

# Topologically Equivariant Artist Model

Hannah Aizenman  
Mikael Vejdemo-Johansson  
Thomas Caswell  
Michael Grossberg

Department of Computer Science, The Graduate Center

January 28, 2026

# About

## Me

- PhD candidate in Computer Science
- Matplotlib Community Manager & Core Developer
- socials/github: story645

## Project

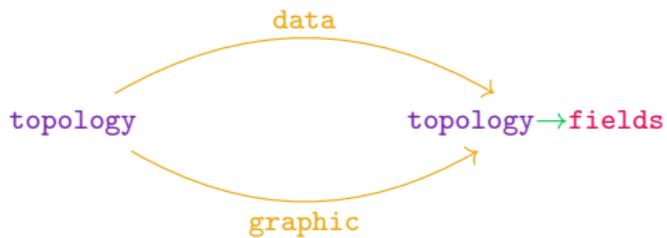
- Funded by Chan Zuckerberg Initiative EOSS 1 & 3

# Roadmap



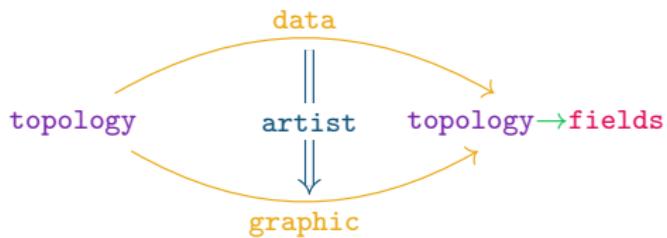
- uniform API for arbitrary datasets in arbitrary data containers

# Roadmap



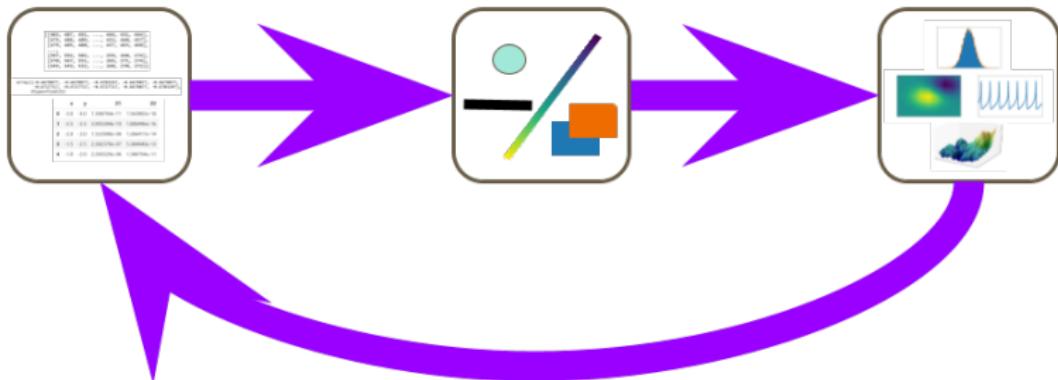
- uniform API for arbitrary datasets in arbitrary data containers
- generalizable methodology for expressing structure preservation

# Roadmap

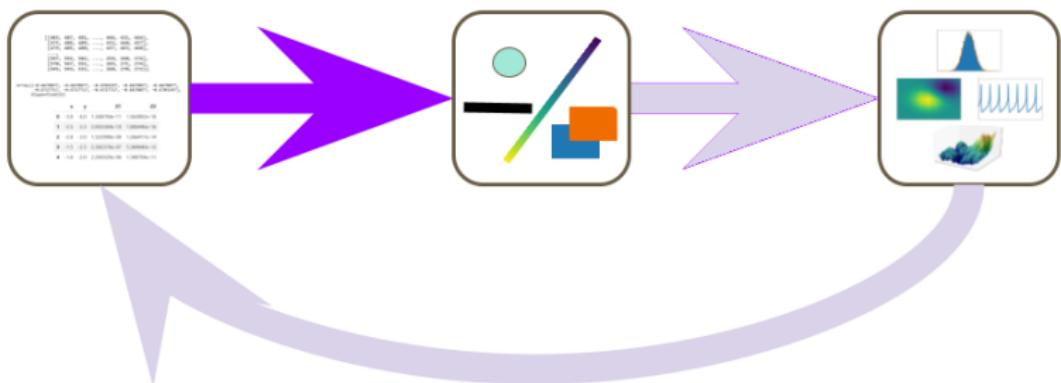


- uniform API for arbitrary datasets in arbitrary data containers
- generalizable methodology for expressing structure preservation
- framework that translates into coding specifications

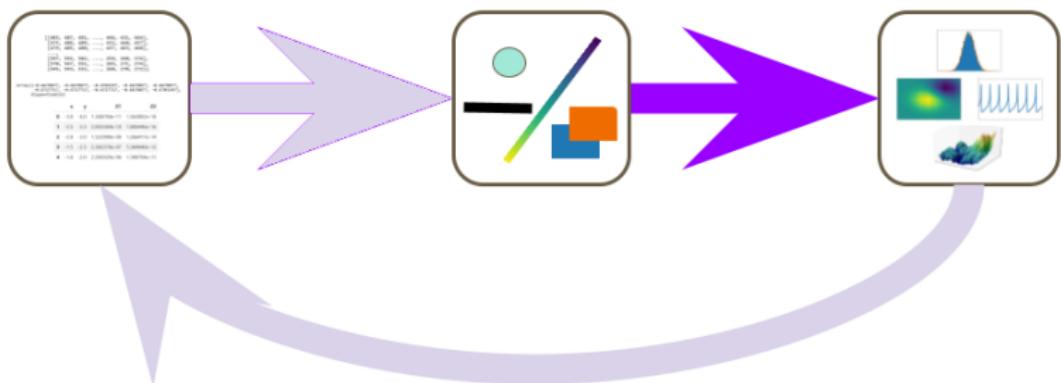
# What do visualization libraries do?



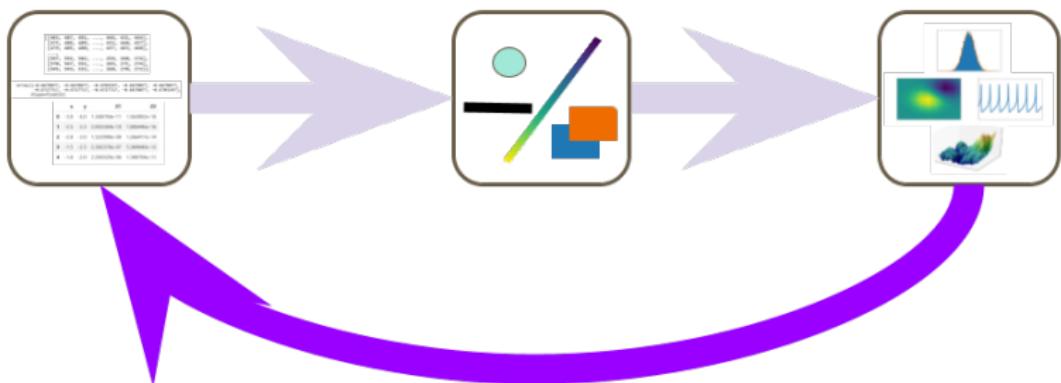
# What do visualization libraries do?



# What do visualization libraries do?



# What do visualization libraries do?



# Matplotlib



# How do we express structure?

**field** is a set of values of the same type, e.g. one column of a table or the pixels of an image

**topology** is the connectivity and relative positioning of elements in a dataset [1].

# Structure Preservation: Groups

## Definition

[2] An **action** of  $G = (G, \circ, e)$  on  $X$  is a function  $\phi : G \times X \rightarrow X$ .  
An action has the properties of identity  $act(e, x) = x$  for all  $x \in X$   
and associativity  $\phi(g, \phi(f, x)) = \phi(f \circ g, x)$  for  $f, g \in G$ .

# Structure Preservation: Groups

$\tau = \text{data}$	$\hat{\phi}_E^* \tau = \text{data.T}$	$\tilde{\phi}_E \tau = \text{data}*2$	$\phi_E \tau = \text{data.T}^2$																								
<table border="1"><tbody><tr><td>0</td><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td><td>5</td></tr></tbody></table>	0	1	2	3	4	5	<table border="1"><tbody><tr><td>0</td><td>3</td></tr><tr><td>1</td><td>4</td></tr><tr><td>2</td><td>5</td></tr></tbody></table>	0	3	1	4	2	5	<table border="1"><tbody><tr><td>0</td><td>2</td><td>4</td></tr><tr><td>6</td><td>8</td><td>10</td></tr></tbody></table>	0	2	4	6	8	10	<table border="1"><tbody><tr><td>0</td><td>6</td></tr><tr><td>2</td><td>8</td></tr><tr><td>4</td><td>10</td></tr></tbody></table>	0	6	2	8	4	10
0	1	2																									
3	4	5																									
0	3																										
1	4																										
2	5																										
0	2	4																									
6	8	10																									
0	6																										
2	8																										
4	10																										

## Definition

[2] An **action** of  $G = (G, \circ, e)$  on  $X$  is a function  $\phi : G \times X \rightarrow X$ .  
An action has the properties of identity  $\text{act}(e, x) = x$  for all  $x \in X$  and associativity  $\phi(g, \phi(f, x)) = \phi(f \circ g, x)$  for  $f, g \in G$ .

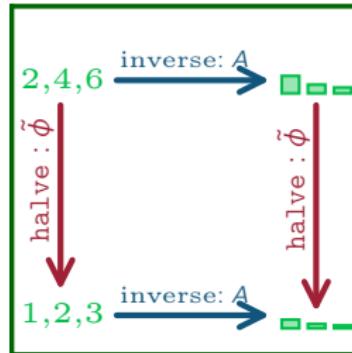
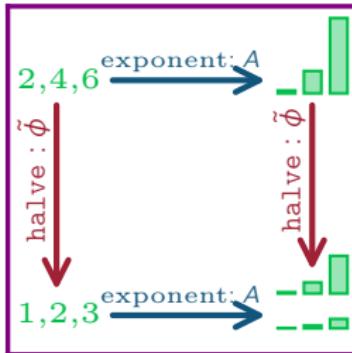
# Fields: Equivariance

## Definition

Given a group  $G$  that acts on both the input  $X$  and the output  $Y$  of a function  $f : X \rightarrow Y$ , a function  $f$  is **equivariant** when  $f(\text{act}(g, x)) = \text{act}(g, f(x))$  for all  $g$  in  $G$  and for all  $x$  in  $X$  [3]

# Fields: Equivariance

Is  $\text{data} \mapsto \square$  equivariant w.r.t. scaling?



## Definition

Given a group  $G$  that acts on both the input  $X$  and the output  $Y$  of a function  $f : X \rightarrow Y$ , a function  $f$  is **equivariant** when  $f(\text{act}(g, x)) = \text{act}(g, f(x))$  for all  $g$  in  $G$  and for all  $x$  in  $X$  [3]

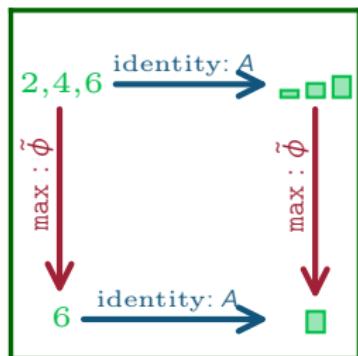
# Fields: Homomorphism

## Definition

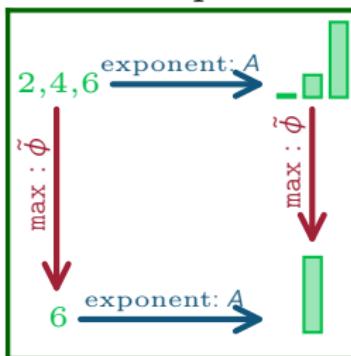
Given the function  $f : X \rightarrow Y$ , with operators  $(X, \circ)$  and  $(Y, *)$ , a function  $f$  is **homomorphic** when  $f(x_1 \circ x_2) = f(x_1) * f(x_2)$  and preserves identities  $f(I_x) = I_y$  all  $x, y \in X$  [2]

# Fields: Homomorphism

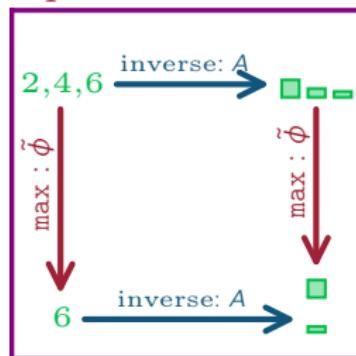
Is  $\text{data} \mapsto \square$  homomorphic w.r.t. partial order?



YES



YES



NO

## Definition

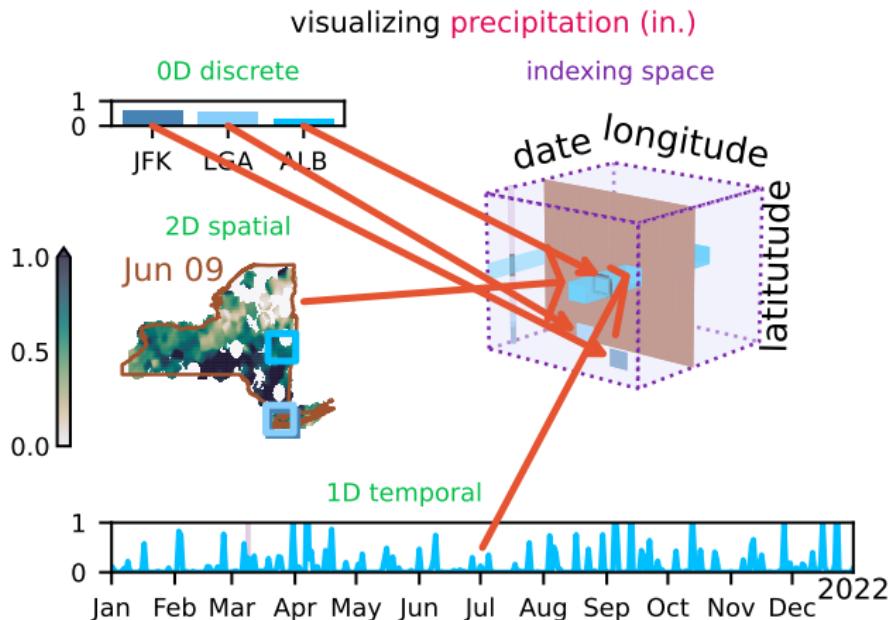
Given the function  $f : X \rightarrow Y$ , with operators  $(X, \circ)$  and  $(Y, *)$ , a function  $f$  is **homomorphic** when  $f(x_1 \circ x_2) = f(x_1) * f(x_2)$  and preserves identities  $f(I_x) = I_y$  all  $x, y \in X$  [2]

# Topology: Homeomorphism

## Definition

A function  $f$  is a *homeomorphism* if it is bijective, continuous, and has a continuous inverse function  $f^{-1}$ .

# Topology: Homeomorphism



## Definition

A function  $f$  is a *homeomorphism* if it is bijective, continuous, and has a continuous inverse function  $f^{-1}$ .

# Structure Preservation

- Bertin: equivalent data and visual field properties [4]
- Mackinlay: homomorphic fields and equivalent topology [5]
- Wilkinson: homomorphic scales and equivalent topology [1]
- Kindleman & Schieidegger: equivariant scales, invariant data representation, equivalent topology [6]

## contribution

explicit homeomorphic topology, equivariant actions on fields and topology

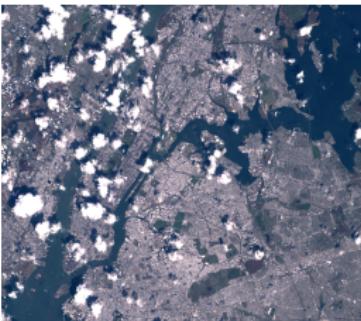
# Domain Specific Library: library assumes structure [19]

DATE	LATITUDE	LONGITUDE	PREC (in)	NAME
2023-01-01	43.1549 <sup>1</sup>	-77.8785 <sup>1</sup>	0.2009	ROCHESTER GTR INTL AP
2023-01-01	45.5 <sup>1</sup>	-35.9 <sup>1</sup>	0.0000	STONYBROOK NEW YORK
2023-01-01	40.7431 <sup>1</sup>	-73.8892 <sup>1</sup>	0.2998	AUSTRALIAN AP
2023-01-01	43.8 <sup>1</sup>	-75.7 <sup>1</sup>	0.0000	SCHROON LAKE NEW YORK
2023-01-01	43.9978 <sup>1</sup>	-73.6511 <sup>1</sup>	0.0000	SARATOGA SPRINGS NEW YORK
2023-01-01	40.7794 <sup>1</sup>	-73.8883 <sup>1</sup>	0.4614	NEW YORK LAGUARDIA AP
2023-01-01	40.7794 <sup>1</sup>	-73.8883 <sup>1</sup>	0.4614	NEW YORK JORDAN AP
2023-01-01	43.1111 <sup>1</sup>	-76.1109 <sup>1</sup>	0.0004	SYRACUSE HAMILTON INTL AP
2023-01-01	40.7939 <sup>1</sup>	-73.1632 <sup>1</sup>	0.2984	ISLE OF MACARTHUR AP
2023-01-01	43.35 <sup>1</sup>	-73.6187 <sup>1</sup>	0.1181	GLENNS FALLS AP

- ggplot[7]
- Vega[10]
- Altair[13]
- Tableau [16]
- [17, 18]

# Domain Specific Library: library assumes structure [19]

DATE	LATITUDE	LONGITUDE	PREC (in)	NAME
2023-01-01	43.15497	-77.87857	0.2000	ROCHESTER GTR INTL AP
2023-01-01	45.5	-35.9	0.0000	STONYBROOK NEW YORK
2023-01-01	40.7431	-73.8892	0.2998	AUSTRALIAN AP
2023-01-01	43.8	-75.7	0.0000	SCHROON LAKE NEW YORK
2023-01-01	43.9978	-73.6511	0.0000	SARATOGA SPRINGS NEW YORK
2023-01-01	40.7794	-73.8867	0.4614	NEW YORK LAGUARDIA AP
2023-01-01	40.7794	-73.8867	0.4614	NEW YORK J F肯尼迪 AP
2023-01-01	43.1111	-76.1109	0.0000	SYRACUSE HAMILTON INTL AP
2023-01-01	40.7939	-73.1632	0.2998	ISLE OF MACARTHUR AP
2023-01-01	43.35	-73.6187	0.1181	GLENNS FALLS AP

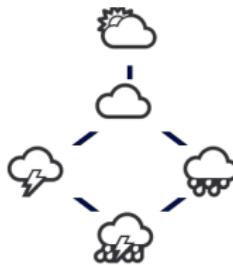


ggplot[7]  
Vega[10]  
Altair[13]  
Tableau [16]  
[17, 18]

ImageJ[8]  
ImagePlot[11]  
Napari[14]

# Domain Specific Library: library assumes structure [19]

DATE	LATITUDE	LONGITUDE	PREC (in)	NAME
2023-01-01	43.15497	-77.87857	0.2209	ROCHESTER GTR INTL AP
2023-01-01	45.5	-35.9	0.0000	STONYBROOK NEW YORK
2023-01-01	42.7431	-73.8892	0.2998	AUSTRALY AP
2023-01-01	43.8	-75.7	0.0000	SCHROON LAKE NEW YORK
2023-01-01	43.9978	-73.6511	0.0000	SARATOGA NEW YORK
2023-01-01	40.7794	-73.8867	0.4614	NEW YORK LAGUARDIA AP
2023-01-01	40.7794	-73.8867	0.4614	NEW YORK JF肯尼迪 AP
2023-01-01	43.1113	-76.1109	0.0000	SYRACUSE HANCOCK INTL AP
2023-01-01	40.7939	-73.1632	0.2984	ISLE OF MACARTHUR AP
2023-01-01	43.35	-73.6187	0.1181	GLENS FALLS AP



ggplot[7]  
Vega[10]  
Altair[13]  
Tableau [16]  
[17, 18]

ImageJ[8]  
ImagePlot[11]  
Napari[14]

Gephi[9]  
Graphviz[12]  
Networkx[15]

Building Block Library[29]: visual algorithms assume structure [30]

1. Matplotlib[20] → Seaborn[21], xarray [22]
2. D3 [23]
3. VTK [24, 25], MayaVi[26] → Titan[27], ParaView[28]

# Design Composable Structure Preserving API

Fiber Bundles Butler: "unified, dimension-independent framework" that expresses data as the mapping between continuity and fields [31, 32]

Simplicial Databases Spivak: Add rich typing for fields to Butler [33]

Category Theory Language express constraints in specifications [34]

Sheaves Ghrist "algebraic data structure" for representing data over topological spaces [35]

# Expressive Types

```
dataset : topology → fields
```

# Fiber Bundle

## Definition

A **fiber bundle**  $(E, K, \pi, F)$  is a structure with topological spaces  $E, F, K$  and bundle projection map  $\pi: E \rightarrow K$  [36].

$$F \hookrightarrow E \xrightarrow{\pi} K \tag{1}$$

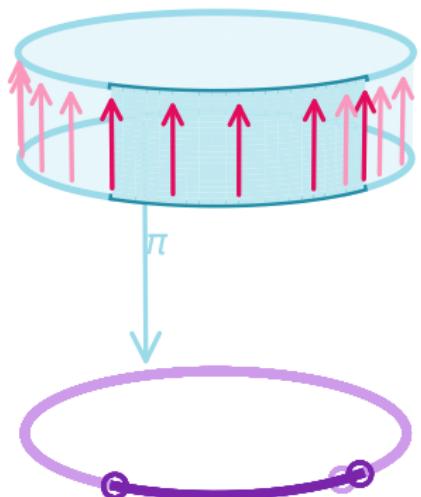
A continuous surjective map  $\pi$  is a **bundle projection** map when

1. all fibers in the bundle are isomorphic. Since all fibers are isomorphic  $F \cong F_k$  for all points  $k \in K$ , there is a uniquely determined **fiber space**  $F$  given by the preimage of the projection  $\pi$  at any point  $k$  in the **base space**  $K$ :  $F = \pi^{-1}(k)$ .
2. each point  $k$  in the **base space**  $K$  has an open neighborhood  $U_k$  such that the **total space**  $E$  over the neighborhood is locally trivial.

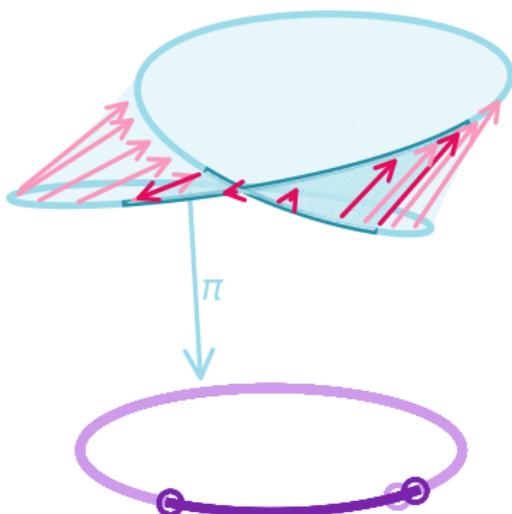
**Local triviality** means  $E|_U = U \times F$ .

# Fiber Bundle

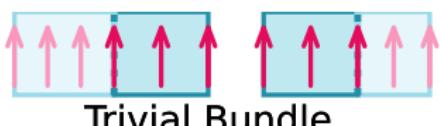
(  $E_c, K_c, \pi, F_c$  )



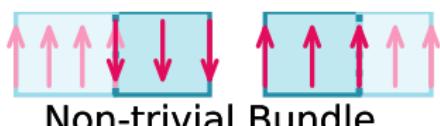
(  $E_m, K_m, \pi, F_m$  )



Transition  
maps



Trivial Bundle

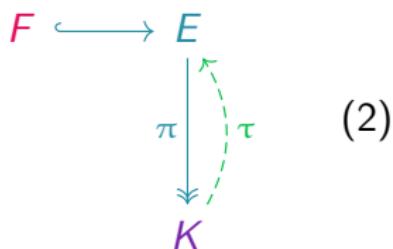


Non-trivial Bundle

# Fiber Bundle: section

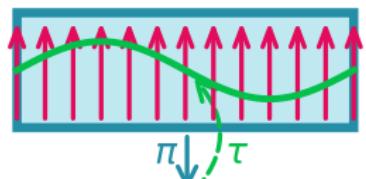
## Definition

A **section**  $\tau: K \rightarrow E$  over a fiber bundle is a smooth right inverse of  $\pi(\tau(k)) = k$  for all  $k \in K$



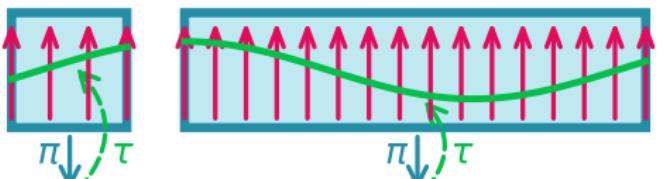
# Fiber Bundle: section

fiber bundle  
 $(E, K, \pi, F)$



$0 \quad \frac{2\pi}{5} \quad \frac{2\pi}{5} \quad 2\pi$

local trivializations  
 $(E_0, K_0, \pi, F_0)$  &  $(E_1, K_1, \pi, F_1)$



$\varepsilon^- \quad \frac{2\pi}{5} \quad \frac{2\pi}{5}$

$\varepsilon^+$

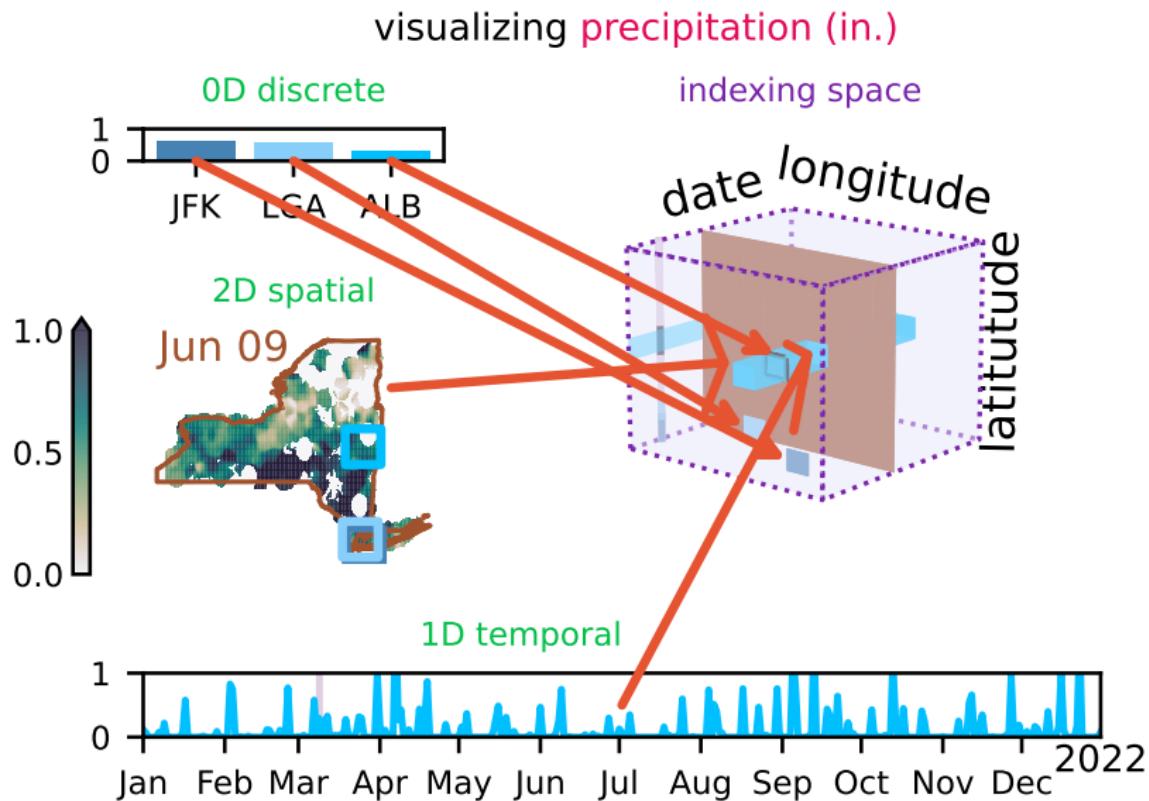
## Base space: topology

The base space of a fiber bundle is a quotient topology[37]. Given a set  $X$  and a function  $\mathcal{N}: X \rightarrow 2^{2^X}$  that assigns to any  $x \in X$  a non-empty collection of subsets  $\mathcal{N}(x)$ , where each element of  $\mathcal{N}(x)$  is a *neighborhood of  $x$* , then  $X$  with  $\mathcal{N}$  is a **topological space** and  $\mathcal{N}$  is a neighborhood *topology* if for each  $x$  in  $X$ : [38]

### Definition

1. if  $N$  is a neighborhood  $N \in \mathcal{N}(x)$  of  $x$  then  $x \in N$
2. every superset of a neighborhood of  $x$  is a neighborhood of  $x$ ; therefore a union of a neighborhood and adjacent points in  $X$  is also a neighborhood of  $x$
3. the intersection of any two neighborhoods of  $x$  is a neighborhood of  $x$
4. any neighborhood  $N$  of  $x$  contains a neighborhood  $M \subset N$  of  $x$  such that  $N$  is a neighborhood of each of the points in  $M$

# Why neighborhood topology?



## Base TYPE $\mathcal{K}$

An **category**  $\mathcal{C}$  consists of the following *data*:

1. a collection of *objects*  $X \in \mathbf{ob}(\mathcal{C})$
2. for every pair of objects  $X, Y \in \mathbf{ob}(\mathcal{C})$ , a set of *morphisms*  
 $X \xrightarrow{f} Y \in \text{Hom}_{\mathcal{C}}(X, Y)$
3. for every object  $X$ , a distinct *identity morphism*  $X \xrightarrow{id_X} X$  in  
 $\text{Hom}_{\mathcal{C}}(X, X)$
4. a *composition function*

$$f \in \text{Hom}_{\mathcal{C}}(X, Y) \times g \in \text{Hom}_{\mathcal{C}}(Y, Z) \rightarrow g \circ f \in \text{Hom}_{\mathcal{C}}(X, Z)$$

such that

1. *unitality*: for every morphism  $X \xrightarrow{f} Y$ ,  $f \circ id_X = f = id_Y \circ f$
2. *associativity*: if any three morphisms  $f, g, h$  are composable,

$$\begin{array}{ccccccc} X & \xrightarrow{f} & Y & \xrightarrow{g} & Z & \xrightarrow{h} & W \\ & \searrow & & & \nearrow & & \\ & & & & h \circ (g \circ f) = (h \circ g) \circ f & & \end{array}$$

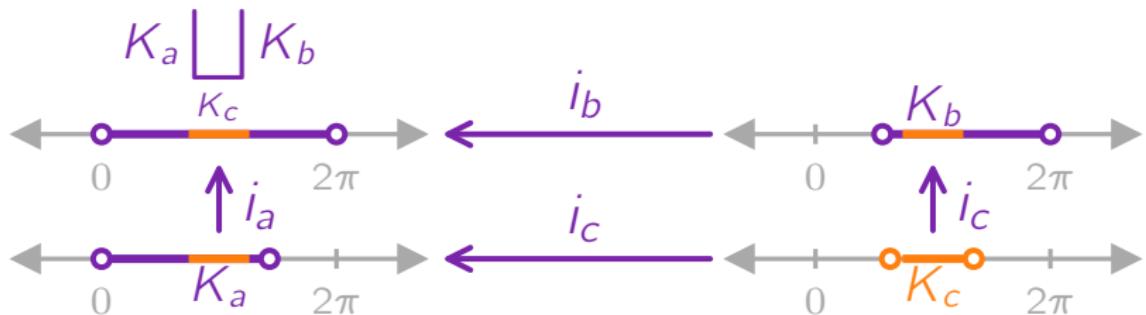
then they are associative such that  $h \circ (g \circ f) = (h \circ g) \circ f$  [39, 40, 41, 42].

## Base TYPE $\mathcal{K}$

The base space category consists of the openset object  $\mathcal{K}$  and morphisms between openset objects  $\hat{\phi} \in \text{Hom}(\mathcal{K}, \mathcal{K})$ .

The base space category  $\mathcal{K}$  is also equipped with functor  $\oplus : \mathcal{K} \sqcup \mathcal{K} \rightarrow \mathcal{K}$  for combining base spaces.

## Add records $\rightarrow$ Add to base space



# Fiber space: Fields

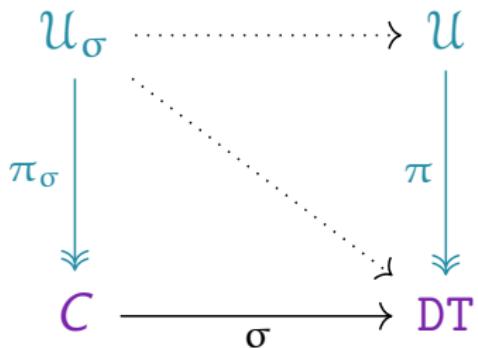
types and schema	record	table
$\mathcal{U}_\sigma \xrightarrow{\cdot} \mathcal{U}$ $\pi_\sigma \downarrow$ $C \xrightarrow{\sigma} DT$	$\mathcal{U}_\sigma$ $\pi_\sigma \downarrow$ $C$	$F := \Gamma^\pi(\sigma)$ $\pi \downarrow$ $K$

$\mathcal{U}$  Space of all possible values, e.g.  $\mathbb{R} \times \mathbb{R}^+$

$C$  Field names, e.g.  $TEMP, PRCP$

$DT$  Data Type, e.g.  $float, PosFloat$

# Fiber space: Fields

types and schema	record	table
$\mathcal{U}_\sigma \xrightarrow{\quad} \mathcal{U}$ 	$\mathcal{U}_\sigma$ $\pi_\sigma \downarrow$ $C$	$F := \Gamma^\pi(\sigma)$ $\pi \downarrow$ $K$

A *schema* consists of a pair  $(C, \sigma)$  where  $C$  is the set of field names and  $\sigma : C \rightarrow \mathbf{DT}$  is a function from field name to field data type[33]. The function  $\sigma$  is composed with  $\pi$  such that  $\pi^{-1}(\sigma(C)) \subseteq \mathcal{U}$ ; this composition induces a domain bundle  $\pi_\sigma : \mathcal{U}_\sigma \rightarrow C$  that associates a field name  $c \in C$  with its corresponding domain  $\pi_\sigma^{-1}(c) \subseteq \mathcal{U}_\sigma$ .

# Fiber space: Fields

types and schema	record	table
$\mathcal{U}_\sigma \xrightarrow{\quad} \mathcal{U}$ $\pi_\sigma \downarrow$ $C \xrightarrow{\sigma} \text{DT}$	$\mathcal{U}_\sigma$ $\pi_\sigma \downarrow$ $C$	$F := \Gamma^\pi(\sigma)$ $\pi \downarrow$ $K$

A **record** is a function  $r : C \rightarrow \mathcal{U}_\sigma$  and the set of records on  $\pi_\sigma$  is denoted  $\Gamma^\pi(\sigma)$ . Records must return an object of type  $\sigma(C) \in \text{DT}$  for each field  $c \in C$ .

# Fiber space: Fields

types and schema	record	table
$\mathcal{U}_\sigma \xrightarrow{\quad} \mathcal{U}$ $\pi_\sigma \downarrow$ $C \xrightarrow{\sigma} \mathbf{DT}$	$\mathcal{U}_\sigma$ $\pi_\sigma \downarrow$ $C$	$F := \Gamma^\pi(\sigma)$ $\pi \downarrow$ $K$

A **record** is a function  $r : C \rightarrow \mathcal{U}_\sigma$  and the set of records on  $\pi_\sigma$  is denoted  $\Gamma^\pi(\sigma)$ . Records must return an object of type  $\sigma(C) \in \mathbf{DT}$  for each field  $c \in C$ .

**Tables** are sections  $\tau : K \rightarrow \Gamma^\pi(\sigma)$  from an indexing space  $K$  to the set of all possible records  $\Gamma^\pi(\sigma)$  on the schema bundle

# Fiber space: Fields

types and schema	record	table
$\mathcal{U}_\sigma \xrightarrow{\quad} \mathcal{U}$ $\pi_\sigma \downarrow$ $C \xrightarrow{\sigma} DT$	$\mathcal{U}_\sigma$ $\pi_\sigma \downarrow$ $C$	$F := \Gamma^\pi(\sigma)$ $\pi \downarrow$ $K$

We define the **fiber space**  $F$  to be the space of all possible data records

$$F := \{r : C \rightarrow \mathcal{U}_\sigma \mid \pi_\sigma(r(C)) = C \text{ for all } C \in C\} \quad (3)$$

such that the preimage of a point is the corresponding data type domain  $\pi^{-1}(k) = F_k = \mathcal{U}_{\sigma_k}$ .

## Fiber category $\mathcal{F}$

The fiber category has a single object  $\mathcal{F}$  of an arbitrary type and morphisms on the fiber object  $\tilde{\phi} \in \text{Hom}(\mathcal{F}, \mathcal{F})$ .

The fiber category  $\mathcal{F}$  is also equipped with a bifunctor  $\otimes : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$  for combining fiber types.

## Fiber morphism $\tilde{\phi}$

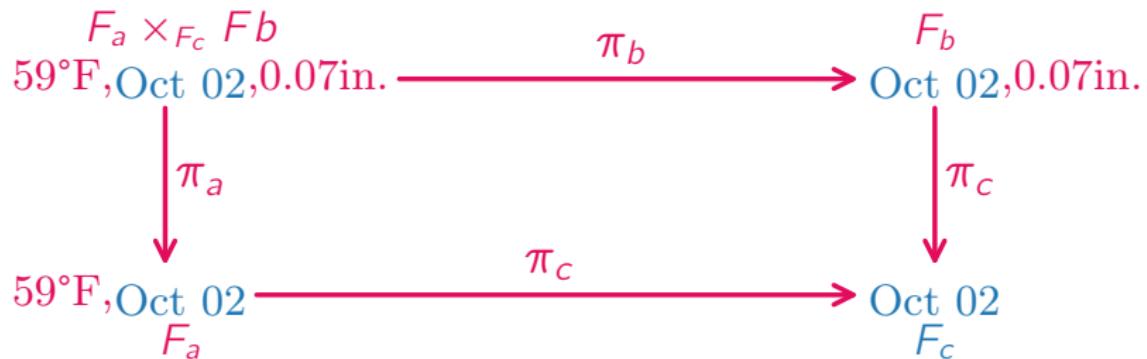
scale	operators	sample constraint
nominal	$=, \neq$	$\tau(\textcolor{violet}{k}_1) \neq \tau(\textcolor{violet}{k}_2) \implies \tilde{\phi}(\tau(\textcolor{violet}{k}_1)) \neq \tilde{\phi}(\tau(\textcolor{violet}{k}_2))$
ordinal	$<, \leqslant, \geqslant, >$	$\tau(\textcolor{violet}{k}_1) \leqslant \tau(\textcolor{violet}{k}_2) \implies \tilde{\phi}(\tau(\textcolor{violet}{k}_1)) \leqslant \tilde{\phi}(\tau(\textcolor{violet}{k}_2))$
interval	$+, -$	$\tilde{\phi}(\tau(\textcolor{violet}{k}) + C) = \tilde{\phi}(\tau(\textcolor{violet}{k})) + C$
ratio	$*, /$	$\tilde{\phi}(\tau(\textcolor{violet}{k}) * C) = \tilde{\phi}(\tau(\textcolor{violet}{k})) * C$

## Constraints as definition: Dates

- year:  $\{y \in \mathbb{I} | 1992 \leq y \leq 2025\}$
- month:  $\{m \in \mathbb{I} | 1 \leq m \leq 12\}$
- day:  $\{d \in \mathbb{I} | 1 \leq d \leq 31\}$
- date:  $\otimes : F_{year} \times F_{month} \times F_{day} \rightarrow F_{date}$

where the composition  $\otimes$  includes the constraint to only return dates that have the right number of days for each month.

Expand records → multiply fiber space



## local sections

We define section functions locally:

$$\Gamma(U, E|_U) \coloneqq \{\tau : U \rightarrow E|_U \mid \pi(\tau(k)) = k \text{ for all } k \in U\} \quad (4)$$

such that each section function  $\tau : k \mapsto r$  maps from each point  $k \in U$  to a corresponding record in the fiber space  $r \in F_k$  over that point.

This encoding can be translated into a set of signatures

{`data-subset` : `topology` → `fields` s.t. `data-subset` ⊂ `dataset`}.

## global sections

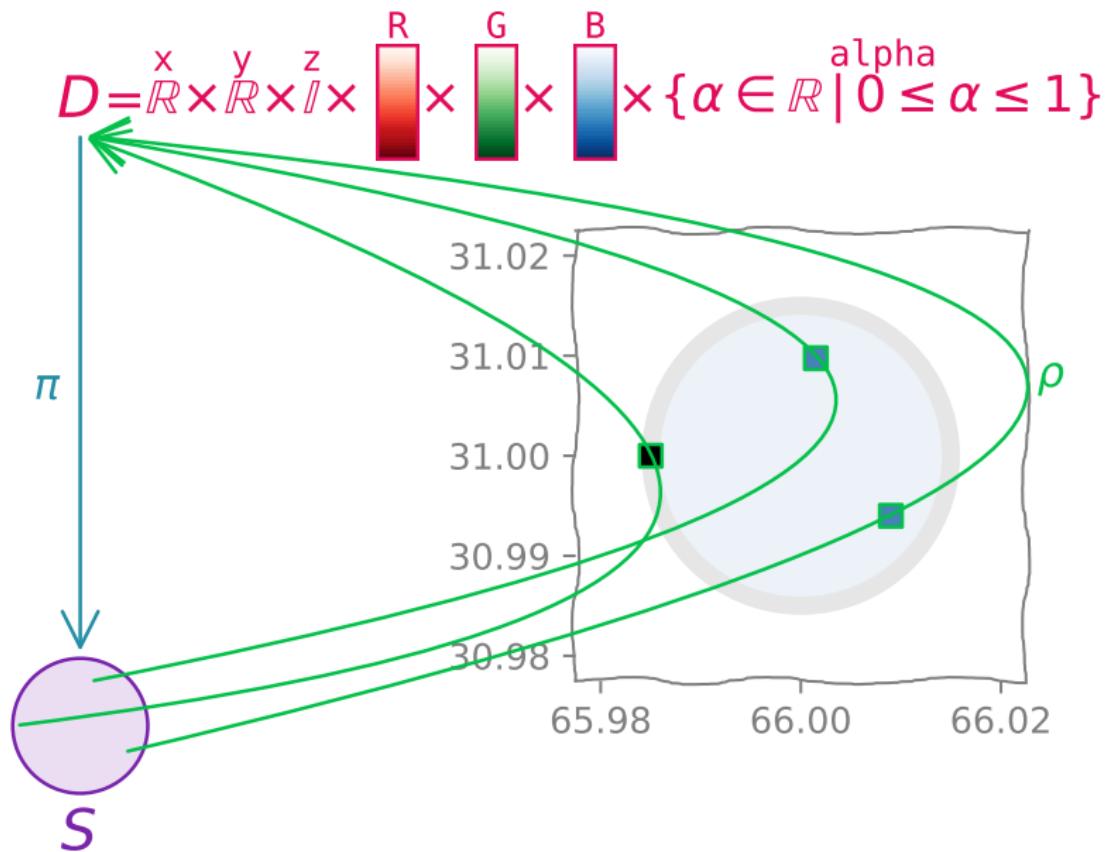
When a bundle is trivial  $E = K \times F$ , we can define a global sections  $\tau : K \rightarrow F \in \Gamma(K, F)$  which we translate into a data signature of the form `dataset : topology → field` where  $\tau = \text{dataset}$ ,  $K = \text{topology}$  and  $F = \text{fields}$

## Graphic bundle and section

$$D \hookrightarrow H \xrightarrow{\pi} S \quad (5)$$

$$\begin{aligned} \Gamma(W, H|_W) &\coloneqq \\ &\{ \rho : W \rightarrow H|_W \mid \pi(\rho(s)) = s \text{ for all } s \in W \} \end{aligned} \quad (6)$$

## Graphic section



## Definition

[43, 44] A **functor** is a map  $F : \mathcal{C} \rightarrow \mathcal{D}$ , which means it is a function between objects  $F : \mathbf{ob}(\mathcal{C}) \mapsto \mathbf{ob}(\mathcal{D})$  and that for every morphism  $f \in Hom(C_1, C_2)$  there is a corresponding function  $F : Hom(C_1, C_2) \mapsto Hom(F(C_1), F(C_2))$ . A **functor** must satisfy the properties

- *identity*:  $F(id_{\mathcal{C}}(C)) = id_{\mathcal{D}}(F(C))$
- *composition*:  $F(g) \circ F(f) = F(g \circ f)$  for any composable morphisms  $C_1 \xrightarrow{f} C_2, C_2 \xrightarrow{g} C_3$

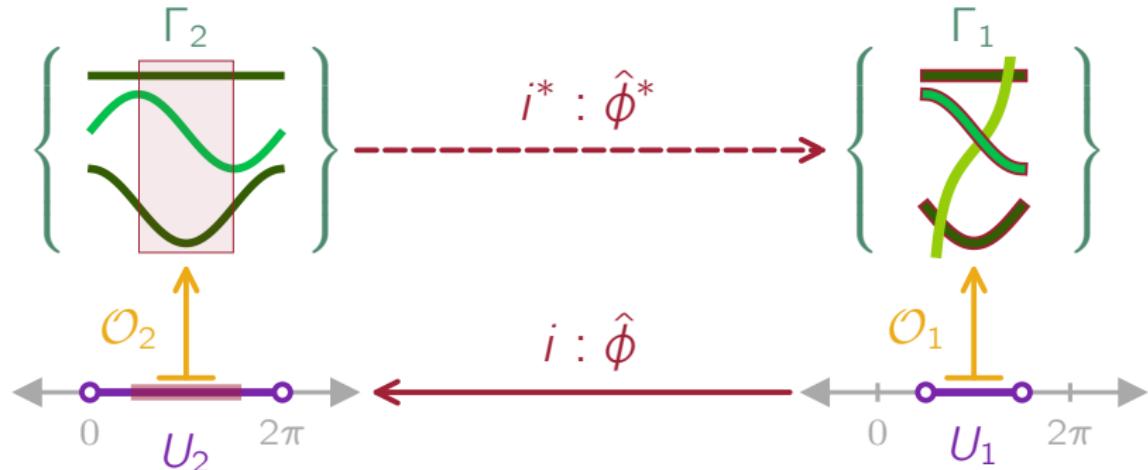
$F(C) \in \mathbf{ob}(\mathcal{D})$  denotes the object to which an object  $C$  is mapped, and  $F(f) \in Hom(F(C_1), F(C_2))$  denotes the morphism that  $f$  is mapped to.

## topological equivariance: presheaf

### Definition

A **presheaf**  $F : \mathcal{C}^{op} \rightarrow \mathbf{Set}$  is a contravariant functor from an object in an arbitrary category to an object in the category  $\mathbf{Set}$ [36].

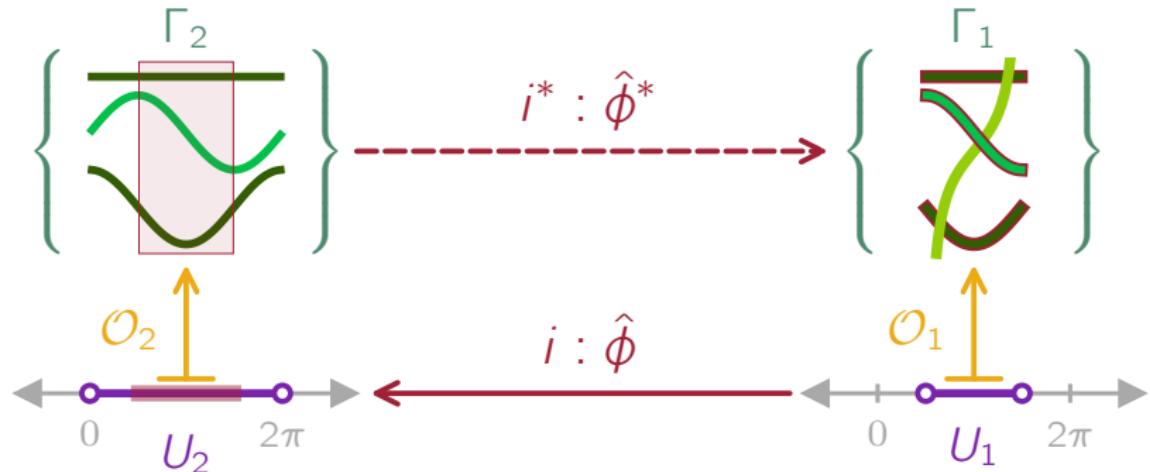
# topological equivariance: presheaf



## Definition

A **presheaf**  $F : \mathcal{C}^{op} \rightarrow \mathbf{Set}$  is a contravariant functor from an object in an arbitrary category to an object in the category  $\mathbf{Set}$ [36].

## topological equivariance: presheaf



The presheaf is contravariant when for every arbitrary morphism between input base spaces  $\hat{\phi} : U_1 \rightarrow U_2$  there exists a corresponding pullback function between the sets of sections

$$\hat{\phi}^* : \Gamma(U_2, E|_{U_2}) \rightarrow \Gamma(U_1, E|_{U_1}).$$

# Container invariance: sheaf

## Definition

[45, 36] A **sheaf** is a presheaf that satisfies the following two axioms

- *locality* two sections in a sheaf are equal  $\tau^a = \tau^b$  when they evaluate to the same values  $\tau^a|_{U_i} = \tau^b|_{U_i}$  over the open cover  $\bigcup_{i \in I} U_i \subset U$  (indexed by  $I$ ).
- *gluing* the union of sections defined over subspaces  $\tau^i \in \Gamma(U_i, E|_{U_i})$  is equivalent to a section defined over the whole space  $\tau|_U = \tau^i$  for all  $i \in I$  if all pairs of sections agree on overlaps  $\tau^i|_{U_i \cap U_j} = \tau^j|_{U_i \cap U_j}$

## Container invariance: sheaf

A **sheaf** is a presheaf that satisfies the following two axioms [45, 36]

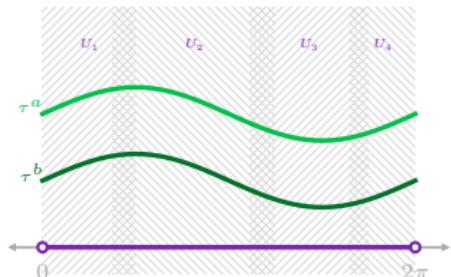
- *locality* two sections in a sheaf are equal  $\tau^a = \tau^b$  when they evaluate to the same values  $\tau^a|_{U_i} = \tau^b|_{U_i}$  over the open cover  $\bigcup_{i \in I} U_i \subset U$  (indexed by  $I$ ).
- *gluing* the union of sections defined over subspaces  $\tau^i \in \Gamma(U_i, E|_{U_i})$  is equivalent to a section defined over the whole space  $\tau|_U = \tau^i$  for all  $i \in I$  if all pairs of sections agree on overlaps  $\tau^i|_{U_i \cap U_j} = \tau^j|_{U_i \cap U_j}$

# Container invariance: sheaf

A **sheaf** is a presheaf that satisfies the following two axioms [45, 36]

- *locality* two sections in a sheaf are equal  $\tau^a = \tau^b$  when they evaluate to the same values  $\tau^a|_{U_i} = \tau^b|_{U_i}$  over the open cover  $\bigcup_{i \in I} U_i \subset U$  (indexed by  $I$ ).

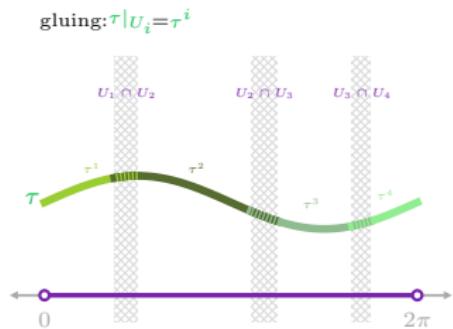
locality:  $\tau^a = \tau^b$



# Container invariance: sheaf

A **sheaf** is a presheaf that satisfies the following two axioms [45, 36]

- *gluing* the union of sections defined over subspaces  $\tau^i \in \Gamma(U_i, E|_{U_i})$  is equivalent to a section defined over the whole space  $\tau|_{U_i} = \tau^i$  for all  $i \in I$  if all pairs of sections agree on overlaps  $\tau^i|_{U_i \cap U_j} = \tau^j|_{U_i \cap U_j}$

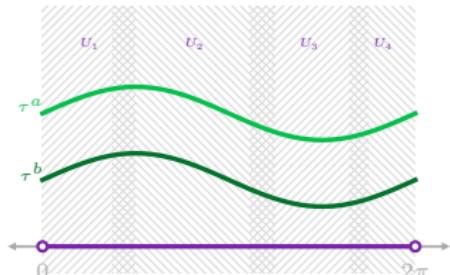


# Container invariance: sheaf

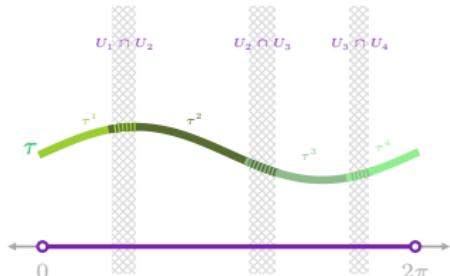
A **sheaf** is a presheaf that satisfies the following two axioms [45, 36]

- *locality* two sections in a sheaf are equal  $\tau^a = \tau^b$  when they evaluate to the same values  $\tau^a|_{U_i} = \tau^b|_{U_i}$  over the open cover  $\bigcup_{i \in I} U_i \subset U$  (indexed by  $I$ ).
- *gluing* the union of sections defined over subspaces  $\tau^i \in \Gamma(U_i, E|_{U_i})$  is equivalent to a section defined over the whole space  $\tau|_{U_i} = \tau^i$  for all  $i \in I$  if all pairs of sections agree on overlaps  $\tau^i|_{U_i \cap U_j} = \tau^j|_{U_i \cap U_j}$

locality:  $\tau^a = \tau^b$



gluing:  $\tau|_{U_i} = \tau^i$



## What about needing more than one record?

The sheaf over a very small region surrounding a point  $k$  is called a *stalk*[46]

$$\mathcal{O}_{K,E}|_k := \lim_{U \ni k} \Gamma(U, E|_U) \quad (7)$$

where the fiber is contained inside the stalk  $F_k \subset \mathcal{O}_{K,E}|_k$ . The *germ* is the section evaluated at a point in the stalk  $\tau(k) \in \mathcal{O}_{K,E}|_k$  and is the data.

# Homeomorphism

We define the mapping  $\xi$  to be a surjective continuous map:

$$\xi : W \rightarrow U \quad (8)$$

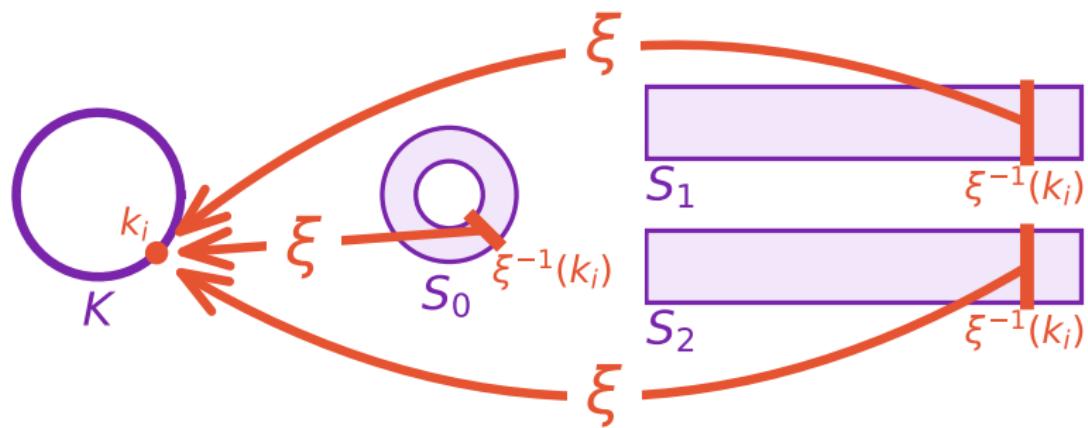
between a graphic subspace  $W \subseteq S$  and data subspace  $U \subseteq K$ .

The set of points in graphic space that correspond to each point in data space is

$$\xi^{-1}(k) = \{s | \xi(s) = k \forall k \in K, s \in S\} \quad (9)$$

such that every point in a graphic space has a corresponding point in data space.

# Homeomorphism



# Pushforward: Graphic for Data

## Definition

Given a sheaf  $\mathcal{O}_{S,H}$  on  $S$ , the **pushforward** sheaf  $\xi_* \mathcal{O}_{S,H}$  on  $K$  is defined as

$$\xi_*(\mathcal{O}_{S,H})(U) = \mathcal{O}_{S,H}(\xi^{-1}(U)) \quad (10)$$

for all opensets  $U \subset K$  [46].

This provides a way to look up which graphic corresponds with a data index

$$\xi_* \rho(k) = \rho|_{\xi^{-1}(k)} \quad (11)$$

such that  $\xi_* \rho(k)(s) = \rho(s)$  for all  $s \in \xi^{-1}(k)$ .

# Pullback: Data at graphic

## Definition

[46] Given a sheaf  $\mathcal{O}_{K,E}$  on  $K$ , the **pullback** sheaf  $\xi^*\mathcal{O}_{K,E}$  on  $S$  is defined as the sheaf associated to the presheaf

$$\xi^*(\mathcal{O}_{K,E})(W) = \mathcal{O}_{K,E}(\xi(W)) \text{ for } \xi(W) \in K.$$

This provides a way to then look up what data values correspond with a graphic index

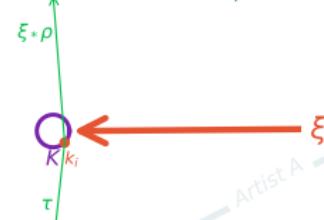
$$\xi^*\tau(s) = \tau(\xi(s)) = \tau(k) \tag{12}$$

As  $\xi$  is surjective, there are many points  $s \in W \subseteq S$  in the graphic space that correspond to a single point  $\xi(s) = k$ .

## Homeomorphism

graphic specification for data at  $k_i$

```
<circle cx=0.71 cy=-0.71 fill="#bcbdb2" stroke="#bcbdb2" r=7/>
  <path d="L 5.50 0.71 stroke="#ff7f0e" fill="transparent"/>
  <path d="L 5.50 -0.71 stroke="#f177b4" fill="transparent"/>
```

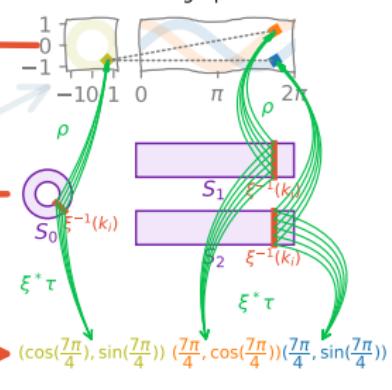


i	cos	sin
$\frac{7\pi}{4}$	$\cos(\frac{7\pi}{4})$	$\sin(\frac{7\pi}{4})$

data at point  $k_i$

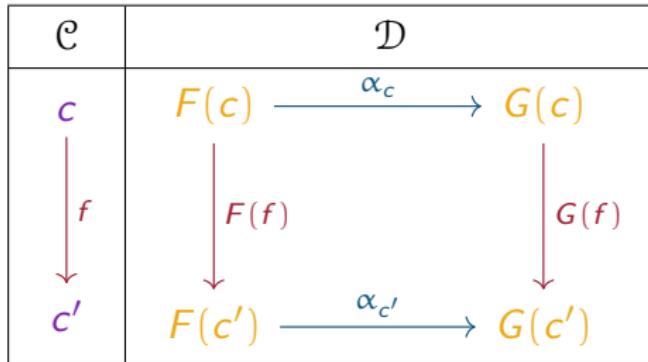
$$-\xi^* -$$

data at graphic region  $\xi^{-1}(k_i)$



## Definition (natural transform)

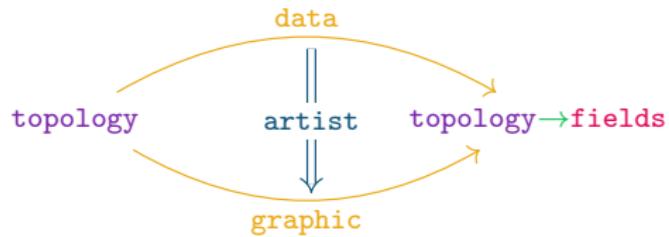
Given two functors  $F, G : \mathcal{C} \rightarrow \mathcal{D}$ , a **natural transformation**  $\alpha : F \rightarrow G$  is a function which assigns to each object  $c$  of  $\mathcal{C}$  a morphism  $\alpha_c : F(c) \rightarrow G(c)$ ,  $G(c) \in \mathcal{D}$ , in such a way that for every morphism  $f : c \rightarrow c'$ ,  $c' \in \mathcal{C}$ , the morphisms in  $\mathcal{D}$  commute such that  $\alpha'_{c'}(F(f)(F(c))) = G(f)(\alpha_c(F(c)))$ . When this holds,  $\alpha_c$  is *natural* in  $c$ .[41].



## Definition (natural transform)

Given two functors  $F, G : \mathcal{C} \rightarrow \mathcal{D}$ , a **natural transformation**  $\alpha : F \rightarrow G$  is a function which assigns to each object  $c$  of  $\mathcal{C}$  a morphism  $\alpha_c : F(c) \rightarrow G(c)$ ,  $G(c) \in \mathcal{D}$ , in such a way that for every morphism  $f : c \rightarrow c'$ ,  $c' \in \mathcal{C}$ , the morphisms in  $\mathcal{D}$  commute such that  $\alpha'_{c'}(F(f)(F(c))) = G(f)(\alpha_c(F(c)))$ . When this holds,  $\alpha_c$  is *natural* in  $c$ .[41].

# Artist



How?

$$\begin{array}{ccc} \Gamma(U, \xi_* H|_U) & \xleftarrow{\xi_*} & \Gamma(W, ,)H|_W \\ Hom_{\mathcal{O}_K} \uparrow & Hom_{\mathcal{O}_K, \mathcal{O}_S} \nearrow & \uparrow Hom_{\mathcal{O}_S} \\ \Gamma(U, E|_U) & \xrightarrow{\xi^*} & \Gamma(W, \xi^* E|_W) \end{array} \quad (13)$$

# Artist

$$A : \Gamma(K, E) \rightarrow Im_A(S, H), Im_A(S, H) \subset \Gamma(S, H) \quad (14)$$

where:

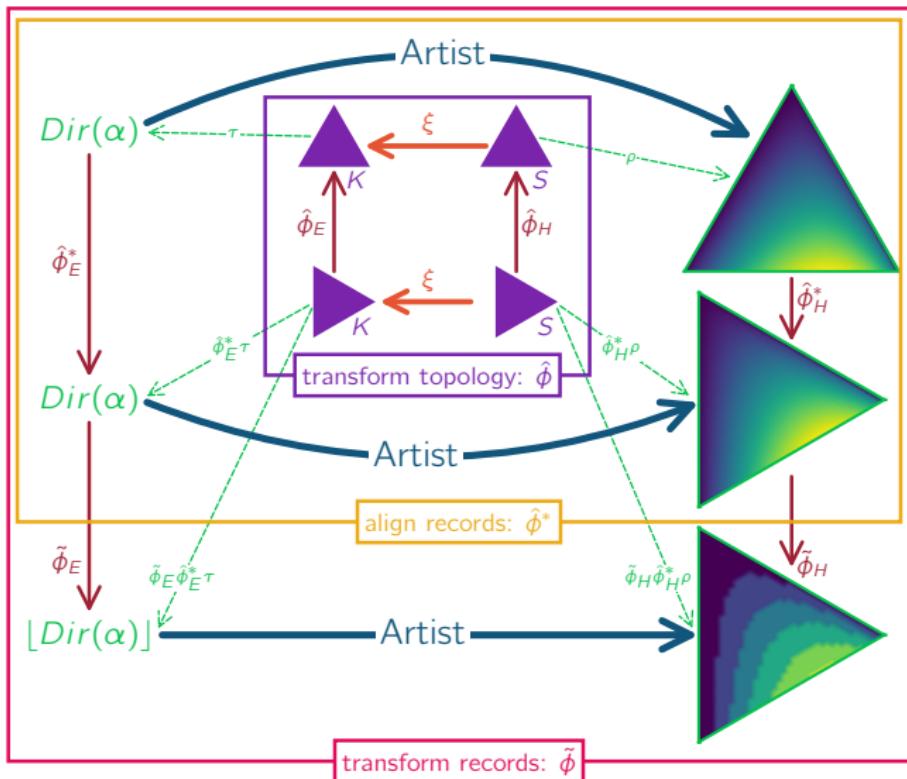
$$Im_A(S, H) \coloneqq \{\rho \mid \exists \tau \in \Gamma(K, E) \text{ s.t. } A(\tau) = \rho, \xi(S) = K\} \quad (15)$$

# Equivariant Artist

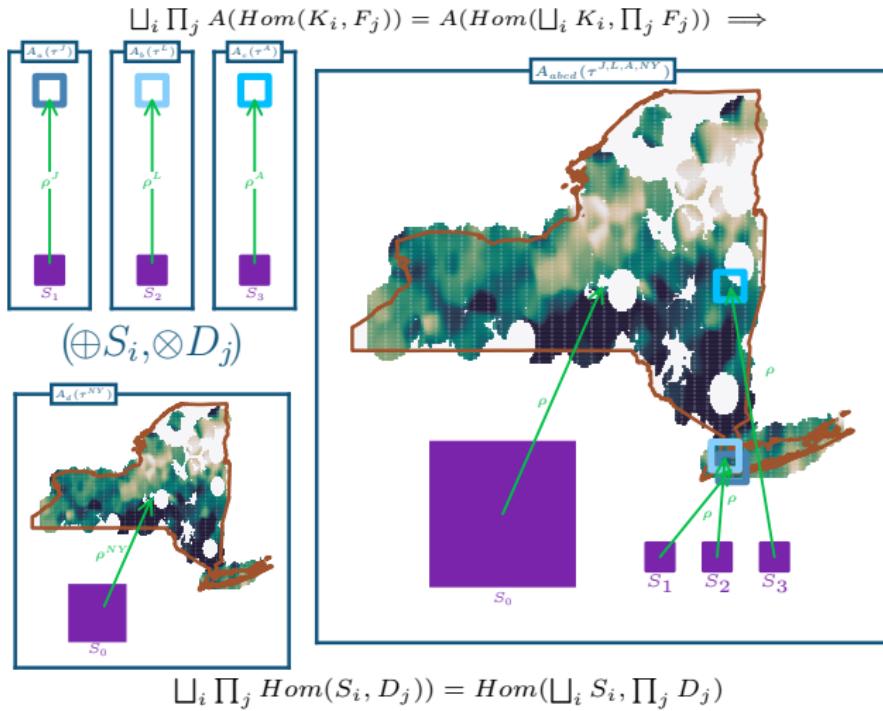
$$\begin{array}{ccc} \Gamma(K, E) & \xrightarrow{A} & \text{Im}_A(S, H) \\ \hat{\phi}^*_E \downarrow & & \downarrow \hat{\phi}^*_H \\ \Gamma(K', \hat{\phi}^*_E E) & K \xleftarrow{\xi} S & \text{Im}_A(S', \hat{\phi}^*_H H) \\ \tilde{\phi}_E \downarrow & \hat{\phi}_E \uparrow & \downarrow \tilde{\phi}_H \\ K' & \xleftarrow{\xi} S' & \\ \Gamma(K', E') & \xrightarrow{A} & \text{Im}_A(S', H') \end{array} \quad (16)$$

# Equivariant Artist

## Visualizing the Dirichlet Distribution



# artist composition

$$\bigsqcup_i \prod_j A(Hom(K_i, F_j)) = A(Hom(\bigsqcup_i K_i, \prod_j F_j)) \implies$$

$$(\oplus S_i, \otimes D_j)$$
$$\bigsqcup_i \prod_j Hom(S_i, D_j) = Hom(\bigsqcup_i S_i, \prod_j D_j)$$

# Animation and Interactivity

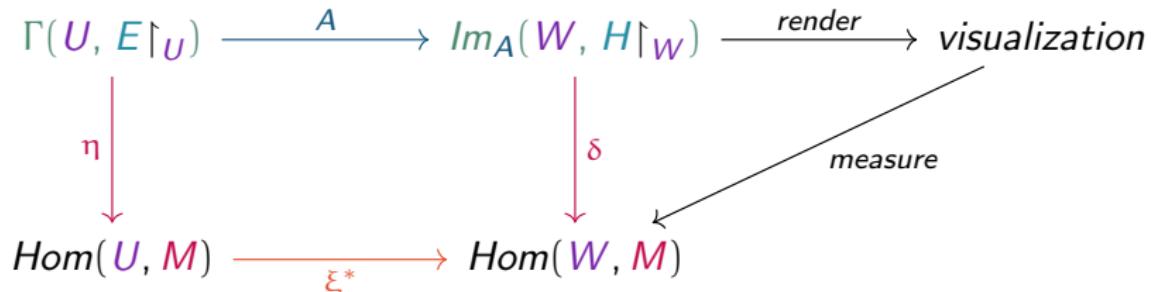
pan, zoom, scroll sheaf: locality + gluing 9

selection and hover pushforward 11, pullback 12

brushing, linking, annotation composition of artists ??

# Testing if $A$ is equivariant

$M$  is a (scalar, vector) measurable component (e.g. color, position, shape, texture, rotation, ) of the rendered visual element.



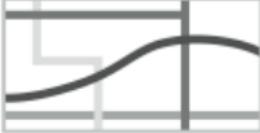
$$\text{input } \eta : \tau \mapsto (U \xrightarrow{\eta_\tau} M)$$

$$\text{output } \delta : \rho \mapsto (W \xrightarrow{\delta_\rho} M)$$

$$\eta_\tau(k) = \delta_\rho(s) \text{ for all } \xi(s) = k, k \in K, s \in S$$

# Visual Space: Specialized data space

$$P \hookrightarrow V \xrightarrow{\pi} K$$

	Points	Lines	Areas	Best to show
Shape		<i>possible, but too weird to show</i>	<i>cartogram</i>	<i>qualitative differences</i>
Size			<i>cartogram</i>	<i>quantitative differences</i>
Color Hue				<i>qualitative differences</i>
Color Value				<i>quantitative differences</i>
Color Intensity				<i>qualitative differences</i>

# Encoding

$$\nu : F_k \rightarrow P_k \quad (17)$$

$$F_k \xrightarrow{\nu} P_k := F'_k \xrightarrow{\nu'} P'_k \quad (18)$$

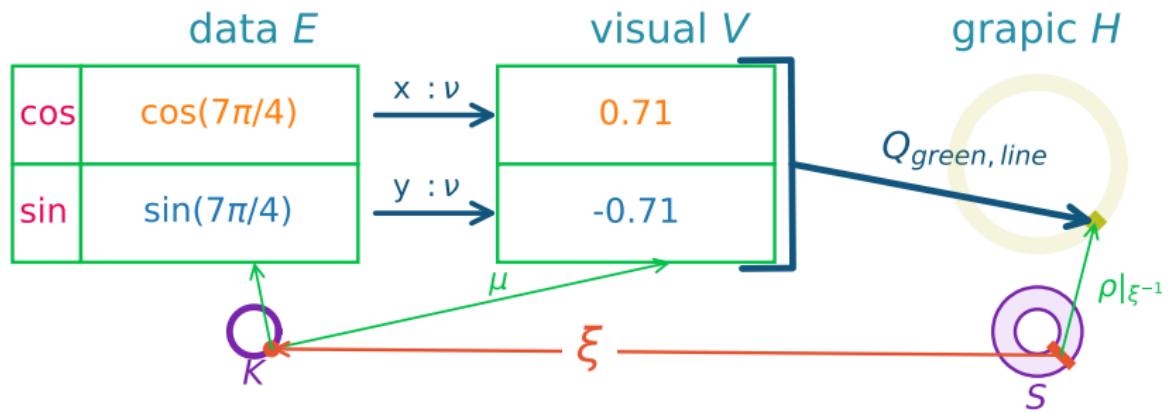
```
graph LR; Fk[F_k] -- "ν" --> Pk[P_k]; Fk -- "ν''" --> Pkp[P'_k]; Fkp[F'_k] -- "ν'" --> Pkp;
```

# Composition

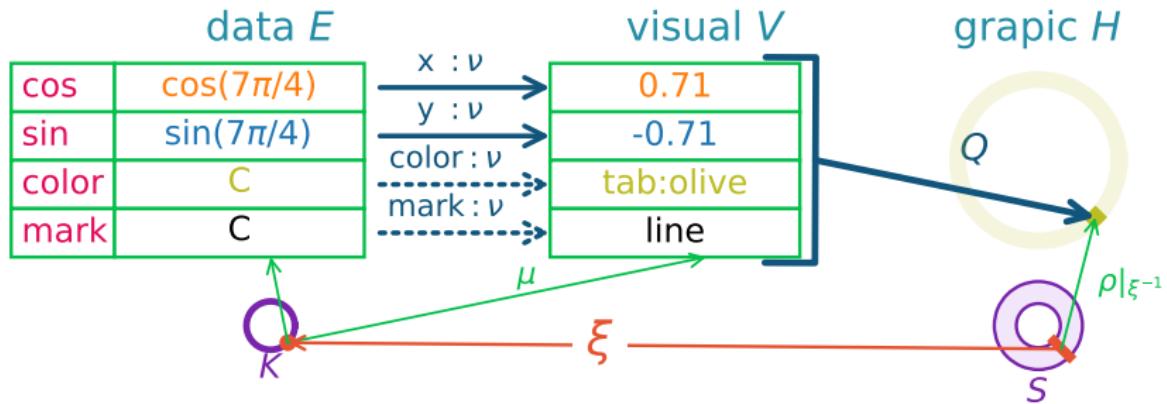
$$Q : \Gamma(K, V) \rightarrow \Gamma(S, H) \quad (19)$$

$$\begin{array}{ccccc} \Gamma(K, V) & \xrightarrow{\gamma} & \Gamma(K, V') & \xrightarrow{Q} & \Gamma(S, H) \\ & \searrow & \downarrow Q' & \nearrow & \\ & & & & \end{array} \quad (20)$$

# Construction Stages



# Construction Stages: composition



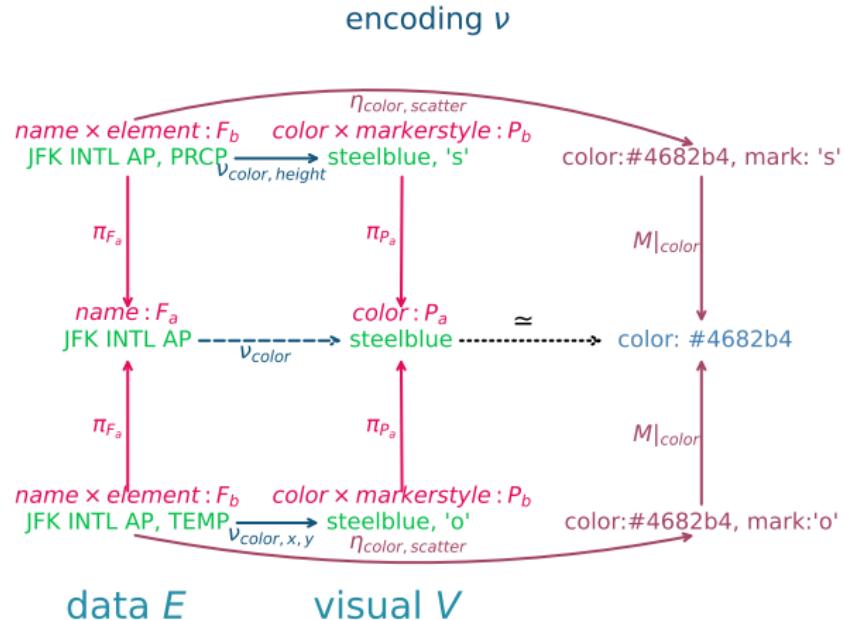
# Verify encoder

$$\begin{array}{ccccc}
 & & \eta_{ab} & & \\
 & F_k^a \times F_k^b & \xrightarrow{\gamma_{ab}} & P_k^a \times P_k^b & M_k^{ab} \\
 \pi_a \downarrow & & & \downarrow \pi_a & \downarrow M \upharpoonright_a \\
 F_k^a & \dashrightarrow^{\gamma_a} & P_k^a & \simeq & M_k^a \\
 \uparrow \pi_a & & \uparrow \pi_a & & \uparrow M \upharpoonright_a \\
 F_k^a \times F_k^c & \xrightarrow{\gamma_{ac}} & P_k^a \times P_k^c & & M_k^{ac}
 \end{array} \tag{21}$$

The diagram illustrates the verification of an encoder across three stages:  $F_k^a \times F_k^b$ ,  $P_k^a \times P_k^b$ , and  $M_k^{ab}$ . It also includes a second row for  $F_k^a \times F_k^c$  and  $P_k^a \times P_k^c$ , which is connected to the first row via  $\gamma_{ac}$ .

- Top Row:**  $F_k^a \times F_k^b \xrightarrow{\gamma_{ab}} P_k^a \times P_k^b \xrightarrow{\eta_{ab}} M_k^{ab}$ . The vertical projection  $\pi_a$  maps  $F_k^a$  to  $F_k^b$ , and the vertical projection  $M \upharpoonright_a$  maps  $M_k^{ab}$  to  $M_k^a$ .
- Bottom Row:**  $F_k^a \times F_k^c \xrightarrow{\gamma_{ac}} P_k^a \times P_k^c \xrightarrow{\eta_{ac}} M_k^{ac}$ . The vertical projection  $\pi_a$  maps  $F_k^a$  to  $F_k^c$ , and the vertical projection  $M \upharpoonright_a$  maps  $M_k^{ac}$  to  $M_k^a$ .
- Connections:**
  - A dashed arrow  $\dashrightarrow^{\gamma_a}$  connects  $F_k^a$  to  $P_k^a$ .
  - A dashed arrow  $\simeq$  connects  $P_k^a$  to  $M_k^a$ .
  - A curved red arrow  $\eta_{ac}$  connects  $M_k^{ab}$  to  $M_k^{ac}$ .

# Verify encoder

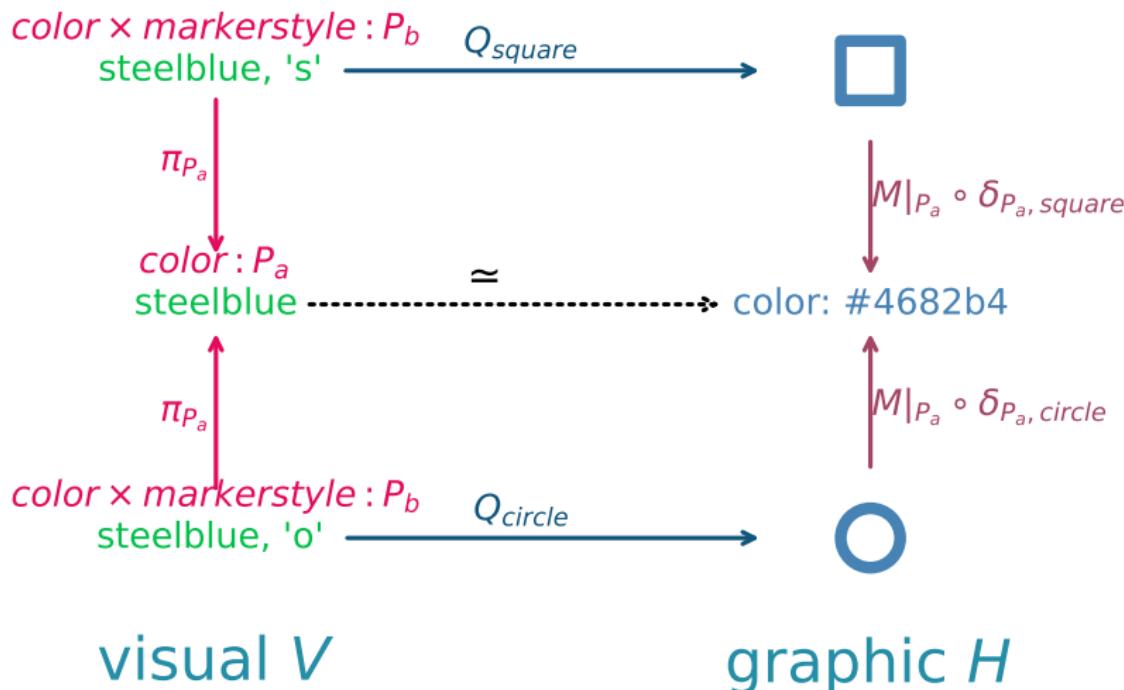


# Verify compositor

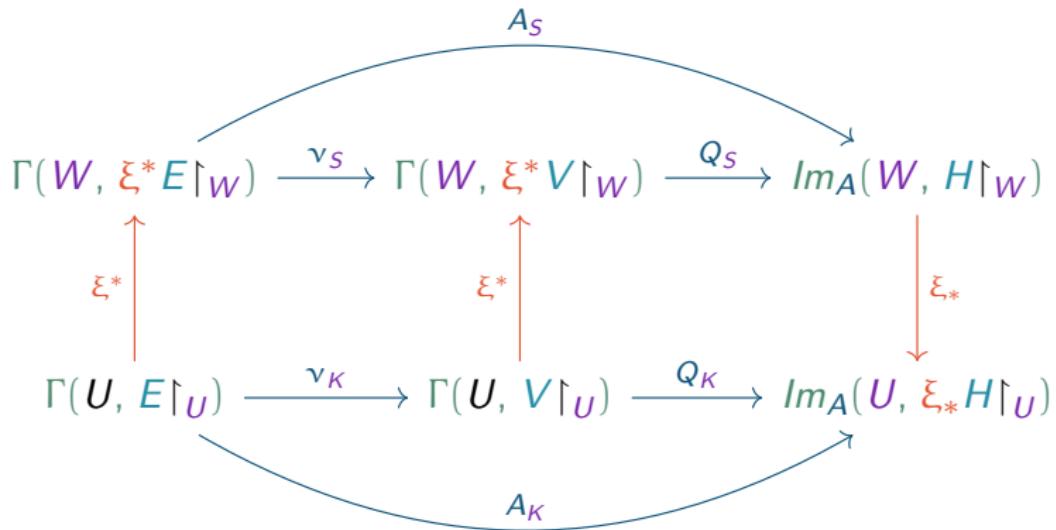
$$\begin{array}{ccc} \Gamma(K, V^a \times V^b) & \xrightarrow{Q_{ab}} & \text{Im}_{A_{ab}}(S, H) \\ \pi_a \downarrow & & \downarrow M|_a \circ \delta_{ab} \\ \Gamma(K, V^a) & \xrightarrow{\simeq} & \text{Hom}(K, M^a) \\ \pi_a \uparrow & & \uparrow M|_a \circ \delta_{ac} \\ \Gamma(K, V^a \times V^c) & \xrightarrow{Q_{ac}} & \text{Im}_{A_{ac}}(S, H) \end{array} \quad (22)$$

# Verify compositor

composition  $Q$



# Implementation Choices: $A_K = A_S$



## References I

- [1] L. Wilkinson. *The Grammar of Graphics*. 2nd ed. Statistics and Computing. New York: Springer-Verlag New York, Inc., 2005.
- [2] R. P. Grimaldi. *Discrete and Combinatorial Mathematics*, 5/e. Pearson Education, 2006.
- [3] A. M. Pitts. *Nominal Sets: Names and Symmetry in Computer Science*. USA: Cambridge University Press, 2013.
- [4] J. Bertin. *Semiology of Graphics : Diagrams, Networks, Maps*. Redlands, Calif.: ESRI Press, 2011.
- [5] J. Mackinlay. “Automatic Design of Graphical Presentations”. Stanford, 1987.
- [6] G. Kindlmann and C. Scheidegger. “An Algebraic Process for Visualization Design”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 2181–2190.

## References II

- [7] H. Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [8] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. "NIH Image to ImageJ: 25 Years of Image Analysis". In: *Nature Methods* 9.7 (July 2012), pp. 671–675.
- [9] M. Bastian, S. Heymann, and M. Jacomy. "Gephi: An Open Source Software for Exploring and Manipulating Networks". In: International AAAI Conference on Weblogs and Social Media. Vol. 3. Proceedings of the International AAAI Conference on Web and Social Media. 2009.
- [10] A. Satyanarayan, K. Wongsuphasawat, and J. Heer. "Declarative Interaction Design for Data Visualization". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. Honolulu Hawaii USA: ACM, Oct. 2014, pp. 669–678.

## References III

- [11] S. Studies. *Culturevis/Imageplot*.
- [12] J. Ellson et al. “Graphviz— Open Source Graph Drawing Tools”. In: *Graph Drawing*. Ed. by P. Mutzel, M. Jünger, and S. Leipert. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 483–484.
- [13] J. VanderPlas et al. “Altair: Interactive Statistical Visualizations for Python”. In: *Journal of Open Source Software* 3.32 (Dec. 2018), p. 1057.
- [14] N. Sofroniew et al. *Napari/Napari: 0.4.5rc1*. Version v0.4.5rc1. Zenodo, Feb. 2021.

## References IV

- [15] A. A. Hagberg, D. A. Schult, and P. J. Swart. “Exploring Network Structure, Dynamics, and Function Using NetworkX”. In: *Proceedings of the 7th Python in Science Conference*. Ed. by G. Varoquaux, T. Vaught, and J. Millman. Pasadena, CA USA, 2008, pp. 11–15.
- [16] C. Stolte, D. Tang, and P. Hanrahan. “Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases”. In: *IEEE Transactions on Visualization and Computer Graphics* 8.1 (Jan. 2002), pp. 52–65.
- [17] P. Hanrahan. “VizQL: A Language for Query, Analysis and Visualization”. In: *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data*. SIGMOD '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 721.

## References V

- [18] J. Mackinlay, P. Hanrahan, and C. Stolte. "Show Me: Automatic Presentation for Visual Analysis". In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (Nov. 2007), pp. 1137–1144.
- [19] J. Heer and M. Agrawala. "Software Design Patterns for Information Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 853–860.
- [20] J. D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science Engineering* 9.3 (May 2007), pp. 90–95.
- [21] M. L. Waskom. "Seaborn: Statistical Data Visualization". In: *Journal of Open Source Software* 6.60 (2021), p. 3021.
- [22] S. Hoyer and J. Hamman. "Xarray: ND Labeled Arrays and Datasets in Python". In: *Journal of Open Research Software* 5.1 (2017).

## References VI

- [23] M. Bostock, V. Ogievetsky, and J. Heer. “D<sup>3</sup> Data-Driven Documents”. In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2301–2309.
- [24] M. D. Hanwell et al. “The Visualization Toolkit (VTK): Rewriting the Rendering Code for Modern Graphics Cards”. In: *SoftwareX* 1–2 (Sept. 2015), pp. 9–12.
- [25] B. Geveci et al. “VTK”. In: *The Architecture of Open Source Applications* 1 (2012), pp. 387–402.
- [26] P. Ramachandran and G. Varoquaux. “Mayavi: 3D Visualization of Scientific Data”. In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 40–51.
- [27] Brian Wylie and Jeffrey Baumes. “A Unified Toolkit for Information and Scientific Visualization”. In: *Proc. SPIE*. Vol. 7243. Jan. 2009.

## References VII

- [28] J. Ahrens, B. Geveci, and C. Law. "Paraview: An End-User Tool for Large Data Visualization". In: *The visualization handbook* 717.8 (2005).
- [29] K. Wongsuphasawat. "Navigating the Wide World of Data Visualization Libraries (on the Web)". Outlier Conf. 2021.
- [30] M. Tory and T. Moller. "Rethinking Visualization: A High-Level Taxonomy". In: *IEEE Symposium on Information Visualization*. 2004, pp. 151–158.
- [31] D. M. Butler and S. Bryson. "Vector-Bundle Classes Form Powerful Tool for Scientific Visualization". In: *Computers in Physics* 6.6 (1992), p. 576.
- [32] D. M. Butler and M. H. Pendley. "A Visualization Model Based on the Mathematics of Fiber Bundles". In: *Computers in Physics* 3.5 (1989), p. 45.

## References VIII

- [33] D. I. Spivak. *Simplicial Databases*. 2009. URL: <https://arxiv.org/abs/0904.2012>. Pre-published.
- [34] V. Wiels and S. Easterbrook. "Management of Evolving Specifications Using Category Theory". In: *Proceedings 13th IEEE International Conference on Automated Software Engineering (Cat. No.98EX239)*. 1998, pp. 12–21.
- [35] R. W. Ghrist. *Elementary Applied Topology*. Vol. 1. Createspace Seattle, 2014.
- [36] E. Spanier. *Algebraic Topology*. McGraw-Hill Series in Higher Mathematics. Springer, 1989.
- [37] J. R. Munkres. *Elements of Algebraic Topology*. Menlo Park, Calif: Addison-Wesley, 1984.
- [38] R. Brown. *Topology and Groupoids*. 2006.

## References IX

- [39] F. W. Lawvere and S. H. Schanuel. *Conceptual Mathematics: A First Introduction to Categories*. Cambridge University Press, 2009.
- [40] E. Riehl. *Category Theory in Context*. Dover Publications, Nov. 16, 2016.
- [41] S. Mac Lane. *Categories for the Working Mathematician*. Vol. 5. Springer Science & Business Media, 2013.
- [42] B. Fong and D. I. Spivak. *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. 1st ed. Cambridge University Press, July 2019.
- [43] T.-D. Bradley. *What Is a Functor? Definitions and Examples, Part 2*. Math3ma. URL: <https://www.math3ma.com/blog/what-is-a-functor-part-2> (visited on 10/07/2021).

## References X

- [44] T.-D. Bradley, J. Terilla, and T. Bryson. *Topology : A Categorical Approach*. MIT Press, 2020.
- [45] M. J. Baker. *EuclideanSpace: Maths - Sheaf*. URL: <https://www.euclideanspace.com/mathematics/topology/sheaf/> (visited on 01/06/2022).
- [46] G. Harder and K. Diederich. *Lectures on Algebraic Geometry I: Sheaves, Cohomology of Sheaves, and Applications to Riemann Surfaces*. Aspects of Mathematics. Vieweg+Teubner Verlag, 2008.