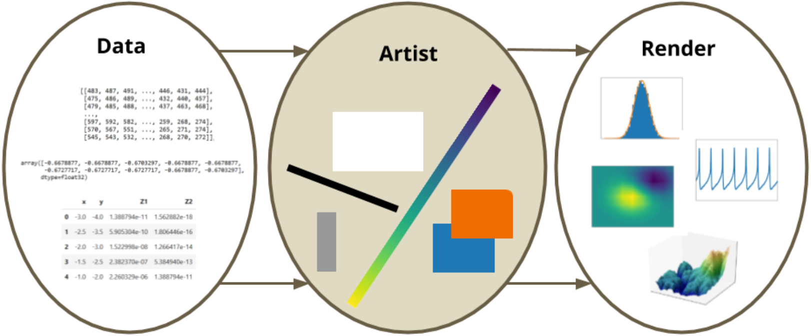
Topological Artist Model

March 5, 2021

Hannah Aizenman
Advisor: Dr. Michael Grossberg
Committee: Dr. Robert Haralick, Dr. Lev Manovich, Dr. Huy Vo
External Member: Dr. Marcus Hanwell

What do visualization libraries do?



Why are we doing this?

1. visualization software design patterns are tuned to data structures[16]
2. visualization types tuned to data structures[35]
3. need to build a tool that supports domain specific structures and algorithms
4. tool needs to be general purpose enough to support lots of domains
5. tools needs to support 2D, 3D, static, dynamic, and interactive viz

Everything is a table

DATA: longitude, latitude = *map*(*source*("World"))
TRANS: bd = *max*(birth-death, 0)
COORD: *project.mercator*()
ELEMENT: *point*(*position*(lon*lat), *size*(bd), *color*(*color.red*))
ELEMENT: *polygon*(*position*(longitude*latitude))

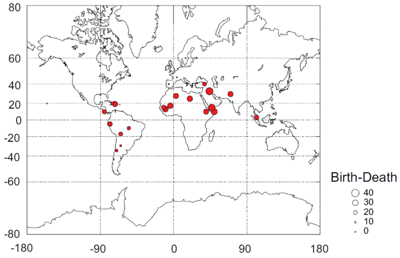


Figure 1.5 *Excess birth (vs. death) rates in selected countries*

Figure: Figure 1.5 in Wilkenson's Grammar of Graphics[40]

1. ggplot[39], protovis[4], D3 [5], vega[27], altair[37]
2. tends to be declarative[17]
3. users describe how to compose visual elements [41]

TAM: build the visual elements

Everything is an image

1. ImageJ[28], ImagePlot[32], Napari[29]
2. build plugins into existing system where image is primary input

TAM: more general than tables and images

Everything is ?

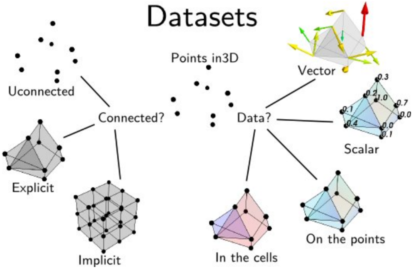


Figure: Image is from the Data Representation chapter of the MayaVi 4.7.2 documentation.[12]

1. Matplotlib[18], VTK [14, 15], MayaVi[26], ParaView[3], Titan[6]
2. each visualization type API is tuned to the data structure

TAM: common data abstraction interface between data and viz

Variables & Topology

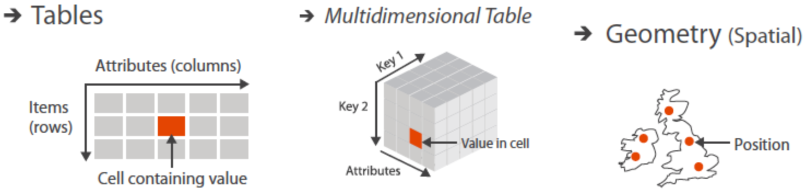


Figure: Image is figure 2.8 in Munzner's Visualization Analysis and Design[24]

	metadata are *keys* with associated *values* (Munzner [24])
topology	Fiber bundles can be a common data abstraction (Butler [7, 8])
variables	Fiber as a database schema (Spivak [30, 31])

Tam: variables are *values* and *keys* locate them in topology

Property Matching

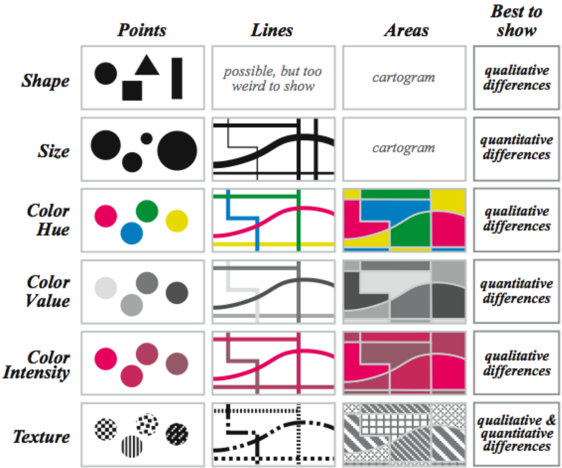


Figure: This tabular form of Bertin's retinal variables is from Understanding Graphics [23] who reproduced it from Krygier and Wood's *Making Maps: A Visual Guide to Map Design for GIS*[20]

Evaluating visualizations

Expressiveness structure preserving mappings from data to graphic (Mackinlay [21])

Effectiveness design choices made in deference to perceptual saliency [**(**]Mackinlay [9–11, 24])

Naturalness easier to understand when properties match (Norman [25])

Graphical Integrity graphs show **only** the data (Tufte [36])

Mathematical Frameworks for evaluating visualization

1. APT: visualization has syntax and semantics like a language (Mackinlay [21, 22])
2. Visualization has functional dependencies that can be represented as graphs (Sugibuchi [33])
3. the semiotics of visualization are commutative in a category theory framework (Vickers [38])

Tam: Framework for building visualizations

Visualization is commutative maps

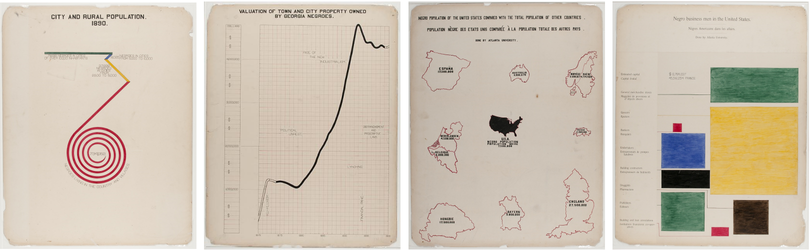
data (α) and viz (ω) symmetries (Kindlmann and Scheidegger [19])

$v \circ r_2 \circ \alpha = \omega \circ v \circ r_1$

$$\begin{tikzcd} D \arrow[r, "r_1"] \arrow[d, "\alpha"'] & R \arrow[r, "\nu"] & V \arrow[d, "\omega"] \\ D \arrow[r, "r_2"'] & R \arrow[r, "\nu"'] & V \end{tikzcd}$$

Tam: Add topology and make it functional

So why?



(a) (b) (c) (d)

Figure: Du Bois' data portraits from the Prints and Photographs collection of the Library of Congress [1, 2, 13, 34]

Framework for building any valid visualization

Contributions

1. a formal description of the topology preserving relationship between data and graphic via continuous maps
2. a formal description of the property preservation from data component to visual representation as equivariant maps that carry a homomorphism of monoid actions
3. abstraction of data structure using fiber bundles with schema like fibers to encode components and topology
4. algebraic sum operator such that more complex visualizations can be built from simple ones
5. a functional oriented visualization tool architecture built on the mathematical model to demonstrate the utility of the model
6. a prototype of the architecture built on Matplotlib's infrastructure to demonstrate the feasibility of the model

Topological Artist Model

$$\mathscr{A} : \mathscr{E} \to \mathscr{H} \tag{1}$$

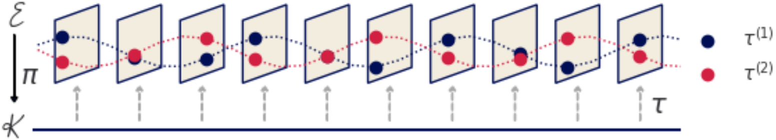
which is a map from data E to graphic H that carries a homomorphism of monoid actions

$$\varphi : M \to M' \tag{2}$$

such that artists are equivariant maps

$$\mathscr{A}(m \cdot r) = \varphi(m) \cdot \mathscr{A}(r) \tag{3}$$

Data Bundle



a fiber bundle is a tuple (E, K, π, F) defined by the projection map π

$$F \hookrightarrow E \xrightarrow{\pi} K \tag{4}$$

Variables: Fiber

$$F = \mathbb{U}_{\sigma(c)} = \mathbb{U}_T \tag{5}$$

where $\mathbb{U}_{\sigma(c)}$ is

$$\begin{tikzcd} \mathbb{U}_\sigma \arrow[r] \arrow[d, "\pi_\sigma"'] & \mathbb{U} \arrow[d, "\pi"] \\ C \arrow[r, "\sigma"'] & \mathbf{DT} \end{tikzcd} \tag{6}$$

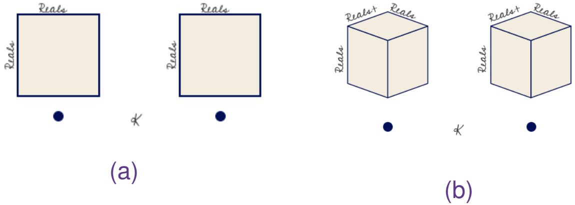
DT data types of the variables in the dataset

$\mathbb{U}$ disjoint union of all values of type $T \in \mathbf{DT}$

C variable names

$\mathbb{U}_\sigma$ $\mathbb{U}$ restricted to the data type of a named variable

Sample Fiber



Figure

6a $F = \mathbb{R} \times \mathbb{R}$, for example (time, temperature)

6b $\mathbb{R} \times \mathbb{R}^{+} \times \mathbb{R}$, for example (time, wind=(speed, direction))

Fiber Components

We can decouple decouple F into components F_i

$$F = F_0 \times \ldots \times F_i \times \ldots \times F_n \tag{7}$$

Structure of Components: Monoids

A monoid M is a set with

associative binary operator $* : M \times M \to M$

identity element $e \in M$ such that $e * a = a * e = a$ for all $a \in M$.

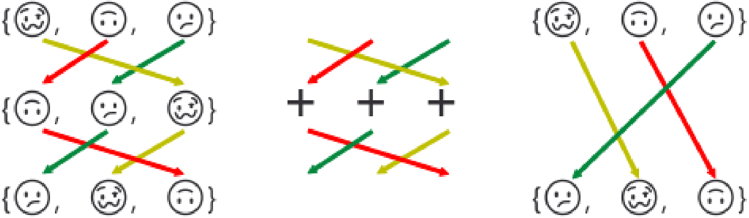
Monoid Actions

A left monoid action of M_i is a set F_i with an action $\bullet : M \times F_i \to F_i$ 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 $x \in F_i, e \in M_i$, $e \bullet x = x$

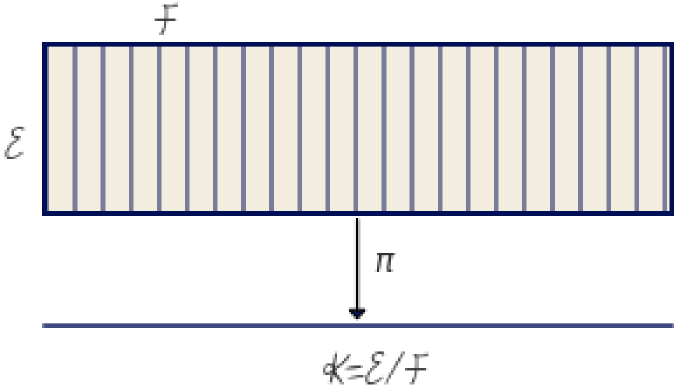
Monoid Composition: Permutation



Monoid Composition: Partial Order

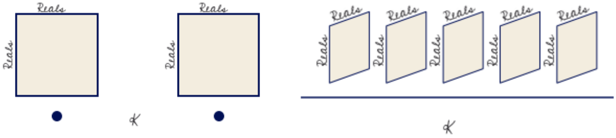
MAKE THIS Hasse Diagram (borrow from theories of composablity)

Continuity: base space



Figure

Sample: Base Space



Values: Section

For any fiber bundle, there exists a map

$$\begin{tikzcd} F \arrow[r, hook] & E \arrow[d, "\pi"'] \\ & K \arrow[u, bend right, "\tau"'] \end{tikzcd} \tag{9}$$

such that $\pi(\tau(k)) = k$.
The set of all global sections is denoted as $\Gamma(E)$.

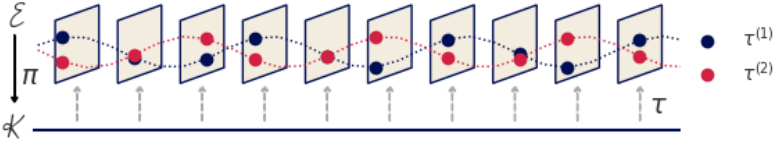
Record

Assuming a trivial fiber bundle $E = K \times F$, the section is

$$\tau(k) = (k, (g_{F_0}(k), \ldots, g_{F_n}(k))) \tag{10}$$

where $g : K \to F$ is the index function into the fiber.

Sample dataset



Fiber Bundle

The graphics bundle is a tuple (H, S, π, D) defined by the projection map π

$$D \hookrightarrow H \xrightarrow{\pi} S \tag{11}$$

Continuity: Base space

The surjective map $\xi : S \to K$

$$\begin{tikzcd} E \arrow[d, "\pi"'] & H \arrow[d, "\pi"'] \\ K & S \arrow[l, "\xi"'] \end{tikzcd} \tag{12}$$

goes from region $s \in S_k$ to its associated point s.

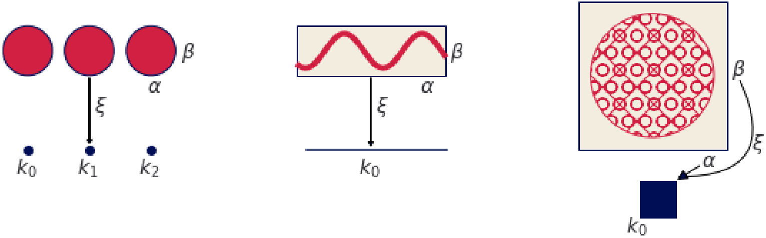
Sample Display Target: Fiber

Assume a 2D opaque image $D = \mathbb{R}^5$ with elements

$$(x, y, r, g, b) \in D \tag{13}$$

such that a rendered graphic only has 2D position and color.

Continuity Retraction Maps



Rendering: Define a Pixel

Given a pixel

$$p = \left[y_{top}, y_{bottom}, x_{right}, x_{left}\right] \tag{14}$$

the inverse map of the bounding box

$$S_p = {\rho_{xy}}^{-1}(p) \tag{15}$$

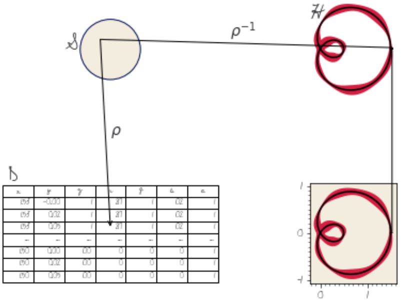
is a region $S_p \subset S$ such that

$$r_p = \iint_{S_p} \rho_r(s) ds^2 \tag{16}$$

$$g_p = \iint_{S_p} \rho_g(s) ds^2 \tag{17}$$

$$b_p = \iint_{S_p} \rho_b(s) ds^2 \tag{18}$$

yields the color of the pixel.



References I

- *[A Series of Statistical Charts Illustrating the Condition of the Descendants of Former African Slaves Now in Residence in the United States of America] Negro Business Men in the United States.* eng. https://www.loc.gov/item/2014645363/. Image.
- *[A Series of Statistical Charts Illustrating the Condition of the Descendants of Former African Slaves Now in Residence in the United States of America] Negro Population of the United States Compared with the Total Population of Other Countries /.* eng. https://www.loc.gov/item/2013650368/. Image.
- J. Ahrens, B. Geveci, and C. Law. "Paraview: An End-User Tool for Large Data Visualization". In: *The visualization handbook* 717.8 (2005).

References II

- 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.
- M. Bostock, V. Ogievetsky, and J. Heer. “D^3 Data-Driven Documents”. In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2301–2309.
- Brian Wylie and Jeffrey Baumes. “A Unified Toolkit for Information and Scientific Visualization”. In: *Proc.SPIE*. Vol. 7243. Jan. 2009.
- D. M. Butler and M. H. Pendley. “A Visualization Model Based on the Mathematics of Fiber Bundles”. en. In: *Computers in Physics* 3.5 (1989), p. 45.
- D. M. Butler and S. Bryson. “Vector-Bundle Classes Form Powerful Tool for Scientific Visualization”. en. In: *Computers in Physics* 6.6 (1992), p. 576.

References III

- J. M. Chambers et al. *Graphical Methods for Data Analysis*. Vol. 5. Wadsworth Belmont, CA, 1983.
- W. S. Cleveland. “Research in Statistical Graphics”. In: *Journal of the American Statistical Association* 82.398 (June 1987), p. 419.
- W. S. Cleveland and R. McGill. “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods”. In: *Journal of the American Statistical Association* 79.387 (Sept. 1984), pp. 531–554.
- *Data Representation in Mayavi — Mayavi 4.7.2 Documentation*. https://docs.enthought.com/mayavi/mayavi/data.html.

References IV

- W. E. B. W. E. B. Du Bois. *[The Georgia Negro] Valuation of Town and City Property Owned by Georgia Negroes.* en. https://www.loc.gov/pictures/item/2013650441/. Still Image. 1900.
- B. Geveci et al. “VTK”. In: *The Architecture of Open Source Applications* 1 (2012), pp. 387–402.
- M. D. Hanwell et al. “The Visualization Toolkit (VTK): Rewriting the Rendering Code for Modern Graphics Cards”. en. In: *SoftwareX* 1-2 (Sept. 2015), pp. 9–12.
- 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.

References V

- J. Heer and M. Bostock. "Declarative Language Design for Interactive Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (Nov. 2010), pp. 1149–1156.
- J. D. Hunter. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science Engineering* 9.3 (May 2007), pp. 90–95.
- G. Kindlmann and C. Scheidegger. "An Algebraic Process for Visualization Design". In: *IEEE transactions on visualization and computer graphics* 20.12 (2014), pp. 2181–2190.
- J. Krygier and D. Wood. *Making Maps: A Visual Guide to Map Design for GIS*. English. 1 edition. New York: The Guilford Press, Aug. 2005.

- J. Mackinlay. “Automating the Design of Graphical Presentations of Relational Information”. In: *ACM Transactions on Graphics* 5.2 (Apr. 1986), pp. 110–141.
- J. D. MACKINLAY. “AUTOMATIC DESIGN OF GRAPHICAL PRESENTATIONS (DATABASE, USER INTERFACE, ARTIFICIAL INTELLIGENCE, INFORMATION TECHNOLOGY)”. English. PhD Thesis. 1987.
- C. Malamed. *Information Display Tips*. https://understandinggraphics.com/visualizations/information-display-tips/. Blog. Jan. 2010.
- T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC press, Oct. 2014.

References VII

- D. A. Norman. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. USA: Addison-Wesley Longman Publishing Co., Inc., 1993.
- P. Ramachandran and G. Varoquaux. “Mayavi: 3D Visualization of Scientific Data”. In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 40–51.
- A. Satyanarayan, K. Wongsuphasawat, and J. 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.
- C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. “NIH Image to ImageJ: 25 Years of Image Analysis”. In: *Nature Methods* 9.7 (July 2012), pp. 671–675.

References VIII

- N. Sofroniew et al. *Napari/Napari: 0.4.5rc1*. Zenodo. Feb. 2021.
- D. I. Spivak. *Databases Are Categories*. en. Slides. June 2010.
- D. I. Spivak. “SIMPLICIAL DATABASES”. en. In: (), p. 35.
- S. Studies. *Culturevis/Imageplot*. Jan. 2021.
- T. Sugibuchi, N. Spyratos, and E. Siminenko. “A Framework to Analyze Information Visualization Based on the Functional Data Model”. In: *2009 13th International Conference Information Visualisation*. 2009, pp. 18–24.
- *[The Georgia Negro] City and Rural Population. 1890.* eng. https://www.loc.gov/item/2013650430/. Image. 1900.

- M. Tory and T. Moller. “Rethinking Visualization: A High-Level Taxonomy”. In: *IEEE Symposium on Information Visualization*. 2004, pp. 151–158.
- E. R. Tufte. *The Visual Display of Quantitative Information*. English. Cheshire, Conn.: Graphics Press, 2001.
- J. VanderPlas et al. “Altair: Interactive Statistical Visualizations for Python”. en. In: *Journal of Open Source Software* 3.32 (Dec. 2018), p. 1057.
- P. Vickers, J. Faith, and N. Rossiter. “Understanding Visualization: A Formal Approach Using Category Theory and Semiotics”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.6 (June 2013), pp. 1048–1061.

References X

- H. Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- L. Wilkinson. *The Grammar of Graphics*. en. 2nd ed. Statistics and Computing. New York: Springer-Verlag New York, Inc., 2005.
- K. Wongsuphasawat. *Navigating the Wide World of Data Visualization Libraries (on the Web)*. 2021.