Data Visualization From a Category Theory Perspective

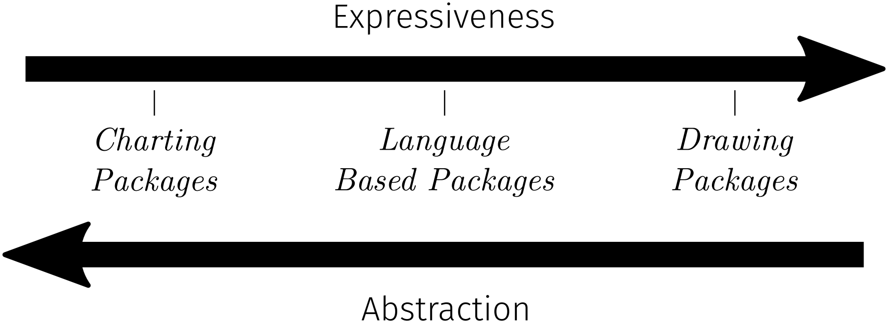
Davi Sales Barreira, Asla Medeiros e Sá

FGV - EMAp, IMPATech

February 4, 2025

Motivation

Balance expressiveness and abstraction in data visualization frameworks.



Motivation

How can we represent complex visualizations without resorting to low-level specifications?

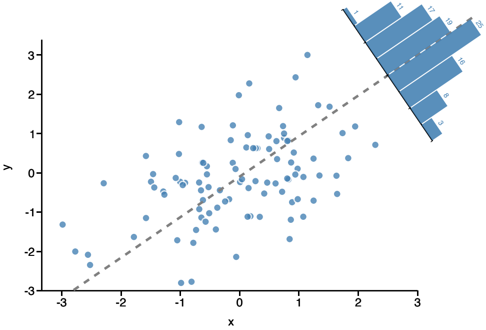


Figure: Rotated histogram aligned with second main PCA axis.

Motivation

How can we represent complex visualizations without resorting to low-level specifications?

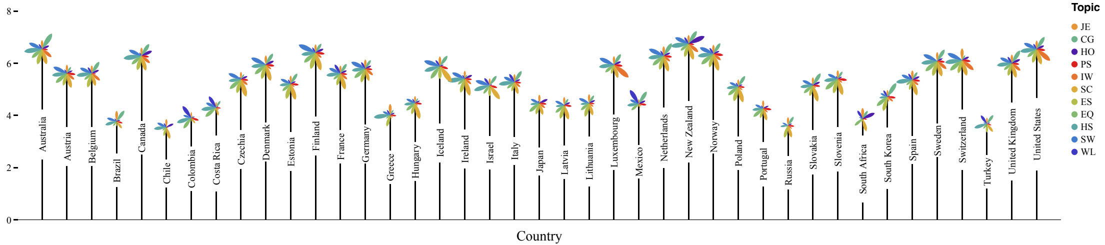


Figure: OECD Better Life Index visualization Stefaner and OECD [3].

Overview

- **Day 1**: Basics of Category Theory
- **Day 2**: Programming with Category Theory
- **Day 3**: Data Visualization Theory
- **Day 4**: Data Visualization + Categorical Programming

Table of contents

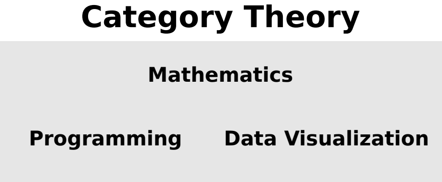
1. What are Categories?
2. Examples of Categories
3. Universal Constructions
4. Functors
5. Natural Transformations
6. Monoids and Monads
7. F-Algebras and F-Coalgebras

Why Category Theory?

Category Theory is a branch of mathematics that studies general abstract structures through their relationships.

Origin: Samuel Eilenberg e Saunders Mac Lane - 1940

As pointed by Fong and Spivak [2], Category Theory is unmatched in its ability to organize and relate abstractions.



What are Categories?

Category Theory is a branch of mathematics that studies general abstract structures through their relationships.

Definition (Category)

- A collection of objects $\mathrm{Ob}_{\mathcal{C}}$.
- A collection of morphisms $\mathrm{Hom}_{\mathcal{C}}$, where each morphism has a source object and a target object. $\mathrm{Hom}_{\mathcal{C}}(A, B)$ is the collection of morphisms going from object A to object B.
- A binary operation $\circ : \mathrm{Hom}_{\mathcal{C}}(A, B) \times \mathrm{Hom}_{\mathcal{C}}(B, C) \to \mathrm{Hom}_{\mathcal{C}}(A, C)$ such that:
    1. **Associative**: $(h \circ g) \circ f = h \circ (g \circ f)$.
    2. **Identity**: Every object has an identity morphism $1_A \in \mathrm{Hom}_{\mathcal{C}}(A, A)$.

What are Categories?

Definition (Small and Locally Small Category)

A category $\mathcal{C}$ is *small* if $\mathrm{Ob}_{\mathcal{C}}$ and $\mathrm{Hom}_{\mathcal{C}}$ are sets. A category $\mathcal{C}$ is *locally small* if for any $A, B \in \mathrm{Ob}_{\mathcal{C}}$, then $\mathrm{Hom}_{\mathcal{C}}(A, B)$ is a set. Note that a small category is also locally small.

Note that when talking about $\mathrm{Ob}_{\mathcal{C}}$ and $\mathrm{Hom}_{\mathcal{C}}$, we didn't say that they were sets, instead we called them *classes*. The reason for this lies in the foundations of Set Theory. There are collections in mathematics that are "larger" than sets, e.g. the "set" of all sets, which itself cannot be a set, otherwise it would incur in a paradox (Russell's Paradox). A way to deal with this is making a distinction between classes and sets. This point is quite technical; readers interested in understanding this nuance can check books such as Borceux [1].

Examples of Categories

The category $\mathbf{1}$ consists of $\mathsf{Ob}_{\mathbf{1}} := \{A\}$ and $\mathrm{Hom}_{\mathbf{1}} = \mathrm{id}_A$.

$$\begin{tikzcd} A \arrow[loop above, "id_A"] \end{tikzcd}$$

Examples of Categories

The category **2** consists of $\mathrm{Ob}_{\mathbf{2}} := \{A, B\}$ and $\mathrm{Hom}_{\mathbf{1}} = \{id_A, id_B, f\}$, where $f : A \to B$. The diagram for such category is shown below.

$$\begin{tikzcd} A \arrow[loop above, "id_A"] \arrow[r, "f"] & B \arrow[loop above, "id_B"] \end{tikzcd}$$

Figure: Category **2**.

Examples of Categories

The category **3** has three morphisms besides the identities. The morphisms are f, g and their composition $g \circ f$. The figure below illustrates the category with all its morphisms.

$$\begin{tikzcd}
A \arrow[loop above, "id_A"] \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & B \arrow[loop above, "id_B"] \arrow[r, "g"] & C \arrow[loop above, "id_C"]
\end{tikzcd}$$

Figure: Category **3** showing all morphisms.

Examples of Categories

When drawing categories, it is common to omit the identity and/or composition morphism. From here on, we do the same, whenever the context is clear.

$$A \xrightarrow{f} B \xrightarrow{g} C$$

Figure: Category **3** omitting morphisms.

Examples of Categories

The discrete category $\underline{\mathbf{N}}$ is the category with N objects and $\text{Hom}\underline{\mathbf{N}} := \{id_1, ..., id_N\}$. An example of this category is illustrated below.

$$\begin{tikzcd}
1 \arrow[loop above, "id_1"] & 2 \arrow[loop above, "id_2"] & \cdots & N \arrow[loop above, "id_N"]
\end{tikzcd}$$

Figure: Category $\underline{\mathbf{N}}$.

Examples of Categories

Given a category $\mathcal{C}$ and an object S of this category, we can define a slice category $\mathcal{C}/S$, where:

- The objects are tuples (A, f) where A is an object in $\mathcal{C}$ and $f : A \to S$ is a morphism.
- A morphism $\varphi_{(A,B)} : (A, f) \to (B, g)$ is equivalent to a morphism $\varphi \in \mathrm{Hom}_{\mathcal{C}}(A, B)$ such that $f = g \circ \varphi$.

$$
\begin{tikzcd}
 & \mathcal{C}/S & & & & \mathcal{C} & \\
(A,\, f) \arrow[rr, "\phi_{A,B}"] & & (B,\, g\,) & \cong & A \arrow[rr, "\phi"] \arrow[dr, "f"'] & & B \arrow[dl, "g"] \\
 & & & & & S &
\end{tikzcd}
$$

Figure: Example of slice category.

Examples of Categories

Definition (Subcategory)

Let $\mathcal{C}$ be a category. A *subcategory* $\mathcal{S}$ of $\mathcal{C}$ is such that

(i) $\mathrm{Ob}_{\mathcal{S}} \subseteq \mathrm{Ob}_{\mathcal{C}}$;

(ii) For every $A, B \in \mathrm{Ob}_{\mathcal{S}}$, we have $\mathrm{Mor}_{\mathcal{S}}(A, B) \subseteq \mathrm{Ob}_{\mathcal{C}}(A, B)$;

(iii) Composition and identity in $\mathcal{S}$ are the same as in $\mathcal{C}$, restricted to morphisms and objects of $\mathcal{S}$.

A subcategory $\mathcal{S}$ is said to be *wide* if $\mathrm{Ob}_{\mathcal{S}} = \mathrm{Ob}_{\mathcal{C}}$, and it is said to be *full* if for every $A, B \in \mathrm{Ob}_{\mathcal{S}}$, then $\mathrm{Mor}_{\mathcal{S}}(A, B) = \mathrm{Ob}_{\mathcal{C}}(A, B)$. Finally, a subcategory is *thin* if for every $A, B \in \mathrm{Ob}_{\mathcal{S}}$ the set $\mathrm{Mor}_{\mathcal{S}}(A, B)$ has only a single morphism.

Examples of Categories

Here are some more interesting categories:

1. **Set** which is the category of sets, where the objects are sets and the morphisms are functions between sets.
2. **Top** is the category where topological spaces are the objects and continuous functions are the morphisms.
3. $\mathbf{Vec}_{\mathbb{F}}$ is the category where vector spaces over field $\mathbb{F}$ are the objects, and linear transformations are the morphisms.
4. **Mon** is the category of monoids, where morphisms are monoid homormophisms.

Universal Constructions

Objects defined in terms of existence and uniqueness of morphisms are known as **universal constructions**.

Definition (Zero, Initial and Terminal)

Let $\mathcal{C}$ be a category.

1. An object $I \in \mathrm{Ob}_{\mathcal{C}}$ is *initial* if for every $A \in \mathrm{Ob}_{\mathcal{C}}$, there is exactly one morphism from I to A. Thus, from I to I there is only the identity.
2. An object $T \in \mathrm{Ob}_{\mathcal{C}}$ is *terminal* if for every $A \in \mathrm{Ob}_{\mathcal{C}}$, there is exactly one morphism from A to T. Thus, from I to I there is only the identity.
3. An object is *zero* if it is both terminal and initial.

Definition (Categorical Isomorphism)

Let $\mathcal{C}$ be a category with $X, Y \in Ob_{\mathcal{C}}$ and $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$.

(i) We say that f is *left invertible* if there exists $f_l \in \mathrm{Hom}_{\mathcal{C}}(Y, X)$ such that $f_l \circ f = id_X$;

(ii) We say that f is *right invertible* if there exists $f_r \in \mathrm{Hom}_{\mathcal{C}}(Y, X)$ such that $f \circ f_r = id_Y$;

(iii) We say that f is invertible if it's both left and right invertible.

Universal Constructions

Theorem

Every initial object is unique up to an isomorphism, i.e. if in a category there are two initial objects, then they are isomorphic. Similarly, terminal objects are unique up to an isomorphism. Moreover, the isomorphism is unique between initial object, and between terminal objects.

Proof.

□

Universal Constructions

Theorem

Every initial object is unique up to an isomorphism, i.e. if in a category there are two initial objects, then they are isomorphic. Similarly, terminal objects are unique up to an isomorphism. Moreover, the isomorphism is unique between initial object, and between terminal objects.

Proof.

Let I_1, I_2 be initial. Then, there exists only $f : I_1 \to I_2$ and $g : I_2 \to I_1$. But since $g \circ f : I_1 \to I_1$ is a morphism from the initial object I_1, it must be equal to id_{I_1}. The same for I_2, which implies that f and g are inverses, and thus the objects are isomorphic. Since both f and g are the only morphisms from I_1 and I_2, this also implies that their are the only isomorphism. The same proof works for terminal objects. □

Universal Constructions

Definition (Product)

Set Category

A very important category is **Set**. This category is used in programming to model types. Some properties:

- Initial object: $\varnothing$;
- Terminal object: any singleton set up to isomorphism;
- For any two objects A and B, the $\mathrm{Hom}(A, B)$ is also an object.

Category Theory Brief Introduction

Definition (Functor)

Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. A functor $F : \mathcal{C} \to \mathcal{D}$ is a pair of mappings with the following properties:

Covariant Functor

$$
\begin{tikzcd}
A \arrow[loop above, "id_A"] \arrow[r, Rightarrow, "F"] & FA \arrow[loop above, "F(id_A) = id_{FA}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "f"] \arrow[dr, "g \circ f"'] & B \arrow[d, "g"] \\
& C
\end{tikzcd}
\overset{F}{\Longrightarrow}
\begin{tikzcd}
FA \arrow[r, "Ff"] \arrow[dr, "Fg \circ Ff"'] & FB \arrow[d, "Fg"] \\
& FC
\end{tikzcd}
$$

Contravariant Functor

$$
\begin{tikzcd}
A \arrow[loop above, "id_A"] \arrow[r, Rightarrow, "F"] & FA \arrow[loop above, "F(id_A) = id_{FA}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "f"] \arrow[dr, "g \circ f"'] & B \arrow[d, "g"] \\
& C
\end{tikzcd}
\overset{F}{\Longrightarrow}
\begin{tikzcd}
FA & FB \arrow[l, "Ff"'] \\
& FC \arrow[u, "Fg"] \arrow[ul, "Ff \circ Fg"]
\end{tikzcd}
$$

Figure: Diagrams showcasing the properties of functors.

Category Theory Brief Introduction

Definition (Natural Transformations)

Let $\mathcal{C}$ and $\mathcal{D}$ be categories, and let $F, G : \mathcal{C} \to \mathcal{D}$ be functors. A natural transformation $\alpha : F \to G$ is such that the following diagram commutes:

$$
\begin{tikzcd}
FA \arrow[r, "\alpha_A"] \arrow[d, "Ff"'] & GA \arrow[d, "Gf"] \\
FB \arrow[r, "\alpha_B"'] & GB
\end{tikzcd}
$$

$$Gf \circ \alpha_A = \alpha_B \circ Ff$$

Figure: Commutative diagram of a natural transformation highlighting the commutative property of the definition.

Category Theory Brief Introduction

Monoids and **Monads** are two ubiquitous constructions both in Category Theory and Functional Programming. These two concepts will be used when talking about data visualization. Therefore, it is required of us to introduce these constructions.
Let's start with the definition of a monoid in the context of Set Theory.

Definition (Monoid - Set Theory)

A monoid is a triple $(M, \otimes, e_M)$ where M is a set, $\otimes : M \times M \to M$ is a binary operation and e_M the neutral element, such that:

1. $a \otimes (b \otimes c) = (a \otimes b) \otimes c$
2. $a \otimes e_M = e_M \otimes a = a$.

An example of a monoid is $(\mathbb{N} \cup \{0\}, +, 0)$. It is easy to check that the summation operator satisfies the associativity neutrality properties.

Category Theory Brief Introduction

Definition (Monoid in the category **Set**)

A monoid in **Set** is a triple (M, μ, η), where $M \in \mathrm{Ob}_{\textbf{Set}}$, $\mu : M \times M \to M$ and $\eta : 1 \to M$ are two morphisms in **Set** satisfying the commutative diagrams below. Note that 1 is the terminal object in **Set**, i.e. the singleton set (which is unique up to an isomorphism).

$$
\begin{tikzcd}
M \times M \times M \arrow[r, "id_M \times \mu"] \arrow[d, "\mu \times id_M"'] & M \times M \arrow[d, "\mu"] \\
M \times M \arrow[r, "\mu"] & M
\end{tikzcd}
\qquad
\begin{tikzcd}
M \arrow[r, "\eta \times id_M"] \arrow[dr, "id_M"'] & M \times M \arrow[d, "\mu"] & M \arrow[l, "id_M \times \eta"'] \arrow[dl, "id_M"] \\
& M &
\end{tikzcd}
$$

Figure: Commutative diagram for monoid.

References

[1] Borceux, F. (1994). *Handbook of categorical algebra: volume 1, Basic category theory*, volume 1. Cambridge University Press.

[2] Fong, B. and Spivak, D. I. (2019). *An invitation to applied category theory: seven sketches in compositionality*. Cambridge University Press.

[3] Stefaner, M. and OECD (2012). Oecd better life index. `http://www.oecdbetterlifeindex.org/`. Accessed on 14 Oct. 2012.

The End