PICK YOUR CHART

A Visualization For
Data Visualization Catalogues

Name: Hanlu Ma
Student ID: 10846507

Supervisor: Michele Mauri

A.A. 2021/2022
Politecnico di Milano
School of Design
LM Communication Design

PICK YOUR CHART

A Visualization For
Data Visualization Catalogues

→ INDEX

ABSTRACT

Benefiting from the development of statistics and computer science, data visualization, a highly specialized field, has more public participation. Nowadays, there is an assortment of data visualization tools that are more user-friendly than ever. However, for non-expert users, there is still confusion while figuring out which visual model is best for the task at hand. Without pre-knowledge in statistical fields, it is necessary for them to refer to the data visualization catalogues. For these catalogues, there is a lack of unified standard and we can not find a perfect one in all aspects. To fix the gap, there is a keen need to review current catalogues.

Therefore, this thesis aims to create a visualization and compare current data visualization catalogues. By collecting catalogues, determining their structure, categories and visual models included, it provides a visual path to help lay users to compare the similarities and differences between existing data visualization catalogues and make wise choices, while also providing experts who will create another data visualization catalogues a comparing perspective.

→ KEYWORDS

Data visualization catalogue;
Taxonomy;
Data multi-experience;

1
BACKGROUNDS

With the development of modern statistics and computer science, data visualization, which has been a highly specialized field, is now relatively available to people without technical training. Many visualization tools have been developed in the past few decades to help. In the process of the general public producing their own visualization, however, there is a barrier between non-expert users and the appropriate visual model of their data. Without pre-knowledge, it is not easy to choose the appropriate technique in current data visualization tools.

1.1 DATA VISUALIZATION

It is common to think of statistical graphics and data visualization as relatively modern developments in statistics. In fact, the graphic portrayal of quantitative information has deep roots.[1] The ancient Babylonians, Egyptians, Greeks and Chinese all developed sophisticated ways of representing information visually to plot the movement of the stars, produce maps to aid navigation, and develop plans for crop planting and city development.[2] In the 17th century, the graphic representation of quantitative information was applied not only in the realm of map-making but also gradually extended to the medical and economic fields. Till the 19th century, which is also known as the "Golden Age" of statistical graphing, many chart

1 Friendly, Michael. 2008. "A Brief History Of Data Visualization". Handbook Of Data Visualization, 15-56. doi:10.1007/978-3-540-33037-0_2.

2 "Dataart - Learning Resources". 2022. Data-Art.Net. http://data-art.net/resources/history_of_vis.php.

types, including histograms, timelines, scatterplots and rose charts have been invented and widely applied in thematic cartography by experts. One of the most cited examples of statistical graphics is Charles Minard's map depicting the process of Napoleon's army retreating from Russia.

With the increasing importance of large data for modern business and the emergence of computer processing in the latter half of the 20th century, data visualization exploded into dozens and even hundreds of focus areas. Almost every STEM field benefits from understanding data—and so do fields in government, finance, education, design, etc. By representing data as graphics, it conveys complex structure and explains patterns and trends, helping people to dig for more insights, look at data more imaginatively and make more intelligent strategies.

Benefiting from the development of statistics and computer graphic technology, the professional threshold for data visualization has been significantly lowered. Nowadays, there is an assortment of dashboards and data discovery tools, analytics suites, and other tools which are more user friendly than ever, like Microsoft Excel, Tableau, D3.js, and Gephi.[3] They are designed to enable businesses, researchers, and even non-expert individuals to create their own visual artifacts based on data. Only with a few clicks of the mouse, along with some editing and formatting, what once took hours of exhausting drawing could be done.

3 "Data Visualization – History And Origins". 2022. https://thinkinsights.net/digital/data-visualization-history/.

1.2 CONFUSION IN VISUAL MODEL CHOOSING

Though thanks to digital tools, now it is more convenient than ever for lay users to involve data visualization practice, there are still some decision making points that can not be skipped or solved by tools automatically. One of the most difficult parts of creating a data visualization is figuring out which visual model is best for the task at hand.[4]

To master these digital tools and translate the datasets into a graphical context, it requires some preliminary knowledge about how visualization techniques work and the ability of choosing fitful techniques to use for analysis, which could be somehow absent for lay users. For instance, a graphic designer who has not touched the field of data visualization might feel confused when he is asked to create a dashboard showing the distribution of computer storage. If the designer is familiar with the statistic, he could recognize that original data is percentages of space used to store each type of files, and the aim of the visualization is to show the audience a part-to-whole relation. But which type of chart should he choose? Pie, donut, gauge, or stacked percentage bars? The data visualization tool can help to map the quantities to graphical features, however, it can not decide the visual model in use automatically for users.

4 "Essential Chart Types For Data Visualization". 2022. Chartio. https://chartio.com/learn/charts/essential-chart-types-for-data-visualization/.

↑ 1 A bad case of visulizaion with wrong visual model

Above is an example of a basic task with almost the simplest data, showing that one of the biggest challenges for lay users is deciding the visual representation to represent the information best. According to the choice, the visualization can make or break the project. It could be a tragedy to put tons of effort into analyzing and modeling the data but end up using the wrong visual model to present the result. The audience will grasp nothing but confusion.

This figure above shows a real case of inefficient data visualization (Figure 1).[5] It is a visualization created by Star Sports, an Indian television sports channel. It just overwhelms the viewers with numbers without giving them any information. Maybe the designer was trying to convey the trend of world cup changing over time, which could be good to show with a simple line plot. Visualized with a table, which is not an appropriate visual model for it, this set of data is almost unrecognizable. Viewers can't get the real dependencies between the result of races and time changing.

5 "Bad Visualisations". 2022. Bad Visualisations. https://badvisualisations.tumblr.com/post/185690817296/this-is-a-screengrab-from-star-sports.

1.3 DATA VISUALIZATION CATEGORIZING IN ACADEMIC LITERATURE

Picking the fitful visual type requires pre-knowledge of statistics, but it does not mean that lay users are excluded from the data visualization field. In this situation, what could be helpful to get over the barrier is a well-curated catalog. A clear taxonomy system can make user's thinking way and application goal clear, helping them to pick the right technique.

Currently, there is some literature focusing on the taxonomy of visual models in data visualization. For instance, in Relationship Types in Visual Analytics Nur Izzati Shabdin divided 59 models in visual analytics into 6 categories (correlation, comparison, distribution, outliers, differences, and reduction) according to the data relationships they suggest.[6] Then he described each of the relationship representations in detail. In this paper, Shabdin collected a total of 64 visual representations, such as the bar chart, box & whisker plot, and histogram, and filtered them into 59 for final analysis. However, there is a lack of details about his collecting and sorting. It does not well report the specific range of current visual models which are common in use.

6 Shabdin, Nur Izzati, Suraya Ya'acob, and Nilam Nur Amir Sjarif. 2020. "Relationship Types In Visual Analytics". Proceedings Of The 2020 6Th International Conference On Computer And Technology Applications. doi: 10.1145/3397125.3397127.

Also, in The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, Ben Shneiderman provides not a specific classification of visual models, but a taxonomical method according to the data types and tasks in interactive data visualization. He offers a proposal of seven data types (one-, two-, three-dimensional data, temporal and multi-dimensional data, temporal and multi-dimensional data, and tree and network data) for seven tasks (overview, zoom, filter, details-on-demand, relate, history, and extracts).[7]

In addition, in Taxonomy of Visualization Techniques and Systems – Concerns between Users and Developers are Different, Qin Chengzhi proposed classification methods based on two factors after reviewing the current status of geo-visualization techniques.[8] He divided visualization techniques according to data type (1D, 2D, 3D, multi-dimensional, hierarchical, graph, and text/hypertext) and representation mode (pixel-oriented, geometric projection, icon-based, hierarchy-based, and graph-based). This study mainly focuses on visualizations in geographical fields, calling for a general study for further steps.

7 Shneiderman, B. 2022. "The Eyes Have It: A Task By Data Type Taxonomy For Information Visualizations". Proceedings 1996 IEEE Symposium On Visual Languages. Accessed July 6. doi:10.1109/vl.1996.545307.

8 Qin, Cheng-Zhi & Chenghu, Zhou & Tao, Pei. (2003). Taxonomy of visualization techniques and systems–concerns between users and developers are different. 10.13140/RG.2.1.3159.2403.

1.4 DATA VISUALIZATION CATALOGUES IN PRACTICE

Besides the taxonomy studies among academic literatures, several attempts were also made in the past to create “families” of visual models by professional designers. One popular and broadly mentioned case is the Data Visualization Catalogue (2013) developed by communication designer Severino Ribecca.[9] It is a library of different information visualization types. Ribecca defined this project as a way to develop his own knowledge of data visualization and create a reference tool for him and anyone else in the field that requires the use of data visualization. In this library, 60 types of visual representations are divided into 16 groups according to their functions or by what the user wants to communicate to an audience. One chart type could be set in multiple groups. He also researched each method, to find the best way to explain how it works and what it is best suited for.

In spite of most catalogues developed by designers individually based on their visualization practices and experience, there are also some catalogues created by commercial institutions or developer communities. For example, Chartopedia (2017) is a catalogue made by AnyChart, a data visualization solution supplier.[10] They provide system engineering and libraries based on JavaScript for

9 "The Data Visualisation Catalogue". 2022. Datavizcatalogue. Com. https://datavizcatalogue.com/.

10 "Usage Type | Chartopedia | Anychart". 2022. Anychart. https://www.anychart.com/chartopedia/usage-type/.

customers to improve their decision-making and enterprise performance. They make Chartopedia as an information resource that allows users to discover as many details about any type of chart supported in their JavaScript charting libraries. To help users find the type of graph they need for the current visualization task, the graphs categorized 70 types of charts based on 8 usage types. Another interesting case of data visualization catalogues is the Carbon Design System (2019).[11] It is a collection of pre-built, reusable patterns, design guidance and code assets, that allow users to minimize the process while designing their own product. Founded by IBM, this open-source community created a reusable D3 charting library called Carbon Charts. As part of the tutorial of their library, Carbon team provides a chart-type catalogue. Users can start by identifying the purpose of the visualization, and then choose the appropriate chart type. There are 6 categories including 24 types of charts, each of them is listed with source code and demo.

Currently, there are many contributors, both individual and institution, who have already shared their own catalogues. However, there could be differences between them which raise the barrier of recognition. For instance, although most of the catalogue contributors sort chart types by visualization purpose, which is also called as “Function”, “Usage” or “You would like to show” in other catalogues, there’s no common standard for its definition. The lack of unified standards leads to a mixing up of categories. Take the three categories mentioned above as an example, they contain 16, 8 and 6 categories respectively. Also, there is no consensus on whether a single chart could be included in multiple classifications. Moreover, many of the catalogues are not full-scale enough to cover all chart types, which can lead to taxonomic confusion for users.

11 "Carbon Design System - Chart Types". 2022. Carbondesignsystem. Com. https://carbondesignsystem.com/data-visualization/chart-types/.

1.5 THE MEANING OF THE PROJECT

At this moment, we cannot find a data visualization catalogue perfect in all aspects. For the existing catalogues, there are neither comparative studies among them, nor sufficient ground of such categorizations in the academic literature. However, as the recognition for the importance of numerical data in commerce and education grows, more people who are not either specialist statisticians or scientists are trying to get involved in data visualization practice. The missing of chart types and the mixing up of classification factors can totally confuse non-expert users. There is a gap between users and the right techniques they need without referencing the appropriate data visualization catalogues.

To start fixing the gap, there is a keen need to review current catalogues. The goal of this thesis is to collect the available data visualization catalogues, visualizing and analyzing them to figure out:

- → Which visual models are common in current data visualization catalogues?
- → What are the categorization methods used by current data visualization catalogues?

By researching these questions, this thesis is aiming to help lay users to access a comparison analysis of the catalogues, clarifying the ambiguity in data visualization catalogues and promoting further development for better solutions.

2
METHODS

This project is based on 13 selected data visualization catalogues. The categories and visual models they contain are recorded as our dataset, trimmed to avoid ambiguity and duplication. Subsequently, the frequencies of the visual models appearing in these catalogues are calculated. After being sorted, these data were interpreted into visual artifacts and analyzed. In order to help the public be aware of this topic, the final results are reorganized and designed into a website.

For the research method, our main goals are to record the structures of selected catalogues in a unified format and determine the range of visual models and categories included in the selected catalogues.

2.1 DATA SOURCE & PRE-ANALYSIS

Searching “data visualization catalogue” on Google, we can get about 15400,000 results, most of which are informational noise. In order to build a reliable list of data visualization catalogues, we referred to expert opinions. Firstly, our eyes were on Dr. Yuri Engelhardt’s meta catalogue, an assistant professor of the multidisciplinary data visualization curriculum at the University of Amsterdam.[12] In his meta catalogue, he mentioned eight catalogues. Then we also referred

12 "Yuri Engelhardt's Twitter". 2022. https://twitter.com/YuriEngelhardt/status/1285885738253537281.

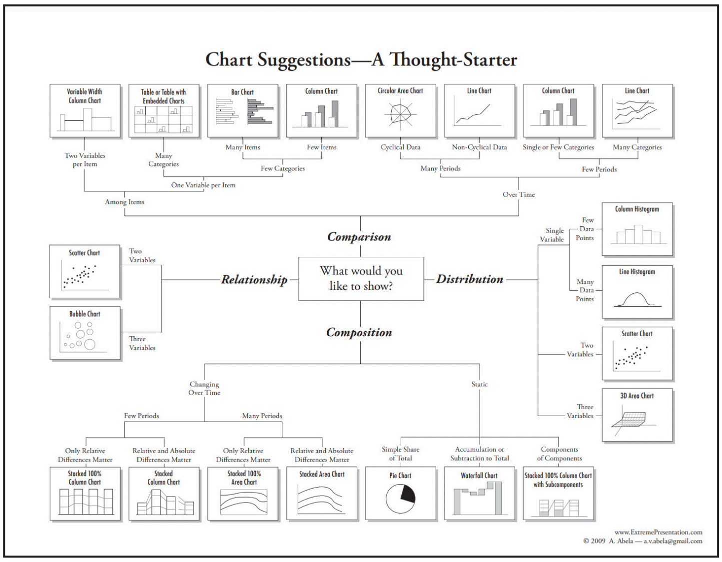
to the Info Guide by George Mason University Library, which aimed to help students find the chart that gets their purpose.[13] It contains 11 catalogues. Since the catalogue is defined as "a systematic list of items," the catalogues without any classification, which are against the definition, are unsuitable for analysis. Also, the duplicated ones are excluded. After sorting, there is a total of 13 catalogues for pre-analysis. These catalogues are:

- → Chart Chooser
  by Andrew Abela (2006)
- → A Periodic Table of Visualization Methods
  by Ralph Lengler & Martin J. Eppler (2007)
- → A Classification of Chart types
  by Jorge Camoes (2013)
- → The Data Visualization Catalogue
  by Severino Ribecca (2013)
- → The Graphic Continuum
  by Jonathan Schwabish and Severino Ribecca (2014)
- → Data Viz Project
  by Ferdio (2015)
- → Visual Vocabulary
  by Financial Times (2015)
- → The Chart Guide
  by Michiel Dullaert (2016)
- → Chartopedia
  by AnyChart (2017)
- → The Chartmaker Directory
  by Andy Kirk (2017)
- → Interactive Chart Chooser
  by Depict Data Studio (2018)
- → From Data to Viz
  by Yan Holtz (2018)

13 "Infoguides: Data Visualization: Visualization Types". 2022. Infoguides. Gmu.Edu. https://infoguides.gmu.edu/data-visualization/types.

→ Carbon Design System - Chart types by Carbon Team (2019).

Looking through the 13 catalogues, we can get a general idea of them. According to the media in use, the chosen catalogues could be classified into static ones and interactive ones. A typical static catalogue, like Chart Chooser, provides a decision diagram to answer the question "What you like to show?".[14] Based on the static table, the whole category system and all visual models could be seen at once. (Figure 2)



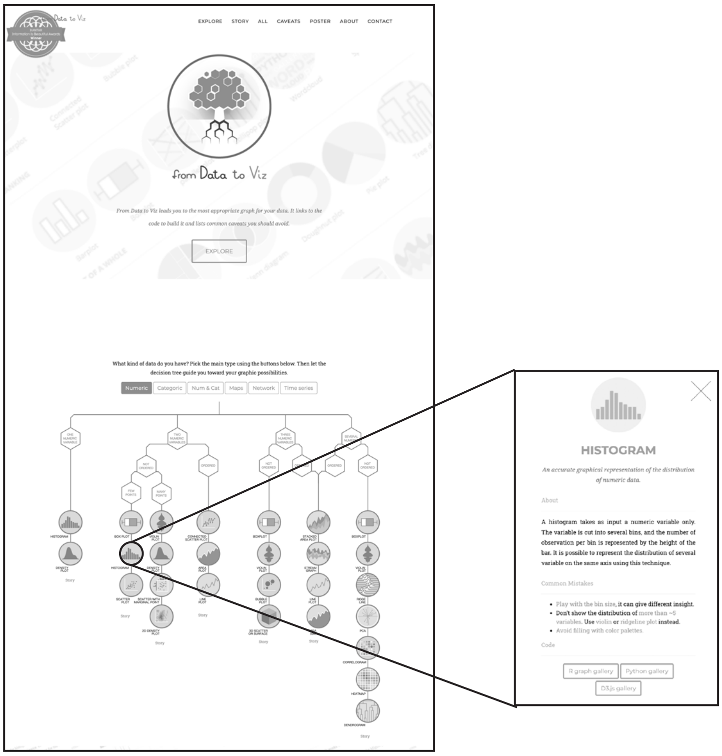
↑ 2 Data Viz Project by Ferdio (2015)

The interactive catalogues, conversely, make the category tree folded and presented as a multiple-leveled website. By choosing one or series of options, users are led to the pages of specific visual models. Compared with the static ones, the interactive catalogues are usually more informative and contain more details of visual models. Like From Data to Viz, it provides not only detailed description, but also

14 Abela, Dr. 2022. "Choosing A Good Chart". The Extreme Presentation(Tm) Method. https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html.

15 Healy, Yan. 2022. "From Data To Viz | Find The Graphic You Need". Data-To-Viz.Com. https://www.data-to-viz.com/.

common mistakes in use and coding solutions for the model. (Figure 3)[15]



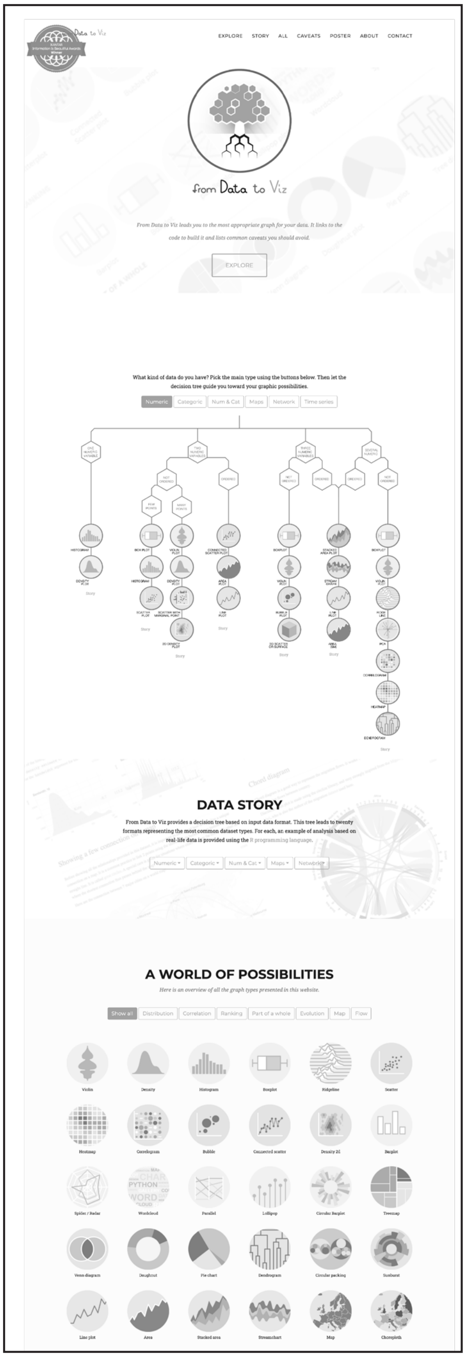
↑ 3 Chart Chooser by Andrew Abela (2006)

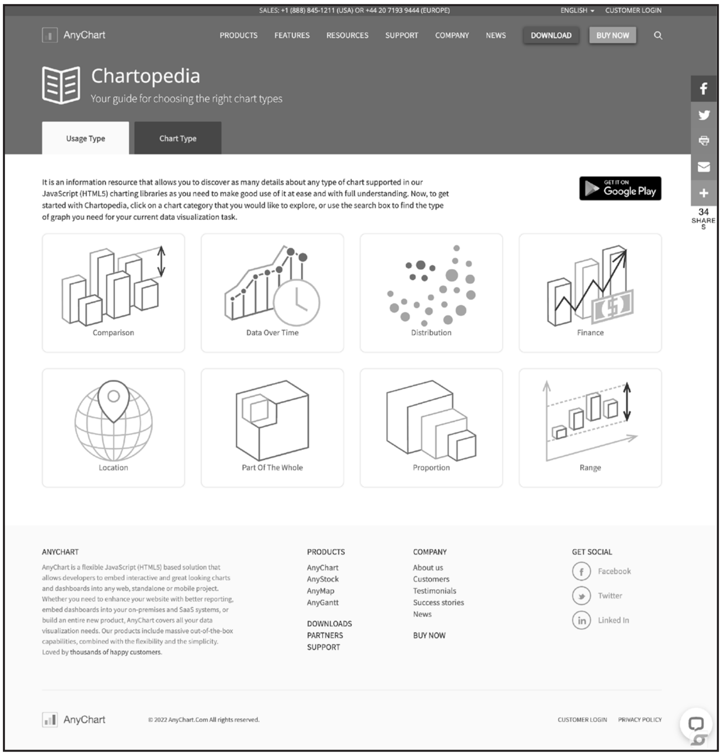
The creators of the catalogues are very diverse. Among those 13 catalogues, ten catalogues are created by individuals for non-profit goals. These individuals are usually designers who focus on data visualization and user experience field. Some of them also have technology or data science background. The other three catalogues are built by institutions. Two of them are commercial institutions, including the famous newspaper Financial Times, and several relatively small companies which provide digital solutions for data visualization as their main business.[16] One of them

16 "Visual-Vocabulary. Pdf /Financial-Times". 2022. Github. https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf.

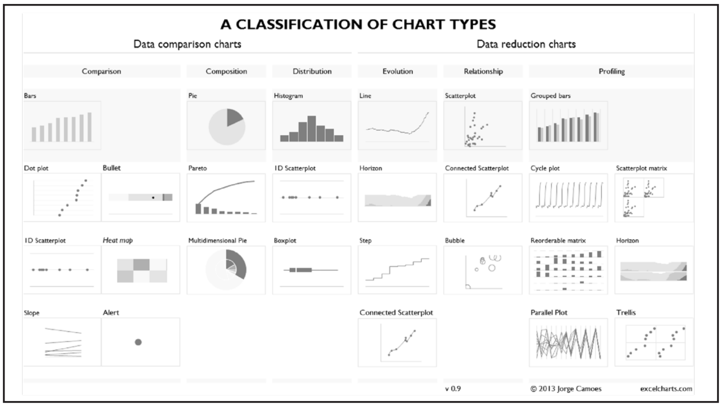
is created as open-source digital assets, while it is sponsored by IBM, the big tech. Due to the diversity of developers, there's no unified standard of how to organize data visualization catalogues. For instance, in some of the catalogues, there's a collection showing all the visual models it includes, while in other catalogues, users can only check on a group of visual models at once. (Figure 4.1) Also, in some catalogues, it provides an overview of the classification system in table or tree diagrams, while in other catalogues, users can only click into each category to find out what is there. (Figure 4.2) Furthermore, some catalogues are published in the format of images, in which the text cannot be recognized automatically. (Figure 4.3)

Pre-analyzed the chosen catalogues, it is obvious to see the ambiguity among data visualization catalogues, which would potentially lead us into a detour while visualizing them. For further processing data, since these selected catalogues are based on different media and share no standard criteria in organizing, we decided to record them manually and restructure them into Microsoft Excel sheets. The detailed configuration of the form will be explained in the following paragraphs.





← 4.1 ↑ 4.2 ↓ 4.3



4.1 Chart Chooser by Andrew Abela (2006)
4.2 Chartopedia by AnyChart (2017)
4.3 A Classification of Chart Types by Jorge Camoes (2013)

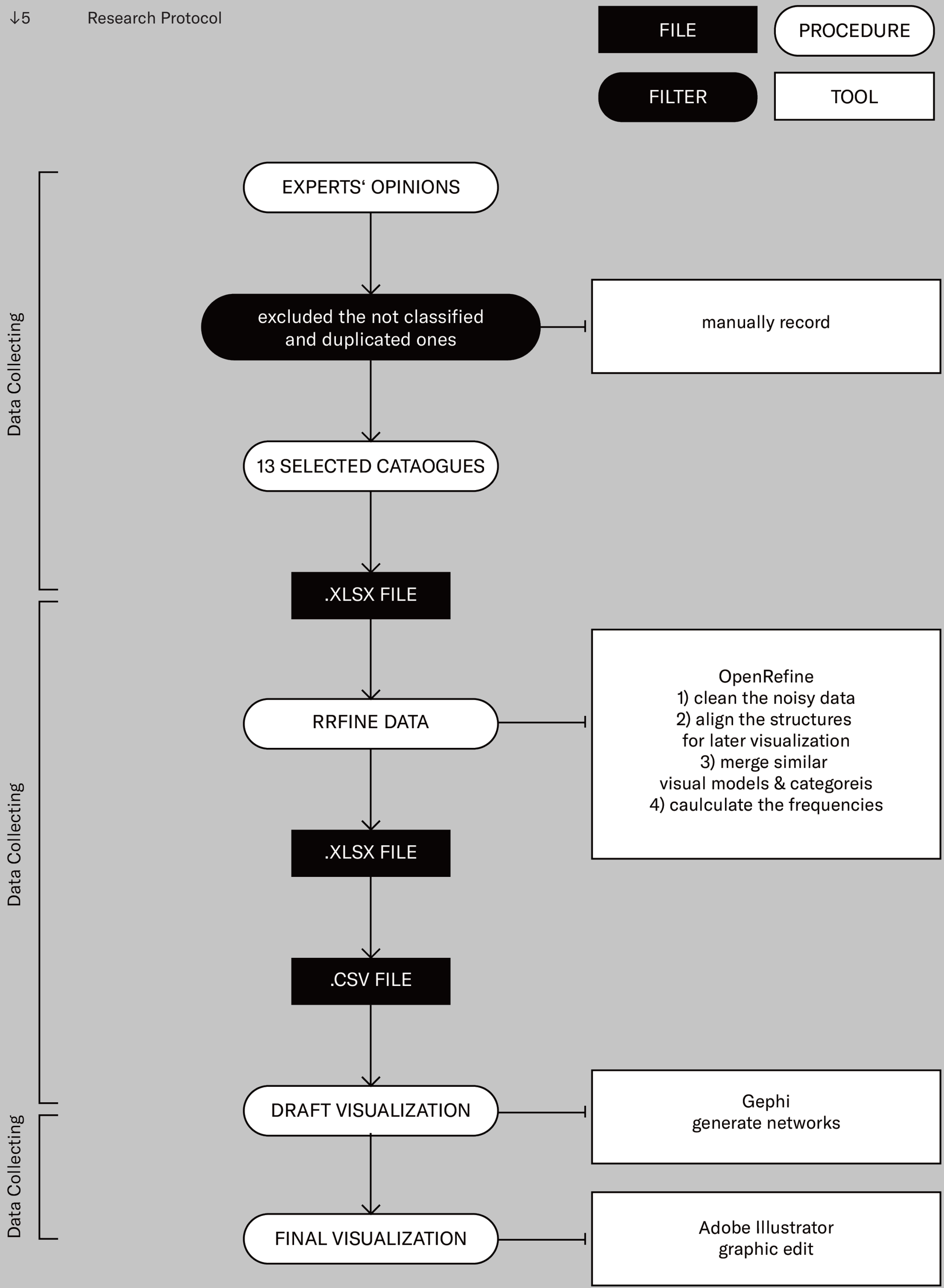
2.2 DATA PROCESSING

From the previous data pre-analysis, we can get that due to the diversity of the catalogue creator's personal experience or business goals, the way they present visual models, the criteria for inclusion or not, and the classification method are very different, which make them difficult to be compared with each other and translate into visualization. Therefore, it requires an approach to unify them into datasets that can be quantified.

Looking through the selected catalogues and pre-analyzing their features, it brings us two goals for the data processing part:

- → Archive the structures of selected catalogues in a unified format.
- → Determine the range of visual models and ctegories that are included in the selected catalogues.

To reach these goals, a data processing protocol is proposed. (Figure 5)


FILE
PROCEDURE
FILTER
TOOL
Data Collecting
EXPERTS' OPINIONS
excluded the not classified and duplicated ones
manually record
13 SELECTED CATAOGUES
.XLSX FILE
Data Collecting
RRFINE DATA
OpenRefine
1) clean the noisy data
2) align the structures for later visualization
3) merge similar visual models & categoreis
4) caulculate the frequencies
.XLSX FILE
.CSV FILE
Data Collecting
DRAFT VISUALIZATION
Gephi
generate networks
FINAL VISUALIZATION
Adobe Illustrator
graphic edit

→ INPUT: filling structured sheets

Firstly, in consideration of the complexity of the taxonomy diversity, we manually checked and recorded all the 14 catalogues. Most of the catalogues only have one level, presented as “category -- visual model”. However, some catalogues, such as the Chart chooser, have more levels, even up to 4. Based on this situation, our form is set like this. (Figure 6) As the screenshot shows, there are six main columns in the form: catalogue name, 1st level, 2nd level, 3rd level, 4th level, and visual model name. Column 5 to column 6 contain other information like established dates and notes, which are not the priority for the analysis. If the recorded catalogue contains four levels of classification, all the cells would be filled; If the catalogue has fewer levels of classification, we use “0” as the placeholder. This strategy is made for the convenience of the following visualizing step, which will be explained later.

	A	B	C	D	E	F	G
1	Name	Year	First level	Second level	Third level	Fourth level	Visual model
2	The graphic	2014	distribution	0	0	0	histogram
3	The graphic	2014	distribution	0	0	0	box & whisker plot
4	The graphic	2014	distribution	0	0	0	violin plot
5	The graphic	2014	distribution	0	0	0	stock
6	The graphic	2014	distribution	0	0	0	confidence interval
7	The graphic	2014	distribution	0	0	0	fan
8	The graphic	2014	time	0	0	0	line graph
9	The graphic	2014	time	0	0	0	sparklines
10	The graphic	2014	time	0	0	0	horizon
11	The graphic	2014	time	0	0	0	stacked area graph
12	The graphic	2014	time	0	0	0	stream graph
13	The graphic	2014	time	0	0	0	history flow
14	The graphic	2014	time	0	0	0	gantt chart
15	The graphic	2014	time	0	0	0	cause-effect
16	The graphic	2014	time	0	0	0	timeline
17	The graphic	2014	time	0	0	0	flow chart
18	The graphic	2014	time	0	0	0	calender
19	The graphic	2014	time	0	0	0	connected scatterplot
20	The graphic	2014	time	0	0	0	arc timeline
21	The graphic	2014	comparing cat	0	0	0	bar chart
22	The graphic	2014	comparing cat	0	0	0	column
23	The graphic	2014	comparing cat	0	0	0	stacked bar graph
24	The graphic	2014	comparing cat	0	0	0	stacked column

↑ 6 The configuration of data collecting sheets

→ REFINE DATA: working with OpenRefine

With the 1241 records, there was a general image of the catalogues, organized in a unified format. The second thing we were interested in is to find out the real range of visual models covered by the 13 catalogues. In order to get deeper into the data, we use OpenRefine, a powerful tool for working with messy data, to filter and clean the dataset.

As the screenshot shows, there are more than five hundred names of visual models, which was exceptionally higher than the range of 70 types of charts (Chartopedia) we found before.

To get a clearer overview of those names, a facet was built in the column of visual model names. Choose “sorted by name,” and we can see a series of visual model names arranged in alphabetical order. While exploring the list, we found that several names can refer to the same visual model, which causes the duplicative and redundant situation. For instance, a chart using vertical bars (column) to show discrete, numerical comparisons across categories can be called “Column” in The Graphic Continuum and Visual Vocabulary, while it is called “Column chart” in Chart Chooser and Chartopedia. Also, it is called “Bar Chart (Vertical)” in the Data Viz Project. In this situation, a single visual model is triple counted. To solve this problem, we first use the cluster function, a feature that helps to find groups of different cell values that might be alternative representations of the same thing. After trying several methods (“key collision” and “nearest neighbor”) and merging the duplicated names, the list of visual model names was trimmed to 514. It excluded the duplication due to singular/plural problems, spelling divergence, and typos. Subsequently, we checked the original web pages of the listed visual model and clustered names refer to the same visual model manually.

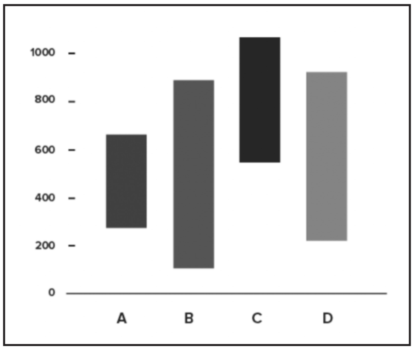
The same workflow is applied to process the categories in the selected catalogues. Since there is no unified visual model classification standard in the field of data visualization, the individuals and institutions mentioned above classify visual models according to their own experience and logic. Therefore, the classification methods they use are not the same. In order to figure out how many potential categories there are and whether there is a range of more commonly used categories, we clustered the first-level root categories that all visual visualization catalogues have, using the same methods described above. The same method. Initially, we got 65 taxonomy names. After correcting singular and plural and spelling differences and merging expressions that refer to the same meaning, the number of valid categories has been reduced to 37. It is a heavy workload, but it is necessary for following data analysis and visualization.

→ CALCULATE THE FREQUENCY: merging like items

After clustering and merging, each name refers to the same visual model is counted as one. There are 314 types of visual models in total. To offer one visual model that has different expressions a unified name, we checked the frequency for each expression and chose the most frequent one. For instance, the visual model in Figure 7 displays a range of data by plotting two Y values per data point. Each Y value used is drawn as the upper, and lower bounds of a column/bar/cylinder.[17] It is mentioned in all the categories for 10 times. It is called as span chart (6 times), hanging rootogram (3 times), range column chart (2 times), range bar chart (2 times), floating bar (1 time), floating

17 "Span Chart | Data Viz Project". 2022. Data Viz Project. https://datavizproject.com/data-type/span-chart/.

bar chart (1 time), column range (1 time), span (1 time). In this case, this visual model is named as “Span chart”, which is mentioned most in all his names, and its frequency is 10.



↑ 7 Span Chart

We also tried to unify the names for categories. The most mentioned category behind each cluster is used to name it. For instance, the category of visualization methods that show data over geographical regions are called spatial (2 times), location (1 time), maps (2 times), location (1 time), geospatial (1 time), and geographical data (1 time). According to our criteria, it could be named “spatial” or “maps.” Since the keyword “spatial” is also mentioned in the other name, “geospatial,” which means it might be a more common expression, we decided it to be the unified name for this category.

After automatically refining and manually editing the original data, we got 13 forms showing the leveled classification of the selected catalogues. Also, there is a final list for all 314 types of visual models and a list of 37 visual model categories, named as their most common name while showing how many times they are mentioned in all catalogues. The selected catalogues are archived in a unified format, and the range of visual models and classification methods that are included in these catalogues are determined.

2.3 VISUALIZATION

The structures of data visualization catalogues and the types of visual models generate thousands of records. By rearranging and analyzing them, they can explain how the catalogues are different from each other and how messy the visual models are named. However, it is not an effective way to convey our idea to the audience. This complex situation requires a better understanding of their relations and hierarchies. It is apparently difficult for users without pre-knowledge about statistics, which is our target audience, to have an understandable and inspiring comparison with those catalogues in such intricate tables. As we mentioned in chapter one, without a clear view of the taxonomy methods of visual models will lead users to detours while they are trying to visualize their own datasets. Also, the complexity of choosing appropriate charts will be a barrier when involving the public in data science practice.

Therefore, the next visualization steps will focus more on how to interpret the structure of the selected data visualization catalogues into a visual artifact that can be understood by the audience. They are required to be sufficiently clear and uniformly formatted so that lay users can easily compare each other and draw similarities and differences. At the same time, the range of the visual models covered by all the 13 data visualization catalogues needs to be clearly demonstrated. Information like the names that each visual model is called, the frequencies of mentioning need to be arranged in a certain order and shown.

Data visualization catalogues are hierarchical structures. As shown in Figure 8, this is the record of the directory Chart Chooser, and a total of 21 visual models have been classified into 4 levels. There are four levels of classification basis, from the more general types of data relationships to the more specific characteristics of the data itself. This is a hierarchy of narrowing targets from broad categories to detailed categories. Therefore, we aimed to find a visual model with the capacity to show complex hierarchical relationships.

	A	B	C	D	E	F	G	H	I
1	Name	Year	First level	Second leve	Third level	Fourth level	Chart types	Reference/C	Frequency
2	Chart choos	2006	comparison	static	one variable	0	single bar chart		14
3	Chart choos	2006	comparison	static	two variable	0	variable width column ch		4
4	Chart choos	2006	comparison	static	three or mo	few items	multiple bar charts		14
5	Chart choos	2006	comparison	static	three or mo	many items	table or table with embe		1
6	Chart choos	2006	comparison	changing ov	many perioc	cyclical data	circular area chart		14
7	Chart choos	2006	comparison	changing ov	many perioc	non-cyclical	line chart		3
8	Chart choos	2006	comparison	changing ov	few periods	single or fev	column chart		7
9	Chart choos	2006	comparison	changing ov	few periods	many items	line chart		3
10	Chart choos	2006	distribution	0	single variat	few items	column histogram		14
11	Chart choos	2006	distribution	0	single variat	many items	line histogram		7
12	Chart choos	2006	distribution	0	two variable	0	scatter chart		14
13	Chart choos	2006	distribution	0	three variab	0	3D area chart		1
14	Chart choos	2006	compositior	static	simple share	0	pie chart		14
15	Chart choos	2006	compositior	static	accumulatio	0	waterfall chart		9
16	Chart choos	2006	compositior	static	components	0	stacked 100% column cha		1
17	Chart choos	2006	compositior	changing ov	few periods	only relative	stacked 100% column cha		1
18	Chart choos	2006	compositior	changing ov	few periods	relative and	stacked column chart		5
19	Chart choos	2006	compositior	changing ov	many perioc	only relative	stacked 100% area chart		2
20	Chart choos	2006	compositior	changing ov	many perioc	relative and	stacked area chart		10
21	Chart choos	2006	relationship	0	two variable	0	scatter chart		14
22	Chart choos	2006	relationship	0	three variab	0	bubble chart		14

↑ 8 The configuration of data collecting sheets

The visual model we considered firstly is dendrogram. A dendrogram is a way to visually represent a hierarchy in a tree-like structure.[18] Generally, the structure of a dendrogram consists of elements such as root nodes, branches, nodes, and leaf nodes. A root node is a member with no superiors/parents; nodes are linked together by lines called branches, representing relationships and links between members. Leaf nodes are members with no children. Dendrogram has a wide range of applications in management, mathematics, and taxonomy.

18 Holtz, Yan. 2022. "Dendrogram | The R Graph Gallery". R-Graph-Gallery. Com. https://r-graph-gallery.com/dendrogram.html.

However, some visual models are multi-vocal categorized. For example, in Ribecca's The Data Visualization Catalogue, the connection map is referenced 6 times, belonging to the six categories of relationship, location, distribution, movement or flow, pattern, and geospatial. This multi-vocal categorized relation can be unreadable in the dendrogram, especially when the catalogue includes a huge amount of visual models. For instance, here's the dendrogram of the Data Visualization Catalogue, which contains 173 visual models. Since the names of visual models are arranged vertically, the diagram is too long to be read, let alone with the crossing links.

The second visualization approach is network. This type of visualization shows how things are interconnected through the use of nodes and link lines to represent their connections and help illuminate the relationships between a group of entities.[19] One node can be attached with several links, which can show the multi-vocal categories well.

In spite of the structure of the 13 data visualization catalogues, we also visualized 314 visual models to highlight how often they were mentioned in all catalogues, i.e., how common they were. We chose a circle. Each circle represents a visual model, and the size of the circle represents the frequency of the visual model being mentioned. We take a similar approach for categories that also need to visualize mention frequency. All these rules and elements together generate the visualization of the data visualization catalogue.

19 "Network Diagram - Learn About This Chart And Tools To Create It". 2022. Datavizcatalogue. Com. https://datavizcatalogue.com/methods/network_diagram.html.

3
RESULT

All 13 data visualization catalogues and the 314 visual models it refers to have been tabulated and represented in the form of visual artifacts. In the end, we have a series of networks and circle graphs. This will provide us with an in-depth look at the current state of the data visualization catalogues, and compare and analyze it.

At the same time, these visualizations will be used as raw materials to make a web-based data multi-experience. This data multi-experience is designed to provide non-expert users with an interactive interface to explore the selected catalogues of data visualizations.

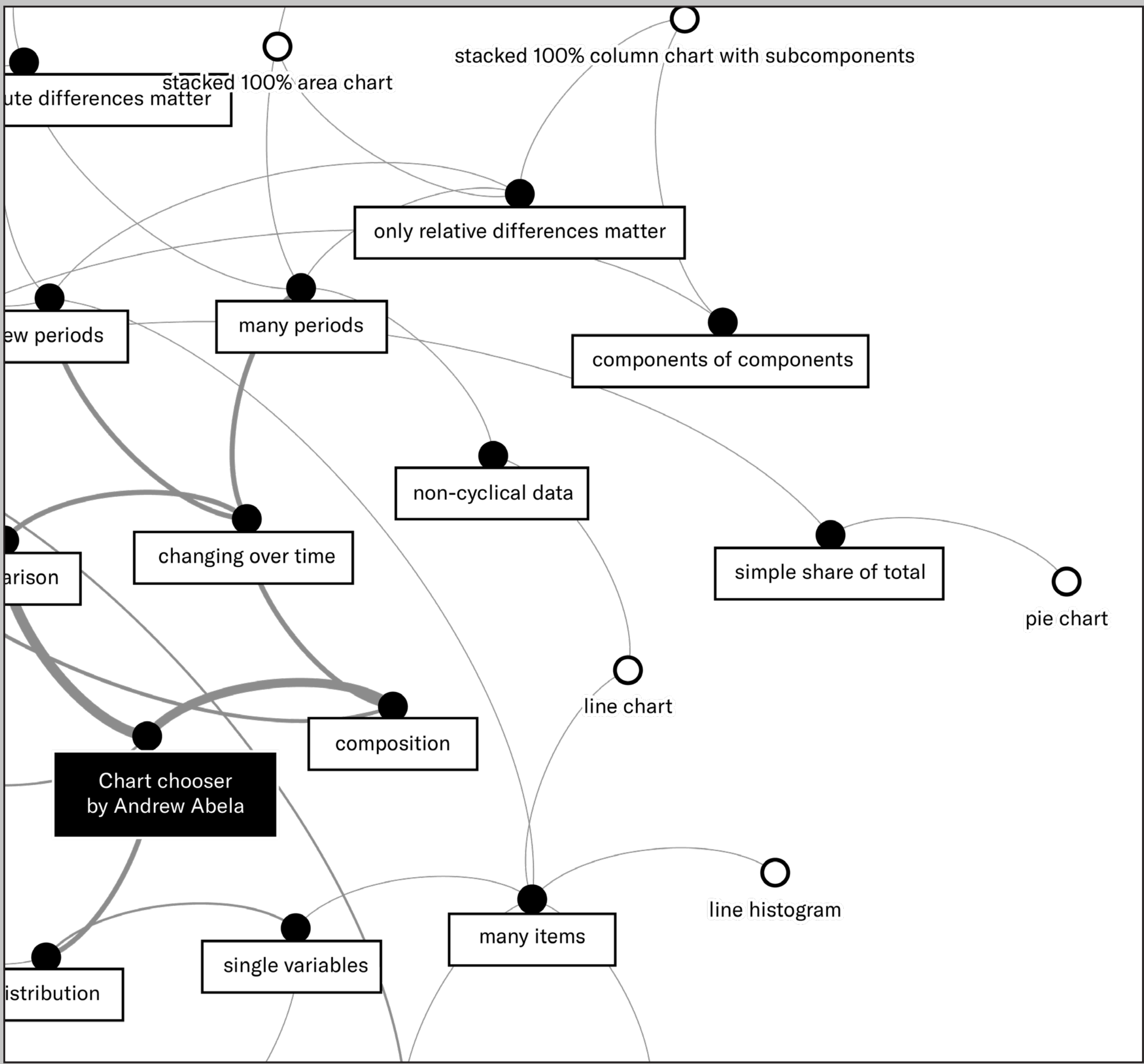
It will explain in a clear and intuitive way the structural differences between the different data visualization catalogues, as well as the range of common visual models available, and allow the user to further view the details of each item.

3.1 VISUALIZATION RESULT

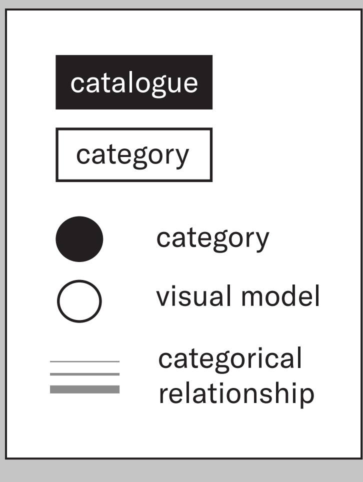
With the development of statistics and computer technology, and the increasing dependence of businesses on big data-based decision-making, data visualization has received more and more attention. However, there is a lack of research on the current visual analysis of relational representations. There is a huge knowledge gap between non-expert users and visual models when using technology to translate information visually. Without classification, it is difficult for users to choose effective techniques to represent data. To make a bridge between them, many data visualization catalogs have been created. However, due to the lack of a unified standard, the expressions of visual models and the types of visual models are easily confused.

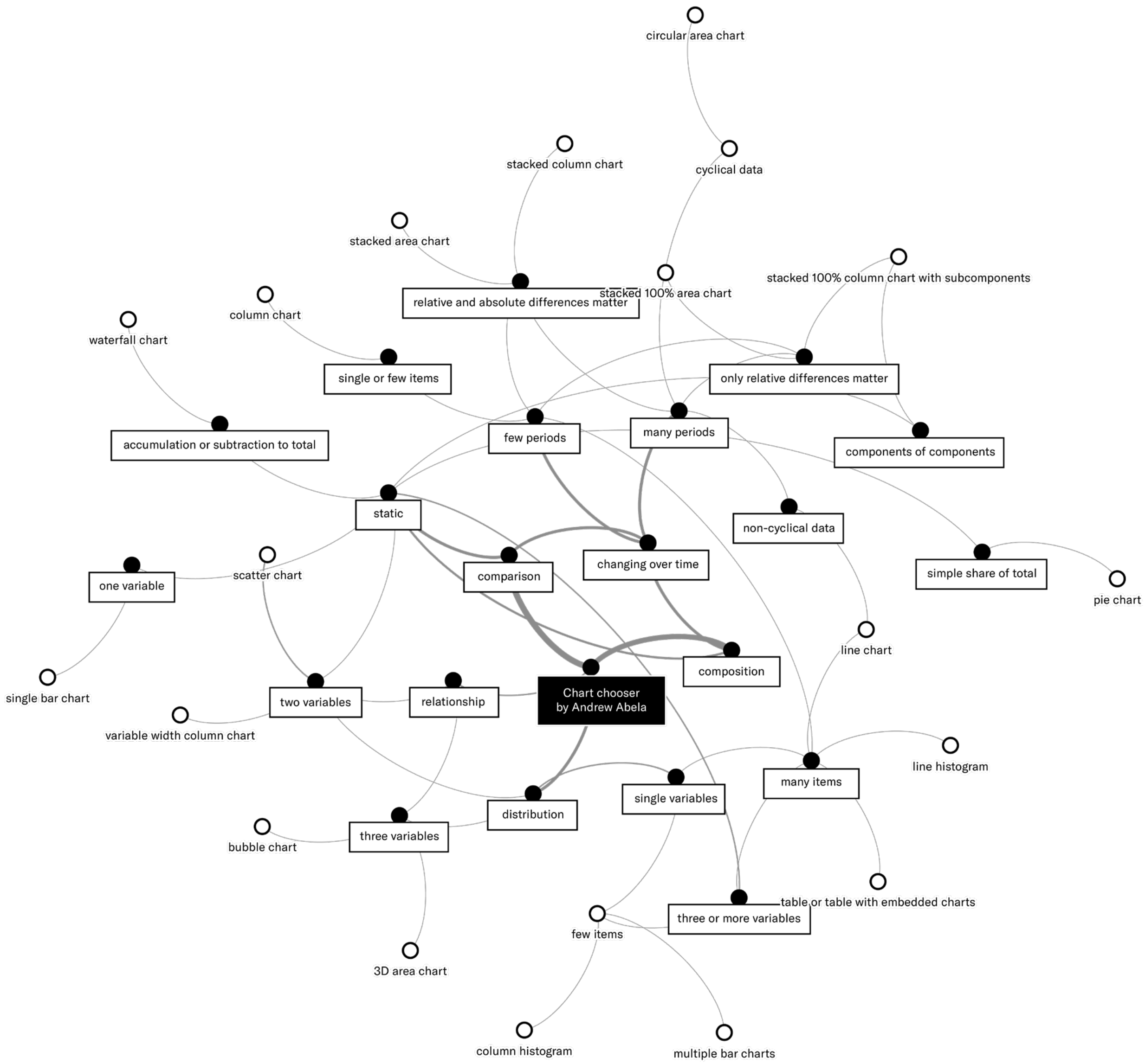
The key to this visualization is how to explain these confusing models and categories in a clear and intuitive way.

Figure 9 shows how a network can be used to represent the structure of a single data visualization catalogue: each solid circle represents a category; each open circle represents a visual model. It can be difficult to categorize structure and visual models, especially when a catalogue is complex. To make the categories stand out from the clusters of visual models, we applied two visual strategies: allowing Gephi to emphasize the edges according to the automatically calculated weights value, and using frames to emphasize the labels of categories. Also, the structures of all 13 data visualization catalogues are shown in following Figure 10-22.



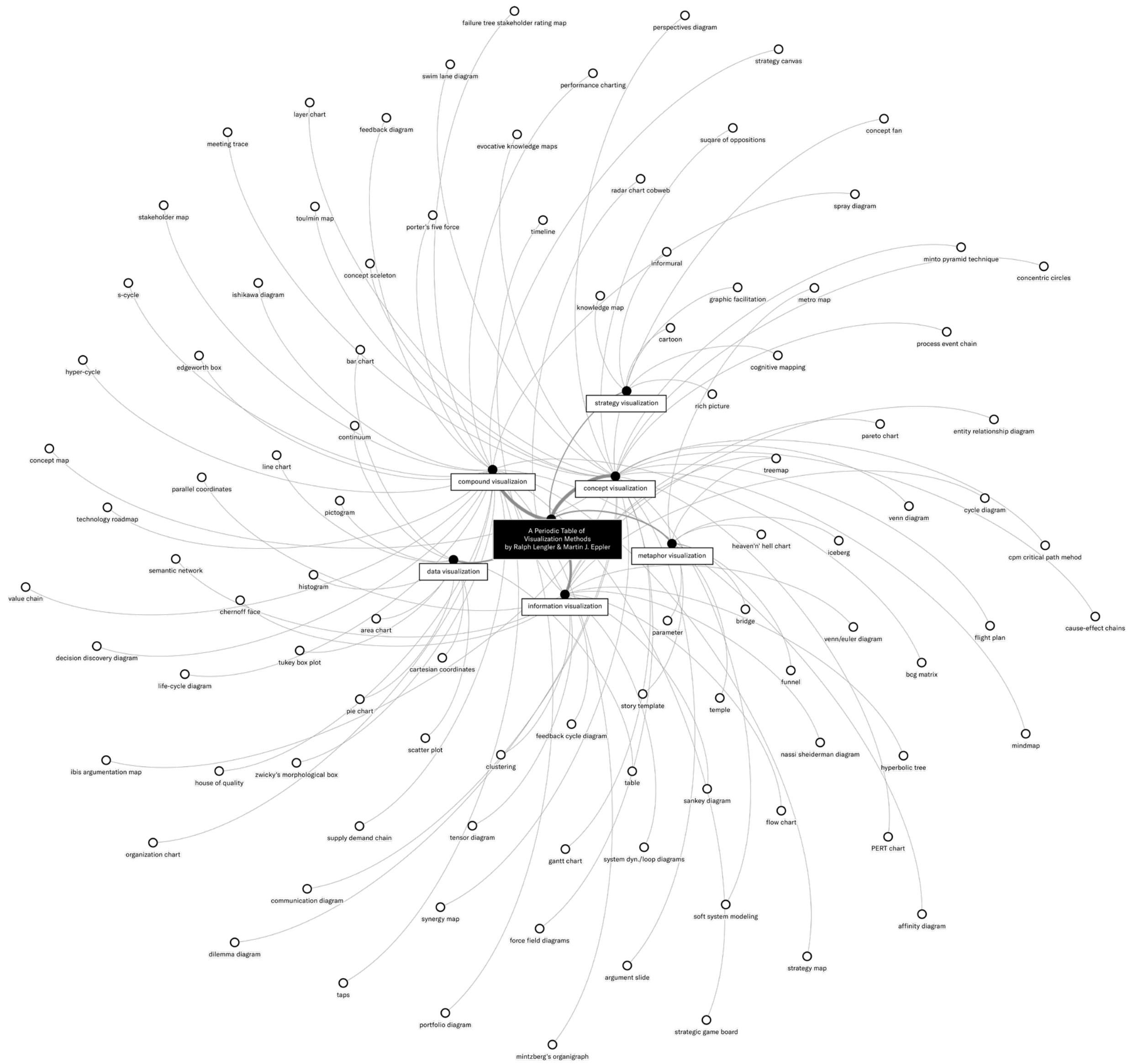
↑ 9 A zoom-in view of the visualization for Chart Chooser



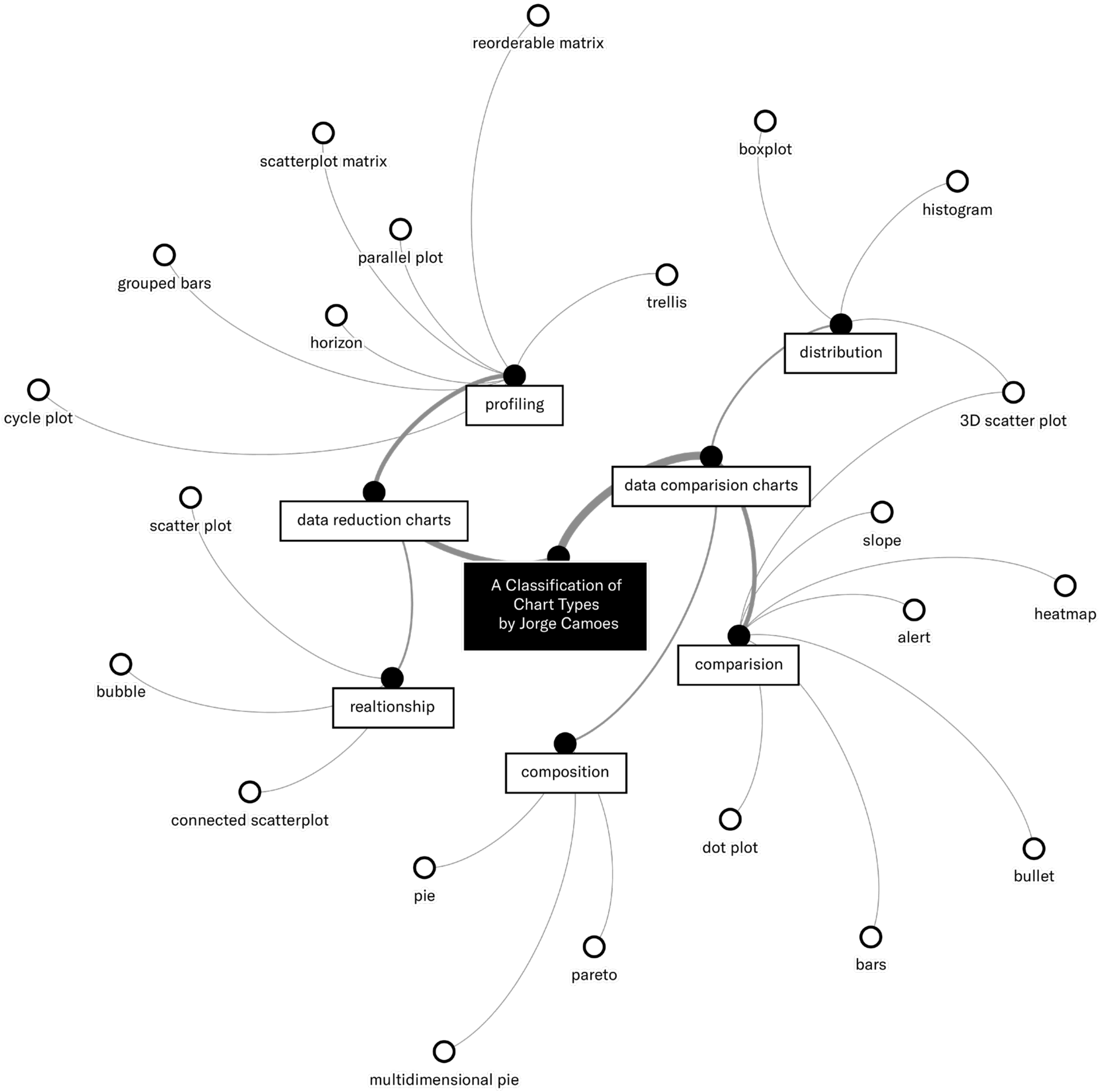


circular area chart
stacked column chart
cyclical data
stacked area chart
stacked 100% column chart with subcomponents
stacked 100% area chart
relative and absolute differences matter
column chart
waterfall chart
single or few items
only relative differences matter
accumulation or subtraction to total
few periods
many periods
components of components
static
non-cyclical data
one variable
scatter chart
comparison
changing over time
simple share of total
pie chart
line chart
composition
single bar chart
two variables
relationship
Chart chooser
by Andrew Abela
variable width column chart
line histogram
many items
single variables
distribution
three variables
bubble chart
table or table with embedded charts
three or more variables
few items
3D area chart
column histogram
multiple bar charts

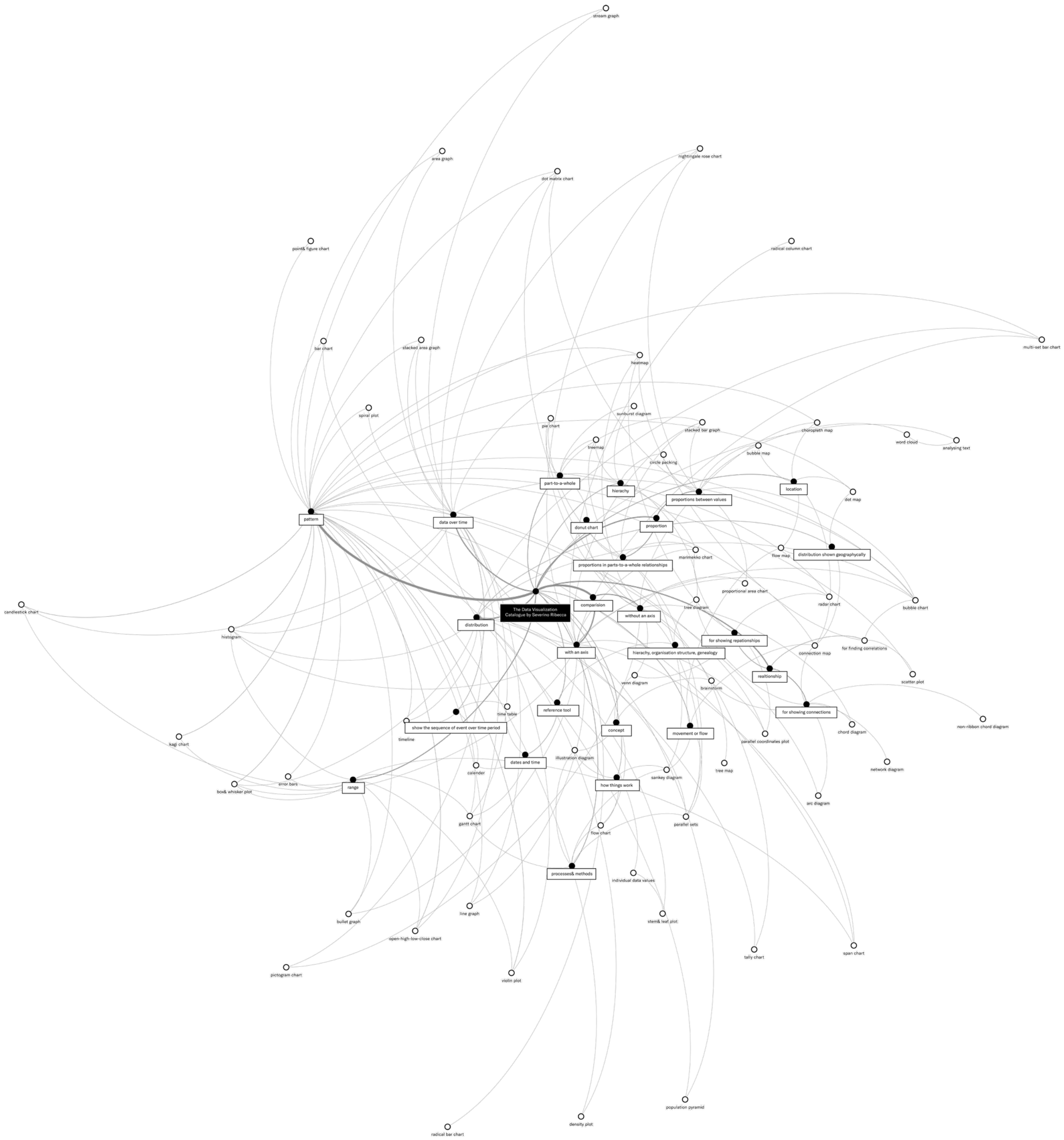
RESULT



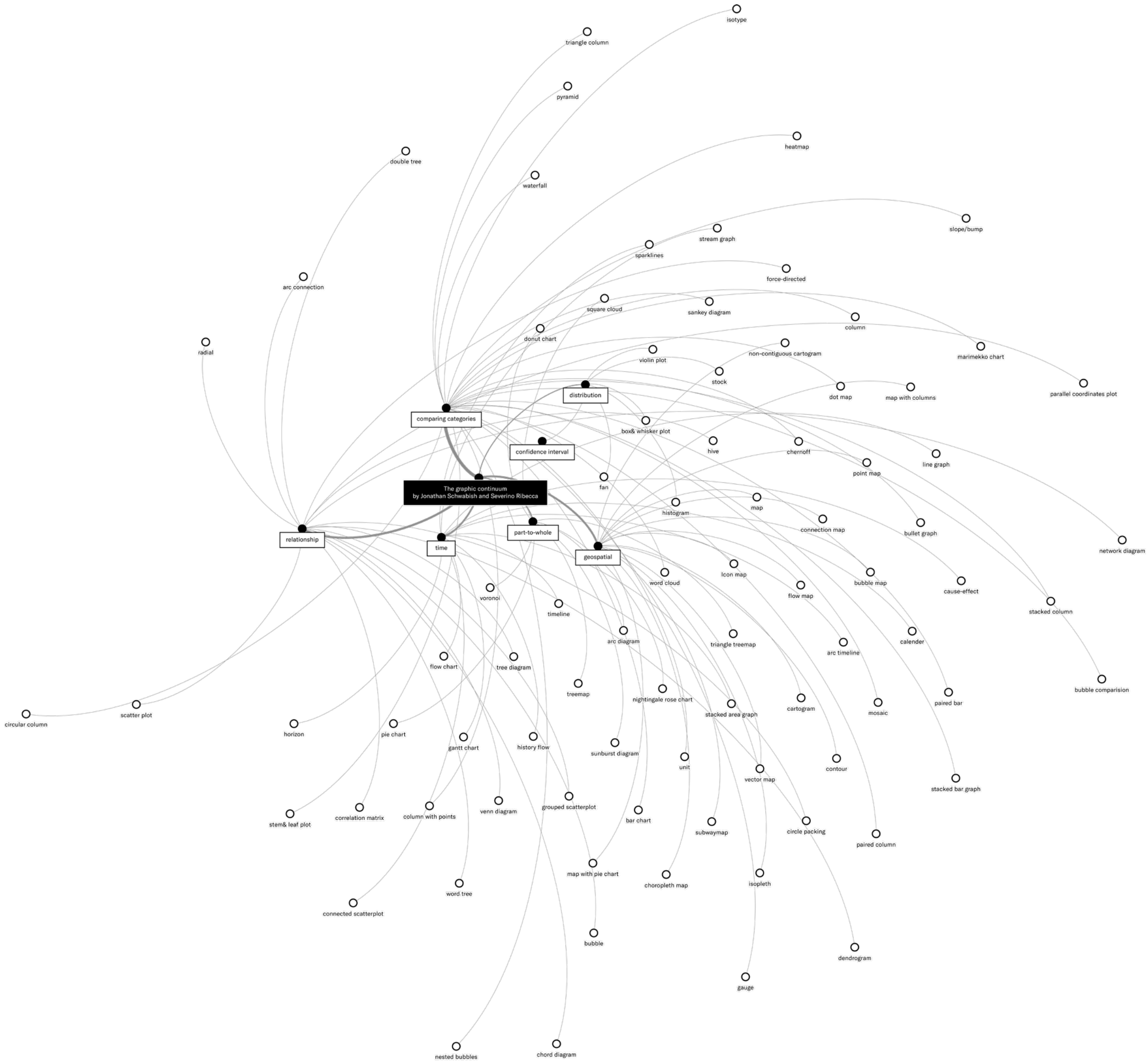
PICK YOUR CHART:
A VISUALIZATION FOR DATA VISUALIZATION CATALOGUES


reorderable matrix
scatterplot matrix
parallel plot
grouped bars
horizon
trellis
cycle plot
profiling
boxplot
histogram
distribution
3D scatter plot
data comparision charts
scatter plot
data reduction charts
A Classification of
Chart Types
by Jorge Camoes
slope
heatmap
alert
comparision
bubble
realtionship
connected scatterplot
composition
pie
pareto
multidimensional pie
dot plot
bars
bullet

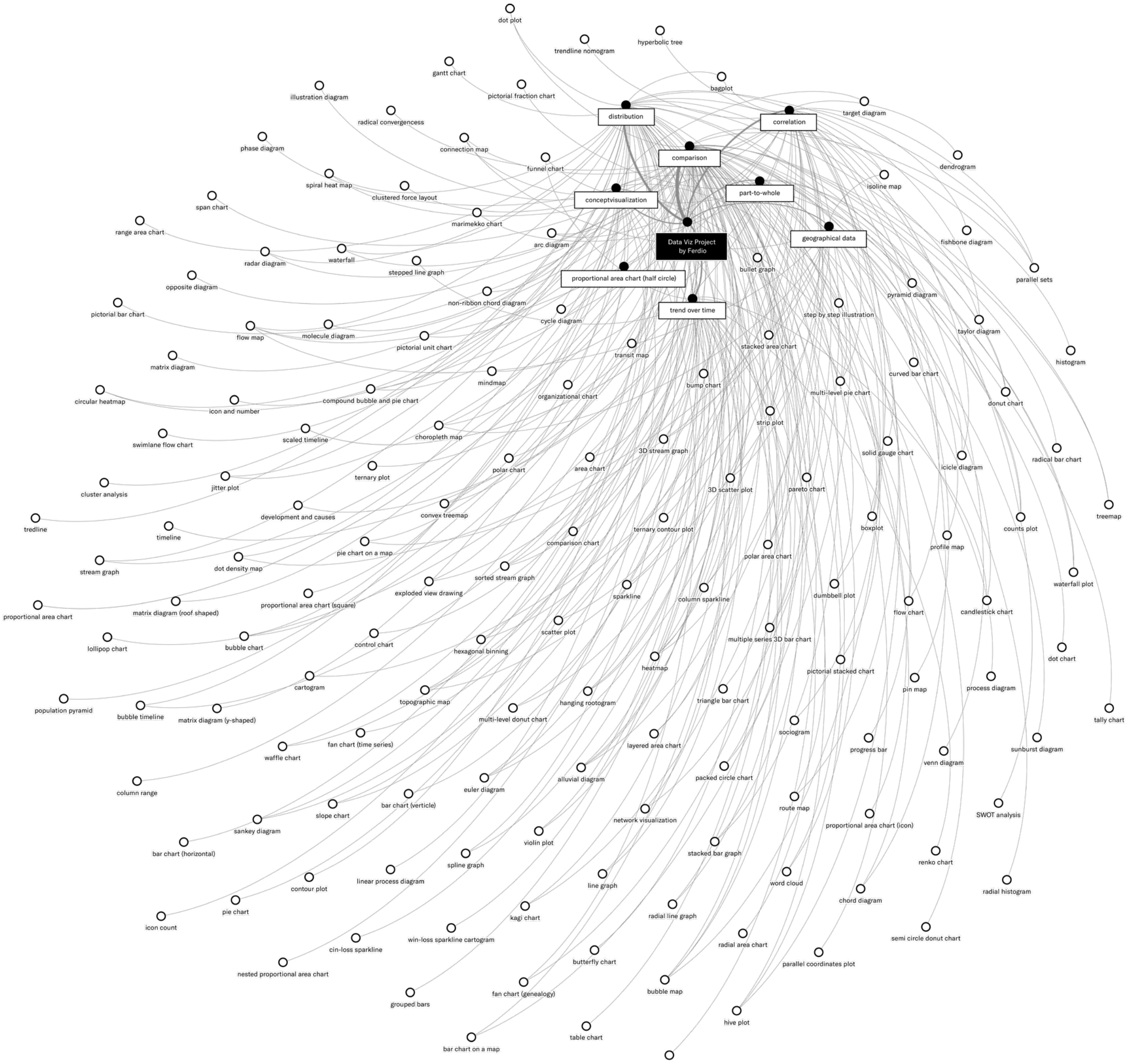
RESULT



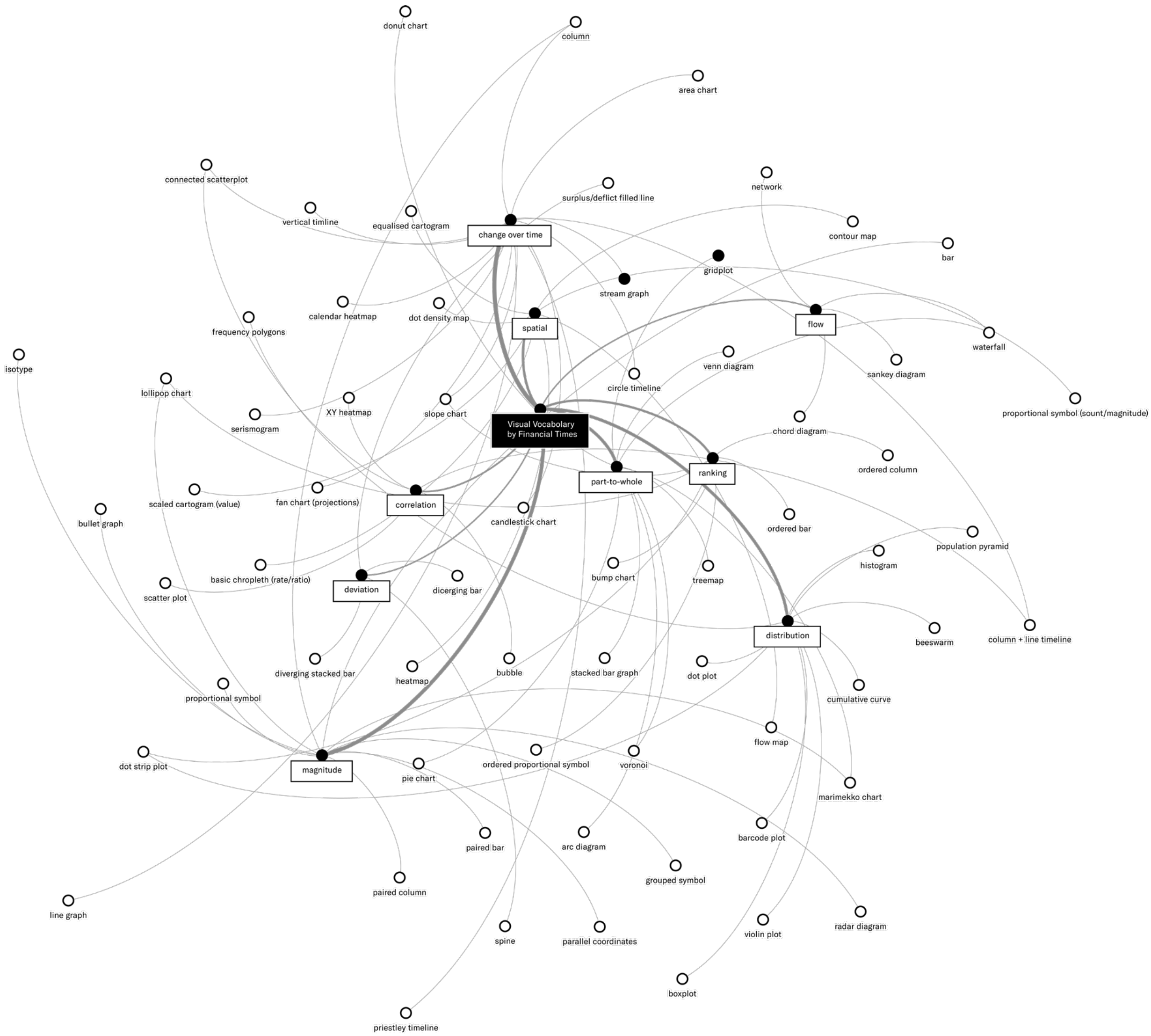
PICK YOUR CHART:
A VISUALIZATION FOR DATA VISUALIZATION CATALOGUES


isotype
triangle column
pyramid
heatmap
double tree
waterfall
slope/bump
stream graph
sparklines
force-directed
arc connection
square cloud
sankey diagram
column
radial
donut chart
marimekko chart
non-contiguous cartogram
violin plot
stock
parallel coordinates plot
distribution
dot map
map with columns
comparing categories
box& whisker plot
confidence interval
hive
chernoff
line graph
point map
fan
The graphic continuum
by Jonathan Schwabish and Severino Ribecca
map
histogram
connection map
bullet graph
network diagram
relationship
part-to-whole
time
geospatial
icon map
word cloud
bubble map
cause-effect
voronoi
flow map
stacked column
timeline
arc diagram
calender
triangle treemap
arc timeline
flow chart
tree diagram
treemap
nightingale rose chart
bubble comparision
paired bar
cartogram
stacked area graph
mosaic
circular column
scatter plot
horizon
pie chart
gantt chart
history flow
sunburst diagram
unit
contour
vector map
stacked bar graph
stem& leaf plot
correlation matrix
column with points
venn diagram
grouped scatterplot
bar chart
subwaymap
circle packing
paired column
map with pie chart
choropleth map
isopleth
word tree
connected scatterplot
bubble
dendrogram
gauge
nested bubbles
chord diagram

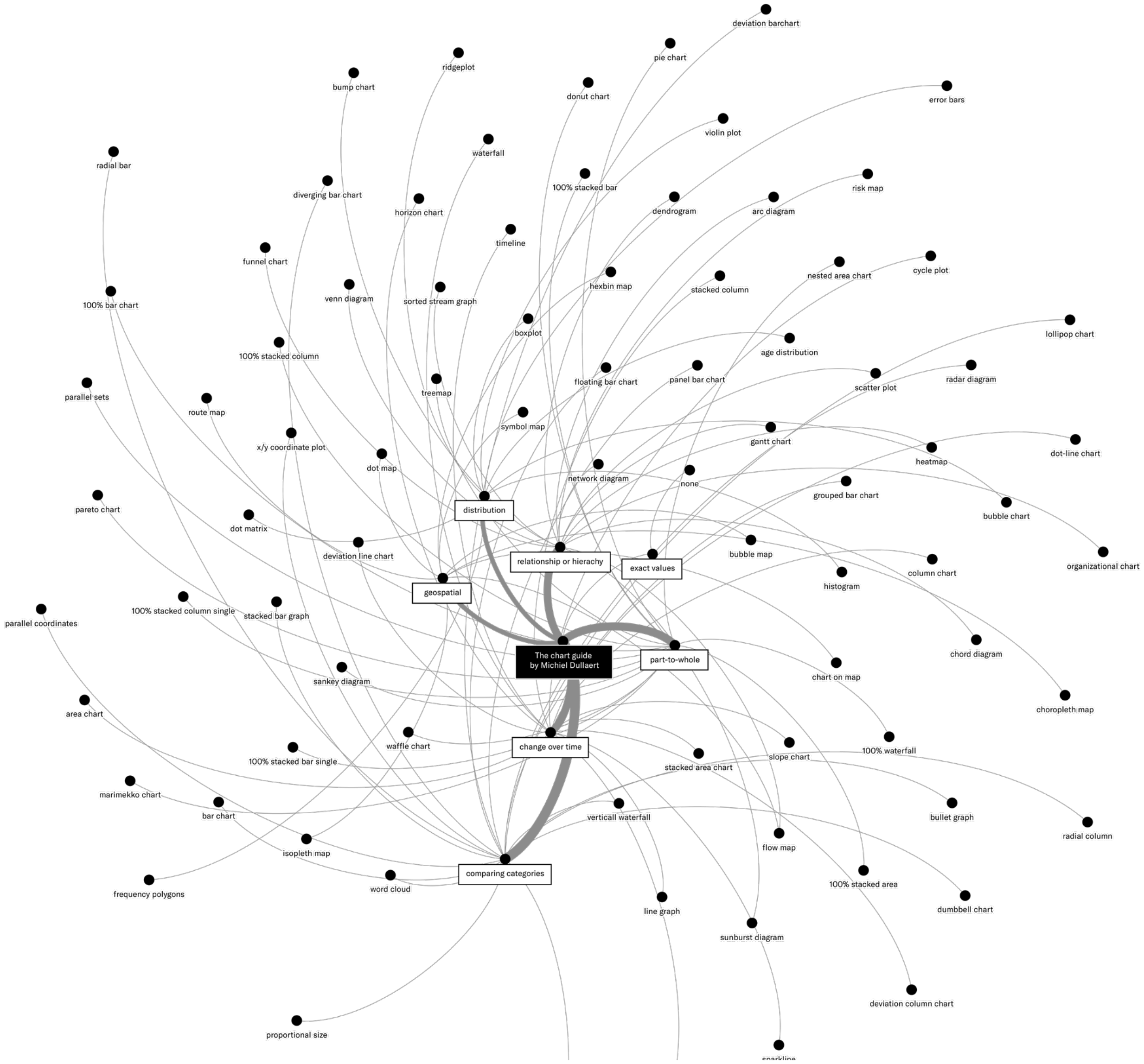
RESULT



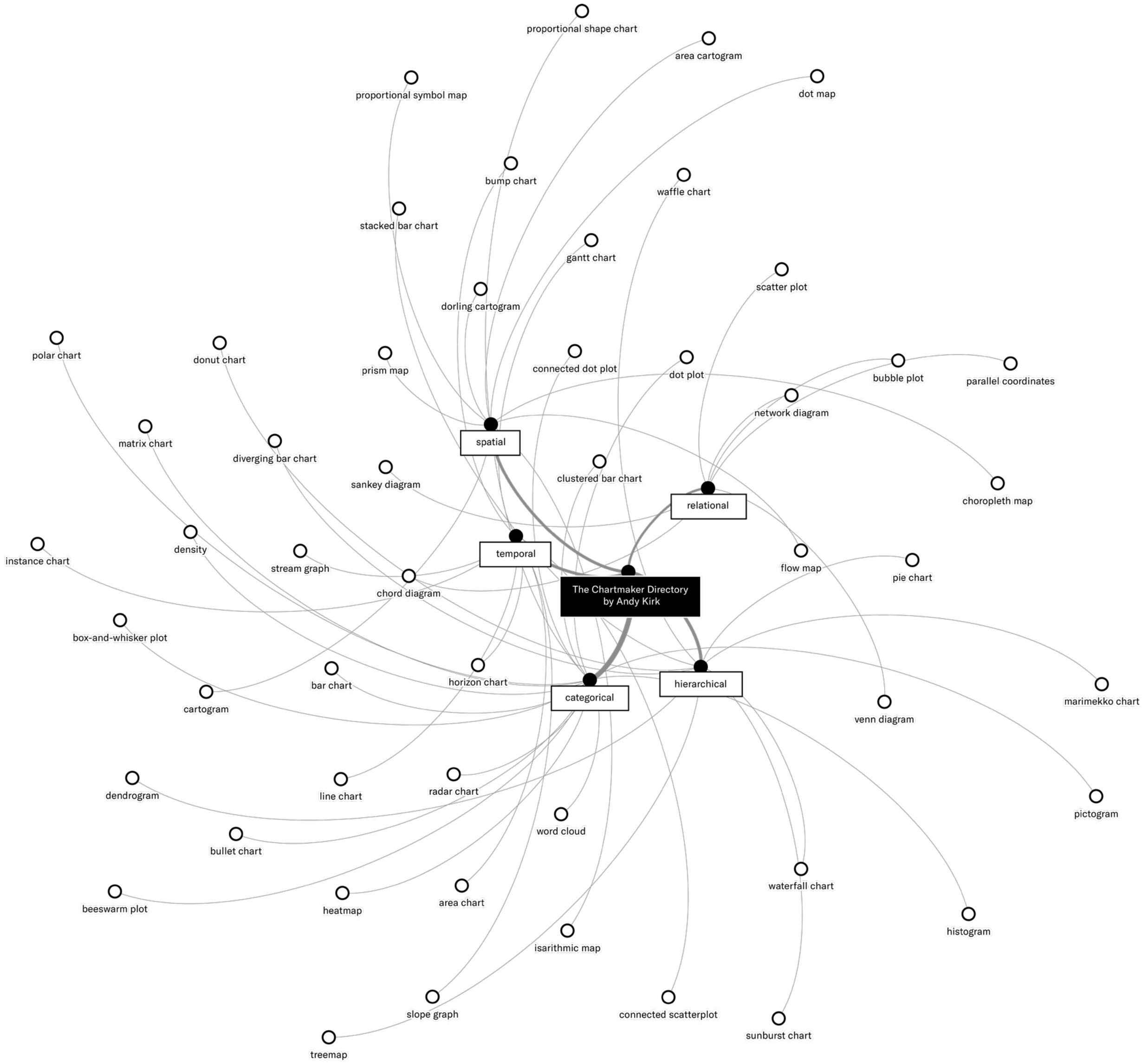
PICK YOUR CHART:
A VISUALIZATION FOR DATA VISUALIZATION CATALOGUES


Visual Vocabolary
by Financial Times
change over time
spatial
flow
ranking
part-to-whole
correlation
deviation
distribution
magnitude
donut chart
column
area chart
connected scatterplot
vertical timline
equalised cartogram
surplus/deflict filled line
network
contour map
bar
gridplot
stream graph
calendar heatmap
dot density map
frequency polygons
waterfall
sankey diagram
proportional symbol (sount/magnitude)
isotype
lollipop chart
serismogram
XY heatmap
slope chart
circle timeline
venn diagram
chord diagram
ordered column
scaled cartogram (value)
fan chart (projections)
candlestick chart
ordered bar
bullet graph
population pyramid
histogram
basic chropleth (rate/ratio)
dicerging bar
bump chart
treemap
scatter plot
beeswarm
column + line timeline
diverging stacked bar
heatmap
bubble
stacked bar graph
dot plot
cumulative curve
proportional symbol
flow map
dot strip plot
pie chart
ordered proportional symbol
voronoi
marimekko chart
barcode plot
paired bar
arc diagram
grouped symbol
paired column
line graph
radar diagram
violin plot
spine
parallel coordinates
boxplot
priestley timeline

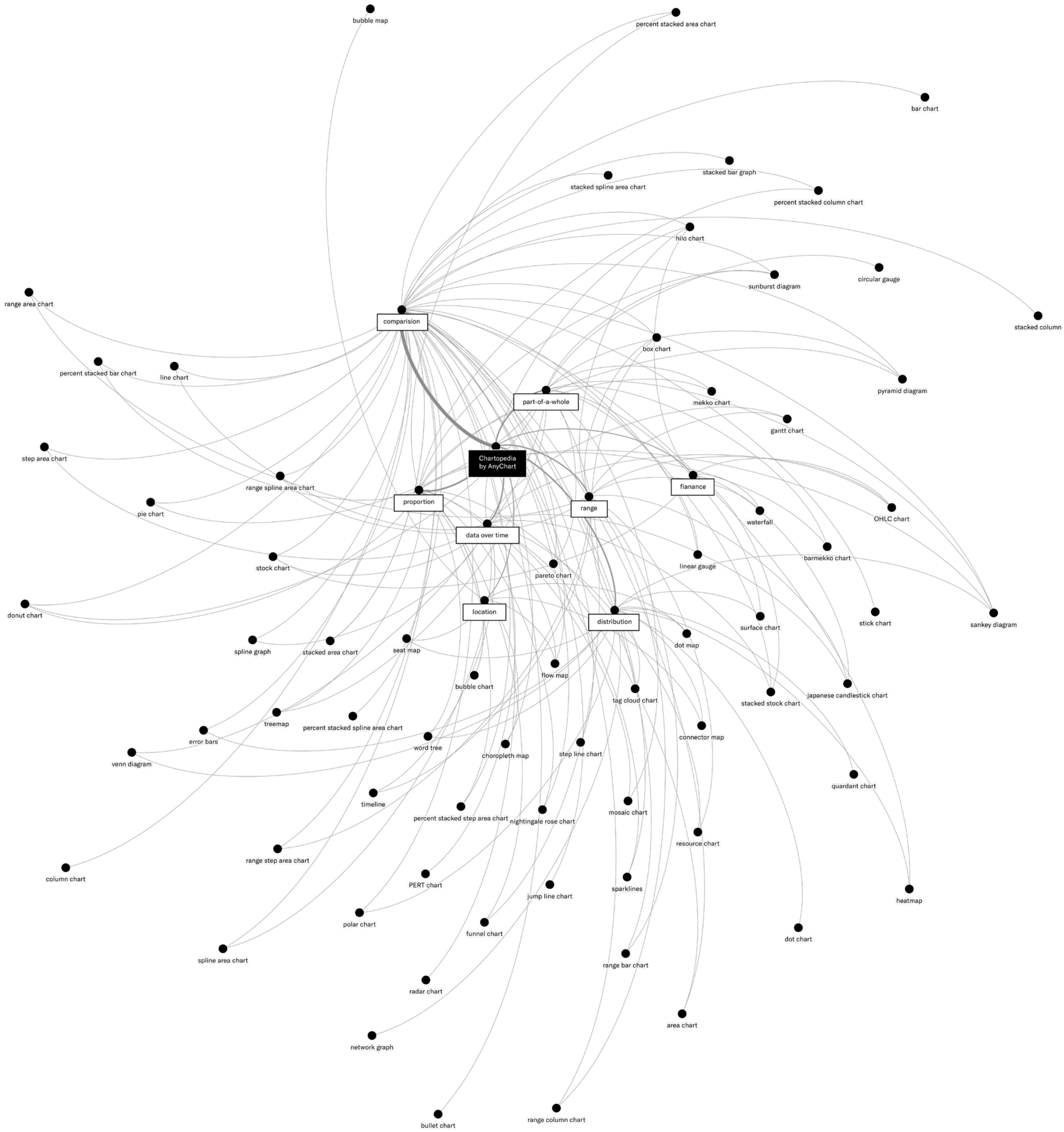
RESULT



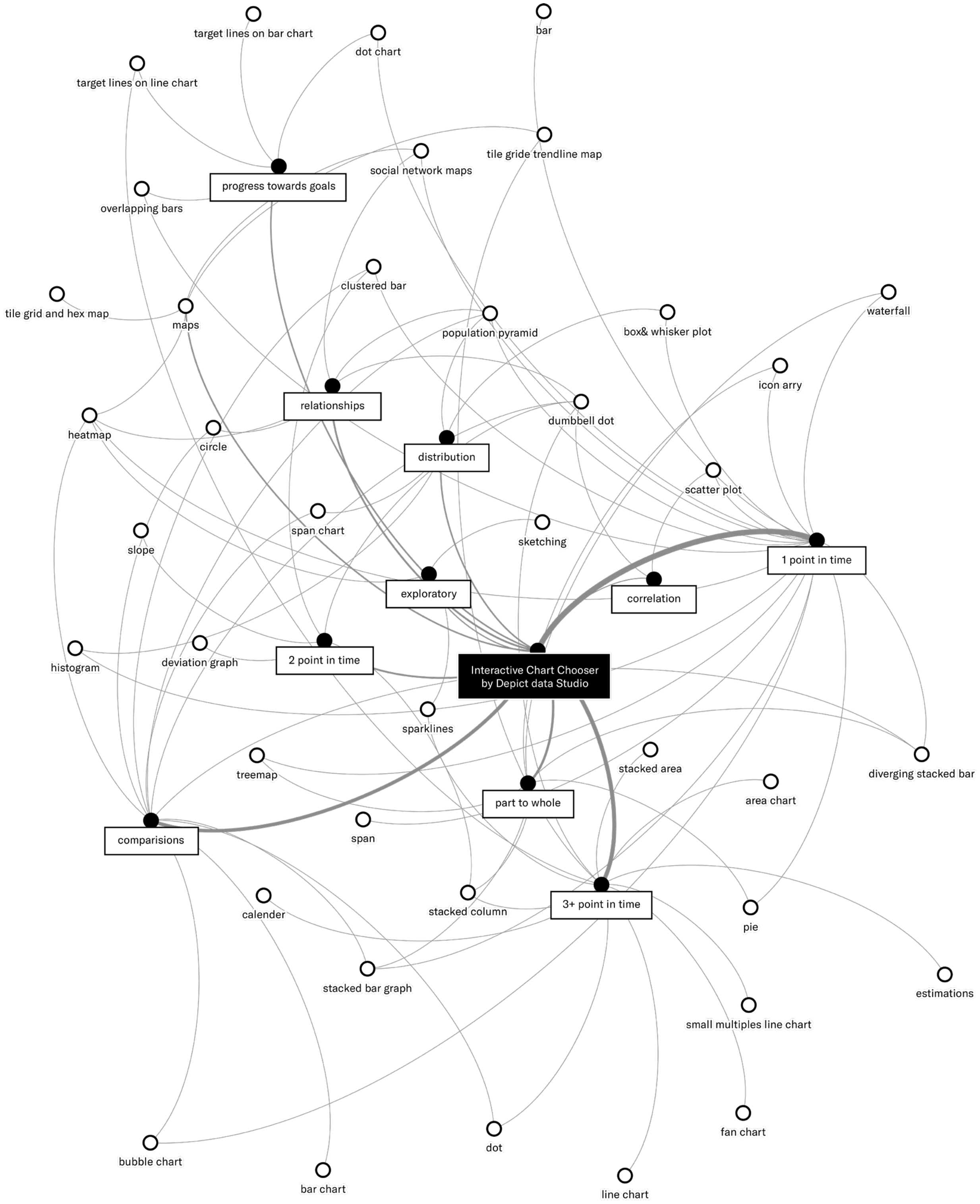
PICK YOUR CHART:
A VISUALIZATION FOR DATA VISUALIZATION CATALOGUES


proportional shape chart
area cartogram
proportional symbol map
dot map
bump chart
waffle chart
stacked bar chart
gantt chart
scatter plot
dorling cartogram
polar chart
donut chart
prism map
connected dot plot
dot plot
bubble plot
parallel coordinates
network diagram
matrix chart
diverging bar chart
spatial
sankey diagram
clustered bar chart
relational
choropleth map
instance chart
density
stream graph
temporal
flow map
pie chart
chord diagram
The Chartmaker Directory
by Andy Kirk
box-and-whisker plot
horizon chart
bar chart
hierarchical
marimekko chart
cartogram
categorical
venn diagram
dendrogram
line chart
radar chart
pictogram
word cloud
bullet chart
waterfall chart
beeswarm plot
heatmap
area chart
histogram
isarithmic map
slope graph
connected scatterplot
sunburst chart
treemap

RESULT

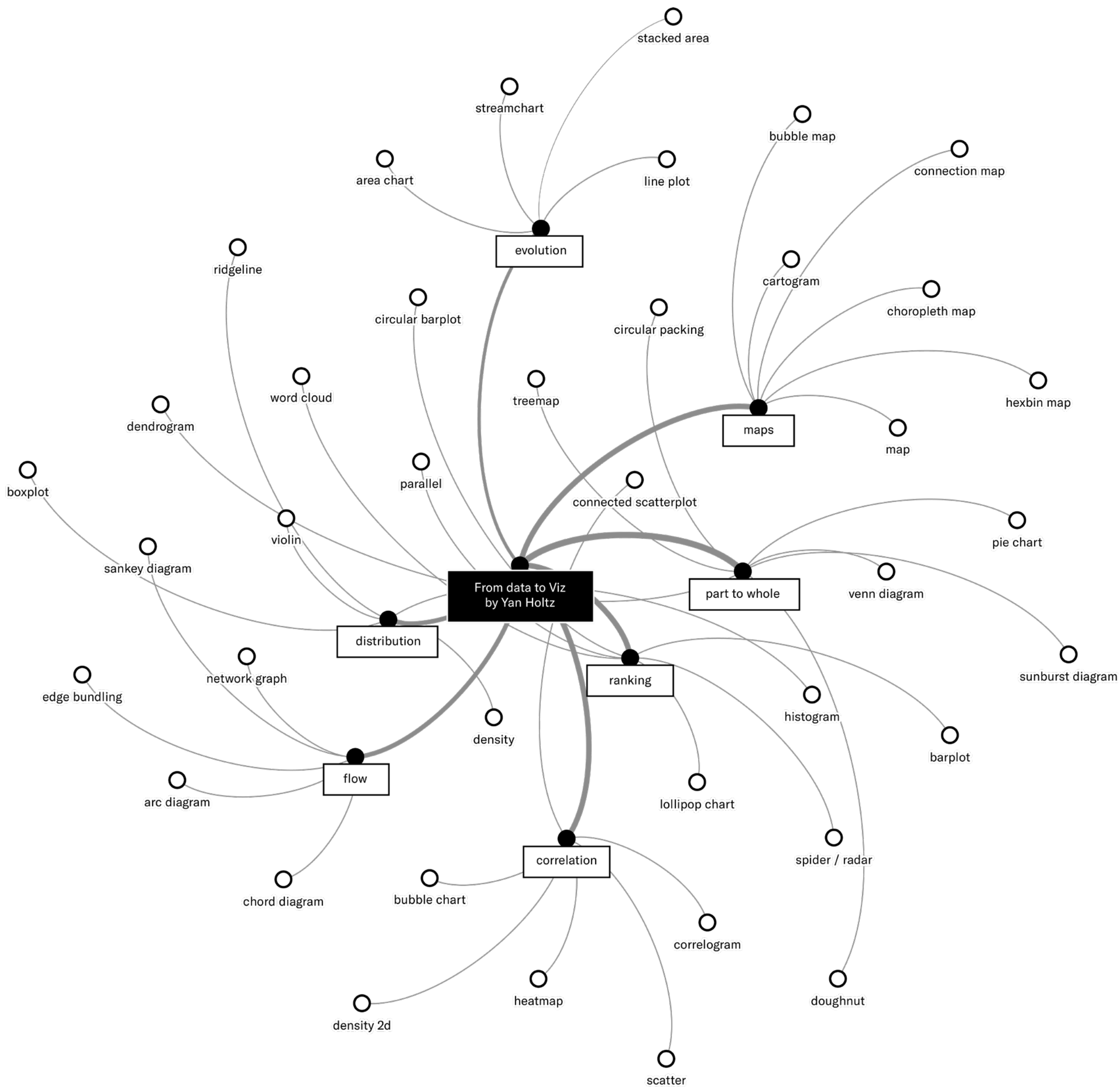


PICK YOUR CHART:
A VISUALIZATION FOR DATA VISUALIZATION CATALOGUES

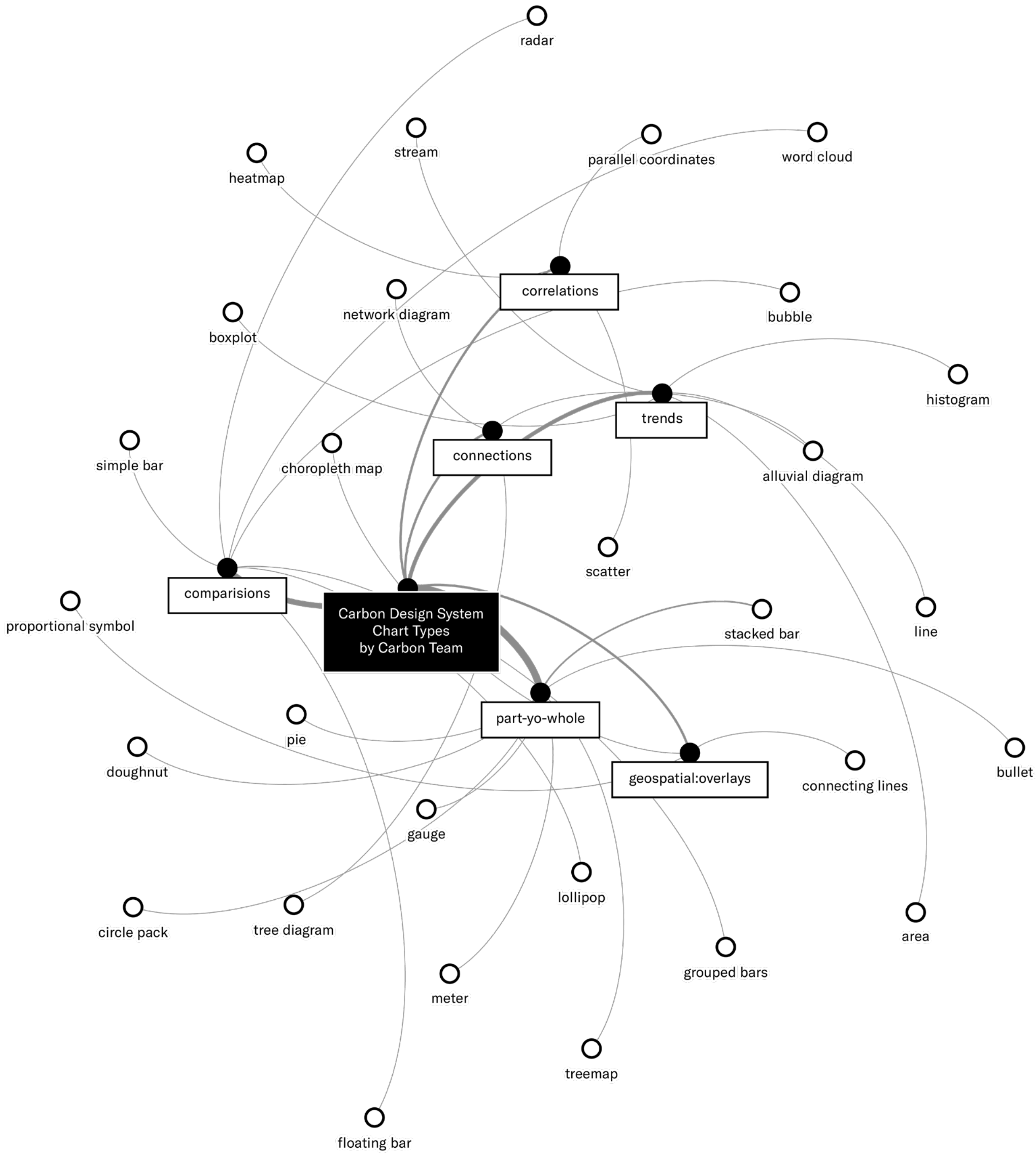


target lines on bar chart
dot chart
bar
target lines on line chart
tile gride trendline map
social network maps
progress towards goals
overlapping bars
clustered bar
tile grid and hex map
maps
population pyramid
box& whisker plot
waterfall
icon arry
relationships
heatmap
circle
dumbbell dot
distribution
scatter plot
span chart
sketching
slope
1 point in time
exploratory
correlation
histogram
deviation graph
2 point in time
Interactive Chart Chooser
by Depict data Studio
sparklines
stacked area
diverging stacked bar
treemap
area chart
part to whole
span
comparisions
calender
stacked column
3+ point in time
pie
stacked bar graph
estimations
small multiples line chart
fan chart
bubble chart
dot
bar chart
line chart

RESULT



PICK YOUR CHART:
A VISUALIZATION FOR DATA VISUALIZATION CATALOGUES


radar
stream
parallel coordinates
word cloud
heatmap
correlations
bubble
network diagram
boxplot
histogram
trends
connections
simple bar
choropleth map
alluvial diagram
scatter
comparisions
Carbon Design System
Chart Types
by Carbon Team
proportional symbol
stacked bar
line
part-yo-whole
bullet
pie
doughnut
geospatial:overlays
connecting lines
gauge
lollipop
circle pack
tree diagram
area
grouped bars
meter
treemap
floating bar

RESULT

Figure 23 shows the categories mentioned under the 37 dat visualization categories. We applied the same translation as above. Each category is represented by a hollow circle, and the size of the circle represents the frequency of the category being mentioned. The highest frequency is 8 times, and the lowest is 1 time.

→ 23 The visulization for categories

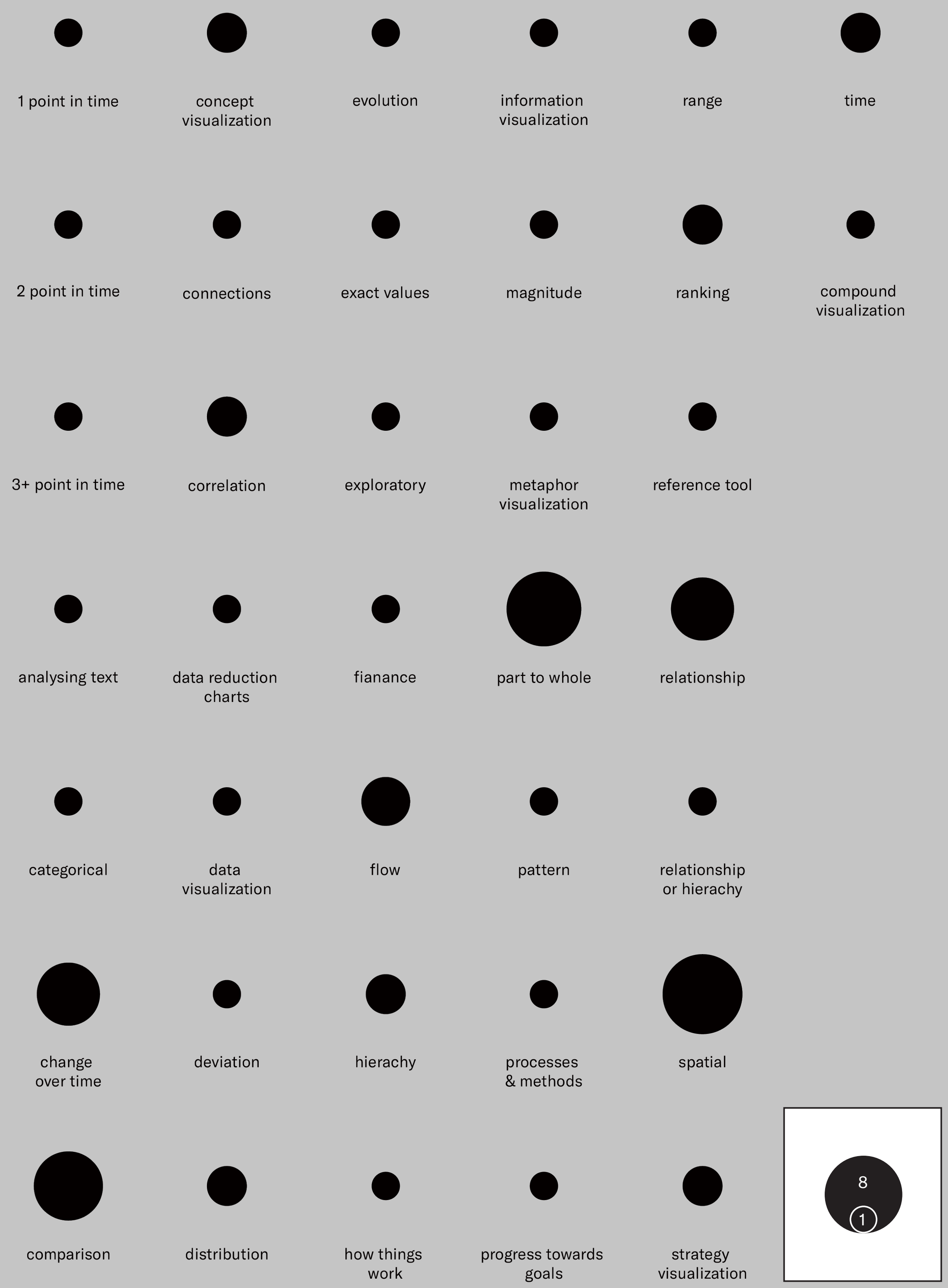

1 point in time
concept visualization
evolution
information visualization
range
time
2 point in time
connections
exact values
magnitude
ranking
compound visualization
3+ point in time
correlation
exploratory
metaphor visualization
reference tool
analysing text
data reduction charts
fianance
part to whole
relationship
categorical
data visualization
flow
pattern
relationship or hierachy
change over time
deviation
hierachy
processes & methods
spatial
comparison
distribution
how things work
progress towards goals
strategy visualization
8
1

Figure 24 shows a total of 314 visual models contained in the selected catalogues. Each visual model is represented by a white circle, and the size of the circle represents the frequency of the visual model being mentioned. The highest frequency is 13 times, and the lowest is 1 time. The frequency of bar chart, grouped bar chart, boxplot, histogram, area cahrt, bubble chart, radar cahrt, scatter plot, treemap, and heatmap are metioned 13 times, which means that they are encluded in all the catalogues.

→ 24 The visulization for visual models


100% bar chart
bar chart
column chart
donut chart
overlapping bars
single bar chart
100% stacked bar
bar chart race
column with points
error bars
pie chart
span chart
100% stacked bar single
barmekko chart
compound bubble and pie chart
grouped bars
pie chart on a map
stacked bar
100% stacked column
barplot
deviation barchart
marimekko chart
population pyrami
stacked column
100% stacked column single
boxplot
deviation graph
multi-level donut chart
radical bar chart
sunburst diagram
100% waterfall
bullet graph
diverging bar chart
ordered bar
semi circle donut chart
verticall waterfall
arc diagram
circular column
diverging stacked bar
ordered column
simple bar
waterfall


100% stacked area
circle packing
dot matrix
line chart
polar chart
slope chart
area chart
clustered force layout
dot plot
lollipop chart
radar chart
sparklines
baseline chart
connected scatterplot
dumbbell dot
nested bubbles
radar chart cobweb
spline area chart
beeswarm
density plot
grouped scatterplot
parallel coordinates
range area chart
spline graph
bubble chart
difference chart
histogram
parallel sets
range step area chart
stacked area chart
bubble map
dot
horizon char
percent stacked spline area chart
scatter plot
step area chart
circle
dot map
jitter plot
percent stacked step area chart
scatterplot matrix
violin plot

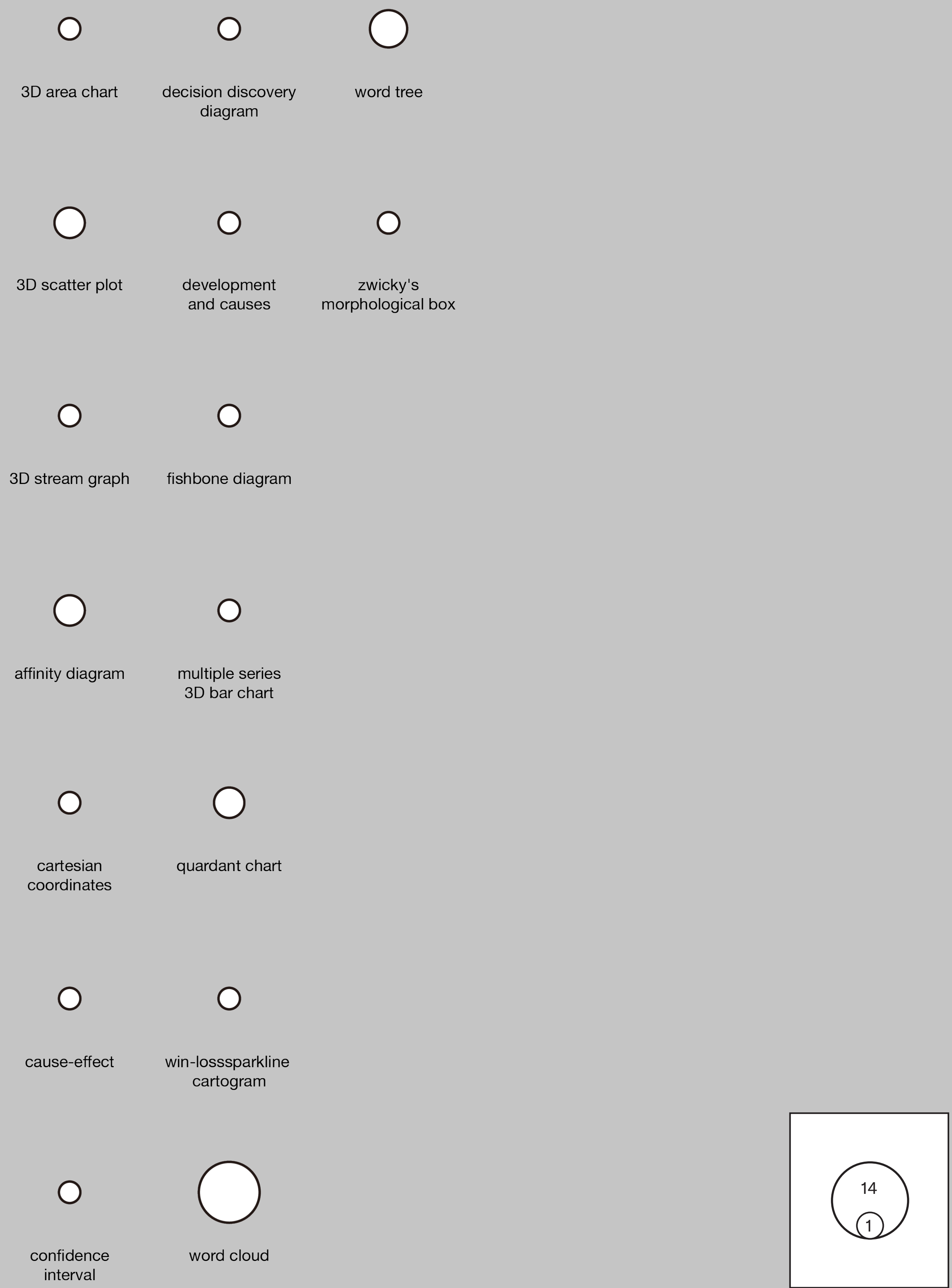

alert
bcg matrix
dendrogram
gauge
non-ribbon chord diagram
spiral heat map
alluvial diagram
bridge
double tree
heatmap
ordered proportional symbol
strip plot
arc timeline
bubble timeline
edge bundling
icon and number
pictogram
time table
argument slide
calendar heatmap
estimations
linear gauge
priestley timeline
timeline
bagplot
cartogram
euler diagram
mindmap
ridgeplot
treemap
barcode chart
chord diagram
fan chart
network diagram
scaled timeline
venn diagram
basic chropleth (rate/ratio)
column + line timeline
gantt chart
nightingale rose chart
solid gauge chart
voronoi


PERT chart
pictorial bar chart
process event chain
proportional symbol (sount/magnitude)
risk map
sociogram
panel bar chart
pictorial fraction chart
profile map
proportional symbol map
s-cycle
soft system modeling
parameter
pictorial stacked chart
proportional area chart
pyramid diagram
sankey diagram
spray diagram
pareto
pin map
proportional area chart (half circle)
radical convergencess
seat map
stacked 100% column chart with subcomponents
performance charting
point & figure chart
proportional area chart (icon)
renko chart
semantic network
stacked stock chart
perspectives diagram
porter's five force
proportional area chart (square)
reorderable matrix
serismogram
stakeholder map
phase diagram
portfolio diagram
proportional symbol
rich picture
sketching
waffle chart


SWOT analysis
story template
surface chart
taps
ternary contour plot
triangle bar chart
sorted stream graph
strategic game board
swimlane flow chart
target diagram
ternary plot
triangle column
stem & leaf plot
strategy canvas
symbol map
target lines on bar chart
tile gride trendlinemap
triangle treemap
step by step illustration
strategy map
synergy map
taylor diagram
toulmin map
table or able with embedded charts
step chart
stream graph
system dyn. loop diagrams
technology roadmap
tredline
value chain
stick chart
supply demand chain
table
temple
trellis
vector map
stock chart
suqare of oppositions
tally chart
tensor diagram
trendline nomogram
vertical timline


100% stacked area
circle packing
dot matrix
line chart
polar chart
slope chart
area chart
clustered force layout
dot plot
lollipop chart
radar chart
sparklines
baseline chart
connected scatterplot
dumbbell dot
nested bubbles
radar chart cobweb
spline area chart
beeswarm
density plot
grouped scatterplot
parallel coordinates
range area chart
spline graph
bubble chart
difference chart
histogram
parallel sets
range step area chart
stacked area chart
bubble map
dot
horizon char
percent stacked spline area chart
scatter plot
step area chart
circle
dot map
jitter plot
percent stacked step area chart
scatterplot matrix
violin plot


100% stacked area
circle packing
dot matrix
line chart
polar chart
slope chart
area chart
clustered force layout
dot plot
lollipop chart
radar chart
sparklines
baseline chart
connected scatterplot
dumbbell dot
nested bubbles
radar chart cobweb
spline area chart
beeswarm
density plot
grouped scatterplot
parallel coordinates
range area chart
spline graph
bubble chart
difference chart
histogram
parallel sets
range step area chart
stacked area chart
bubble map
dot
horizon char
percent stacked spline area chart
scatter plot
step area chart
circle
dot map
jitter plot
percent stacked step area chart
scatterplot matrix
violin plot


3D area chart
decision discovery diagram
word tree
3D scatter plot
development and causes
zwicky's morphological box
3D stream graph
fishbone diagram
affinity diagram
multiple series 3D bar chart
cartesian coordinates
quardant chart
cause-effect
win-losssparkline cartogram
confidence interval
word cloud
14
1

With the above visualization results, even lay users who are not familiar with the field of data visualization can clearly understand and compare different data visualization catalogues. This practice of providing non-expert users with a large number of case comparisons in a specific domain can help them gain a quick understanding of the domain and establish a simple global perspective. By comparing these catalogues, users can get a general idea of which visual models are the most common and which visual model classifications are the most commonly used - which can help users decide which visual model to use in their visualization practice. At the same time, observing and comparing multiple existing data visualization cataloues can enable users to quickly build a general idea of various data visualization catalogues, so that they can selectively use and critically think about such taxonomic cases.

3.2 VISUALIZATION RESULT

In the Data Visualization Result section, we get the networks of the structure of the selected data visualization categories, the visual models contained in the categories, and the circle graphs of the categories. These visualizations are not only joyful visual artifacts but also a path for us and users to dive deep into the situation of current data visualization taxonomy. In the next sections, we will look into these visualizations, trying to answer the two research questions in Chapter 1.

→ VISUAL MODELS

Figure 13 shows the names of the 314 visual models and the frequencies with which they are mentioned in all catalogues. The frequency is up to 13, which means that every directory mentions this visual model. Among them, the visual models whose frequency reaches 13 are:

Bar chart, Grouped bars, Boxplot, Pie chart, Histogram, Area chart, Bubble chart, Radar chart, Scatter plot, Treemap, Heatmap

Since the above visual models are mentioned in each catalogue, they can be considered commonly used models with high recognition.

Of all 313 visual models, 188 appeared only once. We looked and found that most of them are from A Periodic Table of Visualization Method. It is a catalogue representing visual methods in the form of the periodic table of chemical elements. The periodic table received some criticism, saying it abuses the structure of the chemistry periodic table and includes too many useless business analysis diagrams, which are not common. Our finding can also be evidence: visual models contained by it have a little intersection with other catalogues. There are also visual models composed of two or more common models, such as the “bubble map,” which is only mentioned once in the Data Viz Project, which is simply a combination of a bubble chart data visualization and a map used to visualize location and proportion simultaneously. Theoretically, by combining basic visual models, there can be infinite composite models, which might be the reason why they are rarely mentioned.

In addition to the models mentioned the most and least, 2 visual models were mentioned 12 times, 3 visual models were mentioned 11 times, and 74 visual models were mentioned 3-10 times, which means they are also common models with relatively high public acceptance.

As mentioned earlier in the Data Processing section, it is also a notable phenomenon that a visual model corresponds to multiple expressions. The original quantity of visual models' names was over 600 before refining, which is 1.5 times of the actual visual models. Most of the multiple naming is due to spelling or word usage. For example, whether to add suffixes ("Bar" and "Bar Chart") and what suffixes to add (chart, graph, plot, or diagram). These multiple namings are easy to be recognized: the keyword "bar" keeps them consistent. However, in some cases, it's difficult to find the similarity of keywords. For instance, the butterfly chart and population pyramid are both names of a special type of bar chart. The name of the butterfly chart comes from the butterfly-like shape, while it is also called the population pyramid for its function of representing population gender distribution. They do not share any keywords while they are actually used to refer to the same model.

According to this finding, we can assume that there is also a naming problem while users practice visualization since different catalogues or data visualization tools use different expressions to represent. The lack of unified terms can be confusing, especially for beginners of data visualization while they are searching for appropriate visual models.

Moreover, sometimes several visual models similar in the visual dimension can be confused. For example, there is confusion between the bar chart, column chart, and histogram is very common. The bar chart is a visual model using rectangular bars to represent values on the X-axis. The column chart is a vertical version of the bar chart. Only the x-axis and y-axis are switched. For the histogram, although it looks the same as the column chart, is totally different things. The histogram displays continuous developments over an interval, while the bar chart represents discrete data and answers the question of "how many?" in each category. When they are included in data visualization catalogues, in The Data Visualization Catalogues, "Bar Chart" refers to both car chart and column chart, while the Visual Vocabulary distinguishes these two. Looking into the more complex models we mentioned before, and we can also see that the boundaries between columns and histograms are blurring. For example, the span chart is also called "Range Column (in Chartope" ia)" and "Floating bar (in Carbon Design System)."

→ CATEGORIZATION METHODS

For the categorization method analysis, we look into the visualization of the general structure of catalogues and their specific categories.

As shown in Figure 12, after visual interpretation, the 13 data visualization catalogues take on different shapes. The more white nodes a data visualization catalogue has, the more visual models it contains. The figure shows that the directory containing the most significant number of data visualization models is the Data Viz Project created by Ferdio in 2015, which contains 152 branches in total. The more visual models included, the more comprehensive the range of knowledge covered by the catalogue, which means it provides wider choices for users looking for solutions for their data.

In spite of the leaf nodes, the branches are also key elements. The categorization structures are up to 4 levels and down to 1. In order to provide an easy-to-use decision-making path, these structures of catalogues are usually interactive. By sorting specific categories and reading the explanation, users could be aware of which visual model to choose. The more levels of categorizations are, users can choose the more detailed factors.

Another significantly different feature is the pattern of branches in catalogue trees. It shows the structure and complicity of the catalogues. In some visualization, the branches are tangled like twines. The multi-vocal categories cause it. For instance, in The Data Visualization catalogues made by Ribbeca, the catalogues that have the most complex structure, we can see that the visual model “Sankey diagram” is included in 4 different categories, which are proportion, how things work, processes & methods, and movement for flow. The function of the single visual model is not fixed. Depending on the context, the same visual model can be used to explain different events or relationships, which suggests the flexibility of visual interpretations.

These selected data visualization catalogues identify 37 categories in all. Some of the data visualization catalogues explain the reasons for this on their about page. Looking at these categories, we can roughly classify them into three categories:

According to relations
According to data types
According to fields

Some authors say that they classified the models according to the usage of visualization goals ("what would you like to show?", like the Chart Chooser asks). However, we merge them into "according to relations", since their primary purposes are still to show the relations in the dataset. These relations include categories like spatial, comparison, part to whole, change over time, distribution, etc. The categories made according to data types take the feature of the original dataset as the standard. It includes categories like 1 point in time, 2 points in time, and 3+ points in time. Although it is not very common, some catalogues, such as finance, take data source fields as classification evidence.

Also, it is interesting to see that all the categories that are mentioned more than 1 time are all categories made according to relations. It means that relationship is a common categorical standard for these catalogue creators. Most of the catalogues use relation as a factor to classify visual models, while some tacalogues, such as Chartopedia and Interactive Chart Chooser, identify categories of multiple types of classification methods.

According to the findings above, we can see many features of current data visualization catalogues. The range of included visual models is ambiguous. There is a lack of a "must have" list of visual models and a unified standard for naming these models. This problem can make non-expert users lost in the massive amount of models, increasing the difficulty of searching for the proper technique. For the categorization methods used by selected catalogues, they can be divided into three types. The most common method is classifying by relation, while several catalogues also combine other methods, such as classifying by data types or fields, as additional options.

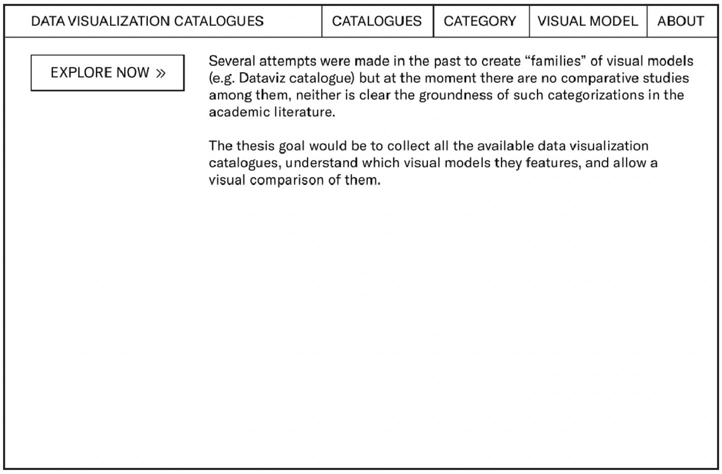
3.3 DATA MULTI-EXPERIENCE

The final visualization contains 13 networks, which show the structures of selected data visualization catalogues, and two groups of circles, representing 314 visual models and 37 categories, respectively. However, more development needs to be done in order to give lay users a chance to access and simply understand the project. A single static visualization is harder to grab the user's attention and is inconvenient for further information. Therefore, it is better to provide users with an available surface to query these visual artifacts and information. Therefore, an interactive website is a suitable medium. It can present our results in a structured and accessible form, making it convenient for non-expert users to explore and disseminate.

To provide an understandable perspective for interpreting the complex and diverse catalogue of data visualization that will appeal even to those who do not know the field of information visualization, showing the process of the project, the website consists of three parts: landing page, interactive visualization, and about page.

→ LANDING PAGE

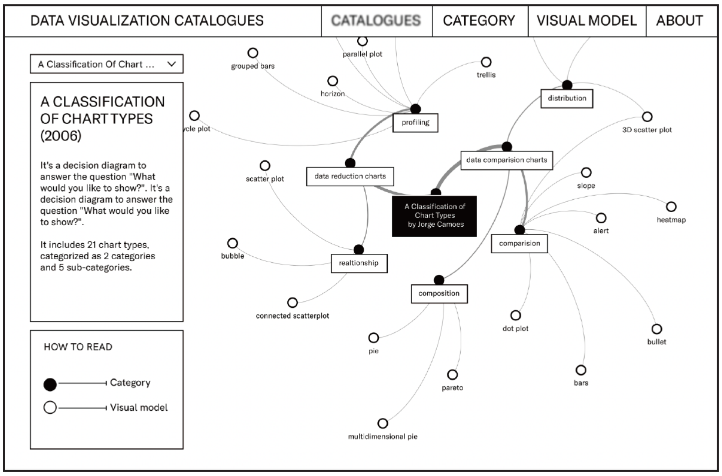
This page contains the current situation of the lack of a unified taxonomy in the field of data visualization and the troubles it brings. It provides background information on the project for non-expert users and provides the rationality of our work. At the same time, this page also involves a brief introduction of the project to give users a general idea. Click on "Visualize it!" and the website will lead to the next section: interactive visualization.



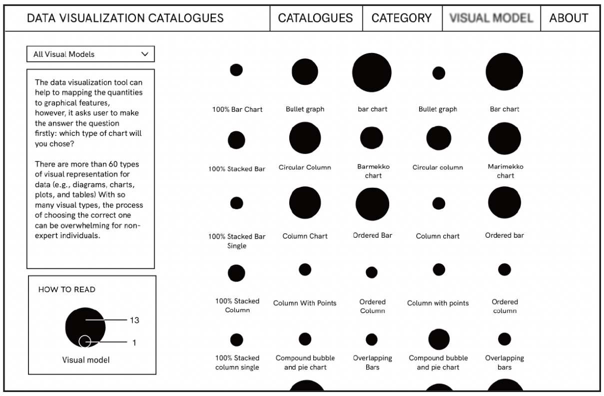
↑ 25 Landing page

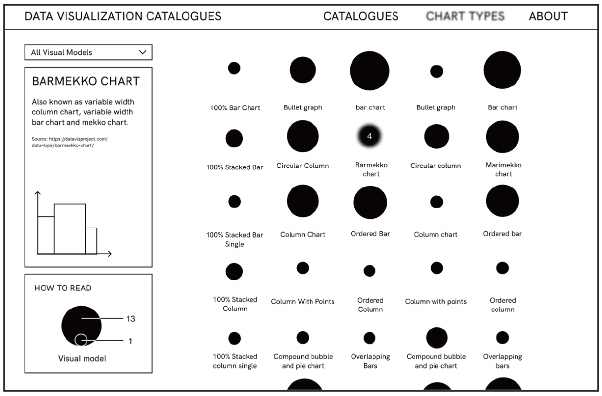
→ INTERACTIVE VISUALIZATIONS

This section consists of three pages: catalogue, visual model, and category. They host all the visualization artifacts we made previously in Chapter 2. They are also the most critical part of the project. This part of the page can be divided into two areas horizontally: the left area is the information bar, which generally contains explanatory pictures and text, such as an introduction and legend. The introduction generally includes the details of the selected item and points to the original Hyperlinks in catalogues. The Legend section tells users how to read these visualizations effectively. Sometimes, this area also provides a drop-down menu that toggles options. The area on the right is the main visualization, which usually contains interactive graphics based on the visualization results above. Users explore the area on the right; further information they request will be displayed in the information bar.



↑ 26 Interactive visualizations - Catalogue
↓ 27 Interactive visualizations - Visual Model





On the page of catalogue structures, users can select the data visualization catalogue they want to explore in the drop-down menu in the upper left corner, the network of the catalogue will be shown on the right. (Figure 15) When users click on black or white nodes with a black stroke, It will jump to the corresponding category or visual model pages, providing users a view of how this model or category looks in the visual model/category matrix.

On the visual model page, when the user selects the specific circle, the column on the left will display detailed information about the visual model it represents, including alternative names of the visual model, an illustration of the model, and the visual model in the selected Raw hyperlinks in catalogues. With this information, users can gain a clearer picture of the potential of a visual model. At the same time, the circle changes from white to black, showing that this model has been selected now. The white numbers on the circles represent the number of times this visual model was mentioned in all catalogues.

The same is true for the category section. Users can also explore the category matrix to get more information about a category they are interested in and view the original source for that category.

→ ABOUT PAGE

The about page contains the background description of this project, the protocol of research., and an excerpt from the dissertation. As a supplement to the project, this part is dominated by text descriptions that explain the project's background lesson intentions, how to collect and analyze data, and 13 data visualization catalogues as raw data sources.

4
CONCLUSION AND DISCUSSION

4 CONCLUSION & DISCUSSION

This thesis provides a visual dimension to look into the current situation of data visualization catalogues. The research focuses on comparing the similarities and differences between existing data visualization catalogues and proposing common standards that may exist between them. Starting from collecting a total of 13 catalogues, through data pre-analysis, data processing, visualization, visual result analysis, and multi-data experience designing, we conducted a comparative study between catalogues through a series of logically related steps.

We choose the structures of the catalogues, categories identified in the catalogues, and visual models included by the catalogues, refining and analyzing them to appropriate data as the basis for the visualization. Next, we obtain quantitative and qualitative findings by observing and analyzing the processed data and visualization results. These results show a lack of unified model inclusion and taxonomy standards in the current data visualization catalogues, which will lead to redundancy and confusion of visual models and unclear classification. This makes it difficult for users, especially non-expert users, to choose techniques suitable for their own database and visualization purposes when practicing visualization.

However, the purpose of this paper is not to criticize existing data visualization catalogues, nor to advise users to distrust them.This paper only shows the gap between users and technology that still exists in catalogues. Due to the increasing importance of data to modern business and education today, more and more lay users without statistical knowledge are involved in visualization practice, which is why we need to find and bridge this gap. In Chapter 3, we redesign artifacts created previously based on the data visualization catalogues into a multi data experience. This web-based experience presents in a simple and clear way the features of data visualization catalogues that lay users may be using or will use in the future, providing a way to compare multiple catalogues. After understanding the similarities, differences, and characteristics of these data visualization catalogues, Users can choose the ones that suit them in future practice with more basis.

As a consolidation thesis, this thesis has certain limitations. For example, we only included 13 common data visualization catalogues, which means the number of samples is relatively limited. If time permits, we can expand the scope of data collection to obtain a more accurate current status of catalogues. In addition, the data sources of this research are mainly from catalogues produced by designers in practice or institutions based on their business experience. However, better data visualization catalogues require more academic support. We did read some papers and took inspiration from them. But they were not included in this comparative study as formal catalogues. Given more time, we hope to expand the data visualization catalogues of academia to provide another perspective.

→ REFERENCE

1 Friendly, Michael. 2008. "A Brief History Of Data Visualization". Handbook Of Data Visualization, 15-56. doi:10.1007/978-3-540-33037-0_2.

2 "Dataart - Learning Resources". 2022. Data-Art. Net. http://data-art.net/resources/history_of_vis.php.

3 "Data Visualization – History And Origins". 2022. https://thinkinsights.net/digital/data-visualization-history/.

4 "Essential Chart Types For Data Visualization". 2022. Chartio. https://chartio.com/learn/charts/essential-chart-types-for-data-visualization/.

5 "Bad Visualisations". 2022. Bad Visualisations. https://badvisualisations.tumblr.com/post/185690817296/this-is-a-screengrab-from-star-sports.

6 Shabdin, Nur Izzati, Suraya Ya'acob, and Nilam Nur Amir Sjarif. 2020. "Relationship Types In Visual Analytics". Proceedings Of The 2020 6Th International Conference On Computer And Technology Applications.doi: 10.1145/3397125.3397127.

7 Shneiderman, B. 2022. "The Eyes Have It: A Task By Data Type Taxonomy For Information Visualizations". Proceedings 1996 IEEE Symposium On Visual Languages. Accessed July 6. doi:10.1109/vl.1996.545307.

8 Qin, Cheng-Zhi & Chenghu, Zhou & Tao, Pei. (2003). Taxonomy of visualization techniques and systems–concerns between users and developers are different. 10.13140/RG.2.1.3159.2403.

9 "The Data Visualisation Catalogue". 2022. Datavizcatalogue.Com. https://datavizcatalogue.com/.

10 "Usage Type | Chartopedia | Anychart". 2022. Anychart. https://www.anychart.com/chartopedia/usage-type/.

11 "Carbon Design System - Chart Types". 2022. Carbondesignsystem.Com. https://carbondesignsystem.com/data-visualization/chart-types/.

12 "Yuri Engelhardt's Twitter". 2022. https://twitter.com/YuriEngelhardt/status/1285885738253537281.

13 "Infoguides: Data Visualization: Visualization Types". 2022. Infoguides.Gmu.Edu. https://infoguides.gmu.edu/data-visualization/types.

14 Abela, Dr. 2022. "Choosing A Good Chart". The Extreme Presentation(Tm) Method. https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html.

15 Healy, Yan. 2022. "From Data To Viz | Find The Graphic You Need". Data-To-Viz.Com. https://www.data-to-viz.com/.

16 "Visual-Vocabulary.Pdf /Financial-Times". 2022. Github. https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf.

17 "Span Chart | Data Viz Project". 2022. Data Viz Project. https://datavizproject.com/data-type/span-chart/.

18 Holtz, Yan. 2022. "Dendrogram | The R Graph Gallery". R-Graph-Gallery.Com. https://r-graph-gallery.com/dendrogram.html.

19 "Network Diagram - Learn About This Chart And Tools To Create It". 2022. Datavizcatalogue.Com. https://datavizcatalogue.com/methods/network_diagram.html.