

VisDecode: Distilling Design Decisions in Visualizations using pixels-to-text Foundation Models

Martín A. Sinnona,¹ Julián Eisenchlos,² Viviana Siless,¹ Emmanuel Iarussi^{1,3}

¹ Universidad Torcuato Di Tella.

² Google DeepMind, Google.

³ Consejo Nacional de Investigaciones Científicas y Técnicas.

✉ martin.a.sinnona@gmail.com

Introduction

Automated Visualization's understanding and features extraction is a fundamental challenge in **Data Visualization**.

Visualizations are crucial for conveying information, insights or messages, but design choices such as **colors, shapes, positioning**, and other visual attributes, can be intricate, vague and lead to misinterpretation and ineffective visuals.

We present **VisDecode**, a novel framework to automatically distill design decisions from charts and uncover the thought process behind them.

Goals

✓ Augmenting visualization interfaces to advocate for **better design practices**.

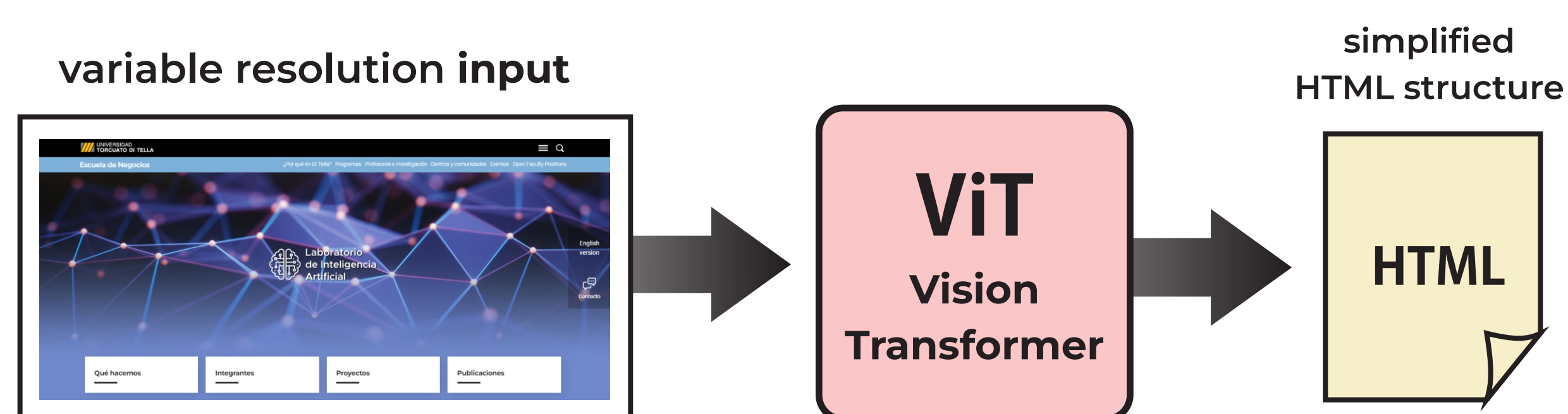
✓ Extract design choices from charts such as **mark, variable types** and **variable names**.

✓ Identify **perceptual attributes** and link them to data dimensions.

Pix2Struct [2]

✓ Base **image-to-text** model for **Visual Language Understanding**.

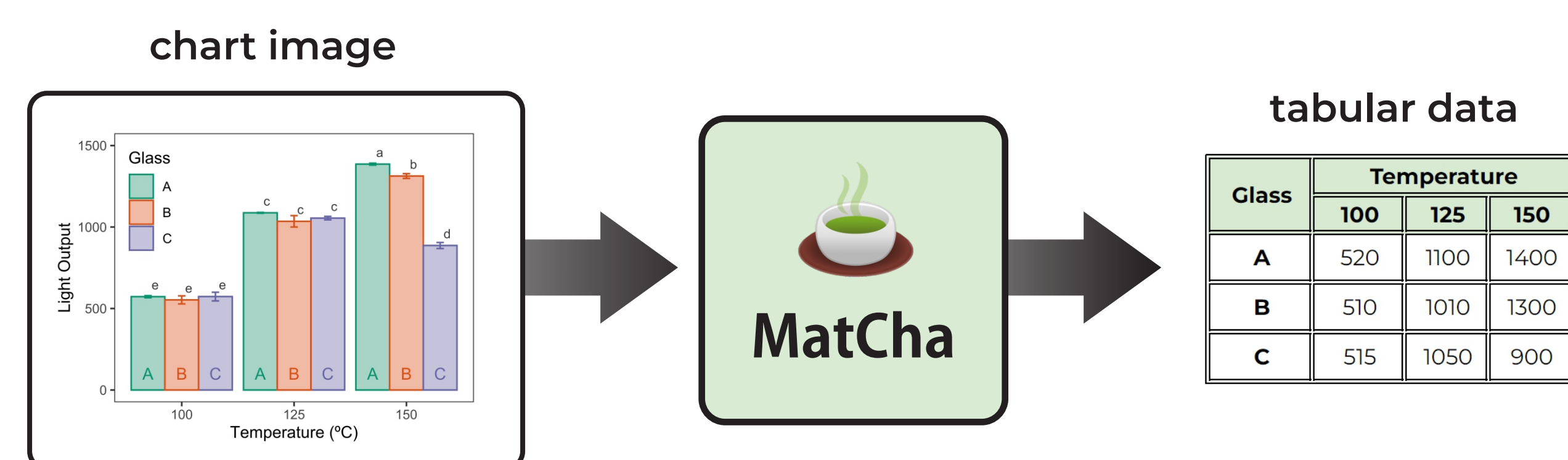
✓ Pretrained by learning to parse web pages screenshots into **HTML**.



MatCha [1]

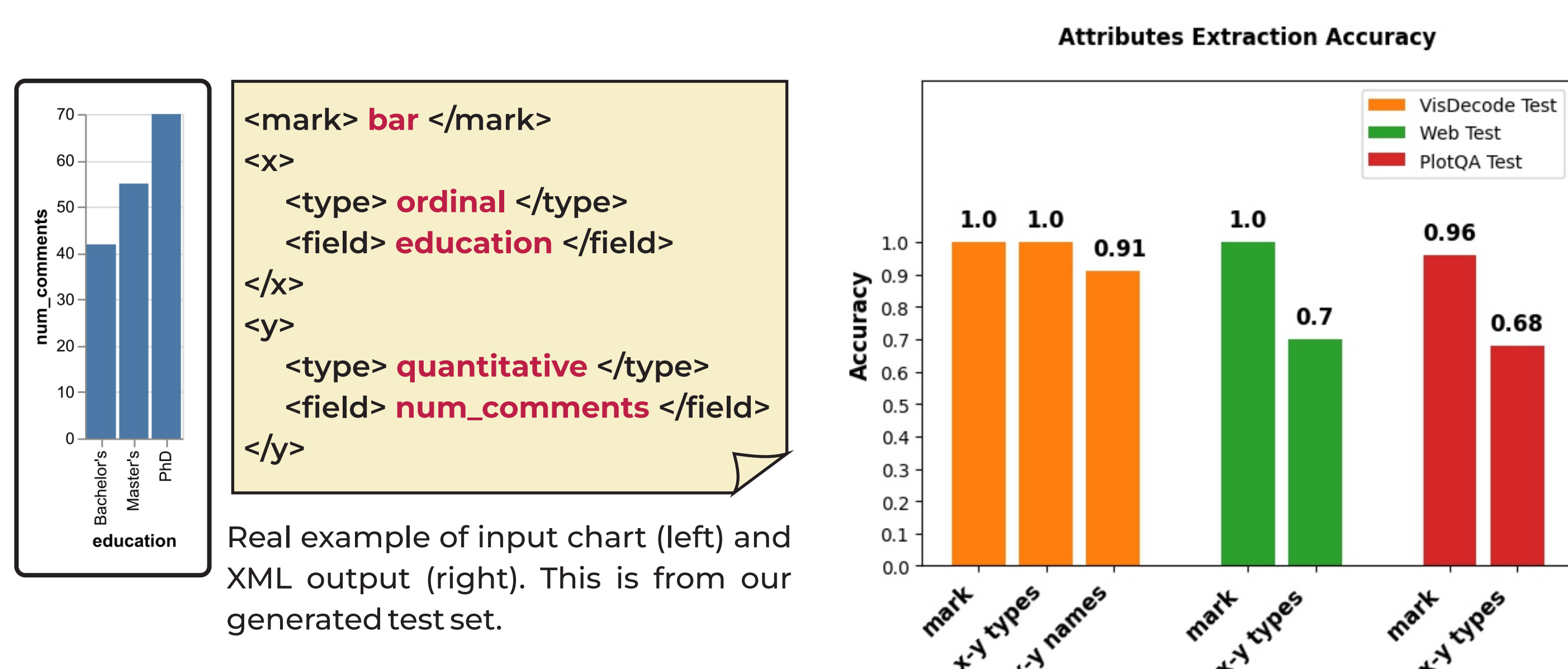
✓ Pretraining initialized from **Pix2Struct**.

✓ Aims to **plot deconstruction** and **numerical reasoning**.



Results

We use **Levenshtein Normalized Distance Accuracy** for variable names extraction and **Exact Match Accuracy** for the others.

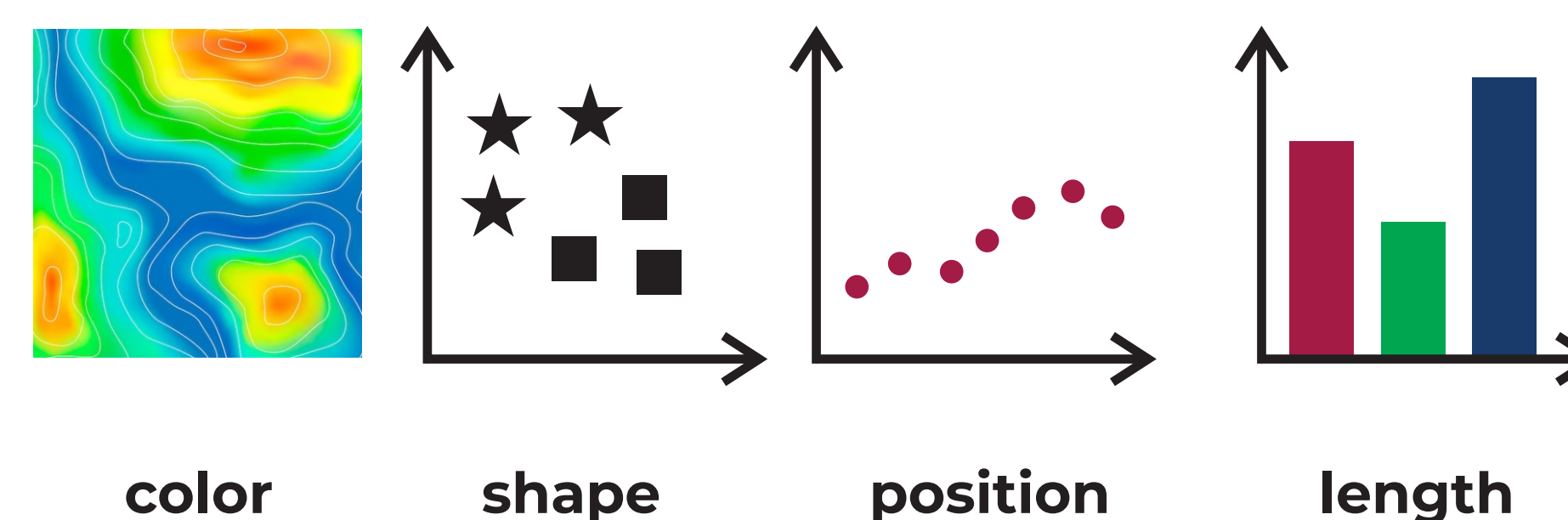


References

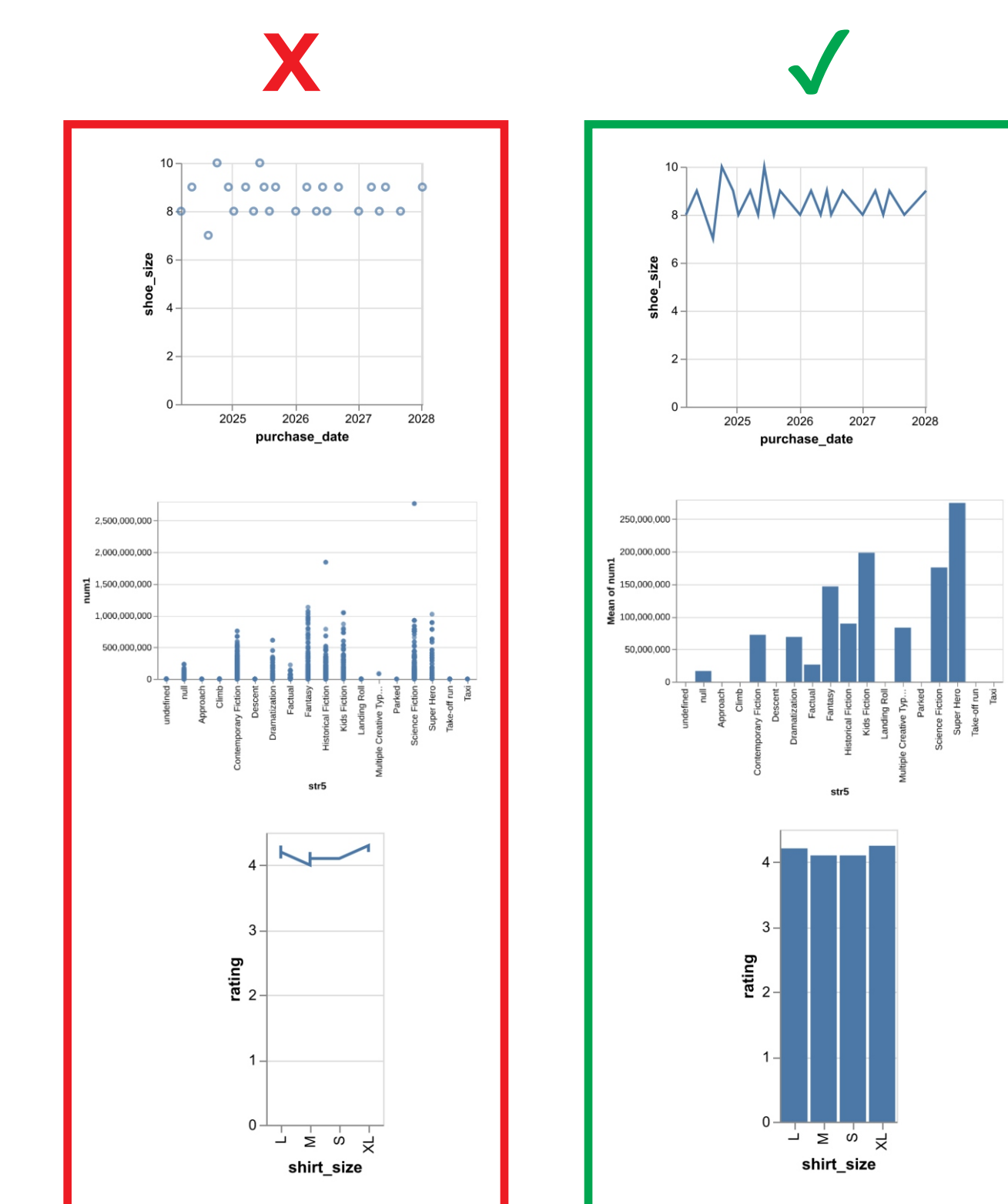
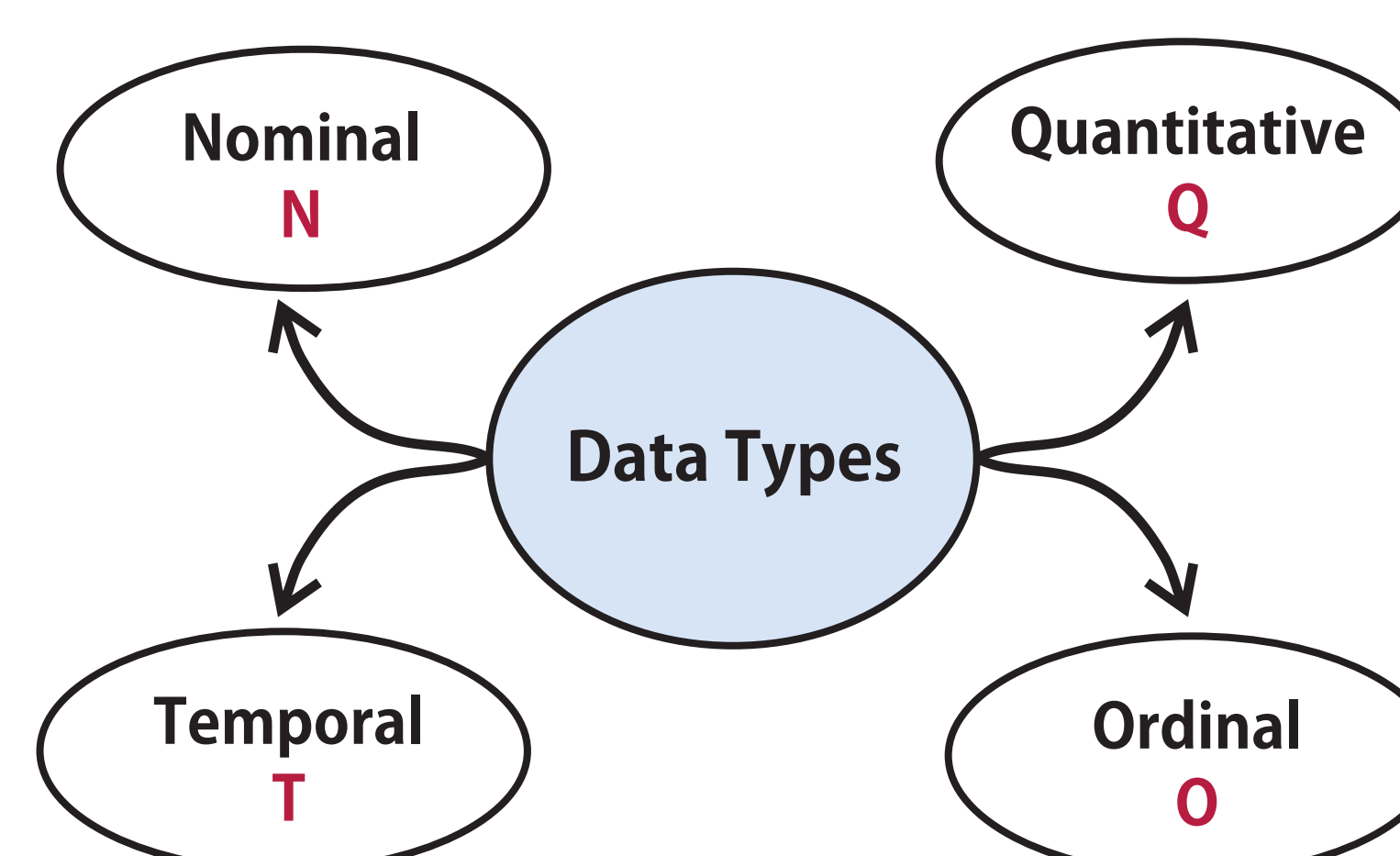
- [1] Liu, F., Piccinno, F., Krichene, S., Pang, C., Lee, K., Joshi, M., ... & Eisenchlos, J. M. (2022). MatCha: Enhancing visual language pretraining with math reasoning and chart derendering.
- [2] Lee, K., Joshi, M., Turc, I. R., Hu, H., Liu, F., Eisenchlos, J. M., ... & Toutanova, K. (2023, July). Pix2struct: Screenshot parsing as pretraining for visual language understanding. In International Conference on Machine Learning.
- [3] Edward R. Tufte, The Visual Display of Quantitative Information.

Good Design Practices

We can use many **visual attributes** to explain data. [3]



We propose to divide and classify **Variable Types** such as:



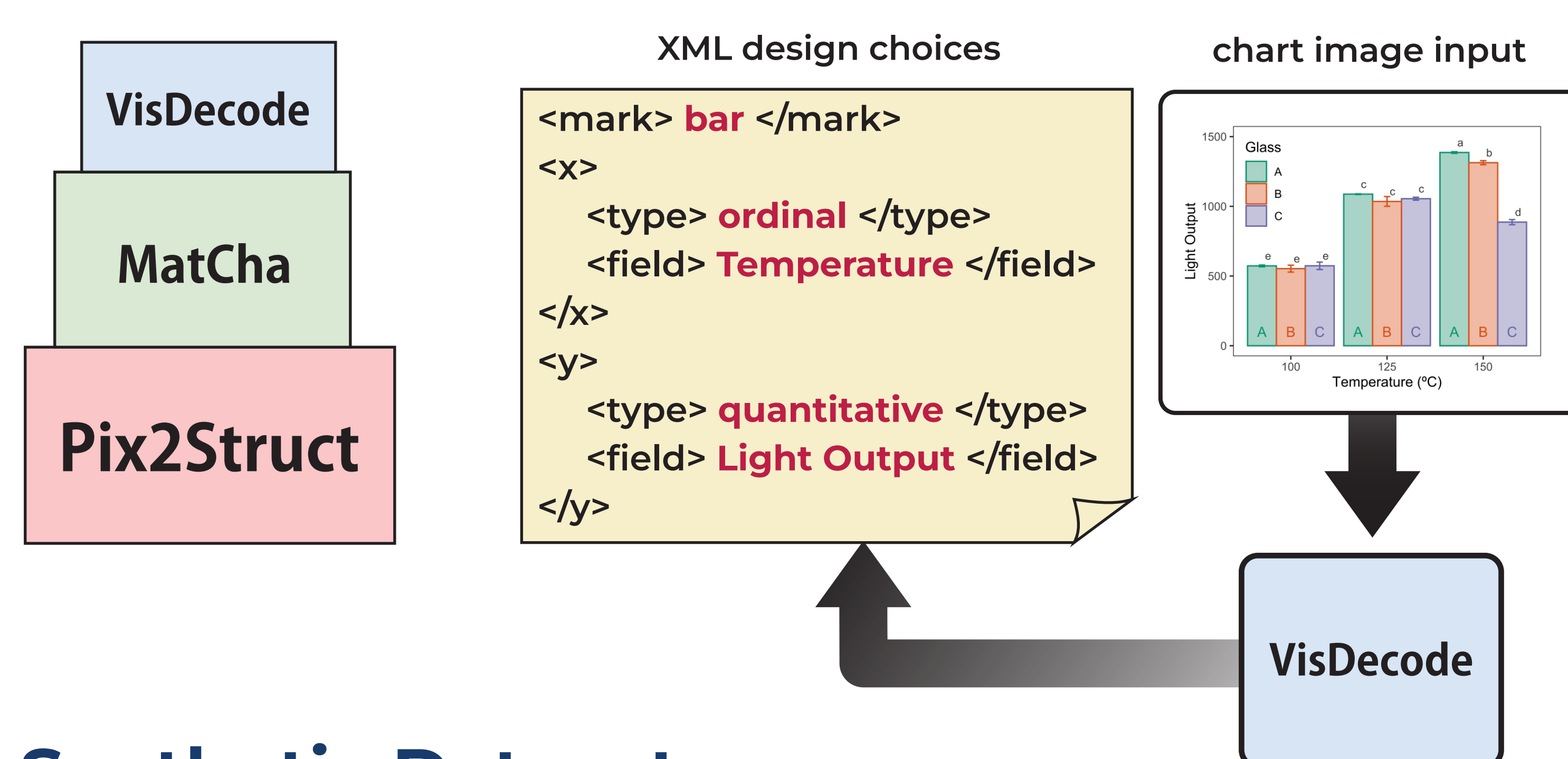
We predict the best suited **chart** based on **variable types**.

	Q	T	N	O
Q	scatter plot	line plot	bar plot	bar plot
T	line plot			
N	bar plot			
O	bar plot			

VisDecode

MatCha [1] is the initial point for **VisDecode** pretraining.

✓ Trained on a **synthetic dataset** comprising **visualization-design choices** pairs.



Synthetic Dataset

✓ Around **18k** unique generated visualizations.

✓ Were rasterized from randomly generated **JSON datasets** using **GPT-3.5** featuring both **quantitative** and **categorical** variables.

