Politecnico di Torino

Master's Degree in
Cinema and Media Engineering

Master's Degree Thesis

Enhancing Data Narratives.

Prototype of "Accurat: Data Home" case study

by

Sabrina Laneve

Supervisor: Prof. Laura Farinetti
Company Supervisor: Cesare Soldini

December 2021

"Either this is madness or it is Hell." "It is neither," calmly replied the voice of the Sphere, "it is Knowledge; it is Three Dimensions: open your eye once again and try to look steadily."

– Edwin A. Abbott, Flatland: A Romance of Many Dimensions

Abstract

Nowadays we are witnessing an exponential rise in the generation, gathering and exchange of data in a broad range of domains: from commerce, education, health, to culture and society. The display of this huge quantity of information is now, more than ever, of great social relevance: data visualization has a great potential in uncovering substantial behaviors and aspects of data, but also in unraveling novel forms of communication, engagement and interactivity. With the advance of theoretical models and tools for handling information, data visualization has developed original languages, and still enormous prospects are opening, thanks to the contamination between different disciplines.

One of the key challenges of data visualization design is to discover ways of addressing the broadness and multiplicity of interpretations of complex phenomena, providing through new visual means a non simplistic eye on the relational dynamics of data. An important aspect to consider in approaching such tasks is that data is a human artifact, and as such it is not impartial, but it is representative of people's culture and behaviors. In the exploration of innovative visualization practices, the creation of dynamic and multi-channel media installations may offer interesting avenues, providing alternative forms of engagement and informativeness, through the combination of different media forms.

In this work, some exploration of this direction is described through the case study "Accurat: Data Home", a data-driven art installation prototyped in collaboration with the data visualization design firm Accurat. Combining a humanistic concept of data and an experiential approach to the visualization design, the installation addresses the narration of the evolution, life and values of Accurat.

In the first part, the concepts of data and its visualization are presented, alongside with an introduction on emerging trends in the field considered relevant to this work. Following an outline of the main principles and methodologies of data visualization, the design and development process of the project is reported. The described project is therefore presented as a demonstration on the benefits of the integration of multiple channels in telling rich data stories.

Acknoledgments

I would like to express my gratitude to my supervisor, prof. Laura Farinetti, for her kindness and promptness in assisting me over these months.
I am fortunate to have been a part of the Accurat team. I would like to thank all the studio team for the warm welcome and the chance to undertake this project, especially my company supervisor Cesare Soldini for his enthusiasm, encouragement and support.

Contents

List of Figures

1 Introduction

We live in the era of big data. The number of information available every day grows exponentially, so much that new and innovative research fields become necessary to be able to analyze, process and treat this large amount of data.
It is equally important to understand how to present this huge flow of information. Hence the concept of data visualization, which designates the graphical representation of quantitative information, with the aim of understanding and presenting it. It has the task of democratizing and making large or small data containers accessible, by turning them into visual elements that are more easily digested by the human intellect.
In the present, we are witnessing a unique and unprecedented growth of societal relevance of the field. "[...] data are becoming increasingly valued and relied upon, as they come to play an ever more important role in decision-making and knowledge about the world" is what H. Kennedy et al. claim [1] to underline how important the visualization of quantitative information is becoming over time.
Thus we begin to speak of "datafication" [2], meaning how much the collection of data, assisted by the technological innovations of recent years, has begun to pervade society, activating a progressive all-round transformation of the reality that surrounds us. Law, education, transport and commerce are some of the numerous areas that require data collection on a daily basis, in order to study behaviors, actions, sentiments and other many aspects that otherwise would not be identifiable.
Given the heterogeneity and complexity of the systems underlying data visualization, the latter develops its own specific semantic structure and its own form of communication, which remains inextricably tied to the cutting-edge technology on which it relies.

One of the primary challenges of data visualization is to address the depth, complexity and feelings of contemporary reality by offering a non-simplistic eye on the relational dynamics of data through innovative visual tools. To approach this task, it is necessary to understand data as a human abstraction of reality, and as such, flawed, biased and indicative of people's culture and habits [3]. Visualization should try to express the richness and diversity of multiple contextualization of a phenomenon without reduction, while still being clear and accessible. To attain this outcome, it is crucial to draw the user's attention to the exploration of this complexity: hence, it is critical to research new tools and languages that stimulate audience interest and participation in the data story. In today's exhibit scenario, data representation is an increasingly adopted practice, with the use of immersive or combined technologies for data-driven projects [4, 5, 6].

The work discussed in this thesis is set in the context of investigating novel forms of data engagement and communication, through the production of an experimental project in partnership with the high-end data science, design and development studio Accurat, which, co-founded in 2011 by the data artist G. Lupi, is regarded as one of the most interesting information design firms in recent years, also thanks to its unorthodox, "imperfect" approach to data. The project, developed during an internship period in the studio carried out by the author, consists in the prototyping of a data-driven art installation for the Accurat Milan headquarter, as part of the "Data-Home" initiative aimed at reshaping the office into an innovative hybrid workspace which augments the employees' activities and represents a defining aspect of the Accurat firm, as well as a local reference in the information design field. The installation addresses the spirit, identity, and works of the company from a human perspective, narrating the evolution over time of the work team organization and relationships. To further attract the audience to the exploration of such a multifaceted theme, the engaging power of animation and the integration of different outputs is exploited. Thus, this thesis investigates how to enhance engagement in narrating evolving complex relationships, such as those between company employees or between projects and their work teams, proposing the combination of a humanistic approach to data with the creation of a dynamic and multi-channel installation to foster attention and involvement.

In chapter 2, a definition of data and its presentation is provided in the coupled with an introduction to current developments in the field deemed important to this work. Following, chapter 3 and chapter 4 describe basic principles and approaches to data visualization design. The project design process is described in chapter 5 and chapter 6, while the prototype implementation is reported in chapter 7. Finally, conclusions and future developments of this work are discussed in chapter 8.

2 Context

"Data visualization projects can be very aesthetically pleasing and we can appreciate them just for that. But their real power lies in the ability to make people discover corners of life they would never have seen otherwise, analyzing reality one subject at a time. What's scary about this? Data science is no longer a question of numerical analysis, but of stories to be told"

– G. Lupi

The work addressed in this thesis is placed in the context of studying innovative forms of data communication through visualization. As a result, in order to better clarify the practices and aims of this thesis contribution, this chapter provides a definition of data and its visualization, as well as an overview of emerging data visualization applications and trends.

2.1 What is Data Visualization?

To describe the scope of data visualization, it is necessary to define the object of the representation, and hence the meaning of data. In the field of Information Visualization, data is typically defined as the outcome of the creation, collection, or registration of processes or objects that may be classified, abstracted, and translated into a graphical representation. Data can be quantitative or qualitative. Quantitative data refer to features shared by many units in a dataset, that can be counted, measured. Qualitative data are non-numeric information, valued for the uniqueness of each unit. Both these categories can be visualized, albeit with different approaches.

The huge quantity of data available grows so much that the term big data has become ubiquitous: it refers to data that is too vast or complex to process it with traditional methods [7]. The operation of acquiring and saving extensive volumes of information for analytics was already known in the past, but the notion of big data acquired momentum in the early 2000s, when D. Laney associated it with the three Vs: volume, variety, and velocity [8]. Central point in big data is that it reveals

the "desirability of unlocking the information hidden within it" [9]. The process of big data analysis is crucial in breaking down large datasets for various investigation reasons, such as behavior analysis or predictive analysis.
Due to the volume and diversity of data sources, several analytical challenges arise in the classification operations: global analysis of information that does not focus on individual data, but elaborates and considers it as a whole, could be superficial and exposed to bias [10]. Many scholars also pointed out that data are necessarily subject to interpretation, which makes them unneutral [10, 11]. Finally data, especially those coming from the Internet, are inevitably subject to errors or losses. Consequently, different perspectives on the study of data arise.

Noteworthy is the concept of thick data, consisting in the inclusion of qualitative, ethnographic and anthropological methodologies in the data analysis process [12], attenuating the context deficit contingent on the quantitative analysis of big data. As M. Cassidy underlines, "Thick data is simply the idea that numbers alone are not enough. To understand data, you often need to consider things like human emotion, which is rarely data-driven" [13]. Together with the growing interest in thick data, data scientists as well as artists are exploring how visualization can represent socio-cultural, historical phenomena or human emotions. Hence, the idea of "humanizing data" is widening greatly.

2.1.1 Data Humanism

The concept of data humanism, formulated by G. Lupi [3], stems from the observation that data is a human artifact, and as such it is subject to different levels of imperfection and ambiguity. She denies a purely impersonal and technical approach, which looks at data as impartial and inclusive of an absolute truth. Although reliability of data is an important factor in evaluating its use, it must be borne in mind that data are a human representation of reality.
Data are not only valuable in their quantity, but also in their interpretability: the activities of extraction, synthesis, and visualization have great potential in bringing back human qualities in the way we read information. The way in which data are collected and selected transforms what is represented. This subjectivity is seen by Lupi as a tool to envision ways of combining and processing data in order to connect with and better understand the social or cultural context that is being addressed. In this regard, Lupi proposes a process of 'design-driven data', in contrast to the classic data-driven design: data visualization should be a visual result that explores both quantitative and qualitative data sources and combines the informa-

tion into an elaborate narrative. Thus, the visual output should not only convey an accurate representation of reality, but also inspire empathy.

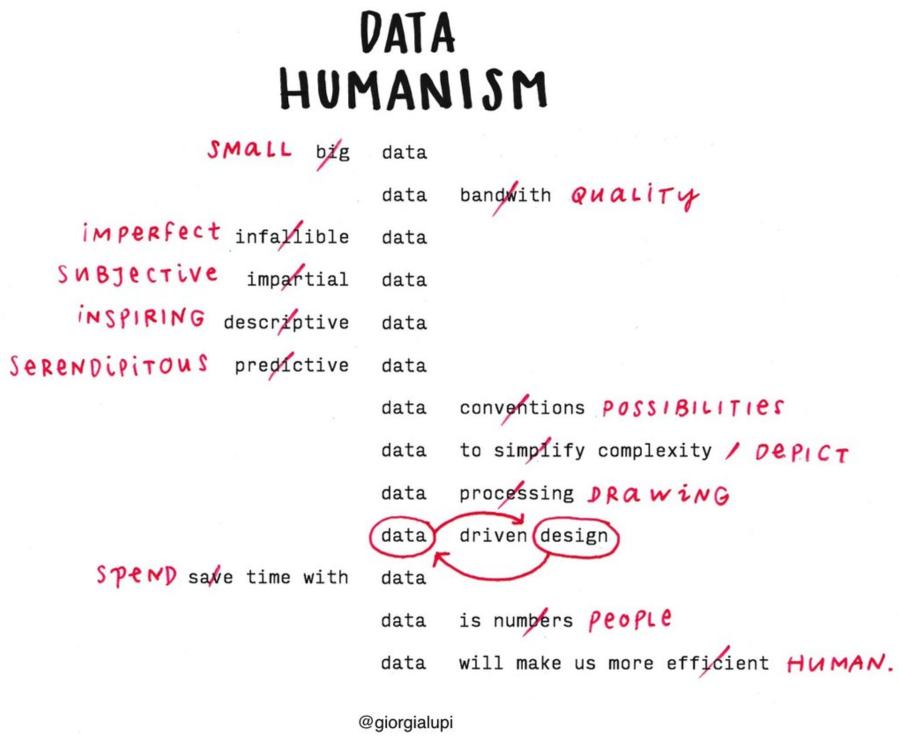


Figure 2.1: Data Humanism Manifesto, Giorgia Lupi

2.1.2 A Definition of Data Visualization

In light of the concepts aforementioned, data visualization (also referred as dataviz) can be more precisely defined as the the graphical representation of data which are mainly, but not exclusively, numeric. The result is a depiction and abstraction of the world, being it the outcome of human decision, cultural and social conventions, as well as high-tech methodologies which consent to collect, filter, and display data. Its aim is to promote knowledge, but also to produce feelings, meanings and engagement [14].

2.2 Trends and emerging fields

As previously mentioned, working with data can aim to different purposes: it can have a narrative, teaching, research and exploration value. Furthermore, it can constitute an alternative means of creating artwork and experiences. In the present, data visualization is pushing its languages and tools beyond standard practices, with several innovative areas emerging from it. In this chapter, emerging fields together with exemplar case studies are presented, showing applications and avenues of the discipline in various contexts.

2.2.1 Data Journalism

Data journalism is a rising discipline, that can be defined as "finding stories in data – stories that are of interest to the public – and presenting these stories in the most appropriate manner for public use and reuse" [15]. The processes of investigation and storytelling typical of journalism are combined with the instruments and tools of data science and information design, which aid journalists to convey in depth complex and rich stories [16]. Data journalism projects can deal with many and heterogeneous fields, but they are generally united by a strong narrative mark and, quite often, by tight production schedules.
Pioneers in the discipline are newspapers such as The New York Times, the Guardian, the Texas Tribune. The New York Times is nowadays one of the main references in the dataviz scenario: some of its historical members were fundamental in advancing the development of tools and techniques for data visualization. Among them, M. Bostock is the author of the JavaScript library D3.js, which is now the global standard for implementing data visualizations in the web and in general (see subsection 7.1.1). He is also the co-founder of Observable, a framework for rapid prototyping and sharing of data-driven designs on the web. One example of innovative elements introduced in the articles of The New York times IS analyzed below. The renowned Yeld Curve, a three-dimensional graph that illustrates the trajectory of the US economy and proposes future projections, was published by the newspaper in 2015. The graph reveals the link between long- and short-term interest rates by displaying how much it costs the federal government to borrow money for a specific period of time. Despite the fact that three-dimensionality is more difficult for the human brain to understand and hence less commonly utilized in data visualization, it works excellently in this situation, illustrating many features of the same phenomena from different perspectives.

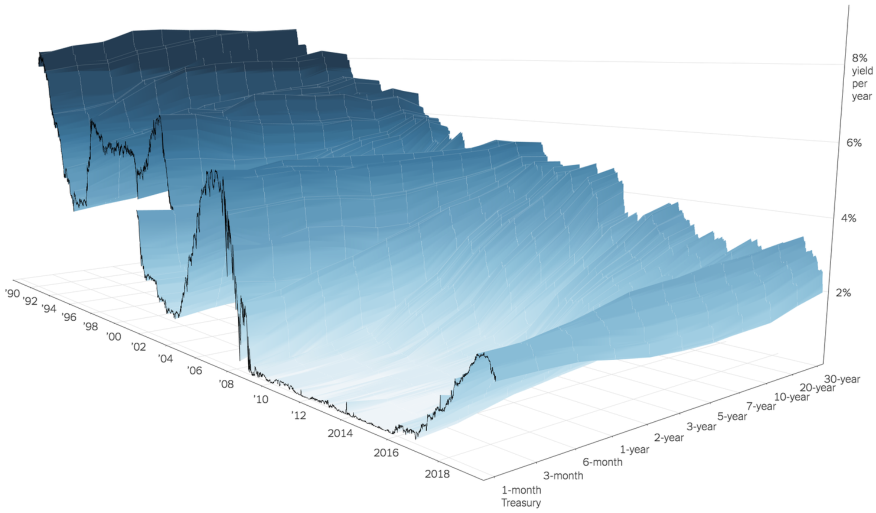


Figure 2.2: 3-D view of the Yeld Curve, The New York Times

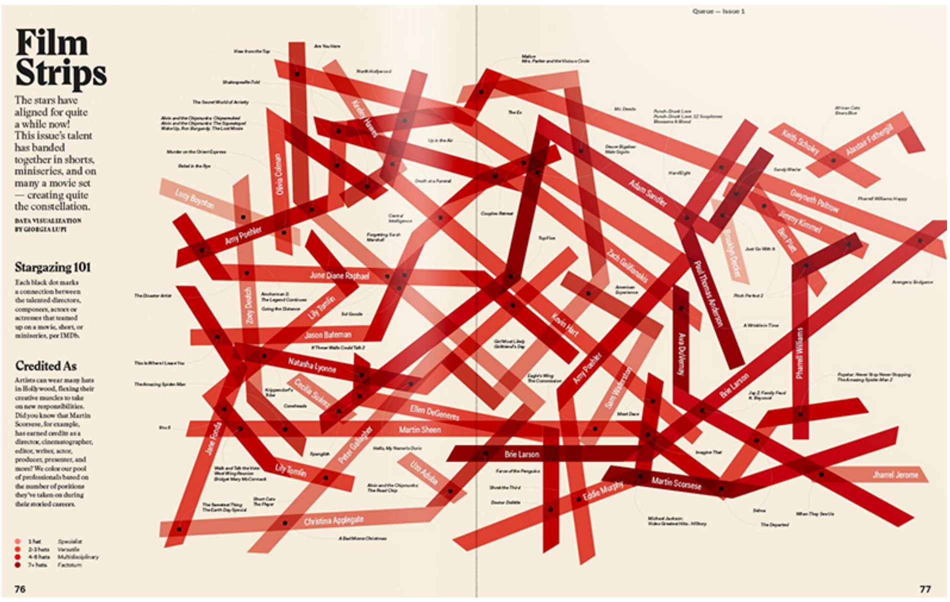


Figure 2.3: Another example of data journalism: Film Strips, Giorgia Lupi for Netflix Queue magazine

From this example, one can also appreciate the social potential of data journalism: those who practice it have a responsibility to provide context, clarity and -especially- find authenticity in the vast amount of information in the digital realm.

2.2.2 Data Art and Data-Driven Exhibitions

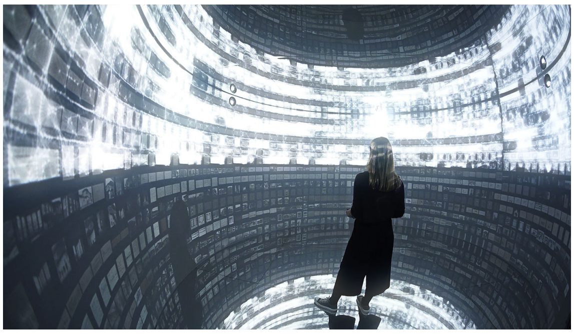
Data art refers to the emerging practice of using data or data-driven processes for the purpose of entertaining or producing an experience that can be evaluated according to aesthetic criteria [17]. The main distinction with the adopted meaning of data visualization is therefore related to the loss of analytical purpose: the goal is not to make accurate and readable graphics, but to communicate emotions and feelings, or to make something fascinating and enjoyable.

Data representation appears to be a practice that is increasingly being pursued in today's exhibit scenario, exploring the application of immersive or combined technologies to data-driven design.
The use of different tools and media in the exhibit data-driven design can be an efficient way to enliven visualizations with small dynamic components that respond to the visitors' interests or the passage of time, resulting in a more engaging system. Innovative and interesting results in the contemporary scene can be observed in the work of the artist Refik Anadol. In 2018 his installation "Melting memories", which explores the theme of the materiality of memory, was exhibited [5]. Visitors are subjected to an electroencephalogram while performing exercises based on the use of memory and the data thus obtained on changes in brain wave activity are rendered in an abstract way: the information is reworked by algorithms designed by the artist to form multidimensional structures displayed through the use of augmented data sculptures and light projections. The installation does not have an informative or didactic purpose, but aims to provide an aesthetic interpretation of neural mechanisms of cognitive control used to store memories and to arouse emotions in the viewers.
The exhibit "Archive Dreaming" offers the visitor with a depiction of the material from Istanbul's historical SALT archive, in which the visitor may filter the links of varying kinds between the various documents using a tablet. In the absence of input, the work "dreams," displaying interactions selected by the artificial intelligence and projecting beams of light and color within a dark cylindrical environment [4].

Other data artworks do not limit themselves only at provoking an emotional impact through the creative representation of data, but also offer clear keys to interpretation and acquire an informative value. This is the case of some projects by Accurat studio, such as Building Hopes and The Room of Change.

The first is an immersive data-art experience that leverages Google's ARCore technologies and that invites people to give their hopes an element of "physicality" by creating permanent augmented reality sculptures made of balancing rocks that represent what they are hopeful for. The project, born in 2018, is accessible both via web application and smartphone: the user selects their hopes and assigns them a weight that determines the size of the rocks that make up their sculpture. The sculptures take on a double meaning: on the one hand they express the emotionality of the individual, on the other hand they create a network of tangible data revealing how people around the world are searching for the same concepts, because each single sculpture considered in relation to the others incorporates in itself Google Trends' data that are expressed through size, rotation, direction and speed of the rocks [6].

(a) Building Hopes, Accurat



(b) Archive Dreaming, Refik Anadol

On the occasion of the 2019 Milan Triennale themed "Broken Nature: Design Takes on Human Survival", Accurat studio creates the installation "The Room of Change", a 30-meters-long hand-crafted data tapestry that horizontally shows the evolution over time of 8 macro subjects (nature, universe, animal kingdom, society, hope, happiness, science, technology), all of which are relevant to humans and related to our environment, while vertically shows a snapshot of a precise moment. Each topic (i.e. animal kingdom) is portrayed using both global data sets to frame large-scale events (e.i. reducing word population of bees, growing population of cattle and chicken) and specific tales that depict the micro implications of the large-scale phenomena, either directly or indirectly (e.i. extinction of animal species in South America). The artwork is accompanied with a legend in the center of the room that explains how to interpret the separate stories and allows people to see our changing world from a unique vantage point [18].

Figure 2.5: Melting Memories, Refik Anadol

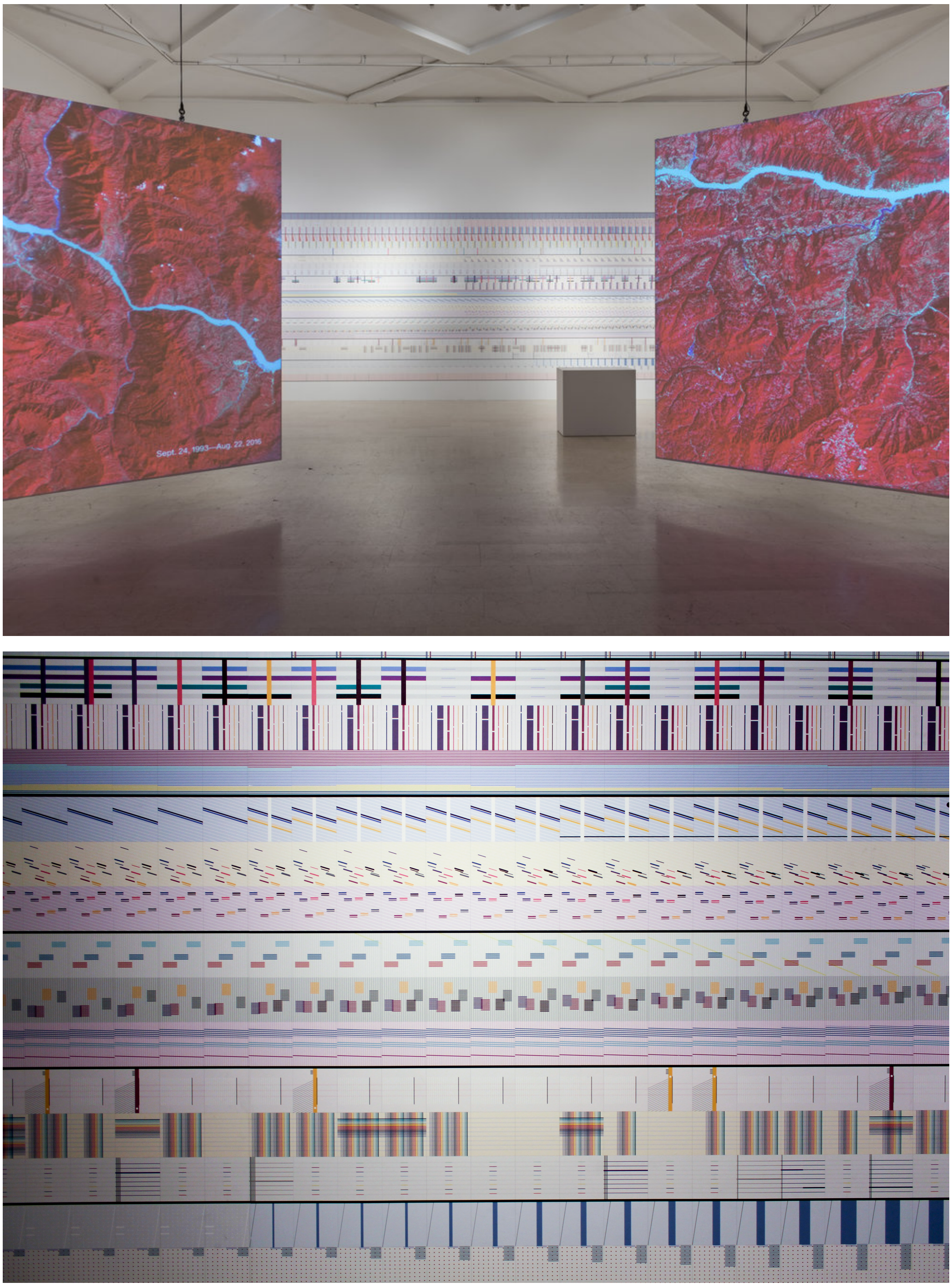


Figure 2.6: The Room of Change, Accurat

3 Data Visualization Methodology

A data visualization design process can be conventionally subdivided in three main stages: data collection, exploratory analysis and representation. This procedure is not to be considered as always sequential as some of its stages may involve multiple iterations [19]. However, the presented simplification is helpful to frame the practices adopted in this work.

3.1 Data collection

The starting point of a data-driven creation is the selection of the data to be used and the sources from which this data is derived. During this stage, one or more data-sets are identified, constructed or combined, in order to be subsequently analyzed. The goal is not only to find data, but to obtain it in a form that can be properly interpreted and manipulated through analysis tools.
Databases can be produced from third parties and obtained through an examination and selection procedure, or can be self-produced, following a process of observation. These two approaches may also be combined. It is undoubtedly important to have a clear understanding of the main goal and questions of the research from the first stages of the visualization process, as the chosen approach often depends on the nature and goals of the project. For example, projects that fall into the category of data jorunalism commonly use databases provided by third parties. This choice allows the data collection phase to be limited in terms of timing, and, with a proper analysis, to bring out interpretations that are independent from the ones provided by official sources [16]. Other types of projects, as in the case of data humanism, consider data collection as the central point of the process, making it an iterative step and part of the final visualization.

3.2 Exploratory Analysis

The data exploration stage has the aim to highlight patterns among the collected data, which will serve as the basis for the narrative structure of the visualization. This phase can be further divided into three steps: data humanism, data mining and exploratory visualization.
The operation of data cleaning consists in identifying and fixing errors in the data-set that may affect the final visualization. This step is particularly important when using data made available by third parties. H. Mueller & J. Freytag [20] subdivide the main types of errors that can be found in a data-set into syntactic errors and semantic errors.
The next step is that of data mining: this practice can be defined as a process aimed at identifying interesting structures within the data [21], or - depending on the interpretation - identifying knowledge within unstructured data [22]. This process can consist in a broad range of techniques and forms of application, but in the context of this work it is intended as the act of manipulating the organizational structure of data in order to find insights and obtain a exploratory visualization. The use of exploratory visualizations can indeed be considered a data mining technique [21], but in the case of designing a data visualization it is so predominant over other approaches as to deserve a distinction.
Exploratory visualization refers to the practice of visualizing data for the purpose of understanding its relationships rather than communicating it. This procedure involves the repeated representation of a data-set in different forms, so that data can be observed according to different criteria. For example, the same data-set or part of it can be visualized from a temporal perspective using a line chart, or according to a quantitative criterion by means of a bar chart: adopting different models allows to prioritize different aspects and to define which ones lead to the most interesting results. Usually, the tools that allow the production of exploratory visualizations (i.e. Tableau, RAWGraphs, R, Python, Observable) make use of data sorted and formatted according to the chosen visualization mode, hence the steps of data mining and exploratory visualization are iterated for a limited number of times.

3.3 Data representation

The creation of the visualization consists in a design process that can have very different formal outcomes, depending on the contents that emerged in the exploratory phase and on the prioritization of some themes over others. According to A. Cairo [23], data-driven visualizations are to be considered technologies, "devices whose goal is to help an audience to complete certain tasks", so the form chosen must be guided by a functional purpose. Thus, the guidelines for designing a visualization are not defined *a priori* but depend on the functional purpose of the visualization. In The New Aesthetics of Data Narrative [24], G. Lupi offers two basic ways that might drive a visualization , labeling her works as "visual-driven" or "data-driven." In the first scenario, the data codification is inspired by a graphical intuition, a visual metaphor for seeing and depicting a certain topic or data. Even if the information is not yet accessible or has not been investigated, this metaphor helps the design process by assuming the data structure and behavior. In other circumstances, the design approach is driven by a narrative: the first focus is on extensive study and analysis of captivating datasets with the goal of uncovering engaging storylines.

This example demonstrates how the three stages indicated in this chapter are not necessarily strict and consecutive. In practice, there is frequently a continual interplay of visual and narrative approaches, with which the various phases are iterated repeatedly, expanding each time: this was also the case in the work examined in this thesis. In the following chapter, general principles of visual representation of data are outlined.

4 Data Visualization Principles

"Graphical elegance is often found in simplicity of design and complexity of data."

– E. Tufte

Visualizations have a strong potential in revealing data, manifesting interesting stories that are hidden within them. For a graphic to be effective, it has to be impactful and carry ideas accurately. E.Tufte in his book "The Visual Display of Quantitative Information" [25] states that "excellence" in statistical graphics is reached when intricate meanings are communicated with clarity, precision and efficiency. He further subdivides this goal in multiple objectives of a visualization, which according to him should: (i) show the data (ii) make the viewer consider the content rather than the methods, design, or technology (iii) avoid distorting the data (iv) display many numbers in a limited space (v) create coherence amongst huge datasets (vi) encourage the eye to compare various pieces of data (vii) reveal data structures at different layers of detail (viii) serve a certain purpose: description, investigation, tabulation, and adornment (ix) be strongly connected with a dataset's statistics and verbal descriptions

Many factors and features have to be considered in order to obtain a functional and powerful visualization. In this chapter the main components and perceptual bases of data visualization are analyzed, together with the introduction of approaches to the visualization of dynamic and complex networks, which is relevant to the project illustrated in this thesis.

4.1 Principles of visual perception

The basis of the components of a visualization is human perception. Indeed, as C. Ware [26] states, data representations are now very important in the study of cognitive systems. Visual representation is the channel with the largest bandwidth of information transmission to the human brain: sight is the sense from which the brain gets the most information. Hence, the design of data visualization has a strong potential in making large quantities of information easily understandable. That is why it is useful to understand basic principles of visual perception, in particular, pre-attentive processing and gestalt principles.

4.1.1 Pre-attentive processing

Pre-attentive processing is a mechanism that occurs before attentional processing and it is carried out by sensory memory, without the activation of a cognitive effort. Studies on perception [27, 28] have identified the properties of a representation that trigger a pre-attentive processing: color, shape, movement, and position. More generally, tasks that can be performed on large representations with many elements in less than 250 milliseconds are considered pre-attentive processing. In contrast, high-level properties of the visual information processing system require focused attention on the item to be "consciously" interpreted [29]. High-level processing can be induced by either overly dense and chaotic representations or by the use of arbitrary symbols: they do not make use of sensory attributes, but are consciously learned or interpreted. It is important to base a visualization on sensory perception: Visualizations based on unconscious properties have a strong expressive power, while more arbitrary elements are powerful according to how well they are learned.

4.1.2 Gestalt principles

Gestalt principles are laws that describe how visual input is simplified and grouped by the human brain into unitary forms. Even if these rules refer to vision, many correspondences can be found in auditory and somatosensory perception. The Gestalt principles were formulated by M. Wertheimer; and were then advanced by W. Köhler, K. Koffka, and W. Metzger [30]. There is no definitive and exhaustive list of Gestalt laws: such principles are primarily employed to describe rules of organization of considerably complicated visual fields. Some of the most often discussed are mentioned and described here. These classifications are robust and unequivocal in certain circumstances, while they are best defined as tendencies in others, especially when diverse elements interact with each other.

- Proximity: Objects and shapes close to each other are grouped together
- Similarity: In the presence of several different elements, there is a tendency to group similar ones

- Continuity: oriented units or groups tend to be perceived as belonging to continuous wholes if they are aligned with each other.
- Closure: elements tend to be grouped if they are parts of a closed figure.
- Common fate: if elements are moving or pointing in a direction, those with a coherent displacement are grouped.
- Good Gestalt: elements tend to cluster if they are part of a good Gestalt pattern, that is the as simple, ordered, cohesive, regular structure as possible.
- Past experience: elements tend to be grouped if the observer has seen them together frequently in the past.
- Figure-ground: objects are perceived as figures in contrast to a background. Despite figure-ground perception being an important aspect of field organization, it is not commonly referred to as a Gestalt rule.

4.2 Visual Variables

J. Bertin, a French cartographer and graphic designer, breaks down and classifies the spectra of adoptable representations for information in his 1967 book "Semiologie Graphique", [31] providing designers with a foundational methodology for finding ways to represent data in a systematic manner. The book is recognized as a fundamental theoretic work for cartography and information visualization. Bertin focuses on static representation, explicitly leaving out cinematography, in which movement and the temporal dimension would alter the rules discussed. He identifies a set of visual variables, graphical objects that constitute the building blocks of every visualization, organizing them in groups: size, value, shape, color, orientation, texture, which are organized in space, a variable transversal to all of them. From this classification, he analyzes the use of visual variables in function of the communication of data.
J. Bertin's work has been developed and integrated by a number of researchers [23, 25, 32, 33, 34, 35], and it is still being worked on nowadays. A very recent overview of the main static visual variables is the one of R.E. Roth [36], which collects the seven variables (splitting position into a pair of x,y variables) introduced by J. Bertin and integrates them with the contributions of other authors: color saturation and arrangement,introduced by Morrison [37]; crispness, resolution and transparency, introduced by A. M. MacEachren [38]. Roth also introduces a hierarchy and categorization of the variables, guiding the designer to their appropriate usage.

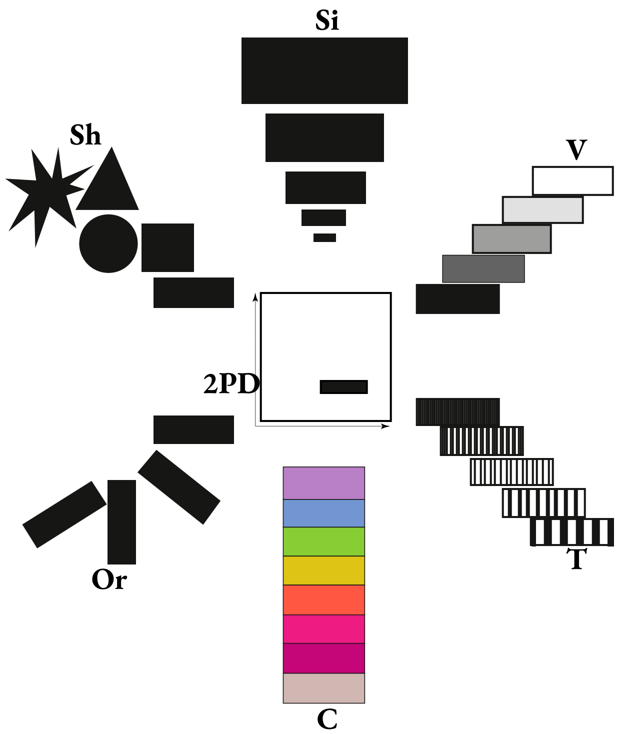


Figure 4.1: Jacques Bertin's Visual Variables

The taxonomy of static visual variables is divided into two types of classifications: by "levels of organization," postulated by Bertin himself, and by the kind of information that variables can encode. The first classification identifies four levels:

(i) associative perception: shape, orientation, color hue, texture. Color size and intensity are instead related to dissociative perception.

(ii) selective perception. All variables except shape fall into this category.

(iii) sorting perception: position, size, color intensity, texture. MacEachren later added also color saturation, crispness and transparency, and argued that texture is only slightly suitable for sorting.

(iv) quantitative perception: position and dimension.

The second classification, proposed by Roth, resumes the division established by MacEachren:

- Unordered visual variables: appropriate for encoding nominal information, that is unordered qualitative data: color hue, orientation, and shape.
- Ordered but non-quantitative visual variables: appropriate for encoding ordinal information: color intensity, color
- Quantitative visual variables: position and size. These variables are appropriate for encoding numeric information but can also be used for ordinal and nominal information, due to their visual dominance.

The variables collected by E. Roth are briefly described below.

- Position: defines the symbol's placement in reference to a coordinate system. The location is regarded as the fundamental visual variable, with precedence over all others. Elements positioned in the optical center of the image tend to appear in the foreground, whilst objects positioned towards the page's perimeter tend to fall into the background.
- Shape: describes the object's outward geometry, or trace. Form is extremely suited in the design of qualitative representations. Shapes might range from highly abstract (circles, squares, or triangles) to very distinctive (directly imitating the thing being symbolized). Symbols that are more complicated or less compact tend to emerge in the forefront.

- Orientation: defines the direction or rotation of the sign with relation to a baseline condition. Neighboring clusters of symbols with the same orientation tend to appear as a single group in the front. Otherwise, visual components that are not aligned to the baseline tend to show in the background.
- Color Hue: defines the symbol's primary wavelength in the visible section of the electromagnetic spectrum. Red items are more likely to emerge in the foreground, whereas blue items are more likely to fade into the background.
- Color Intensity: defines the amount of energy that the sign emits or reflects. Color intensity modification is essential in representations of ordinal or numerical data. The color intensity figure-background connection is linked to the quantity of light and dark regions present on the display: dark items generally appear in the forefront on primarily light backgrounds, whereas light things appear in the foreground on largely dark backgrounds.
- Texture: represents the roughness of the fill pattern of the object. Figures with richer textures are more likely to emerge in the forefront.
- Color saturation: the extent to which the sign's color seems washed out over the visible spectrum of light. Colors that are brilliant or intense transmit or reflect light in a narrow band of the spectral region and appear in the front, whereas pale or desaturated colors transmit or reflect energy equally over the visible spectrum and appear in the background.
- Layout: specifies the arrangement of the visual symbols that form the representation's object. The layout's visual variability ranges from regular (graphic signs exactly aligned in a grid system) to irregular (graphic signs ordered freely or connected in clusters). Signs having an uneven layout, especially if grouped, are more likely to show in the front.
- Focus: defines the sharpness of the sign's edges. It's also known as "fuzziness," and it's a great way to portray uncertainty. Sharp-edged items tend to emerge in the foreground, whereas fuzzy-edged objects tend to blend into the background.
- Resolution: defines the spatial accuracy with which the sign is presented. The resolution of a raster representation relates to the coarseness of the grid size, while the resolution of a vector representation refers to the level of detail in terms of vertices and edges. Items with a high degree of precision are more likely to emerge in the forefront.

- Transparency: defines the level of visual blending between the sign and the underlying ground or objects. Signs that are opaque or non-transparent likely emerge in the front.

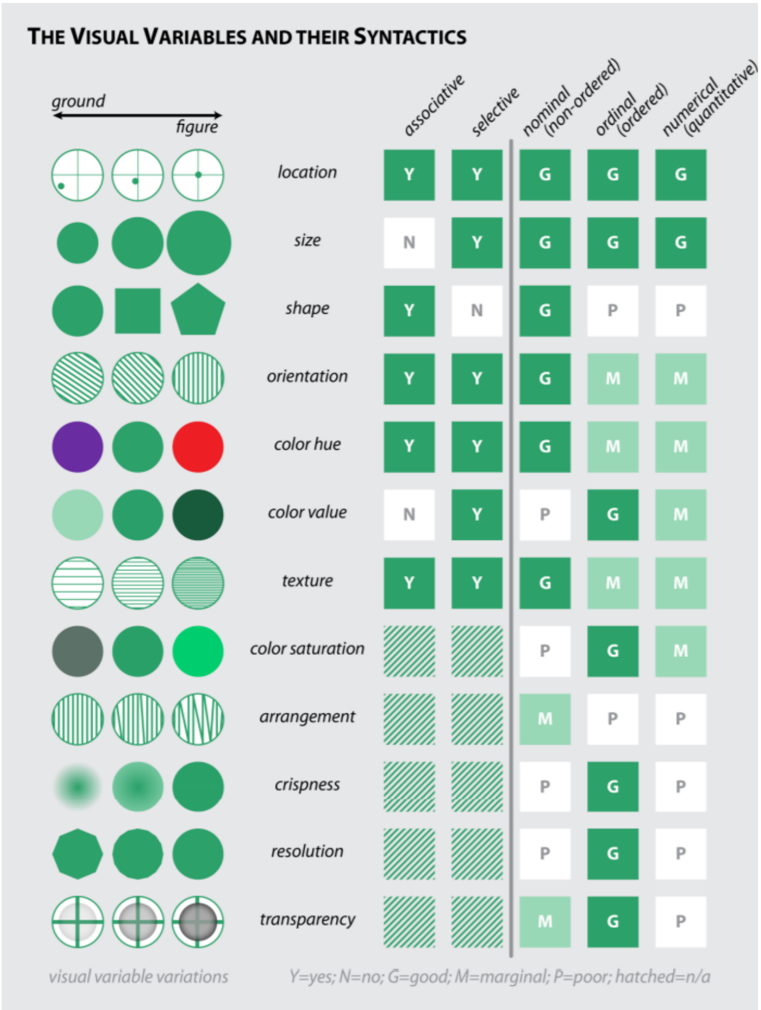


Figure 4.2: Roth's Visual Variables Synthesis

4.3 Chart Types

Over the years, many attempts have been made to categorize graphs and visual patterns, among which it is relevant to mention the table created by A.V. Abela [39]. It is one of the best known classifications even though, compared to more modern classifications, it is more convoluted and less rigid. According to this division, 4 macro-categories can be identified based on chart function: Relation, Comparison, Distribution, Composition. Within these 4 functions there are sub-functions: this division leads to similar graphs in different sections. Standard contemporary classifications [40, 41, 42] group charts according to their goals, aesthetics or visual features. Categories inherent chart functions adopt a more linear systematization than the traditional one previously cited, avoiding sub-functions. The following categories can be identified:

Comparisons Comparisons Charts allow to highlight the differences between two or more elements, the areas where they are similar or where they differ. It enables to understand which one among two or more values is higher or which one is lower. It is based on the idea that people are easily able to compare between different shapes; if these two shapes are put side by side, the comparison should be easier. Their purpose is best exploited for time-based data such as units sold per day or for categorized data, for instance sold units by team.

Trends over time Trend charts are mainly applied when working with data using time dimension, as they keep track of alterations over a definite amount of time. They show the evolution and development of set values through direction and can be affected by the cultural environment they found themselves in.

Part to whole The part to whole charts aim at showing the different subcategories of a value. They are applied for categorized data, for instance the subdivision of revenue by product as they work with both percentages and absolute values.

Correlations Correlation charts convey a particular kind of relationship. However, these are usually considered a separate category, as this distinction is extremely important in influencing the correct meaning to convey and, consequently, what chart, among many, is more effective. Correlation occurs when there is relationship between two or more indicators and the correlation charts aims at making it easier to show combined behaviors.

Relationships and connections These charts aim to specify the hierarchical role of an element in a system or to examine the nature of a subject throughout various states of a process. This chart can be used for both categorized data and multidimensional data. An example of the first can be the analysis of the country of origin of asylum seekers and their gender while an example of the latter is number of active users by testing phase.

Maps Maps are the best means through which geolocated information is conveyed. Thanks to the location of the elements, they easily allow to identify places or to grasp the geographical context. These charts are applied for geographical data, for instance voters by county or average wage by neighborhood.

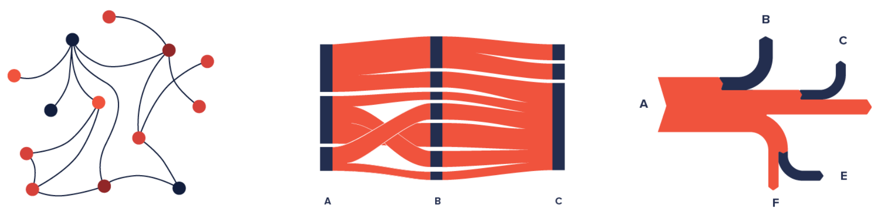
4.4 Visualizing Relationships

This section delves into graphs and approaches for visualizing relationships, which is an essential element of the project described in this thesis.

Network graph Network visualization finds often use to describe the multiple interactions we live daily. It is based on two units called nodes and lines. Nodes represents the various entities and the connection between said entities are established through the aforementioned lines, thus creating different spaces within the graph depending on how many interactions an entity develops. This disposition is carried out thanks to algorithms that organize said space. According to this, nodes at the center are more relevant and have more relationships than the ones on the outer sides. The disposition of the nodes may also underline unexpected data discrepancies in an immediate way.

Alluvial diagram The alluvial diagram is a type of flow chart which is used to represent alterations in a network system over a stretched period of time, therefore it points out trends. Its name derives from the alluvial fans which originate from the soil settled from a stream of water, thus the term relates to both the how the diagrams appear and their flow. The categorical or qualitative variables the diagram uses are located on parallel vertical axes in form of blocks with each their own size and stream. The height of the block refers to the size of cluster while the height of the stream refers to the size of the components held in different blocks linked by the stream field. It is possible to put two or more block of indicators next to each other to form larger alluvial diagrams, but this type of graph does not show correlation between indicator that are not directly connected to each other. This is emphasized by using different sets of colors for different sets of blocks.

Sankey Diagram Sankey diagrams are a variety of flow diagram that help visualize transfer of cost, material or energy in various processes, emphasizing major transfers and flow. They can visualize energy on a community level but also energy of an isolated system. The arrows symbolize the magnitude of these transfers therefore the wider the arrow, the bigger the quantity of flow. These arrows can both get together of divide during a process. Usually, color becomes helpful in the process of dividing Sankey diagrams in different groups or highlighting the various transitions during a certain process. Both the Sankey and the alluvial diagrams are flowcharts and utilize flow streams and blocks, but as B. Peterson stresses [43] they also have some differences. The alluvial diagram use values derived from a component or a flow stream thanks to categorical variables. The Sankey, instead, focuses on the flow, on its derivation and respective quantities.



(a) Network Graph (b) Alluvial Diagram (c) Sankey Diagram

Figure 4.3: Relationships standard charts

Expressing and reading complex systems, especially networks, is a demanding practice, particularly since these systems are increasingly employed as depiction of reality. The visualization of networks is a relatively new but rapidly increasing activity that confronts significant hurdles. S. Ortiz [44] points out that networks cannot be properly interpreted if observed exclusively as static systems. Over time, networks' structures alter, their constituents change, information flows, and high-level features and patterns arise. Moreover networks, by definition nonlinear, are interpreted in nonlinear and dynamic ways. There is no gestalt principle for networks: they cannot be grasped at first glance, but can be navigated and analyzed, and visualization should reflect this. S. Ortiz in his article "Living Networks" proposes through some examples helpful methodologies to address dynamic qualities of networks: his main suggestion is that "dynamism should be addressed with dynamism", therefore, extensive and efficient networks visualization necessitates motion, exploration and interaction.

4.5 Animation and Interactivity

Current technologies can capture more visual attention than in the past because they incorporate numerous dimensions: visual, temporal, kinetic, aural, tactile, haptic, interactivity and sometimes immersiveness. Motion, in particular, is extremely successful in capturing attention when it occurs on the edge of the visual field [26]; as a result, its use is optimal to communicate changes on screen. The research data highlight the risks and potentials of using movement in visual communication: on the one hand, it can be distracting by conveying incomplete information; on the other hand, it has the potential to allow not only faster completion of comparison tasks [45] and identification of patterns, groups, and subsets through filtering and grouping functions [46], but also the transmission of emotions and aesthetic impressions [47]. Numerous studies have demonstrated that humans are extremely sensitive to relative kineticism: for example, it is possible to clearly see the outlines of a region among fields of random dots defined simply by differential motion [48]. Although its potential remains largely untapped, motion is therefore regarded in the context of data visualisation as an attribute of a visual object, on par with features such as size, color, and location.

Despite the fact that various theoretical works address the use motion for information visualization, no complete theory of animation visual variables has been established. A. Zotta compiles a synthesis of the various taxonomies [49], proposing a list of good practices in approaching animation applied to data visualization that exploits movement to filter, group, and highlight the elements represented. The following principles are proposed:

(i) Small, brief, and basic movements are clearly visible and sufficient to isolate from a mass components that differ yet have the same kinetic pattern. It is advisable to employ the minimum amount of movement necessary to convey the intended destination for the sake of visual economy.

(ii) To be considered unrelated, kinetic components with equivalent time (frequency, phase, cycle) and routes (direction, shape) must have sufficiently large movement variations. It is recommended that the visual property used to communicate a relevant attribute be as obvious as possible when it comes to element grouping.

(iii) Objects exhibiting constant in-phase motion (flicker and oscillation) can be identified as groups.

(iv) Target items must differ in motion directions by at least a 20° visual angle in order to be distinguished from others.

(v) The data must be encoded at a rate of between 0.5° and 4° per second of viewing angle. To separate kinetic entities, the minimum speed differential between them must be at least 0.43°.

(vi) The flicker frequency must be less than 120 milliseconds.

(vii) Motion form is more effective than amplitude and phase of motion in differentiating or grouping items.

Interactivity is a significant element that encourages data exploration and improves user comprehension, hence reducing the learning curve. When a visualization allows users to choose their own course of investigation, it becomes extremely valuable. Charts include interaction patterns that enable users to control the data shown, allowing users to focus on certain values or aspects of a chart. Below are introduced some standard interaction patterns [41] that can enhance user interpretation of visual information. An important concept is the one of progressive disclosure: the process of gradually revealing chart data gives a clear path to providing details on demand. Examples of direct manipulations that allow users to immediately operate on UI components are zooming and panning, pagination and data controls. Zooming and panning are common chart interactions that influence how closely users may analyze data and navigate the chart UI. Zooming alters if the UI is viewed from a closer or a wider distance. The way of zooming is determined on the device type: for instance by clicking and dragging or scrolling on desktop or by pinching or double tapping on mobile. Panning enables the user to explore content that extends beyond the screen by panning. It should be limited in ways that are appropriate for the data shown. For instance, if one dimension of a chart is more significant than the other, the panning orientation might be limited to that dimension exclusively. Panning is frequently combined with zooming. Moreover toggles, tabs, and drop-down menus can be used to filter or alter data. These controls can also show stats as users alter them.

4.6 Multilayered Storytelling

Lastly, the key premise guiding the visual design approach of this work is Accurat's principle of embracing complexity, depicting all the intricacies and flavors of data while delivering a comprehensible story. To maintain this balance of depth and readability, Accurat's design technique, described by G. Lupi in "The New Aesthetics of Data Narrative" [24] suggests dividing the visualization (and hence the data) into numerous layers of varying character and importance, which will be distributed throughout the visualization's various spatial hierarchies. The steps of this methodology are outlined in Figure 4.4. From the fundamental skeleton to the connections between pieces and the final aesthetic tweaking, the many layers combine to create a finished result that should entice the audience to explore and become lost in the visualization.

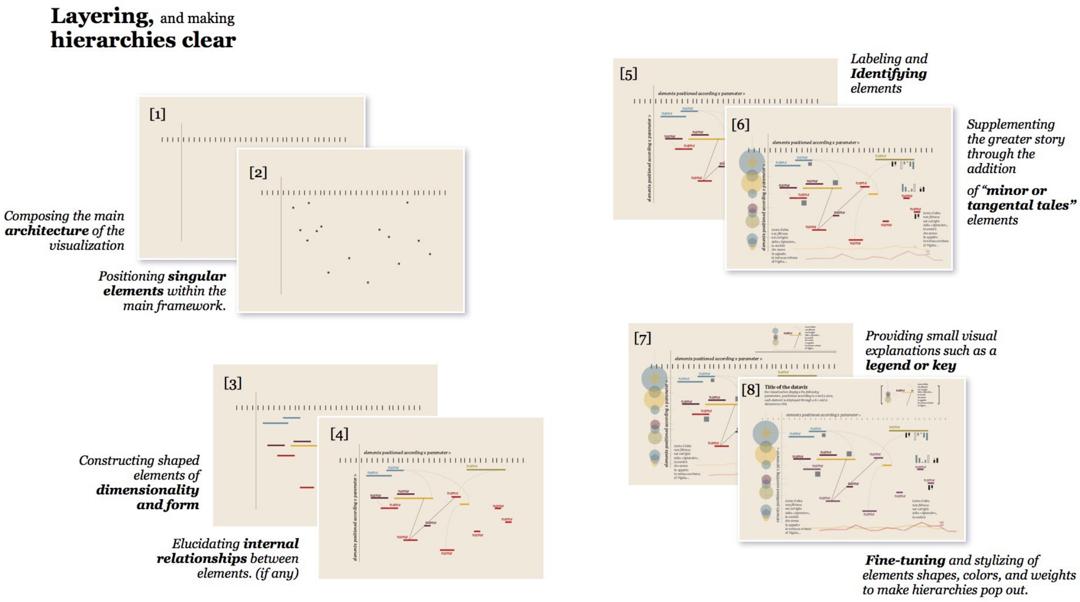


Figure 4.4: G. Lupi design methododology for multilayered storytelling

5 The Story to tell

5.1 The client: Accurat

Accurat is an independent firm focused on data-driven research and design based in Milan and New York City. A diversely experienced team of 40 data scientists, analysts, designers and developers, based on the contamination of different skills and backgrounds, collaborate to ideate and build intelligent, user-centric solutions for global leaders in business, government, and media, through a combination of high-end design with various types of data. Accurat's products range from data-driven art to services for enterprise analytics, embodying the design philosophy of data humanism, which rejects the impersonality of a purely technological approach to data, working to relate statistics to what they truly represent: knowledge, behaviors, and people.

Figure 5.1: Accurat Logo

Since its foundation in 2011 by S. Quadri, G. Lupi and G. Rossi, Accurat has changed its shape from a small studio composed of a few people to a structured organization, capable of collaborating with major international brands. In the process of definition of the near future of the company identity, the workplace is imagined to establish a central role: Accurat will continue to be an organization, but it will also have to be a place.

5.2 Project's Objectives

From this idea comes "Data Home", a project which aims to the definition of a new hybrid workspace that will improve the experience of work teams and that will be an identifying element of Accurat and a reference for the city of Milan. The new space is intended to be a door to external stakeholders, mainly existing for potential new clients, talented designers, developers and data analysts, businesses and people from the neighborhood. In connection with the intent of being a reference point in the information design field, the studio is intended to grow as an innovation hub, hosting exhibitions and encouraging research and experimentation in immersive technologies applied to data visualization.

In this context, the proposal a data-driven installation for the Accurat office, which tells the history, work and identity of the company to the office visitors in an engaging way.

5.3 Themes and Data Selection

During the brainstorming phase carried out with the company for the selection of the themes, one of the first inputs was to collect the data necessary to create and communicate the archive of projects implemented by the studio in its 10 years. However, the main declination chosen for the project convincingly stemmed from the company's own anthropocentric philosophy. If data humanism tries to bring back context to the data represented, a similar approach could be extended to the research of the context behind the projects. Following an analysis of the company's main values and interviews with company members, it emerged that the most authentic way of looking at Accurat was to start from the stories of the people who are part of it. This led to the idea of observing the projects from the perspective of the people that worked on them. In other words, projects are conceived as the meeting of skills and people, a living and evolving entity existing even before its final result, which is not only the outcome presented to each client.

Despite the fact that a direct emphasis on the external outcome of each project would give a clear and immediate idea of the domain in which Accurat operates, a people-centered motif offers some advantages. Firstly, the focus on the human side behind the projects is an empathizing factor that can contribute to the involvement of those approaching the installation, fostering new dialogues and connections between the studio team, but also between the office and external visitors. In addition it can become a motivational factor for those who are or have been part of the studio, giving visibility to the work done by each individual and also show-

ing the path of personal and interpersonal growth undertaken by each individual. Finally, by showing how people have worked together in different times and on various projects, it can serve as an analysis of company procedure, representing the evolution of the anatomy of the teams and internal organization of the company.
Thus, the main information represented is the evolution of each individual's collaboration on the various projects. This data is supplemented with a description of both the persons and the projects. The installation uses a dual perspective to convey the story of the company, emphasizing the journey of the person on the one hand and the progress of the project team on the other. Internal dynamics can be detected across a wide range of time scales. For example, considering a smaller time frame, the course of each project is highlighted: the relative team composition, who took over, when and for how long, and the overall duration of the project. On a larger scale, however, it is possible to see regularities or shuffling in the formation of project teams, how their size varies according to the project itself or with the evolution of the size of the company itself, the diversity of the outputs achieved.
Since its origin, Accurat has been characterized by being composed of people with heterogeneous and hybrid backgrounds. This peculiarity is taken into account in the characterization of the members of the firm: the installation intends to highlight how the characteristics of individuals influence the output of projects, determine the formation of the team and the internal relational dynamics. In this perspective, several elements are considered for the characterization of each person, such as the role in the context of the firm's hierarchy and the specific task, but also the soft skills and personal background. Each member of a team contributes to the project according to their skills and qualities that go beyond the specific role they plays. For this reason, it is important to consider how the various teams change over time or are repeated, taking into account the training path of each individual from before they joined the company up to the moment in question.

The information needed for this installation comes from a number of sources and is related to the projects and who worked on them. The primary data is derived from the timesheet, which is a daily record of the time spent on each project by each worker and therefore expresses the link between the studio members and what they have worked on. Except for the first years of the studio, the timesheet of each year can be acquired mainly from the company's accounting system software. Specific project information, such as work type, output type, and client sector, can be acquired from employment contracts, discussions with team leaders, and the company's GitHub and Figma repositories. Particular attention is paid to the collection of information relating to people: because for this work it is relevant to consider each person's background rather than just their formal role within the firm, the collection should involve not only company data relating to the job performed, but also data obtained through questionnaires, gathering for each person a set of key-words that more fully define their interests and skills. Special care must be given to the fact that this kind of data can have privacy concerns, contain personal information, or be particularly sensitive for the company. This lead to a pre-emptive anonymization of the dataset, making it unlinked from people and clients names. Moreover, the installation is intended to be extended beyond the company's 10 years by drawing on real-time information about new projects and personnel.

In its 10-year history, Accurat has grown in size, and many people have succeeded it. A similar consideration can be made for the projects: an average of 50 projects per year is estimated, resulting in a total of about 500 projects. The use of so many different data and the adoption of different scales for the purpose of data visualization makes necessary a stratification and selection of micro-themes to bring to light the multiple meanings contained in the data. The thematic micro-narratives arise from different focuses such as the evolution of individual skills, interpersonal relationships or the organization of a project team in terms of size, turnover within it and correlation between members and results obtained.

Finally, it is fundamental to clarify that the purpose of the visualization is certainly aesthetic, with the goal of communicating the dynamic, diverse and human spirit of the studio. However, the work is not intended as exclusively ornamental, but as an hybridization of a data art with a more descriptive and analytic visualization, which should therefore keep some readability: in this case, a better understanding of the data underlying the aesthetics is believed to favor engagement with the work.

6 Towards the Visual Model

In this chapter, the main design choices of the project are outlined. The outcome of the steps discussed here is to be considered a proof of concept, and as such it is subject to some constraints. The value of the different steps and tests reported consists in the demonstration of the methodology and of the feasibility of the design solutions adopted to approach the storytelling and visualization of a dynamic and complex system such as a company's work and life. Specific problems outside of the scope of the prototype are not discussed with detail; however, some considerations regarding the general picture and thus the obtainment of a scalable and coherent prototype were taken into account to drive design choices and are briefly addressed.

6.1 Discarded Options

As explained in the previous chapter, the final installation should display data from the company's ten-year history. However, only 1 year was considered in the prototype, exactly from January to December 2020. This limitation saved on data collection and cleaning time, while still allowing for a meaningful study of their behavior and applications to a visual model, potentially replicable on data from other years.

A first relevant consideration concerns the structure of the relationship between projects and people. This aspect was among the main factors taken into account in the evaluation of the coding options considered. Each person generally carries out multiple projects, and contextually each project is carried out by multiple people. Thus, the relationship between these two elements can be thought of as a bipartite network. A bipartite graph, also called a bi-graph, is a set of graph vertices divided into two disjoint sets, that is, they have no element in common, such that no two graph vertices within the same set are adjacent (i.e. connected by an edge). Therefore, a way to represent such system could be the disposition of the two types of nodes (projects and people) on two opposite sides of the viewport, connecting them according to the projects collaborations happened in the visualized time range. Considered statically, this model does not allow to observe a temporal evolution, that in this case would be represented exploiting dynamism, with disappearance and appearance of nodes and connections according to the active people and projects.

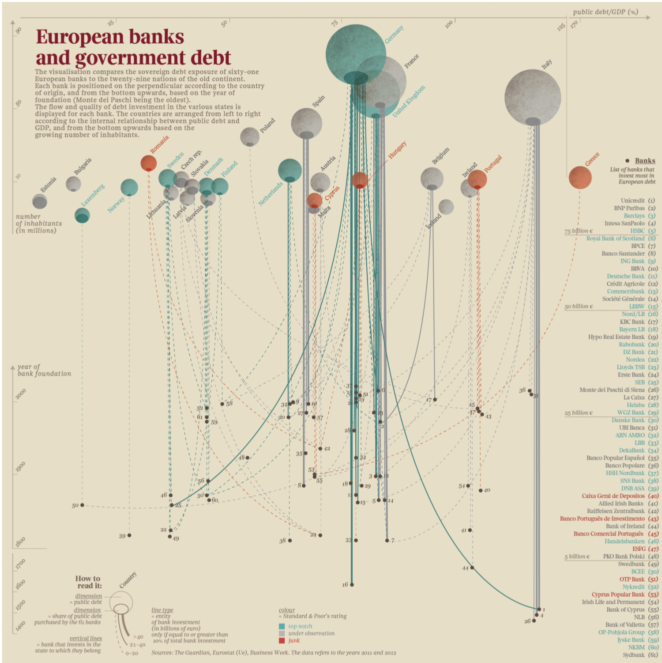


Figure 6.1: Example of bipartited graph visualization: European banks and government debt, Accurat

An interesting aspect of this model is that people and projects can be rearranged on the axis in which they are placed, in order to highlight significant patterns. For example, people with multiple projects in common and vice versa can be arranged according to a system of forces that attract them, thus showing affinities present within the two categories of nodes. From this arrangement, however, it is not immediate to observe projects as a set of people, as the number of connections between nodes tends to be very large.
A very similar approach involves creating a force-driven graph with two types of nodes, corresponding to projects and people. A project node constitutes an attractor of person nodes, resulting in a convergence of connections, one for each collaboration. Since an individual performs several projects simultaneously, the corresponding person-node will be located at an intermediate position between the active collaborations. The person-nodes closest to a project are therefore not nec-

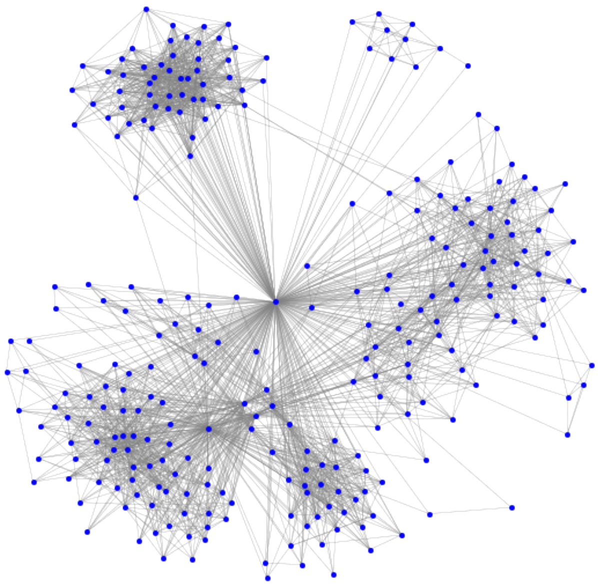


Figure 6.2: Example of force-driven network graph

essarily those inherent in the collaborations of that project, but the ones resulting from the overall sum of forces that could interfere with each other. The complexity of the data in question could lead to a confusing representation.
Although there are positioning algorithms able to mitigate similar phenomena, in this case it would be necessary to manage at the same time the position of the nodes-persons with respect to each node-project, the relative position of the projects and the relative position of the people. Even operating on the reciprocal positions of project-nodes or limiting the length of the connections, there is the risk of obtaining a system in which the positions of the nodes-persons are not significant and the reader's perception of the graph would be very dependant on the initial positions of the graph, which are completely random. The time evolution management would happen in a similar way to that described in the previous hypothesis.
Another alternative examined, although still employing the more classic force directed network graph, was to depict only one type of node rather than a bipartite network. For example, representing all individuals as nodes, the connections are given by the projects in common. The closest nodes then correspond to the people who have spent the most time collaborating together. The same model can be re-

versed, considering the projects as nodes.
This visualization is interesting because it clearly highlights insights into the data, such as significant relationships between people or projects. This is achieved at the price of separating information about projects and people into two separate views. To maintain the duplicity of perspective enunciated at the beginning of this section, a different approach from the classical network representation was sought. In fact, spatial representations of the temporal dimension were devised, in particular by arranging the projects along a temporal exis. A different visualization effort depicts the projects by placing lines on the horizontal axis according to their completion dates. People are represented as semi-transparent colorful regions that span and surround these lines, overlapping in the case of concurrent cooperation. Other experiments involve arranging persons and projects in a grid-like pattern: projects are placed in a horizontal order, as vertical lines of varying lengths. Each cooperation is represented by colorful symbols on the project line, and each individual is represented by a line linking their many collaborations. Even if the previously described visual models have various flaws in terms of readability or coherence of visual coding, some of the features utilized in them have been used as a starting point for the building of the final visual model, which is detailed in the next section.

6.2 Main Architecture

The chosen visual model maps time on the horizontal dimension. Each horizontal line represents a person's collaboration on a project. Thus, the arrangement on the horizontal axis of each line indicates the time interval in which the collaboration occurred.
Collaboration-lines are aggregated by projects. That is, all collaboration-lines to the same project are stacked vertically, so each project appears as an aggregation of a series of collaboration-lines.
At the same time, each collaboration-line has connections: all lines connected to each other represent collaborations of the same individual to different projects. Therefore, a person is represented by a set of interconnected lines that, read from left to right, represent the temporal succession of projects that person has worked on. Connection-lines always start from the last project started by the person to the one immediately following. Therefore, if a collaboration begins before the previous one is finished, the corresponding connection-line will be vertical.
The arrangement of the projects - that is, the aggregations of collaboration-lines - takes place on several rows in the case of simultaneity, hence the visualization expands with respect to the vertical axis according to the number of simultaneous

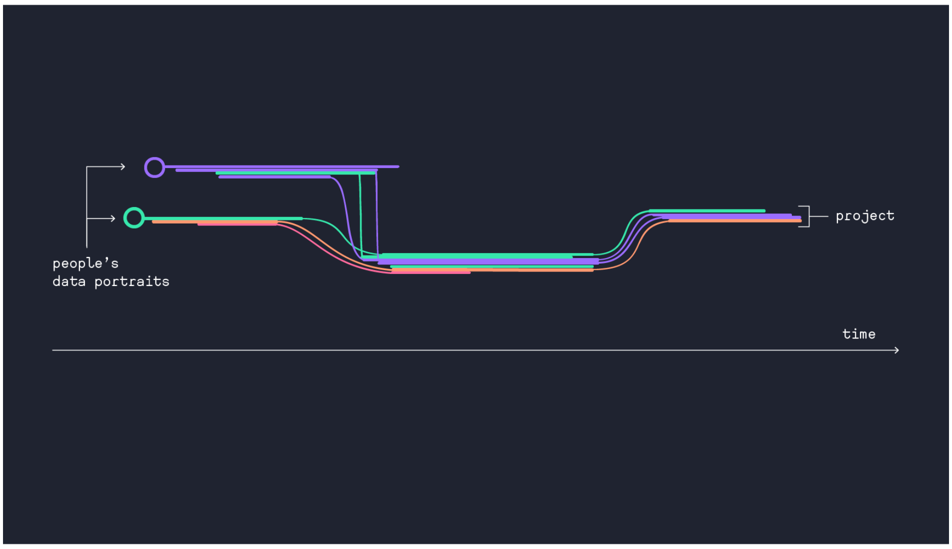


Figure 6.3: Simplified design of the visual model

projects observed in the window considered.
This model makes it possible to look at the data according to the twofold perspective previously mentioned: people are represented as a flow of collaborations(a set of collaboration-lines and connection-lines), which represents the steps of the individual's journey in the company. The projects in turn result from the combination of different moments in the history of each person. As the horizonal axis represents time, applying the visual model to the 10 years data produces a long linear horizontal sequence of groups of projects. As the time dimension is potentially always expanding, it is reasonable to imagine that only a selection of the total time range is displayed, employing the digital nature of the installation to enable the exploration of the horizontal axis via a panning movement (see subsection 6.3.1). To compare data belonging to more extended time, it may be necessary to decrease the horizontal resolution (i.e. representing weeks and not days of work). The model designed is compatible with this eventuality, as explained in subsection 6.2.2) Superimposed on this general structure is the level of data representation that characterizes people and projects. In the developed prototype, people are colored according to a simplification of their main role, grouping everyone in 4 main categories: Development, Dataviz Design, Experience Design, Strategy/Management. However, this categorization can be considered as a placeholder for a more complex representation that has been drafted and is introduced in the following section.

6.2.1 Data Portraits

For the visualization of people, the idea was to go beyond the label of the role, showing more qualitative data in a "personalized" way, so that people may still recognize themseleves and be represented in a unique way, but without exposing private data. The chosen approach is a practice frequently adopted by Accurat, to the point that it can be considered almost emblematic of the studio's methodology: data portraits. Data portraits consist in generally tiny artworks based on the individual replies to qualitative surveys, translating answers into a unique combination of shapes, colors or other visual variables. This singular kind of soft data representation is an great technique to encourage connections and engagement.

Figure 6.4: Examples of data portraits created by Accurat for the Target space at TED 2017 Vancouver

Considering that the prototype focuses primarily on the overall architecture of the visualization, the visual arrangement of the signs forming the data portrait has not been explored in depth. However, the integration of the data portrait in the general architecture has been defined.
Each data portrait (set of shapes) is placed at the start of the first collaboration-line belonging to the corresponding person, i.e. it is placed with respect to the horizontal axis at the point representing the first day on which that person began his or her

journey in the company. These shapes are combined with the whole person's line aesthetics to form the complete portrait. The person line (set of collaboration-lines and connection-lines belonging to one person) should include part of this portrait through the use of texture, transparency and color gradients, showing growth of the person over time.
For project descriptions, a similar approach is used. Each block containing the collaboration-lines corresponding to a project team is marked by a set of symbols representing the project team's main characteristics such as type of client and output.
This addition of characterizing elements is also graphically useful in visually grouping the lines into the representation's main blocks: people and projects. Since data portraits are inherently compact elements, they provide an immediate synthesis of these two building blocks. Moreover, the highlighting of the beginning of a person's line through their portrait identifies for the user the beginning of a possible path of exploration of the visualization, that is, the one from that same person's perspective. It also encourages the comparison of the paths of people who began at similar times and may therefore share different types of links. Similarly, a project, represented by a series of collaboration-lines representing its team members, becomes much more recognizable and distinct as a result of a set of signs that frame it.

6.2.2 Exploratory Visualizations

To test on the actual data the validity of the various levels of the hypothesized architecture, several statistics and exploratory visualizations were operated.

First, simple histograms were used to observe trends over time in the number of projects carried out simultaneously. These statistics were necessary to understand if the height resulting from applying the data to the thought visual model was manageable. Indeed, the number of concurrent projects in the time range considered determines the number of rows on which they can be arranged. In 2020, the number of simultaneous projects per day is in average around 19, growing in the last months up to a maximum of 26, and decreasing significantly in the very last and first weeks of the year, when - as expected based on the company's organizational rules - most of the projects have already reached a closure or have not yet started. A number of 26 rows is acceptable for the visualization, and it could be estimated that this number does not grow significantly in other years, given that over the ten-year period the size of the firm has tremendously grown. This behavior was observed at different time scales: concurrent projects for each day, week, or month were visualized. This allowed to observe whether the visual model could be used at different temporal resolutions. Visualizing this dataset at different levels of temporal detail is quite salient, as it may bring out different patterns observable only for each time scale. For the prototype, a daily resolution was chosen, but the statistics showed that a weekly resolution would keep a similar behavior in terms of chart height (in 2020, the weekly maximum is still 26), and even a monthly resolution would not be excessively problematic to handle, with a maximum of 30 simultaneous projects. Issues relative to more compact time resolutions could be addressed with a simplification or removal of the details showed in the representation, combined with a filtering of the project displayed. Anyhow, these options will not be further investigated for the sake of the prototype.

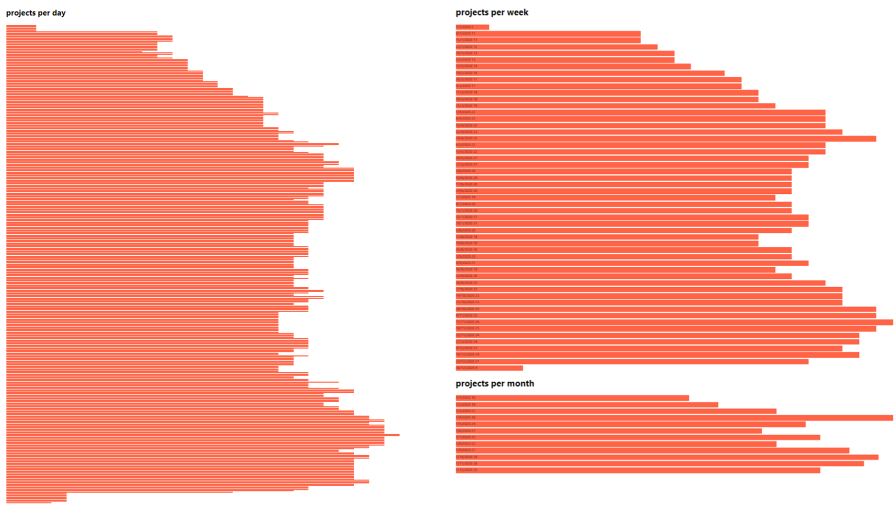


Figure 6.5: Histograms showing the number of simultaneous projects over time

A second visualization was created to examine the duration and sequentiality of the projects over time. In this visualization, the horizontal dimension indicates time, and each project is represented on one row as a horizontal line of specific width and height. The positions of the left and right endpoints of each project line represent respectively the first and last day someone worked on the project. Each project line only connects beginning and ending extremes, without representing any breaks during the work on the project. This simplification is sufficient to have an approximate idea of the duration of the various projects. The height of each line represents how many people the project team is made of, without any representation of individual temporalities. This simplified representation of height was also useful to consider the vertical space occupied by each row in the visualization. Project durations result quite heterogeneous: extremely short projects tend to have a contained height, while longest projects have variegated heights.

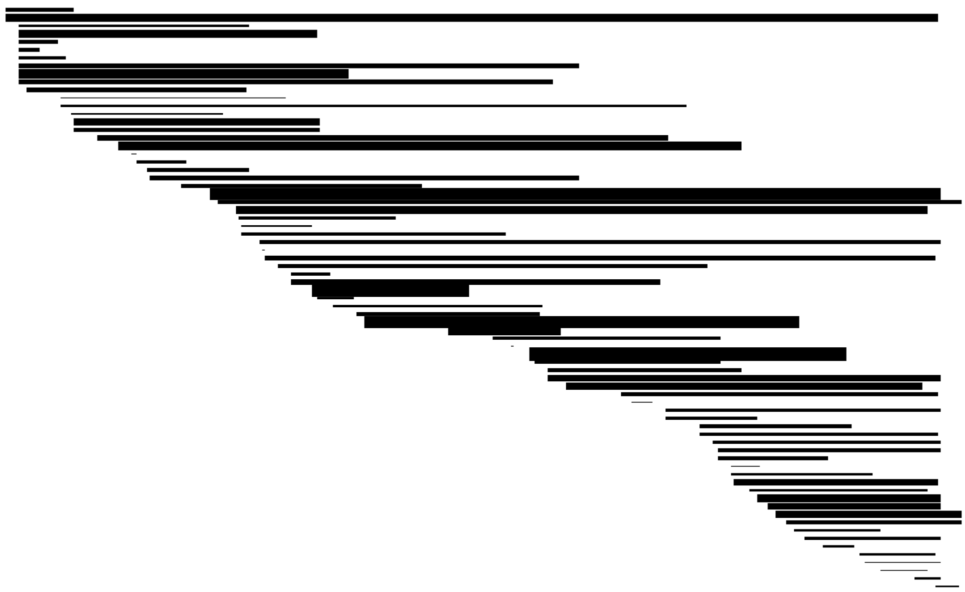


Figure 6.6: Exploratory visualization of projects over time

Further visualizations(Figure A.1 and Figure A.2 in the Appendix) included for each project a distinction between of each person's contribution-lines, marked with colored lines. This allowed to observe the time spent by each person on a project. Contrary to expectations, people's work schedules did not always turn out to be compact with respect to a certain project: people start and finish a project often at very different times. This led to the idea that observing the data on a small, project-sized scale could bring out interesting patterns not only because of the interplay of different skills in a team, but also because of the roles distribution over the course of the project.

6.2.3 Visual Metaphor

Given the constraints on data gathering for the prototype, the identification of visual metaphors that could drive the overall structure of the project was sought during the design process. Accurat's reality is seen as a dynamic, multifaceted, eclectic, and energetic organism in which people's qualities interact to generate projects.

Hence, this concept is visually imagined as a pool in which colored fluids of energy, representing people, are released and flow. Although at an early stage the project was also inspired by the association of relationships to a frame of knots and interlacing, the metaphor that was eventually adopted is inspired by and blends with the aesthetic lines of the colorful office space, as well as the overall layout designed for the installation (see section 6.4). In Figure 6.8 some of aesthetic references for the visual model are presented.

6.3 Handling complexity

Because of the density of the data to include and the amount of connections to portray, the chosen visual design inevitably tends to be quite intricate. Even if this complexity adds a substantial aesthetic character to the message to be delivered, the goal of the work is not merely ornamental but also to lead to user involvement with the data: it is therefore critical to maintain a certain degree of readability. Complexity is mainly generated by the fact that such a large number of projects and people causes several overlapping connection-lines between the collaboration-lines of one person and the ones belonging to other people. Therefore, some strategies have been developed to limit this phenomenon.

6.3.1 Animation

The main idea was to exploit the dynamic component of a digital representation, explicating the layering of visualization through the use of animation. The more lines that appear on the screen, the more intricate the visualization becomes, therefore effective data filtering helps in breaking the visualization down into its parts and showing its different paths. This thought led to the concept of usefully sequencing several types of filtering in order to construct animated narratives of the dataviz itself. For example, using an animated sequence to focus the viewport on a single project, showing that the project consists of collaboration-lines through the sequential appearance of each correspondent line, and finally showing how the project is contextualized in a wider time range through a reverse zoom transition would allow an approaching user to immediately understand the different layers of the visualization. This observation was expanded upon. Another meaning was added to the filtering process in order to contextualize it: the animated sequences became emblematic of the thematic micro-narratives outlined in section 5.3. The filtering sequences not only stratify the visual elements, but also change the perspective from which the data is perceived, i.e. the micro-theme handled, highlight-

Figure 6.7: Example of visual references: Selection of Tyler Hobbs' generative artworks

Figure 6.8: Accurat’s Milan office interiors

ing the most significant patterns for each theme. Filtering sequences can be based on statistics, revealing data insights and behaviors that would be missed in a static representation.
One very simple demonstrative example of meaningful sequence may be: (i) showing the person-line with highest number of collaboration in the visualized time range (ii) adding the person-line that has highest number of collaborations in common with the first one shown (iii) adding all collaboration-lines that are related to the two people's shared projects (iv) filling the visualization with the total set of lines.
The visualization depicting the whole set of lines serves as an arrival and reference point for the animation sequence (and for the user), allowing short loops of different filtering sequences to be started from it. To summarize, animation is used with a twofold narrative level: the description of the visualization's graphic components and the exposition of the micro-themes underlying the overall theme.

As a general approximation, the following types of movement are designed to be employed:

- horizontal panning, to explore the time progression of content that extends beyond the boundaries of the displayed time range
- panning and zoom, to focus the viewport on specific projects
- line drawing, for the appearance/disappearance of lines belonging to a person
- change of opacity, to highlight different kinds of filtered project lines or elements

In terms of animation aesthetics, the movements and appearance/disappearance of the elements displayed should adhere to the visual metaphor used, hence the lines visualized are imagined to be liquid, flowing or floating.

To check the effectiveness of using filtering to improve the readability of the visual model, static exploratory visualizations, with various filtering applied, were created. In Figure **??** examples of tests carried out filtering by a single person or by role are shown. These visualizations demonstrated a significant increase in legibility.

6.3.2 Sorting and Rearrangement

The connection-lines are without a doubt the components that add the most intricacy to the visual model, due to their proclivity to overlap with one another: their length and displacement have a significant impact on the legibility of the visualization. As a result, readability may be improved by utilizing the vertical axis of the

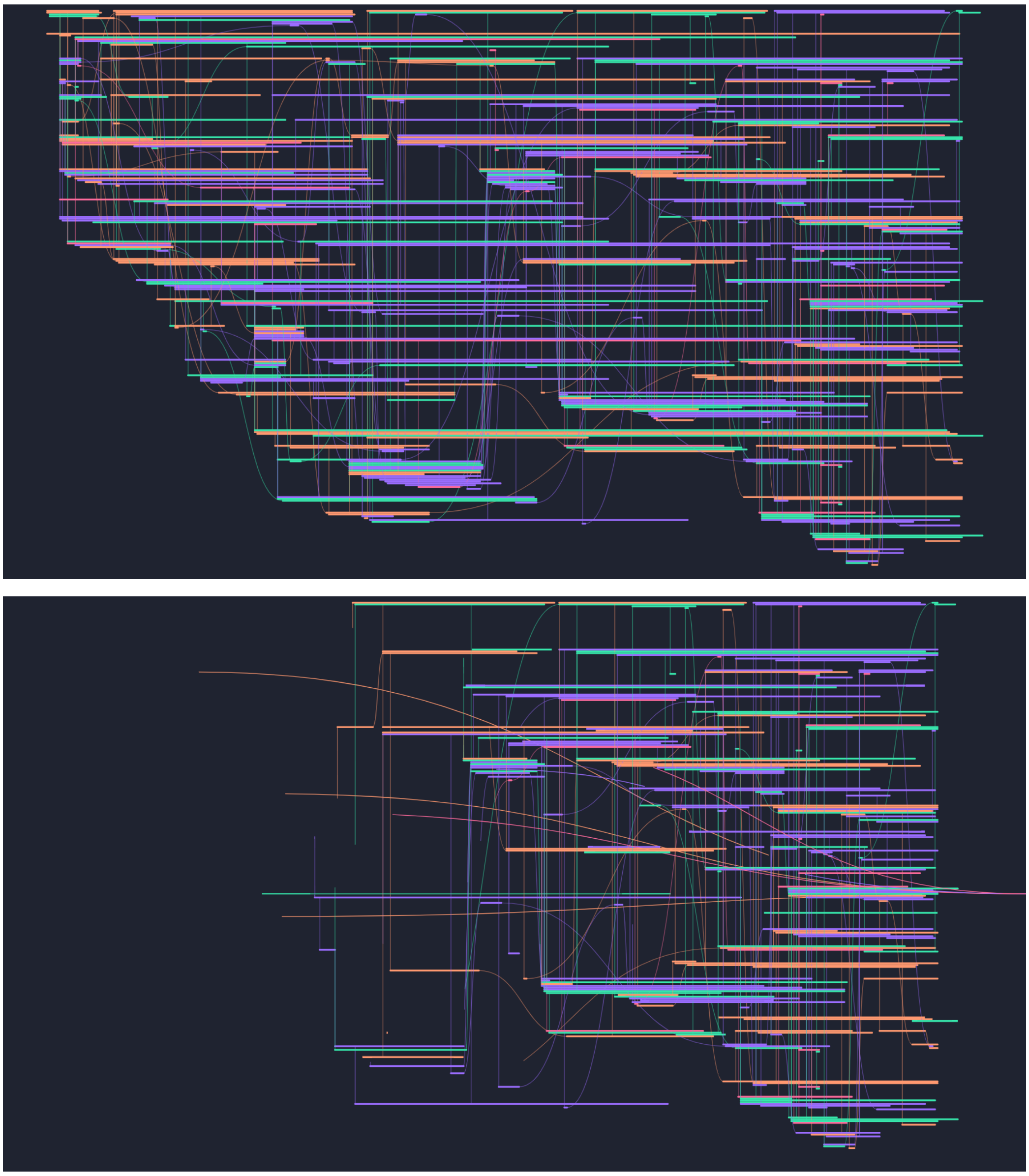


Figure 6.9: Animation frames of the data-driven installation: complete visualization

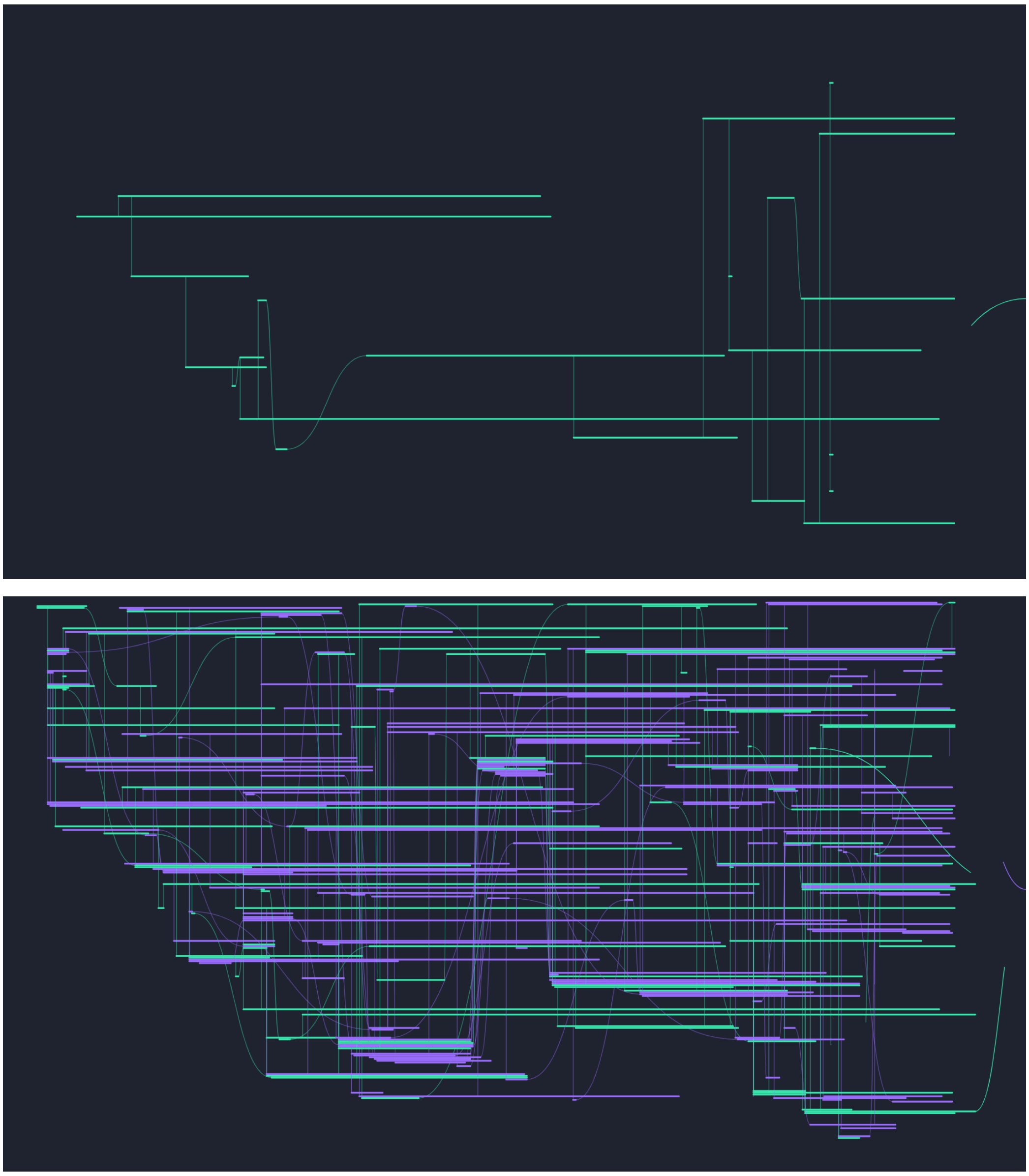


Figure 6.10: Animation frames of the data-driven installation: examples of filtering

visualization and modifying the rows allocated to projects that occur in the same horizontal range. The development of a force-driven algorithm that moves project lines in order to reduce the distance between them might be achieved, but given the richness of the data, it may be very hard to produce acceptable results. However, certain less complex but more targeted algorithms might make improvements. For example, by stacking projects with a bigger team in the middle of the vertical axis, a greater number of connection-lines might be intercepted in the middle of the screen, potentially shortening their vertical length by avoiding connections that fill a large portion of the screen.
Rearranging the collaboration-lines inside each project might also contribute to a more compact and clear picture. Besides the order of each collaboration inside a project, it may be useful to always have lines stacked: for instance, lines that extend horizontally beyond a line above should glide up to fill the gap that would otherwise be empty.

6.4 Installation Prototype

One of the work's goals was to leverage the potential of combining different media to draw the audience's attention to the visualization. An aesthetic consideration was added to this need: in the installation, it was sought to incorporate some elements that characterize the style of the office's interior design, such as the many neon lights and tubes. At an early stage of the design process, the concept of basing the visualization on a combination of tubular neon light arrays has been developed, conveying information through movement, light, and color plays which would immediately drag the user's attention. Figure 6.13, Figure 6.14 and Figure 6.15 show some of the aesthetic references that were considered for the installation and visual model.

The depth of information that needs to be communicated, however, contrasts with the limited number of manipulable variables (primarily color, opacity, and tilt) in systems of this type, which would also provide a much lower resolution than a screen. The combination of these necessities resulted in the creation of a system that integrates and put into dialogue different outputs: screen and neon lights(simulated by LED lights). The installation consists in a wide high resolution screen displaying the visual architecture as described in section 6.2, leaning against the wall. On the left and right sides of the screen, two LED tubes simulating neon lights are positioned, as shown in Figure 6.11 The lights serve to synthesize the filtering events described in the preceding section while also drawing attention to the visualization. Given that a classification - albeit simplified - has been made for peo-

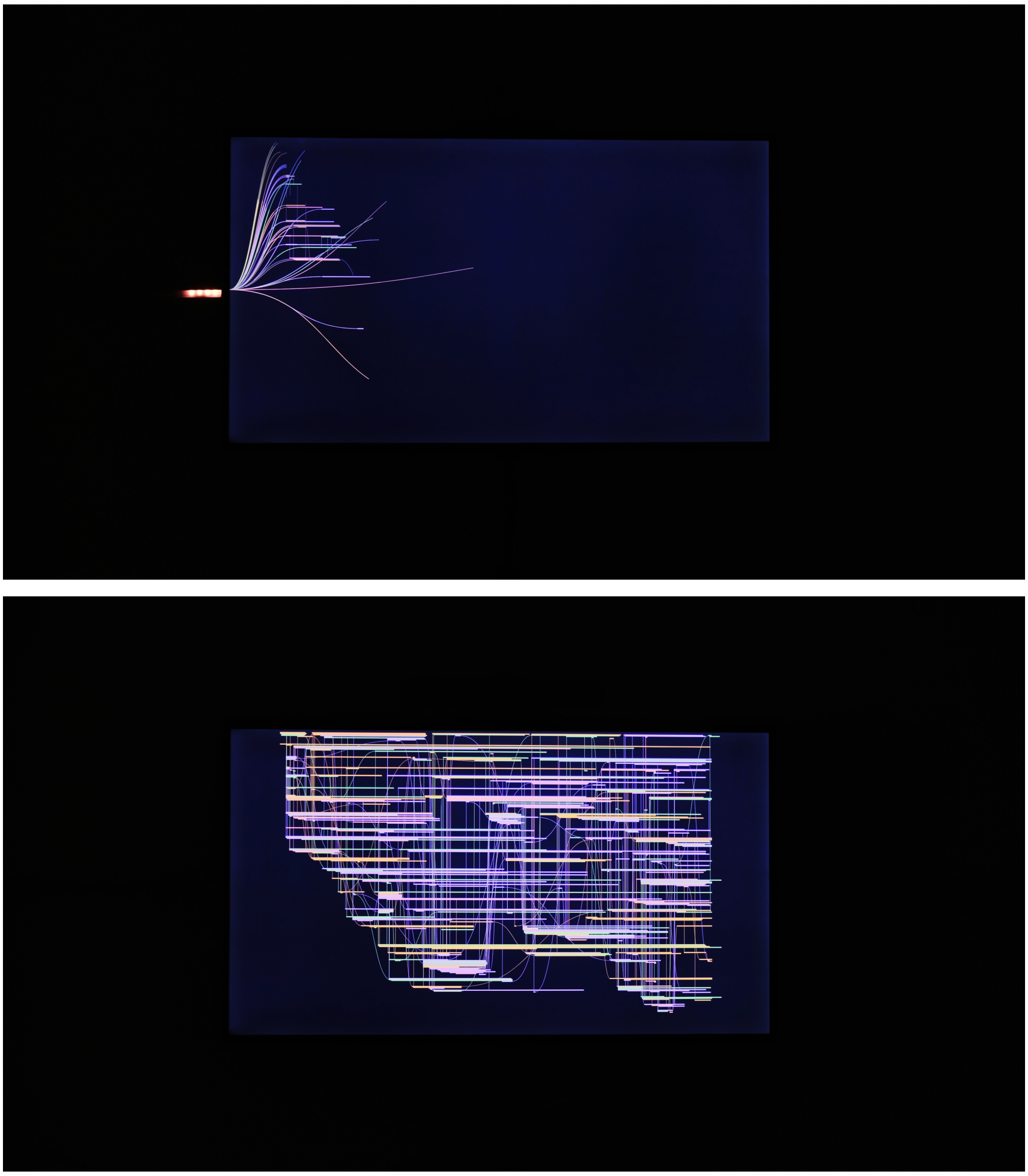


Figure 6.11: LED's animation sequence: entering lines

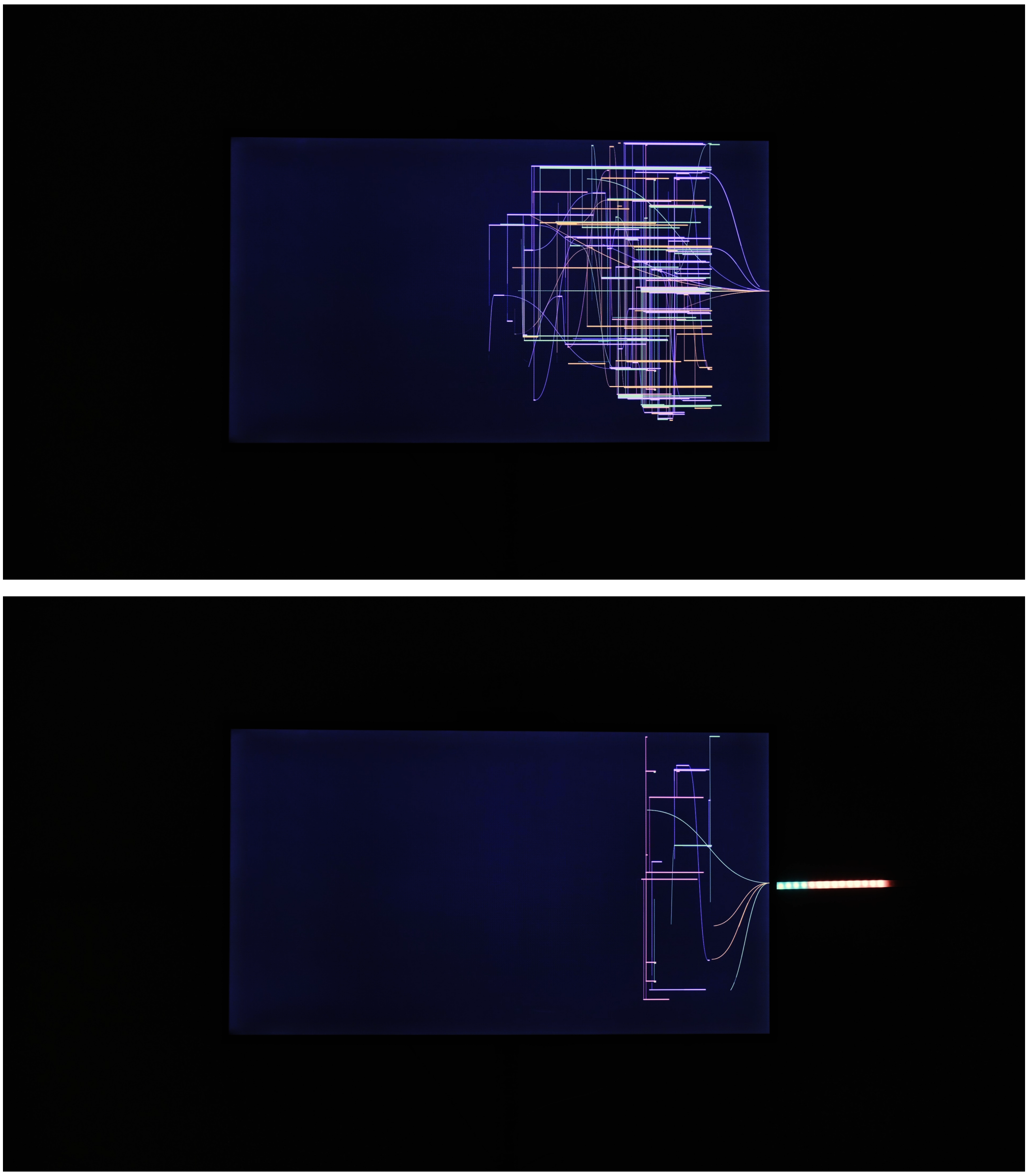


Figure 6.12: LED's animation sequence: outcoming lines

ple, the behavior of lights in the case of filtering by people has been developed for demonstration purposes.
Tracing the style of the screen animation, the entry of the lines into the screen is represented on the LED tubes as a stream of colored light flowing from right to left, on the tube positioned to the left of the screen. The animation of the filtered lines is displayed on the screen when the colored light comes to virtually "touch" the left border of the screen, as if the stream of light entering from the light tube is dispersed across the screen to form the display. The color of the scrolling light shows a preview of the filtered lines: the light is colored according to the category of people filtered and the number of people in each category, giving a general idea of the type of filtering that is taking place. Specular behavior occurs for visualization elements that are not filtered, and thus for outcoming animations. The outcoming lines all convey to the far right of the screen, at the right light tube, where a stream of colored lights flows from left to right based on the number and category of people corresponding to the excluded lines.
As previously mentioned, compared to the screen neon light tubes have a more limited but impactful capability of conveying information: this feature makes the chosen medium well suited to drag instant attention to one of the main aspects of the functioning of the visualization: the filtering sequences. In this way, the eye-catching effect of the LED lights encourages the user to explore the various levels of such a complex visual model.

The artwork was designed with an environmental connotation in mind: to immerse the user, the installation is supposed to be distributed across the workplace area, as an ubiquitous system underpinning the office architecture. As a result, the suggested screen and light model is designed to be reproducible in a variety of screen sizes, orientations, and LED displacements and shapes. These numerous sets of displays and lights might be synced to generate a complementing image or a dialogue between multiple screens. Some ideas include assigning each component of the installation a role relating to the purpose of the area in which it is placed, or using light animation to build coherence across several displays.

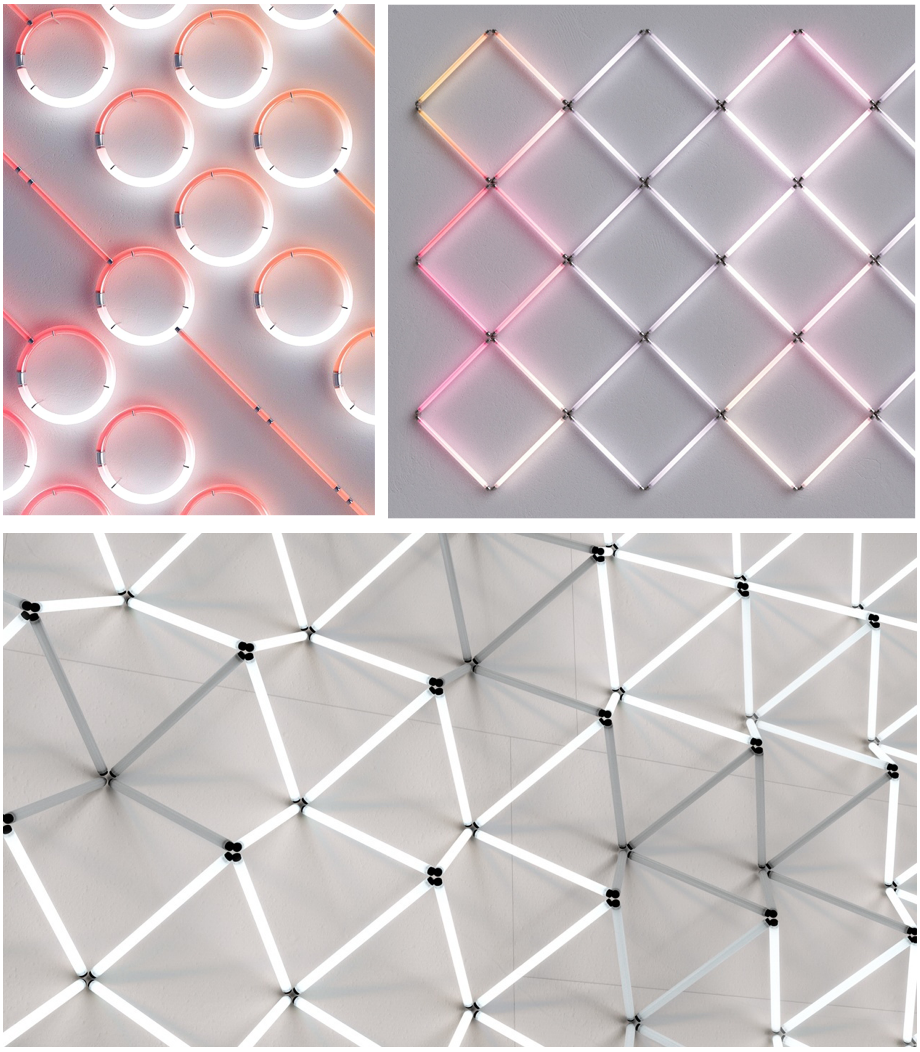


Figure 6.13: Installation References: Samsung Ambient digital artwall, Onformative

Figure 6.14: Installation references: Selection of Dan Flavin's light sculptures

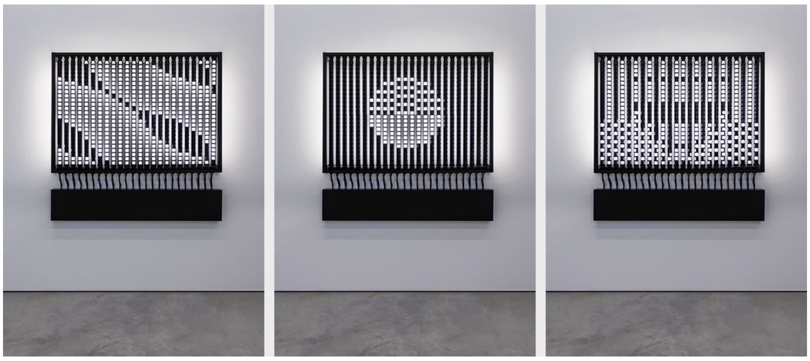


Figure 6.15: Installation references: TrueFalse kinetic light installation, Onformative

7 Installation Development

The goal of the implementation step was to test the viability of the combination of light tubes and screen behavior. As a result, a sequence of filters with different people selections related to the whole year 2020 was animated on a screen and coordinated with two LED strips that simulate neon lights. The LEDs', screen's, and web visualization's aesthetics are only demonstrative. Similarly, the animated sequence implemented does not match to a certain micro-narrative, but rather displays many filtering combinations that may be achieved. The main frameworks and methodology adopted to construct the visualization are explained, followed by a description of the physical environment, architecture, and logic of the system deployed.

7.1 Data Visualization Technologies

7.1.1 D3.js

D3.js (Data-Driven Documents) is a versatile open source JavaScript library for constructing dynamic, interactive data visualizations in web browsers, created in 2011 by M. Bostock as the successor to his Protovis library [50]. D3 is made to leverage on SVG, HTML5 and CSS standards to obtain an extremely flexible system that enables to apply data-driven transformation to the DOM (Document Object Model), selecting and editing its nodes with a sequence of modifier functions. The DOM is a tree-like structure of the data contained in an XML document that is used to graphically portray web pages. It is possible to modify visual elements and sub-elements, as well as their attributes such as color, position, text, and images, by modifying it.

D3 uses a declarative and not an imperative syntax. Declarative programming is a programming paradigm in which the logic of a computation is expressed without the control flow being described. [51] Instead, imperative programming is a programming paradigm that employs statements to modify the state of a program: code is based on declaring variables and changing their values. In other words, declarative programming is concerned with what the program should achieve, whereas imperative programming is concerned with how the program should achieve the

outcome. Among the benefits of declarative programming are improved code optimization and cohesion.
D3 enables the use of a CSS-style selector to choose a specific set of Document Object Model (DOM) nodes, followed by the use of operators to modify them in a way similar to jQuery. Attribute getting/setting, style, and element addition/removal are all examples of operations. Moreover, values for attributes and styles can be seamlessly interpolated over time by defining a transition. These procedures can be linked to data, which is the fundamental idea behind D3. Like jQuery, D3 allows to modify the DOM, but it also has the important feature of data join. A data join is often used to link n data items to n DOM nodes of the same kind, establishing a connection between an array of data and a set of HTML or SVG items. When a dataset is bound to a document, D3 is often used in a pattern consisting of an explicit enter method, an implicit update method, and an explicit exit method. The enter function chain is executed on all items of the dataset that are not already present in the selected DOM nodes. Similarly, all operations corresponding to the update function are called for each existing item of the DOM nodes selection, and exit functions are run for all the selected existing nodes which are not bound to any element of the dataset. This allows to update and animated the graph when the data changes.

7.1.2 React

React is an open source front-end Javascript declarative framework for building user interfaces based on UI components. [52] Its first version was released in 2013, and it is maintained by Meta (previously Facebook)'s web development team. [53] React is often used as a foundation for single-page or mobile apps. However, it only carries out state management and display to the DOM, therefore constructing React apps typically necessitates the usage of extra libraries for routing and also some client-side functions. React is typically described as the V in the MVC (Model View Controller) architecture [54], a popular software design pattern for implementing web applications. A MVC architecture consists of: (i) the Model, which processes data and is linked only to the controller; (ii) the View, which renders the visualization of the Model's data to the user; (iii) the Controller, which functions as a bridge between View and Model, performing server-side logic and regulating data flow.

React is built around the concept of virtual DOM, a lightweight replica of the real DOM. When something in the application's state is modified, the virtual DOM, not the real one, is recreated. In this way, React constructs an in-memory data-structure cache, computes the resultant differences, and then changes the browser's

displayed DOM based solely on those differences, without redrawing all of the screen's components. This selective rendering delivers a significant efficiency increase by eliminating the need to recalculate the CSS style, page layout, and rendering for the full page. This results in a significant simplification of the logic: the developer has the impression of altering the page as if it were being totally rebuilt, whereas React handles recognizing differences and rendering efficiently.
Components are the core of a React application. A component is a self-contained block that accept arbitrary inputs (called "props") and renders a piece of the user interface. React splits the UI into independent, reusable parts that can be processed separately. Components are composable and a component can include other components in its output. These Components can be nested with other components to allow complex applications to be built of simple building blocks.
Moreover, React follows a unidirectional data flow. In general, this idea implies that data has just one route to be transported to other areas of the program. In React, the data kept in the parent component is referred to as a state, and it may be passed to its children components in read-only mode through a prop. The view is the output of the application's state. The user triggers actions in the view, which might update the state. The state update is then sent to the view and child components. This process cycles each time a new action is triggered. Unidirectional data flow prevents data from flowing the opposite way, so children are denied the possibility to modify data from their parent, unless explicitly given. This makes the code more efficient, easier to debug and less prone to errors, as there is greater control over data and the system's elements [55]. Finally, React components are typically written using JSX, although they do not have to be. JSX or Javascript XML is a Javascript syntax extension that enables to write components in Javascript in a way that looks similar to HTML.

When building a React app with complex visualization, it is necessary to rely on a dedicated library such as D3. Combining React and D3 however is not straightforward, as both frameworks operate by modifying the DOM: D3 via the select-enter-exit-update sequence and React via the virtual DOM mechanism. The co-existence of both DOM-modifying operations could be accomplished by keeping a reference to the container element in a component [56] and running D3 code when the component mounts. D3 will then use its selection methods on the referenced object, to operate on the DOM by adding, deleting, or changing components. However, this procedure may lead to imperative, lengthy and less performing code. [57]
Accurat's approach, and hence the one suggested in this study, employs a different technique. Instead of running D3 code on mount, all DOM structure is constructed within the React render method utilizing HTML and SVG elements. D3

is used exclusively to compute all the data-driven drawing instructions, translating data into a format that can be mapped to React components. In other words, code goes over each data point in a dataset, generating DOM objects with rendering characteristics determined using D3 functions. This enables to manage the visualization state using the React component life-cycle rather than D3's update, exit, or enter methods. This solution has certain issues when it comes to producing animated transitions, which in D3 are handled smoothly and easily, but it does have some advantages that make it preferable:

- it follows a declarative paradigm: Instead of describing the steps to draw something, the code explains what is being drawn.
- it is more performant and mantainable. React is fundamentally a rendering framework, with several optimizations to mantain web applications performant. Adding elements to the DOM with D3 causes a "disassembly" of the React rendering cycle, which is something to avoid when wanting to keep React's optimization properties.
- code tends to be clearer and more concise: In complex applications, where developers time is critical, the less code is written the better. Hence, a framework allowing developers to express more logic with less code is a huge advantage.

7.2 Installation Technologies

The following physical components were used to build the prototype:

- HD Monitor
- Raspberry Pi 3
- Individually Addressable LED Strip WS2812b, 60 LEDs per meter
- Power supply 5V, cable connectors, and voltage-conversion electronic components

7.3 Architecture and Logic

The hardware architecture comprises a single computational node, the Raspberry Pi, which communicates with two outputs: the screen and the LED strip, via HDMI

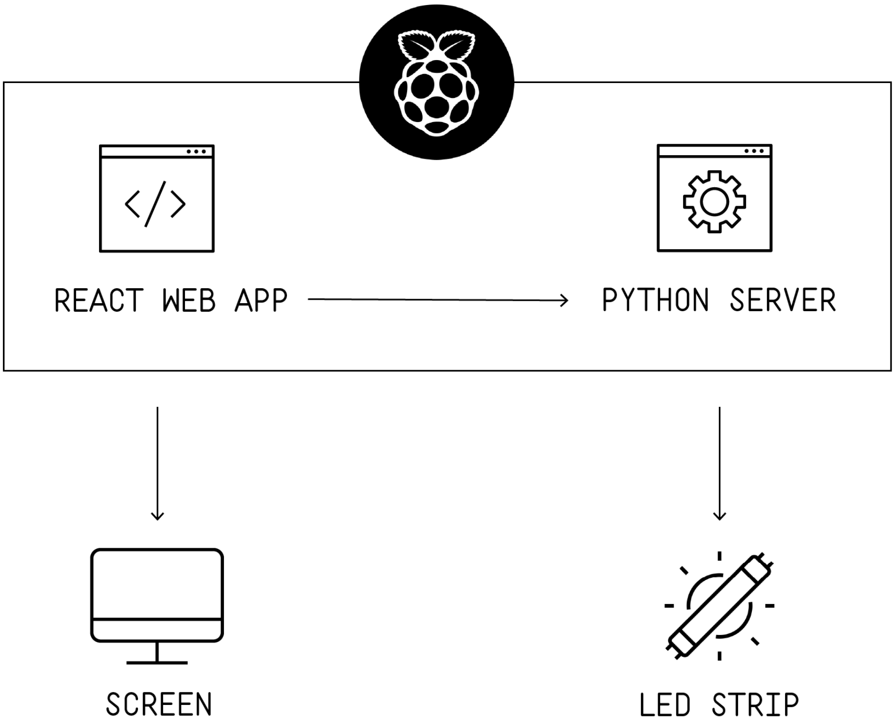


Figure 7.1: Hardware Architecture of the Installation prototype

and JST connectors, respectively. As the Raspberry can control only one strip at a time, the two LED strips are wired together and addressed as a single strip. The Raspberry Pi wiring with the LED strip is illustrated in the scheme of Figure 7.2 This configuration employs a 1N4001 power diode to reduce the 5V power supply voltage enough for the LED strip to read the Pi's GPIO 3.3V output [58].

The Raspberry Pi runs two programs:

- Javascript React App: It handles on-screen rendering, showing the various components and running the animation filtering sequence at regular intervals. For each filtering event, it sends LED animation requests to the server.
- Python Flask Server: Gets animation data relevant to the filtering from the React App and utilizes it to control the LED lighting.

Following are the steps that describe the logic of the system, as schematized in Figure 7.3.

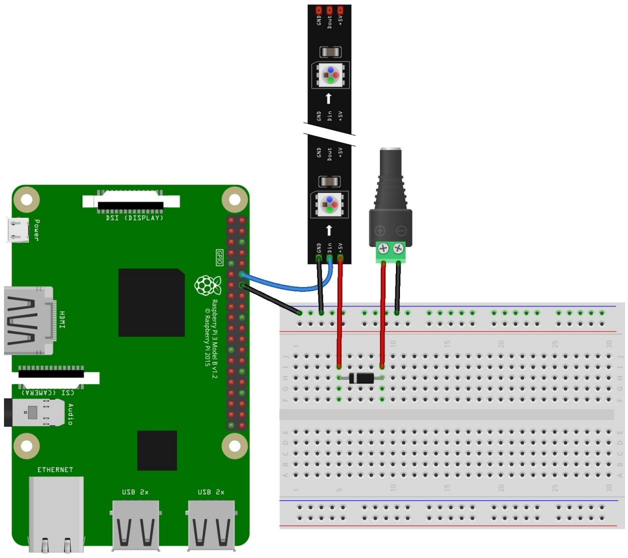


Figure 7.2: LED strip and Raspberry Pi Wiring

Data Formatting The Javascript program accepts two JSON files as input: the timesheet and the assignment of each employee to a role category. As previously stated, the timesheet specifies the projects on which each individual worked as well as the number of hours spent on each day. More exactly, it is an array in which each member is an object describing a person's hours worked on a project in a day, in the form:

```
{
  "TB_TIME_DATE": "2020-01-07T00:00:00.000Z",
  "PE_NAME": "NWE6TTGTCT",
  "JO_JOB_TITLE": "oZZU9TmfKA",
  "TB_TOT_TIME": 6
}
```

The data formatting section is in charge of synthesizing and reprocessing this information into a structure that explains for each project the duration of each person's cooperation. This structure is required to create the information needed to draw the lines. The Lodash utility library [59] was used for the generation and manipulation of complex objects and arrays. As a first approximation for the prototype, just the initial and last day of a person's work on a project are considered, therefore no gaps in a person's work on a project are accounted. To make working

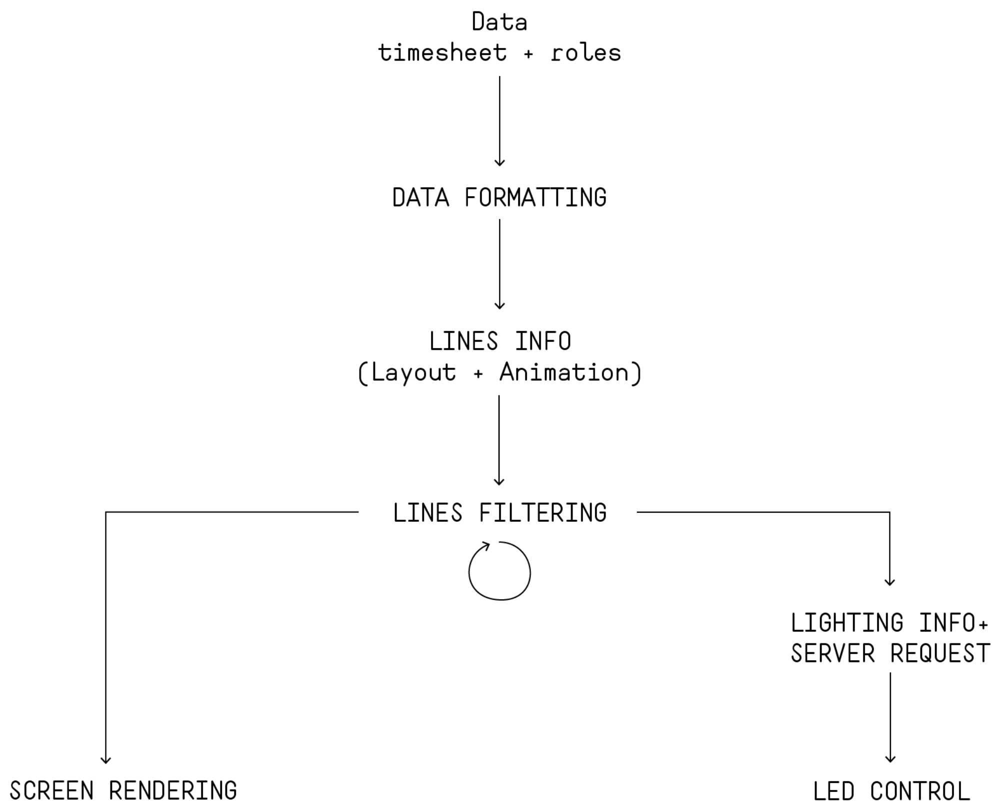


Figure 7.3: Logic blocks of the installation

with dates easier, the Luxon library was utilized, which is a robust, contemporary, and user-friendly wrapper for JavaScript dates and times [60].

Lines Information Combining the formatted timesheet data with information relative to roles, this section generates a structure containing the main drawing properties of each segment, regarding both the static display and the animation. Then, two types of lines are computed: project lines (which reflect each person's contributions to projects) and connection lines (which connect the same person's contributions). Because of their differing behavior, static line characteristics are computed separately for these two categories: `x/y` coordinates, line path and length, stroke color, width, and opacity. D3's scaling functions are utilized to map the data to line information: `ScaleLinear` to correlate time to the continuous `x` axis and `ScaleOrdinal` to associate roles to colors. Other D3 modules, line and `curveBumpX`, are constructed to build a smooth curve from the two endpoints that correspond to the coordinates of each segment. This structure is then grouped by persons, and animation information is calculated: duration, based on line length, and delay, calculated as the total of all delays in the preceding segments belonging to the delay's person plus the duration of the previous segment. The final result of this step is an array of all the lines derived from the dataset, each one carrying all the essential information to be drawn and animated appropriately.

Lines Filtering The whole collection of lines is periodically filtered based on what the narrative sequence is highlighting. Filtering by groups of individuals, projects, or roles has been implemented. The filtered lines' information is utilized to produce the visuals on the screen as well as to light the LED tubes. This action is periodically reiterated, with a different filtering selection.

Screen Rendering This step employs React components to draw each of the filtered segments, using the lines information previously generated. It is preferable to precompute as much drawing information as possible outside of the component, in order to optimize the number of rendering cycles. The main component that is utilized translate the data into SVG lines and animate them is `AnimatedDataset`, a React component for data animation released in 2020 in by Accurat's development team [61]. `AnimatedDataset` makes use of the capabilities of D3 data join to produce robust animations, but uses D3 only below the curtains, without leaving the comfortable declarative style of React. For each value of a given dataset, it generates an SVG element, connecting each attribute name to its value or to a function that returns a value from a single dataset element(datum). By giving a different dataset or changing attributes, animation is automatically performed. `AnimatedDataset` lets the developer act on the single attributes that should be animated for each of the three animations phases: enter(datum is added), update(datum is transformed), exit(datum is erased). It is possible to modify animation states, defining attribute values for each entering or outgoing datum. The CSS path attributes `stroke-dasharray` and `stroke-dashoffset` were used to generate the line drawing animation, changing the dasharray value to the same length as the line and animating the dashoffset property from the same length as the path (displaying an empty gap) to 0 (all line drawn).

Lighting info and Server Request The filtered lines information is also used to compute the necessary data to light up the LED. An array of objects containing information about color and the number of LEDs to light up for each color is created from the number of person visualized for each role. This array is then sent to the Python Server via HTTP Post Request.

LED Control To handle http requests and control the LEDs, a Python server was built using the microframework Flask. The server receives through `POST` request information regarding color and number of LEDs to animate, direction and duration of the animation as well as which tube to address. To address the LEDs, the Adafruit CircuitPython NeoPixel library was used. The PWM (pulse-width modulation) module can generate a signal with a specific duty cycle, for example to dim a LED. By using DMA(Direct Access Memory) to send a specific sequence of bytes to the PWM module, the NeoPixel data signal can be generated without being interrupted by the Raspberry Pi's operating system. The NeoPixel driver renders the strip as a series of pixels, each of which can hold its color as a tuple of RGB values. The two strips each contain 60 LEDs. However, because only one strip can be controlled at a time, they are wired together and addressed as a single strip of 120 LEDs, thus addressing the first or second strip implies showing the animation on a separate range of LEDs, 0-59 and 60-120, respectively. The NeoPixel driver presents the strip as a sequence of pixels, which can store their color as tuples of RGB values. Each of the two strips contains 60 LEDs, but they are wired together and are addressed as a single strip of 120 LEDs, so addressing the first or second strip means displaying the animation on a different range of LEDs, 0-59 and 60-120 respectively.

8 Conclusion

This work resulted in the design, development, and building of a data-driven installation prototype that addresses the complexities of a company's evolution, work relationships, and projects. With a focus on the human context underlying the company's work outcomes, the contribution of each employee was meant to be the major element shaping the projects made and the overall company's history. Particular care was taken to preserve the richness of the various perspectives and stories arising from team work activities and values: while avoiding oversimplification, the visualization aims to deliver the multiple shades of the human side behind Accurat's work through the narration of various micro-themes.
To highlight the different layers arising from the presentation of such rich and detailed data, substantial filtering was used, proposing the ideation of animated filtering sequences centered on the various selected micro-themes. Furthermore, several modifications to the pieces' vertical placement were recommended to improve readability in the overall static visualization. To encourage engagement and draw attention to the filtering events, a multi-channel media system was designed to display the visual model: the installation consists in the integration and dialogue of a dynamic visualization displayed on a screen and the animation of LED lights inside tubes(simulating neon lights) positioned near the screen. The weaknesses of one medium are compensated for by the other in this system: the aesthetic impact of LED lights synthesizes the filtering event while enticing the viewer to the visualization, and their limited capability of information portrayal is compensated for by the animated visualization on the screen.
The project code repository and documentation can be found at this link: `https://github.com/sisbet/data-home-thesis`

8.1 Issues and future development

As previously stated, the work detailed in the thesis is to be seen as a proof of concept, and as such displays purposeful restrictions, demonstrating an idea from the discussion of all its elements, even if they are incomplete. As a result, there are several possibilities for the project's future improvement and growth: given that a large portion of the data must be acquired, the most relevant improvement directions in terms of information presentation are listed here.

- Micro-Narratives. The combination of filtering and animation was judged as one of the main elements that made the visualization successful, thanks to its capability of layering the visualization and encouraging attention. Being it an indispensable element, it is crucial to develop interesting narrative sequences, going deep in each of the micro-themes proposed(i.e. evolution of individual skills, interpersonal relationships, composition of working groups), finding insights and patterns that are less evident in the current visualization. This work would greatly improve the installation in two ways: making the narratives explicit would give much more meaning to the visualization, promoting an empathic connection with the user; and the targeted use of filtering would allow the created visualization to be used as an analytical tool, usable both internally and towards a potential customer, in order to understand and show relevant internal working dynamics.

- Temporal data. As described in section 6.1, the prototype focused on the visualization of the year 2020, but the intended installation should encompass Accurat's 10 years history and integrate it with the ongoing activities of the work team, resulting in a dynamic product that always communicates something new. The proposed visualization maps time on the horizontal axis, allowing users to view time ranges outside the screen window via a panning motion. Although considerable attention was given to making the visual model adaptable to multiple temporal resolutions, there is still the need to handle specific challenges coming from actual data.

- Environmental Scalability. The prospect of having various sets of screen and lights of varying dimension and form that would be positioned in different areas of the workplace was considered in the design of the prototype, leading to the idea of a distributed and pervasive installation. While some sort of complementarity between the material exhibited on each distinct screen has been imaged, the particular distribution of the visuals and how the various components of the installation interact with each other has not yet been clarified. Investigation in this direction might lead to some intriguing possibilities, such as providing each part of the installation a meaning and function related to the space's purpose in which it is located, or employing light animation to establish coherence across several screens.

- Sonification. The profoundly emotional and immensely expressive nature of sound makes it a powerful means to convey data in a very different way from graphics. The installation's future evolution will also include the addition of a sound dimension to communicate sentiments, build connection, and immerse the user in the work, exploring ways of codifying data into sound.

- Projects Positioning. Some different project displacements and sortings were suggested to increase readability. They have not, however, been tested in the prototype implementation stage. Other, more sophisticated alternatives, such as the construction of a force-driven system, may be tried, but given the intricacy of the many interactions, this would not necessarily generate satisfying results, and in certain circumstances, a manual approach may be preferred. Moreover, the compacting of the collaboration lines within a project and the representation of extensive time gaps has not been implemented yet. Furthermore, the compacting of cooperation lines within a project and the portrayal of large time gaps have yet to be implemented.

- Data Portraits. The usage of data portraits has only been considered in this work in terms of its relationship to the fundamental architecture of the visual model. Specific visual codifying of the individual's distinctive features has not been addressed: given the major significance of people in the installation conception, it is undoubtedly an essential aspect of the projects' future objectives. An intriguing development might result from the conception of dynamic data portraits that alter in accordance with the micro-narrative exhibited or with the person's evolution.

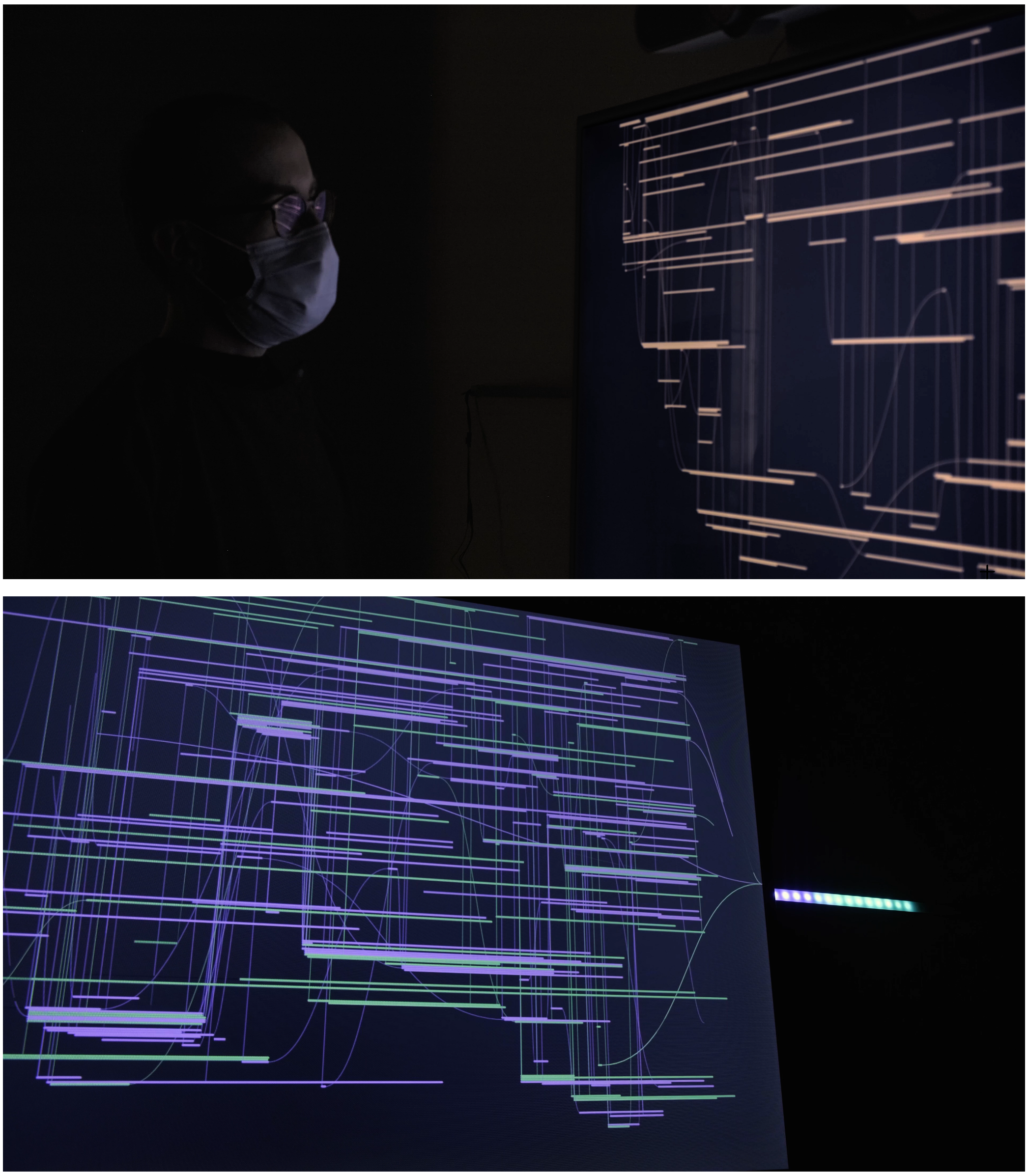


Figure 8.1: Details of the data-driven installation

Bibliography

1. H. Kennedy and W. Allen. "Data Visualisation as an Emerging Tool for Online Research". In: vol. The SAGE Handbook of Online Research Methods AU. SAGE Publications Ltd, 55 City Road; 55 City Road, London, 2017. DOI: 10.4135/9781473957992. URL: https://sk.sagepub.com/reference/the-sage-handbook-of-online-research-methods-2e.
2. V. Mayer-Schönberger and K. Cukier. *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. 2013.
3. G. Lupi. "Data Humanism: The Revolution will be Visualized". *PrintMag*, 2016. URL: https://www.printmag.com/article/data-humanism-future-of-data-visualization/.
4. *Archive Dreaming*. en-US. URL: https://refikanadol.com/works/archive-dreaming/ (visited on 12/08/2021).
5. *Melting Memories*. en-US. URL: https://refikanadol.com/works/melting-memories/ (visited on 12/08/2021).
6. *Accurat - Sculpting Data with Google Trends*. en. URL: https://www.accurat.it (visited on 12/08/2021).
7. T. Breur. "Statistical Power Analysis and the contemporary "crisis" in social sciences". 4:2-3, 2016, pp. 61–65. DOI: 10.1057/s41270-016-0001-3. URL: https://doi.org/10.1057%2Fs41270-016-0001-3.
8. D. Laney. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Technical report. META Group, 2001. URL: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.
9. F. X. Diebold. "A Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline, Second Version", 2012. DOI: 10.2139/ssrn.2202843. URL: https://doi.org/10.2139%2Fssrn.2202843.

10. D. Boyd and K. Crawford. "Critical questions for Big Data". 15:5, 2012, pp. 662–679. DOI: 10.1080/1369118x.2012.678878. URL: https://doi.org/10.1080%2F1369118x.2012.678878.
11. D. Bollier. *The Promise and Peril of Big Data*. Technical report. The Aspen Institute, 2010.
12. T. Wang. "Why Big Data Needs Thick Data." *Medium*, 2016.
13. M. Cassidy. "Big data is yielding to thick data and that's a good thing". *Bloom-Reach*, 2014.
14. H. Kennedy and M. Engebretsen. *Introduction : The relationships between graphs, charts, maps and meanings, feelings, engagements*. Amsterdam University Press, 2020, pp. 19–32. DOI: 10 . 1515 / 9789048543137 - 005. URL: https://doi.org/10.1515%2F9789048543137-005.
15. B. R. Heravi. *Teaching Data Journalism*. 2017.
16. J. Gray, L. Bounegru, and L. Chambers. *The Data Journalism Handbook*. O'Reilly Media, Sebastopol, CA, 2012.
17. S. Few. *Data art vs. data visualization: why does a distinction matter?* URL: http://www.perceptualedge.com/blog/?p=1245.
18. *Accurat - The Room of Change at the 2019 Milan Triennale*. en. URL: https://www.accurat.it (visited on 12/08/2021).
19. A. Kirk. *Data Visualization: a successful design process*. Packt Publishing, 2012.
20. H. Mueller and J.-C. Freytag. "Problems , Methods , and Challenges in Comprehensive Data Cleansing". In: 2005.
21. R. Roiger. *Data Mining A Tutorial-Based Primer (Second Edition)*. 2016. ISBN: ISBN 9781498763974.
22. H. L. Capri. *Data mining : principles, applications and emerging challenges / Harold L. Capri, editor.* eng. Computer Science, Technology and Applications. Nova Publishers, New York, New York, 2015 - 2015. ISBN: 1-63463-770-4.
23. A. Cairo. *The Functional Art An introduction to information graphics and visualization*. New Riders, 2013.
24. G. Lupi. "The New Aesthetic of Data Narrative". en. In: ed. by D. Bihanic. Springer, London, 2015, pp. 57–88. ISBN: 9781447165965. DOI: 10.1007/978-1-4471-6596-5_3. URL: https://doi.org/10.1007/978-1-4471-6596-5_3 (visited on 11/30/2021).

25. E. R. Tufte. *The Visual Display of Quantitative Information*. 2nd ed. Graphics Press, Cheshire, CT, 2001. ISBN: 978-0-9613921-4-7.

26. C. Ware. *Information Visualization: Perception for Design*. 3rd ed. Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann, Amsterdam, 2012. ISBN: 978-0-12-381464-7. URL: http://www.sciencedirect.com/science/book/9780123814647.

27. C. Healey and J. Enns. "Attention and Visual Memory in Visualization and Computer Graphics". *IEEE Transactions on Visualization and Computer Graphics* 18:7, 2012, pp. 1170–1188. DOI: 10.1109/TVCG.2011.127.

28. D. Huber and C. Healey. "Visualizing data with motion". In: *VIS 05. IEEE Visualization, 2005.* 2005, pp. 527–534. DOI: 10.1109/VISUAL.2005.1532838.

29. L. R. Bratram. *Enhancing Information Visualization with Motion*. 2001.

30. D. Todorovic. "Gestalt principles". *Scholarpedia* 3:12, 2008, p. 5345. DOI: 10.4249/scholarpedia.5345. URL: https://doi.org/10.4249%2Fscholarpedia.5345.

31. J. Bertin. *Sémiologie graphique. Les diagrammes, les réseaux, les cartes.* Mouton Gauthier-Villars, Paris ;La Haye ;Paris, 1967.

32. D. A. Morrison. "The Book of Trees: Visualizing Branches of Knowledge. â€" By Manuel Lima.Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. â€" By Isabel Meirelles." *Systematic Biology* 64:2, 2014, pp. 363–365. ISSN: 1063-5157. DOI: 10.1093/sysbio/syu092. eprint: https://academic.oup.com/sysbio/article-pdf/64/2/363/24586872/syu092.pdf. URL: https://doi.org/10.1093/sysbio/syu092.

33. I. Meirelles. "Dynamic Visual Formation." The Language of Dynamic Media. Works from the Dynamic Media Institute at the Massachusetts College of Art, 2005, pp. 80–85.

34. M. Botta. *Design dell'informazione. Tassonomie per la progettazione di sistemi grafici auto-nomatici*. Artimedia, Trento, 2006, p. 272. ISBN: 8887980373.

35. S. Rendgen and J. Wiedemann. *Information Graphics*. Taschen, Köln, 2012. ISBN: 9783836528795.

36. R. E. Roth. *Visual Variables*. 2017. DOI: 10.1002/9781118786352.wbieg0761. URL: https://doi.org/10.1002%2F9781118786352.wbieg0761.

37. J. Morrison. "Theoretical Framework for Cartographic Generalization With Emphasis on the Process of Symbolization". *International Yearbook of Cartography*, 1974.

38. A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. "Visual Semiotics & Uncertainty Visualization: An Empirical Study". *IEEE Transactions on Visualization and Computer Graphics* 18:12, 2012, pp. 2496–2505. DOI: 10.1109/tvcg.2012.279. URL: https://doi.org/10.1109%2Ftvcg.2012.279.

39. A. V. Abela. "The presentation: A Story About Communicating Successfully With Very Few Slides". PhD thesis. 2010.

40. *IBM Design Language | Data Visualization: Visualization*. en. 2016. URL: http://www.ibm.com/design/language/experience/data-visualization/visualization (visited on 12/08/2021).

41. *Google Material Design- Data Visualization*. URL: https://material.io/design/communication/data-visualization.html.

42. *Ferdio Data Viz Project*. URL: https://datavizproject.com/.

43. *Alluvial vs Sankey Diagrams – the difference explained*. en-US. URL: http://www.datasmith.org/2020/05/02/alluvial-plots-vs-sankey-diagrams/ (visited on 12/06/2021).

44. S. Ortiz. "Living Networks". In: *New Challenges for Data Design*. Springer London, 2014, pp. 159–174. DOI: 10.1007/978-1-4471-6596-5_8. URL: https://doi.org/10.1007%2F978-1-4471-6596-5_8.

45. I. de la Torre - Arenas and P. Cruz. "A Taxonomy of Motion Applications in Data Visualization". In: *Proceedings of the Symposium on Computational Aesthetics*. CAE '17. Association for Computing Machinery, Los Angeles, California, 2017. ISBN: 9781450350808. DOI: 10.1145/3092912.3122798. URL: https://doi.org/10.1145/3092912.3122798.

46. L. Bartram and C. Ware. "Filtering and Brushing with Motion". en. *Information Visualization* 1:1, 2002, pp. 66–79. ISSN: 1473-8716. DOI: 10.1057/palgrave.ivs.9500005. URL: https://doi.org/10.1057/palgrave.ivs.9500005 (visited on 12/07/2021).

47. L. Bartram and A. Nakatani. "What Makes Motion Meaningful? Affective Properties of Abstract Motion". In: *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*. 2010, pp. 468–474. DOI: 10.1109/PSIVT.2010.85.

48. L. Bartram, C. Ware, and T. Calvert. "Moticons:: detection, distraction and task". en. *International Journal of Human-Computer Studies*. Notification User Interfaces 58:5, 2003, pp. 515–545. ISSN: 1071-5819. DOI: 10.1016/S1071-5819(03)00021-1. URL: https://www.sciencedirect.com/science/article/pii/S1071581903000211 (visited on 12/07/2021).

49. A. Zotta. *La comunicazione empatica dei dati. Studio delle variabili cinetiche per il coinvolgimento emotivo del lettore nella rappresentazione dei dati*. 2019. URL: http://hdl.handle.net/10589/147454.

50. M. Bostock, V. Ogievetsky, and J. Heer. "D³: Data-Driven Documents". eng. *IEEE transactions on visualization and computer graphics* 17:12, 2011, pp. 2301–2309. ISSN: 1941-0506. DOI: 10.1109/TVCG.2011.185.

51. J. Lloyd. "Practical advantages of declarative programming". *Unknown*, 1994, pp. 3–17.

52. *React – A JavaScript library for building user interfaces*. en. URL: https://reactjs.org/ (visited on 11/30/2021).

53. Z. Hemel. *Facebook's React JavaScript User Interfaces Library Receives Mixed Reviews*. 2013. URL: https://www.infoq.com/news/2013/06/facebook-react/ (visited on 11/30/2021).

54. D. Deutsch. *Understanding MVC Architecture with React*. en. 2020. URL: https://medium.com/createdd-notes/understanding-mvc-architecture-with-react-6cd38e91fefd (visited on 11/30/2021).

55. F. Copes. *Unidirectional Data Flow in React*. en. URL: https://flaviocopes.com/react-unidirectional-data-flow/ (visited on 11/30/2021).

56. *Refs and the DOM – React*. en. URL: https://reactjs.org/docs/refs-and-the-dom.html (visited on 11/30/2021).

57. *Using React with D3.js*. en. URL: https://wattenberger.com/blog/react-and-d3 (visited on 11/30/2021).

58. *NeoPixels on Raspberry Pi*. en-US. URL: https://learn.adafruit.com/neopixels-on-raspberry-pi/raspberry-pi-wiring (visited on 11/30/2021).

59. *lodash*. original-date: 2012-04-07T04:11:46Z. 2021. URL: https://github.com/lodash/lodash (visited on 12/08/2021).

60. *Luxon*. original-date: 2015-11-30T12:48:48Z. 2021. URL: https://github.com/moment/luxon (visited on 12/08/2021).

61. *AnimatedDataset*. original-date: 2020-01-16T17:04:57Z. 2021. URL: https://github.com/accurat/react-animated-dataset (visited on 11/30/2021).

Appendix

This appendix includes additional pictures that complete the presentation of the exploratory visualizations, as well as extra details of specific frames of the produced animation.

Figure A.1: Exploratory visualization: project blocks with colored people

Figure A.2: Exploratory visualization: projects with collaboration-lines(time distinction for each collaboration)

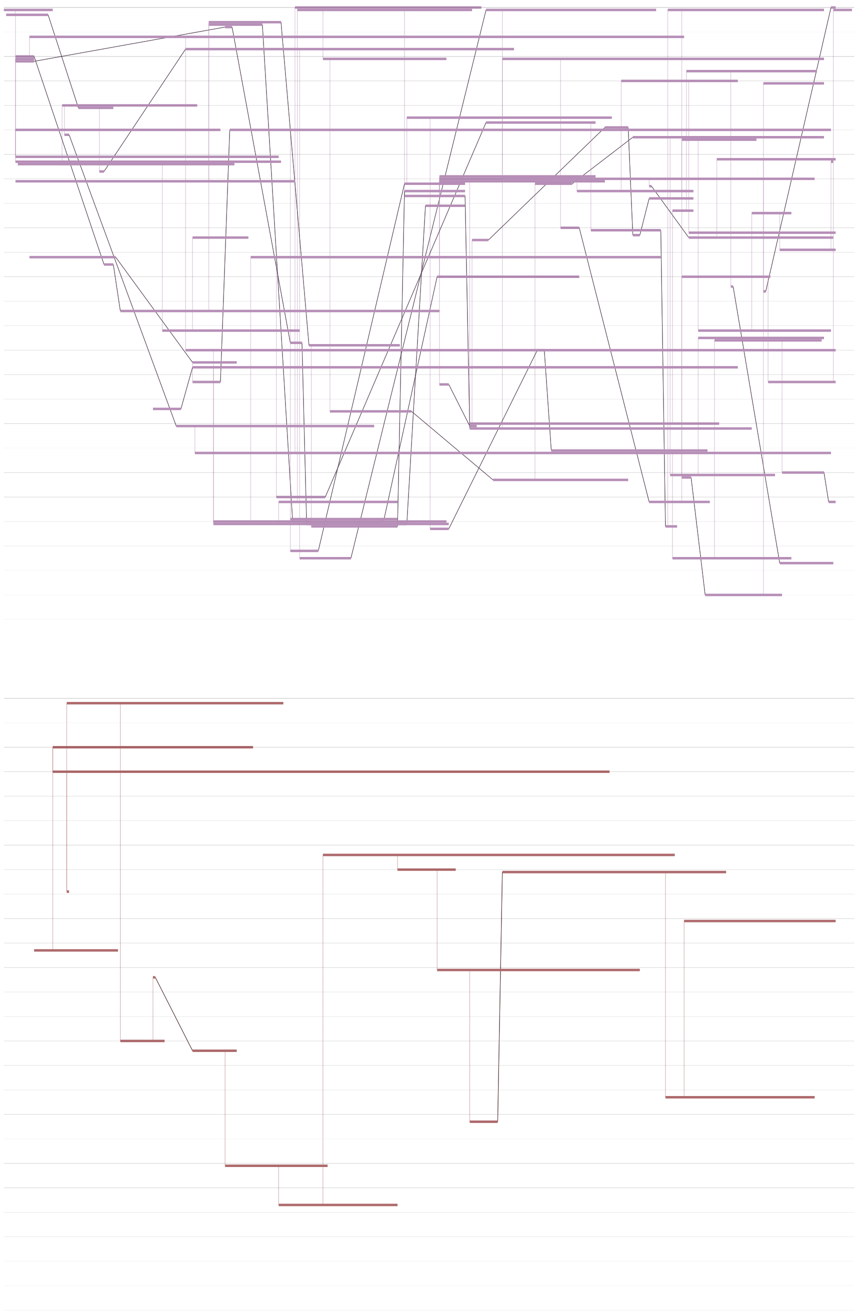
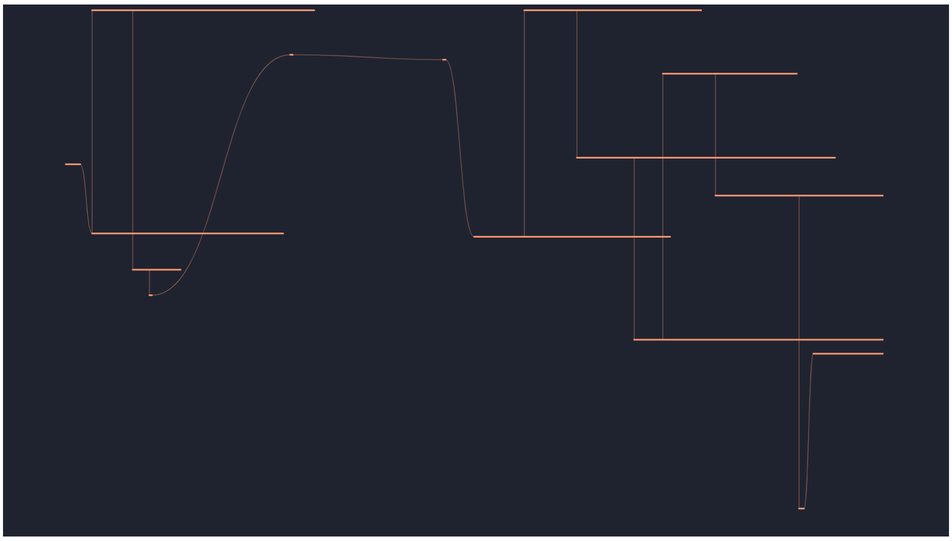


Figure A.3: Exploratory visualizations of filtering



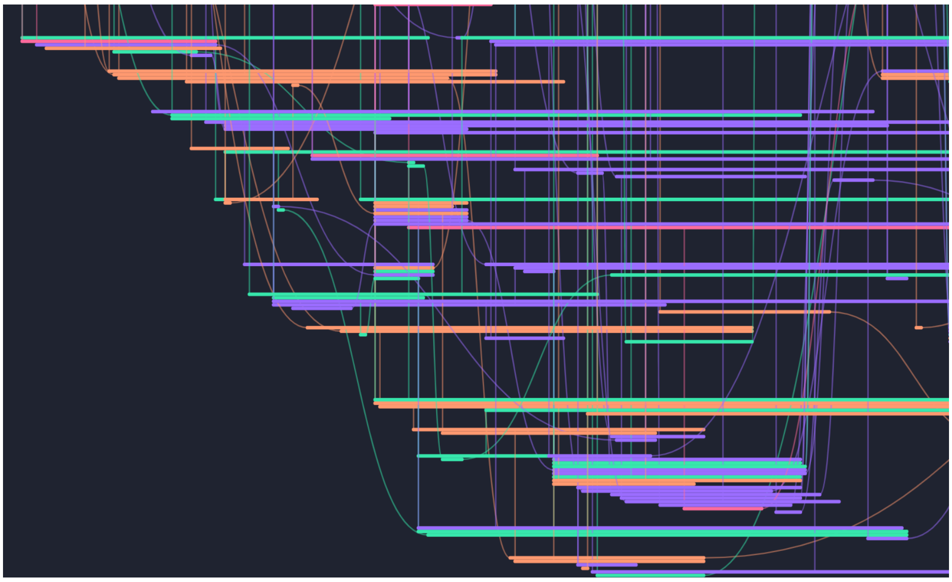


Figure A.4: More animation frames