Cartography M.Sc.

Master thesis

Mapping the Relationships among Data Visualization Types

Bo-Yuan Chen

Supervisor: Dr. Yuri Engelhardt

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITY OF TWENTE.

2025

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"Mapping the Relationships of Data Visualization Types"

Artificial Intelligence (AI) tools were used only in a secretarial/assistant role, specifically for:

- supporting coding in searching for relevant literature references,
- helping to refine initial versions of some difficult-to-organize paragraphs (which were subsequently reviewed and revised by myself),
- grammar and language corrections,
- technical assistance in plotting some of the figures,
- organization of the textual feedback from online survey.

All research ideas, arguments, analyses, and conclusions presented in this thesis are entirely my own work. I have fully referenced the ideas and work of others, whether published or unpublished. All literal or analogous citations are clearly marked as such.

Enschede, 08.29.2025 Bo-Yuan, Chen

ABSTRACT

Understanding the relationships among various data visualization types is essential for both beginners and experts, yet few representations of such relationships exist to support this understanding. This study focuses on the relationships with respect to visual encoding, that is, the ways in which data are visualized. We first explore the relationships among visualization types in a spatial way via a force-directed graph, leading to three representations—a Node-Link Diagram, a redesigned Euler Diagram, and an UpSet plot—to facilitate comprehension. Second, user testing was conducted to evaluate understanding. Results indicate that "Positioning along a coordinate axis," "Proportional space-filling," "Extending along a coordinate axis," "Connecting," and "Sizing" are the most frequently used encodings. Each representation has its advantages and limitations: The Node-Link Diagram is familiar but can become cluttered, the redesigned Euler Diagram is visually appealing but may become overloaded with information, and the UpSet plot offers detailed insights but necessitates user training. This study provides spatially explicit and more comprehensive representations of relationships among visualization types, offering guidance for beginners and insights for experts in visualization design and education.

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to my supervisor, Dr. Yuri Engelhardt, who showed great patience during the early stages of our collaboration, provided honest and professional feedback on my study, and consistently demonstrated passion for both his students and my thesis topic. I truly enjoyed our discussions on mutual interests and greatly benefited from his youthful enthusiasm and motivating spirit.

I am also very grateful to everyone who participated in my online survey, the majority of whom were my friends in Taiwan and classmates from the same intake of the Erasmus Mundus Programme, as well as people I had never met before, who still provided valuable perspectives that I would not have discovered on my own, allowing me to complete the thesis from a more holistic perspective.

In addition to the study, I would like to give special thanks to the programme coordinator, Juliane Cron, who, despite being always busy with many matters, was still willing to provide both bureaucratic and administrative support to help us settle in. I would also like to thank the local coordinators of the other three universities for their assistance.

I would also like to thank the Selection Committee of the M.Sc. Cartography programme for fully funding my two-year study abroad in Europe, which not only made this research possible but also allowed me to make wonderful international friends from diverse backgrounds, explore different cultures, and further develop my technical skills.

Last but not least, I thank God for making all these seemingly impossible things possible, which could not have been achieved by my effort alone.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Visualization was born out of human needs. Since historic times, paintings were used to capture memorable yet sophisticated events or phenomena in simplified figures, while symbols were created for counting in the rising commercial activities. Over time, the appearance of the division of labor and the development of technology, architecture, census, land surveys, long-distance trade and even military operations gave rise to the invention of maps.

Data visualization, which mainly followed the emergence of statistics, did not come into place until after the Scientific Revolution in the 18th century. Since then, various data visualization types (may be abbreviated as visualization type in this thesis) serving distinctive purposes thrived, from the earliest basic bar chart, pie chart to today's word clouds, treemaps, sophisticated composite charts, or some specific chart types used in certain fields like the Skew-T diagram in atmospheric science.

Surprisingly, the systematic analysis of data visualization did not begin until the second half of the 20th century. Bertin, in his book *Semiology of Graphics*, was the first to investigate the composition of existing two-dimensional static visualization types in terms of the data types they used. He also discussed the effectiveness of each visual encoding based on human perception — a concept describing different ways to transform data into visual components — which he referred to as 'visual variables', based on his own experience (Bertin, 1983). Since then, research in the field of data visualization has flourished.

This article aims to organize and present different types of data visualization, using current insights from data visualization analysis, through three distinct representation designs. To organize these visualizations in a logical and meaningful manner—for example, determining the relative positions of a word cloud and a treemap, or how to include additional visualization types—a concept called the 'relationship of data visualization types' is introduced. In this thesis, this concept describes how many features are shared between two visualization types.

To build background knowledge on the relationships among different data visualization types—and to clarify why this topic warrants further exploration—relevant previous works are reviewed. Bertin's pioneering contributions are discussed in detail in Section 1.1. To verify and extend his work, subsequent research has primarily focused on analyzing the effectiveness of visual encoding techniques and their connection to data types in greater depth and from multiple perspectives. However, these efforts have largely remained confined to the so-called 'low-level' components (or variables) of graphical representation and human perception. This means, for example, researchers at that time paid more attention to which kinds of visual encodings were better perceived with certain data types, rather than tackling data, visualization types, or even real-life tasks, which belong to 'higher-level' components. Studies addressing these higher-level aspects only emerged decades later. Several classic papers in this area are briefly summarized in Section 1.2. Drawing from this body of fundamental and scientific knowledge, various attempts have been made to organize the visualization types according to different principles. These include guidelines and representations for novice visual designers and for scientific purposes, which are discussed in Section 1.3. Finally, Section 1.4 outlines the research motivation, objectives, and key research questions.

1.1. The pioneering work of Bertin in Data Visualization

"Semiology of Graphics" represents the first systematic analysis of graphical representations, focusing on spatial configurations and visual encodings. In this book, Bertin identified eight 'visual variables.' Two of them belong to the category of 'planar variables' (which he termed 'imposition'), referring to the two axes available on a plane, such as paper. These were further differentiated according to the orientation of the axes (e.g., Cartesian coordinates, polar coordinates, etc.). The remaining six belong to retinal variables, a metaphor representing a third dimension perpendicular to the plane. These include size, value, texture, color, symbol, and orientation. Bertin then examined how each of these variables performed perceptually when applied to different data types—namely quantitative, ordinal, qualitatively selective (the ability to focus on one specific category), and qualitatively associated (the ability to group several subcategories).

Beyond these so-called 'low-level components of visualization, Bertin also proposed a classification of visualization types themselves into four groups: diagrams, networks, maps, and symbols. The classification is based on where the meaning of the visualization originates. For example, the meaning of a diagram arises from the relationships between two different datasets, whereas that of a network emerges from relationships within the same dataset. (Bertin, 1983)

1.2. The Two Aspects of Data Visualization

As Bertin implied, data visualization is not just about converting data into graphs, but one step further: It is also about effectively conveying the messages from data via graphs to humans. The interactions between data visualization, data type, and human perception have been long and heavily discussed since Bertin's time, with even the appearance of some mature commercial software recommending suggested visualization types automatically (Mackinlay et al., 2007). On top of that, Shneiderman, in his classic work, emphasized the importance of incorporating user tasks into the design of data visualizations (Shneiderman, 1996). Building upon this, Amar and Stasko further argued that visual design should serve higher-level analytical purposes, namely, "higher-level" tasks, such as decision-making and insight generation (Amar & Stasko, 2005). From that point onward, two primary lines of research in data visualization began to take shape.

1.2.1 Research based on Data and Human Perception

Research in this perspective mostly focuses on either the establishment of converting given data type into the configuration of graph, preferably automatically, or the effectiveness of different visual encodings concerning different perception tasks given by the researchers.

Building on Bertin's work on the perceptual effectiveness of the eight visual variables, Mackinlay proposed an automatic method for selecting effective visualization types by choosing visual encodings according to data type and perceptual principles (Mackinlay, 1986). This approach is also regarded as a precursor to modern visualization recommendation systems. Similarly, following Bertin's four categories of data types and visual variables, Card and Mackinlay formulated a transformation table that systematically relates data types to the corresponding visual encodings—and vice versa—which can be applied both in figure creation and in analytical contexts (Card & Mackinlay, 1997). More recently, Engelhardt and Richards extended Bertin's set of visual variables by introducing the Languages of Visualization system (abbreviated as LangVIS system), a visual language system that identifies and employs more comprehensive "Building Blocks." These building blocks encompass three aspects—data type questions, visual encodings, and axis orientations—

allowing visualizations to be decomposed and analyzed in terms of their connections to data and potential tasks (Engelhardt & Richards, 2024).

In contrast, several studies concerning human perception have examined how different low-level design choices affect user performance across various tasks. Cleveland and McGill's classic study examined how accurately people perceive different visual encodings. Based on the findings, he proposed a ranking to help designers prioritize visual encodings when accuracy is critical in visualization design (Cleveland & McGill, 1984). Simkin and Hastie extended their work and investigated how people perform various low-level tasks depending on the type of visual encoding presented (Simkin & Hastie, 1987). Compared to the previous analysis or classification framework, Rodrigues et al. proposed a novel framework for data visualization that emphasizes and integrates human perception to analyze visualization techniques (Rodrigues et al., 2007).

1.2.2 Research based on Task

Perhaps the most influential research in this area is the study by Saket et al., which systematically examined the effectiveness of five basic visualization types—bar chart, line chart, pie chart, scatterplot, and table—across three dimensions: time, accuracy, and user preference. The study revealed several key insights. For instance, bar charts were found to be most effective for identifying clusters, whereas line charts were more suitable for analyzing correlations (Saket et al., 2019). Although this aspect of studies provides valuable and interesting insights, we do not delve into its details here, as its specific focus is not directly relevant to the scope of our research, and this aspect of the field remains relatively immature and under development. Nevertheless, a recent systematic review by Quadri and Rosen provides a comprehensive overview of the current studies in this area, offering a broader understanding of the topic (Quadri & Rosen, 2022).

1.3. Representations of Visualization Type Classifications

Besides research on the 'analytical approach' to data visualization, which examines the essence of visualization, many studies have also attempted to construct taxonomies (Rodrigues et al., 2007), sometimes accompanied by representations to show the relationships between visualization types. Interestingly, most of these are not found in scientific literature. Rather, their classifications—a simplified form of taxonomy—are primarily intended as design guidelines and representations for novice designers and often rely less on strict scientific rigor. This implies that such classifications simplify the overwhelming number of visualization options. Although these classifications are often more experiential, our focus is on the design decisions they employ in representing classifications to distinguish between different visualization types while keeping their guidelines readable and practical.

Three common perspectives for classifying visualization types align with the research mentioned above: the input data type (data type), the dominant shape (visual encoding), and the function the visualization serves (task). However, there are also two scientific literatures that aim to organize various visualization types with strict scientific rigor

It is important to note that, unlike in machine learning—where classification aims to automate image categorization—the classification of visualization types here is based on explicit principles.

1.3.1 Classifications and Representations based on Data Type

Four sources we found online classify visualization types based on the nature of the data input. One of the most important attempts to lay out and categorize a wide range of visualization types is *A Tour through the Visualization Zoo*, which groups 18 popular visualization types into five categories: Hierarchies, Maps, Networks, Statistical Distributions, and Time-Series Data (Heer et al., 2010). The *Which Visualization? A Quick Reference* by Franconeri (See Figure 1) presents a table-like poster that guides designers by recommending the visualization types (Franconeri, 2019) based on the data type they have (e.g., discrete category, ordered category, or continuous metrics, whether representing absolute numbers or proportions). *From Data to Viz* (See Figure 2) also provides a poster-based guide. Instead of simply grouping by data types, it organizes the classification into six decision trees: Categoric, Categoric and Numeric, Map, Numeric, Relational, and Time Series. Once users select the relevant branch, it expands further based on the number and characteristics of the dataset to be visualized (Holtz & Healy, 2018). The *Data Viz Project* is a website that collects many visualization types, aiming at inspiring other visualization designers. It classifies visualizations across three aspects. Regarding data type (which they refer to as "input"), sample data types are illustrated in matrices. Users can select the data structure they are working with and be directed to a corresponding set of suitable visualization options (Ferdio, 2025).

1.3.2 Classifications and Representations based on Dominant Shape

Two notable websites classify visualizations by their dominant shape. The *Data Viz Project* includes a "Shape" tab that organizes visualization types into ten general categories: Area, Bar, Circle, Dot, Icon, Line, Map, Pyramid, and Square (Ferdio, 2025). *Graphopedia*, another comprehensive resource, serves a similar purpose but focuses specifically on identifying the visualization techniques employed by each chart type, rather than grouping them by shape (Graphopedia, n.d.). Its approach draws extensively on the LangVIS system (Engelhardt & Richards, 2024).

1.3.3 Classifications and Representations based on Function

Both the *Data Viz Project* and *Graphopedia* provide a function-based classification. Similar to previous methods, the former uses a 'Function' tab to group visualizations into seven categories: Comparison, Concept Visualization, Correlation, Distribution, Geolocation, Part-to-Whole, and Trend over Time (Ferdio, 2025). The latter treats these functions more as descriptive attributes of each visualization type (Graphopedia, n.d.). Beyond its main classification by data type, *From Data to Viz* also uses color to indicate the intended function of each visualization—such as Correlation, Distribution, Evolution, Flow, Maps, Ranking, and Part-to-Whole (Holtz & Healy, 2018). The *Visual Vocabulary*, designed by the Financial Times (See Figure 3), organizes visualization types according to their functionality in a list format: Change over Time, Correlation, Deviation, Distribution, Flow, Magnitude, Part-to-Whole, Ranking, and Spatial (Times, 2016-2019). The *Graphic Continuum* (See Figure 4) adopts a spatial reasoning approach by laying out five major functional categories—Comparing Categories, Distribution, Time, Geospatial, Part-to-Whole, and Relationship—while visually connecting related types with lines (Ribecca, 2025). More recently, *A Friendly Guide to Choosing a Chart Type* presents a storytelling-oriented guideline with six categories: Absolute Values, Correlations, Developments Over Time, Flows, Maps, and Shares (Schwab, 2025).

1.3.4 Scientific Classifications and Representations

In the scientific field, there is increasing awareness of the importance of organizing a comprehensive classification or representation of various data visualization types (Oliveira et al., 2017). One of the earliest attempts is the well-known *Periodic Table of Visualization Methods* (See Figure 5), developed to support the selection of appropriate visualization types, particularly in management contexts (Lengler & Eppler, 2007). This concept was later refined through a treemap-based interface, which introduced interactive features allowing users to filter out irrelevant visualization types—thereby enhancing usability and making the system more user-centered (Oliveira et al., 2017).

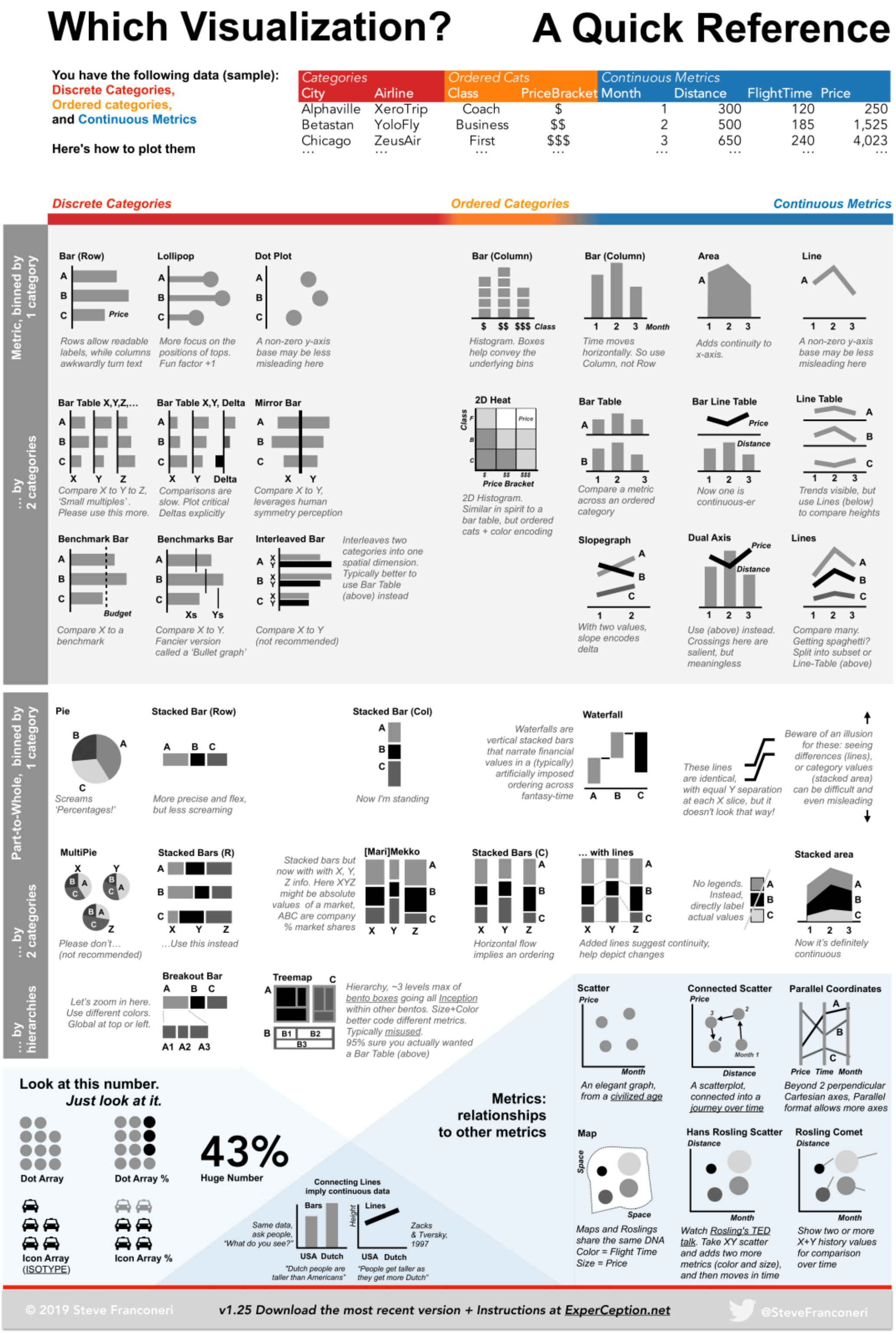

Figure 1: *Which Visualization? A Quick Reference*, a representation of visualization types categorized by adopted data type.

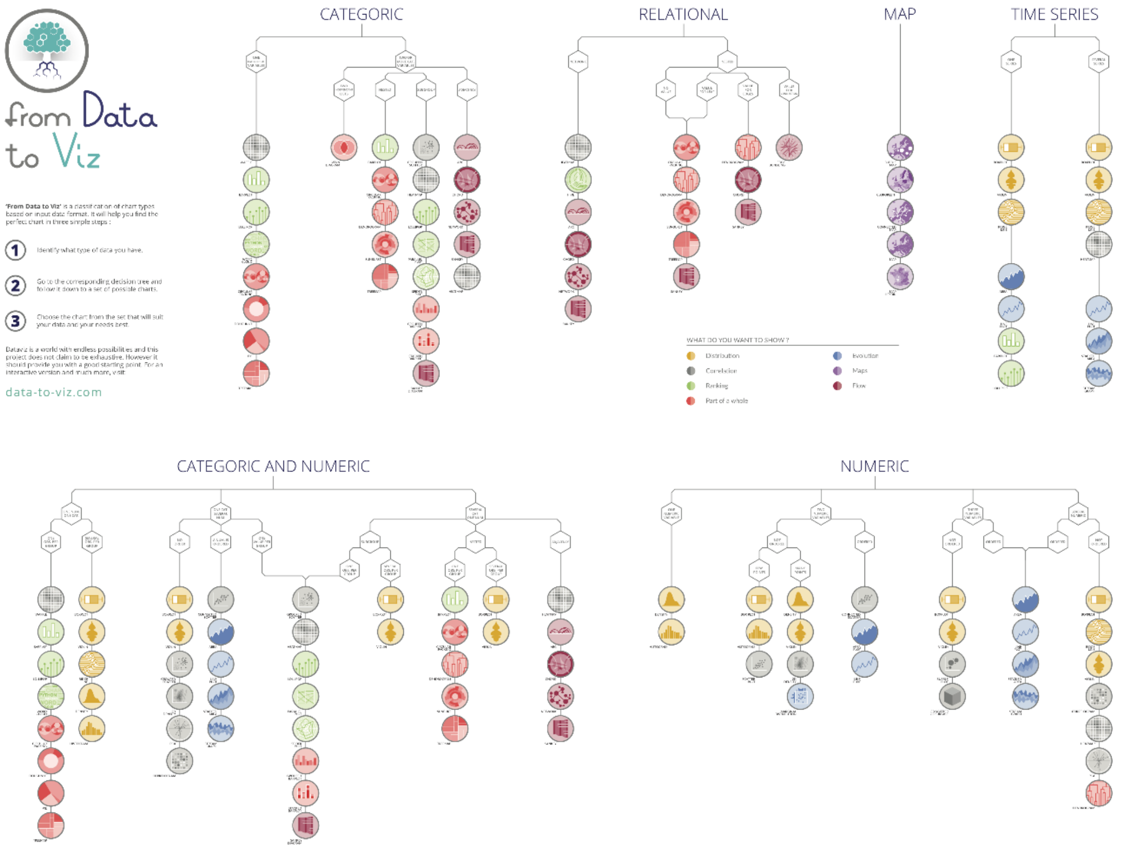

Figure 2: *From Data to Viz*, a representation of visualization types categorized by adopted data type and served function.

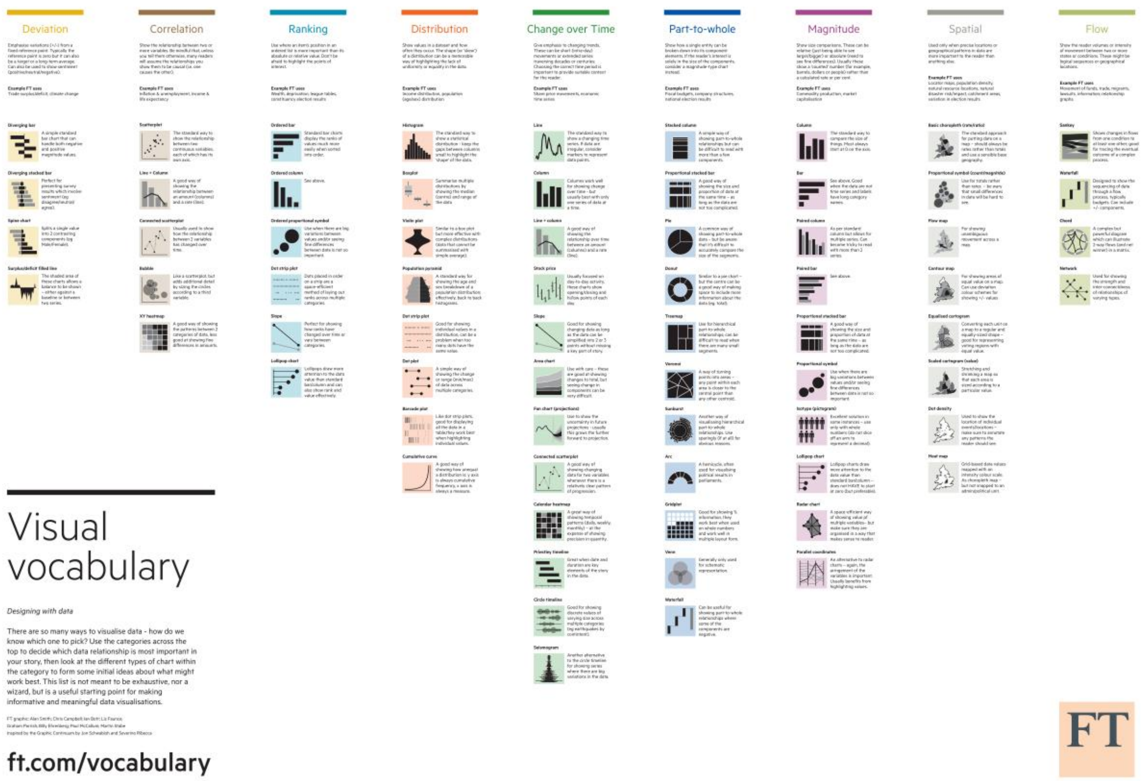

Figure 3: *Visual Vocabulary*, a representation of visualization types categorized by served function.

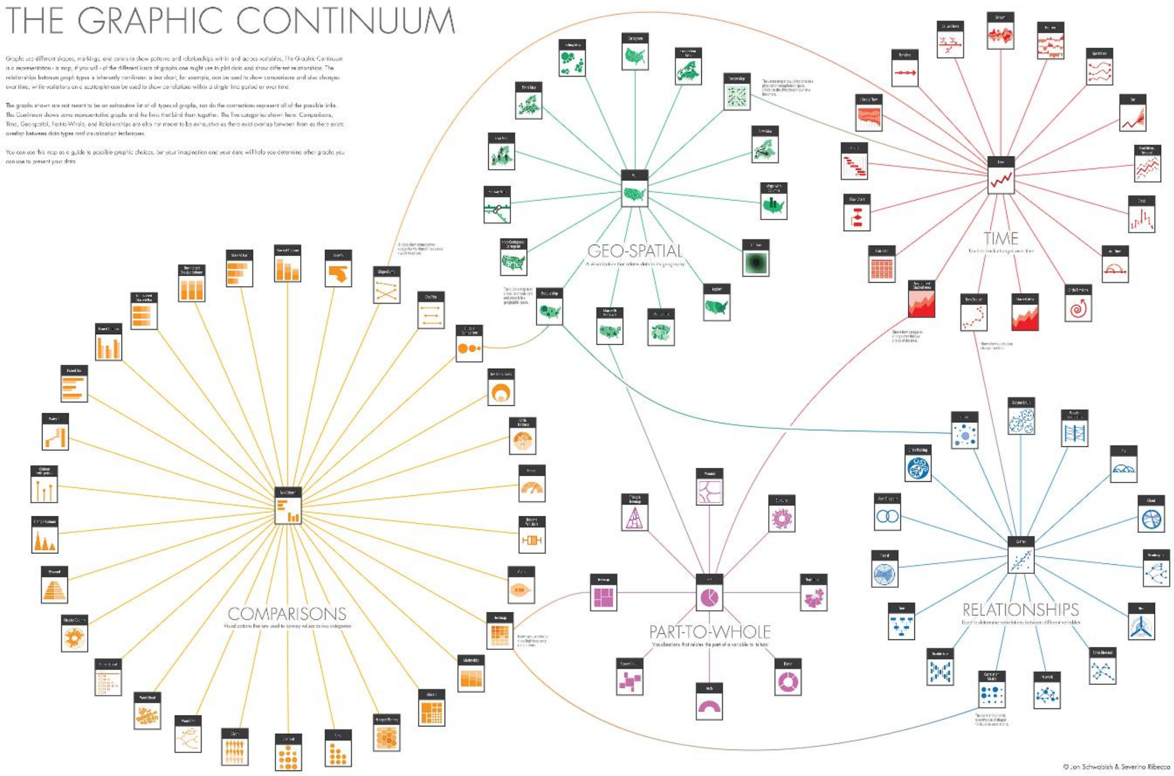

Figure 4: *Graphic Continuum*, a representation of visualization types categorized by served function.

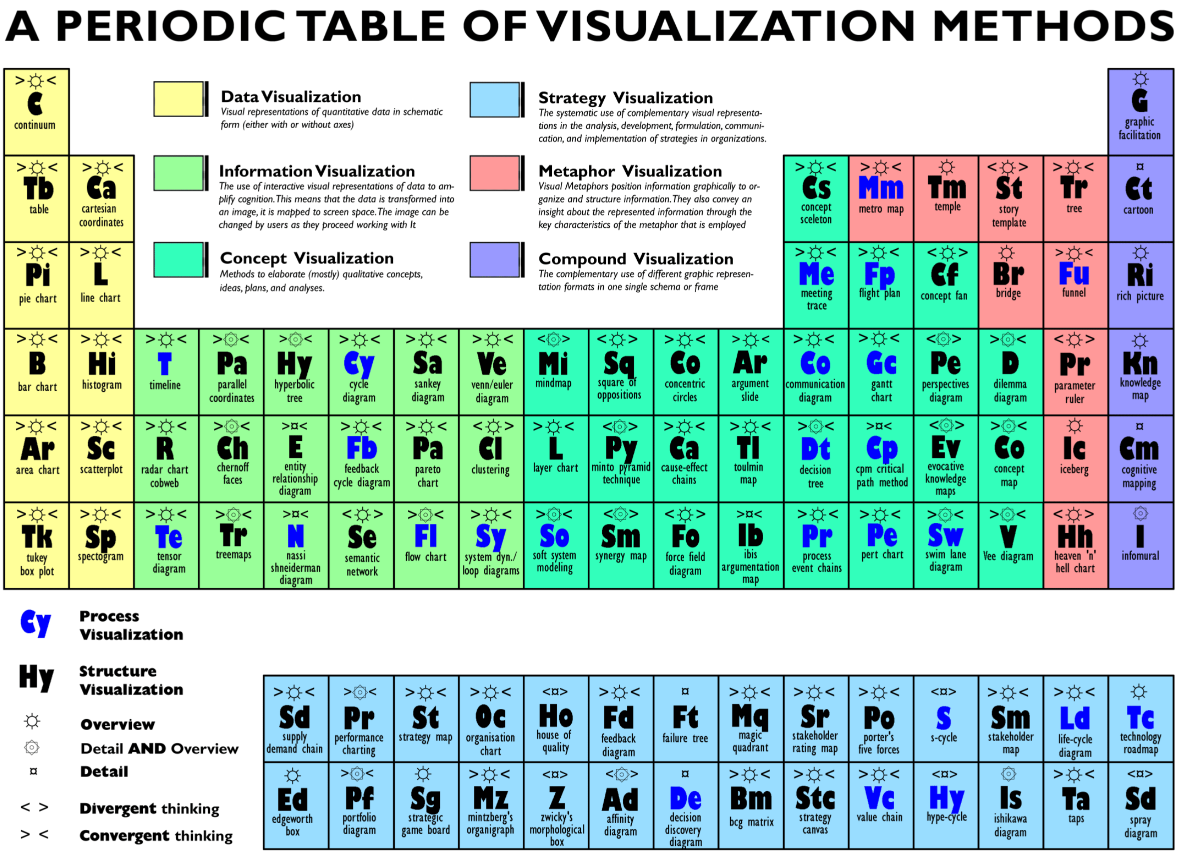

Figure 5: *A Periodic Table of Visualization Methods*, a representation of visualization types in management contexts.

1.4. Research Background

1.4.1 Motivation

Although a significant amount of research has been conducted in the field of data visualization, relatively little attention has been given to the representation of visualization types in the scientific literature, as introduced in the previous section. The analytical studies reviewed in Section 1.2 mainly focus on specific visual encodings, a limited subset of visualization types, or structure them in text or tables, rather than analyzing the majority of the visualization types in a visual manner. Meanwhile, the taxonomies presented in Section 1.3 prioritize user-friendly guidelines, often covering only a limited range of visualization types and lacking scientific rigor.

Preliminary attempts in the scientific field—such as periodic table-like visualizations—have mainly focused on management contexts rather than on visualization types themselves, and have been heavily criticized for their weak spatial metaphor compared to the original periodic table in chemistry (Few, 2007; Kosara, 2009). Subsequent attempts, including those using treemap-based visualizations, were largely built upon this earlier approach.

These observations highlight a gap in the literature: there is a need for scientifically grounded representations that can systematically capture the relationships among visualization types, particularly with respect to visual encoding techniques. Addressing this gap could help data visualization designers better understand, compare, and draw inspiration from the overwhelming variety of visualization options available today.

1.4.2 Research Objectives

This research aims to construct spatial representations of data visualization types with a focus on visual encoding, inspired by the successful model of the periodic table in chemistry. The main objective is to map and design the structure of relationships among various visualization types through spatial reasoning. Such representations are intended to provide an intuitive "big picture" of how visualization types are distributed and interconnected according to their visual encoding techniques. In turn, this enables a better understanding of which kinds of representation designs enhance comprehension, while also allowing insights to be drawn from the mere observation of these relational structures.

To achieve this, this can be broken into two sub-objectives:

1. To create representations that boost our understanding of the relationships between visualization types and their visual encodings.
2. Evaluate both user comprehension and preferences, as well as the outcomes of these representations.

This research targets two groups: (i) supporting novice designers in understanding various visualization types, and (ii) offering insights for researchers interested in visualization grammar and complex spatial reasoning design.

This study focuses on static 2D data visualization types, excluding 3D graphs, animations, interactive visualizations, and complex composite visualizations, in line with Bertin's limitation of scope (Bertin, 1983). Although factors such as data input type, aesthetics, target audience, specific tasks, and domain-specific contexts also play important roles in shaping different types of visualizations (Kandogan & Lee, 2016), we temporarily exclude these for simplification. These factors, however, can certainly be incorporated in future work.

1.4.3 Research Questions

To address these two objectives, the following three research questions are formulated:

RQ1: Structure Aspect
How can the relationships between visualization types and their visual encodings be systematically represented?

RQ2. Design Aspect
Which design approaches are most effective for supporting the understanding of the overall structure of the relationships?

RQ3: Analysis Aspect
What insights can be derived from the representations? For instance, which visual encodings or combinations of encodings are most frequently employed across visualization types?

1.4.4 Research Section Layout

Chapter 2 outlines the research methodology. Chapter 3 presents and discusses the results of user testing. Chapter 4 presents the discussion of the findings. Chapter 5 concludes the study, highlights its contributions and limitations, and outlines possible directions for future work.

2. METHODOLOGY

To construct our methodology, we adopt the classic Nested Model for Visualization Design and Validation (Munzner, 2009) which consists of four levels: Domain Problem Characterization, Data and Task Abstraction, Visual Encoding and Interaction Design, and Algorithm Design. This model provides a structured approach to designing and evaluating visualization systems.

- **Domain Problem (Section 2.1):** This level addresses the target user group and their needs. In our case, we highlight the lack of effective representations for understanding the relationships among different data visualization types and the visual encodings they employ.
- **Abstraction Design (Section 2.2):** At this level, we construct the skeleton of the relationships between visualization types and visual encodings, laying the foundation for the representation designs described in the following section. We justify the selection of visualization types and visual encoding schemas and explain how these relationships are organized to form a preliminary relational structure.
- **Visual Encoding (Section 2.3):** Here, we develop three distinct static visual representations based on the relational structure constructed in Section 2.2, each reflecting different design decisions. These representations are evaluated through an online user study to assess how effectively each supports user understanding. The design of the user study is described in Section 2.4.
- **Algorithm Design:** The underlying structure was implemented through coding via Python, which will be mentioned within the description part of the structuring and designing phases. However, since our study uses static images without interactive features, interaction design is not included in the scope of this research.

Although user testing was only conducted in the Visual Encoding section, our methodology ensures conceptual clarity by explicitly following the Nested Model framework.

2.1 The Domain Level

Our target audience includes both novices and experts in data visualization, as well as individuals who work with datasets involving complex relationships. Among these groups, novices and experts are the primary focus of this study—and where its main contribution lies. As previously noted, there is a lack of sufficient educational resources for beginners to get started (Unwin, 2020). For advanced users, this research aims to support the development of a more structured and comprehensive theory of data visualization design, an area where progress remains limited. At the same time, the methods we propose for addressing the complex relationships between various visualization types may also be beneficial to the third group.

2.2 The Abstraction Design Level

2.2.1 The Selection of the Data Visualization Types

The visualization types included in this study were selected based on two main criteria: first, their popularity; and second, the need to ensure diversity in the aspect of the visual encoding techniques used by supplementing common types with less conventional ones.

Surprisingly, only one research study has been found that focused on the popularity of the visualization types used online. By analyzing online visualization platforms and libraries, only four categories of visualization types are considered popular: bar chart, line chart, scatter plot, and geographic map. There is a sharp drop of usage beyond these four, followed by pie chart, box plot, donut chart, and area chart (Battle et al., 2018). Note that unlike the other part of this study, the category of types here treats grouped bar chart and stacked bar chart as a single bar chart category. Since only one study has been found, we also refer to the categories of types in the field of chart mining, by which the training datasets are classified. This is because, except for datasets where category sizes are artificially balanced, differences in category sizes can reflect accessibility and serve as an indicator of popularity. Surprisingly, the statistics are similar. Among the public accessible datasets, UB-PMC, DocFigure and Chart-OCR also contain more bar charts, line charts, and scatter plots in general, with pie charts as an addition, and with much smaller amounts of geographic maps, box plots, heat maps, scatter-lines, interval diagrams, Venn diagrams, tables, flowcharts, and block diagrams (Dhote et al., 2023).

To ensure the diversity of visualization types, we also include all examples listed in the LangVIS system, demonstrating how various visual encodings are applied in different visualization types (Engelhardt & Richards, 2024), which also serves our purpose.

In total, this study involves up to 60 visualization types, as shown in Table 2.

The icons for the visual encodings in this study were created by Yuri Engelhardt and Clive Richards in the LangVIS system and are licensed under CC BY-NC-SA 4.0. The icons for the visualization types were designed by Anna Vital and Mark Vital for the Graphopedia project and are used here with permission.

2.2.2 The Selection of Visual Encoding System

Besides serving as one of the sources of visualization types, the visual encodings identified in the LangVIS system (Engelhardt & Richards, 2024) are also defined. The reason is that this system systematically and rigorously identifies visual encodings that can be found in almost all data visualizations and is also extended directly from Bertin's work of classifying visual variables and adjusted for the modern context.

The LangVIS system develops 19 visual encodings to construct the grammar used in every visualization type. These encodings are further divided into three groups: arranging, linking, and varying (see Table 1). The first two groups correspond to Bertin's groups of imposition: arranging relates to diagrams, maps, and symbols, which position along the two planar axes, while linking relates to networks. The last group, varying, corresponds to retinal variables, which Bertin referred to as "the third axis".

LangVIS also discusses other aspects such as data type, the semantic structure that forms a visualization, and the axis orientations, all of which can potentially also be used in a definition of the relationships between visualization types and visual encodings. However, in this study, we only focus on the visual encoding aspect for simplicity (the exception is the concept of axis orientation, which has been integrated into the design of the visualization type icon as its outer shape, whose design idea is briefly mentioned in the caption of Table 2). For example, Figure 8 shows a connected scatter plot. In the aspect of data type it uses, it is comprised of two "HMuchs (How much and how many, namely, quantitative data)" and one "Catego (Categorical data)," while in the aspect of visual encoding, it contains two "Positionings on an axis" and one "Connecting." As for the aspect of axis orientation, it uses vertical and horizontal axes to position the data. And Figure 6 is the description of a connected scatter plot.

2.2.3 The Definition of the Visualization Type Relationships

To integrate visualization types and visual encodings, it is necessary to define the concept of "relationship" in this study. A relationship can be described at three levels:

1. Between two visualization types: This indicates how many visual encodings are shared between the two types, reflecting the similarity of design strategies applied. For example, a relationship exists between a "connected scatter plot" (Figure 8) and a "network diagram" Figure 7 through the shared visual encoding "connecting."
2. Between two visual encodings: This describes how many visualization types employ both visual encodings, highlighting which visual encodings frequently co-occur. For instance, the relationship between "connecting" and "positioning on an axis" can be observed through the visualization type "Connected Scatter Plot" (Figure 8).
3. Between a visualization type and a visual encoding: This shows how often a specific visualization type uses a particular visual encoding. For example, the "connected scatter plot" employs the visual encoding "positioning on an axis" twice, while "connecting" is used only once (Figure 8).

By distinguishing these three levels of relationships, we can systematically represent the interactions between visualization types and visual encodings, supporting both the design and analysis of visualizations.

2.2.4 The General Structure of the Relationships among Data Visualization Types

Since our goal is to decipher the complex relationships regarding visualization types and visual encodings, especially from a spatial reasoning perspective, a representation that can reflect their topology is a top priority. Network diagrams are well suited for this purpose (Munzner, 2014). We therefore chose a force-directed graph, which intuitively conveys relationship strength through spatial distance. For example, if visualization type A shares two visual encodings with type B, but type B shares only one encoding with type C, then A and B will be positioned closer together than B and C, with A and C at the greatest distance.

This spatial arrangement arises from the mechanism of force-directed layouts, which simulate physical systems. Nodes repel each other, while links attract, similar to magnetic poles. Initially, nodes are placed randomly, and their positions are iteratively adjusted based on repulsion and attraction forces until the system reaches a stable equilibrium.

To construct the force-directed graph, the meanings of nodes and links must be defined. Nodes represent either a visualization type or a visual encoding, while links exist only between a visualization type and a visual encoding. A link is created whenever a visualization type employs a visual encoding. This design is inspired by the compact structure of the Graphic Continuum (Ribecca, 2025). We also experimented with an alternative design in which only visualization types served as nodes and links indicated shared visual encodings between two types. However, as the number of nodes and links grew — along with the frequency of link crossings — the layout became unstable, causing the force-directed iteration to fail to converge and enter an infinite loop.

We also account for cases where a visualization type uses the same visual encoding multiple times (the relationship between a visualization type and a visual encoding). In this representation, the width of a link denotes the frequency of usage, with wider links indicating more repetitions. The resulting structure is illustrated in Figure 9.

We used Pyvis, an open-source Python library that wraps the JavaScript-based VisJS engine, to visualize force-directed graphs in an interactive web-based environment.

Scatter plot	Range chart	Bar chart	Flow chart	Gauge chart	Proportional symbol map
Shape-coded scatter plot	Dumbbell chart	Ordered bar chart	Vertical tree diagram	Clock face	Pictorial map
Rosling bubble chart	Gantt chart	Grouped bar chart	Bump chart	Spiral timeline	Quantitative flow map
Small multiple of scatter plot	Range area chart	Stacked bar chart	Arc diagram	Radar chart	Connection map
Line plot	Box plot	100% stacked bar chart	Proportional area chart	Circular bar chart	Gradient scale map
Connected scatter plot	Streamgraph	Variable width bar chart	Dot plot	Polar bar chart	Network diagram
Parallel coordinates	Pyramid diagram	Area chart	Isotype chart	Radial area chart	Venn diagram
Icicle chart	Population pyramid	Stacked area chart	Warming stripes	Sunburst chart	Nested circle packing
Heatmap	Waffle chart	100% stacked area chart	Pie chart	Picture	Proportional Venn diagram
Marimekko chart	Treemap	Table	Doughnut chart	Area cartogram	Word clouds

Table 1: List of Visualization Types with their corresponding icons involved in this study. The outer shape of each icons represents the underlying structural or spatial logic used in different types of visualizations. Specifically: ■Square: Rectangular (Cartesian) coordinates ●Circle: Polar coordinates ⬢Octagon: Visual-spatial resemblance (e.g., geographic or picture) ⬟Pentagon: Topological relationships (e.g., networks, clusters)

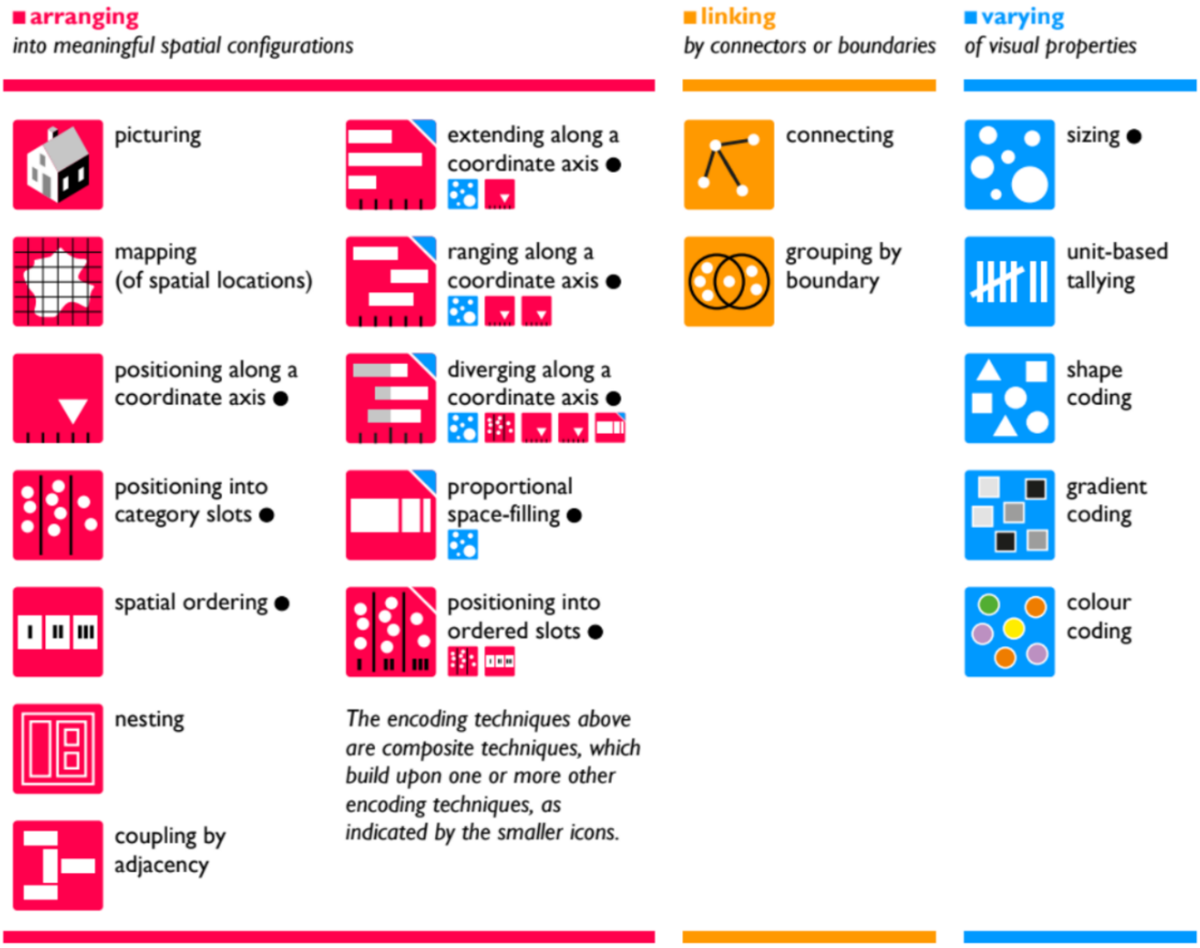

Table 2: Identified Visual Encodings in the LangVIS system.

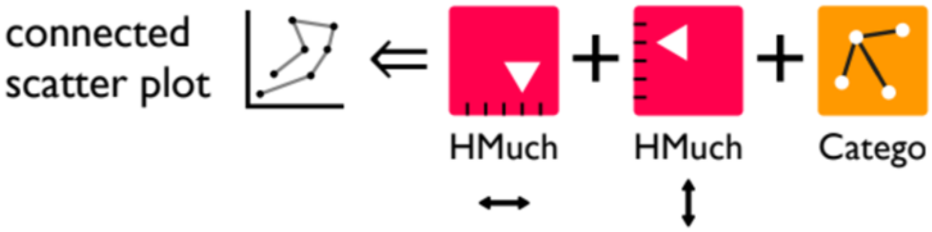

Figure 8: An example analyzed by the LangVIS system.

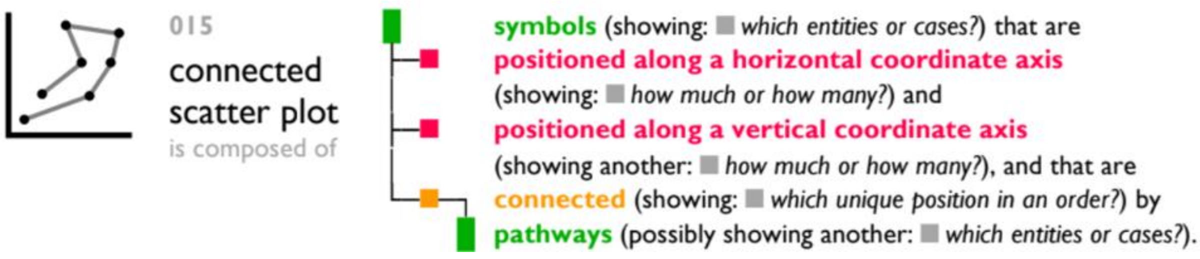

Figure 6: A description example analyzed by the LangVIS system.

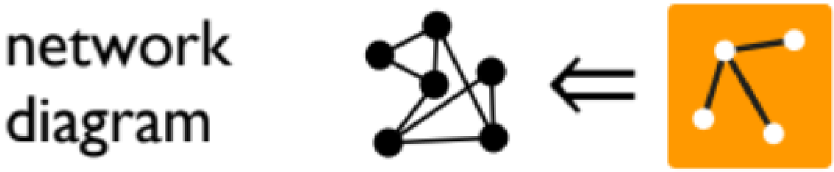

Figure 7.: Another example analyzed by the LangVIS system.

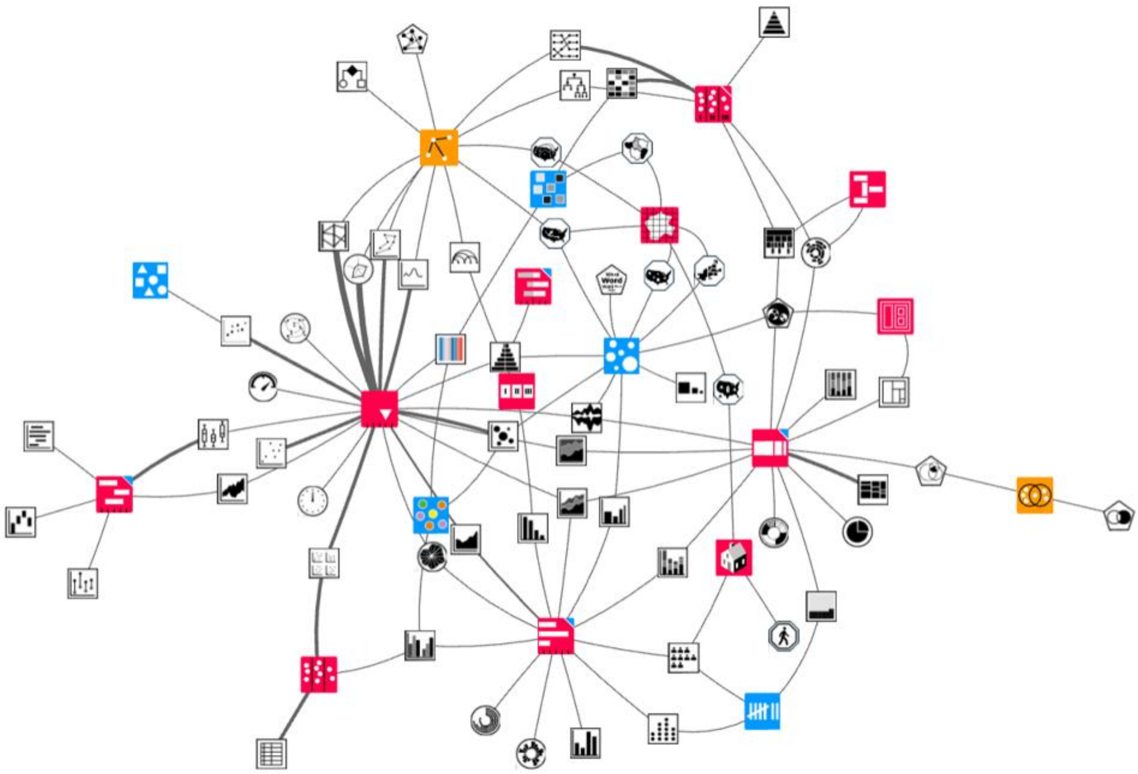

Figure 9: A force-directed graph representing the relationships among data visualization types.

2.3 The Visual Encoding Level

2.3.1 Three Representations of the Relationships among Data Visualization Types

Though the relative positions of individual visualization types and visual encodings are determined, several drawbacks can be easily observed in the automatically generated result. Notably, overlapping occurs not only in the links but also in the icons, which may lead to confusion and highlights the need for redesign.

One potential improvement is to show relationships by grouping elements within shared boundaries—such as Euler Diagrams—instead of connecting them with lines. Other visualization techniques suitable for representing relational datasets have also been organized (Alsallakh et al., 2014), including overlap-based, matrix-based, aggregation-based, and scatterplot-based approaches.

In this study, we selected three visualization design directions: the Node-Link Diagram, the redesigned Euler Diagram, and a hybrid of matrix-based and aggregation-based techniques called *UpSet plot* (Lex et al., 2014). The reasons for this selection are twofold.

First, these techniques are well-suited to our data and goals:

1. Node-link and Euler Diagrams are generally familiar to users.
2. Matrix-based and aggregation-based techniques, such as UpSet, are effective for large and complex relational datasets like ours.

Second, we excluded certain approaches due to mismatches with our data structure:

3. Our dataset is not of secondary importance to other contextual information, making overlap-based techniques inappropriate.
4. The topology of our data does not encode real spatial distance, which limits the applicability of scatterplot-based techniques.

2.3.2 The Node-Link Diagram

The first representation of the relationships among data visualization types uses “edges” to link related visual encodings and visualization types. This approach is based on the previously constructed force-directed graph, with only a few adjustments applied before finalization (See Figure 10)

To preserve the overall topology as much as possible, only four refinements were implemented:

1. Reducing line crossings: To improve readability, minimizing edge crossings was prioritized, even if it required slight compromises in the original topology, consistent with prior research (Purchase, 1997).
2. Edge reduction: When two visualization types share the exact same set of visual encodings, the nodes were merged and the types were positioned adjacently to simplify the network.
3. Representing multiple uses: Rather than varying edge width, double or triple edges indicate cases where a single visualization type employs the same visual encoding multiple times.
4. Label placement: Abbreviated names of visual encodings are positioned below their respective icons to enhance clarity, with a legend linking each abbreviation to its full name, as listed in Table 1.

2.3.3 The redesigned Euler Diagram

The second representation uses "grouping by boundary" to illustrate the relationships between related visual encodings and visualization types.

Before describing the design, the terminology must be clarified. Consider a simple classic Venn diagram with two elements, A and B (a Venn diagram is a special case of a redesigned Euler Diagram in which every possible intersection must be presented, even if redundant). The spaces enclosed by A and B are called "zones", while the two distinct borders surrounding A and B are called curves. There are two base sets (A and B) and an intersection set corresponding to the overlap between A and B. However, it should be noted that the relationships among visualization types in this study are far more complex.

To create a well-designed redesigned Euler Diagram in terms of aesthetics and comprehensibility, both "well-formedness" and "well-matchedness" must be considered (Alsallakh et al., 2014). Among the many properties of non-well-formed diagrams, four have been shown to significantly hinder understanding (Blake et al., 2013; Rodgers et al., 2012).

1. Non-circle or non-symmetric shape: curves in circle shape are most understandable.
2. Disconnected Zones: A zone is split into multiple subzones by other zones.
3. Concurrency: two curves run parallel and overlap.
4. Non-Simple Curve: a curve intersects itself.

Well-matchedness, which requires presenting only zones that contain data, has been shown to be more critical than well-formedness (Chapman et al., 2014).

Applying well-matchedness, redundant intersections were removed. For well-formedness, circular curves were preferred, with intersections made perpendicular whenever possible and crossings kept to a minimum. Inspired by the designs of *Untangling Euler Diagrams* (Riche & Dwyer, 2010) and *Kelp diagrams* (Dinkla et al., 2012*)*, we created the second representation of the relationships among data visualization types (See Figure 11)

The transformation steps from Node-Link Diagram to redesigned Euler Diagram are as follows:

1. Remove all elements except for the icons and their labels of the visualization types and visual encodings.
2. Draw circles centered at each visual encoding and its name text to define zones. Fill each zone with the corresponding icon color: red for *Arranging*, orange for *Linking*, and blue for *Varying*.
3. Place each visualization type into the appropriate visual encoding zone based on its primary visual encoding. Although the positioning is not strictly defined due to exceptions, the encoding that determines the spatial logic of a visualization type generally dictates its priority. The order of prioritization is as follows: *Mapping*, *Diverging*, *Axis-positioning*, *Ranging*, *Extending*, *Nesting*, *Ordered/Category Slots*, *Spatial Ordering*, *Picturing*, *Proportioning, Linking* (orange), and *Varying*(blue).
4. Colored edges, matching the color of each zone, are used to connect detached visualization types that employ more than one visual encoding to their corresponding visual encoding

zones. Visually, these edges resemble tentacles extending toward their targets. All edges are drawn orthogonally (i.e., using horizontal and vertical segments).
5. Adjust the shape of the container at the end of each extended edge based on the spatial logic described in Table 2.
6. Modify the fill color's lightness (value) to distinguish between different visual encoding zones.
7. Embed visual encoding icons as watermark-styles in the background of each zone, using a darker shade of the same hue with reduced opacity.
8. Align the zones both horizontally and vertically and distribute the visualization types evenly along the zone curves where possible.
9. When multiple containers overlap around a detached visualization type, or when edges must share space for clarity, apply the following layering order: *Varying* (blue) on top, followed by *Linking* (orange), and *Arranging* (red). This implies:
    - *Varying* encodings represent a third axis,
    - *Linking* encodings usually serve as additional information when *Arranging* encodings are also used.
    - *Arranging* encodings occupy the planar axes.
10. Group visualization types that share the same visual encodings within a single container.
11. Add one black outline for each additional use of the same visual encoding in a visualization type.
12. Include a legend that maps each abbreviation to its corresponding full name, as listed in Table 1.

2.3.4 The UpSet plot

The third representation is an UpSet plot, developed by Lex et al., designed to visualize complex relational datasets in a readable and compact format (Lex et al., 2014) —aligning with the objectives of this study.

A standard UpSet plot consists of three main components:

1. Central matrix: Displays all sets (base sets and intersection sets) in a dataset, including base sets and intersection sets, using connected black dots, with light grey dots indicating absence from a set. In this study, it visualizes all combinations of visual encodings across the 60 visualization types.
2. Top bar chart: illustrates the size of each set. In this study, it aggregates the number of visualization types that employ the same combination of visual encodings, which are linked to the central matrix below.
3. Left bar chart: Presents the frequency of each element. In this study, it shows how often each visual encoding is used across all visualization types, with each row corresponding to a specific encoding and linked to the central matrix on the right.

To enhance readability, a consistent color scheme matching the visual encoding icons was applied across both the matrix and the left bar chart. Each row is accompanied by a corresponding icon and abbreviated name to aid recognition. Similarly, icons representing visualization types are placed above the top bars to indicate which types employ specific combinations of visual encodings (See Figure 12)

This plot was generated using the Python package Upset Plot, based on the original design by Lex et al..

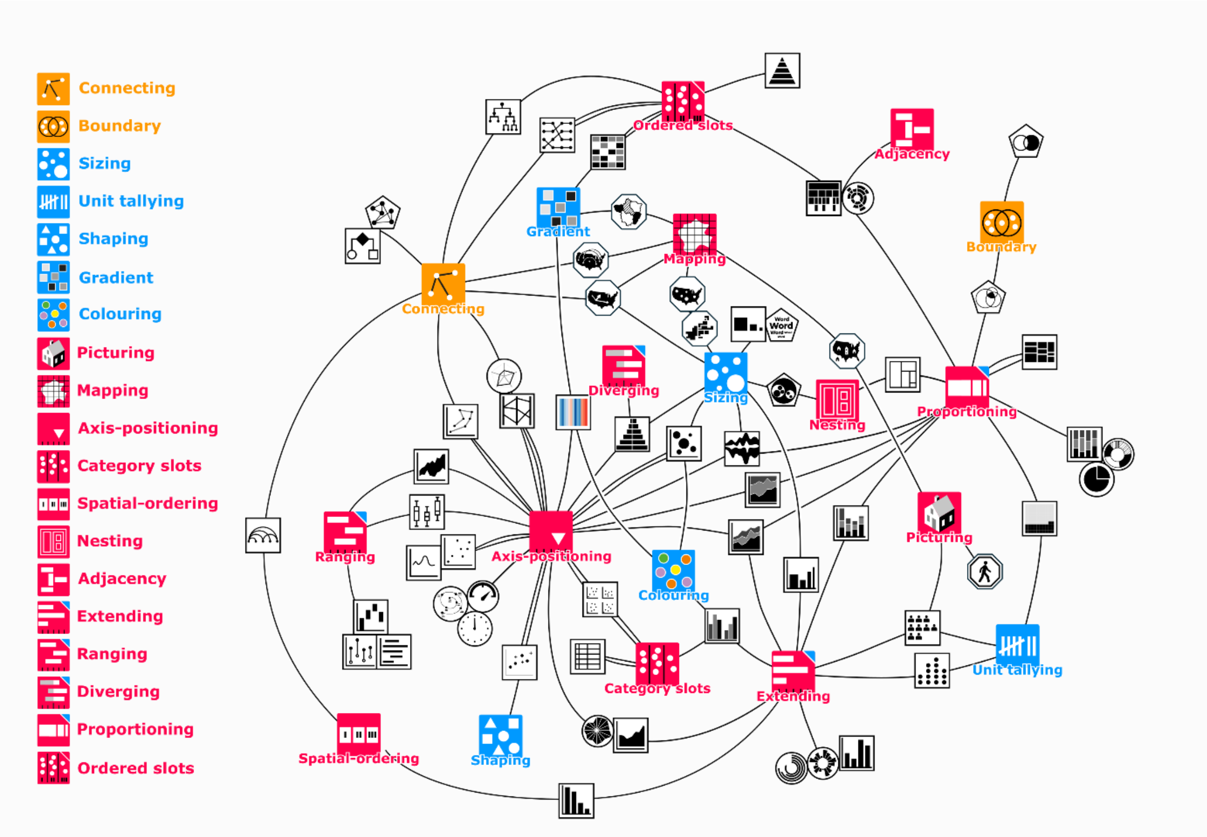

Figure 10: A Node-Link diagram representing the relationships among data visualization types.

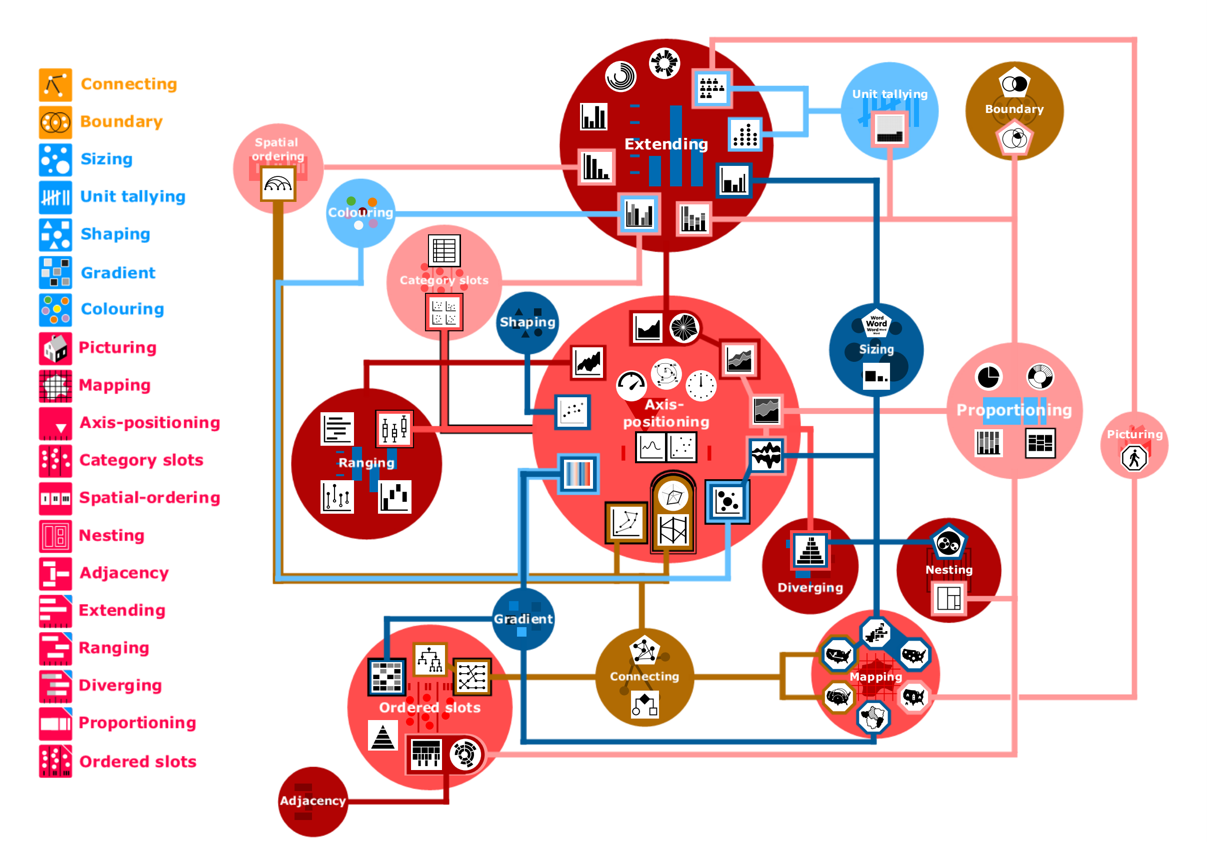

Figure 11: A redesigned Euler Diagram representing the relationships among data visualization types.

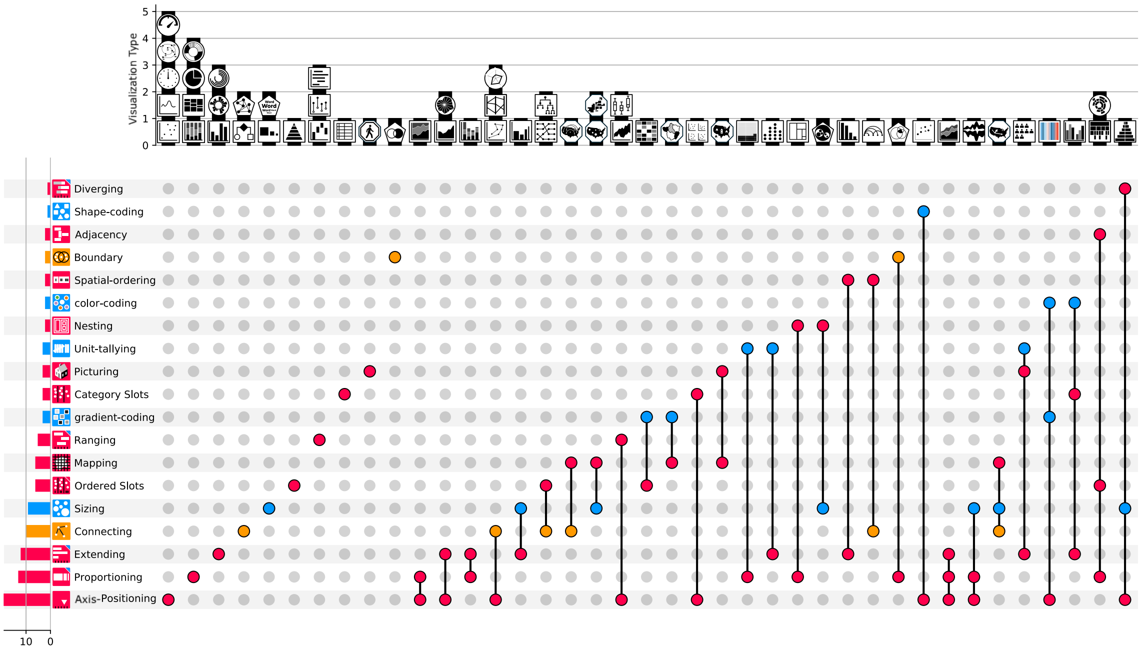

Figure 12: An UpSET plot representing the relationships among data visualization types.

2.4 The Design of the Online User Study

The purpose of the survey is to evaluate how well users understand the relationships among different visualization types based on the visual encoding they share, as presented through three distinct representational designs. To achieve this, an online user study was conducted to collect participants' opinions and impressions after completing a set of tasks involving these representations. In addition, their response accuracy was measured to assess response quality. Due to the presence of multiple manipulated variables across the three representations, it would be difficult to determine which variables contributed to any deviations in response time. Moreover, because our primary focus was on users' comprehension accuracy and subjective feelings of understanding, we chose not to measure response time in this study.

2.4.1 Study Questions

In the survey, we designed three types of questions to help participants become familiar with the three representational designs. Each question type included three variations—one for each representation—which, together with the random order of representations, were intended to mitigate potential learning effects.

The three types of questions were designed to help respondents become familiar with the representations and to assess their understanding of the different types of relationships present in each representation, as outlined below:

1. **Direct Connection**: This question type assesses participants' understanding of **direct** relationships between visual encodings and visualization types. The question is framed as: "The visual encoding icon below is **. According to the representation in this page, *which of the following visualization types use it?"*
2. **Indirect Connection**: This tests the understanding of **indirect** relationships between two visualization types. The question is framed as: "The visualization type icon below is **. According to the representation in this page, *which of the following visualization types share [how many] visual encodings with it?"*
3. **Statistics**: This assesses the understanding of **overall** relationships between visual encodings and visualization types. Example question formats include: "According to the representation in this page, *which of the following visual encodings are used less/more than * times by the visualization types?*

This order is fixed and applies to all representations with only one representation presented per page. In the survey, we denote the Node-Link Diagram (Figure 10) as Representation A, the redesigned Euler Diagram (Figure 11) as Representation B, and the UpSet plot (Figure 12) as Representation C.

2.4.2 The Structure of the Survey

The process of the survey is divided into five steps (or pages), listed as follows:

1. **Cover Page**: Briefly introduces the purpose of our study, what is expected in the survey, and the general instructions and information about the survey.
2. **Participant Background of Education**: Due to the concern that any kind of education may already provide training of data visualization, general questions like age, education level, education in data visualization and the purpose of creating data visualizations are listed.

3. **Participant Experience of data visualization design**: Likert scales are applied in this page to evaluate the abstract experience of designing data visualizations. The four questions are ordered as below, with abbreviations for use in the subsequent chapter's plots:
    1. I regularly create data visualizations for work, study, or personal use (abbreviated as Viz Use),
    2. I have independently designed data visualizations beyond standard templates (e.g., not relying on Excel defaults) (abbreviated as Viz Design),
    3. I am familiar with the design features that distinguish different types of data visualizations (e.g., how bar charts differ from heatmaps) (abbreviated as Viz Know),
    4. When creating data visualizations, I think about which visual encodings best suit my message. (* If you don't even know what this means, simply select "disagree".) (abbreviated as Viz Think).
4. **Task to familiarize the representations**: The concepts of visual encoding, visualization types, and the relationships among them are explained in more detail before presenting three representations, each on a separate page. No additional explanation about the representations themselves is given, in order to observe how participants react to and interpret the relationships. During the tasks, higher-resolution versions of the representations are also available via hyperlinks if needed.
5. **Feedback**: This section begins with a question asking participants which representation they favored overall. It is followed by five ranking questions, each presenting the same three representations as options. Participants are asked to rank them based on the following criteria, which have been abbreviated for use in the subsequent chapter's plots:
    1. Rank the representations according to how easily you can understand how different visualization types relate to each other (abbreviated as Easier),
    2. Rank the representations according to how much they encourage you to discover new ideas (abbreviated as Inspiring),
    3. Rank the representations according to how much they help you choose a suitable visualization type (abbreviated as Helpful),
    4. Rank the representations according to how eye-pleasing they are (abbreviated as Eye-pleasing),
    5. Rank the representations according to how burdensome you find them to understand. (abbreviated as Burdensome).

    Additionally, four short-answer questions were included to collect qualitative feedback:
    1. What did you like most about the representations?
    2. What aspects did you find confusing or unnecessary?
    3. In what context would you find these tools most useful?
    4. Anything you want to add?

2.4.3 The Execution of the Survey

The online survey was disseminated via social media platforms, such as LinkedIn, as well as through personal networks. It remained open for three weeks, from 02 August 2025 to 22 August 2025. All submitted responses were complete and considered eligible for further analysis. It is worth noting that, on the feedback page, Ranking Question 5 and Short-Answer Question 4 were added on 05 August 2025.

3. RESULT OF THE ONLINE SURVEY

3.1 General Statistics

In total, 120 complete responses were collected (85 responses for the feedback question regarding Burdensome). The participants were generally younger, well educated, and possessed at least some background knowledge of data visualization, primarily applying it in academic and research contexts. Among them, 63% were aged between 25 and 34, followed by 21% between 18 and 24, and 16% above 35 years old. Regarding the highest level of education, 50% held at least a master's degree, 42% held a bachelor's degree, 8% a doctorate, and 1% reported less than a bachelor's degree. A total of 43% of respondents had studied data visualization in some capacity, with an additional 27% reporting partial exposure. However, 30% indicated that they had no educational background in this area. Concerning applications, 65% used data visualization for academic and research purposes, while 53% applied it in their professional work. Other uses included educational purposes (28%) and personal interests (20%).

Regarding participants' experience in producing data visualizations, measured using a Likert scale, between 52% and 78% rated themselves as above average across the four self-assessment questions.

Figure 13 shows the number of respondents who answered each question fully correctly, grouped by question type and representation. Questions 1–3 correspond to Representation A (Node-Link Diagram), Questions 4–6 to Representation B (redesigned Euler Diagram), and Questions 7–9 to Representation C (UpSet plot). Among these, Representation A (Q 1–3) were answered more accurately than the other representations. Within Representation B, the indirect connection question (Q5) was answered more accurately than the other question types, including those testing comprehension of direct relationships, while in Representation C, the statistical question (Q9) achieved the highest accuracy. Overall, statistical questions exhibited the highest accuracy across all representations.

Figure 14 illustrates the distribution of the number of questions answered fully correctly. Notably, 38 participants answered no questions correctly, forming the main peak in the distribution. However, a secondary peak occurs at 5 questions, suggesting a subgroup of participants who may have responded more carefully.

For the feedback page, after familiarizing themselves with the three representations, the redesigned Euler Diagram was the most preferred (39%), followed by the UpSet plot (35%) and the Node-Link Diagram (26%).

Finally, Figure 15 presents the ranking of the three representations across the five feedback questions. The UpSet plot was the least inspiring for generating new design ideas, while the redesigned Euler Diagram was rated as the most visually appealing. The differences were statistically significant according to pairwise Mann–Whitney tests (non-parametric) and denoted by a star symbol in red, with $p = 0.05$.

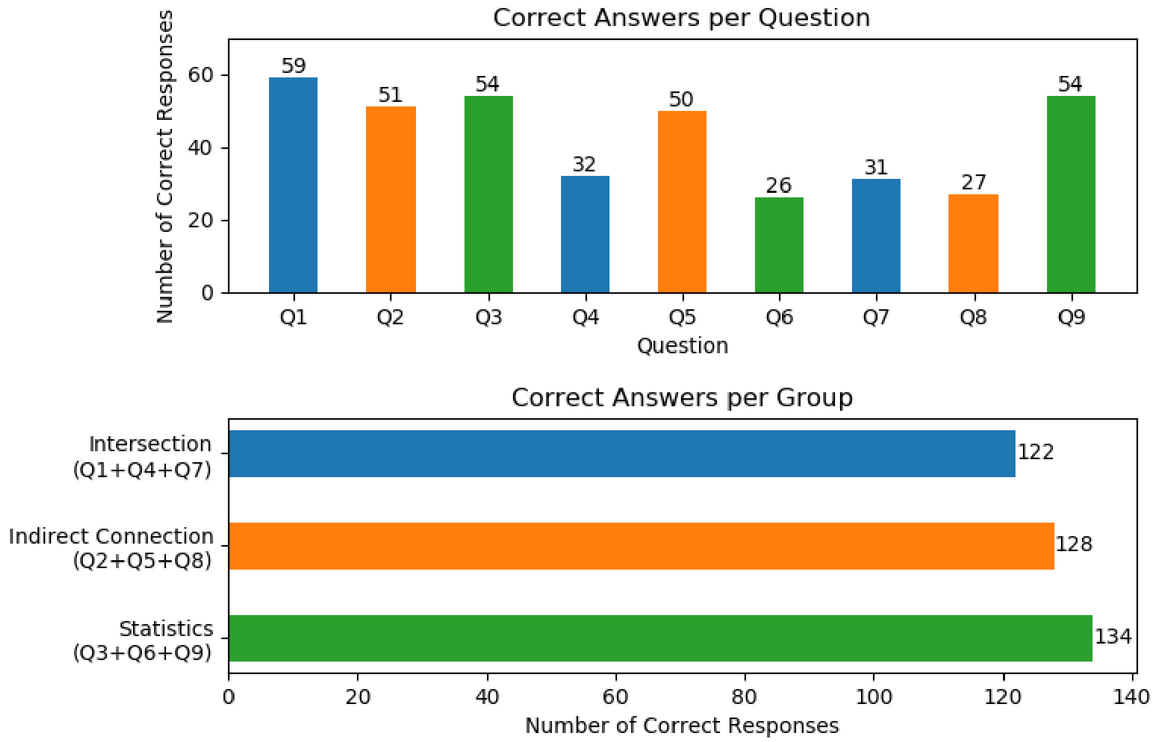

Figure 13: The number of Fully Correct Answers per question and per group.

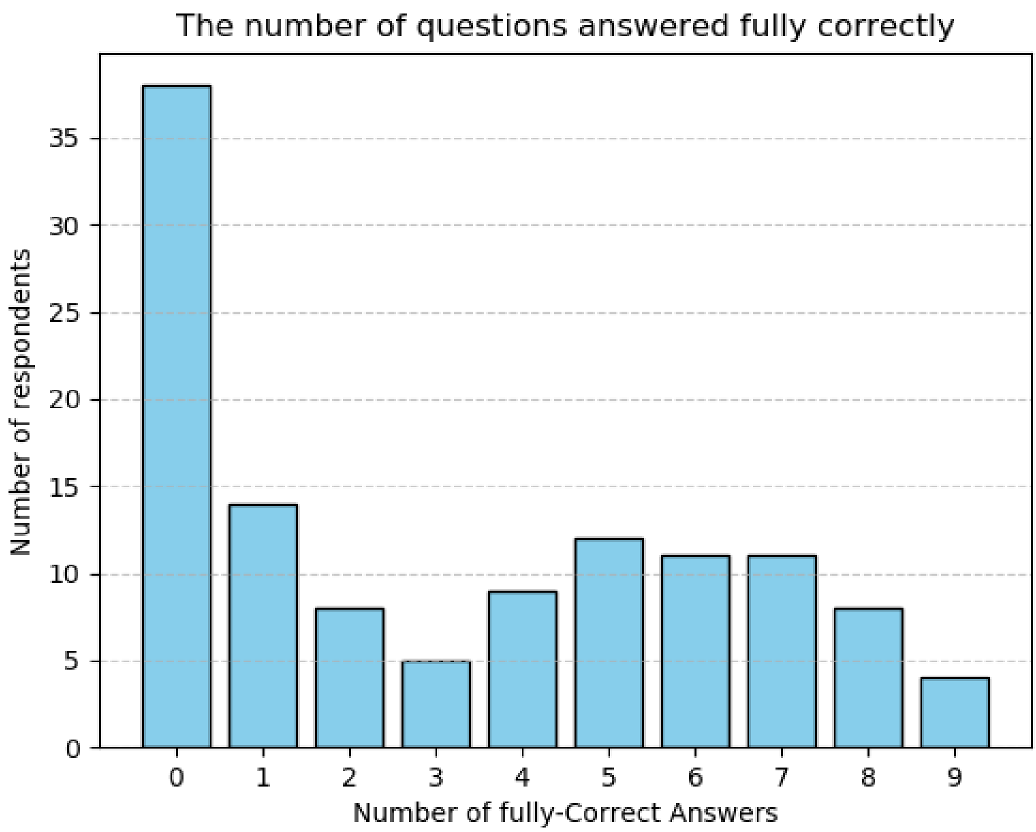

Figure 14: The number of questions that are answered fully correctly.

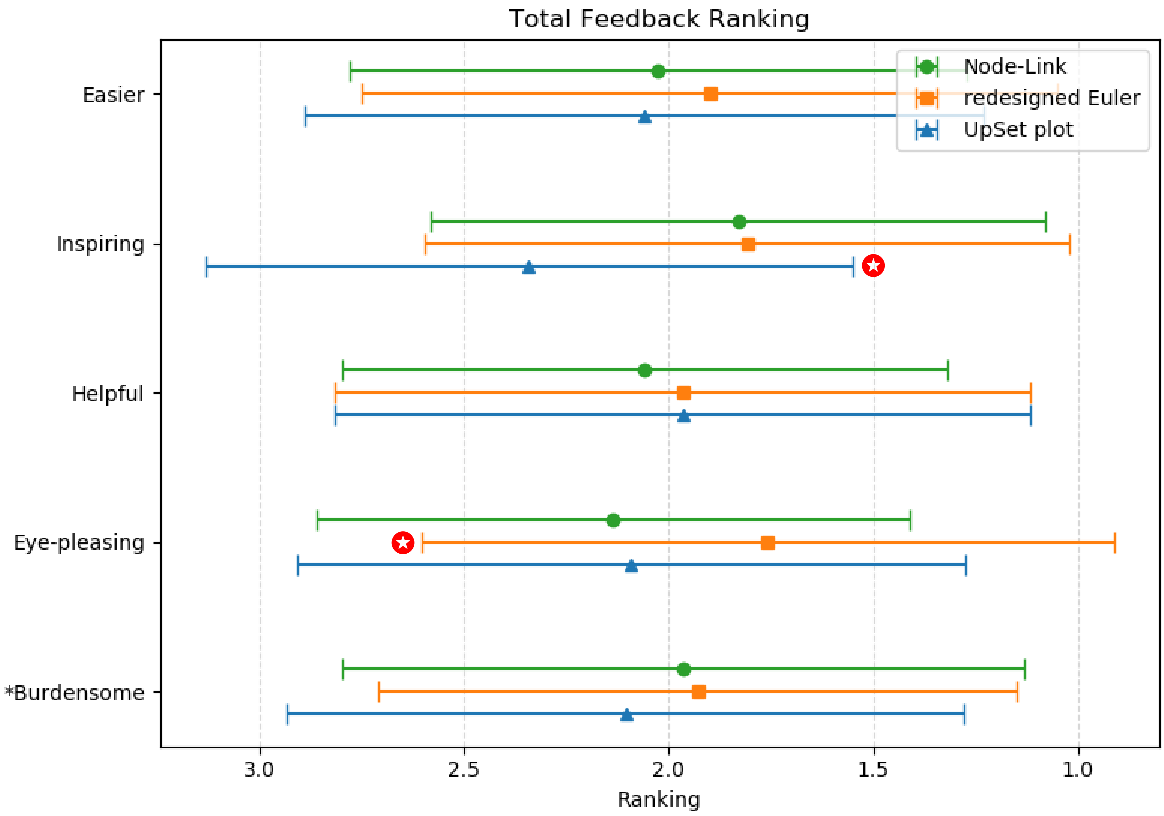

Figure 15: Overall participant feedback on the rankings of three representations across five ranking-type questions in the feedback section. Red star icons are placed next to the bars that show a significant difference compared to the other two bars, according to pairwise Mann–Whitney tests (non-parametric) with a significance level of $p = 0.05$. The ranking feedback questions are abbreviated for space constraints. *Note that only 85 responses were collected for the last ranking question: " Rank the representations according to how burdensome you find them to understand."

3.2 Criteria for a Qualified Respondent

Due to the presence of unqualified responses, we applied a criterion to filter out participants who completed the survey carelessly—that is, those who rushed through it without sufficient understanding of the questions or representations. While many methods exist to mitigate the effects of poor-quality responses (Curran, 2016), we chose not to apply them for three reasons:

1. Our survey mainly contains questions about personal background and experience, which require little comprehension, and multiple-choice questions, which are not commonly addressed in the scenarios considered by those methods.
2. The order of options in each multiple-choice question was not recorded, which is necessary for methods that rely on detecting behavioral patterns in responses.
3. The nine multiple-choice questions were designed to help participants familiarize themselves with the three representations, ensuring that their feedback is meaningful. Thus, a response can be considered qualified if the participant answers these questions reasonably correctly.

Because requiring "fully correct" answers across 8 or 9 questions would leave too few responses for analysis (See Figure 14), we redefined a "correct" answer for each question: a question is correct if the number of correct options selected exceeds the number of incorrect options. This approach tolerates minor mistakes while ensuring that participants generally understood the representations.

Figure 16 shows a comparison of the number of participants under two criteria for defining a question as "correctly answered." Under the redefined criterion, 37 participants selected more correct than incorrect options for all nine questions, while 18 participants did so for eight out of nine questions.

Ultimately, 55 out of 120 responses were qualified for further analysis. The general statistics of these participants are similar to the overall sample: 66% are aged 25–34, and 18% are aged 18–24; 46% hold a Master's degree, and 45% hold a Bachelor's degree; 45% have studied subjects relevant to data visualization, while 29% have not; and 68% use data visualization in academic or research settings. Regarding experience, 51% to 89% of participants rated themselves as above average across the four assessment questions.

3.3 Results of the Qualified Respondents

In this section, the preferences of qualified respondents (55 out of 120) regarding the comprehension of the three representations are evaluated quantitatively in terms of their experience in data visualization design and their responses to the feedback questions, while qualitative feedback is discussed in the next chapter.

Figure 17 shows the general preferences for the three representations across the four Likert-scale questions on data visualization experience. Although differences across questions are subtle, there is a slight trend: experienced visualization producers tend to prefer Representation C (UpSet plot), while Representation A (Node-Link Diagram) is favored by participants with average experience in the field. Representation B, on the other hand, shows relatively little variation across experience levels.

Figure 18 applies the same plotting method as Figure 15 but includes only qualified responses. As a result, Representation C (UpSet plot) remains the least inspiring, consistent with Figure 15, indicating that this finding is relatively robust. Additionally, two new significant results emerged: Representation A (Node-Link Diagram) is significantly the easiest to understand and the least burdensome.

To further explore how the participants' experience in data visualization production affects their responses to the ranking questions, the results from Figure 18 are disaggregated by each respondent's Likert-scale rating, as shown in Figure 19. Interestingly, the Node-Link Diagram is generally the easiest and least burdensome to understand across most Likert-scale levels. Regarding the other two representations, two notable tendencies emerged. First, among less experienced participants, Representation B (redesigned Euler Diagram) is perceived as easier to understand and more inspiring, whereas Representation C (UpSet plot) is viewed in this way by the more experienced participants. Second, in the Viz Think question (whether participants consider visual encodings when designing a visualization), more experienced participants regard the UpSet plot as more helpful than the redesigned Euler Diagram for choosing a suitable visualization type, whereas the opposite pattern is observed among less experienced participants.

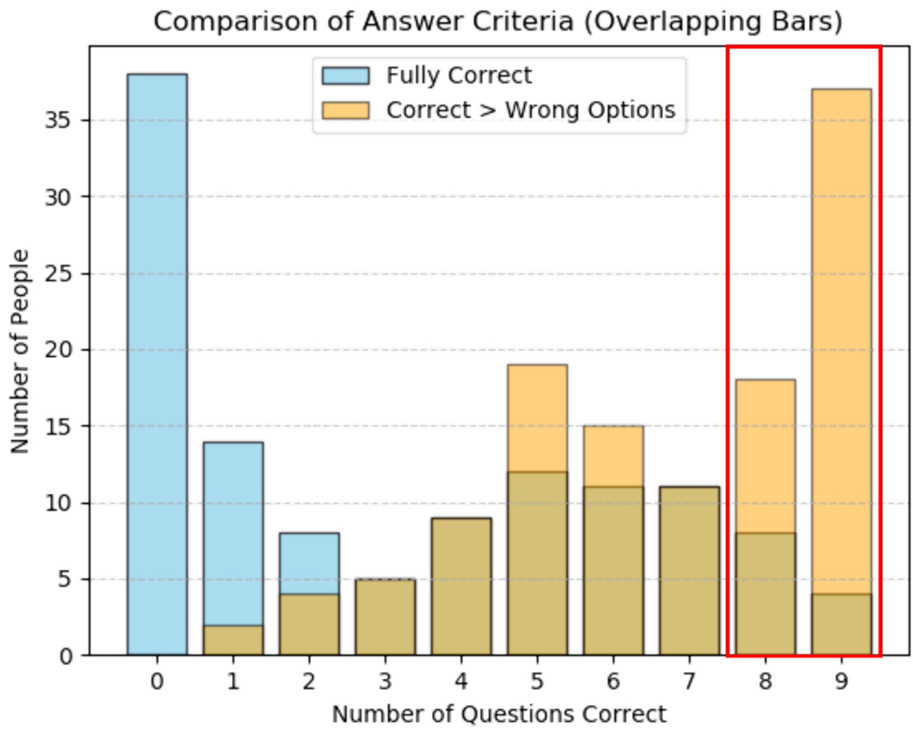

Figure 16: Comparison of two criteria for defining a correct answer. The blue bars represent responses counted as correct only if all correct options were selected, while the orange bars count responses as correct if the number of correct options selected exceeds the number of incorrect options. The two orange bars inside the red box indicate the qualified responses selected for further analysis.

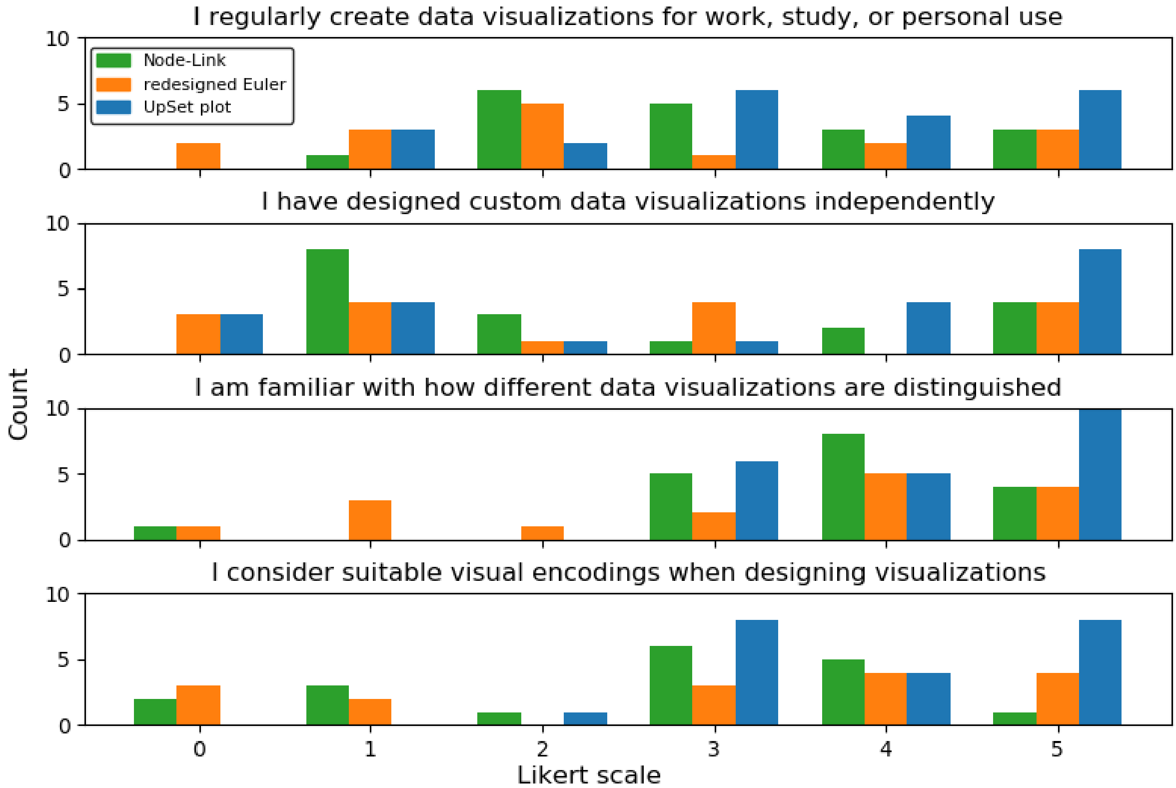

Figure 17: General preferences for the three representations across the four Likert-scale questions on data visualization experience.

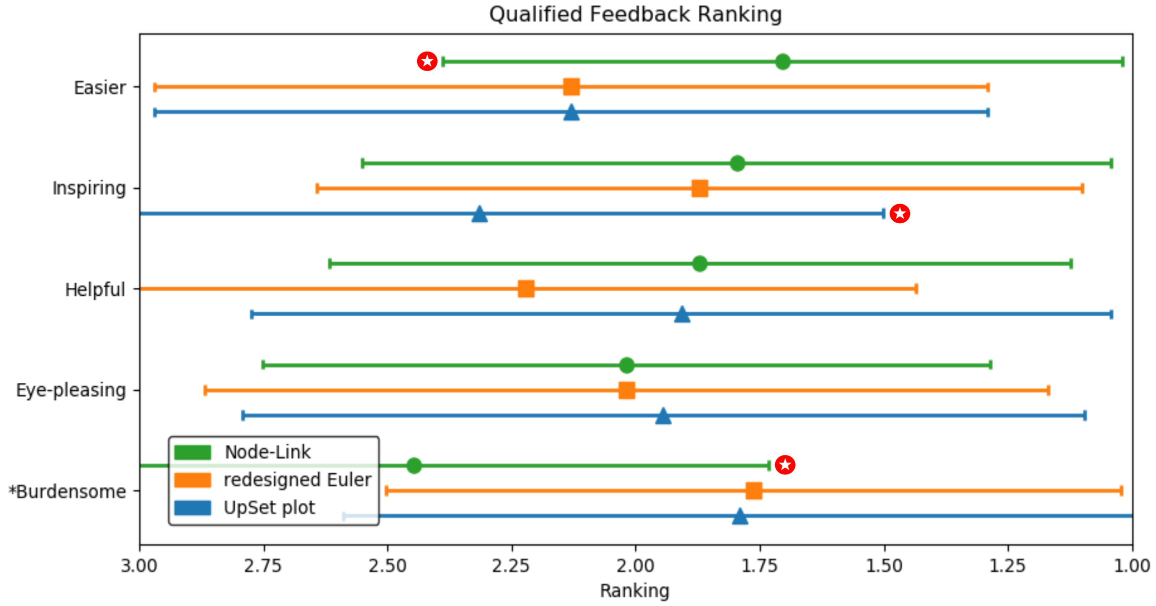

Figure 18: Qualified participant feedback on the rankings of three representations across five ranking-type questions in the feedback section. Red star icons are placed next to the bars that show a significant difference compared to the other two bars, according to pairwise Mann–Whitney tests (non-parametric) with a significance level of p = 0.05. *Note that only 38 responses were collected for the last ranking question: "Rank the representations according to how burdensome you find them to understand."

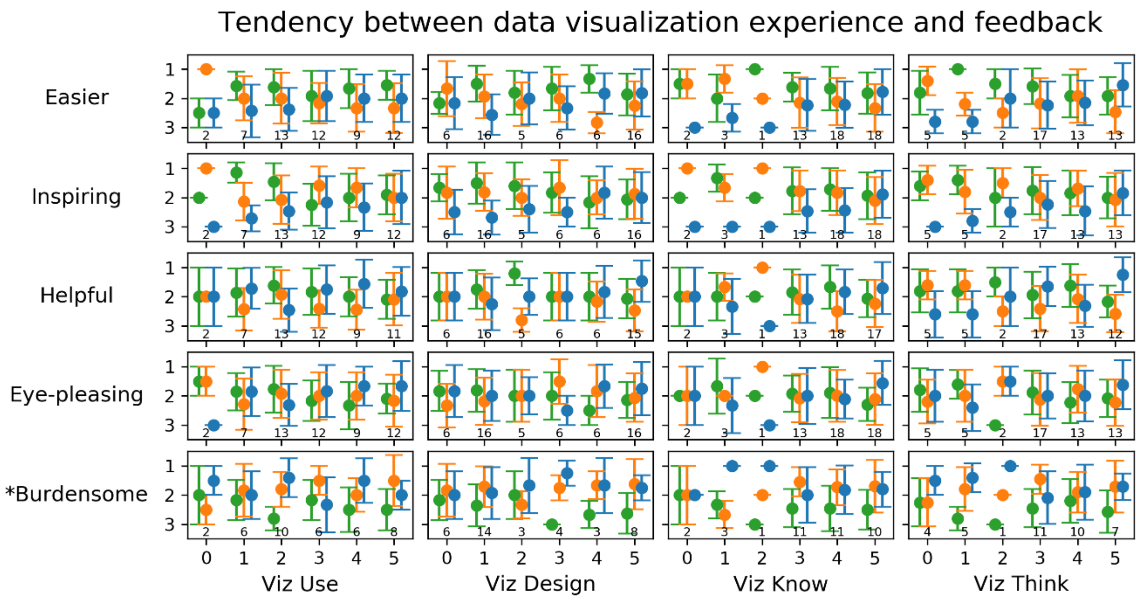

Figure 19: Qualified participant feedback on the rankings of three representations across four Likert-scale questions related to data visualization production experience and five ranking-type questions. The small number above each x-axis value indicates the number of responses for that specific Likert-scale value. *Note that only 38 responses were collected for the last ranking question: "Rank the representations according to how burdensome you find them to understand." The color scheme is the same as in the previous figures (green: Node-Link; orange: redesigned Euler; blue: UpSet plot), although it is omitted here due to layout constraints.

4. DISCUSSION

The online survey results and the outcomes of the three representations will be discussed separately, corresponding to the two research questions concerning design and analysis aspects.

4.1. Discussion of the Online Survey

Qualitative feedback is also collected to give more insight into respondents' comprehension of the three representations.

4.1.1 Node-Link Diagram (Representation A)

Respondents appreciated the Node-Link Diagram for being simple, vivid, and easy to read. People can quickly and directly perceive the relationships among different visualization types. However, using only lines to represent relationships sometimes confuses readers, as it can be hard to track which two nodes are actually connected, giving an impression of a messy layout. Double and triple black lines are also confusing without further explanation. This feedback aligns with the quantitative finding that it was the most accurately answered, easiest, and least burdensome representation to comprehend, with no significant differences observed in other aspects of reaction. The result is intuitive, as the Node-Link Diagram resembles force-directed or network graphs, which are already familiar to most users.

4.1.2 Redesigned Euler Diagram (Representation B)

Respondents appreciated the clear grouping and color scheme, which helps relate items in a logical and visually appealing way. On the other hand, encoding too much information with multiple lines, having insufficiently distinguishable colors, and employing a complex layout were noted as drawbacks. These issues can lead to visual overload and make it difficult to find icons. This reconfirms the traits of the (redesigned) Euler Diagram: compared to the Node-Link Diagram, it groups nodes more clearly and can be visually pleasing with an appropriate color scheme, but it becomes burdensome as the amount of information increases.

Regarding the unexpected result that the indirect connection question (Q5) was answered more accurately than the direct connection question (Q4), this may be due to the relatively simpler visualization used in Q5. Specifically, Q5 involved identifying visualization types that share visual encodings with a scatter plot, where all options were positioned within the zone *Positioning along a coordinate axis*, likely facilitating comprehension.

4.1.3 UpSet plot (Representation C)

The UpSet Plot is well-organized and structured using listings, colors, and icons, providing a comprehensive overview and helping users better understand visual encoding, as reflected by the high accuracy in the statistical questions. Once familiar, it allows users to locate specific visualization types and visual encodings easily. However, some respondents also noted that its large-scale layout can make information search cognitively demanding. Users often need to locate the desired icon, trace along the corresponding row or column, then shift 90 degrees to continue tracking their eyes to the final icon, which may create a sense of clutter despite the underlying organization. Vertical black lines can also be confusing without further explanation. While the positive feedback aligns with advantages mentioned in previous literature (Alsallakh et al., 2014; Lex et al., 2014), the negative feedback is newly observed, especially when relating two requested icons, likely because the layout and design of the UpSet Plot are less familiar to users initially.

4.1.4 Conclusion

While it is difficult to name a single "best" representation for showing relationships among data visualization types, each approach has its pros and cons. The Node-Link Diagram is familiar to most users but becomes messy as the number of lines increases. The redesigned Euler Diagram clearly shows groups and is visually appealing, but quickly becomes complex when too much information is included, which can be challenging during design. The UpSet Plot can handle relatively large amounts of relational information, but users need prior training to use it effectively.

4.2 Discussion of the Outcome of the representations themselves

In addition to the participants' preferences regarding the design of the three representations, observing the structures of the representations themselves—how visualization types and visual encodings are distributed and related—may also shed light on our understanding of data visualization. Here, we focus on the Force-Directed Graph and the UpSet plot, since the Node-Link Diagram and the redesigned Euler Diagram can be considered structural variations of the force-directed graph, only enhancing readability and aesthetics without providing additional information.

4.2.1 Observation of the Structure of the Force-Directed Graph

A force-directed graph highlights relationships through edges and distances, allowing affinity among nodes to emerge. Based on this property, several features can be observed (Figure 9).

1. Arranging (red icons) and Varying (blue icons) encodings appear relatively evenly across the spatial layout, indicating that they are often used in combination.
2. Encodings with more edges and closer to the center likely indicate more frequent usage. This includes *Positioning along a coordinate axis*, *Proportional space-filling*, *Extending along a coordinate axis*, *Connecting*, and *Sizing*.
3. Encodings positioned close to each other suggest a higher potential for co-usage, while distant ones tend not to be combined. For instance, *Positioning along a coordinate axis* and *Connecting* are often used together, whereas *Ranging along a coordinate axis* and *Grouping by boundary* are rarely, if ever, combined.

Although this study includes only 60 visualization types and some apparent neighbors may be "false" (nodes pushed together rather than truly attracted (Munzner, 2014)), the graph still offers a first overview of how visualization design operates.

4.2.2 Observation of the Structure of the UpSet plot

The UpSet plot (Figure 12), a combination of matrix-based and aggregation-based relational graphs, is a particularly effective method for disentangling complex relational data, offering both an overview and the ability for detailed exploration (Alsallakh et al., 2014). Its structured layout reveals several statistical characteristics:

1. The frequencies of individual encodings and their combinations are readily apparent in the left and upper bar charts. The top five most frequently used encodings are *Positioning along a coordinate axis*, *Proportional space-filling*, *Extending along a coordinate axis*, *Connecting*, and *Sizing*.
2. Among combinations on the top bar chart, *Positioning along a coordinate axis* and *Connecting* emerge as the most common (appearing in three visualization types).
3. Combinations of two encodings are more frequent overall than combinations of one or three.

4.2.3 Conclusion

Across the 60 visualization types analyzed, *Positioning along a coordinate axis*, *Proportional space-filling*, *Extending along a coordinate axis*, *Connecting*, and *Sizing* are the most frequently used encodings, often appearing in combinations of two.

5. CONCLUSION

5.1. Findings

To address the lack of spatial and comprehensive representations that support the understanding of relationships among data visualization types, we identified key characteristics of these relationships and designed three representations, which were further evaluated through user testing.

From the statistical perspective, *Positioning along a coordinate axis, Proportional space-filling, Extending along a coordinate axis, Connecting,* and *Sizing* emerged as the most frequently used visual encodings. These are often employed in combination with each other or alongside other encodings.

From the user testing, none of the three representations outperformed the others across all aspects. Each has its strengths and limitations: the Node-Link Diagram is intuitive but quickly becomes chaotic with many edges; the redesigned Euler Diagram is visually appealing and highlights clusters, but it can easily become overloaded with information; the UpSet plot offers both overview and detail, yet it is less familiar and inspiring to the general public.

5.2. Contribution

This study makes two main contributions. First, it introduces three alternative representations designed to enhance the understanding of relationships among visualization types through spatial reasoning, and evaluates them based on user feedback. Second, it identifies characteristics of these relationships by observing their spatial structure and statistical characteristics. Together, these contributions support both beginners, who seek to learn how visualizations are constructed and explore new design possibilities, and experts, who aim to advance theoretical integration and the taxonomy of visualization types for academic and educational purposes.

5.3. Limitations

This study focuses on 2D, static, simple composite designs for both the selection of visualization types and the design of the representations. In constructing the relationship structure, only 60 types of data visualization were included; expanding this set could potentially alter the results. Additionally, visual encoding was used as the sole factor to define relationships, and incorporating or replacing it with other factors may lead to different outcomes.

5.4. Future Work

Future research could explore 3D graphs, animations, interactive visualizations, and complex composite visualizations, as well as the design of representations to illustrate the relationships among them, given the increasing prevalence of these visualization types in the digital world. Different definitions of relationships could also be examined, including, but not limited to, data input types and specific tasks. Representations to show such relationships could be further explored for different target audiences, domain-specific contexts, or alternative aesthetic designs.

LIST OF REFERENCES

Alsallakh, B., Micallef, L., Aigner, W., Hauser, H., Miksch, S., & Rodgers, P. (2014). *Visualizing sets and set-typed data: State-of-the-art and future challenges* 16th Eurographics Conference on Visualization (EuroVis 2014), https://kar.kent.ac.uk/39007/1/STAR_paper.pdf

Amar, R. A., & Stasko, J. T. (2005). Knowledge Precepts for Design and Evaluation of Information Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *11*(4), 432-442. https://doi.org/10.1109/TVCG.2005.63

Battle, L., Duan, P., Miranda, Z., Mukusheva, D., Chang, R., & Stonebraker, M. (2018). *Beagle: Automated Extraction and Interpretation of Visualizations from the Web* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, https://doi.org/10.1145/3173574.3174168

Bertin, J. (1983). *Semiology of graphics* (W. J. Berg, Trans.). University of Wisconsin Pres.

Blake, A., Stapleton, G., Rodgers, P., Cheek, L., & Howse, J. (2013). *Improving user comprehension of Euler diagrams* 2013 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), https://ieeexplore.ieee.org/document/6645262

Card, S. K., & Mackinlay, J. (1997). The Structure of the Information Visualization Design Space. IEEE Symposium on Information Visualization (InfoVis '97),

Chapman, P., Stapleton, G., Rodgers, P., Micallef, L., & Blake, A. (2014). Visualizing Sets: An Empirical Comparison of Diagram Types. In T. P. Dwyer, H.; Delaney, A. (Ed.), *Diagrammatic Representation and Inference* (Vol. 8578, pp. 146–160). Springer. https://doi.org/10.1007/978-3-662-44043-8_18

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554. https://doi.org/10.2307/2288400

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4-19. https://doi.org/10.1016/j.jesp.2015.07.006

Dhote, A., Javed, M., & Doermann, D. S. (2023). A Survey and Approach to Chart Classification. In M. F. Coustaty, A. (Ed.), *Document Analysis and Recognition – ICDAR 2023 Workshops* (Vol. 14193, pp. 59–74). Springer, Cham. https://doi.org/10.1007/978-3-031-41498-5_5

Dinkla, K., van Kreveld, M. J., Speckmann, B., & Westenberg, M. A. (2012). Kelp Diagrams: Point Set Membership Visualization. *Computer Graphics Forum*, *31*(3pt1), 875–884. https://doi.org/10.1111/j.1467-8659.2012.03080.x

Engelhardt, Y., & Richards, C. (2024). A Building-Block Approach to the Diversity of Visualization Types: Each Type Expressed Visually, and as a Systematically Generated Sentence. In J. Lemanski, M. W. Johansen, E. Manalo, P. Viana, R. Bhattacharjee, & R. Burns (Eds.), *Diagrammatic Representation and Inference* (Vol. 14981, pp. 28–43). Springer. https://diagrams-2024.diagrams-conference.org/

Ferdio. (2025). *Data Viz Project*. Retrieved June 14 from https://datavizproject.com/

Few, S. (2007). Periodic table of visualization methods exhibits poor visual design. https://www.perceptualedge.com/blog/?p=81

Franconeri, S. L. (2019). *Data Visualization Quick Reference*. Northwestern University. https://bpb-us-e1.wpmucdn.com/sites.northwestern.edu/dist/8/4164/files/2020/01/Franconeri_ExperCeption DotNet_DataVisQuickRef.pdf

Graphopedia. (n.d.). *Graphopedia*. Retrieved 14 June from

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, *53*(6), 59–67. https://doi.org/10.1145/1743546.1743567

Holtz, Y., & Healy, C. (2018). *from Data to Viz*. Retrieved June 14 from https://www.data-to-viz.com/

Kandogan, E., & Lee, H. (2016). *A Grounded Theory Study on the Language of Data Visualization Principles and Guidelines* Human Vision and Electronic Imaging (HVEI),

Kosara, R. (2009). Visualization is not Periodic, Period! https://eagereyes.org/blog/2009/visualization-is-not-periodic-html

Lengler, R., & Eppler, M. J. (2007, September). Towards a Periodic Table of Visualization Methods for Management. I-KNOW '07: International Conference on Knowledge Management, Graz, Austria.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, *5*(2), 110–141. https://doi.org/10.1145/22949.22950

Mackinlay, J. D., Hanrahan, P., & Stolte, C. (2007). Show Me: Automatic Presentation for Visual Analysis. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1137–1144. https://doi.org/10.1109/TVCG.2007.70594

Munzner, T. (2009). A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 921–928. https://doi.org/10.1109/TVCG.2009.111

Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press. https://www.cs.ubc.ca/~tmm/vadbook/

Oliveira, E. C., Oliveira, L. C., Cardoso, A., Mattioli, L., & Lamounier, E. A. (2017). Meta-model of information visualization based on Treemap. *Universal Access in the Information Society*, *16*(4), 903–912. https://doi.org/10.1007/s10209-016-0477-9

Purchase, H. C. (1997). *Which aesthetic has the greatest effect on human understanding?* Proceedings of the 5th International Symposium on Graph Drawing (GD '97), https://link.springer.com/content/pdf/10.1007/3-540-63938-1_67.pdf

Quadri, G. J., & Rosen, P. (2022). A Survey of Perception-Based Visualization Studies by Task. *IEEE Transactions on Visualization and Computer Graphics*, *28*(12), 5026–5048. https://doi.org/10.1109/TVCG.2021.3098240

Ribecca, S. (2025). *The Graphic Continuum*. Retrieved 14 June from https://www.severinoribecca.one/portfolio-item/the-graphic-continuum/

Riche, N. H., & Dwyer, T. (2010). Untangling Euler Diagrams. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1090–1099. https://doi.org/10.1109/TVCG.2010.210

Rodgers, P., Zhang, L., & Purchase, H. (2012). Wellformedness Properties in Euler Diagrams: Which Should Be Used? *IEEE Transactions on Visualization and Computer Graphics*, *18*(7), 1089-1100. https://doi.org/10.1109/TVCG.2011.143

Rodrigues, J. F., Jr., Traina, A. J. M., de Oliveira, M. C. F., & Traina, C., Jr. (2007). The spatial-perceptual design space: A new comprehension for data visualization. *Information Visualization*, *6*(4), 261–279. https://doi.org/10.1145/1375935.1375937

Saket, B., Endert, A., & Demiralp, Ç. (2019). Task-Based Effectiveness of Basic Visualizations. arXiv:1709.08546. Retrieved 14.04.2025, from https://arxiv.org/abs/1709.08546

Schwab, L. (2025). *A friendly guide to choosing a chart type*. Retrieved June 17 from https://www.datawrapper.de/blog/chart-types-guide

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.

Simkin, D., & Hastie, R. (1987). An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, *82*(398), 454-465. https://doi.org/10.2307/2289447

Times, F. (2016-2019). Visual Vocabulary. In.

Unwin, A. (2020). Why Is Data Visualization Important? What Is Important in Data Visualization? *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.8ae4d525

APPENDIX

The code for the interactive force-directed graph, the link to the online survey, the survey responses, and the three representations in SVG format are available at the following GitHub repository:

https://github.com/bryanCycwalker/Cartography-Master-Thesis-Mapping-the-Relationships-among-Data-Visualization-Types