

1

TOPOLOGICAL ARTIST MODEL

2

HANNAH AIZENMAN

3

A DISSERTATION PROPOSAL SUBMITTED TO

4

THE GRADUATE FACULTY IN COMPUTER SCIENCE IN PARTIAL FULFILLMENT OF THE

5

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY,

6

THE CITY UNIVERSITY OF NEW YORK

7

COMMITTEE MEMBERS:

8

DR. MICHAEL GROSSBERG (ADVISOR), DR. ROBERT HARALICK, DR. LEV MANOVICH,

9

DR. HUY VO, DR. MARCUS HANWELL

10

JUNE 2021

Abstract

This work presents a functional model of the structure-preserving maps from data to visual representation to guide the development of visualization libraries. Our model, which we call the topological equivariant artist model (TEAM), provides a means to express the constraints of preserving the data continuity in the graphic and faithfully translating the properties of the data variables into visual variables. We formalize these transformations as actions on sections of topological fiber bundles, which are mathematical structures that allow us to encode continuity as a base space, variable properties as a fiber space, and data as binding maps, called sections, between the base and fiber spaces. This abstraction allows us to generalize to any type of data structure, rather than assuming, for example, that the data is a relational table, image, data cube, or network-graph. Moreover, we extend the fiber bundle abstraction to the graphic objects that the data is mapped to. By doing so, we can track the preservation of data continuity in terms of continuous maps from the base space of the data bundle to the base space of the graphic bundle. Equivariant maps on the fiber spaces preserve the structure of the variables; this structure can be represented in terms of monoid actions, which are a generalization of the mathematical structure of Stevens' theory of measurement scales. We briefly sketch that these transformations have an algebraic structure which lets us build complex components for visualization from simple ones. We demonstrate the utility of this model through case studies of a scatter plot, line plot, and image. To demonstrate the feasibility of the model, we implement a prototype of a scatter and line plot in the context of the Matplotlib Python visualization library. We propose that the functional architecture derived from a TEAM based design specification can provide a basis for a more consistent API and better modularity, extendability, scaling and support for concurrency.

Contents

36	Abstract	ii
37	1 Introduction	1
38	2 Background	2
39	2.1 Structure:	3
40	2.2 Tools	5
41	2.3 Data	8
42	2.4 Contribution	10
43	3 Topological Equivariant Artist Model	10
44	3.1 Data Space E	11
45	3.1.1 Variables in Fiber Space F	11
46	3.1.2 Measurement Scales: Monoid Actions	13
47	3.1.3 Continuity of the Data K	15
48	3.1.4 Data τ	18
49	3.1.5 Sheafs	19
50	3.1.6 Applications to Data Containers	20
51	3.2 Graphic Space H	21
52	3.2.1 Idealized Display D	21
53	3.2.2 Continuity of the Graphic S	22
54	3.2.3 Graphic ρ	23
55	3.3 Artist	26
56	3.3.1 Visual Fiber Bundle V	27
57	3.3.2 Visual Encoders ν	28
58	3.3.3 Visualization Assembly	31
59	3.3.4 Assembly Q	33
60	3.3.5 Composition of Artists: $+$	37
61	3.3.6 Equivalence class of artists A'	39

62	4 Prototype: Matplottoy	40
63	4.1 Scatter, Line, and Bar Artists	42
64	4.2 Visual Encoders	46
65	4.3 Data Model	47
66	4.4 Case Study: Penguins	49
67	5 Discussion	54
68	5.1 Limitations	54
69	5.2 Future Work	55
70	6 Conclusion	57

1 Introduction

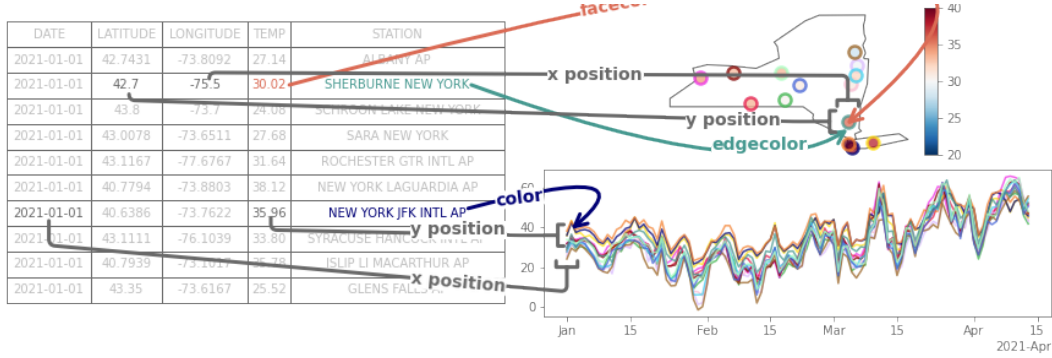


Figure 1: Visualizations are made up of transformations from data into visual representation. These functions transform individual data values to visual representation, such as date to x position or latitude to y position. These functions are composed into the assembly of all these transformations into a visual mark, such as a line or point. The same variable can be mapped in different ways, for example line is mapped to a color in the scatter plot and to y position in the line plot.

Visualization is the transformation of data into visual representation. As illustrated by Figure 1, these translations are both at the level of the individual variable and the entire record. In the case of the scatter plot, the latitude and longitude are encoded as the x and y position, respectively, while the temperature and station are represented by the face and edge colors. A row in the table is collectively encoded as a point mark. None of these encodings are fixed, as evidenced by temperature being translated into the y value in the case of the line plot. The station is now the source of the color of the entire line, and the date is the x position. As with scatter, the encodings of the individual transformations, which again are on values from the same record in the table, are composited into a line mark. It is these raw transformations from data space to visualization space that are implemented by building block level visualization libraries, named as such because the functions provided by the library can be composited in any number of ways to yield visualizations [1]. We propose that like physical building blocks, building block libraries must provide a collection of well defined pieces that can be composed in whichever ways the blocks fit together. We specify that a valid visualization block is a structure preserving transformation from data to visual space, and we define structure in terms of *continuity* and *equivariance*. We

88 then use this model to develop a design specification for the components of a building block
89 visualization library. The notion of self contained, inherently modular, building blocks lends
90 itself naturally to a functional paradigm of visualization [2]. We adopt a functional model
91 for a redesign because the lack of side effects means functional architecture can be evaluated
92 for correctness, functional programs tend to be shorter and clearer, and are well suited to
93 distributed, concurrent, and on demand tasks[3].

94 This work is strongly motivated by the needs of the Matplotlib[4, 5] visualization library.
95 One of the most widely used visualization libraries in Python, since 2002 new components
96 and features have been added in a some what adhoc, sometimes hard to maintain, manner.
97 Particularly, each new component carries its own implicit notion of how it believes the data is
98 structured-for example if the data is a table, cube, image, or network - that is then expressed
99 in the API for that component. In turn, this yields an inconsistent API for interfacing with
100 the data, for example when updating streaming visualizations or constructing dashboards[6].
101 This entangling of data model with visual transform also yields inconsistencies in how visual
102 component transforms, e.g. shape or color, are supported. We propose that these issues can
103 be ameliorated via a redesign of the functions that convert data to graphics, named *Artists* in
104 Matplotlib, in a manner that reliably enforces *continuity* and *equivariance* constraints. We
105 evaluate our functional model by implementing new artists in Matplotlib that are specified
106 via *equivariance* and *continuity* constraints. We then use the common data model introduced
107 by the model to demonstrate how plotting functions can be consolidated in a way that makes
108 clear whether the difference is in expected data structure, visual component encoding, or
109 the resulting graphic.

110 2 Background

111 There are many formalisms of the notion that visualization is structure preserving maps
112 from data to visual representation, and many visualization libraries that attempt to pre-
113 serve structure in some manner; this work bridges the formalism and implementation in a
114 functional manner with a topological approach at a building blocks library level to propose

a new model of the constraints visual transformations must satisfy such that they can be composed to produce visualize representations that can be considered equivalent to the data being represented.

2.1 Structure:

Visual representations of data, by definition, reflect something of the underlying structure and semantics[7], whether through direct mappings from data into visual elements or via figurative representations that have meaning due to their similarity in shape to external concepts [8]. The components of a visual representation were first codified by Bertin[9], who introduced a notion of structure preservation that we formally describe in terms of *equivariance* and *continuity*.

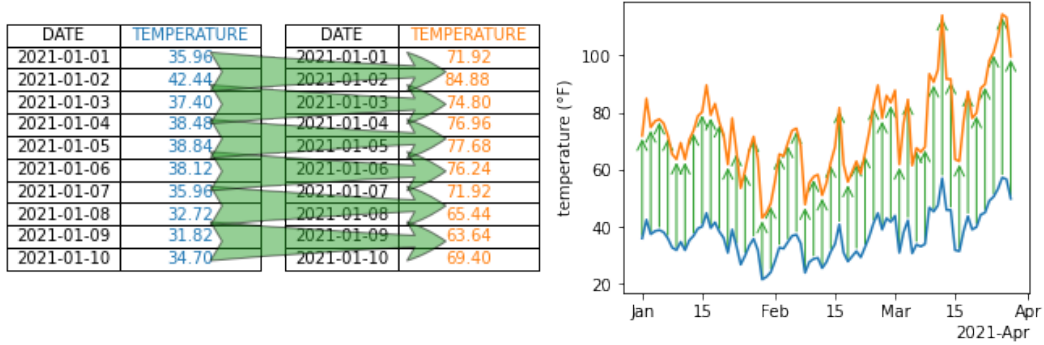


Figure 2: The data in blue is scaled by a factor of two, yielding the data in orange. To preserve *equivariance*, the blue line plot representation of the unscaled data is also scaled by a factor of two, yielding the orange line plot that is equivalent to the scaled data.

Bertin proposes that there are classes of visual encodings-such as position, shape, color, and texture-that when mapped to from specific types of measurement, quantitative or qualitative, will preserve the properties of that measurement type. For example, in Figure 2, the data and visual representation are scaled by equivalent factors of two, resulting in the change illustrated in the shift from blue to orange data and lines. The idea of equivariance is formally defined as the mapping of a binary operator from the data domain to the visual domain in Mackinlay’s *A Presentation Tool*(APT) model[10, 11]. The algebraic model of

132 visualization proposed by Kindlmann and Scheidegger uses equivariance to refer generally
 133 to invertible binary transformations[12], which are mathematical groups [13]. Our model
 134 defines *equivariance* in terms of monoid actions, which are a more restrictive set than all bi-
 135 nary operations and more general than groups. As with the algebraic model, our model also
 136 defines structure preservation as commutative mappings from data space to representation
 137 space to graphic space, but our model uses topology to explicitly include continuity.

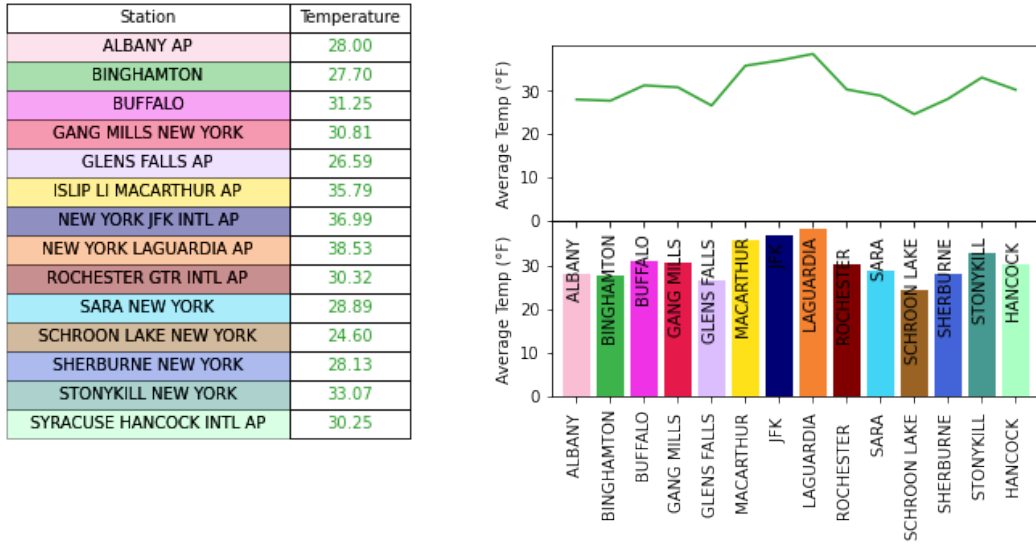


Figure 3: The line plot does not preserve *continuity* because it implies that the average temperature at each station lie along a 1D continuous line, while the bar plot preserves *continuity* by representing the average temperatures at each station as the discrete values they are.

138 Bertin proposes that the visual encodings be composited into graphical marks that match
 139 the *continuity* of the data - for example discrete data is a point, 1D continuous is the line,
 140 and 2D data is the area mark. In Figure 3, the line plot does not preserve continuity because
 141 the line connecting the discrete categories implies that the frequency of weather events is
 142 sampled from a continuous interval and the categories are points on that interval. But,
 143 when the continuity is preserved, as in the bar chart, then the graphic has not introduced
 144 new structure into the data.

Structure

continuity How records in the dataset are connected to each other, e.g. discrete rows, networked nodes, points on a continuous surface

equivariance if an action is applied to the data or the graphic—e.g. a rotation, permutation, translation, or rescaling—there must be an equivalent action applied on the other side of the transformation.

The notion that a graphic should be equivalent to the data has been expressed in a variety of ways. Informally, Norman’s Naturalness Principal[14] states that a visualization is easier to understand when the properties of the visualization match the properties of the data. This principal is made more concrete in Tufte’s concept of graphical integrity, which is that a visual representation of quantitative data must be directly proportional to the numerical quantities it represents (Lie Principal), must have the same number of visual dimensions as the data, and should be well labeled and contextualized, and not have any extraneous visual elements[15]. expressing, as defined by Mackinlay, is a measure how much of the mathematical structure in the data that can be expressed in the visualizations; for example that ordered variables can be mapped into ordered visual elements. We propose that a graphic is an equivalent representation of the data when *continuity* and *equivariance* are preserved.

2.2 Tools



Figure 4: Visualization libraries, especially ones tied to specific domains, tend to be architected around a core data structure, such as tables, images, or networks.

158 One of the reasons we developed a new formalism rather than adopting the architecture
159 of an existing library is that most information visualization software design patterns, as
160 categorized by Heer and Agrawala[16], are tuned to very specific data structures. These
161 libraries can often assume the expected data structure because they are domain specific,
162 and that is the common data structure in that domain. For users who generally work in
163 one domain, such as the data, networks, or graphs shown in Figure 4, this well defined data
164 space (and corresponding visual space[17]) often yields a very coherent user experience[18].
165 But, for developers who want to build new visualizations on top of these libraries, they must
166 work around the existing assumptions, sometimes in ways that break the model the libraries
167 are developed around.

168 For example, many domain specific libraries integrate computation into the visualization,
169 for example libraries based that assume all data is a relational database. This assumption is
170 core to tools influenced by APT, such as Tableau[19–21] and the Grammar of Graphics[22],
171 such as ggplot[23], protovis[24], vega[25] and altair[26]. Since these libraries represent all
172 data as a table, and computations on tables are fairly well defined[27], they can include
173 computations on the table with a fair bit of confidence that the computation is accurate.
174 Since most computations are specific to domains, general purpose block libraries can not
175 make this assumption; instead a goal of this model is to identify which computations are
176 specifically part of the visual encoding - for example mapping data to a color-and which
177 are manipulations on the data. Disentangling the computation from the visual transforms
178 allows us to determine whether the visualization library needs to handle them or if they can
179 be more efficiently computed by the data container.

180 A different class of user facing tools are those that support images, such as ImageJ[28]
181 or Napari[29]. These tools often have some support for visualizing non image components
182 of a complex data set, but mostly in service to the image being visualized. These tools are
183 ill suited for general purpose libraries that need to support data other than images because
184 the architecture is oriented towards building plugins into the existing system [30] where
185 the image is the core data structure. Even the digital humanities oriented ImageJ macro
186 ImagePlot[31], which supports some non-image aggregate reporting charts, is still built

around image data as the primary input. The need to visualize and manipulate graphs has spawned tools like Gephi[32], Graphviz[33], and Networkx[34]. As with tables and images, extending network libraries to work with other types of data either require breaking their internal model of how data is structured and what transformations of the data are allowable or growing a model for other types of data structures alongside the network model.

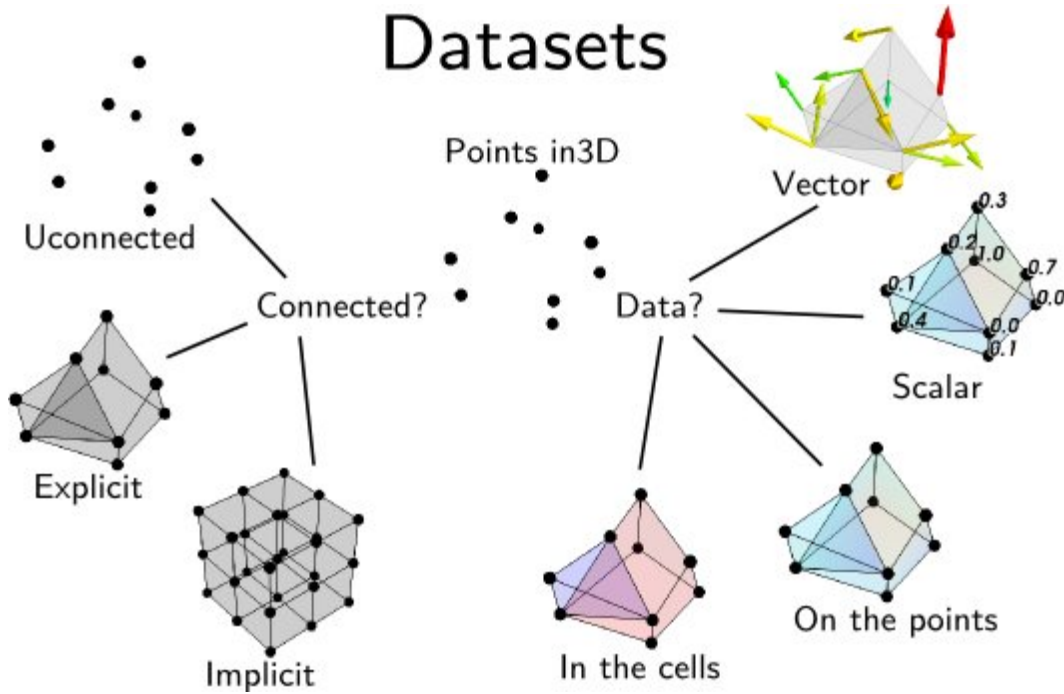


Figure 5: One way to describe data is by the connectivity of the points in the dataset. A database for example is often discrete unconnected points, while an image is an implicitly connected 2D grid. This image is from the Data Representation chapter of the MayaVi 4.7.2 documentation.[35]

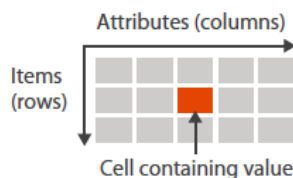
Many building block libraries carry multiple models of data internally because they cannot assume a data structure. Algorithms are designed such that the structure of data is assumed, as described in Tory and Möller’s taxonomy [ToryRethinkingVisualization2004], and by definition building block libraries try to provide the components to build any sort of visualization. Matplotlib, D3[36], and VTK [geveci2012vtk, 37] and its derivatives such as MayaVi[38] and extensions such as ParaView[39] and the infoviz themed Titan[40]. Where GoG and ImageJ type libraries have coherent APIs for their visualization tools because the

199 data structure is the same, the APIs for visualizations in Matplotlib, D3, and VTK are
 200 significantly dependent on the structure of the data it expects. VTK has codified this in
 201 terms of *continuity* based data representations, as illustrated in figure 5. This API choice
 202 can lead to visualizations that break *continuity* when fed into visualizations with different
 203 assumptions about structure. The lack of consistent data model can also mean no consistent
 204 way of updating the data and therefore no way of guaranteeing that the views are in sync,
 205 in visualizations that consist of multiple views of the same datasource, such as dash-
 206 boards[6, 41]. To resolve this issue, our functional model takes as input a structure aware
 207 data abstraction general enough to provide a common interface for many different types of
 208 visualization.

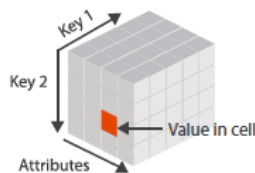
209 2.3 Data

210 One such general abstraction are fiber bundles, which Butler proposed as a core data struc-
 211 ture for visualization because they encode data continuity separately from the variable
 212 properties and are flexible enough to support discrete and ND continuous datasets [42, 43].
 213 Since Butler’s model lacks a robust way of describing variables, we can encode a schema
 214 like description of the data in the fiber bundle by folding in Spivak’s topological description
 215 of data types [44, 45]. In this work we will refer to the points of the dataset as *records*
 216 to indicate that a point can be a vector of heterogeneous elements. Each *component* of the
 217 record is a single object, such as a temperature measurement, a color value, or an image.
 218 We also generalize *component* to mean all objects in the dataset of a given type, such as
 219 all temperatures or colors or images. The way in which these records are connected is the
 220 *connectivity*, *continuity*, or more generally *topology*.

→ Tables



→ Multidimensional Table



→ Geometry (Spatial)



Figure 6: Values in a dataset have keys associated with them that describe where the value is in the dataset. These keys can be indexers or semantically meaningful; for example, in a table the keys are the variable name and the row ID. In the data cube, the keys are the row, column, and cell ID, and in the map the key is the position in the grid. Image is figure 2.8 in Munzner’s Visualization Analysis and Design[46]

The *continuity* can often be described by some variables in the dataset; this is formalized Munzner’s notion of metadata as *keys* into the data structure that return associated *values*[47]. As shown in Figure 6, keys can be labeled indexes, such as the attribute name and row ID, or physical entities such as locations on a map. We propose that information rich metadata are part of the components and instead the values are keyed on coordinate free structural ids. In contrast to Munzner’s model where the semantic meaning of the key is tightly coupled to the position of the value in the dataset, our model considers keys to be a pure reference to topology. This allows the metadata to be altered, without imposing new semantics on the underlying structure, for example by changing the coordinate systems or time resolution. This value agnostic model also supports encoding datasets where there may be multiple independent variables - such as a measure of plant growth given variations in water, sunlight, and time - without having to assume any one variable is inducing the change in growth. For building block library developers, this means the components are able to fully traverse the data structures without having to know anything about the values or the semantic meaning of the structure. Since these components are by design *equivariant* and *continuity* preserving, domain specific library developers in different domains that both rely on the same continuity, for example 2D continuity, can then safely reuse the components to build tools that can safely make domain specific assumptions.

239 2.4 Contribution

240 The contribution of this work is

- 241 1. formalization of the topology preserving relationship between data and graphic via
242 continuous maps [subsubsection 3.2.2](#)
- 243 2. formalization of property preservation from data component to visual representation
244 as monoid action equivariant maps [subsubsection 3.3.2](#)
- 245 3. functional oriented visualization architecture built on the mathematical model to
246 demonstrate the utility of the model [subsubsection 3.3.3](#)
- 247 4. prototype of the architecture built on Matplotlib’s infrastructure to demonstrate the
248 feasibility of the model. ??

249 3 Topological Equivariant Artist Model

To guide the implementation of structure preserving building block components, we develop a mathematical formalism of visualization that specifies how these components preserve *continuity* and *equivariance*. Inspired by the somewhat analogous component in Matplotlib[5], we call the transformation from data space to graphic that these building block components implement the *artist*.

$$\mathcal{A} : \mathcal{E} \rightarrow \mathcal{H} \tag{1}$$

250 The *artist* \mathcal{A} is a map from the data \mathcal{E} to graphic \mathcal{H} fiber bundles. To explain how
251 the *artist* is a structure preserving map from data to graphic, we first describe how we
252 model data ([subsection 3.1](#)) and graphics ([subsection 3.2](#)) as topological structures that
253 encapsulate component types and continuity. We then discuss the maps from graphic to
254 data ([subsubsection 3.2.2](#), data components to visual components ([subsubsection 3.3.2](#)), and
255 visual components into graphic ([subsubsection 3.3.3](#)) that make up the artist.

256 3.1 Data Space E

Building on Butler’s proposal of using fiber bundles as a common data representation structure for visualization data[42, 43], a fiber bundle is a tuple (E, K, π, F) defined by the projection map π

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

257 that binds the components of the data in F to the continuity of the data encoded in K .
 258 The fiber bundle models the properties of data component types F (subsection 3.1.1),
 259 the continuity of records K (subsection 3.1.3), the collections of records τ (??), and the
 260 space E of all possible datasets with these components and continuity. By definition fiber
 261 bundles are locally trivial[48, 49], meaning that over a localized neighborhood U the total
 262 space is the cartesian product $K \times F$. We use fiber bundles as the data model because they
 263 are inclusive enough to express all the types of structures of data described in subsection 2.2

264 3.1.1 Variables in Fiber Space F

To formalize the structure of the data components, we use notation introduced by Spivak [45] that binds the components of the fiber to variable names. This allows us to describe the components in a schema like way. Spivak constructs a set \mathbb{U} that is the disjoint union of all possible objects of types $\{T_0, \dots, T_m\} \in \mathbf{DT}$, where \mathbf{DT} are the data types of the variables in the dataset. He then defines the single variable set \mathbb{U}_σ

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

which is \mathbb{U} restricted to objects of type T bound to variable name c . The \mathbb{U}_σ lookup is by name to specify that every component is distinct, since multiple components can have the same type T . Given σ , the fiber for a one variable dataset is

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

where σ is the schema that binds a variable name c to its datatype T . A dataset with multiple components has a fiber that is the cartesian cross product of \mathbb{U}_σ applied to all the columns:

$$F = \mathbb{U}_{\sigma(c_1)} \times \dots \times \mathbb{U}_{\sigma(c_i)} \times \dots \times \mathbb{U}_{\sigma(c_n)} \quad (5)$$

which is equivalent to

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

265 which allows us to decouple F into components F_i .

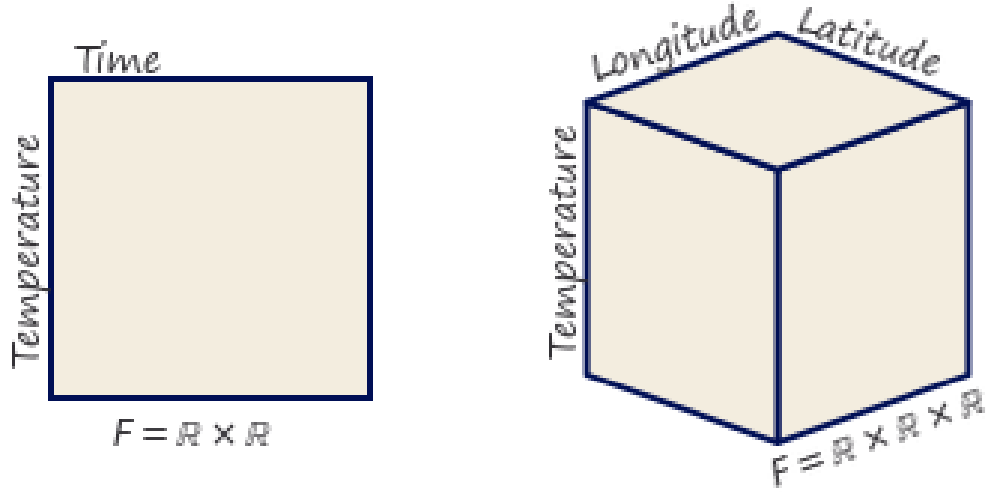


Figure 7: The fiber space is the set of all combinations of all theoretically possible values of the components. The 2D fiber $F = \mathbb{R} \times \mathbb{R}$ encodes the properties of *time* and *temperature* components. One dimension of the fiber encodes the range of possible values for the time component of the dataset, which is a subset of the \mathbb{R} , while the other dimension encodes the range of possible values \mathbb{R} for the temperature component. This means the fiber is the set of points $(temperature, time)$ that are all the combinations of *temperature* \times *time*. The 3D fiber encodes points at all possible combinations of *temperature*, *latitude*, and *longitude*.

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

$$F = \mathbb{R} \times \mathbb{R}^+$$

where the first component F_0 is the set of values specified by $(c = \text{time}, T = \text{datetime}, \mathbb{U}_\sigma = \mathbb{R})$ and F_1 is specified by $(c = \text{temperature}, T = \text{float}, \mathbb{U}_\sigma = \mathbb{R})$ and is the set of values $\mathbb{U}_\sigma = \mathbb{R}$. In the 3D fiber in ??, time is replaced with location. This location variable is of type **point** and has two components *latitude* and *longitude* $\{(lat, lon) \in \mathbb{R}^2 \mid -90 \leq lat \leq 90, 0 \leq lon \leq 360\}$. The fiber for this dataset is

$$F = \mathbb{R} \times \mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$$

266 where the dimensionality of the fiber does not change, but the components of the fiber
 267 can be coupled. For example, *location* can either be specified as $(c = \text{location}, T =$
 268 **point**, $\mathbb{U}_\sigma = \mathbb{R}^2)$ or $(c = \text{latitude}, T = \text{float}, \mathbb{U}_\sigma = \mathbb{R})$ and $(c = \text{longitude}, T =$
 269 **float**, $\mathbb{U}_\sigma = \mathbb{R})$.

270 As illustrated in [Figure 7](#), Spivak's framework provides a consistent way to describe
 271 potentially complex components of the input data.

272 3.1.2 Measurement Scales: Monoid Actions

Implementing expressive visual encodings requires formally describing the structure on the components of the fiber, which we define by the actions of a monoid on the component. In doing so, we specify the properties of the component that must be preserved in a graphic representation. A monoid [50] M is a set with a binary operation $* : M \times M \rightarrow M$ that satisfies the axioms:

associativity for all $a, b, c \in M$ $(a \bullet b) \bullet c = a \bullet (b \bullet c)$

identity for all $a \in M$, $e \bullet a = a$

As defined on a component of F , a left monoid action [51, 52] 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 $x \in F_i, e \in M_i$, $e \bullet x = x$

273 The identity and associativity properties of the action denote that the action is a monoid
 274 homomorphism, which means that the group operation is preserved on both sides of the
 275 action[53].

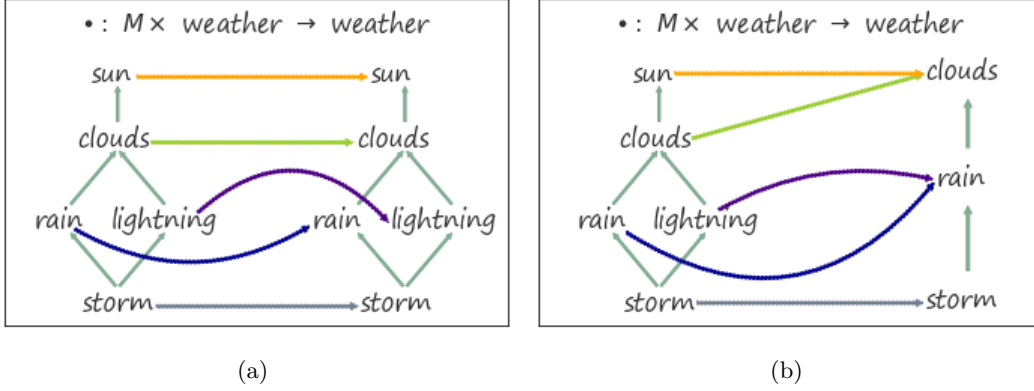


Figure 8: The action \bullet in ?? is the arrows from the partial order diagram of weather states on the left to the diagram of weather states on the right. Since the action maps the weather states to themselves, the ordering defined by the monoid $*$ is preserved on both sides of the action. The action in ?? is monoid homomorphism because the ordering of the weather states is the same as the ordering of the elements they are mapped to. Given $sun \geq clouds \geq rain/lightning$ on the right, the action $sun clouds \rightarrow clouds$, and $rain/lightning \rightarrow rain$ is structure preserving because on the left $cloud \geq rain$ so the relative ordering of elements is the same as the elements they are mapped to.

One example of monoids are partial orderings on a set, such as seen in . Each hasse diagram of the set of weather states describes an ordering on the set; the arrow goes from the lesser value to the greater one. For example, $storm \leq rain$. In ??, the action \bullet maps the elements of a set of weather states into itself by mapping them into other elements of the weather states. The action in Figure 8a, represented as the arrows between the hasse diagrams of the weather states, maps the weather states to themselves; therefore the

ordering of the weather states is identical on both sides of the action and it is therefore homomorphic. The action \bullet in Figure 8b is a monotone map[54]

$$if\ a \leq b\ then\ \bullet(a) \leq \bullet(b) \mid a, b \in F_i$$

276 where the structure the action preserves is the relative, rather than exact, ordering. Since
 277 groups are monoids with invertible operations, this definition of structure is also broad
 278 enough to include the Steven's measurment scales[55, 56]. Monoids are also commonly
 279 found in functional programming since the core property of monoids is composability [57].

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

$$M = M_0 \times \dots \times M_i \times \dots \times M_n \quad (7)$$

280 of each monoid M_i on F_i . The monoid is also added to the specification of the fiber
 281 $(c_i, T_i, \mathbb{U}_\sigma M_i)$

282 3.1.3 Continuity of the Data K



Figure 9: The topological base space K encodes the continuity of the data space, for example if the data is discrete points or lies on a plane or a sphere

283 The base space K acts as an indexing space, as emphasized by Butler[42, 43], to express
 284 how the records in E are connected to each other. As shown in Figure 9, K can have
 285 any number of dimensions and can be continuous or discrete. The base space also does
 286 not describe anything about the dataset besides the continuity. While the base space may
 287 have components to identify the continuity, such as *time*, *latitutde*, *longitude*, these labels

288 are indexed into from K the same as any other component. This is similar to the notion of
 289 structural *keys* with associated *values* proposed by Munzner[46], but our model treats keys as
 290 a pure reference to topology. Decoupling the keys from their semantics allows the metadata
 291 to be altered; this provides for coordinate agnostic representation of the continuity and
 292 facilitates encoding of data where the independent variable may not be clear. For example
 293 the amount of snow on the ground is dependent on time of day and how much snow has
 294 fallen, and changing the coordinate system or time resolution should have no effect on how
 295 the records are connected to each other.

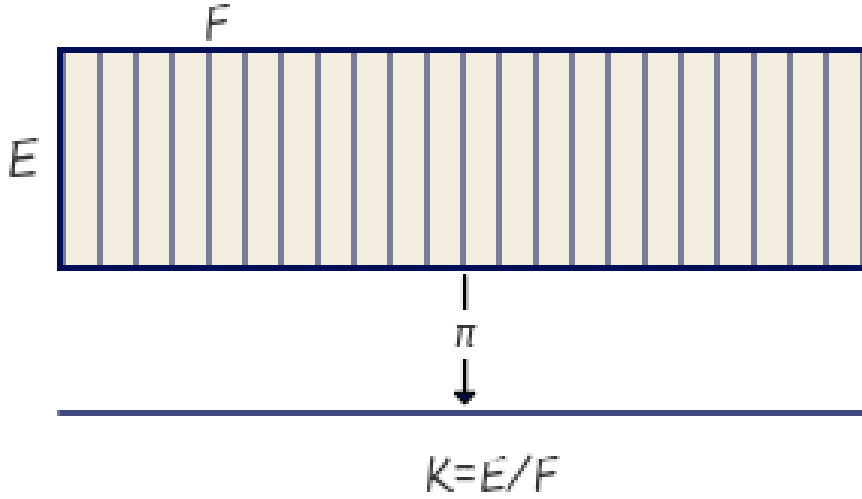


Figure 10: The base space E is divided into fiber segments F . The base space K acts as an index into the records in the fibers.

296 Formally K is the quotient space [58] of E meaning it is the finest space[59] such that
 297 every $k \in K$ has a corresponding fiber F_k [58]. In Figure 10, E is a rectangle divided by
 298 vertical fibers F , so the minimal K for which there is always a mapping $\pi : E \rightarrow K$ is the
 299 closed interval $[0, 1]$.

As with Equation 6 and Equation 7, we can decompose the total space into component bundles $\pi : E_i \rightarrow K$ where

$$\pi : E_1 \oplus \dots \oplus E_i \oplus \dots \oplus E_n \rightarrow K \quad (8)$$

such that M_i acts on component bundle E_i . The K remains the same because the connectivity of records does not change just because there are fewer components in each record. By encoding this continuity in the model as K the data model now explicitly carries information about its structure such that the implicit assumptions of the visualization algorithms are now explicit.

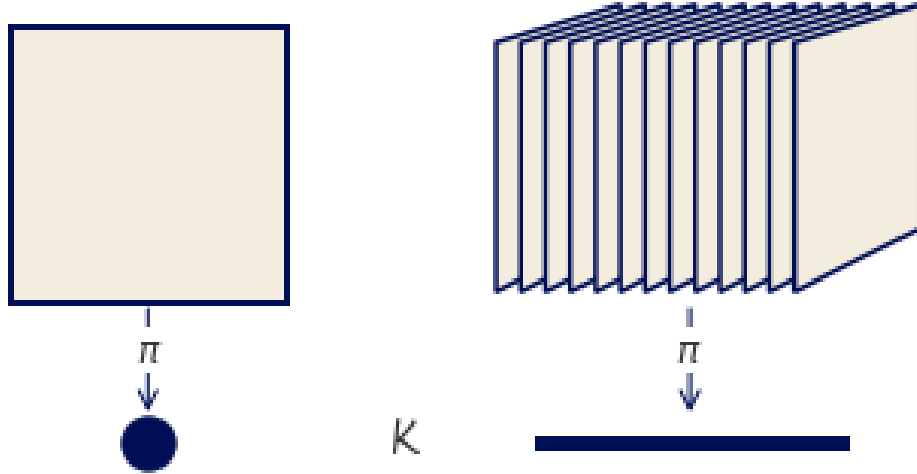


Figure 11: These two datasets have the same $(time, temperature)$ fiber, but different continuities. The dataset on the left consists of discrete records, while the records in the dataset on the right sampled from a continuous space.

The datasets in Figure 11 have the same fiber of $(temperature, time)$. The dot represents a discrete base space K , meaning that every dataset encoded in the fiber bundle has discrete continuity. The line is a representation of a 1D continuity, meaning that every dataset in the fiber bundle is 1D continuous. By encoding this continuity in the model as K the data model now explicitly carries information about its structure such that the implicit assumptions of

the visualization algorithms are now explicit. The explicit topology is a concise way of distinguishing visualizations that appear identical, for example heatmaps and images.

3.1.4 Data τ

While the projection function $\pi : E \rightarrow K$ ties together the base space K with the fiber F , a section $\tau : K \rightarrow E$ encodes a dataset. A section function takes as input location $k \in K$ and returns a record $r \in E$. For example, in the special case of a table [45], K is a set of row ids, F is the columns, and the section τ returns the record r at a given key in K . For any fiber bundle, there exists a map

$$\begin{array}{ccc} F & \hookrightarrow & E \\ & \pi \downarrow \nearrow \tau & \\ & K & \end{array} \quad (9)$$

such that $\pi(\tau(k)) = k$. The set of all global sections is denoted as $\Gamma(E)$. Assuming a trivial fiber bundle $E = K \times F$, the section is

$$\tau(k) = (k, (g_{F_0}(k), \dots, g_{F_n}(k))) \quad (10)$$

where $g : K \rightarrow F$ is the index function into the fiber. This formulation of the section also holds on locally trivial sections of a non-trivial fiber bundle. Because we can decompose the bundle and the fiber (Equation 8, Equation 6), we can decompose τ as

$$\tau = (\tau_0, \dots, \tau_i, \dots, \tau_n) \quad (11)$$

where each section τ_i maps into a record on a component $F_i \in F$. This allows for accessing the data component wise in addition to accessing the data in terms of its location over K .

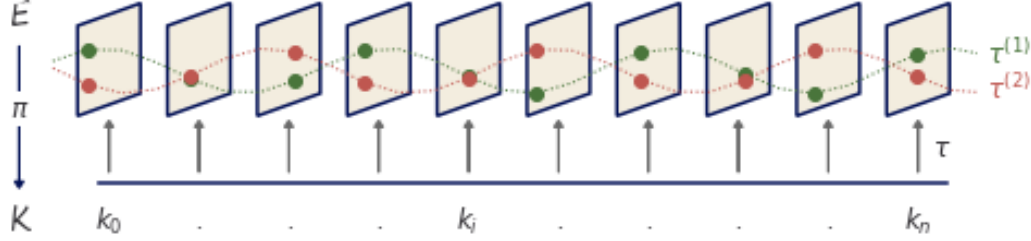


Figure 12: Fiber (time, temperature) with an interval K basespace. The sections $\tau^{(1)}$ and $\tau^{(2)}$ are constrained such that the time variable must be monotonic, which means each section is a timeseries of temperature values. They are included in the global set of sections $\tau^{(1)}, \tau^{(2)} \in \Gamma(E)$

In Figure 12, the fiber is the same encoding of *(time, temperature)* illustrated in Figure 7, and the base space is the interval K shown in Figure 11. The section $\tau^{(1)}$ is a function that for a point k returns a record in the fiber F . The section applied to a set of points in K resolves to a series of monotonically increasing in time records of *(time, temperature)* values. Section $\tau^{(2)}$ returns a different timeseries of *(time, temperature)* values. Both sections are included in the global set of sections $\tau^{(1)}, \tau^{(2)} \in \Gamma(E)$.

3.1.5 Sheafs

Many types of dynamic visualizations require evaluating sections on different subspaces of K , and a sheaf, denoted \mathcal{O} provides a way to do so. A sheaf is a mathematical structure for defining collections of objects[60–62] on mathematical spaces. On the fiber bundle E , we can describe a sheaf as the collection of local sections $\iota^* \tau$

$$\begin{array}{ccc} \iota^* E & \xleftarrow{\iota^*} & E \\ \pi \downarrow \Big) \iota^* \tau & & \pi \downarrow \Big) \tau \\ U & \xleftarrow{\iota} & K \end{array} \quad (12)$$

which are sections of E pulled back over local neighborhood $U \subset E$ via the inclusion map $\iota : E \rightarrow U$. The collation of sections enabled by sheafs is necessary for navigation techniques such as pan and zoom[63] and dynamically updated visualizations such as sliding windows[64, 65].

3.1.6 Applications to Data Containers

This model provides a common formalism for widely used data containers without sacrificing the semantic structure embedded in each container. For example, the section can be any instance of a univariate numpy array[66] that stores an image. This could be a section of a fiber bundle where K is a 2D continuous plane and the F is $(\mathbb{R}^3, \mathbb{R}, \mathbb{R})$ where \mathbb{R}^3 is color, and the other two components are the x and y positions of the sampled data in the image. This position information is already implicitly encoded in the array as the index and the resolution of the image being stored. Instead of an image, the numpy array could also store a 2D discrete table. The fiber would not change, but the K would now be 0D discrete points. These different choices in topology indicate, for example, what sorts of interpolation would be appropriate when visualizing the data.

There are also many types of labeled containers that can richly be described in this framework because of the schema like structure of the fiber. For example, a pandas series which stores a labeled list, or a dataframe[67] which stores a relational table. A series could store the values of $\tau^{(1)}$ and a second series could be $\tau^{(2)}$. We could also fatten the fiber to hold two temperature series, such that a section would be an instance of a dataframe with a time column and two temperature columns. While the series and dataframe explicitly have a time index column, they are components in our model and the index is assumed to be data independent references such as hashvalues, virtual memory locations, or random number keys.

Where this model particularly shines are N dimensional labeled data structures. For example, an xarray[68] data that stores temperature field could have a K that is a continuous volume and the components would be the temperature and the time, latitude, and longitude the measurements were sampled at. A section can also be an instance of a distributed data container, such as a dask array [69]. As with the other containers, K and F are defined in terms of the index and dtypes of the components of the array. Because our framework is defined in terms of the fiber, continuity, and sections, rather than the exact values of the data, our model does not need to know what the exact values are until the renderer needs to fill in the image.

355 3.2 Graphic Space H

To establish that the artist is structure preserving map from data E to graphic H we construct a graphic bundle so that we can define *equivariance* in terms of maps on the fiber spaces and *continuity* in terms of maps on the base space. As with the data, we can represent the target graphic as a section ρ of a bundle (H, S, π, D) .

$$D \hookrightarrow H \quad \begin{array}{c} \pi \downarrow \\ S \end{array} \quad \begin{array}{c} \nearrow \\ \searrow \end{array} \quad \rho \quad (13)$$

356 The graphic bundle H consists of a base S ([subsubsection 3.2.1](#)) that is a thickened form of
 357 K a fiber D ([subsubsection 3.2.2](#)) that is an idealized display space, and sections ρ ([??](#)) that
 358 encode a graphic where the visual characteristics are fully specified.

359 3.2.1 Idealized Display D

To fully specify the visual characteristics of the image, we construct a fiber D that is an infinite resolution version of the target space. Typically H is trivial and therefore sections can be thought of as mappings into D . In this work, we assume a 2D opaque image $D = \mathbb{R}^5$ with elements

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

360 such that a rendered graphic only consists of 2D position and color. To support overplotting
 361 and transparency, the fiber could be $D = \mathbb{R}^7$ such that $(x, y, z, r, g, b, a) \in D$ specifies the
 362 target display. By abstracting the target display space as D , the model can support different
 363 targets, such as a 2D screen or 3D printer.

364 **3.2.2 Continuity of the Graphic S**

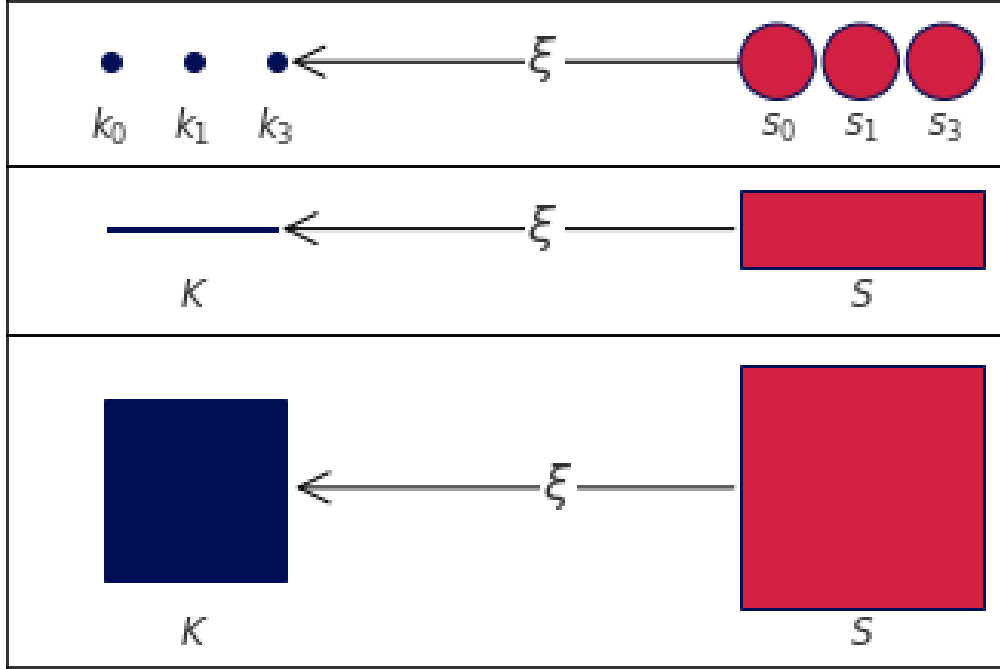


Figure 13: For a visualization component to preserve continuity, it must have a continuous surjective map $\xi : S \rightarrow K$ from graphic continuity to data continuity. The scatter and line graphic base spaces S have one more dimension of continuity than K so that S can encode physical aspects of the glyph, such as shape (a circle) or thickness. The image has the same dimension in S as in K .

365 To establish that a visualization component preserves continuity, we propose that their must
 366 be a continuous map $\xi : S \rightarrow K$ from the graphic base space to the data space. For example,
 367 consider a S that is mapped to the region of a 2D display space that represents K . For
 368 some visualizations, K may be lower dimension than S . For example, a point that is 0D in
 369 K cannot be represented on screen unless it is thickened to 2D to encode the connectivity
 370 of the pixels that visually represent the point. This thickening is often not necessary when
 371 the dimensionality of K matches the dimensionality of the target space, for example if K is
 372 2D and the display is a 2D screen. We introduce S to thicken K in a way which preserves
 373 the structure of K .

Formally, we require that K be a deformation retract[70] of S so that K and S have the same homotopy, meaning there is a continuous map from S to K [71]. The surjective map $\xi : S \rightarrow K$

$$\begin{array}{ccc} E & & H \\ \pi \downarrow & & \pi \downarrow \\ K & \xleftarrow{\xi} & S \end{array} \quad (14)$$

374 goes from region $s \in S_k$ to its associated point s . This means that if $\xi(s) = k$, the record at
 375 k is copied over the region s such that $\tau(k) = \xi^* \tau(s)$ where $\xi^* \tau(s)$ is τ pulled back over S .
 376 When K is discrete points and the graphic is a scatter plot, each point $k \in K$ corresponds
 377 to a 2D disk S_k as shown in Figure 13. In the case of 1D continuous data and a line plot,
 378 the region β over a point α_i specifies the thickness of the line in S for the corresponding
 379 τ on k . The image has the same dimensions in data space and graphic space such that no
 380 extra dimensions are needed in S .

381 The mapping function ξ provides a way to identify the part of the visual transformation
 382 that is specific to the the connectivity of the data rather than the values; for example it
 383 is common to flip a matrix when displaying an image. The ξ mapping is also used by
 384 interactive visualization components to look up the data associated with a region on screen.
 385 One example is to fill in details in a hover tooltip, another is to convert region selection (such
 386 as zooming) on S to a query on the data to access the corresponding record components on
 387 K .

388 3.2.3 Graphic ρ

The section $\rho : S \rightarrow H$ is the graphic in an idealizes prerender space and also acts as a specification for rendering the graphic to an image. It is sufficient to sketch out how an arbitrary pixel would be rendered, where a pixel p in a real display corresponds to a region S_p in the idealized display. To determine the color of the pixel, we aggregate the color values

over the region via integration:

$$\begin{aligned} r_p &= \iint_{S_p} \rho_r(s) ds^2 \\ g_p &= \iint_{S_p} \rho_g(s) ds^2 \\ b_p &= \iint_{S_p} \rho_b(s) ds^2 \end{aligned}$$

389 For a 2D screen, the pixel is defined as a region $p = [y_{top}, y_{bottom}, x_{right}, x_{left}]$ of the rendered
 390 graphic. Since the x and y in p are in the same coordinate system as the x and y components
 391 of D the inverse map of the bounding box $S_p = \rho_{xy}^{-1}(p)$ is a region $S_p \subset S$. The color is
 392 the result of the integration over S_p .

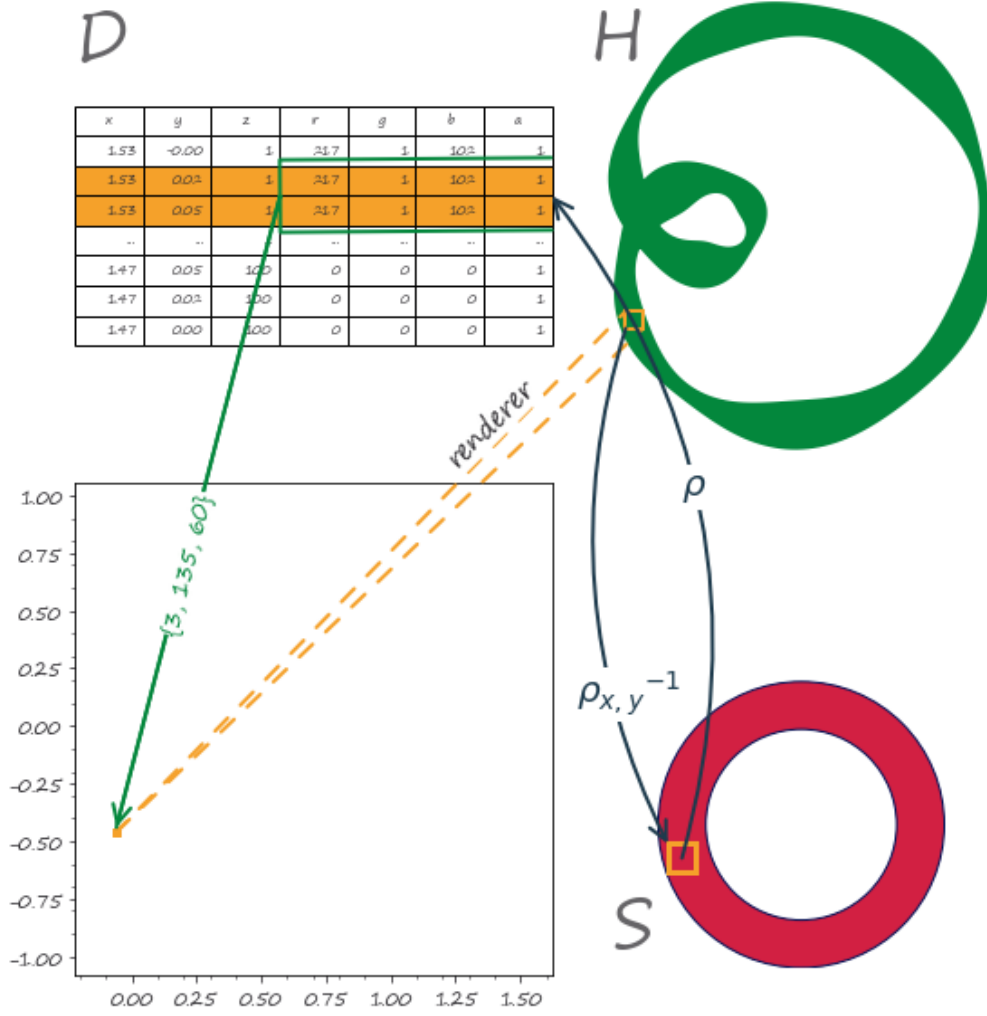


Figure 14: To render a graphic, a pixel p is selected in the display space, which is defined in the same coordinates as the x and y components in D via the renderer. In H the inverse mapping $\rho_{xy}(p)$ returns a region $S_p \subset S$. $\rho(S_p)$ returns a set of points $(x, y, r, g, b) \in D$ that lie over S_p . The integral over the (r, g, b) pixels specifies that the pixel should be green

As shown in Figure 14, a pixel p in the output space, drawn in yellow, is selected and mapped, via the renderer, into a region on H . The region on H corresponds to a region $S_p \subset S$ via the inverse mapping $\rho_{xy}(p)$. The base space S is an annulus to match the topology of the graphic idealized in H . The section $\rho(S_p)$ then maps into the fiber D over S_p to obtain the set of points in D , here represented as a table, that correspond to that

398 section. The integral over the pixel components of this set of points in the fiber yields
 399 $\{3, 135, 60\}$ the actual color of the pixel. In general, ρ is an abstraction of rendering.
 400 In very broad strokes ρ can be a specification such as PDF[72], SVG[73], or an OpenGL
 401 scene graph[74]. Alternatively, ρ can be a rendering engine such as cairo[75] or AGG[76].
 402 Implementation of ρ is out of scope for this work,

403 3.3 Artist

The topological artist A is how we model the building block component that transforms data into a graphic. The artist A is a map from the sheaf on a data bundle E which is $\mathcal{O}(E)$ to the sheaf on the graphic bundle H , $\mathcal{O}(H)$.

$$A : \mathcal{O}(E) \rightarrow \mathcal{O}(H) \quad (15)$$

The artist preserves *continuity* through the ξ map discussed in [subsubsection 3.2.2](#) and is an *equivariant* map because it carries a homomorphism of monoid actions [77]

$$\varphi : M \rightarrow M' \quad (16)$$

Given M on data \mathcal{E} and M' on graphic \mathcal{H} , we propose that artists \mathcal{A} are equivariant maps

$$A(m \cdot r) = \varphi(m) \cdot A(r) \quad (17)$$

404 such that applying a monoid action $m \in M$ to the data input $r \in \mathcal{E}$ of the artist \mathcal{A} is
 405 equivalent to applying a monoid action $\varphi(m) \in M'$ to the graphic $A(r) \in \mathcal{H}$ output of the
 406 artist.

The monoid equivariant map has two stages: the encoders $\nu : E' \rightarrow V$ convert the data components to visual components, and the assembly function $Q : \xi^*V \rightarrow H$ composites the

fiber components of ξ^*V into a graphic in H .

$$\begin{array}{ccccc}
 E' & \xrightarrow{\nu} & V & \xleftarrow{\xi^*} & \xi^*V & \xrightarrow{Q} & H \\
 & \searrow \pi & \downarrow \pi & & \xi^* \downarrow \pi & \swarrow \pi & \\
 & & K & \xleftarrow{\xi} & S & &
 \end{array} \tag{18}$$

407 ξ^*V is the visual bundle V pulled back over S via the equivariant continuity map $\xi : S \rightarrow K$
 408 introduced in [subsubsection 3.2.2](#). The functional decomposition of the visualization artist
 409 in [Equation 18](#) facilitates building reusable components at each stage of the transformation
 410 because the equivariance constraints are defined on ν , Q , and ξ . We name this map the artist
 411 as that is the analogous part of the Matplotlib[5] architecture that builds visual elements.

412 3.3.1 Visual Fiber Bundle V

We introduce a visual bundle V to store the mappings of the data components into components of the graphic. The visual bundle (V, K, π, P) is the space of possible parameters of a visualization type, such as a scatter or line plot. As with the data and graphic bundles, the visual bundle is defined by the projection map π

$$\begin{array}{ccc}
 P & \hookrightarrow & V \\
 & & \pi \downarrow \Big)^\mu \\
 & & K
 \end{array} \tag{19}$$

413 where μ is the visual variable encoding, as described by Bertin [9], of the data section τ .
 414 The visual fiber P is defined in terms of the input parameters of the visualization library's
 415 plotting functions; by making these parameters explicit components of the fiber, we can
 416 build consistent definitions and expectations of how these parameters behave.

ν_i	μ_i	$\text{codomain}(\nu_i) \subset P_i$
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	$(\mathbb{R}, \mathbb{R}^{+n, n\%2=0})$

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

417 A section μ is a tuple of visual values that specifies the visual characteristics of a part
 418 of the graphic. For example, given a fiber of $\{x, y, color\}$ one possible section could be
 419 $\{.5, .5, (255, 20, 147)\}$. The $\text{codomain}(\nu_i)$ determines which monoids can act on P_i . These
 420 fiber components are implicit in the library, as seen in [Table 1](#), and by making them explicit
 421 as components of the fiber we can build consistent definitions and expectations of how these
 422 parameters behave.

423 3.3.2 Visual Encoders ν

We define the visual transformers ν

$$\{\nu_0, \dots, \nu_n\} : \{\tau_0, \dots, \tau_n\} \mapsto \{\mu_0, \dots, \mu_n\} \quad (20)$$

as the set of equivariant maps $\nu_i : \tau_i \mapsto \mu_i$. Given M_i is the monoid action on E_i and that there is a monoid M_i' on V_i , then there is a monoid homomorphism from $\varphi : M_i \rightarrow M_i'$ that ν must preserve. As mentioned in [subsubsection 3.1.2](#), monoid actions define the structure on the fiber components and are therefore the basis for equivariance. A validly constructed ν is one where the diagram of the monoid transform m commutes

$$\begin{array}{ccc} E_i & \xrightarrow{\nu_i} & V_i \\ m_r \downarrow & & \downarrow m_v \\ E_i & \xrightarrow{\nu_i} & V_i \end{array} \quad (21)$$

such that applying equivariant monoid actions to E_i and V_i preserves the map $\nu_i : E_i \rightarrow V_i$. In general, the data fiber F_i cannot be assumed to be of the same type as the visual fiber P_i and the actions of M on F_i cannot be assumed to be the same as the actions of M' on P ; therefore an equivariant ν_i must satisfy the constraint

$$\nu_i(m_r(E_i)) = \varphi(m_r)(\nu_i(E_i)) \quad (22)$$

424 such that φ maps a monoid action on data to a monoid action on visual elements. However,
 425 we can construct a monoid action of M on P_i that is compatible with a monoid action of
 426 M on F_i . We can compose the monoid actions on the visual fiber $M' \times P_i \rightarrow P_i$ with the
 427 homomorphism φ that takes M to M' . This allows us to define a monoid action on P of M
 428 that is $(m, v) \rightarrow \varphi(m) \bullet v$. Therefore, without a loss of generality, we can assume that an
 429 action of M acts on F_i and on P_i compatibly such that φ is the identity function.

430 The translation from weather state data to visual representation as umbrella emoji in
 431 [Figure 15a](#) is an invalid visual encoding map ν because it is not homomorphic. This is
 432 because the monotonic condition $rain \geq storm \implies \nu(rain) \geq \nu(storm)$ is not met since
 433 $\nu(rain) \leq \nu(storm)$. To satisfy the monotonic condition for $rain \geq storm$, either red arrow
 434 in [Figure 15a](#) would have to go in a different direction. On the other hand, the mapping
 435 from weather state to umbrella in [Figure 15b](#) is a homomorphism since $\nu(rain) = \nu(storm)$
 436 satisfies the monotonic condition of $rain \geq storm$. [Figure 15](#) is an example of how the

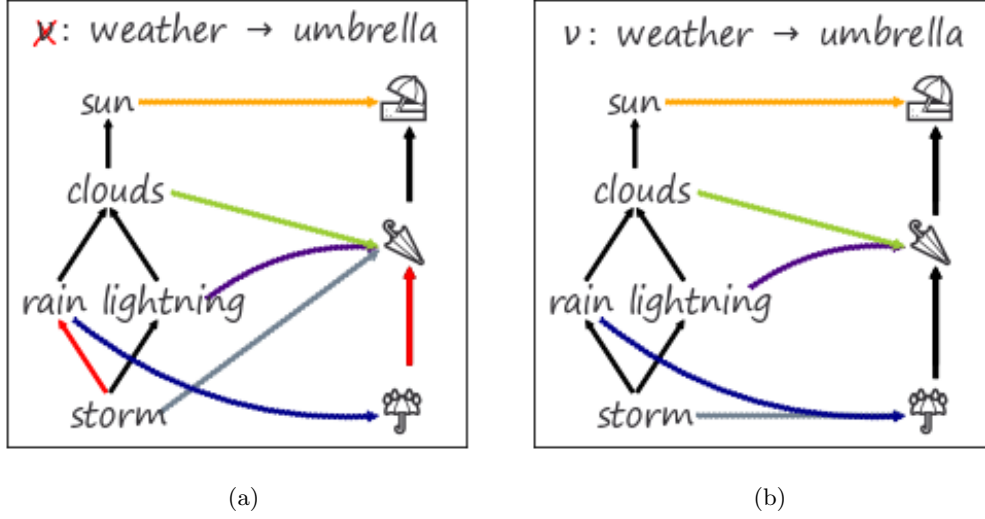


Figure 15: The map from data component to visual component in Figure 15a is not homomorphic, and therefore invalid, because $\text{rain} \geq \text{storm}$ is mapped to elements with the reverse ordering $\nu(\text{storm}) \geq \nu(\text{umbrella})$. In contrast, the mapping in Figure 15b is valid since $\nu(\text{storm}) = \nu(\text{rain})$ satisfies the condition $\nu(\text{storm}) \geq \nu(\text{umbrella})$.

437 model supports partially ordered data components, which was a motivation for defining
 438 equivariance as monoid homomorphisms.

scale	group	constraint
nominal	permutation	if $r_1 \neq r_2$ then $\nu(r_1) \neq \nu(r_2)$
ordinal	monotonic	if $r_1 \leq r_2$ then $\nu(r_1) \leq \nu(r_2)$
interval	translation	$\nu(x + c) = \nu(x) + c$
ratio	scaling	$\nu(xc) = \nu(x) * c$

Table 2

439 The Stevens measurement types[55], listed in Table 2, are specified in terms of groups,
 440 which are monoids with invertible operations[78]. Despite critiques of the scales[79, 80], we
 441 believe it is critical for the model to include the measurement scales since they are com-
 442 monly used in visualization to classify components [22, 46]. By specifying the equivariance
 443 constraints on ν we can guarantee that the stage of the artist that transforms data compo-
 444 nents into visual representations is equivariant. These constraints guide the implementation
 445 of reusable component transformers ν that are composed when generating the graphic.

446 3.3.3 Visualization Assembly

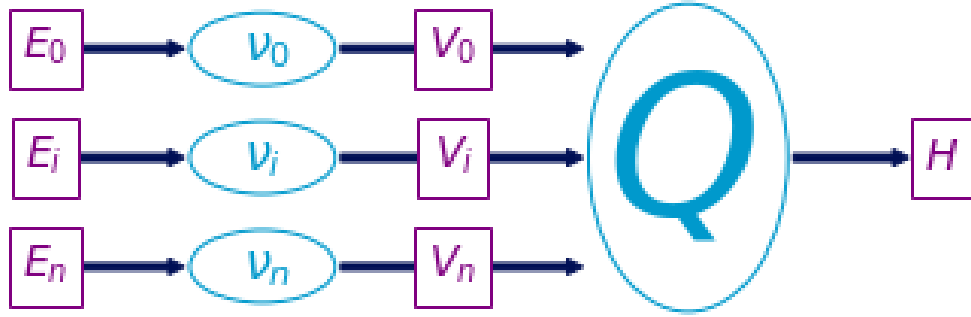


Figure 16: The transform functions ν_i convert data $\tau_i \in E$ to visual characteristics $\mu_i \in V$, then Q assembles μ_i into a graphic $\rho \in H$.

447 The transformation from data into graphic is analogous to a map-reduce operation;
 448 as illustrated in ??, data components E_i are mapped into visual components V_i that are
 449 reduced into a graphic in H . The space of all graphics that Q can generate is the subset
 450 of graphics reachable via applying the reduction function $Q(\Gamma(V)) \in \Gamma(H)$ to the visual
 451 section $\mu \in \Gamma(V)$. The full space of graphics is not necessarily equivariant; therefore we
 452 formalize the constraints on Q such that it produces structure preserving graphics.

453 We formalize the expectation that visualization generation functions parameterized in
 454 the same way should generate the same functions as the equivariant map $Q : \mu \mapsto \rho$. We

455 then define the constraint on Q such that if Q is applied to two visual sections μ and μ'
 456 that generate the same ρ then the output of μ and μ' acted on by the same monoid m
 457 must be the same. We do not define monoid actions on all of $\Gamma(H)$ because there may be
 graphics $\rho \in \Gamma(H)$ for which we cannot construct a valid mapping from V . Lets call the

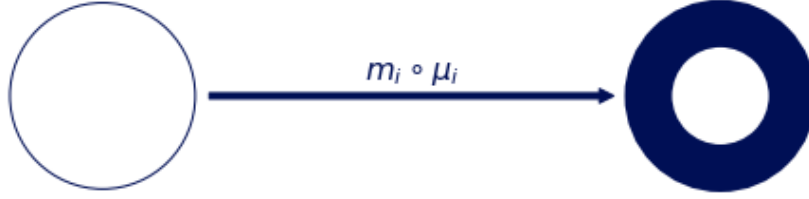


Figure 17: These two glyphs are generated by the same annulus Q function. The monoid action m_i on edge thickness μ_i of the first glyph yields the thicker edge μ_i' in the second glyph.

458

459 visual representations of the components $\Gamma(V) = X$ and the graphic $Q(\Gamma(V)) = Y$

Proposition 1. *If for elements of the monoid $m \in M$ and for all $\mu, \mu' \in X$, we define the monoid action on X so that it is by definition equivariant*

$$Q(\mu) = Q(\mu') \implies Q(m \circ \mu) = Q(m \circ \mu') \quad (23)$$

460 then a monoid action on Y can be defined as $m \circ \rho = \rho'$. If and only if Q satisfies
 461 *Equation 23*, we can state that the transformed graphic $\rho' = Q(m \circ \mu)$ is equivariant to a
 462 monoid action applied on Q with input $\mu \in Q^{-1}(\rho)$ that must generate valid ρ .

463 For example, given fiber $P = (xpos, ypos, color, thickness)$, then sections $\mu = (0, 0, 0, 1)$
 464 and $Q(\mu) = \rho$ generates a piece of the thin hollow circle. The action $m = (e, e, e, x + 2)$,
 465 where e is identity, translates μ to $\mu' = (e, e, e, 3)$ and the corresponding action on ρ causes
 466 $Q(\mu')$ to be the thicker circle in [Figure 17](#).

We formally describe a glyph as Q applied to the regions k that map back to a set of path connected components $J \subset K$ as input

$$J = \{j \in K \text{ s. t. } \exists \gamma \text{ s.t. } \gamma(0) = k \text{ and } \gamma(1) = j\} \quad (24)$$

where the path[81] γ from k to j is a continuous function from the interval $[0,1]$. We define the glyph as the graphic generated by $Q(S_j)$

$$H \xrightleftharpoons[\rho(S_j)]{} S_j \xrightleftharpoons[\xi^{-1}(J)]{\xi(s)} J_k \quad (25)$$

such that for every glyph there is at least one corresponding region on K , in keeping with the definition of glyph as any visually differentiable element put forth by Ziemkiewicz and Kosara[82]. The primitive point, line, and area marks[9, 83] are specially cased glyphs.

3.3.4 Assembly Q

Given the continuities described in 13, we illustrate a minimal Q that will generate the most minimal visualizations associated with those continuities: non-overlapping scatter points, a non-infinitely thin line, and an image.

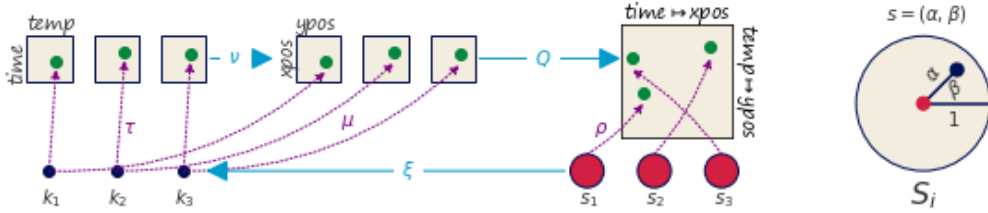


Figure 18: The data is discrete points (temperature, time). Via ν these are converted to (xpos, ypos) and pulled over discrete S . These values are then used to parameterize ρ which returns a color based on the parameters (xpos,ypos) and position α, β on S_k that ρ is evaluated on.

The scatter plot in Figure 18 can be defined as

$$Q(xpos, ypos)(\alpha, \beta) \quad (26)$$

with a constant $size$ and color $\rho_{RGB} = (0, 0, 0)$ that are defined as part of Q . The position of this swatch of color can be computed relative to the location on the disc $(\alpha, \beta) \in S_k$ as

shown in Figure 18

$$x = size * \alpha \cos(\beta) + xpos$$

$$y = size * \alpha \sin(\beta) + ypos$$

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

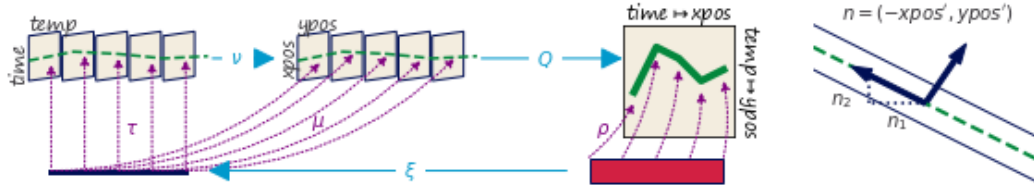


Figure 19: The line fiber $(time, temp)$ is thickened with the derivative $(time', temperature')$ because that information will be necessary to figure out the tangent to the point to draw a line. This is because the line needs to be pushed perpendicular to the tangent of $(xpos, ypos)$. The data is converted to visual characteristics $(xpos, ypos)$. The α coordinates on S specifies the position of the line, the β coordinate specifies thickness.

In contrast, the line plot

$$Q(xpos, \hat{n}_1, ypos, \hat{n}_2)(\alpha, \beta) \quad (27)$$

in ?? has a ξ function that is not only parameterized on k but also on the α distance along k and corresponding region in S . As shown in ??, 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; therefore the artist takes as input the jet bundle [84, 85] $\mathcal{J}^2(E)$ which is the data E and the first and second derivatives of E . The magnitude of the slope is $|n| = \sqrt{n_1^2 + n_2^2}$ such that the normal is $\hat{n}_1 = \frac{n_1}{|n|}$, $\hat{n}_2 = \frac{n_2}{|n|}$ which yields components of ρ

$$x = xpos(\xi(\alpha)) + width * \beta \hat{n}_1(\xi(\alpha))$$

$$y = ypos(\xi(\alpha)) + width * \beta \hat{n}_2(\xi(\alpha))$$

where (x,y) look up the position $\xi(\alpha)$ on the data and the derivatives \hat{n}_1, \hat{n}_2 . The derivatives are then multiplied by a *width* parameter to specify the thickness.

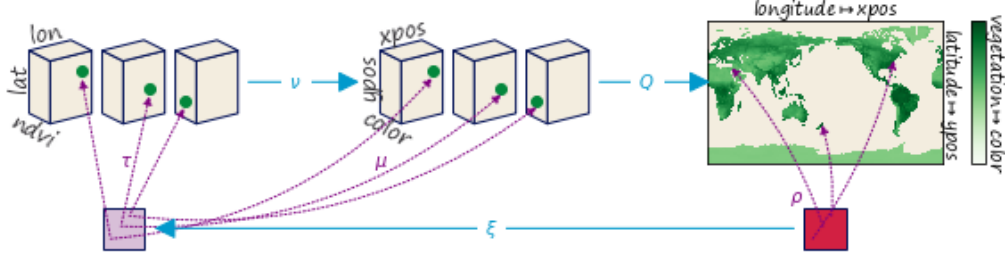


Figure 20: The only visual parameter an image requires is color since ξ encodes the mapping between position in data and position in graphic.

In Figure 20, the image

$$Q(xpos, ypos, color) \quad (28)$$

is a direct lookup into $\xi : S \rightarrow K$. The indexing variables (α, β) define the distance along the space, which is then used by ξ to map into K to lookup the color values

$$R = R(\xi(\alpha, \beta)), \quad G = G(\xi(\alpha, \beta)), \quad B = B(\xi(\alpha, \beta))$$

In the case of an image, the indexing mapper ξ may do some translating to a convention expected by Q , for example reorientng the array such that the first row in the data is at the bottom of the graphic.

The graphic base space S is not accessible in many architectures, including Matplotlib; instead we can construct a factory function \hat{Q} over K that can build a Q . As shown in Equation 18, Q is a bundle map $Q : \xi^*V \rightarrow H$ where ξ^*V and H are both bundles over S .

The preimage of the continuity map $\xi^{-1}(k) \subset S$ is such that many graphic continuity points $s \in S_K$ go to one data continuity point k ; therefore, by definition the pull back of μ

$$\xi^*V|_{\xi^{-1}(k)} = \xi^{-1}(k) \times P \quad (29)$$

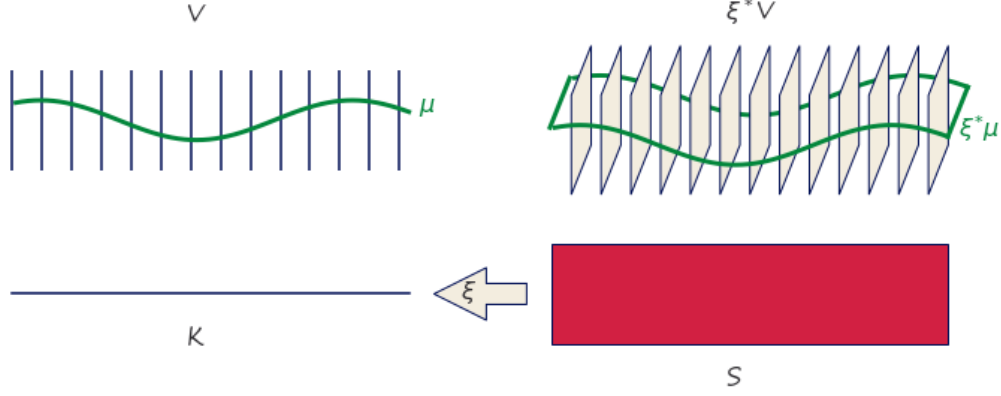


Figure 21: Because the pullback of the visual bundle ξ^*V is the replication of a μ over all points s that map back to a single k , we can construct a \hat{Q} on μ over k that will fabricate the Q for the equivalent region of s associated to that k

483 copies the visual fiber P over the the points s in graphic space S that correspond to one k
 484 in data space K . This set of points s are the preimage $\xi^{-1}(k)$ of k .

485 As shown in **Figure 21**, given the section $\xi^*\mu$ pulled back from μ and the point $s \in \xi^{-1}(k)$,
 486 there is a direct map $(k, \mu(k)) \mapsto (s, \xi^*\mu(s))$ from μ over k to the section $\xi^*\mu$ over s . This
 487 means that the pulled back section $\xi^*\mu(s) = \xi^*(\mu(k))$ is the section μ copied over all s such
 488 that $\xi^*\mu$ is identical for all s where $\xi(s) = k$. In **Figure 21** each dot on P is equivalent to
 489 the line on $P^*\mu$.

Given the equivalence between μ and $\xi^*\mu$ defined above, the reliance on S can be factored out. When Q maps visual sections into graphics $Q : \Gamma(\xi^*V) \rightarrow \Gamma(H)$, if we restrict Q input to $\xi^*\mu$ then the graphic section ρ evaluated on a visual region s

$$\rho(s) := Q(\xi^*\mu)(s) \quad (30)$$

is defined as the assembly function Q with input $\xi^*\mu$ evaluated on s . Since the pulled back section $\xi^*\mu$ is the section μ copied over every graphic region $s \in \xi^{-1}(k)$, we can define a Q factory function

$$\hat{Q}(\mu(k))(s) := Q((\xi^*\mu)(s)) \quad (31)$$

where \hat{Q} with input μ is defined to Q that takes as input the copied section $\xi^*\mu$ such that both functions are evaluated over the same location $\xi^{-1}(k) = s$ in the base space S . Factoring out s from Equation 31 yields

$$\hat{Q}(\mu(k)) = Q(\xi^*\mu) \quad (32)$$

where Q is no longer bound to input but \hat{Q} is still defined in terms of K . In fact, \hat{Q} is a map from visual space to graphic space $\hat{Q} : \Gamma(V) \rightarrow \Gamma(H)$ locally over k such that it can be evaluated on a single visual record $\hat{Q} : \Gamma(V_k) \rightarrow \Gamma(H \mid_{\xi^{-1}(k)})$. This allows us to construct a \hat{Q} that only depends on K , such that for each $\mu(k)$ there is part of $\rho \mid_{\xi^{-1}(k)}$. The construction of \hat{Q} allows us to retain the functional map reduce benefits of Q without having to majorly restructure the existing pipeline for libraries that delegate the construction of ρ to a backend such as Matplotlib.

3.3.5 Composition of Artists: +

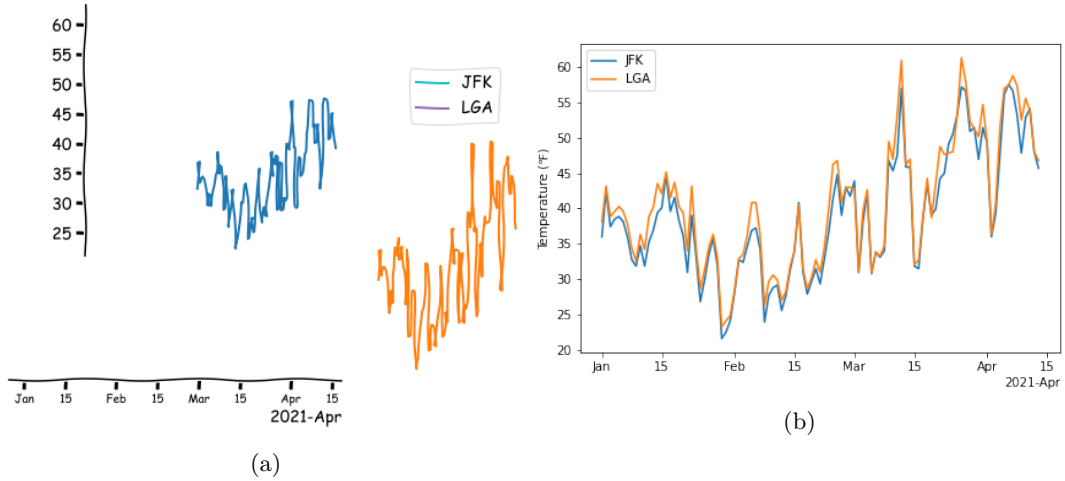


Figure 22: Each of the visual elements in ?? is generated via a unique artist A . In Figure 22a, they are added to the image independent of the other elements, creating an incoherent visualization. In Figure 22b, these artists are composited before being added to the image. Disjoint union of E aligns the two timeseries with the x and y axis so all these elements use a shared coordinate system. A more complex composition dictates that the legend is connected to the E such that it must use the same color as the data it is identifying.

We describe the constraints for compositing artists by defining addition operators. Given the family of artists $(E_i : i \in I)$ that are rendered to the same image, the $+$ operator

$$+ := \sqcup_{i \in I} E_i \tag{33}$$

defines a simple composition of artists. For example, the components in [Figure 22a](#) are each generated by different artists, and a visualization of solely the x axis is rarely all that useful. In [Figure 22a](#), these artists are all added to the image independently of the other and therefore there are no constraints on how they are generated in conjunction with each other. In [Figure 22b](#), the data is joined via disjoint union; doing so aligns the components in F such the ν to the same component in P targets the same coordinate system. When artists share a base space $K_2 \hookrightarrow K_1$, a composition operator can be defined such that the artists are acting on different components of the same section. This type of composition is important for visualizations where elements update together in a consistent way, such as multiple views [\[86, 87\]](#) and brush-linked views[\[88, 89\]](#).

3.3.6 Equivalence class of artists A'

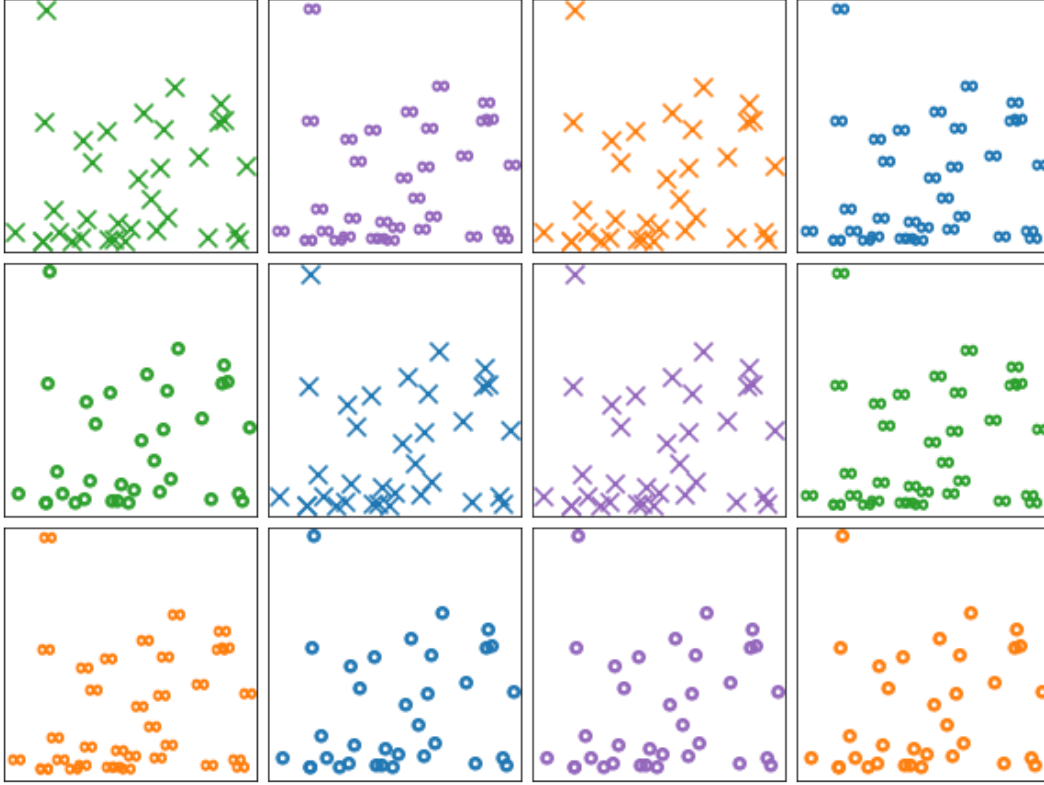


Figure 23: Each scatter plot is generated via a unique artist function A_i , but they only differ in aesthetic styling. Therefore, these artists are all members of an equivalence class $A_i \in A'$

Representational invariance, as defined by Kindlmann and Scheidegger, is the notion that visualizations are equivalent if changing the visual representation, such as colors or shapes, does not change the meaning of the visualization[12]. We propose that visualizations are invariant if they are generated by artists that are members of an equivalence class

$$\{A \in A' : A_1 \equiv A_2\}$$

For example, every scatter plot in Figure 23 is a scatter of the same datasets mapped to the x position and y position in the same way. The scatter plots only differ in the choice of constant visual literals, differing in color and marker shape. Each scatter is generated by an

512 artist A_i , and every scatter is generated by a member of the equivalence class $A_i \in A'$. Since
 513 it is impractical to implement a new artist for every single graphic, the equivalence class
 514 provides a way to evaluate an implementation of a generalized artist. Given equivalent, but
 515 no necessarily identical, ν , Q , and ξ , two artists are equivalent. This criteria also allows for
 516 comparing artists across libraries.

517 4 Prototype: Matplotlibtoy

518 To evaluate the feasibility of the model described in ??, we built prototypes of a **point**,
 519 **line**, and **bar** artist. We make use of the Matplotlib figure and axes artists [4, 5] so that
 520 we can initially focus on the data to graphic transformations and exploit the Matplotlib
 521 transform stack to convert data coordinates into screen coordinates. While the artist is
 522 specified in a fully functional manner in Equation 18, we implement the prototype in a
 523 heavily object oriented manner. We do so mostly to more easily manage function inputs,
 524 especially parameters that are passed through to methods that are structurally functional.

```

1 fig, ax = plt.subplots()
2 artist = Artist(E, V)
3 ax.add_artist(artist)

```

525 Building on the current Matplotlib artists which construct an internal represen-
 526 tation of the graphic, **ArtistClass** acts as an equivalence class artist A' as de-
 527 scribed in ??. The visual bundle V is specified as the **V** dictionary of the form
 528 **{parameter:(variable name, encoder)}** where parameter is a component in P , variable
 529 is a component in F , and the ν encoders are passed in as functions or callable objects. The
 530 data bundle E is passed in as a **E** object. By binding data and transforms to A' inside
 531 **__init__**, the **draw** method is a fully specified artist A as defined in Equation 15.

```

1 class ArtistClass(matplotlib.artist.Artist): #A'
2     def __init__(self, E, V, *args, **kwargs):
3         # properties that are specific to the graphic
4         self.E = E
5         self.V = V
6         super().__init__(*args, **kwargs)
7
8     def hat_Q(self, **args):
9         # set the properties of the graphic
10
11    def draw(self, renderer):
12        # returns K, indexed on fiber then key
13        tau = self.E.view(self.axes)
14        # visual channel encoding applied fiberwise
15        mu = {p: nu(tau(c))
16              for p, (c, nu) in self.V.items()}
17        self.hat_q(**mu)
18        # pass configurations off to the renderer
19        super().draw(renderer)

```

532 The data is fetched in section τ via a `view` method on the data because the input to the
533 artist is a section on E . The `view` method takes the `axes` attribute because it provides the
534 region in graphic coordinates S that can be used to query back into data to select a subset
535 as described in [subsubsection 3.1.5](#). To ensure the integrity of the section, `view` must be
536 atomic, which means that the values cannot change after the method is called in `draw` until
537 a new call to `draw`[27]. We put this constraint on the return of the `view` method so that we
538 do not risk race conditions.

539 The ν functions are then applied to the data, as describe in Equation 20, to generate
 540 the visual section μ that here is the object V . The conversion from data to visual space
 541 is simplified here to directly show that it is the encoding ν applied to the component.
 542 The `q_hat` function that is \hat{Q} , as defined in Equation 32, is responsible for generating a
 543 representation such that it could be serialized to recreate a static version of the graphic.
 544 This artist is not optimized because we prioritized demonstrating the separability of ν
 545 and \hat{Q} . The last step in the artist function is handing itself off to the renderer. The
 546 extra `*arg, **kwargs` arguments in `__init__`, `draw` are artifacts of how these objects are
 547 currently implemented.

548 4.1 Scatter, Line, and Bar Artists

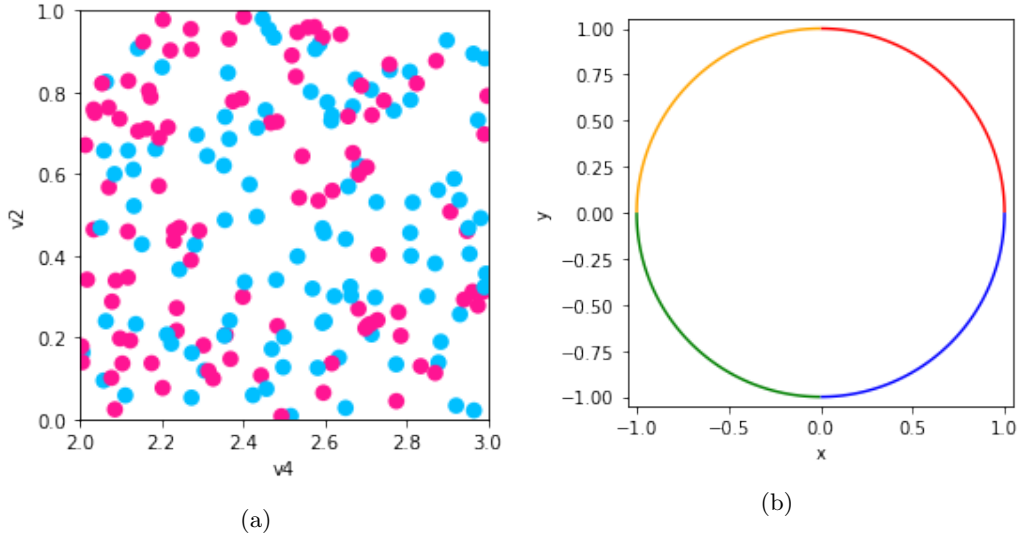


Figure 24: Scatter plot and line plot implemented using prototype artists and data models, building on Matplotlib rendering.

549 The figure in Figure 24a is described by Equation 26. This is implemented via a `Line` object
 550 where the scatter marker shape is fixed as a circle, and the visual fiber components are x
 551 and y position and the facecolor and size of the marker. We only show the `q_hat` function
 552 here because the `__init__`, `draw` are identical the prototype artist.

553 The `view` method repackages the data as a fiber component indexed table of vertices.
554 Even though the `view` is fiber indexed, each vertex at an index k has corresponding values
555 in section $\tau(k_i)$. This means that all the data on one vertex maps to one glyph.

```

1 class Point(ArtistClass, mcollections.Collection):
2     def q_hat(self, x, y, s, facecolors): #\hat{Q}
3         # construct geometries of circle glyphs
4         self._paths = [mpath.Path.circle((xi,yi), radius=si)
5                         for (xi, yi, si) in zip(x, y, s)]
6         # set attributes of glyphs, these are vectorized
7         # circles and facecolors are lists of the same size
8         self.set_facecolors(facecolors)

```

556 In `q_hat`, the μ components are used to construct the vector path of each circular marker
557 with center (x,y) and size x and set the colors of each circle. This is done via the
558 `Path.circle` object.

```

1 class Line(ArtistClass, mcollections.LineCollection):
2     def q_hat(self, x, y, color): #\hat{Q}
3         #assemble line marks as set of segments
4         segments = [np.vstack((vx, vy)).T for vx, vy
5                         in zip(x, y)]
6         self.set_segments(segments)
7         self.set_color(color)

```

559 To generate [Figure 24b](#), the `Line` artist `view` method returns a table of edges. Each edge
560 consists of (x,y) points sampled along the line defined by the edge and information such as
561 the color of the edge. As with `Point`, the data is then converted into visual variables. In
562 `q_hat`, described by [Equation 27](#), this visual representation is composed into a set of line

563 segments, where each segment is the array generated by `np.vstack((vx, vy))`. Then the
 564 colors of each line segment are set. The colors are guaranteed to correspond to the correct
 565 segment because of the atomicity constraint on the view. The current implementation of
 566 line in Matplotlib does not have this functionality because it has no notion of an atomic
 567 view τ , instead the user keeps track of which values are input into a single call of \hat{Q} by
 568 creating a loop where every color is a new call to the line plotting function.

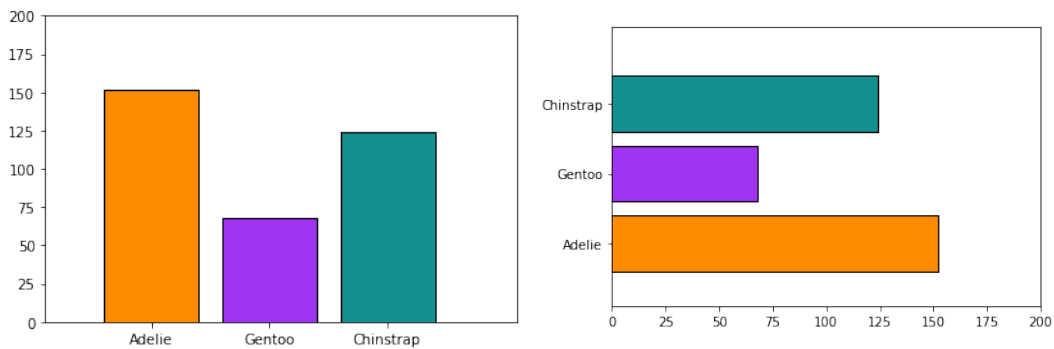


Figure 25: Frequency of Penguin types visualized as discrete bars.

569 The bar charts in figure 25 are generated with a **Bar** artist. The artist has required
 570 visual parameters P of (position, length), and an additional parameter **orientation** which
 571 controls whether the bars are arranged vertically or horizontally. This parameter only applies
 572 holistically to the graphic and never to individual data parameters, and highlights how the
 573 model encourages explicit differentiation between parameters in V and graphic parameters
 574 applied directly to \hat{Q} .

```

1 class Bar(ArtistClass, mcollections.Collection):
2     def __init__(self, data, transforms, orientation='v', *args, **kwargs):
3         """
4         orientation: str, optional
5             v: bars aligned along x axis, heights on y
6             h: bars aligned along y axis, heights on x
7         """
  
```



```

8         self.orientation = orientation
9         # set E and V
10
11     def q_hat(self, position, length, floor=0, width=0.8,
12               facecolors='C0', edgecolors='k', offset=0):
13         #set some defaults
14         width = itertools.repeat(width) if np.isscalar(width) else width
15         floor = itertools.repeat(floor) if np.isscalar(floor) else (floor)
16
17         # offset is passed through via assemblers such as multigroup,
18         # not supposed to be directly tagged to position
19         position = position + offset
20
21     def makeBars(xval, xoff, yval, yoff):
22         return [[(x, y), (x, y+yo), (x+xo, y+yo), (x+xo, y), (x, y)]
23                 for (x, xo, y, yo) in zip(xval, xoff, yval, yoff)]
24         #build bar glyphs based on graphic parameter
25         if self.orientation in {'vertical', 'v'}:
26             verts = makeBars(position, width, floor, length)
27         elif self.orientation in {'horizontal', 'h'}:
28             verts = makeBars(floor, length, position, width)
29
30         self._paths = [mpath.Path(xy, closed=True) for xy in verts]
31         self.set_edgecolors(edgecolors)
32         self.set_facecolors(facecolors)

```

575 As with Point and scatter, q_hat function constructs bars and sets their properties, face
576 and edge colors. The makeBars function converts the input position and length to the
577 coordinates of a rectangle of the given width. Defaults are provided for 'width' and 'floor'

578 to make this function more reusable. Typically the defaults are used for the type of chart
 579 shown in figure 25, but these visual variables are often set when building composite versions
 580 of this chart type as discussed in section 4.4.

581 4.2 Visual Encoders

582 The visual parameter serves as the dictionary key because the visual representation is con-
 583 structed from the encoding applied to the data $\mu = \nu \circ \tau$. For the scatter plot, the mappings
 584 for the visual fiber components $P = (x, y, facecolors, s)$ are defined as

```

1 cmap = color.Categorical({'true': 'deeppink',
2                           'false': 'deepskyblue'})
3 # {P_i name: {'name': c_i, 'encoder': \nu_i}}
4 V = {'x': {'name': 'v4', 'encoder': lambda x: x},
5      'y': {'name': 'v2', 'encoder': lambda x: x},
6      'facecolors': {'name': 'v3', 'encoder': cmap},
7      's': {'name': None,
8            'encoder': lambda _: itertools.repeat(.02)}}

```

585 where `lambda x: x` is an identity ν , `{'name': None}` maps into P without corresponding
 586 τ to set a constant visual value, and `color.Categorical` is a custom ν implemented as a
 587 class for reusability.

```

1 # \nu_i(m_r(E_i)) = \varphi(m_r)(\nu_i(E_i))
2 def test_nominal(values, encoder):
3     m1 = list(zip(values, encoder(values)))
4     random.shuffle(values)
5     m2 = list(zip(values, encoder(values)))
6     assert sorted(m1) == sorted(m2)

```

588 As described in Equation 22, a test for equivariance can be implemented trivially. It is
589 currently factored out of the artist for clarity.

590 4.3 Data Model

591 The data input into the `Artist` will often be a wrapper class around an existing data
592 structure. This wrapper object must specify the fiber components F and connectivity K
593 and have a `view` method that returns an atomic object that encapsulates τ . To support
594 specifying the fiber bundle, we define a `FiberBundle` data class[90]

```
1 @dataclass
2 class FiberBundle:
3     K: dict #{'tables': []}
4     F: dict # {variable name: type}
```

595 that asks the user to specify the the properties of F and the K connectivity as either discrete
596 vertices or continuous data along edges. To generate the scatter plot and the line plot, the
597 distinction is in the `tau` method that is the section.

```
1 class PointData:
2     def __init__(self):
3         self.FB = FiberBundle({'tables': ['vertex']},
4                                {'v1': float, 'v2': str, 'v3': float})
5     def tau(self, k):
6         return # tau evaluated at one point k
7
8 class LineData:
9     def __init__(self):
10        self.FB = FiberBundle({'tables': ['edge']},
```

```

11         {'x': float, 'y': float, 'color': str})
12     def tau(self, k):
13         return # tau evaluated on interval k

```

598 The discrete `tau` method returns a record of discrete points, akin to a row in a table, while
599 the `linetau` returns a sampling of points along an edge k .

```

1 def view(self, axes):
2     table = defaultdict(list)
3     for k in self.keys:
4         table['index'].append(k)
5         for (name, value) in zip(self.FB.fiber.keys(),
6                                   self.tau(k)[1]):
7             table[name].append(value)
8     return table

```

600 In both cases, the `view` method packages the data into a data structure that the artist can
601 unpack via data component name, akin to a table with column names when K is 0 or 1 D.

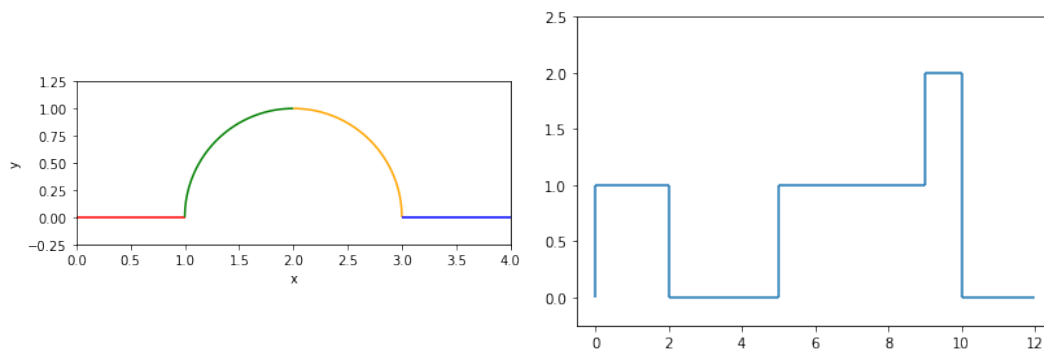


Figure 26: Continuous and discontinuous lines as defined via the same data model, and generated with the same `A'Line`

602 The graphics in figure [Figure 26](#) are made using the `Line` artist and the `GraphData`
603 data source where if told that the data is connected, the data source will check for that

connectivity by constructing an adjacency matrix. The multicolored line is a connected graph of edges with each edge function evaluated on 100 samples,

```
1 GraphData(FB, edges, verticies, num_samples=100, connect=True)
```

which is an arbitrary choice made to display a smooth curve. The axes can also be used to choose an appropriate number of samples. In contrast, the stair chart is discontinuous and only needs to be evaluated at the edges of the interval

```
1 GraphData(FB, edges, verticies, num_samples=2, connect=False)
```

such that one advantage of this model is it helps differentiate graphics that have different artists from graphics that have the same artist but make different assumptions about the source data.

4.4 Case Study: Penguins

Building on the `Bar` artist in [subsection 4.1](#), we implement grouped bar charts as these do not exist out of the box in the current version of Matplotlib. Instead, grouped bar charts are often achieved via looping over an implementation of `bar`, which forces the user to keep track which values are mapped to a single visual element and how that is achieved. For this case study, we use the Palmer Penguins dataset[91, 92], packaged as a pandas dataframe[93] since that is a very commonly used Python labeled data structure. The wrapper is very thin because there is explicitly only one section.

```
1 class DataFrame:
2     def __init__(self, dataframe):
3         self.FB = FiberBundle(K = {'tables': ['vertex']},
4                                 F = dict(dataframe.dtypes))
```

```

5         self._tau = dataframe.iloc
6         self._view = dataframe
7
8     def view(self, axes=None):
9         return self._view

```

620 Since the aim for this wrapper is to be very generic, here the fiber is set by querying the
 621 dataframe for its metadata. The **dtypes** are a list of column names and the datatype of
 622 the values in each column; this is the minimal amount of information the model requires to
 623 verify constraints. The pandas indexer is a key valued set of discrete vertices, so there is no
 624 need to repackage for the data interface.

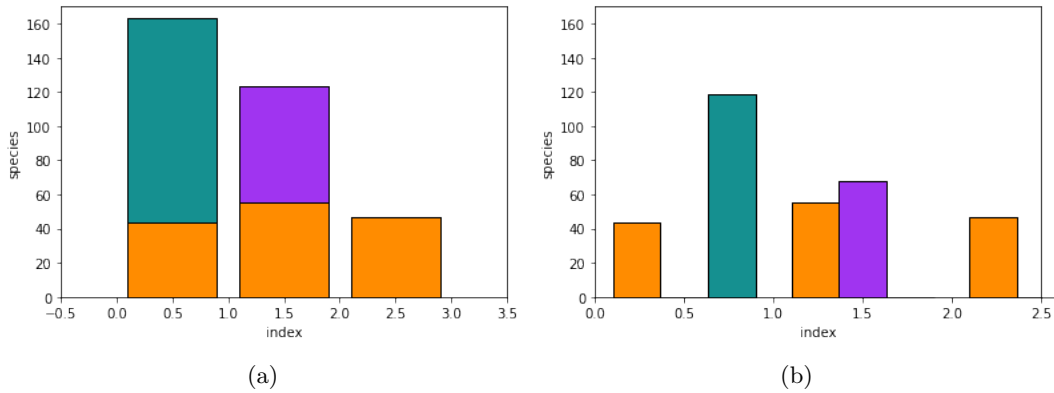


Figure 27: Penguin count disaggregated by island and species

625 The stacked and grouped bar charts in figure 27 are both composites of **Bar** artists such
 626 that the difference between **StackedBar** and **GroupedBar** is specific to the ways in which the
 627 **Bar** are stitched together. These two artists have identical constructors and **draw** methods.
 628 As with **Bar**, the orientation is set in the constructor. In both these artists, we separate
 629 the transforms that are applied to only one component (column) from transforms applied
 630 to multiple components (columns). We encode the one component case as **V** dictionary, and
 631 the multiple components case as **MV**. This convention allows us to, for example, map one
 632 column to position, but multiple to length.

```

1 class StackedBar(martist.Artist):
2     def __init__(self, E, V, MV, orientation='v', *args, **kwargs):
3         """
4         Parameters
5         -----
6
7         orientation: str, optional
8             vertical: bars aligned along x axis, heights on y
9             horizontal: bars aligned along y axis, heights on x
10        """
11        super().__init__(*args, **kwargs)
12        self.E = E
13        self.orientation = orientation
14        self.V = V
15        self.MV = MV
16
17    def assemble(self):
18        tau = self.data.view(self.axes)
19        self.children = [] # list of bars to be rendered
20        floor = 0
21        for group in self.MV:
22            # pull out the specific group transforms
23            bar = Bar(self.E, {**group, **self.V, 'floor':floor},
24                    self.orientation, transform=self.axes.transData)
25            self.children.append(bar)
26            floor += view[group['length']]['name']]
27
28
```

```

29     def draw(self, renderer, *args, **kwargs):
30         # all the visual conversion gets pushed to child artists
31         self.assemble()
32         #self._transform = self.children[0].get_transform()
33         for artist in self.children:
34             artist.draw(renderer, *args, **kwargs)

```

Since all the visual transformation is passed through to Bar, the draw method does not do any visual transformations. In StackedBar the view is used to adjust the floor for every subsequent bar chart, since a stacked bar chart is bar chart area marks concatenated together in the length parameter. In contrast, GroupedBar does not even need the view, but instead keeps track of the relative position of each group of bars in the visual only variable offset.

```

1  class GroupedBar(martist.Artist):
2      def assemble(self):
3          self.children = [] # list of bars to be rendered
4          ngroups = len(self.mtransforms)
5
6          for gid, group in enumerate(self.mtransforms):
7              group.update(self.transforms)
8              width = group.get('width', .8)
9              gwidth = width/ngroups
10             offset = gid/ngroups*width
11             bar = Bar(self.E, {**group, **self.V, 'width':gwidth, 'offset':offset},
12                       self.orientation, transform=self.axes.transData)
13             self.children.append(bar)

```

639 Since the only difference between these two glyphs is in the composition of `Bar`, they take
 640 in the exact same transform specification dictionaries. The `transform` dictionary dictates
 641 the position of the group, in this case by island the penguins are found on.

```

1 transforms = {'position': {'name': 'island',
2                       'encoder': position.Nominal({'Biscoe':0.1, 'Dream':1.1, 'Torgersen':2.1})}}
3 group_transforms = [{'length': {'name': 'Adelie'},
4                       'facecolors': {'name': "Adelie_s", 'encoder': cmap}},
5                      {'length': {'name': 'Chinstrap'},
6                      'facecolors': {'name': "Chinstrap_s", 'encoder': cmap}},
7                      {'length': {'name': 'Gentoo'},
8                      'facecolors': {'name': "Gentoo_s", 'encoder': cmap}}]

```

642 `group_transforms` describes the group, and takes a list of dictionaries where each dictionary
 643 is the aesthetics of each group. That `position` and `length` are required parameters is
 644 enforced in the creation of the `Bar` artist. These means that these two artists have identical
 645 function signatures

```

1 artistSB = bar.StackedBar(bt, ts, group_transforms)
2 artistGB = bar.GroupedBar(bt, ts, group_transforms)

```

646 but differ in assembly \hat{Q} . By decomposing the architecture into data, visual encoding,
 647 and assembly steps, we are able to build components that are more flexible and also more self
 648 contained than the existing code base. While very rough, this API demonstrates that the
 649 ideas presented in the math framework are implementable. For example, the `draw` function
 650 that maps most closely to A is functional, with state only being necessary for bookkeeping
 651 the many inputs that the function requires. In choosing a functional approach, if not
 652 implementation, we provide a framework for library developers to build reusable encoder

653 ν assembly \hat{Q} and artists A . We argue that if these functions are built such that they
654 are equivariant with respect to monoid actions and the graphic topology is a deformation
655 retraction of the data topology, then the artist by definition will be a structure and property
656 preserving map from data to graphic.

657 5 Discussion

658 This work contributes a functional model of the structure-preserving maps from data to
659 visual representation to guide the development of visualization libraries, thereby providing
660 a means to express the constraints that visualization must preserve continuity and faithfully
661 translate the properties of the data variables into visual variables. Combining Butler’s
662 proposal of a fiber bundle model of visualization data with Spivak’s formalism of schema
663 lets this model support a variety of datasets, including discrete relational tables, multivariate
664 high resolution spatio temporal datasets, and complex networks. By decomposing the artist
665 into encoding ν , assembly Q , and reindexing ξ maps, the model expresses the specifications
666 that graphic and data must have equivalent *continuity* equivalent to the data, and that
667 the visual characteristics of the graphics are *equivariant* to their corresponding components
668 under monoid actions. TEAM defines these constraints on the transformation function
669 such that they are not specific to any one type of encoding or visual characteristic. The toy
670 prototype built using this model validates that it is usable for a general purpose visualization
671 tool since it can be iteratively integrated into existing architecture rather than starting from
672 scratch. Factoring out graphic formation into assembly functions allows for much more
673 clarity in how they differ. This prototype demonstrates that this framework can generate
674 the fundamental point (scatter plot) and line (line chart) marks.

675 5.1 Limitations

676 Our model and prototype are deeply tied to Matplotlib’s existing architecture, so it has
677 not yet been worked through how the model generalizes to libraries such as VTK or D3.
678 Even though the model is designed to be backend and format independent, it has only been

679 tested using PNGs rendered with AGG[76]. It is unknown how this framework interfaces
680 with high performance rendering libraries such as OpenGL[74] that implement different
681 models of ρ . While our model supports equivariance of figurative glyphs [8] generated
682 from data components[94, 95], it cannot evaluate the semantic accuracy of the figurative
683 representation. Effectiveness criteria[11, 96] are out of scope.

684 5.2 Future Work

685 More work is needed to formalize the composition operators, equivalence class A' , and
686 the mathematical model of interactivity. We also need to implement artists that demon-
687 strate that the model can underpin a minimally viable library, foremost an image[97, 98],
688 a heatmap[99, 100], and an inherently computational artist such as a boxplot[101]. In
689 summary, the proposed scope of work is

work period	milestones & tasks
April - July 2021	<p>prepare and submit conference presentation on new functionality enabled by model for <i>SciPy</i>:</p> <p>artists that do not inherit from existing Matplotlib artists, computational artists such as histograms, non tabular data, composite interactive artist</p>
June - Sept 2021	<p>prepare and submit theory paper on interactivity to <i>TCVG or Eurovis 2022</i>:</p> <p>fully work out and describe math of addition operators and lookups from graphic to data space, implement brush linked artist (shared base space) and artist that exploits sheafs</p>
May - Nov 2021	<p>prepare and submit applications paper on high dimensional to <i>TCVG</i>:</p> <p>math and implementation of computational artists, concurrent artists and data sources, non-trivial data bundles</p>
Aug 2021 - Feb 2022	<p>prepare and submit systems paper on building domain specific libraries based on this model to <i>Infoviz 2022</i>:</p> <p>domain specific structured data, composite artists, inference of meta data components, mathematical notion of a visualization (labeled, multiple artists, etc)</p>
March 2022	<p>dissertation writing:</p> <p>synthesize previous work on climate data, compile topological equivariant artist model work</p>
April 2022	<p>defense 56</p>

Table 3

690 In acknowledgement that the schedule is optimistic, the data applications of this work
 691 could instead be revisited after the defense as that work is not integral to visualization li-
 692 brary architecture. The data applications could be further integrated with topological[102]
 693 and functional [103] data analysis methods. Since this model formalizes notions of structure
 694 preservation, it can serve as a good base for tools that assess quality metrics[104] or in-
 695 variance [12] of visualizations with respect to graphical encoding choices. While this paper
 696 formulates visualization in terms of monoidal action homomorphisms between fiberbundles,
 697 the model lends itself to a categorical formulation[54, 105] that could be further explored.

698 6 Conclusion

699 A TEAM driven refactor of visualizations libraries could produce more maintainable,
 700 reusable, and extensible code, leading to better building blocks for the ecosystem of tools
 701 built on top of TEAM architected libraries. Building block libraries could better support
 702 downstream, including domain specific, libraries without having to explicitly incorporate
 703 the specific data structure and visualization needs of those domains back into the base
 704 library. Adopting this model would induce a separation of data representation and visual
 705 representation that, for example, in Matplotlib is so entangled that it has lead to a brittle
 706 and sometimes incoherent API and internal code base. A refactor that incorporated
 707 the generalized data model and functional transforms presented in TEAM would lead to
 708 building block libraries that provide a more consistent, reusable, flexible, collection of
 709 blocks.

710 References

- 711 [1] Krist Wongsuphasawat. *Navigating the Wide World of Data Visualization Libraries*
 712 *(on the Web)*. 2021.
- 713 [2] J. Hughes. “Why Functional Programming Matters”. In: *The Computer Journal* 32.2
 714 (Jan. 1989), pp. 98–107. ISSN: 0010-4620. DOI: [10.1093/comjnl/32.2.98](https://doi.org/10.1093/comjnl/32.2.98).

- [3] Zhenjiang Hu, John Hughes, and Meng Wang. “How Functional Programming Mat-
tered”. In: *National Science Review* 2.3 (Sept. 2015), pp. 349–370. ISSN: 2095-5138.
DOI: [10.1093/nsr/nwv042](https://doi.org/10.1093/nsr/nwv042).
- [4] J. D. Hunter. “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science
Engineering* 9.3 (May 2007), pp. 90–95. ISSN: 1558-366X. DOI: [10.1109/MCSE.2007.
55](https://doi.org/10.1109/MCSE.2007.55).
- [5] John Hunter and Michael Droettboom. *The Architecture of Open Source Applications
(Volume 2): Matplotlib*. <https://www.aosabook.org/en/matplotlib.html>.
- [6] A. Sarikaya et al. “What Do We Talk About When We Talk About Dashboards?”
In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019),
pp. 682–692. ISSN: 1941-0506. DOI: [10.1109/TVCG.2018.2864903](https://doi.org/10.1109/TVCG.2018.2864903).
- [7] Michael Friendly. “A Brief History of Data Visualization”. en. In: *Handbook of Data
Visualization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 15–56. ISBN:
978-3-540-33036-3 978-3-540-33037-0. DOI: [10.1007/978-3-540-33037-0_2](https://doi.org/10.1007/978-3-540-33037-0_2).
- [8] L. Byrne, D. Angus, and J. Wiles. “Acquired Codes of Meaning in Data Visualization
and Infographics: Beyond Perceptual Primitives”. In: *IEEE Transactions on Visual-
ization and Computer Graphics* 22.1 (Jan. 2016), pp. 509–518. ISSN: 1077-2626. DOI:
[10.1109/TVCG.2015.2467321](https://doi.org/10.1109/TVCG.2015.2467321).
- [9] Jacques Bertin. *Semiology of Graphics : Diagrams, Networks, Maps*. English. Red-
lands, Calif.: ESRI Press, 2011. ISBN: 978-1-58948-261-6 1-58948-261-1.
- [10] Jock Mackinlay. “Automating the Design of Graphical Presentations of Relational
Information”. In: *ACM Transactions on Graphics* 5.2 (Apr. 1986), pp. 110–141. ISSN:
0730-0301. DOI: [10.1145/22949.22950](https://doi.org/10.1145/22949.22950).
- [11] Jock Mackinlay. “Automatic Design of Graphical Presentations”. English. PhD The-
sis. Stanford, 1987.
- [12] 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. ISSN: 1941-0506. DOI: [10.1109/TVCG.2014.2346325](https://doi.org/10.1109/TVCG.2014.2346325).

- [13] Ricky Shadrach. *Introduction to Groups*. <https://www.mathsisfun.com/sets/groups-introduction.html>. 2017.
- [14] *Naturalness Principle - InfoVis:Wiki*. https://infovis-wiki.net/wiki/Naturalness_Principle.
- [15] Edward R. Tufte. *The Visual Display of Quantitative Information*. English. Cheshire, Conn.: Graphics Press, 2001. ISBN: 0-9613921-4-2 978-0-9613921-4-7 978-1-930824-13-3 1-930824-13-0.
- [16] 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. DOI: [10.1109/TVCG.2006.178](https://doi.org/10.1109/TVCG.2006.178).
- [17] E. H. Chi. “A Taxonomy of Visualization Techniques Using the Data State Reference Model”. In: *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*. Oct. 2000, pp. 69–75. DOI: [10.1109/INFVIS.2000.885092](https://doi.org/10.1109/INFVIS.2000.885092).
- [18] Jeffrey Heer and Michael Bostock. “Declarative Language Design for Interactive Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (Nov. 2010), pp. 1149–1156. ISSN: 1077-2626. DOI: [10.1109/TVCG.2010.144](https://doi.org/10.1109/TVCG.2010.144).
- [19] 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. ISSN: 1941-0506. DOI: [10.1109/2945.981851](https://doi.org/10.1109/2945.981851).
- [20] Pat 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. ISBN: 1-59593-434-0. DOI: [10.1145/1142473.1142560](https://doi.org/10.1145/1142473.1142560).
- [21] 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. ISSN: 1941-0506. DOI: [10.1109/TVCG.2007.70594](https://doi.org/10.1109/TVCG.2007.70594).
- [22] Leland Wilkinson. *The Grammar of Graphics*. en. 2nd ed. Statistics and Computing. New York: Springer-Verlag New York, Inc., 2005. ISBN: 978-0-387-24544-7.

- [23] Hadley Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4.
- [24] M. Bostock and J. Heer. “Protovis: A Graphical Toolkit for Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (Nov. 2009), pp. 1121–1128. ISSN: 1941-0506. DOI: [10.1109/TVCG.2009.174](https://doi.org/10.1109/TVCG.2009.174).
- [25] Arvind Satyanarayan, Kanit Wongsuphasawat, and Jeffrey Heer. “Declarative Interaction Design for Data Visualization”. en. In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. Honolulu Hawaii USA: ACM, Oct. 2014, pp. 669–678. ISBN: 978-1-4503-3069-5. DOI: [10.1145/2642918.2647360](https://doi.org/10.1145/2642918.2647360).
- [26] Jacob VanderPlas et al. “Altair: Interactive Statistical Visualizations for Python”. en. In: *Journal of Open Source Software* 3.32 (Dec. 2018), p. 1057. ISSN: 2475-9066. DOI: [10.21105/joss.01057](https://doi.org/10.21105/joss.01057).
- [27] Jeffrey D. Ullman and Jennifer. Widom. *A First Course in Database Systems*. English. Upper Saddle River, NJ: Pearson Prentice Hall, 2008. ISBN: 0-13-600637-X 978-0-13-600637-4.
- [28] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. “NIH Image to ImageJ: 25 Years of Image Analysis”. In: *Nature Methods* 9.7 (July 2012), pp. 671–675. ISSN: 1548-7105. DOI: [10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089).
- [29] Nicholas Sofroniew et al. *Napari/Napari: 0.4.5rc1*. Zenodo. Feb. 2021. DOI: [10.5281/zenodo.4533308](https://doi.org/10.5281/zenodo.4533308).
- [30] *Writing Plugins*. en. https://imagej.net/Writing_plugins.
- [31] Software Studies. *Culturevis/Imageplot*. Jan. 2021.
- [32] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. “Gephi: An Open Source Software for Exploring and Manipulating Networks”. en. In: *Proceedings of the International AAAI Conference on Web and Social Media* 3.1 (Mar. 2009). ISSN: 2334-0770.

- 797 [33] John Ellson et al. “Graphviz— Open Source Graph Drawing Tools”. In: *Graph Draw-*
798 *ing*. Ed. by Petra Mutzel, Michael Jünger, and Sebastian Leipert. Berlin, Heidelberg:
799 Springer Berlin Heidelberg, 2002, pp. 483–484. ISBN: 978-3-540-45848-7.
- 800 [34] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. “Exploring Network Struc-
801 ture, Dynamics, and Function Using NetworkX”. In: *Proceedings of the 7th Python*
802 *in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman.
803 Pasadena, CA USA, 2008, pp. 11–15.
- 804 [35] *Data Representation in Mayavi — Mayavi 4.7.2 Documentation*. <https://docs.entthought.com/mayavi/mayavi/d>
- 805 [36] M. Bostock, V. Ogievetsky, and J. Heer. “D³ Data-Driven Documents”. In: *IEEE*
806 *Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2301–
807 2309. ISSN: 1941-0506. DOI: [10.1109/TVCG.2011.185](https://doi.org/10.1109/TVCG.2011.185).
- 808 [37] Marcus D. Hanwell et al. “The Visualization Toolkit (VTK): Rewriting the Rendering
809 Code for Modern Graphics Cards”. en. In: *SoftwareX* 1-2 (Sept. 2015), pp. 9–12. ISSN:
810 23527110. DOI: [10.1016/j.softx.2015.04.001](https://doi.org/10.1016/j.softx.2015.04.001).
- 811 [38] P. Ramachandran and G. Varoquaux. “Mayavi: 3D Visualization of Scientific Data”.
812 In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 40–51. ISSN: 1558-366X.
813 DOI: [10.1109/MCSE.2011.35](https://doi.org/10.1109/MCSE.2011.35).
- 814 [39] James Ahrens, Berk Geveci, and Charles Law. “Paraview: An End-User Tool for
815 Large Data Visualization”. In: *The visualization handbook* 717.8 (2005).
- 816 [40] Brian Wylie and Jeffrey Baumes. “A Unified Toolkit for Information and Scientific
817 Visualization”. In: *Proc.SPIE*. Vol. 7243. Jan. 2009. DOI: [10.1117/12.805589](https://doi.org/10.1117/12.805589).
- 818 [41] Stephen Few and Perceptual Edge. “Dashboard Confusion Revisited”. In: *Perceptual*
819 *Edge* (2007), pp. 1–6.
- 820 [42] D. M. Butler and M. H. Pendley. “A Visualization Model Based on the Mathematics
821 of Fiber Bundles”. en. In: *Computers in Physics* 3.5 (1989), p. 45. ISSN: 08941866.
822 DOI: [10.1063/1.168345](https://doi.org/10.1063/1.168345).

- [43] David M. Butler and Steve Bryson. “Vector-Bundle Classes Form Powerful Tool for Scientific Visualization”. en. In: *Computers in Physics* 6.6 (1992), p. 576. ISSN: 08941866. DOI: [10.1063/1.4823118](https://doi.org/10.1063/1.4823118).
- [44] David I Spivak. *Databases Are Categories*. en. Slides. June 2010.
- [45] David I Spivak. “SIMPLICIAL DATABASES”. en. In: (), p. 35.
- [46] Tamara Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC press, Oct. 2014. ISBN: 978-1-4665-0891-0.
- [47] Tamara Munzner. “Ch 2: Data Abstraction”. In: *CPSC547: Information Visualization, Fall 2015-2016* ().
- [48] E.H. Spanier. *Algebraic Topology*. McGraw-Hill Series in Higher Mathematics. Springer, 1989. ISBN: 978-0-387-94426-5.
- [49] *Locally Trivial Fibre Bundle - Encyclopedia of Mathematics*. https://encyclopediaofmath.org/wiki/Locally_trivial_fibre_bundle
- [50] “Monoid”. en. In: *Wikipedia* (Jan. 2021).
- [51] “Semigroup Action”. en. In: *Wikipedia* (Jan. 2021).
- [52] nLab authors. “Action”. In: (Mar. 2021).
- [53] Eric W. Weisstein. *Group Homomorphism*. en. <https://mathworld.wolfram.com/GroupHomomorphism.html>. Text.
- [54] Brendan Fong and David I. Spivak. *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. en. First. Cambridge University Press, July 2019. ISBN: 978-1-108-66880-4 978-1-108-48229-5 978-1-108-71182-1. DOI: [10.1017/9781108668804](https://doi.org/10.1017/9781108668804).
- [55] S. S. Stevens. “On the Theory of Scales of Measurement”. In: *Science* 103.2684 (1946), pp. 677–680. ISSN: 00368075, 10959203.
- [56] W A Lea. “A Formalization of Measurement Scale Forms”. en. In: (), p. 44.
- [57] Brent A Yorgey. “Monoids: Theme and Variations (Functional Pearl)”. en. In: (), p. 12.

- [58] “Quotient Space (Topology)”. en. In: *Wikipedia* (Nov. 2020).
- [59] Professor Denis Auroux. “Math 131: Introduction to Topology”. en. In: (), p. 113.
- [60] Robert W. Ghrist. *Elementary Applied Topology*. Vol. 1. Createspace Seattle, 2014.
- [61] Robert Ghrist. “Homological Algebra and Data”. In: *Math. Data* 25 (2018), p. 273.
- [62] David Urbanik. “A Brief Introduction to Schemes and Sheaves”. en. In: (), p. 16.
- [63] Dmitry Nekrasovski et al. “An Evaluation of Pan & Zoom and Rubber Sheet Navigation with and without an Overview”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 11–20. ISBN: 1-59593-372-7. DOI: [10.1145/1124772.1124775](https://doi.org/10.1145/1124772.1124775).
- [64] Michael S. Crouch, Andrew McGregor, and Daniel Stubbs. “Dynamic Graphs in the Sliding-Window Model”. In: *European Symposium on Algorithms*. Springer, 2013, pp. 337–348.
- [65] Chia-Shang James Chu. “Time Series Segmentation: A Sliding Window Approach”. In: *Information Sciences* 85.1 (July 1995), pp. 147–173. ISSN: 0020-0255. DOI: [10.1016/0020-0255\(95\)00021-G](https://doi.org/10.1016/0020-0255(95)00021-G).
- [66] Charles R Harris et al. “Array Programming with NumPy”. In: *Nature* 585.7825 (2020), pp. 357–362.
- [67] Jeff Reback et al. *Pandas-Dev/Pandas: Pandas 1.0.3*. Zenodo. Mar. 2020. DOI: [10.5281/zenodo.3715232](https://doi.org/10.5281/zenodo.3715232).
- [68] Stephan Hoyer and Joe Hamman. “Xarray: ND Labeled Arrays and Datasets in Python”. In: *Journal of Open Research Software* 5.1 (2017).
- [69] Matthew Rocklin. “Dask: Parallel Computation with Blocked Algorithms and Task Scheduling”. In: *Proceedings of the 14th Python in Science Conference*. Vol. 126. Citeseer, 2015.
- [70] “Retraction (Topology)”. en. In: *Wikipedia* (July 2020).

- [71] Eric W. Weisstein. *Homotopy*. en. <https://mathworld.wolfram.com/Homotopy.html>. Text.
- [72] Tim Bienz, Richard Cohn, and Calif.) Adobe Systems (Mountain View. *Portable Document Format Reference Manual*. Citeseer, 1993.
- [73] A. Quint. “Scalable Vector Graphics”. In: *IEEE MultiMedia* 10.3 (July 2003), pp. 99–102. ISSN: 1941-0166. DOI: [10.1109/MMUL.2003.1218261](https://doi.org/10.1109/MMUL.2003.1218261).
- [74] George S. Carson. “Standards Pipeline: The OpenGL Specification”. In: *SIGGRAPH Comput. Graph.* 31.2 (May 1997), pp. 17–18. ISSN: 0097-8930. DOI: [10.1145/271283.271292](https://doi.org/10.1145/271283.271292).
- [75] *Cairographics.Org*. <https://www.cairographics.org/>.
- [76] Maxim Shemanarev. *Anti-Grain Geometry*. <https://antigrain.com/>.
- [77] A. M. Cegarra. “Cohomology of Monoids with Operators”. In: *Semigroup Forum*. Vol. 99. Springer, 2019, pp. 67–105. ISBN: 1432-2137.
- [78] Christian Remling. *Algebra (Math 5353/5363) Lecture Notes*. Lecture Notes. University of Oklahoma.
- [79] H. M. Johnson. “Pseudo-Mathematics in the Mental and Social Sciences”. In: *The American Journal of Psychology* 48.2 (1936), pp. 342–351. ISSN: 00029556. DOI: [10.2307/1415754](https://doi.org/10.2307/1415754).
- [80] M. A. Thomas. *Mathematization, Not Measurement: A Critique of Stevens’ Scales of Measurement*. en. SSRN Scholarly Paper ID 2412765. Rochester, NY: Social Science Research Network, Oct. 2014. DOI: [10.2139/ssrn.2412765](https://doi.org/10.2139/ssrn.2412765).
- [81] “Connected Space”. en. In: *Wikipedia* (Dec. 2020).
- [82] Caroline Ziemkiewicz and Robert Kosara. “Embedding Information Visualization within Visual Representation”. In: *Advances in Information and Intelligent Systems*. Ed. by Zbigniew W. Ras and William Ribarsky. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 307–326. ISBN: 978-3-642-04141-9. DOI: [10.1007/978-3-642-04141-9_15](https://doi.org/10.1007/978-3-642-04141-9_15).

- [83] Sheelagh Carpendale. *Visual Representation from Semiology of Graphics by J. Bertin*. en.
- [84] “Jet Bundle”. en. In: *Wikipedia* (Dec. 2020).
- [85] Jana Musilová and Stanislav Hronek. “The Calculus of Variations on Jet Bundles as a Universal Approach for a Variational Formulation of Fundamental Physical Theories”. In: *Communications in Mathematics* 24.2 (Dec. 2016), pp. 173–193. ISSN: 2336-1298. DOI: [10.1515/cm-2016-0012](https://doi.org/10.1515/cm-2016-0012).
- [86] Yael Albo et al. “Off the Radar: Comparative Evaluation of Radial Visualization Solutions for Composite Indicators”. In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Jan. 2016), pp. 569–578. ISSN: 1077-2626. DOI: [10.1109/TVCG.2015.2467322](https://doi.org/10.1109/TVCG.2015.2467322).
- [87] Z. Qu and J. Hullman. “Keeping Multiple Views Consistent: Constraints, Validations, and Exceptions in Visualization Authoring”. In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), pp. 468–477. ISSN: 1941-0506. DOI: [10.1109/TVCG.2017.2744198](https://doi.org/10.1109/TVCG.2017.2744198).
- [88] Richard A. Becker and William S. Cleveland. “Brushing Scatterplots”. In: *Technometrics* 29.2 (May 1987), pp. 127–142. ISSN: 0040-1706. DOI: [10.1080/00401706.1987.10488204](https://doi.org/10.1080/00401706.1987.10488204).
- [89] Andreas Buja et al. “Interactive Data Visualization Using Focusing and Linking”. In: *Proceedings of the 2nd Conference on Visualization '91*. VIS '91. Washington, DC, USA: IEEE Computer Society Press, 1991, pp. 156–163. ISBN: 0-8186-2245-8.
- [90] *Dataclasses — Data Classes — Python 3.9.2rc1 Documentation*. <https://docs.python.org/3/library/dataclasses>.
- [91] Kristen B. Gorman, Tony D. Williams, and William R. Fraser. “Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus *Pygoscelis*)”. In: *PLOS ONE* 9.3 (Mar. 2014), e90081. DOI: [10.1371/journal.pone.0090081](https://doi.org/10.1371/journal.pone.0090081).

- [92] Allison Marie Horst, Alison Presmanes Hill, and Kristen B Gorman. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. Manual. 2020. DOI: [10.5281/zenodo.3960218](https://doi.org/10.5281/zenodo.3960218).
- [93] Muhammad Chenariyan Nakhaee. *Mcnakhaee/Palmerpenguins*. Jan. 2021.
- [94] F. Beck. “Software Feathers Figurative Visualization of Software Metrics”. In: *2014 International Conference on Information Visualization Theory and Applications (IVAPP)*. Jan. 2014, pp. 5–16.
- [95] Lydia Byrne, Daniel Angus, and Janet Wiles. “Figurative Frames: A Critical Vocabulary for Images in Information Visualization”. In: *Information Visualization* 18.1 (Aug. 2017), pp. 45–67. ISSN: 1473-8716. DOI: [10.1177/1473871617724212](https://doi.org/10.1177/1473871617724212).
- [96] John M Chambers et al. *Graphical Methods for Data Analysis*. Vol. 5. Wadsworth Belmont, CA, 1983.
- [97] Robert B Haber and David A McNabb. “Visualization Idioms: A Conceptual Model for Scientific Visualization Systems”. In: *Visualization in scientific computing* 74 (1990), p. 93.
- [98] Charles D Hansen and Chris R Johnson. *Visualization Handbook*. Elsevier, 2011.
- [99] Leland Wilkinson and Michael Friendly. “The History of the Cluster Heat Map”. In: *The American Statistician* 63.2 (May 2009), pp. 179–184. ISSN: 0003-1305. DOI: [10.1198/tas.2009.0033](https://doi.org/10.1198/tas.2009.0033).
- [100] Toussaint Loua. *Atlas Statistique de La Population de Paris*. J. Dejeu & cie, 1873.
- [101] Hadley Wickham and Lisa Stryjewski. “40 Years of Boxplots”. In: *The American Statistician* (2011).
- [102] C. Heine et al. “A Survey of Topology-Based Methods in Visualization”. In: *Computer Graphics Forum* 35.3 (June 2016), pp. 643–667. ISSN: 0167-7055. DOI: [10.1111/cgf.12933](https://doi.org/10.1111/cgf.12933).
- [103] James O Ramsay. *Functional Data Analysis*. Wiley Online Library, 2006.

- 954 [104] Enrico Bertini, Andrada Tatu, and Daniel Keim. “Quality Metrics in High-
955 Dimensional Data Visualization: An Overview and Systematization”. In: *IEEE*
956 *Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2203–2212.
- 957 [105] Bartosz Milewski. “Category Theory for Programmers”. en. In: (), p. 498.