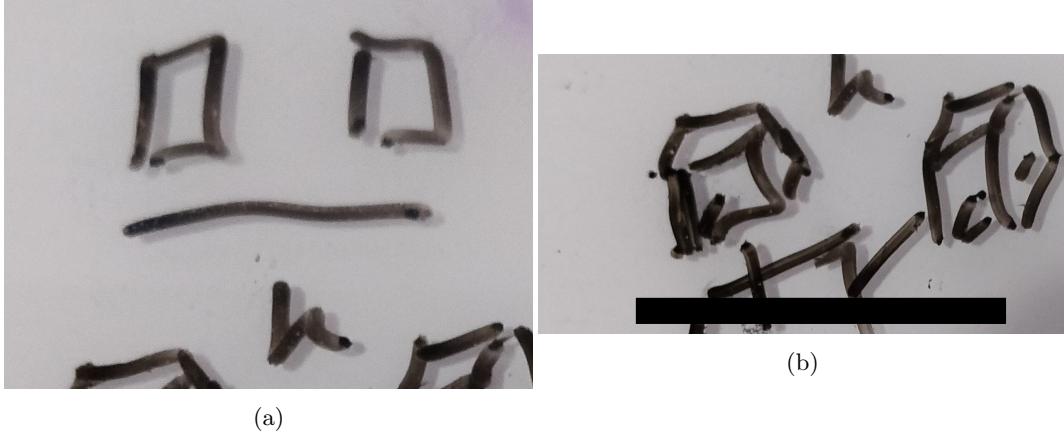


Figure 2: These two datasets have the same base space  $K$  but figure 2a has fiber  $F = \mathbb{R} \times \mathbb{R}$  which is (time, temperature) while figure 2b has fiber  $\mathbb{R}^+ \times \mathbb{R}^2$  which is (time, wind=(speed, direction))

For example, the data in figure 2a is a pair of times and °C temperature measurements taken at those times. Time is a positive number of type `datetime` which can be resolved to positive floats  $U_{\text{datetime}} = \mathbb{R}^+$ . Temperature values are real numbers  $U_{\text{float}} = \mathbb{R}$ . The fiber is

$$\mathbb{U} = \mathbb{R}^+ \times \mathbb{R} \quad (7)$$

where the first component  $F_0$  is the set of values specified by ( $c_0 = \text{time}$ ,  $T_0 = \text{datetime}$ ,  $\mathbb{U}_\sigma = \mathbb{R}^+$ ) and  $F_1$  is specified by ( $c_1 = \text{temperature}$ ,  $T_1 = \text{float}$ ,  $\mathbb{U}_\sigma = \mathbb{R}$ ). In figure 2b, temperature is replaced with wind. This wind variable is of type `wind` and has two components speed and direction  $\{(s, d) \in \mathbb{R}^2 \mid 0 \leq s, 0 \leq d \leq 360\}$ . Therefore, the fiber is

$$F = \mathbb{R}^+ \times \mathbb{R}^2 \quad (8)$$

- <sup>33</sup> such that  $F_1$  is specified by ( $c_1 = \text{wind}$ ,  $T_1 = \text{wind}$ ,  $\mathbb{U}_\sigma = \mathbb{R}^2$ )

### <sup>34</sup> 1.1.2 Measurement Scales: Monoid Actions

After specifying  $F$  we next describe the ways in which we can transform the values by identifying the monoid actions  $M$  on the  $F$ . We use monoids as the abstraction because they encode compositiblty, which maps well to the data transformation process in a software library [31].

A monoid [17]  $M_i$  is a set with an associative binary operator  $* : M_i \times M_i \rightarrow M_i$ . A monoid has an identity element  $e \in M_i$  such that  $e * a = a * e = a$  for all  $a \in M_i$ . A left monoid action [1, 23] of  $M_i$  is a set  $F_i$  with an action  $\bullet : M \times F_i \rightarrow F_i$  with the properties:

- associativity** for all  $f, g \in M_i$  and  $x \in F_i$ ,  $f \bullet (g \bullet x) = (f * g) \bullet x$
- identity** for all  $x \in F_i, e \in M_i$ ,  $e \bullet x = x$

As with the fiber  $F$  the total monoid space  $M$  is the cartesian product

$$M = M_0 \times \dots \times M_i \times \dots \times M_n \quad (9)$$

39 of each monoid  $M_i$  on  $F_i$ . The monoid is also added to the specification of the fiber  
 40 ( $c_i, T_i, \mathbb{U}_\sigma M_i$ )

41 Steven's described the measurement scales[14, 27] in terms of the monoid actions on the  
 42 measurements: nominal data is permutable, ordinal data is monotonic, interval data is trans-  
 43 latable, and ratio data is scalable [29]. For example, given the fiber ( $c = \text{temperature}$ ,  $T =$   
 44  $\text{float}$ ,  $\mathbb{U}_\sigma = \mathbb{R}$ ) which is interval data:

- 45 • monoid operator addition  $* = +$
- 46 • monoid operations:  $f : x \mapsto x + 1$ ,  $g : x \mapsto x + 2$
- 47 • monoid action operator composition  $\bullet = \circ$

then the translation monoid actions on temperature satisfy the condition

$$\begin{array}{ccc} \mathbb{R} & & \\ \downarrow_{x+1^\circ} & \searrow^{(x+1^\circ) \circ (x+2^\circ)} & \\ \mathbb{R} & \xrightarrow{x+2^\circ} & \mathbb{R} \end{array} \quad (10)$$

48 where  $1^\circ$  and  $2^\circ$  are valid distances between two temperatures  $x$ .

#### 49 1.1.3 Continuity: Base Space $K$

50 The advantage of fiber bundles is they provide a way to encode the continuity in a dataset  
 51 as the base space  $K$  without making assumptions as to what that continuity is. In turn this  
 52 representation of continuity can then be used to keep track of how the data fits together,  
 53 for example if a visualization of a very large dataset calls for parallelization.

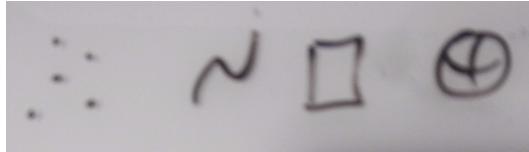


Figure 3: The topological base space  $K$  encodes the connectivity of the data space, for example if the data is independent points or a map or on a sphere

54 As illustrated in figure 3,  $K$  is akin to an indexing space into  $E$  that describes the  
 55 structure of  $E$ .  $K$  can have any number of dimensions and can be continuous or discrete.

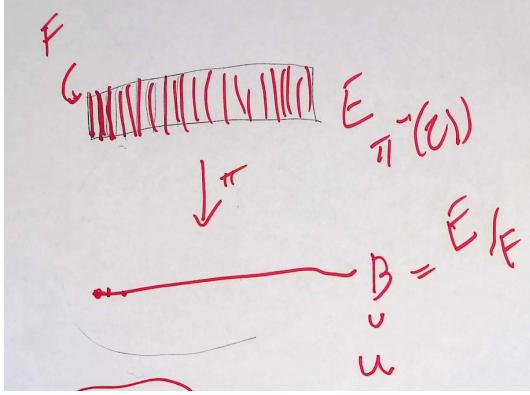


Figure 4: The base space  $E$  is divided into fiber segments  $F$ . The base space  $K$  acts as an index into the records in the fibers. **this figure might be good all the way up top to lay out the components of fb**

56     Formally  $K$  is the quotient space [20] of  $E$  meaning it is the finest space[2] such that  
 57     every  $k \in K$  has a corresponding fiber  $F_k$ [20]. In figure 4,  $E$  is a rectangle divided by  
 58     vertical fibers  $F$ , so the minimal  $K$  for which there is always a mapping  $\pi : E \rightarrow K$  is the  
 59     line.

As with fibers and monoids, we can decompose the total space into components  $\pi : E_i \rightarrow K$  where

$$\pi : E_1 \oplus \dots \oplus E_i \oplus \dots \oplus E_n \rightarrow K \quad (11)$$

60     which is a decomposition of  $F$ . The  $K$  remains the same because the connectivity of  
 61     records does not change just because there are fewer elements in each record.

62     The datasets in figure 5 have the same fiber of (temperature, time). In figure 5a the  
 63     fibers lie over discrete  $K$  such that the records in the datasets in the fiber bundles are  
 64     discrete. The same fiber in figure 5b lies over a continuous interval  $K$  such that the records  
 65     are samples from a continuous function defined on  $K$ .

#### 66 1.1.4 Data: Sections $\tau$

While the fiber and base space describe the general structure of all data that lives in the fiber bundle, the sections  $\tau : K \rightarrow E$  define the datasets that live in the fiber. We generalize Spivak's description of the section as a table of records [25] to any sort of structured dataset such that

$$\begin{array}{ccc} F & \hookrightarrow & E \\ & \pi \downarrow \nearrow \tau & \\ & K & \end{array} \quad (12)$$

such that there is always a map  $\pi(\tau(k)) = k$ . There can be many sections  $\tau$ ; the space of global sections is  $\Gamma(E)$ . For a trivial fiber bundle, the section is

$$\tau(k) = (k, (g_{F_0}(k), \dots, g_{F_n}(k))) \quad (13)$$

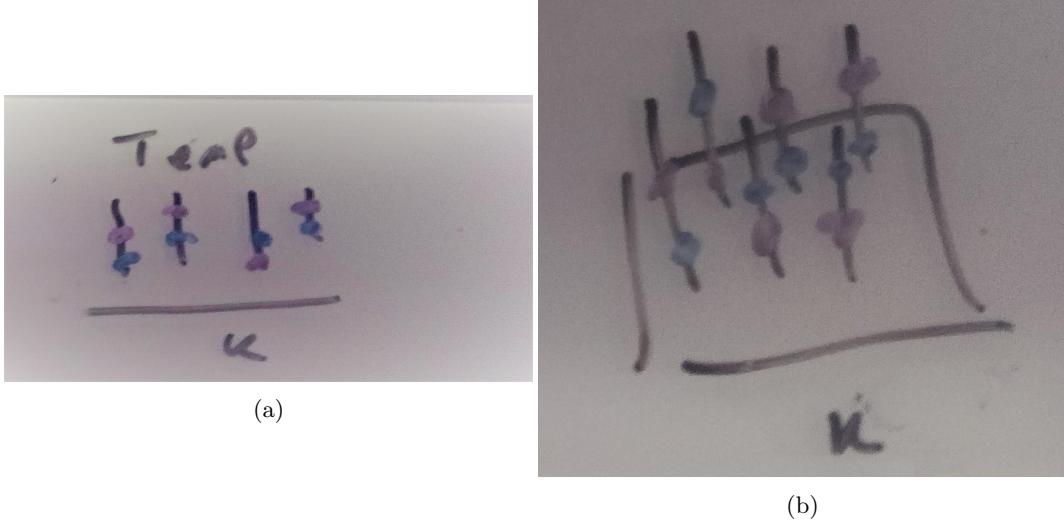


Figure 5: These two datasets have the same (time, temperature) fiber. In figure ?? the total space  $E$  is discrete over points  $k \in K$ , meaning the records in the fiber are also discrete. In figure ??  $E$  lies over the continuous interval  $K$ , meaning the records in the fiber are sampled from a continuous space. *revamp figure: F=Plane, k1 = dots, k2=line*

where  $g : K \rightarrow F$  is the index function into the fiber. Because we can decompose the bundle and the fiber, we can formulate  $\tau$  as

$$\tau = (\tau_0, \dots, \tau_i, \dots, \tau_n) \quad (14)$$

<sup>67</sup> where each section  $\tau_i$  is a variable or set of variables.

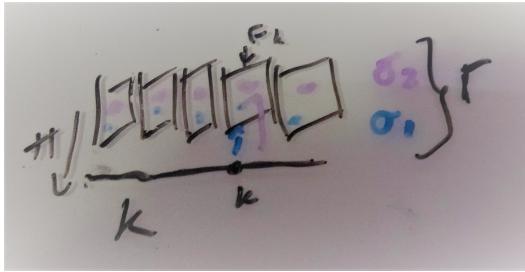


Figure 6: Fiber (time, temperature) with an interval  $K$  basespace. The sections  $\tau_i$  and  $\tau_j$  are constrained such that the time variable must be monotonic, which means each section is a timeseries of temperature values. They are included in the global set of sections  $\tau_1, \tau_2 \in \Gamma(E)$

<sup>68</sup> In the example in figure 6, the fiber is *(time, temperature)* as described in figure 2  
<sup>69</sup> and the base space is the interval  $K$ . The section  $\tau_i$  resolves to a series of monotonically  
<sup>70</sup> increasing in time records of (time, temperature) values. Section  $\tau_j$  returns a different  
<sup>71</sup> timeseries of (time, temperature) values. Both sections are included in the global set of  
<sup>72</sup> sections  $\tau_1, \tau_2 \in \Gamma(E)$ .

73    **1.1.5 Sheaf and Stalk**

74    Often a graphic may need to be updated with live data or support zooming in on a segment  
 75    of the dataset; to support working with a subset of data, we can use the sheaf  $\mathcal{O}(E)$ . All fiber  
 76    bundles are locally trivial, which means that  $E$  restricted over a small enough neighborhood  
 77     $U \subset K$  is a locally trivial bundle over  $U$ [15]. The sheaf  $\mathcal{O}(E)$  is the localized section of  
 78    fibers  $\iota^*\tau : U \rightarrow \iota^*E$

$$\begin{array}{ccc} \iota^*E & \xleftarrow{\iota^*} & E \\ \pi \downarrow \lrcorner^{\iota^*\tau} & & \pi \downarrow \lrcorner^\tau \\ U & \xleftarrow{\iota} & K \end{array} \quad (15)$$

79    pulled back over the neighborhood  $U$  via the inclusion map  $\iota : U \rightarrow K$ . The localized section  
 80    is the germ  $\xi^*\tau$ . The neighborhood of points  $k_i$  surrounding the point  $k$  the sheaf lies over  
 81    is the stalk  $\mathcal{F}_b$  [24, 26]. While  $E$  is only the fiber  $F_k$  over a specific  $k$ , the stalk includes  
 82    nearby records because the sheaf lies over the neighborhood  $U$ . While this can be useful  
 83    for visual transforms, often the extra needed information can be found in the smaller jet  
 84    bundle  $\mathcal{J}$  [13, 19]. For example, line thickness requires the derivative of the given position  
 85    to be rendered, which can be found in  $E' = E + \mathcal{J}(E)$

86    **1.2 Graphic:  $H$**

87    We can separate the structure of the graphic from the properties of the output format by  
 88    modeling the space of graphics as a fiber bundle  $(H, S, \pi, D)$ . As with data, the fiber bundle  
 89    is for a class of graphics with shared base space  $S$ ( 1.2.1) and fiber  $D$ ( 1.2.2) and the sections  
 90     $\rho$ ( 1.2.3) encode a graphic where the visual characteristics are fully specified.

91    **1.2.1 Idealized Display  $D$**

The fiber  $D$  is an idealized infinite resolution version of the target display space, for example  
 a 2D screen or 3D printer. In this work, we assume a 2D opaque image  $F = \mathbb{R}^5$  with elements

$$(x, y, r, g, b) \in D \quad (16)$$

92    such that a rendered graphic only consists of 2D position and color. To support overplotting  
 93    and transparency, the fiber could be  $F = \mathbb{R}^7$  such that  $(x, y, z, r, g, b, a) \in D$  specifies the  
 94    target display. The location coordinates  $x$  and  $y$  are defined in terms of the display, while  
 95    the  $(r, g, b)$  values are filled in via a lookup on  $S$ .

96    **1.2.2 Continuity of the Graphic  $S$**

An assumption of graphical representations is that they match the continuity of the data[8,  
 28], but the underlying topology  $S$  of a graphic may need more dimensions than the data  
 topology  $K$  so that the glyph can be defined in  $F$ . Therefore we define the base space  
 mapping from graphic  $S$  to data  $K$

$$\begin{array}{ccc} E & & H \\ \pi \downarrow & & \pi \downarrow \\ K & \xleftarrow{\xi} & S \end{array} \quad (17)$$



Figure 7: The scatter and line graphic base spaces have one more dimension of continuity than  $K$  so that  $S$  can encode physical aspects of the glyph, such as shape (a circle) or thickness. The heatmap has the same continuity in the graphic  $S$  as in the data  $K$ . **add  $\alpha, \beta$  coordinates to figures**

as the deformation retraction [21]  $\xi : S \rightarrow K$  that goes from a region  $s \in S_k$  to its associated point  $s$ , such that when  $\xi(s) = k$ ,  $\xi^*\tau(s) = \tau(s)$ . While dimensions can be added to  $S$ , it retains the same continuity as  $K$ .

In figure 7 each disk  $S_k$  indexes how elements in  $D$  are glued together to generate a single glyph that is the visual representation of a single record in  $F_k$ . For the line, the region  $\beta$  over a point  $\alpha_i$  specifies the thickness of the line in  $S$  for the corresponding  $\tau$  on  $K$ . The heatmap has the same continuity in data space and graphic space such that no extra dimensions are needed.

### 1.2.3 Renderable Glyph $\rho$

A section  $\rho : S \rightarrow H$  defines a piece of the graphical representation of the data. Evaluated on a single  $s$   $\rho$  returns a single element in  $H$ . For a 2D screen, the pixel is defined as a region  $p = [y_{top}, y_{bottom}, x_{right}, x_{left}]$  of the rendered graphic. Since the  $x$  and  $y$  in  $p$  are in the same coordinate system as the  $x$  and  $y$  components of  $D$  the inverse map of the bounding box  $S_p = \rho_{xy}^{-1}(p)$  is a region  $S_p \subset S$ . Integrating over this region on  $S$

$$r_p = \iint_{S_p} \rho_r(s) ds^2 \quad (18)$$

$$g_p = \iint_{S_p} \rho_g(s) ds^2 \quad (19)$$

$$b_p = \iint_{S_p} \rho_b(s) ds^2 \quad (20)$$

yields the color of the pixel  $p$ .

As shown in figure 8, the output space queries into the graphic bundle to render the image. We select a pixel  $p$  in the output space, inverse map the region of the pixel into  $S_p \subset S$ , then compute the section  $gsection$  over the region  $S_p$ . The section yields the set of elements in  $D$  that specify the  $(r, g, b)$  values corresponding to the region  $p$ . The color of the pixel is then obtained by taking the integral of  $\rho_{rgb}(S_p)$ .

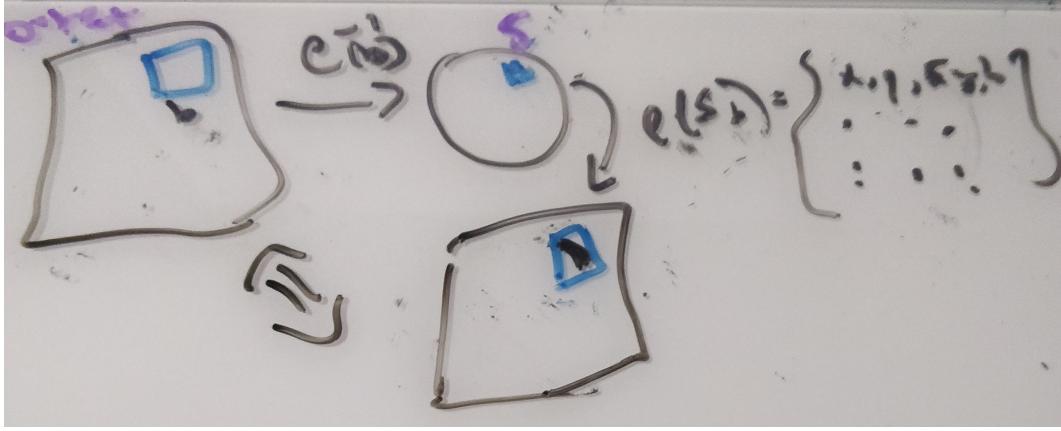


Figure 8: To render a graphic, a pixel  $p$  is selected in the display space, which is defined in the same coordinates as the  $x$  and  $y$  components in  $D$ . The inverse mapping  $\rho_{xy}(p)$  returns a region  $S_p \subset S$ .  $\rho(S_p)$  returns the list of elements  $(x, y, r, g, b) \in D$  that lie over  $S_p$ . The integral over the  $(r, g, b)$  elements is the color of the pixel.

### 112 1.3 Artist

The artist is the function that converts data into graphics; its name is taken from the analogous part of Matplotlib[12] that builds visual elements to pass off to the renderer. The artist  $A$  is a mapping from  $E$  padded with data from  $\mathcal{J}(E)$  to a graphic that is a section  $\rho$  in  $\Gamma(H)$

$$\begin{array}{ccccc}
 E' & \xrightarrow{\nu} & V & \xleftarrow{\xi^*} & \xi^* V \\
 & \searrow \pi & \downarrow \pi & & \downarrow \xi^* \pi \\
 & & K & \xleftarrow{\xi} & S
 \end{array} \quad (21)$$

113 with an intermediate fiber bundle  $V$  to hold visual representations and stages

- 114 1.  $\xi$  binding the continuity in the graphic to the continuity in the data ( 1.2.2)
- 115 2.  $\nu$  conversion of data into visual characteristics ( 1.3.2)
- 116 3.  $Q$  assembly of visual variables into a glyph ( 1.3.3)

117 of the visual transformation illustrated in figure 1. The functions  $\xi$   $\nu$  and  $Q$  are defined  
 118 such that they can be evaluated on a single section  $\tau$ , which allows the artist to be imple-  
 119 mented such that it does not need all the data. This allows for artists tuned to distributed  
 120 and streaming data.

#### 121 1.3.1 Visual Fiber Bundle $V$

122 The visual fiber bundle  $(V, K, \pi, P)$  has section  $\mu : V \rightarrow K$  that resolves to a visual  
 123 variable [3, 18] in fiber  $P$ . The fiber space  $P$  is defined in terms of the parameters of  
 124 the visualization specification- for example aesthetics in ggplot [30], channels in vega[22] or  
 125 parameters in VTK[9] and Matplotlib.

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

Table 1: Some possible components of the fiber  $P$  for a visualization function implemented in Matplotlib

126      Table 1 is a sample of the fiber space for Matplotlib [11]. A section  $\mu$  is a tuple of  
 127      visual values that specifies the visual characteristics of a part of the graphic. For example,  
 128      given a fiber of  $\{xpos, ypos, color\}$  one possible section could be  $\{.5, .5, (255, 20, 147)\}$ . The  
 129       $\text{codomain}(\nu_i)$  determines the monoid actions on  $\mu_i$ . These fiber components are implicit  
 130      in the library, by making them explicit as components of the fiber we can build consistent  
 131      definitions and expectations of how these parameters behave.

### 132      1.3.2 Visual Channels

As introduced in section ??, there are many ways to encode data visually. We define the visual transformers  $\nu$  as the set of independent conversion functions

$$\{\nu_0, \dots, \nu_n\} : \{\tau_0, \dots, \tau_n\} \mapsto \{\mu_0, \dots, \mu_n\} \quad (22)$$

where  $\nu_i : \tau_i \mapsto \mu_i$  is an equivariant map such that there is a monoid homomorphism from  $F_i$  to  $v\text{fiber}_i$ . A validly constructed  $\nu$  is one where the diagram of the monoid transform  $m$

$$\begin{array}{ccc} E_i & \xrightarrow{\nu_i} & V_i \\ m_x \downarrow & & \downarrow m_v \\ E_i & \xrightarrow{\nu_i} & V_i \end{array} \quad (23)$$

commutes such that  $\nu_i(m_x(E_i)) = m_v(\nu_i(E_i))$ . This equivariance constraint yields guidance on what makes for an invalid transform. For example, the conversion  $\nu_i(x) = .5$  does not commute under translation monoid action  $t(x) = x + 2$

$$\nu(t(x + 2)) \stackrel{?}{=} \nu(x) + \nu(2) \quad (24)$$

$$.5 \neq .5 + .5 \quad (25)$$

133      On the other hand figure ?? illustrates a valid  $\nu$  mapping from **Strings** to symbols. The  
 134      group action on these sets is permutation, so shuffling the words must have an equivalent

135 shuffle of the symbols they are mapped to. To preserve ordinal and partial order monoid  
 136 actions,  $\nu$  must be a monotonic function such that given  $x_1, x_2 \in E_i$ , if  $delement_1 \leq$   
 137  $delement_2$  then  $\nu(x_1) \leq \nu(x_2)$ . For interval scale data,  $\nu$  is equivariant under translation  
 138 monoid actions if  $\nu(x + c) = \nu(x) + \nu(c)$ . For ratio data, there must be equivalent scaling  
 139  $\nu(xc) = \nu(x)\nu(c)$ .

140 **1.3.3 Assembling Marks**

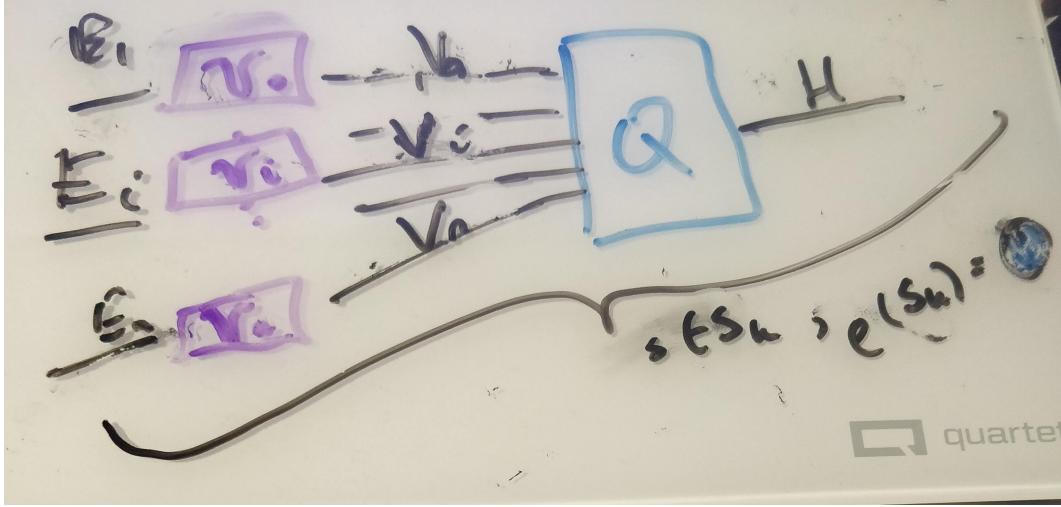


Figure 9:  $\nu$  functions convert data  $\tau_i$  to visual characteristics  $\mu_i$ , then  $Q$  assembles  $\mu_i$  into a graphic  $\rho$  such that there is a map  $\xi$  preserving the continuity of the data.  $\rho$  applied to a region of connected components  $S_j$  generates a graphical mark.

141 As shown in figure 9, the assembly function  $Q$  combines the fiber  $F_i$  wise  $\nu$  transforms into a  
 142 single glyph. Together,  $\nu$  and  $Q$  are a map-reduce operation: map the data into their visual  
 143 encodings, reduce the encodings into a glyph. As with  $\nu$  the constraint on  $Q$  is that for  
 144 every monoid actions on the input  $\mu$  there is a corresponding monoid action on the output  
 145  $\rho$ .

146 Since we define the equivariant map as  $Q : \mu \mapsto \rho$ , we define an action on the subset  
 147 of graphics  $Q(\Gamma(V)) \in \Gamma(H)$  that  $Q$  can generate. We then define the constraint on  $Q$  such  
 148 that if  $Q$  is applied to  $\mu, \mu'$  that generate the same  $\rho$  then the output of both sections acted  
 149 on by the same monoid  $m$  must be the same.

Lets call the visual encodings  $\Gamma(V) = X$  and the graphic  $Q(\Gamma(V)) = Y$ . If for all monoids  
 $m \in M$  and for all  $\mu, \mu' \in X$ , the output is equivalent

$$Q(\mu) = Q(\mu') \implies Q(m \circ \mu) = Q(m \circ \mu') \quad (26)$$

150 then a group action on  $Y$  can be defined as  $m \circ \rho = \rho'$ . The transformed graphic  $\rho'$  is  
 151 equivariant to a transform on the visual bundle  $\rho' = Q(m \circ \mu)$  on a section that  $\mu \in Q^{-1}(\rho)$   
 152 that must be part of generating  $\rho$ .

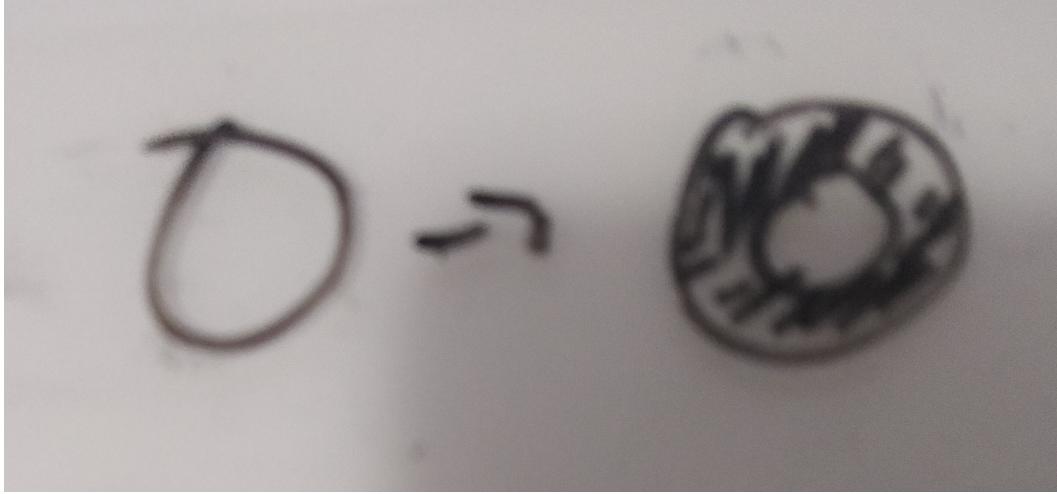


Figure 10: These two glyphs are generated by the same  $Q$  function, but differ in the value of the edge thickness parameter  $\mu_i$ . A valid  $Q$  is one where a shift in  $\mu_i$  is reflected in the glyph generated by  $\rho$ .

153     The glyph in figure 10 has the following characteristics  $P$  specified by  $(xpos, ypos, color, thickness)$   
 154    such that one section is  $\mu = (0, 0, 0, 1)$  and  $Q(\mu) = \rho$  generates a piece of the thin hollow  
 155    circle. The equivariance constraint on  $Q$  is that the action  $m = (e, e, e, x + 2)$ , where  $e$  is  
 156    identity, applied to  $\mu$  such that  $\mu' = (e, e, e, 3)$  has an equivalent action on  $\rho$  that causes  
 157     $Q(\mu')$  to be equivalent to the thicker circle in figure 10.

158     DEGENERATE Q - drawing a blank if this is necessary and if so how  
 159     Check a well defined map  $M \times Y \rightarrow Y$

To output a mark [3, 6],  $Q$  is called with all the regions  $s$  that map back to a set of connected components  $J \subset K$ :

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

160    where the path[7]  $\gamma$  from  $k$  to  $j$  is a continuous function from the interval  $[0, 1]$ . We define  
 161    the mark as the graphic generated by  $Q(S_j)$

$$H \xrightleftharpoons[\rho(S_j)]{\xi(s)} S_j \xrightleftharpoons[\xi^{-1}(J)]{} J_k \quad (28)$$

162    such that for every mark there is at least one corresponding section on  $K$ .

#### 163    1.3.4 Sample Qs

164    In this section we formulate the minimal  $Q$  that will generate distinguishable graphical  
 165    marks: non-overlapping scatter points, a non-infinitely thin line, and a heatmap.

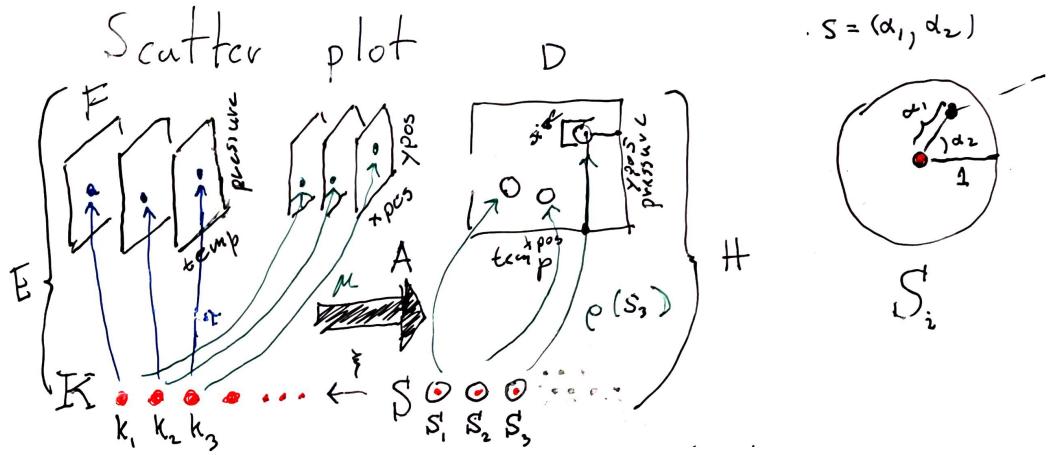


Figure 11: The data is discrete points (temperature, time). Via  $\nu$  these are converted to (xpos, ypos) and pulled over discrete  $S$ . These values are then used to parameterize  $\rho$  which returns a color based on the parameters (xpos,ypos) and position  $\alpha, \beta$  on  $S_k$  that  $\rho$  is evaluated on.

The scatter plot in figure ?? can be defined as  $Q(xpos, ypos)(\alpha, \beta)$  where color  $\rho_{RGB} = (0, 0, 0)$  is defined as part of  $Q$  and  $s = (\alpha, \beta)$  defines the region on  $S$ . The position of this swatch of color can be computed relative to the location on the disc  $S_k$  as shown in figure 11:

$$x = size \bullet \alpha \bullet \cos(\beta) + xpos \quad (29)$$

$$y = size \bullet \alpha \bullet \sin(\beta) + ypos \quad (30)$$

such that  $\rho(s) = (x, y, 0, 0, 0)$  colors the point (x,y) black.

166

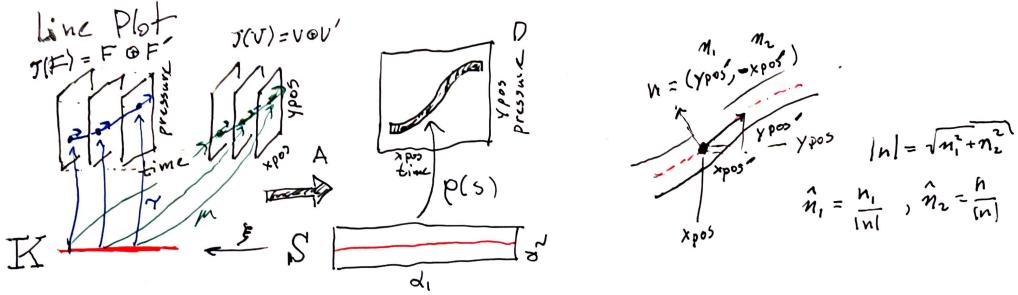


Figure 12: The line fiber (*time, temp*) is thickened with the derivative (*time', temperature'*) because that information will be necessary to figure out the tangent to the point to draw a thick line. This is because the line needs to be pushed perpendicular to the tangent of (*xpos, ypos*). *this is gonna move once this gets regenerated w/ labels* The data is converted to visual characteristics (*xpos, ypos*). The  $\alpha$  coordinates on  $S$  specifies the position of the line, the  $\beta$  coordinate specifies thickness.

The line plot  $Q(xpos, \hat{n}_1, ypos, \hat{n}_2)(\alpha, \beta)$  shown in fig 11 exemplifies the need for the jet discussed in section ???. The line needs to know the tangent of the data to draw an envelope above and below each (*xpos,ypos*) such that the line appears to have a thickness. The magnitude of the thickness is

$$|n| = \sqrt{n_1^2 + n_2^2} \quad (31)$$

such that the normal is

$$\hat{n}_1 = \frac{n_1}{|n|}, \quad \hat{n}_2 = \frac{n_2}{|n|} \quad (32)$$

which yields components of  $\rho$

$$x = xpos(\xi(\alpha)) + \hat{\beta}(n_1)(\xi(\alpha)) \quad (33)$$

$$y = ypos(\xi(\alpha)) + \hat{\beta}(n_2)(\xi(\alpha)) \quad (34)$$

167 where (x,y) look up the position  $\xi(\alpha)$  on the data and then apply thickness  $\beta$  at that  
168 location.

169 **Q: heatmap**

170

Figure 14: Each of these graphics is generated by a different artist  $A$  which is the equivalence class of scatter plots  $A'$   
this is gonna be a whole bunch of scatter plots

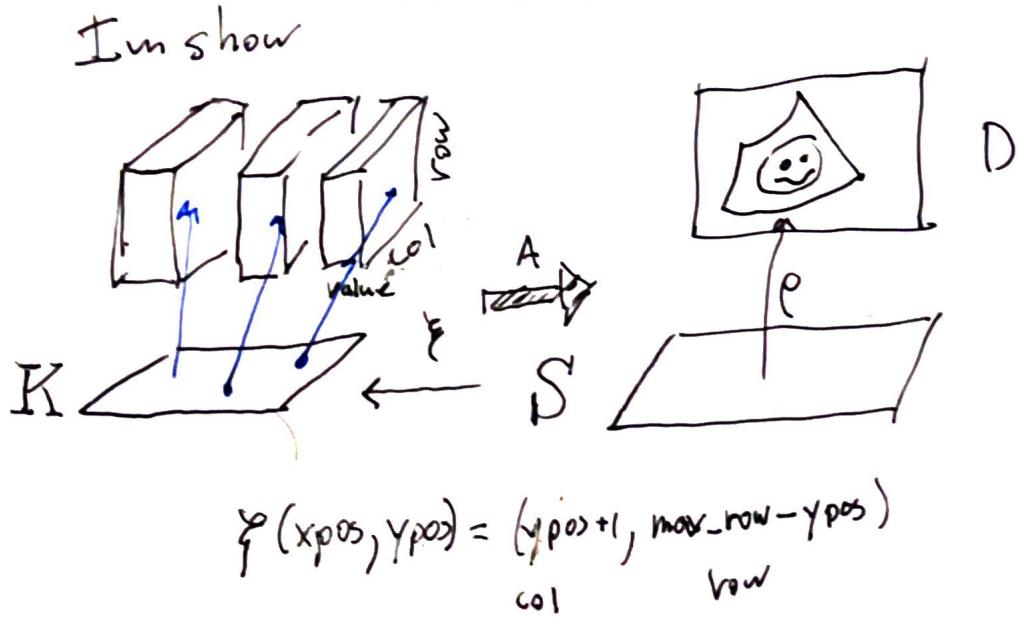


Figure 13: The only visual parameter a heatmap requires is color since  $\xi$  encodes the mapping between position in data and position in graphic.

<sup>171</sup> The heatmap  $Q(\text{color})$  in figure 13 is a direct lookup  $\xi : S \rightarrow K$  such that

$$R = R(\xi(\alpha, \beta)) \quad (35)$$

$$G = G(\xi(\alpha, \beta)) \quad (36)$$

$$B = B(\xi(\alpha, \beta)) \quad (37)$$

<sup>172</sup> where  $\xi$  may do some translating to a convention expected by  $Q$  for example reorientng the  
<sup>173</sup> array such that the first row in the data is at the bottom of the graphic.

### <sup>174</sup> 1.3.5 Equivalence class of artists $A'$

<sup>175</sup> As formulated above, every artist function  $A$  has fixed  $\nu$  and  $Q$  which generates a distinct  
<sup>176</sup> graphic  $\rho$ . It is impractical to implement an artist for every single graphic; instead we  
<sup>177</sup> implement the equivalence class of artists  $\{A \in A' : A_1 \equiv A_2\}$ . Equivalent artists have  
<sup>178</sup> the same fiber bundle  $V$  and same assembly function  $Q$  but act on different sections  $\mu$ .  
<sup>179</sup> To further simplify implementation, we identify a minimal  $P$  associated with each  $A'$  that  
<sup>180</sup> defines what visual characteristics of the graphic must originate in the data  
<sup>181</sup> needs citation, maybe friendlys history or acquired codes of meaning. For example, a scatter plot of red

182 circles is the output of one artist, a scatter plot of green squares the output of another, as are  
 183 the rest of the graphs in figure ???. These two artists are equivalent since their only difference  
 184 is in the literal visual encodings (color, shape). Shape and color could also be defined in  $Q$   
 185 but the position must come from the fiber  $P = (xpos, ypos)$  since fundamentally a scatter  
 186 plot is the plotting of one position against another[8]. We also use this criteria to identify  
 187 derivative types, for example the bubble chart[28] is a type of scatter where by definition  
 188 the glyph size is mapped from the data.

#### 189 1.4 Making the fiber bundle computable

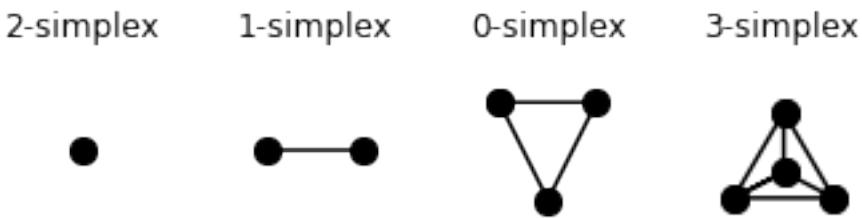


Figure 15: Simplices can encode the connectivity of the data, from fully disconnected (0 simplex) records to all records are connected to at least 3 others

190 One way of expressing the connectivity of records in a dataset is to implement  $K$  as a  
 191 simplicial complex, which is a set of simplices such as those shown in figure 15. The  
 192 advantage of triangulation is that it is general enough to work for more complex topology  
 193 based visualization methods [10] while also providing a consistent interface of vertices, edges,  
 194 and faces for  $\xi$  to map into. When triangulated, the simplices encode the continuity in the  
 195 data

simplex	continuity	$\tau$
vertex	discrete	$\tau(k)$
edge	1D	$\tau(k, \alpha)$
face	2D	$\tau(k, \alpha, \beta)$

Table 2

196 such that each section is bound to a simplex  $k \in K$ . As shown in table 2, in a 1D  
 197 continuous spaces each  $\tau$  lies distance  $\alpha$  along edge  $k$ , while in a 2D continuous space each  
 198  $\tau$  lies at coordinate  $\alpha, \beta$  on the face  $k$ . This is directly analogous to indexing to express  
 199 connectivity in N-D arrays, while also natively supporting graphs and trees as they are  
 200 simplicial complexes of nodes and edges. Path connected components are then sections  
 201 where edges or faces meet.

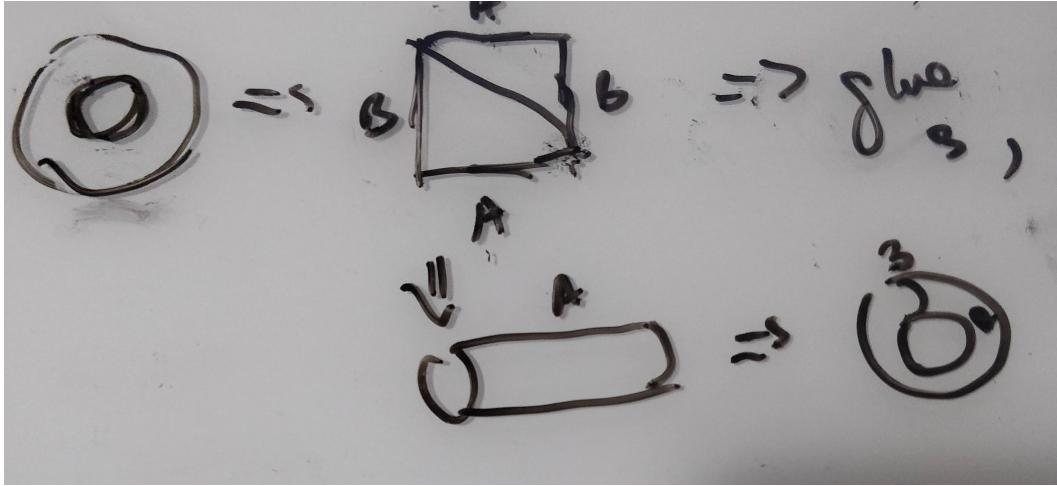


Figure 16: The torus  $E$  is unraveled into a simplicial complex of 2 faces  $K$ . Transition functions are defined on the edges of  $K$  such that surface can be glued back into the torus.  
add cross sections  $a$  and  $b$  to ring and color same as edges in complex

202 One way of encoding the torus in figure 16 while retaining the continuity of both cross  
 203 sections  $a, b$  is to unravel it into a simplicial complex of two triangles with labeled edges.  
 204 Transition functions  $\delta$  are defined on the edges of  $K$  such that  $a$  can be glued to  $a'$  and  $b$  to  $b'$  to  
 205 reconstruct the torus. This simplicial complex is then used as the base space encoding the  
 206 continuity of data that lies in the torus. A constraint on the transition functions is that the  
 207 monoid actions on the fibers on the edges of  $E$  are commutative  $M * F \mapsto \delta(MF) = M * \delta(F)$

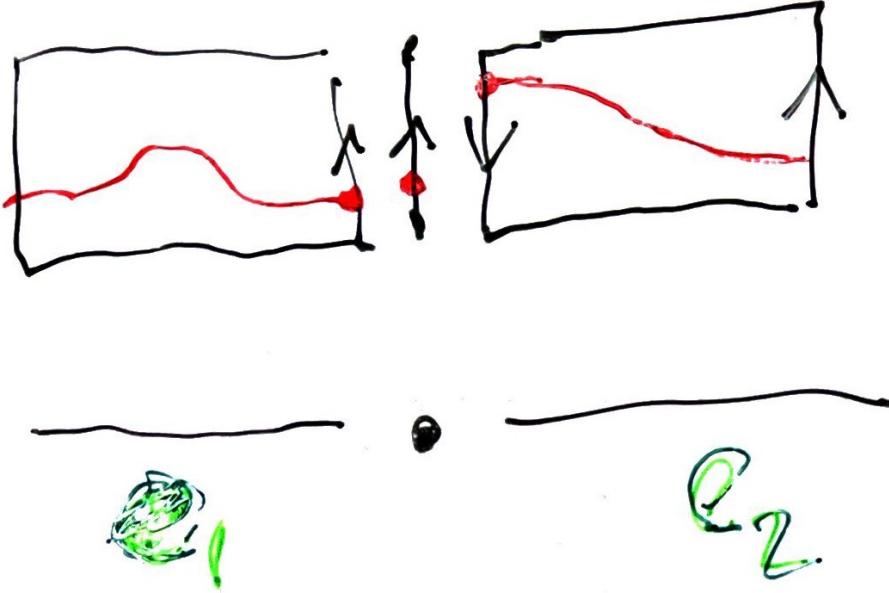


Figure 17: Many non-trivial spaces can be made locally trivial by dividing  $E$  into locally trivial subspaces and defining transition functions between the edges on  $K$  for how to glue the two subspaces such that the  $\tau$  are continuous.

208     Another advantage of triangulation is that it provides a way to encode non-trivial  
 209     structures such as the Möbius strip[16]. As shown in figure 17, one way of making the  
 210     Möbius strip trivial is to separate it into two spaces  $E_1$  and  $E_2$  and then define transition  
 211     functions that specify that the edges of  $E_1$  need to be reversed to line up with  $E_2$  such that  
 212     the sections along the edges meet. As with the torus, the transition functions must preserve  
 213     monoid commutativity.