LUISS

Department of Business and Management

Data Science and Management

Course of Data Visualization

The importance of Data visualization in Healthcare: GIMEMA’s Dashboard

Prof. Blerina Sinaimeri
SUPERVISOR

Prof. Irene Finocchi
Co-SUPERVISOR

758151 Gabriele Di Palma
CANDIDATE

Academic Year 2023/2024

Table of Contents

Introduction

In this thesis, I will discuss the significance of data science and data visualization research in several disciplines of study, with an emphasis on healthcare and, specifically, hematological investigations. The dissertation will address many areas of data visualization, attempting to study and define why I believe it is a critical field of analysis that enables data scientists and other figures to better express the various occurrences discovered throughout their analyses. We will examine the history of data visualization and, through examples, discover why visualizing data is such an important component of data analysis and how it may have a significant impact on people's perceptions of events.

The study will address the major areas of hematology, attempting to demonstrate the value of statistical studies and, more specifically, how Data Science may improve the many studies conducted by researchers. The value of data visualization stems from its capacity to transform complex data into easily interpretable visual formats. This skill is critical in healthcare, as timely and precise data analysis is frequently required for decision making. By converting raw data into visual insights, healthcare practitioners can discover trends, track patient outcomes, and improve research efficiency. This thesis digs into the fundamentals of data visualization and how they are used to produce effective visual tools that support clinical and research activities.

The thesis will also focus on the studies undertaken by GIMEMA Onlus, providing practical examples of how and why I feel that data and the use of various analytical methods can aid better research in this area.

Later in the dissertation, I will demonstrate a real example of how data visualization may assist researchers simplify and enhance their studies by displaying a Dashboard that I designed for GIMEMA using R and, in particular, the Shiny library. The dashboard offers various information, both managerial and research-related, that can assist GIMEMA and its employees in conducting their studies more effectively due to the numerous analytical visualizations available in the various sections of the dashboard.

The dissertation demonstrates how a tool such as a Dashboard can improve many people's quality of life while also increasing the effectiveness and efficiency of studies in the hematological sector, as well as in all disciplines of labor.

The dissertation was completed with the assistance of GIMEMA, which provided me with the tools and expertise necessary to develop the dashboard and learn firsthand the value of statistics and data analytics in such a critical field of research as hematology.

This introduction lays the framework for a detailed investigation of data visualization in healthcare by describing the development process, challenges, and benefits of creating a data visualization dashboard. The subsequent chapters will delve deeper into the theoretical basis of data visualization, the technical design of the GIMEMA dashboard, and its application in hematology research and clinical practice.

1. What is Data visualization and why is it so important?

To start my study I will first need to give some basic definition of what is data visualization and what is not data visualization; also I will show why I, and many other researchers, believe that this is a critical field of research and any data scientist must deepen their understanding of this science and try to give a focal point to the visualization of their data.

1.1 Definition of Data visualization

Data visualization can be defined in many ways but for this dissertation, I will follow the definition given by Hal Varian, Googl's Chief Economist, in 2009 where he defines it as the ability to take data and be able to understand it, process it, extract value from it, visualize it and communicate it. This means that 'data visualization is a complementary scarce factor in understanding the data and extracting value from it' (H. Varian).

Data Visualization might seem like an objective science but the designer of visualization has great control over which message he wants to convey. Using accurate data, he can manipulate the choices that will then be made by those looking at the data and can also make us feel different things by creating illusions of correlation even where none exists.

When designing our visualizations we can decide to show every data point (scatterplots) or show instead only some statistical summaries (histograms). Moreover, we can choose to represent Raw data or Transformed data, the distinction is usually made by those analyzing the data and based on what they have to do. The distinction is thus merely subjective and depends on the situation. Indeed, it is always important to have a base of concepts and terminology to build on since, as stated before, we need to convey information with visualizations, thus we need to simplify as much as possible our charts or graphs to help the interpretation of the shown information.

Another crucial point that we need to define when talking about Data visualization is the difference between Presentation and Exploratory graphics. Antony Unwin defines two macro areas of graphics, those that are used to present your results where graphs should be well drawn with an effective accompanying exploratory text; and graphics used to explore data that are not for the audience but only for yourself. These exploratory graphics should not be perfect but should be used to provide alternative views and additional information to statistics and analysis.

Presentation graphics should convey known information and are usually designed to attract attention, while Exploratory graphics are useful for finding new information and should direct attention to information (A. Unwin, 2020).

1.2 The importance of data visualization

Data visualization is useful for many aspects of the data like cleaning, exploring different data structures, identifying outliers, identifying trends, and clusters, spotting patterns, evaluating modeling output, and lastly presenting results.

Graphics reveal data features that statistics and models may miss, like some unusual distributions, local patterns, clustering, missing values, and more. This means graphics are key to all the steps of data analysis and raise questions that can stimulate different ideas and improve the research process.

A key point when discussing Data visualization is the focus we need to put on the reasons for drawing the graphics. Which message does the designer want to send? How should we interpret the graphics? How has the designer drawn them? Which other graphics might be drawn?

Because of these questions, it is important to say that Graphics on their own are insufficient but they should be complemented with text. This synergy of text and graphics allows the designer to talk through its graphics and explain them to others, allowing them to answer all the important questions we defined before.

In the last year thanks to the development of new and better hardware, we tend to require more precise reproduction and faster drawings. Computer scientists have become much more involved in data visualization thanks to the progress in developing many theories of graphics. The constant research to find problems of color and perception is a key point in data visualization.

Today very large datasets can be easily analyzed and visualized thanks to the improved hardware, thus graphics play a valuable role in finding strengths and weaknesses of the complex models created to analyze the data.

Data visualization is now found everywhere and is a huge improvement from 20 years ago.

1.3 Principles of Data Visualization and Design Techniques

In the world of Data visualization there exist many different techniques that allow the designer to represent data and to convey different messages. The simplest example is the drawing of two numbers, they could be shown in a pie chart, a bar chart, a scatter plot, or even just by writing the numbers.

In this section of my dissertation, I would like to review some of the most common data visualization techniques and try to show the most common use cases and the different scenarios in which one technique is better. Moreover, I will try to give some basic Principles that should always be respected when drawing graphics.

I would like to start by citing a very common website that shows many ways to communicate two simple quantities, the number 75 and the number 37. The author of the blog post states that for a simple data structure, there are many different and convincing ways in which we can communicate the information we need to convey. In her test, she found 45 different ways to represent the two numbers 75 and 37.

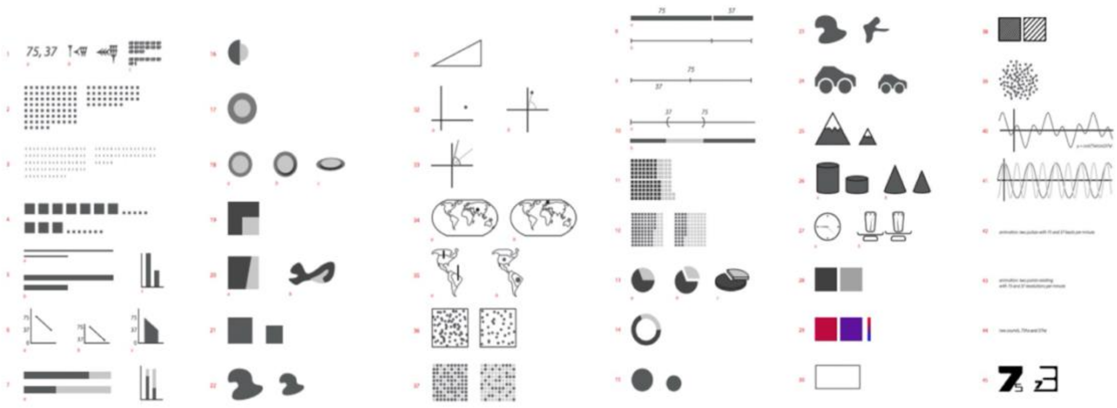


Figure 1

The picture above (Fig. 1) shows some of the different ways in which the crowd and the author of the blog post were able to draw two simple quantities. The results of the test were incredible and she believes that there are many other possible ways to draw data. The author states that by adding more numbers to the list the amount of possible combinations rises exponentially and if we move to datasets that do not contain only values but also units, we will have many more possible ways to convey our message. She defines Data visualization as a language that can explain the world, point out specific facts, and elaborate ambiguous messages. As a language, it works with combinatoric and generative rules and it's thus impossible to delineate the borders or enlist all the objects it can produce.

Moving forward it is important to understand the role of 'perception' when we design some graphics. We, as humans, tend to perceive what we expect thus, our perception is biased based on our past, present, and future expectations. For this reason when designing something we constantly need to avoid ambiguity, be consistent, and understand the goal.

These rules are defined by what are known as Gestalt (form or pattern, from German philosophy) principles and they should always be respected in data visualization.

They are:

- **Proximity:** The human vision system tends to build relationships between elements that are close to each other.
- **Similarity:** Elements sharing visual characteristics (e.g. shape, color, etc…) will be seen as belonging to the same class.
- **Focal point:** The human vision system draws attention toward contrast, thus this can be used to draw attention to outlier elements.
- **Common Region:** Elements tend to be perceived as groups if they share the same area with a clearly defined boundary.
- **Continuity:** Humans are biased to see continuous form even adding missing data if necessary, morever we tend to perceive it as related elements that are arranged in the same orientation.
- **Closure:** We are biased to see whole objects even where they are incomplete
- **Symmetry:** Our vision system is optimized to see structures, resolving complex scenes into a combination of simple and symmetrical elements.
- **Figure/Ground:** We always parse a scene into foreground and background elements.

All these principles show how powerful psychology is when we visualize something. This means that we need to be very careful when designing something because we never know how it could be perceived by the myriad of people looking at our design.

After defining the basic principles we always need to follow when designing a visualization for our data, we need to understand how to choose the right chart for the Right Task.

1.5 Different Charts for Different Tasks and Graphical Integrity

In Data visualization there exist many types of charts and graphs that are very useful to fulfill different tasks. Some of them are Bar Graphs, Line Charts, Scatter Plots, Dual-Ais Charts, Pie Charts and more. Deciding which chart should be used to complete your task might be a very challenging task but we need to remember that choosing the right chart is very important.

The most important thing to analyze before choosing a chart is understanding our main goal, are we comparing values? Are we trying to show proportions? Are we analyzing trends?

Once we understand why we need charts we can choose which chart to use. There exist many online tools that are helpful for the designer to understand which chart best fits the scope of the analysis like the Financial Times Visual Vocabulary and the Data visualization catalog.

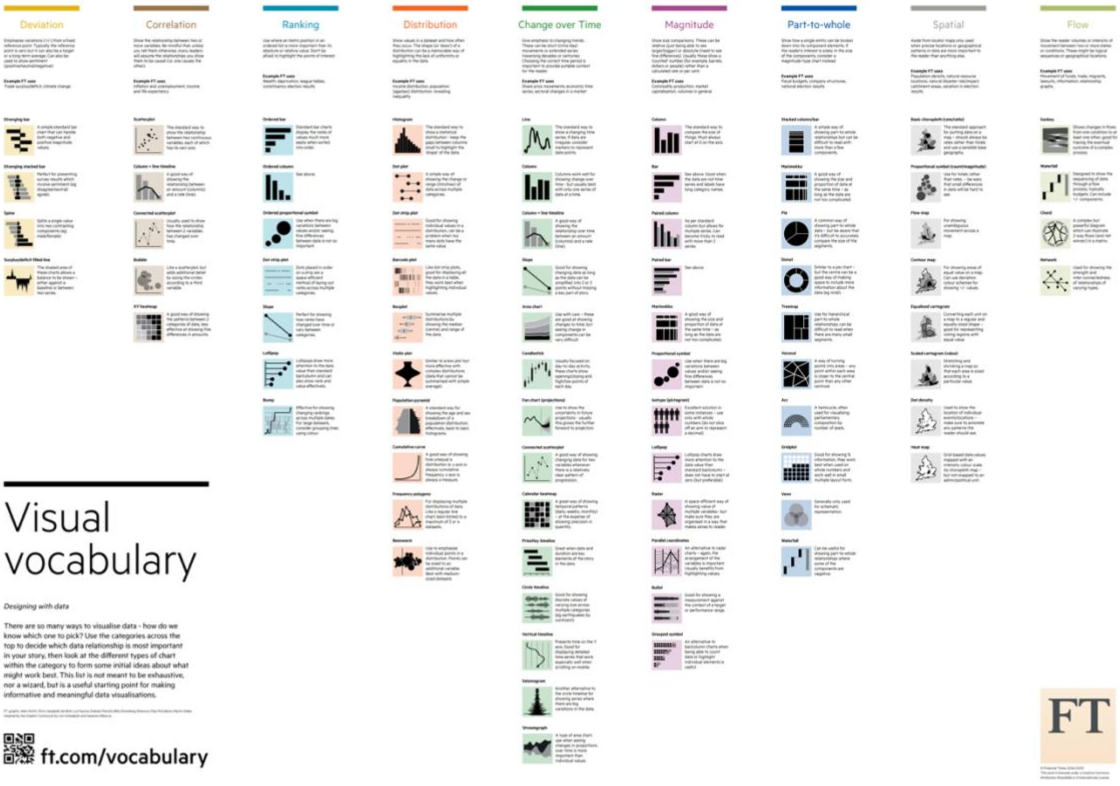


Figure 2

This table (Fig 2.) provided on the Financial Times website provides 9 different classes and the different types of charts that best fit each class. For example, if we are computing a Ranking task, we should use Ordered Bars, Ordered Columns, Slopes, and all those charts that visually transmit a sense of order to the data points we are distributing. Similarly, if we are computing a correlation task we need to show this correlation between data points, thus charts like scatter plots, column + line timelines, heatmaps, and connected scatterplots perfectly contribute to the fulfillment of our task.

The choice of the right chart is only the beginning of the problem because we also need to be very careful in how we design those charts. I want to stress this aspect of data visualization because I believe that today many people misunderstand the importance of showing correct charts. This aspect was very central when showing data during COVID-

19 since many people, also non-experts, were looking at many charts and representations and at that time, showing a wrong chart, or a chart in the wrong way, could have been very harmful both for the population and the analyses. Barcharts, for example, which seem very easy to understand and design are very easy also to misrepresent. Any bar chart should always have a zero-base line because we are comparing two (or more) lengths (or portions) starting from the baseline. Thus if we exclude the baseline it might look like the difference is much higher than it is in reality. This is a common mistake (or bad faith behavior) done when showing election polls, or any political poll since if we zoom in and show the difference between two columns it might look like the difference is much bigger than it is. This again shows how important and impactful the job of a designer is and how much it can have an impact on the perception of people, especially non-experts.

When analyzing a chart there is a very important factor that we could consider looking at, to understand if the chart is showing real facts or is inflating (or deflating) the real data. This factor is called the 'Lie Factor' and is a key aspect of the Graphical Integrity principle defined before. To understand the Lie Factor I want to show an example of the Rt Data for Italian Regions during COVID. This example was provided to us during the course in Data Visualization.

Figure 3

This chart (Fig 3.) shows the Rt Data for each Region in January 2021. During the COVID-19 pandemic, the term "RT data" refers to the virus's effective reproduction number or effective transmission rate. This value indicates how many other individuals an infected person can, on average, transmit the virus to. For example, if the RT data is 2, it means that each infected person, on average, infects two other people. It thus was an essential aspect of the pandemic and each Region was doing its best to reduce it. The chart presents a big mistake we have already discussed before the absence of a zero baseline. Excluding the baseline inflates the difference in RT between each region to show how, in this case, the Lombardia region was doing much better than the others. If we look at the correct chart (Fig 4.) we will see indeed, that the difference is not this large thanks to the presence of the zero baseline that gives a much better and precise representation of the data.

Figure 4

In this case, the Lie Factor can be computed by dividing the 'Size of effect in graphic' by the 'Size of effect in data'. This means that we manually measure the difference between the RT of two regions in the first chart, and then we divide it by the correct difference between the two regions that we get from the data. If the lie Factor is greater than 1 it means that we are inflating our analysis and visualizations, if it is 1, it means that it is graphically correct, if it is smaller than 1 we are deflating the visual representation. For example, if we had a chart of a street showing some distances between two points, we measure with the ruler the distance between the two and then we divide it with the actual distance from the data.

Graphical integrity is thus a key aspect to look at when designing any type of chart. With Graphical Integrity, it is also important to reduce the quantity of information shown to the minimum possible removing all junk from our visualization and reducing as much as possible the ink used to show the data if possible. Moreover, we should never use more dimensions that we need to show our data (e.g. avoid 3D pie charts, 2 information, 3 dimensions). All these rules must be followed to create a chart that is as clear and detailed as possible without confusing the readers.

This concludes my introduction to Data Visualization. In this section I wanted to convince the readers that Data visualization is important, showing all the different aspects of it and the big impact it has on society as a whole. The following chapters of my dissertation will focus more on the Healthcare sector trying to have a more precise look at a practical application of data visualization by deep diving into Healthcare and more specifically Hematology. Later in the thesis, we will also see the concept of Dashboards and a practical example, that I created, showing how charts can be helpful in specific contexts.

2. Data science and Data visualization in Healthcare

In this section of my dissertation, I want to cover a practical application of Data visualization, more specifically we will dive into the world of Data visualization in the Healthcare sector by looking at the history, of how we started implementing charts and graphs in this context, possible real applications and criticisms that might appear when implementing all the rules and techniques we defined in the previous chapter.

2.1 History of Data Visualization and Healthcare

The first case of Data Visualization in Healthcare can be traced back as far as the 1800's with the first example being the work of Florence Nightingale that used different visualization to examine the causes of mortality in hospitals during the Crimean War. Florence Nightingale witnessed the overcrowding in hospitals because of the war and the poor conditions that wounded the British soldiers. Nightingale was a Nurse who constantly tried to improve hygiene, cleanliness, and staff organization in the hospitals. Thanks to the health reform made by the Sanitary Commission, there was an improvement in reducing infections in hospitals.

Florence Nightingale used her statistical expertise to create a polar area chart (Fig 5.) to present the data at the time. She was helped by William Farr (an epidemiologist) and with him they were able to plot the counts of death by month with each area of the polar chart representing the number of deaths and using color to show the cause of mortality, Red for soldiers that died from wounds, blue for deaths caused by preventable diseases and Black other cases.

From this study, it came out that most of the soldiers died from preventable diseases, then from wounds and other causes. Commonly we refer to this polar chart as a 'coxcomb' or 'rose' diagram and is very helpful since it allows us to enable multiple comparisons of data that can be seen in a single illustration. The figure below represents the original chart created by Florence. In Red, we have the soldiers who died from wounds, in Blue the deaths caused by preventable diseases and in Black the other cases. Around the different sectors, we have the months in which the deaths were recorded in the hospitals.

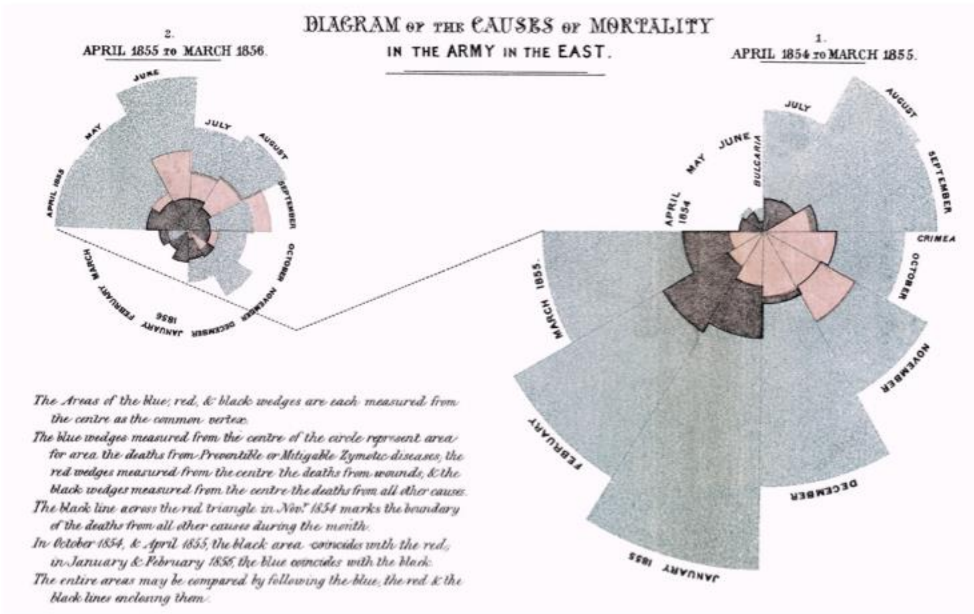


Figure 5

Florence Nightingale used this chart and other statistics to report to the British parliament on the Health of the army. In 1858 she published a report on mortality of the British army that contained numerous charts and tables that explained the health of soldiers at home and in the military hospital overseas. This helped persuade the government officials to improve care for British armed forces active in India by introducing better drainage, clean water, and ventilation to save the lives of thousands of soldiers. She was then elected to the Royal Statistical Society in 1859 as a recognition for her contribution to the field of statistics.

Florence's work and that of others paved the way for modern approaches to visualizing data in healthcare. Digital data has since then become the cornerstone of contemporary evidence-based practice and it increased the necessity to go back to the literature to evaluate the different frameworks that can be used in healthcare to consolidate the best practice. Nightingale's work is still used today and governments, organizations, and citizens still today use some of her techniques to change and affect the visualization in healthcare and many other areas of society. I believe that this historical context underscores the longstanding importance of visualizing data to understand patterns and inform decisions in healthcare.

Today, healthcare data visualization encompasses a wide array of techniques, from traditional charts and graphs to interactive dashboards and advanced machine learning algorithms. These techniques enable healthcare professionals to extract insights from complex datasets efficiently.

2.2 Challenges in the modern world of data visualization

After seeing the first use case of data visualization in the world of healthcare we now must go towards a more modern approach trying to understand all the different improvements, and challenges, that have been appearing in recent years in this scenario.

The study by Stadler et al. (2016) discusses the significance of data visualization dashboards in enhancing the efficiency and ease of healthcare analysis, by also highlighting the challenges faced by healthcare professionals in navigating vast amounts of data and the complex analytical demands involved in healthcare analysis. In their comprehensive exploration, Stadler et al. (2016) delve into the critical significance of data visualization dashboards as transformative tools within the healthcare landscape. They address the formidable challenge confronted by healthcare professionals in managing the vast and complex datasets inherent in modern healthcare systems, coupled with the intricate analytical demands necessary for extracting actionable insights.

A common word used in today's world is Big Data, which is not only a large amount of data but also very varied and valuable data that can be extracted from many different sources. Big data has the potential to transform medical practice by using information generated every day to improve the quality and efficiency of care.'' (*Stadler et al. 2016)*. However, Big Data also increases exponentially the amount of information that we have and this abundance can sometimes be overwhelming.

In healthcare, it is of crucial importance to give immediate feedback on data and this can be done by analyzing only meaningful and helpful data. Sometimes the utility of the data can be easily masked by the noise of numbers and it can require hours of work to mine for value and thus result in prohibitive work.

Another important aspect is the removal of the boundaries of data from a particular facility or system to open a cross-facility analysis that can bring a huge value to data exploration allowing us to compare trends comprehensively and improve a context that is industry-wide to any findings. However while increasing the possible breadth and depth of insight, the removal of boundaries, also increases the complexity of analysis and understanding. "Humans have a limited capacity to process information, and without proper tools, as complexity increases, accuracy of interpretation decreases. The amount of data contained within the EHR raises the possibility of overlooking or misinterpreting the value within the data". (*Stadler et al. 2016)*

One of the solutions that the author proposes to solve this problem is using an accessible, consumable, and meaningful approach which is interactive visualizations. Our goal is to produce easily understood visuals to aid the users in quickly and accurately processing all the information and concluding. The authors emphasize how data visualization dashboards serve as indispensable tools for converting raw healthcare data into actionable insights. These dashboards not only streamline the analytical process but also empower healthcare professionals to make informed decisions that directly impact patient care. By visually representing data trends, patterns, and correlations, these dashboards enable rapid identification of critical information, leading to more efficient and effective healthcare analysis. The authors emphasize the pivotal role played by data visualization dashboards in streamlining healthcare analysis processes. These dashboards serve as dynamic platforms capable of translating raw and often intricate healthcare data into easily comprehensible visual representations. By doing so, they empower healthcare professionals to swiftly discern trends, patterns, and correlations within datasets, facilitating expedited decision-making and bolstering the quality of patient care. From improving operational workflows to enhancing clinical decision-making and patient outcomes, these dashboards demonstrate their versatility and impact in driving positive change within healthcare organizations. From optimizing operational workflows to informing clinical decision-making and ultimately enhancing patient outcomes, these dashboards emerge as versatile tools capable of catalyzing positive change within healthcare organizations.

2.2.1 Types of Data Visualization in Healthcare

In today's world, as we have seen in the previous chapter, there exist many different types and examples of data visualization. In healthcare, it must represent data in a more human-readable form thus, many various types of visualizations can be used.

- Charts and Graphs: These are the most common types of data visualizations and they include all the examples we underwent in the previous chapter:
  - Line charts: Used to show the trend of a variable over time;
  - Bar charts: Used to compare different values across categories;
  - Scatter plots: Used to show the relationship between two variables.
- Infographics: These are graphical representations of data and information that can engagingly communicate complex data.
- Geographic Maps: Maps visualize geographic data, such as disease prevalence or health outcomes, and identify patterns or trends across different regions.

- Dashboards: These are visual displays of different Key performance indicators (KPIs) and metrics that provide an immediate view of the data. They can be used for real-time monitoring of patient health status that can for example track clinical quality metrics or monitor population health trends. They include a combination of charts, graphs, tables, and other visual elements to provide a comprehensive view of the data. We will explore more in detail the concept of dashboards in the next chapter since they will be a key aspect of the last chapter of my dissertation and in my opinion the most important tool for visualizing healthcare data.

2.3 Importance of data visualization for healthcare

The benefits that we can find in the healthcare sector thanks to the introduction of new data visualization tools are numerous and significant. Thanks to the adoption of different visualization tools, healthcare providers can increase the provision of healthcare, improve patient care, and also improve the organization of the healthcare system.

As discussed in the previous chapter Data visualization alone is important but by joining it with Data Analysis and Machine learning we can create new impactful ways to identify disease trends and patterns among specific populations. For example, it can help patients make informed decisions about their healthcare provision and raise their awareness about modifying their lifestyles for better health outcomes.

Some of the most common improvements that we have seen in Healthcare thanks to data visualization are *Improvements in overall patient care, Disease trend, and pattern recognition, Simplified Data Presentation, Accelerated Performance,* and *Healthcare fraud visualization and error detection*. I will go through all of them to have a much deeper understanding of the importance of Data Visualization and Analysis in this field by following the article published on Kodjin in April 2023 on Healthcare Data Visualization: Examples and Benefits.

2.3.1 Improvement in overall patient care

Data visualization in healthcare data can positively impact the general provisioning of healthcare, support its providers in clinical decision-making, and facilitate their ability to predict and react to potential threats. Visualizing health data in real-time can improve the quality of care and enable healthcare providers to make necessary clinical decisions based on their situation. An example is the usage of different visualization techniques to monitor various parameters such as oxygen saturation to evaluate in real-time the patient's response to the treatment provided. Real-time monitoring can significantly help providers detect abnormal parameters to facilitate the intervention that each patient needs and improve the overall healthcare quality in the healthcare facility

2.3.2 Disease trend and pattern recognition

Determining specific trends in healthcare can be vital for making decisions regarding healthcare provision. Data visualization can help identify disease patterns among specific populations and thus raise awareness about modifying some lifestyles to improve health and address specific health threats.

Data analysis and visualization can for example be used on obesity where trend and pattern recognition are of fundamental importance. Identifying the factors behind obesity can help raise awareness about it and help patients modify their lifestyles.

2.3.3 Simplified Data Presentation

Data visualization can help present data for various audiences and this is a critical aspect of healthcare since the data can be more challenging to interpret among providers who work in different disciplines compared to individuals with medical backgrounds.

Data presentation helps presenting the complex data so that it can be easily interpreted by any audience regardless of their background. We can illustrate disease prevalence and the factors affecting it to make medical data easily accessible to any intended audience.

2.3.4 Accelerated performance

Data visualization accelerates the performance of healthcare providers with the use of different measures such as ensuring prompt clinical decision-making in critical situations to impact patient prognosis and health status. Accelerated performance can also reduce the inadequacies of the provided care guaranteeing better performance from healthcare organizations and building a good reputation for efficiency and better patient outcomes.

2.3.5 Healthcare Fraud Visualization and Error Detection

Implementing different analytics and visualization for health data can enable the detection of frauds and errors that occur in organizations (e.g. medical bills). For example, the most common fraud in billing comes from duplicate billing, phantom billing, and false prescriptions. Having a clear and proper correlation among stakeholders can enhance the integrity of the billing process and decrease fraudulent schemes. Thus implementing data visualization techniques has significantly improved transparency in the healthcare sector, leading to better healthcare outcomes for patients and more efficient and effective healthcare provisions.

2.4 Dashboards in Healthcare Data Visualization

One of the most powerful tools in data visualization is dashboards. As we have seen before, they implement different interactive visualizations that are easily accessible, consumable, and meaningful for the user giving them the possibility not only to immediately access different insights but also to play with the visualizations and analyze with a quick view as many different insights as they want.

The creation of these dashboards is not as easy as people may think, they require the designers to have cross-knowledge on data visualization, data analysis, and website development; we will analyze more deeply how a dashboard can be created in the next chapter where I will show how I created one for GIMEMA, going through all the different steps of creation, personalization and visualization by also looking at the challenges I faced during the creation and the benefits it may bring to the organization.

In this chapter, I will focus more on the theoretical aspect of dashboards understanding what they can do in the health world, and showing why they are a very useful tool with many use cases that can be found in the literature.

The paper from Iftikhar A. et al. *Role of dashboards in improving decision-making in healthcare: A review of the literature*, is a great starting point to understand the basic concepts of dashboards and look at some of the benefits that they bring to the Healthcare world. The authors highlight the growing importance of data-driven decision-making in healthcare, driven by the need for improved efficiency, quality of care, and patient outcomes.

A dashboard is defined as "an important data visualization method to display a visual summarization of information to facilitate new accessible insights for improved decision making or management". Dashboards are a new way to present and easily analyze information with different visual elements. They save time and costs in the decision-making process since they can synthesize a large amount of information.

The authors divide the dashboards into three major applications which are *Monitoring, Analysis,* and *Management* all of which facilitate or aid decision-making. In Healthcare dashboards have been used to integrate different data sources like patient information, ward information, management and more so to have in a single place all the key performance metrics that may be helpful in a healthcare unit.

Dashboards aggregate the available data and provide a summary with different visual elements to help healthcare professionals make more accurate decisions. They are for example useful to present patient history and treatment at a glance, like patient records, and thus observe patterns during specific periods.

Data analytics also has a pivotal role in dashboards since it can show different statistical analyses and patterns to help professionals make better decisions.

Iftikhar et al. show the example of a dashboard that was developed to improve efficiency and accuracy for the heart disease medical team by providing the team with a visualization of different possible treatment strategies based on the available data. Another example showed the effects of implementing a dashboard for STEMI patients (patients suffering from heart-attack diseases) to calculate door-to-balloon time (the interval that starts with the patient's arrival in the emergency department, and ends when a catheter guidewire crosses the culprit lesion in the cardiac cath lab). It was highlighted that the use of the dashboard improved the overall performance and process of approximately 90% of the patients which were able to achieve a 60-minute (optimal time) of door-to-balloon time.

In general, optimal dashboard design provides functions that enable physicians to make decisions with high accuracy and confidence.

The paper goes through an extensive review of existing research and case studies, identifying several key themes and benefits associated with the use of dashboards in healthcare decision-making. These include:

1. Enhanced Visibility and Accessibility: Dashboards allow professionals to have a consolidated view of relevant data, enabling them to monitor performance metrics, track patient outcomes, and identify areas for improvement more effectively.
2. Improved Decision-Making: By presenting data in a visually intuitive manner, dashboards facilitate rapid decision-making by enabling stakeholders to quickly identify trends, patterns, and anomalies within the data.
3. Empowerment of Stakeholders: Dashboards empower frontline staff, managers, and executives alike to access timely and relevant information, fostering a culture of transparency, accountability, and collaboration within healthcare organizations.
4. Support for Performance Monitoring and Quality Improvement: Dashboards enable healthcare organizations to monitor key performance indicators in real-time, allowing them to proactively identify and address issues related to patient safety, resource utilization, and operational efficiency.

5. Facilitation of Continuous Learning and Innovation: Dashboards are useful to forecast data and improve efficiency within healthcare organizations by providing professionals with access to timely feedback and performance data.

Creating a Dashboard can thus result in many improvements for hospitals and healthcare in general, but on the other side, many challenges can arise when a designer wants to implement this interactive visualization for an organization. In the paper *"Requirements and challenges of hospital dashboards: a systematic literature review"*(Reza Raibei and Sohrab Almasi), the authors identify four different groups of challenges that can affect the creation of a dashboard, and they are **data sources and data generation, dashboard content**, **dashboard design,** and **implementation and integration**. They all relate to some of the problems that we identified earlier when talking about the challenges of the new world because of Big Data, for example, data sources and data generation challenges appear when the creator of the dashboard has where to take the data from and how to create the data that is missing, dashboard content relates to the fact that we need to be create a dashboard that is easy and with only the content needed thus it results in a hard job to identify the correct content to display and to show to the users; dashboard design is related to a visual aspect of the dashboard like were to place specific visualizations and how to display them; lastly, implementation and integration is related to the fact that not always it's easy to implement a tool like a dashboard inside an already present system that each organization has. Regarding these challenges, we will see in the next chapter some practical examples when I will show the challenges that I faced when creating a dashboard and how I tried to solve them. In this section, I want to present you the solutions that the authors of the paper proposed to alleviate the different challenges and then later I will compare them with what I did.

Proposed solutions to tackle the issues of finding data sources and generating data include things like putting accurate data into the dashboard, setting up a standardized data warehouse to make sharing easier and cut down on processing time, and updating queries straight from the dashboard. They also suggest using Web services and middleware architecture, and using automatic methods for pulling data instead of manual input. For sorting out what should be on the dashboard, they recommend involving stakeholders in its development and keeping an eye on indicators that match up with what the organization wants to achieve. It's also about not overwhelming the dashboard with too many indicators. To make dashboards easier to understand and use, they suggest matching graphics to the purpose, showing data promptly, using colors effectively, and allowing customization options. It's also important to organize information well using visualization tools so users can get what they need quickly, and understand how older systems fit into clinical activities. To get the dashboard up and running smoothly, they recommend integrating it with other systems, following data standards, training users, securing it properly, and rolling out changes gradually.

The paper cited above also defines some requirements that need to be met, in their opinion, when designing a dashboard for Hospitals. Functional requirements identified in the literature for both quality and clinical dashboards include:

1. Customization: This feature allows users to modify the displayed indicators on the dashboard, tailoring the view to their specific needs and preferences. Customization is crucial for quality dashboards to monitor performance indicators in real-time, and for clinical dashboards to display essential clinical information for patient assessment.
2. Alert Creation: Users can create alerts for indicators that necessitate real-time monitoring, triggering notifications when indicator values exceed defined standards. This functionality is vital in quality dashboards to flag indicators surpassing predefined thresholds, and in clinical dashboards to notify healthcare providers when test or imaging results are available.
3. Tracking: This feature facilitates real-time monitoring of patient locations, aiding in identifying crowded wards and ensuring optimal resource allocation. Tracking assists quality dashboards in monitoring patient locations and ward occupancy, while clinical dashboards, help locate physician instructions and track their progress until results are obtained.
4. Performance Indicator Measurement: Users can compare indicators against standards or national averages and track indicators over time. This functionality enables quality dashboards to measure performance indicators and clinical dashboards to display patients' clinical information.
5. Reporting: Dashboards should have the capability to generate visual reports based on clinical or managerial performance indicators. Additionally, they should support the creation of output files in various formats such as Excel, Word, and PDF.
6. Reminders: The system should generate reminders for maintenance and inspection schedules of hospital equipment, ensuring timely upkeep and safety protocol adherence.

After seeing the improvements and challenges that dashboards bring to healthcare systems, it's important to look for some practical examples of dashboard implementation. One example that I want to show is the use of a dashboard for proactive and intelligent continuous control monitoring (CCM). This example is described thoroughly in the paper *"Continuous Monitoring with Machine Learning and Interactive Data Visualization: An Application to a Healthcare Payroll Process"* (G. Zhang et al. 2022).

To provide a continuous monitoring framework, the article explores the combination of machine learning techniques and interactive data visualization, with a particular emphasis on its application to a hospital payroll process. The authors discuss the drawbacks of conventional payroll monitoring techniques, namely their inability to identify irregularities and discrepancies quickly, and they offer a novel solution that makes use of cutting-edge technology.

The authors urge the implementation of a proactive and automated monitoring framework since they understand how crucial payroll accuracy and integrity are for healthcare organizations, where labor costs often account for a sizeable amount of total expenses. By incorporating machine learning into the monitoring process, predictive models that can recognize patterns and trends in the payroll data and highlight any departures from expected norms can be developed. These algorithms get better at spotting possible problems by regularly examining historical data and picking up on previous trends, which increases the monitoring system's efficacy. The study also highlights how crucial interactive data visualization is for making complicated payroll statistics easier to understand and analyze. Users can dynamically explore and interrogate payroll data through interactive visualizations, getting greater insights into patterns, trends, and potential anomalies. This facilitates prompt actions to solve recognized concerns and improves the decision-making process. By improving decision-making skills, this interactive approach helps stakeholders take prompt, well-informed remedial action and intervention. The healthcare payroll industry is an excellent example of how cutting-edge technological solutions may be applied to optimize workflows and enhance organizational performance using machine learning and interactive data visualization. Healthcare companies can increase their capacity to identify and address payroll process hazards by

incorporating these technologies into a continuous monitoring framework. This will ultimately lead to better financial management and operational effectiveness.

This example is of fundamental importance when analyzing Dashboard usage in healthcare systems since it shows how important interactive visualization can be, not only to monitor patient data but also for the management of the organization. In this way, by simplifying some of the management tasks, we can reduce time for some operations and improve efficiency in many other important aspects of the healthcare organization. In my dashboard in the next chapter, you will see how I implemented not only patient data but also management data to simplify both processes for GIMEMA and increase efficiency in every aspect of the organization.

The next example is more focused on patients' data and how a dashboard with some machine Learning can help a hospital provide a real-time display of the anticipated disease progression and response to treatment for children diagnosed with Crohn's disease. The paper "*Real-time Tool to Display the Predicted Disease Course and Treatment Response for Children with Crohn's Disease"* (Corey A. Siegel et al. 2010) although being an older paper, shows a real-time tool that can help predict and communicate individualized risks of Crohn's disease complications and how treatment may modify those risks.

Biologic agents and immunomodulators can be used to treat Crohn's disease in children. However one of the biggest challenges in using these agents is explaining to patients and their families how the disease is projected to progress both with and without treatment. To develop a visualization tool that forecasts and conveys individual risks of complications from Crohn's disease as well as how therapy may alter those risks, this research makes use of system dynamics analysis (SDA). System dynamics analysis (SDA) is an approach that may produce straightforward graphs to illustrate anticipated outcomes across time and can offer real-time, personalized outcome forecasts. A model to forecast the unique risk of complex Crohn's disease and how treatment affects this risk was created using SDA. Providers can better explain disease outcomes and treatment options to patients and their families by using a graphic display of these results. Human-centered development of interactive information visualization in the emergency room that is integrated with the electronic health record employing resources for quick healthcare interoperability.

Healthcare has been changed by the creation of electronic health records (EHRs), yet their accessibility and potential to contribute to physician burnout have become serious concerns. The authors of this study created an asthma timeline application that used contemporary health data standards like FHIR (Fast Healthcare Interoperability Resources) by utilizing human-centered design (HCD) principles. The chart below (Fig. 6) is the part of the dashboard that has been created to track asthma and different possible risks in the patients based on some complications.

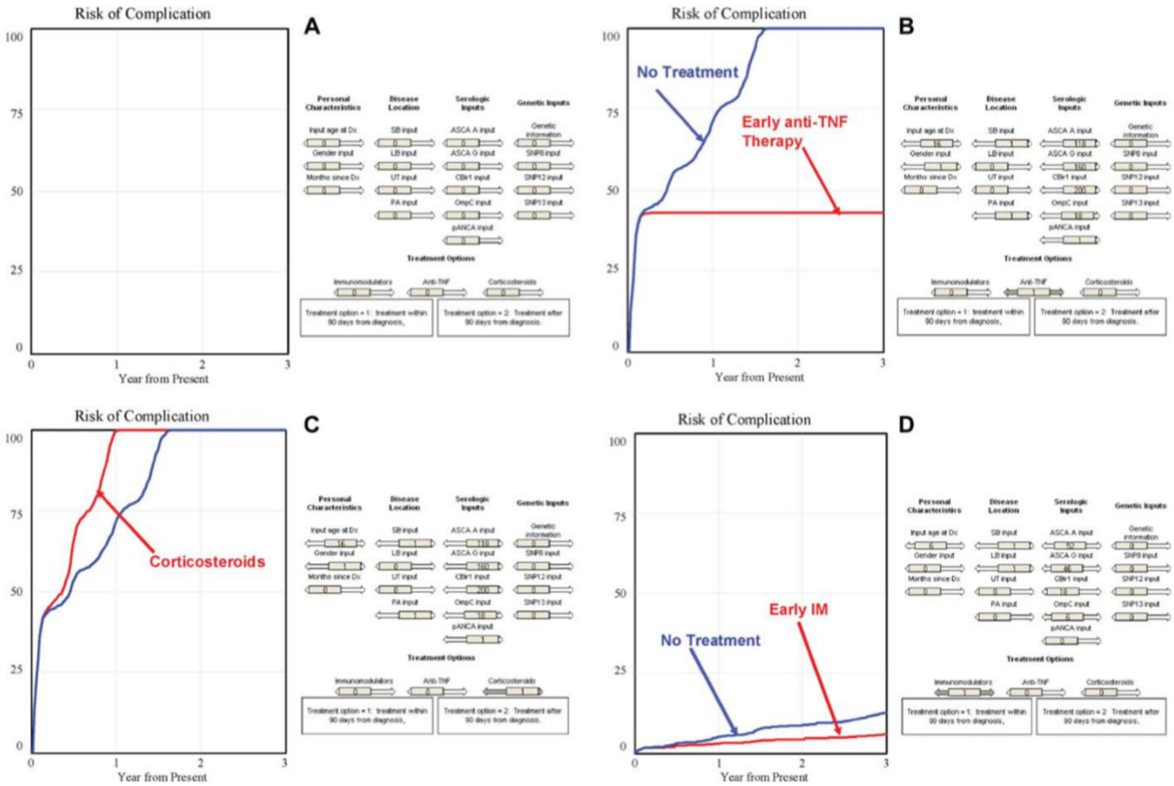


Figure 6, Real-time Tool to Display the Predicted Disease Course and Treatment Response for Children with Crohn's Disease" (Corey A. Siegel et al. 2010)

A standard called FHIR makes it possible for data to be effortlessly transferred across various healthcare systems when exchanging healthcare information electronically. The application's usage of FHIR made it possible to integrate data from the EHR (Electronic Health Record) system, which is sometimes difficult to use and divided into silos. The Asthma Timeline Application can now be implemented in organizations other than the one where the study was done thanks to FHIR-based web services. The paper also emphasizes how FHIR might facilitate application sharing between several EHR platforms without requiring vendor-specific code, which could be advantageous. When compared to standard EHR functionality, the Asthma Timeline Application's usability testing showed a considerable decrease in doctors' time on task and cognitive load as well as an improvement in the accuracy of information retrieval for specific asthma-related data points. Positive feedback regarding the visualization was also found in a post-implementation survey. This visualization can help identify problems and provide a quick overview of the patient's history.

In the next chapter of my thesis, I will go through the creation of a dashboard using the R programming language and a very famous library that aids the development of such dashboards, which is Shiny. I will delve into some of the basic mechanics of dashboard creation and some fundamentals step that I had to take when creating it. Moreover, I will focus on the challenges faced and improvements that I believe the dashboard has brought the organization.

3. Creating a Dashboard: GIMEMA's Dashboard

Before diving into the details of dashboard creation, in this chapter I will give a presentation of GIMEMA, to understand the sector in which it operates and why they decided to use a Dashboard to solve some of their daily problems.

3.1 GIMEMA

I worked as an intern for two months in May 2023 at GIMEMA (Gruppo Italiano Malattie EMatologiche dell'Adulto), an Italian Onlus foundation that was established in 1982 by Professor Franco Mandelli and is responsible for promoting, supervising, and organizing various independent clinical research protocols related to blood diseases and hematology. With the construction of the Data Centre in 1985, a central office devoted to the operational and scientific oversight of all the Foundation's various activities was established. The Foundation has worked with numerous national and international research groups over the years. Specifically, the first experimental procedures with the European Organisation for Research and Treatment of Cancer (EORTC) Leukaemia Group were initiated in 1988. Among its Turning Points, Gimema was one of the first three autonomous research structures to receive the ISO 9001-2000 accreditation in 2002. This certification is a worldwide recognized standard for quality management. As of right now, the foundation is a highly integrated reality that oversees a vast national network in which the majority of Italian hematological centers take part. In addition, it oversees numerous global partnerships. My internship was at a data center, which is responsible for carrying out clinical research projects. As such, it serves as the hub for all research, which is then analyzed. Because of its management and experimental activities in the field, the data center has achieved national and international recognition in the clinical sector.

Gimema deals with most of the blood diseases that we know today, constantly trying to reduce both the number of deaths and improving the quality of life of all the patients. Their contribution to medical research has brought some great improvements in the field, such as:

- Chronic Myeloid Leukemia: The most common form of leukemia; thanks to the introduction of new therapies, and the studies conducted by GIMEMA, the 10-year survival rate has increased from 20% to 80% with a growing patient's expectation.
- Acute Myeloid Leukemia: In the early 1990s, the role of transplant procedures in treating Adult Acute Myeloid Leukemia (AML) was unclear. However, a collaboration between GIMEMA and EORTC provided clarity, with bone marrow

transplantation established as a cornerstone therapy for AML. Subsequent studies, including AML10 and AML12, demonstrated the significance of cytogenetic alterations, the use of peripheral blood stem cells for transplantation, and the effectiveness of high-dose drugs in improving chemotherapy response rates. Additionally, genetic modifications influencing therapy sensitivity were identified. These advancements nearly doubled the number of successfully treated AML patients.

- Acute Promyelocytic Leukemia: Over the past two decades, GIMEMA's contributions have revolutionized the treatment landscape for Acute Promyelocytic Leukemia (APL), also known as LAP. Initially treated with a combination of chemotherapy and retinoic acid, recent breakthroughs involving arsenic and retinoic acid have further improved outcomes. GIMEMA-led studies have redefined standard therapy, minimized chemotherapy side effects, and achieved a remarkable 100% positive response rate in patients. Consequently, APL, once considered highly lethal, is now regarded as the most frequently curable form of acute leukemia, thanks to GIMEMA's transformative efforts.
- Acute Lymphoblastic Leukemia Ph+: Recent years have seen significant progress in treating Acute Lymphoid Leukemia (ALL), particularly the Philadelphia-positive subtype, thanks to studies led by the GIMEMA Foundation (LAL 0201 and LAL 1205). The use of new drugs has resulted in nearly complete disease remission in most treated patients. Despite the persistently high incidence of relapses, these findings indicate positive strides forward in research.

The Data Center is the operative center that coordinates all of the research activities of the Foundation, ensuring that all the activities are conducted following the international norms of good clinical practice. The operational structure of the GIMEMA Data Center comprises four main working units: management, design and analysis, clinical trial management, and secretariat. Since 2015 the center has been one of 5 Clinical Trial Units that have been certificated, by ECRIN (European Clinical Research Infrastructure Network), at the European level for adhering to the strictest standards of clinical research. The main activities that are conducted in the center are:

- Design and organization of clinical studies, from conception to protocol drafting, and management of regulatory affairs and communications with Ethics Committees and competent authorities.

- Conducting clinical trials, coordinating participating centers, and managing data collection;
- Pharmacovigilance, statistical analysis, and publication of reports and scientific papers;
- Provision of training courses and scientific outreach activities.
- During my internship, I collaborated in most of the four sectors, by creating a Dashboard that not only helps coordinate all the participating centers and manage data collection but also allows all the people in the foundation to have a quick view of some statistical analysis and their publications.

3.2 Designing the Dashboard

Understanding GIMEMA's demands was the first step in creating the dashboard. Many people may believe that this is a simple process, but in actuality, it is the most difficult and time-consuming since you have to comprehend not only what they require, but also how to develop a visualization that may help close any gaps in their knowledge without being deceptive or raising additional questions. The primary assignment I was given was to redesign their current data visualization method, making it far more user-friendly and enabling people to engage with their data in one convenient location.

In the beginning, their visualizations were made up of numerous PDFs that they manually created each day containing all the information they would require. This made them inefficient in terms of updates as well as visualization since they had to update the website and PDF for each need. Additionally, they had a very basic PowerBI, an interactive data visualization software, with a few bar plots and line charts for very basic data visualization (http://powerbi.microsoft.com/). My job was to take all of their resources and combine them into a dashboard so that the various visualizations could be combined and enhanced in terms of interactivity.

The R-Shiny Library and the R programming language were used to construct the dashboard from the ground up. Since R powers all their systems, I wanted to stay true to what they already have to make the transition process go as smoothly as possible and avoid any potential misconceptions.

The Shiny Library documentation from Posit's official website (www.posit.co), which produces open-source software for data scientists like RStudio and many more libraries to enrich and improve their employment, was the first step in the dashboard's design process.

There is a section on the Shiny Website that details every action a designer needs to take to create a Dashboard using the Shiny Library. This module, which integrates R and Python with Bootstrap (a free and open-source CSS framework aimed at front-end web development that prioritizes mobile responsiveness) is accessible for both languages. In this manner, the designer can develop the dashboard's overlay and structure using standard HTML frameworks, he can also analyze data, and create different visualizations using the programming language of choice. Since I had never used HTML or CSS previously, this integration posed a significant difficulty for me. However, with the help of the numerous available resources, I was able to quickly resolve this problem and produce a fully functioning dashboard from scratch.

Recognizing the kind of data I was dealing with was a crucial step in the dashboard's design. My main responsibility was to build a dashboard based on their data structure while keeping the functionality largely intact so as not to interfere with the operation of many other systems within the company. This step was challenging for me mostly because it was my first time working with a real database with so much data. In addition, their database was set up so that anyone with access could alter the name or remove variables, changing the database's overall structure. This was a known problem inside the workspace, but GIMEMA was unable to handle it because, as an ONLUS, they are unable to spend money on anything other than research, and thus they are forced to manage their data using free technologies and open source softwares. It was discovered that the best way to solve this issue was to slice a data frame slightly by always using names rather than indexes.

The database is a relational database built on MariaDB that enables access to and modification of the data using standard MySQL commands and syntax. The initial phase in comprehending their structure was examining each of their numerous data frames by executing them in R and gaining a knowledge of the various variables and information contained there. Additionally, I connected to the database and quickly did a database maintenance task using an application called "TablePlus."

Here, there were primarily two challenges. I had to learn how to access the database first, and then I had to learn how to merge the various tables so that I could tell which variables were main keys and which weren't. With a Patient ID, Hospital ID, or Doctor ID, depending on the scope of the database, and numerous other variables linked to that identity that served as the main key, the tables were organized in an extremely simple manner. One of the first things I did was clean the tables by deleting all empty columns and all duplicate columns with different names because certain variables were empty or duplicated. An important attribute in the table was the PID; this variable is a primary key that identified Studies that had some important information in the Database from those that did not have that information. In simple terms, it means that for some studies we have more information than for other studies.

3.3 Creating a Dashboard in R

I will walk through the complete process of making a dashboard with the R programming language in this section of my thesis. The chapter will be divided into three primary subsections, each of which will focus on a distinct dashboard component. The generation of the data frames, which are subsequently read by the dashboard and written on each page, will take up the first section. The UI and the Server, which are a dashboard's two primary components, will be covered in detail in the second and third steps. These two dashboard components are distinct but closely related; we shall see later why developing code for both components needs to be cohesive and what each component performs specifically.

3.3.1 Creating the data tables

This section will go over how to construct and access charts and data frames in an R environment for the dashboard. Tables and charts provide the foundation of dashboards, but to make them, you must first retrieve the data from a database and modify it to provide the optimal user experience. The Gimema Database contained five main tables that I had to work with: one that listed the centers, one that listed the studies, one that listed the status of each study, one that listed the patient and physician records, and a final table that linked the status of the study in a particular center that was conducting that hematological research.

In my example I have separated the tables and charts based on what they were trying to show, more specifically I had to show four main aspects in the dashboard:

1. The information on all the research
2. Information only on open research
3. Information on centers of research (centers are hospitals or clinics pursuing the research)
4. Information on patients (how many patients in each specific research, type of disease, patients in each center...).

Most of the effort in this stage of the dashboard's construction was done to figure out how to shape the data to provide a deeper analysis of their data by attempting to both produce something novel that could surpass their expectations and display what the firm needed. The biggest obstacle was figuring out where the information was in the database, how to get to it, and, if the information I wanted was not present in the database, how to create it using the data I already had. As an example consider a chart showing the patients managed by each center per year and showing both an individual and a cumulative view year by year. Managed patients are those that are still under research, meaning that they are in a center that still has an open study. This information was not in the database so I had to create it by hand. I dedicated to this step an entire R script that first found the number of enlisted patients in each center per year that are unique (no repetitions of patient ID) and then when this number was created I had to merge this data frame with a table having all the information on the research like ‘starting date’ and ‘closing date’. Then for each center, I dropped studies that were closed, subtracted the number of patients that were in each closed study from the total enlisted studies, and found the total number of managed patients in each center. Then I added to this dataframe also the information on the type of study that is being researched. The chart below (Fig 7.) is an example of output for the center of Bologna.

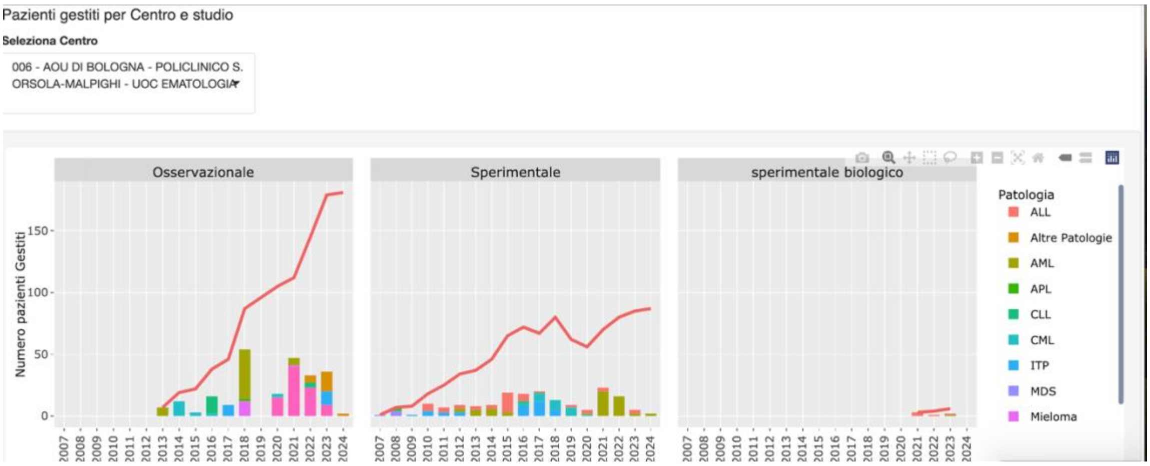


Figure 7, Managed patients for the Bologna Centre.

As you can see, on the top left part of the table you can choose with a dropdown menu the center that you want to look at, then the charts appear for that specific center. There are three different charts, one for each type of study ‘Osservazionale’ (observational), ‘Sperimentale’ (Experimental), and ‘Sperimentale biologico’ (Biological Experimental). In each chart you have two pieces of information, the patients managed year by year with the barplot, that are colored based on the pathology of the patient. Moreover, you can see a red line showing the number of patients in a cumulative way in the center. For example, this means that in 2019 the center of Bologna managed 50 new observational patients, the majority being AML (Acute Myeloid Leukemia) and in total, it had around 90 patients that were managed in their center. On the experimental side, you can see how there are drops in the cumulative sum. This is because patients that are of research that is not being studied anymore are removed from the count and thus the cumulative sum of managed patients decreases like in the years 2019-2020 where we go from a peak of 70 managed patients in 2018 to around 50 in 2020. The charts are all dynamic meaning that the user can choose on the right a specific pathology and it will be shown only the chosen pathology in the charts. The user can hover with the mouse on each column and get the exact number of managed patients.

The charts were created thanks to the data frame built in the first part of the dashboard using R and SQL, as can be seem from a snapshot of the code that found below (Fig 8.).

```
  summarise(arr = sum(arr),
            .groups = 'drop')%>% drop_na()





gest_centro = left_join(gest_centro, centri_sottrai, by=c('Anno'='anno_chiusura', 'linktitle', 'tipo_sperimentazione'))
gest_centro[is.na(gest_centro)] <- 0

gest_centro$cs <- with(gest_centro, ave(arr + sottrai, tipo_sperimentazione,linktitle, FUN = cumsum))


gest_centro = gest_centro[,c(1,2,3,6)]


studi_centro_anno = arr_anno[,c(4, 8,9,12,14)]

studi_centro_anno= studi_centro_anno%>% group_by(Anno,linktitle, tipo_sperimentazione, patologia) %>%
  summarise(arr = sum(arr),
            .groups = 'drop')%>% drop_na()




gestiti_centro_studio = left_join(gest_centro, studi_centro_anno, by=c('linktitle', 'Anno',"tipo_sperimentazione"))





gest_centri = unique(test$linktitle)

```

Figure 8, Code representing the creation of a data table used for managed patients

The code that you can find above (Fig 8.) is used to merge the managed patientis in each center with those that are not being managed anymore so to obtain the exact number per year and then by grouping per year, pathology, type of experimentation, and ‘link title’ (name of the center) I created the grouped information for each center per year presented in Figure 7. The charts on the other hand are not created in the initial R script but are created in the UI/server part of the dashboard since they are an object that is shown thus instead of creating it somewhere else and then accessing it, you first create the visual object (charts or tables) in the UI and then feed the data to that object (data frame) in the server side of the dashboard. In the next subchapter, I will go through the UI so to show how each visual object is created. On the server side instead, you program to tell each object which exact information must be shown.

3.3.2 The UI of the dashboard

In this section of the paragraph, I will go through the creation of the UI (User Interface) of the Dashboard. The UI is the point of interaction between the user of a service and the programmer of such service. The UI is very useful for programmers and designers to transmit to the users of their services easily and visually, with the use of data tables, charts, and more, all the information that they analyzed and created.

In the UI of a Dashboard, the designer needs to create all the interactive buttons and elements that change both the dashboard view and the different charts so to allow the user to interact with them. For example, on the main page of my dashboard (Fig.9) many different interactive buttons and elements allow the user to move around the dashboard in a very efficient way.

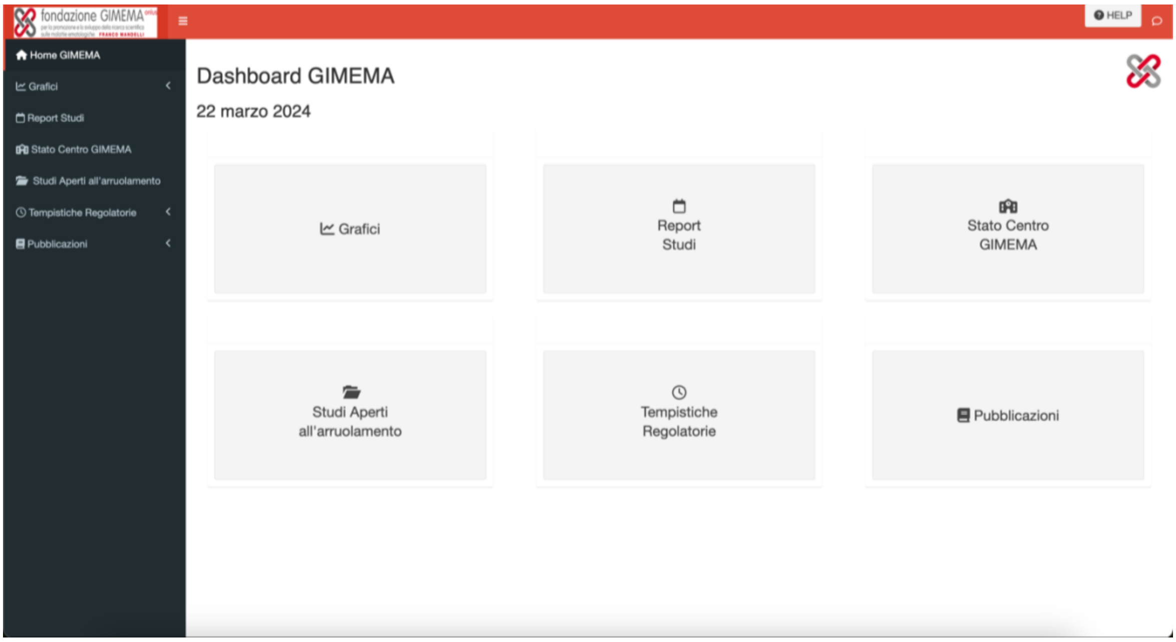


Figure 9, Home section of my Dashboard.

In this main tab (Fig.9), the user can move around the dashboard quickly by clicking the button with the corresponding section in which he wants to move. For example, if the user needs to see some charts he clicks on the ‘Grafici’ button that will move him to the home section for Grafici, the page dedicated entirely to charts. Similarly for all the other buttons, the Dashboard view will change based on what the user clicks. Another interaction that can be done in the home section, but also in any other view of the dashboard is the ‘Help’ button and the ‘chat’. These two buttons allow the user to request help by sending an e-mail directly to the system administrator with the chat . with the help button the user can have a quick view of what each symbol represents.(Fig. 10)

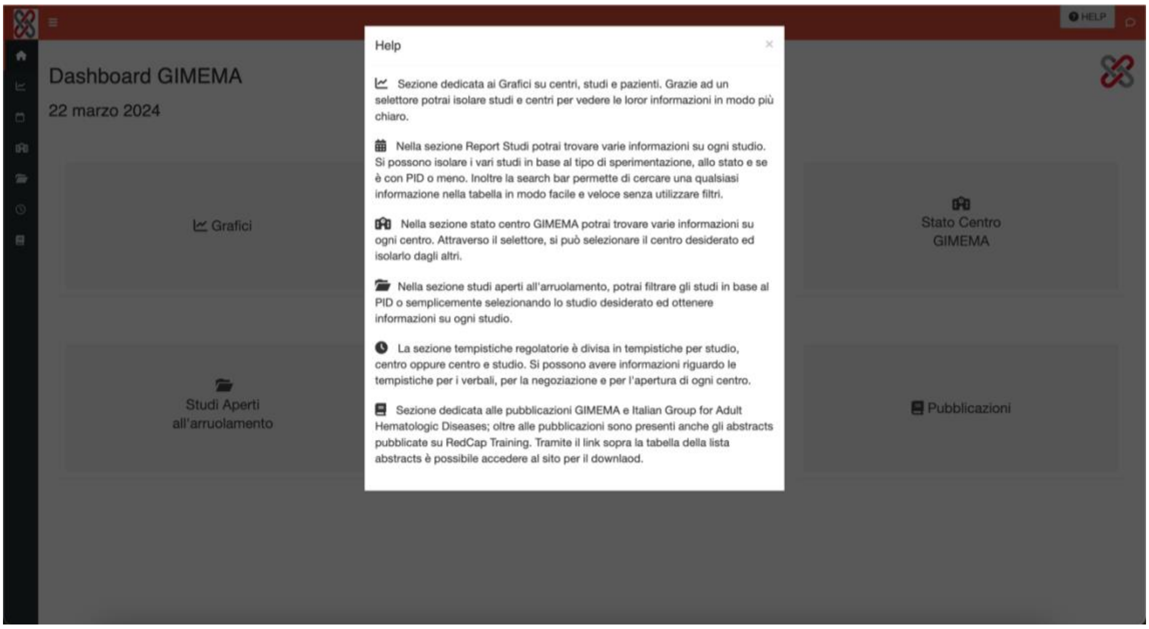


Figure 10, Help button

Other interactions refer to the selection of different views in the charts and tables. For example, in the chart shown before of the ‘Gestiti annui’ I used a select input to allow the user to change the Centre for which they want the information. These buttons and inputs are present next to each chart and data table in the dashboard, as you can see below with another example of a chart present in the dashboard (Fig. 11).

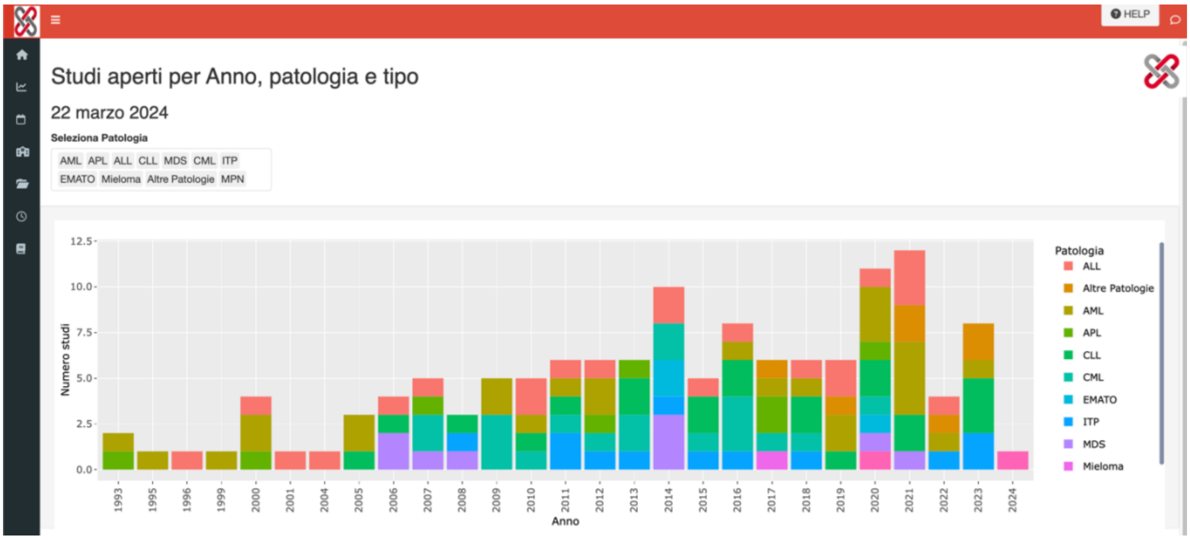


Figure 11, Sidebar with buttons for moving around the dashboard

The box called ‘Seleziona Patologia’ allows the user to select the pathology for which they want to see all the open studies year by year. Moreover, thanks to ggplotly (a library in R that allows the creation of interactive charts) the user can also directly interact with the Chart by selecting the Pathology directly on the chart or zooming in or out to have a closer view of some specific years.

All these interactions were created in the UI section of my code with the use of basic HTML and CSS for a more personalized view of the different buttons and objects present in the Dashboard.

A well-designed dashboard streamlines the complex data into digestible visualizations, allowing users to quickly grasp trends, patterns, and outliers. By presenting information clearly and intuitively, we enable users to make informed decisions swiftly. This not only enhances productivity but also fosters a deeper understanding of the data, leading to more effective strategies and actions. Moreover, a thoughtfully crafted dashboard can facilitate collaboration and communication among stakeholders, ensuring everyone is on the same page and working towards common goals. Ultimately, the goal of a well-designed dashboard is to empower users with actionable insights that drive success and innovation.

On the other hand, creating a well-structured UI is not an easy task since it requires a mix of knowledge that spans from basic programming to a high level of proficiency in CSS and HTML. As a data scientist, my expertise primarily lies in data analysis, machine learning, and statistical modeling rather than front-end web development.

Initially, I spent a significant amount of time familiarizing myself with the fundamentals of HTML and CSS. I started with online tutorials and courses, gradually progressing from basic syntax to more advanced concepts such as layout design, responsive design, and CSS animations. It was a steep learning curve, but I was determined to acquire the necessary skills to bring my vision to life.

As I began building the UI for my dashboard, I encountered numerous challenges along the way. From aligning elements properly to ensuring consistency in design across different screen sizes, each step presented its own set of hurdles. However, with perseverance and continuous learning, I gradually gained confidence in my ability to create and adjust UI elements effectively.

One of the most valuable lessons I learned throughout this process was the importance of iteration. As I iteratively refined the design of my dashboard based on user feedback and usability testing, I discovered new techniques and best practices that further enhanced its intuitiveness and efficiency.

3.3.3 The Server of the Dashboard

The server side of any Dashboard is the place where all the modifications and changes happen. The server renders all the objects I created on the UI side of the dashboard and makes them interactable by capturing the choices made by the User and modifying the data tables or the charts based on those choices by using *reactive tables* and *reactive charts*. Both capture the tables from the data frames created in the initial phase of the dashboard and then capture the choice of the user from the selection made through the use of the UI and lastly modify such data frames. If the data frame is needed for a table it is then passed into any data table and visualized directly on the dashboard. If instead it is used for a chart, you pass it as data to a dynamic *ggplot* chart and render it with a library called '*Plotly'* that is used to modify the charts based on the variable chosen by the user.

In conclusion, while the UI is used to make the dashboard as customizable as possible, the Server side is used to make it interactive and dynamic by combining the UI and the Data frame section.

Below you can find a snapshot of the code of how a reactive (dynamic and interactive) table is created and how it changes based on the given input. (Fig. 12)

```
})

tot_patology <- reactive({  #bar dat è uasta come tabella per disegnare il barplot su stato per malattia
  tot_by_patology %>%
    filter_at(vars("stato_studio"), all_vars(. %in% input$stato)) #filtra la tabella in base allo stato selezionato
})
```

Figure 12, snapshot of how an interactive and dynamic table is created in R

In this case, the table created is used to create a chart for the number of patients based on the state of the study. The way this works is that with the reactive({}) function we tell the dashboard that the data frame '*tot_by_patology'* must be dynamic and it should change based on the input given in the variable 'stato_studio' in the button (or more generally input) called '$stato'.

This type of code must be created for any chart or table that we have in the dashboard to make them as interactive as possible.

After seeing the different practical parts of any dashboard we will now go through all the different sections of my dashboard exploring all the different information that can be extracted and analyzed.

3.4 The Dashboard sections

We will now go through every section of the dashboard, beginning with the home page. We will examine all the tables and charts there, and I will walk you through each one step-by-step, explaining how the user can interact with them, what they can do, and why I believe it was crucial to put that information into practice. I will also discuss the difficulties encountered and how I overcame them in the process of creating some of those visual items.

3.4.1 The Home Page

On the home page of my Dashboard, I wanted to put all the different sections that the users can find while exploring it. The user experience is mainly dependent on how they approach for the first time with the Dashboard, so a good Home Page makes the entire experience better and more enjoyable.

On this page, (Fig. 13) you can find a single white canvas with 6 different buttons each one redirecting the user to a different section of the Dashboard. The 6 sections are: ‘Grafici’ (*Charts)*, ‘Report studi’ (*Study Reports*), ‘Stato Centro GIMEMA’ (*State of GIMEMA’s center)*, ‘Studi aperti all’arruolamento’ (*Studies open to enrollment*), ‘Tempistiche Regolatorie’ (*Regulatory Timing*) and ‘Pubblicazioni’ (*Pubblications*).

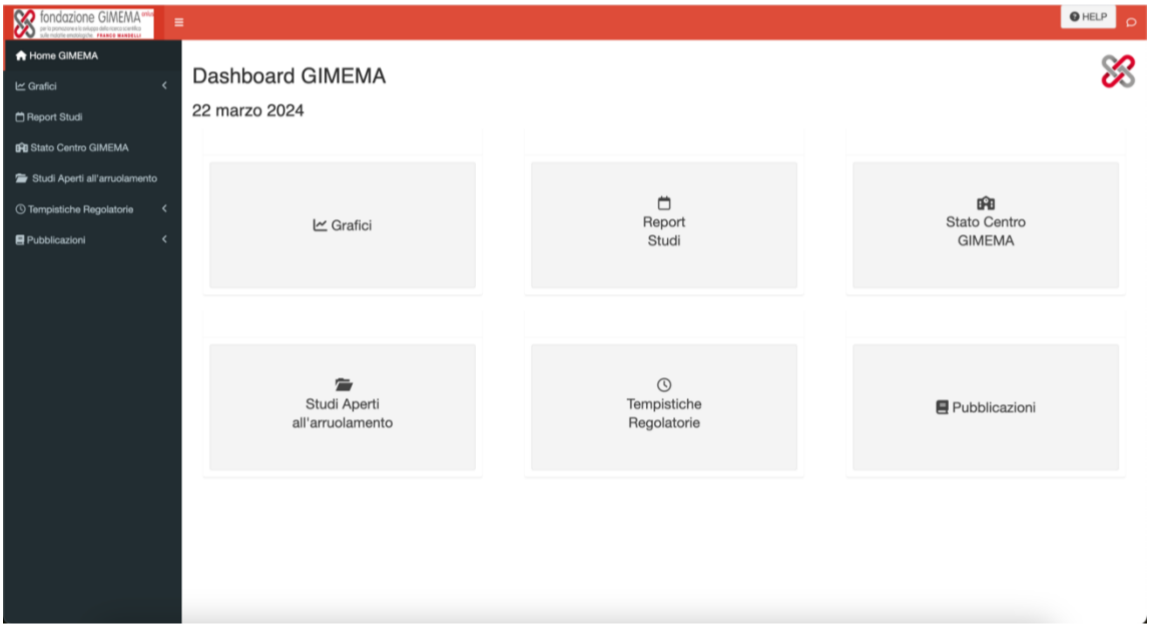


Figure 13, Home section of the dashboard

By clicking any of the buttons the users are redirected to that section of the dashboard. If the section is then subdivided into other parts, the buttons redirect to a new Home Page of that section. For example, the ‘Grafici’ section of the Dashboard has many charts in it so the users will not go directly to a specific chart, but they will be redirected to the ‘Grafici Home’ (Fig.14) that has other buttons that will then display the specific charts requested.

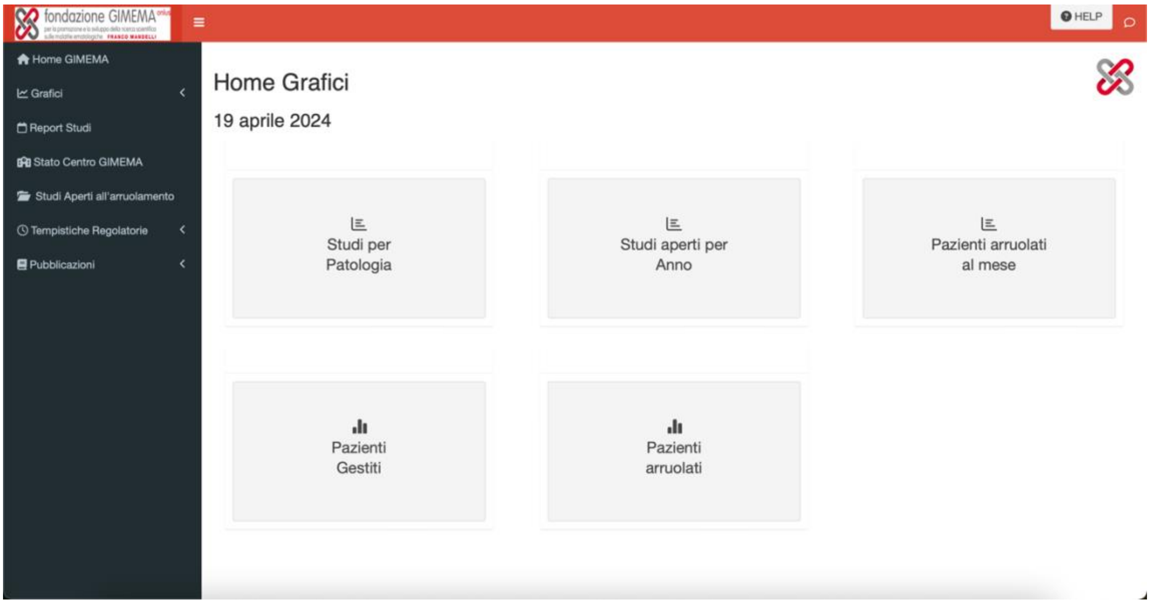


Figure 14, home section dedicated to charts

On the Home Page, but also in any other section of the dashboard, the users can find two main objects, the Sidebar with some shortcuts to the different parts of the dashboard and the Header that has 3 main features: on the left, there is a link that redirects to Gimema’s website, on the right there are two buttons, the Help button that displays some quick guide for the users and a chat button that allows the users to directly send the administrator an E-mail with the requested help.

3.4.2 Charts

In the charts section of my Dashboard, there are plenty of visualizations that are fundamental for studying and analyzing the current situation in any study or center managed by GIMEMA. I wanted to focus on three different things, simplicity, efficiency, and effectiveness. With these three main pillars in mind, I wanted to create charts that were able to give a quick look but very powerful to the information that together with my colleagues believed was crucial for them. With my knowledge of Data visualization, I wanted to give them as many visualizations as possible but also keep them simple and interactive to not create confusion and misunderstandings when using them.

As you have seen before (Fig.14) the Chart section has its dedicated home section that allows the users to immediately go where they need to. Five main charts can be found in my dashboard: ‘Studi per patologia’ (*Studies by pathology)*, ‘Studi aperti per anno’ (*Open studies per year)*, ‘Pazienti arruolati al mese’(*Enrolled patients by month)*, ‘Pazienti Gestiti’(*Managed Patients)* and ‘Pazienti arruolati’(*Enrolled Patients).*

In the section ‘Studi per Patologia’ the users can find two main charts, the first one (Fig.15), is a barplot, showing the number of patients per pathology, with the bars that have been color-coded based on the state of the study (‘Open’, ‘Closed’, ‘New Proposal’…). In this first chart, the user can interact with it by selecting the states that he wants to look at or selecting all of them if he wants to have a view of the overall situation thanks to the selection box in the top left corner.

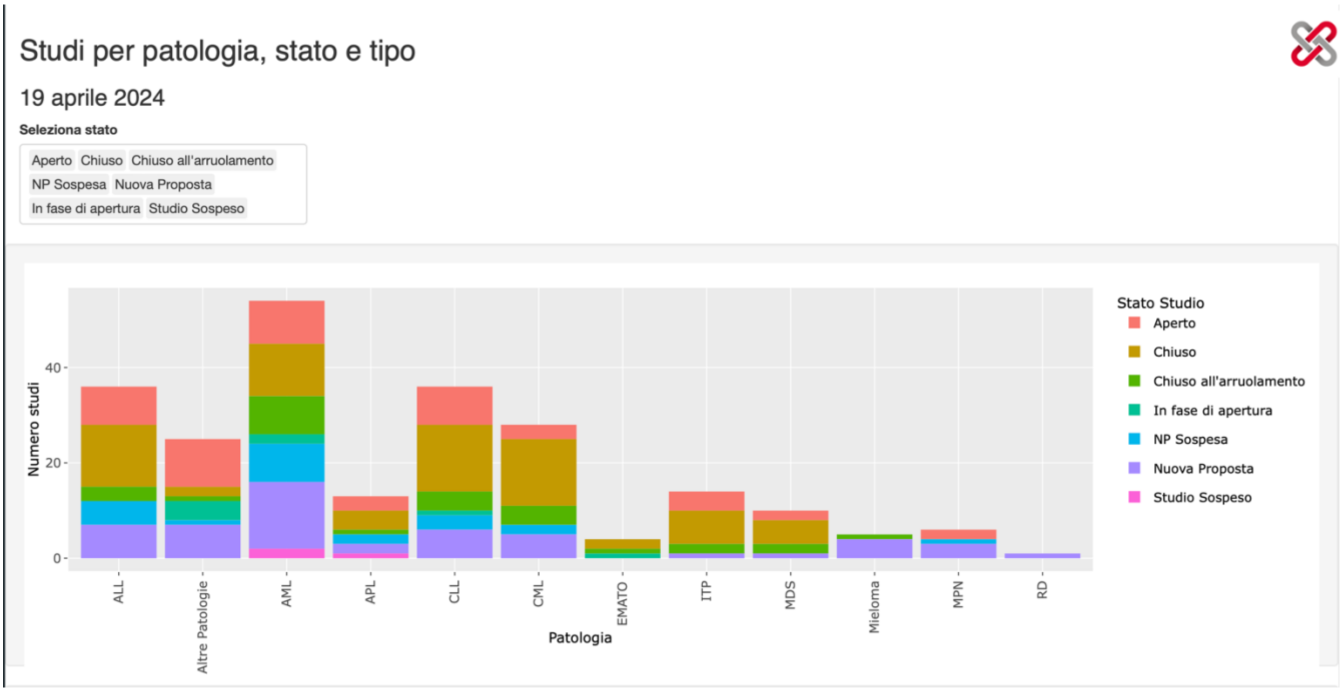


Figure 15, studies for pathology and state

The second one (Fig.16) is instead a differentiation of studies based on the typology, which are ‘Observational’, ‘Experimental’, and ‘Biological Experimental’. This distinction allows the user to better understand how the studies are distributed and not only the state in which it is in but also the type of study that has been brought forward and the number of patients. I wanted to create two separate charts since I believe that is always

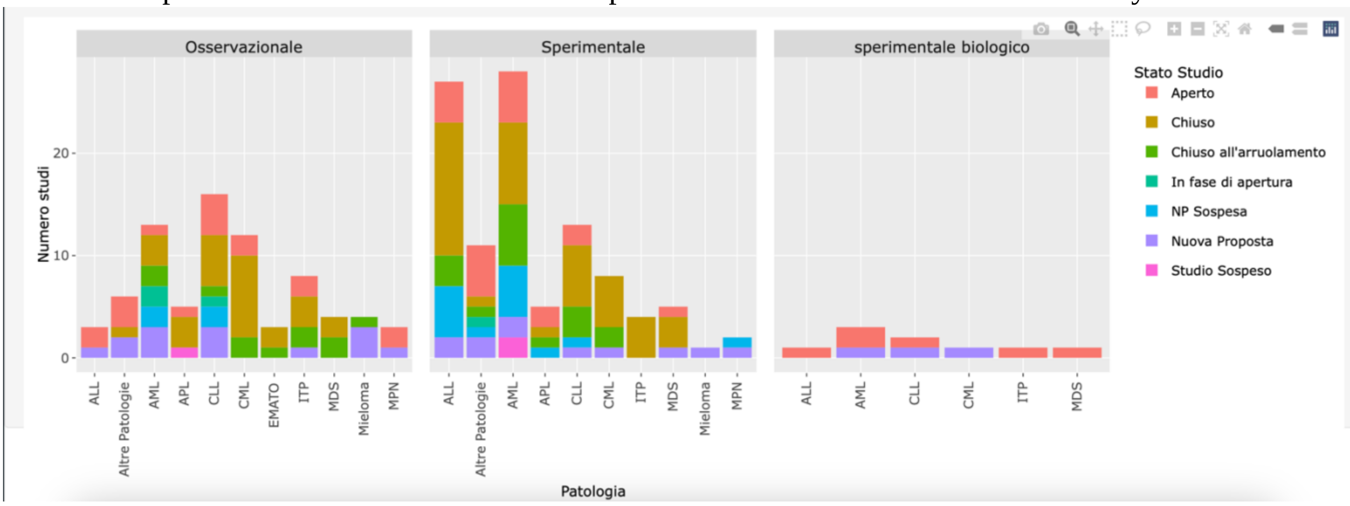


Figure 16, charts differentiating the studies based on typology good to have first a Macro view of the data and then dig down on the more specific sections by augmenting the dimensionality of data with a new variable.

In the section ‘Studi Aperti per anno’ the user can find a similar view of Fig.15 and Fig.16 with the difference that now it also shows the users the year in which the study was opened. Similarly to the other section I also divided these charts in two spaces one with an overall view (Fig. 17) and the other with studies divided by type of analysis (Fig. 18).

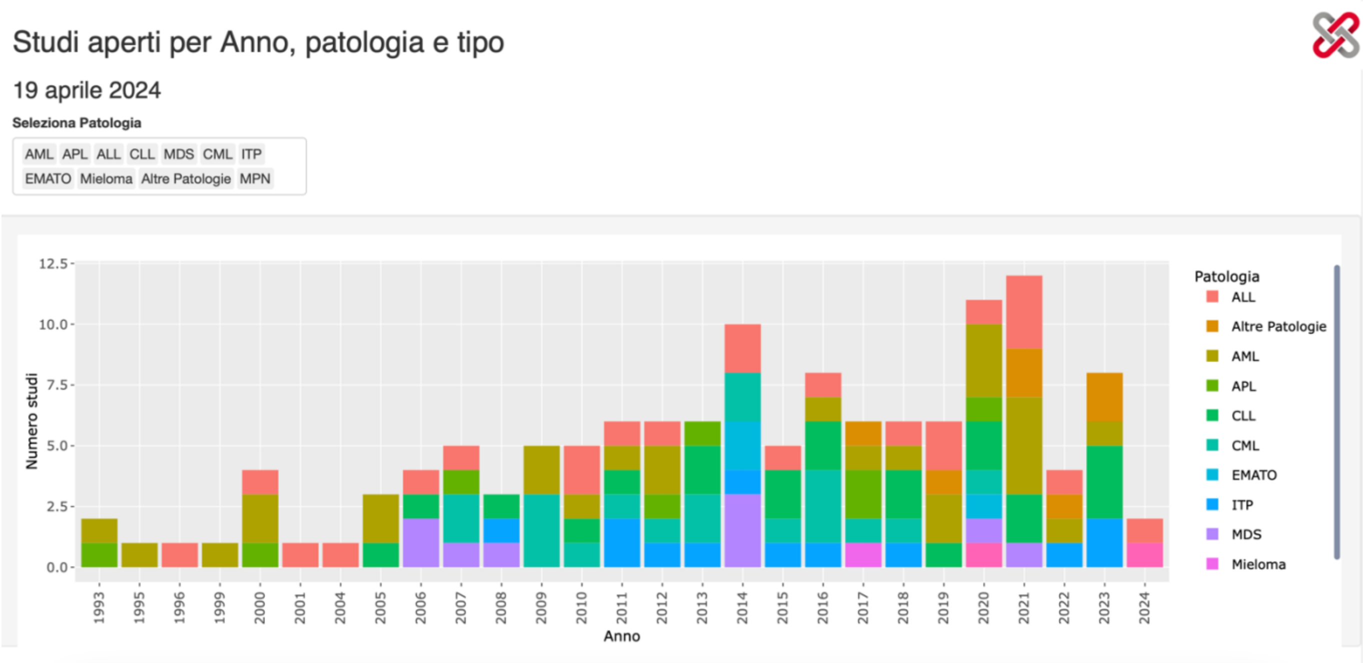


Figure 17, Open studies per Year with differentiation based on pathology

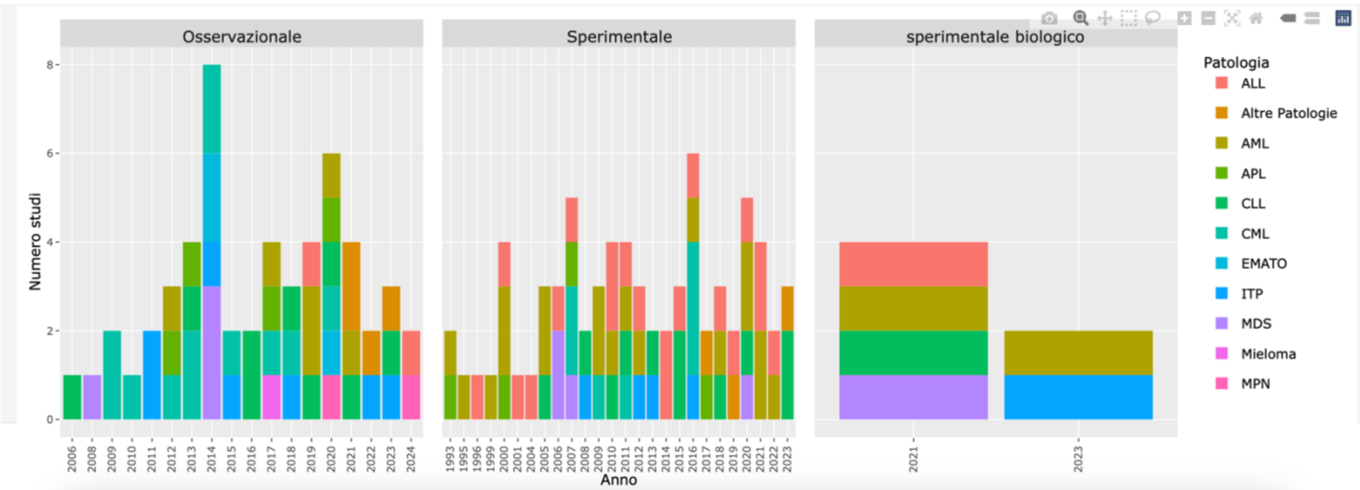
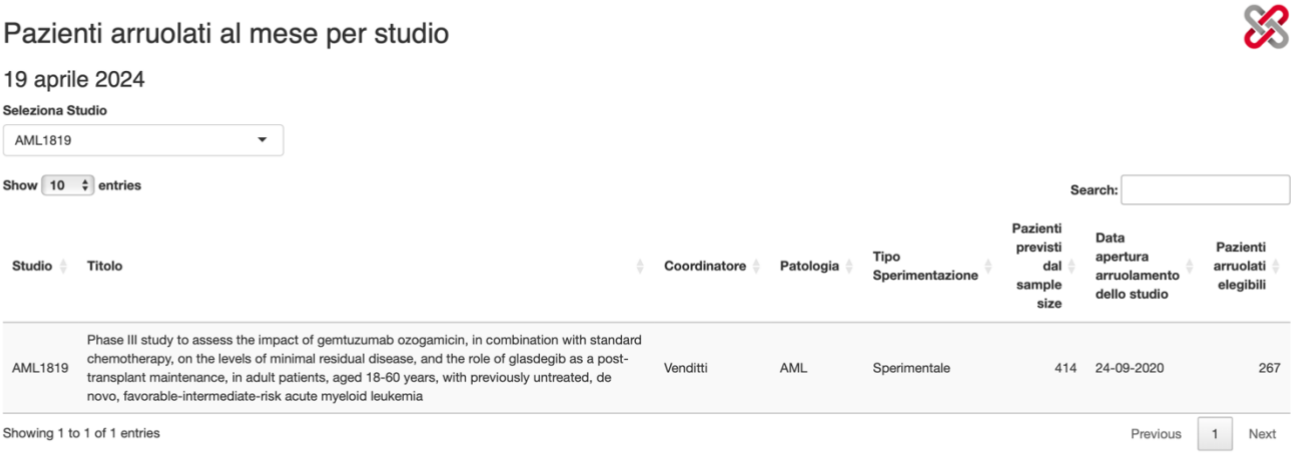


Figure 18, open studies with differentiation based on pathology and type of study

In the section dedicated to ‘Pazienti arruolati al mese’ the user can find a new view, where he can find two main things. First, on the top left corner he has a selection menu where he can choose the study he wants to look at, then below two main things appear, a Table (Fig. 19) with the information regarding that study and the number of enrolled patients and the expected number of patients enrolled, and a chart (Fig.20) showing month by month the number of patients that have been enrolled in that study, also he can find a cumulative look at the same chart.


Pazienti arruolati al mese per studio

19 aprile 2024

Seleziona Studio: AML1819

Show 10 entries

Search:

Studio	Titolo	Coordinatore	Patologia	Tipo Sperimentazione	Pazienti previsti dal sample size	Data apertura arruolamento dello studio	Pazienti arruolati elegibili
AML1819	Phase III study to assess the impact of gemtuzumab ozogamicin, in combination with standard chemotherapy, on the levels of minimal residual disease, and the role of glasdegib as a post-transplant maintenance, in adult patients, aged 18-60 years, with previously untreated, de novo, favorable-intermediate-risk acute myeloid leukemia	Venditti	AML	Sperimentale	414	24-09-2020	267

Showing 1 to 1 of 1 entries

Previous 1 Next

Figure 19, enrolled patients table

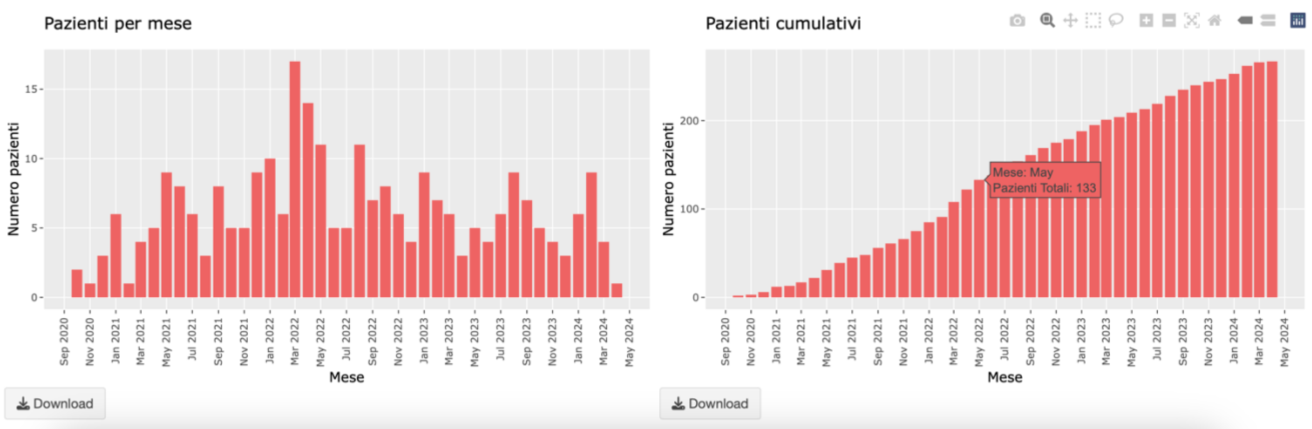


Figure 20, enrolled patients per month, chart

The last two sections ‘Pazienti Gestiti’ and ‘Pazienti Arruliolati’ have been constructed in the same way, with both having two main charts one with the Data year by year and the other with the distinction between type of study. The first section is dedicated to managed patients and we have seen it before in my thesis (Fig. 7); the other section is instead again dedicated to enrolled patients but this time it's not monthly based but Yearly, moreover it contains all the typologies of studies in the same chart (Fig. 21)

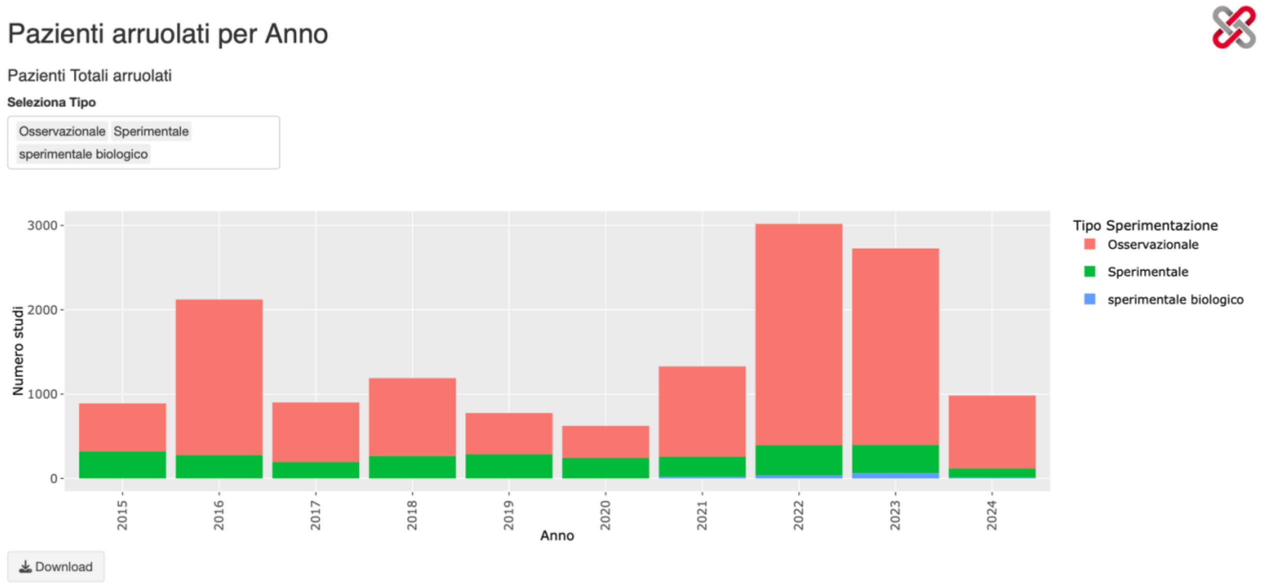


Figure 21, enrolled patients per year and type

The second chart (Fig.22) dedicated to yearly enrolled patients also contains a box that allows the user to choose the center that they want to look at, the barplot is also split based on the type of study as we have seen before (Fig.16). These charts conclude the section and we will now move to the other parts of my Dashboard.

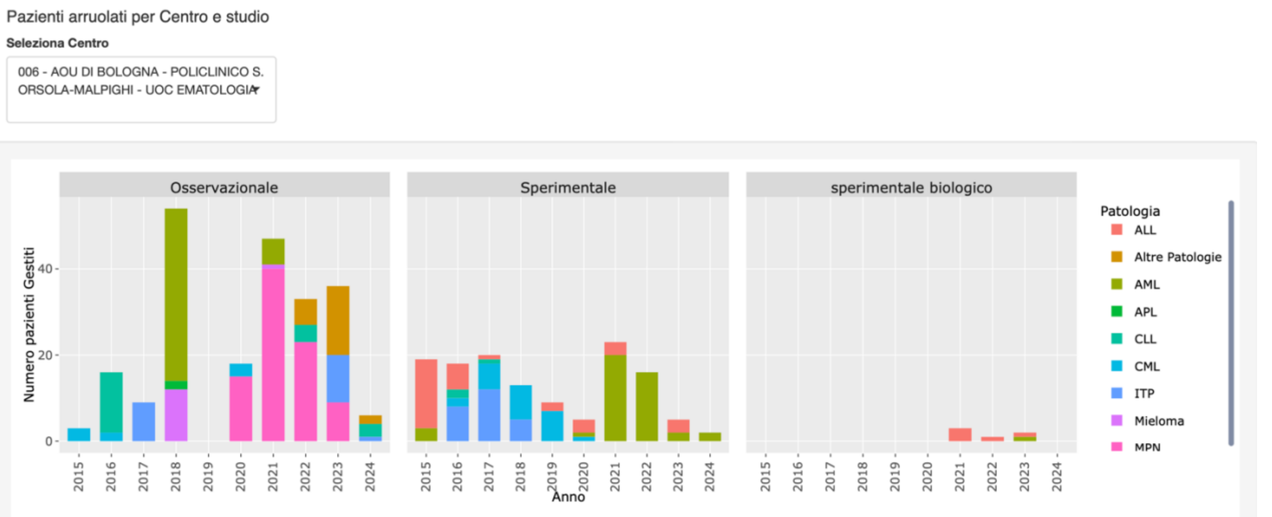


Figure 22

3.4.3 Study Report

The study report section contains a table (Fig. 23) with much information that I believe was of fundamental importance when trying to analyze the state of a study, like the number of expected patients, the dates, and the number of centers that are conducting the study. There is also information regarding doctors and researchers that for privacy reasons I decided to exclude from my thesis.

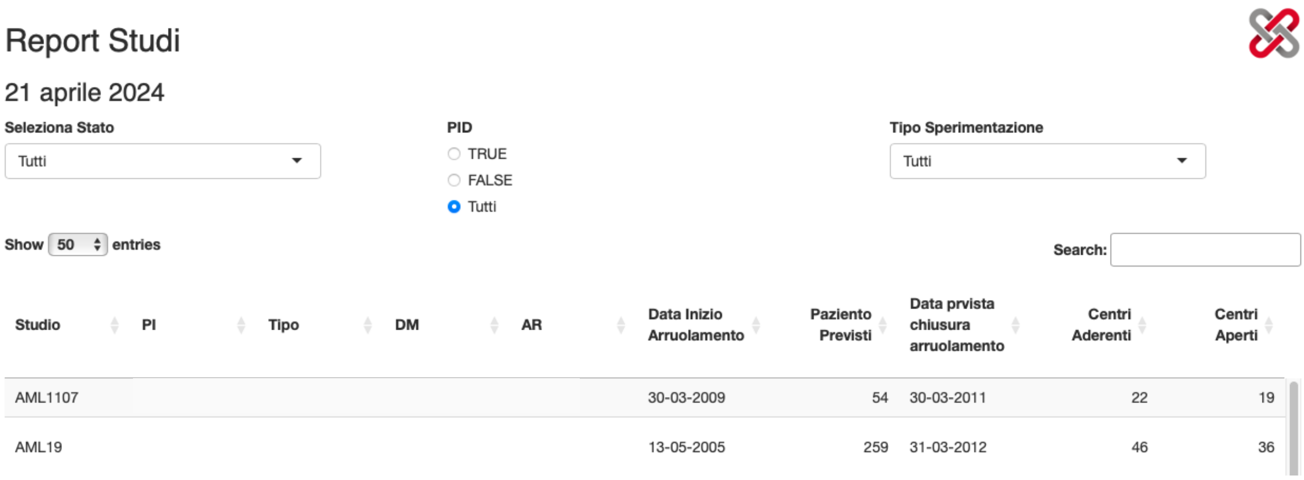


Figure 23, study report

3.4.4 State of GIMEMA's center

As explained before, one of the other types of data that could be found in the database regarded the state of their Centers. A center is a Hospital or clinic where one of the research studies is being brought forward; for example, if Gimema is researching AML (Acute Myeloid Leukemia), all the hospitals that want to bring forward the study with their patients will be registered as Gimema's centers and the patients will be monitored as such.

In this section of the Dashboard, I placed all of the information that Gimema has on their centers, number of patients, state of the different studies, timeliness, and more. This information is fundamental mainly for Internal monitoring to understand which center is performing best and how to redistribute their resources to improve performance and efficiency.

In the figure below (Fig. 24) you can see an example of the information that can be found in this table. In the top left corner, the users can select the center for which they want the information, in this case, 'Università degli Studi di Roma Sapienza', or select 'Tutti' (*All*)to have the information for all the centers aggregated. In the space below you can see the table with some of the information that can be found, like the study that is being researched, the pathology, the type of experimentation, data about patients, and the dates of enrolment in the center. Other information can. Also be found in the table if the user scrolls left that mostly regards Patients and dates of end of treatment.

Centro GIMEMA

21 aprile 2024

Seleziona Centro

028 - UNIVERSITÀ DEGLI STUDI DI ROMA "SAPIENZA" - DIPARTIMENTO DI MEDICINA TRASLAZIONALE E DI PRECISIONE - U.O.C. EMATOLOGIA

Situazione per studio aperto

Show 10 entries

Search:

Studio	Patologia	Tipo sperimentazione	Pazienti previsti dal sample size	Pazienti arruolati nello studio	Data apertura arruolamento dello studio	Apertura arruolamento nel centro	Data apertura arruolamento nel centro	Pazienti previsti nel centro per anno
ALL2418	ALL	Sperimentale	76	64	30-10-2019	Si	27-04-2020	2

Figure 24, state of stuies based on centre selected

Below the table with the information on the center (Fig.24), the users can find a very useful legend (Fig. 25) that gives a brief description of the studies that are being carried out in the selected center. Besides the short description, the user can also find the name of the main doctor conducting the treatment but for privacy, I decided not to include this information in my thesis.

Legenda Studi

Show 10 entries

Search:

Studio	Titolo
ALL2418	A Phase IIA Study of Feasibility and Effectiveness of Inotuzumab Ozogamicin (IO) in Adult Patients with B-Cell Acute Lymphoblastic Leukemia with positive Minimal Residual Disease before any Hematopoietic Stem Cell Transplantation
ALL2620	Ponatinib for the management of minimal residual disease (MRD) and hematological relapse in adult Ph+ acute lymphoblastic leukemia (Ph+ ALL) patients
ALL2922	Combination of Ponatinib plus chemotherapy as frontline approach for the treatment of patients with BCR/ABL1-like acute lymphoblastic leukemia (BCR/ABL1-like ALL)
APL0618	A Long-Term Retrospective and Prospective Safety Study of Arsenic Trioxide in Patients with newly diagnosed, Low- to Intermediate-Risk Acute Promyelocytic Leukemia (APL)
AML2120	A prospective multicentre observational study for the evaluation of clinical-hematological characteristics of familial Acute Myeloid Leukemia (AML) and myelodysplastic syndromes (MDSs)

Figure 25, legend defining all aspects of the studies present in the selected centre

3.4.5 Studies open to enrollment

As we have seen so far, for Gimema is very important to constantly monitor the status of each study, for this reason, besides giving them many charts I also decided to include a section that could give them the possibility to instantly read some of the most important information on the studies open to enrollment.

In this section the users can find the same information present in the chart section for studies open to enrollment but in a different format inserted in a data table.

Stato Studio GIMEMA

21 aprile 2024

Seleziona Studio

AML2120

Studio Gimema

Show 10 **entries**

Search:

Studio	Titolo	Coordinatore	Patologia	Tipo Sperimentazione	Pazienti previsti dal sample size	Data apertura arruolamento dello studio	Pazienti arruolati elegibili
AML2120	A prospective multicentre observational study for the evaluation of clinical-hematological characteristics of familial Acute Myeloid Leukemia (AML) and myelodysplastic syndromes (MDSs)		AML	Osservazionale	237	16-12-2020	72

Showing 1 to 1 of 1 entries

Previous 1 Next

Figure 26, table for state of a study selected by the user

In the top part of this section the users can choose the study that they want to observe, and it will print out a small legend (Fig. 26) with the basic information on that study.

Below this table, it is possible to see the Situation of the study (Fig.27) with information regarding the different centers in which the study is being brought forward, like the dates of enrolment, the status in each center, the doctor carrying the treatment, and more. Once again, for privacy reasons, I decided to white out the names of the Doctors.

Situazione Studio

Show 50 **entries**

Search:

Studio	cic	Denominazione Centro	PI	Apertura arruolamento nel centro	Data apertura arruolamento
AML2120	6	006 - AOU DI BOLOGNA - POLICLINICO S. ORSOLA-MALPIGHI - UOC EMATOLOGIA		Si	25-02-2021
AML2120	1	001 - AOU OSPEDALI RIUNITI "UMBERTO I - G.M. LANCISI - G. SALESI"- ANCONA - SOD CLINICA EMATOLOGICA		Si	07-04-2021
AML2120	101	101 - ASL SALERNO, PRESIDIO OSPEDALIERO TORTORA PAGANI - EMATOLOGIA		Si	07-01-2021

Figure 27, state of the study

3.4.6 Regulatory Timing

In this section of the Dashboard, the User can find three main sections one dedicated to the timing of the Studies, one for the centers, and one for studies in the centers. These sections are easily accessible through a dedicated Home Page like the one found in the chart section (Fig. 28).

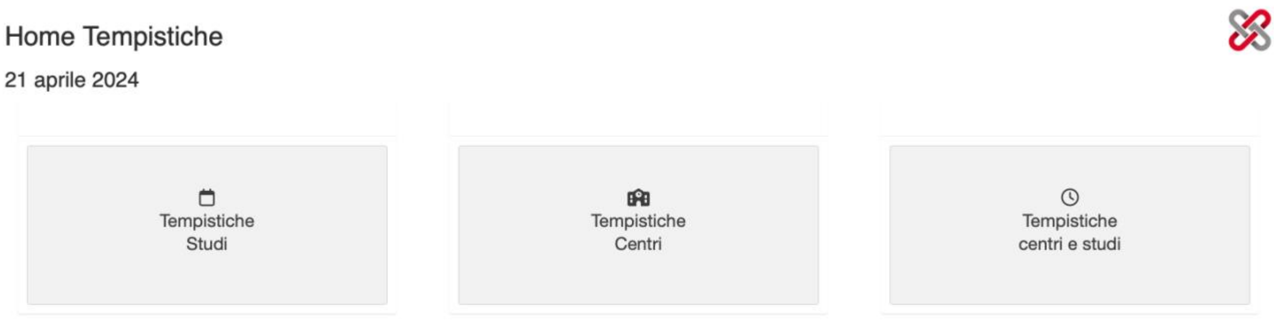


Figure 28, regulatory timing home page

The regulatory timing refers to the time that a specific study takes to complete some regulatory events. There are three main regulatory events that each study must go through before having the possibility to be researched and they are: *Completion of the report, Negotiation,* and *Integration.* For each study, the user can find the number of studies and the maximum, minimum, average, and median time that it takes to complete that event. The figure below (Fig. 29) shows an example of the study AML2220.

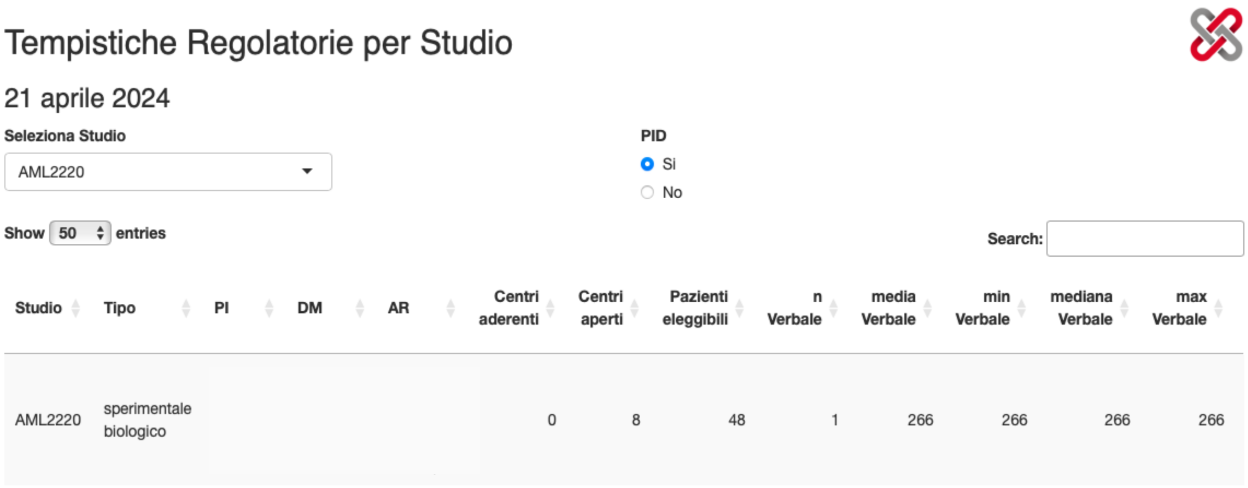


Figure 29

Other information contained in this section refers to the doctors and centers conducting the specific treatment.

The regulatory timing section continues with similar information also on the centers. In this section, the user can find the same information we saw above but dedicated to a specific center. The figure below (Fig. 30) is an example of the Università degli Studi di Roma Sapienza.

Tempistiche Regolatorie per Centro

21 aprile 2024

Seleziona Centro

028 - UNIVERSITÀ DEGLI STUDI DI ROMA "SAPIENZA" - DIPARTIMENTO DI MEDICINA TRASLAZIONALE E DI PRECISIONE - U.O.C. EMATOLOGIA

Show 50 **entries** **Search:**

Centro Gimema	Citta	Adesioni fornite	Studi aperti	n Verbale	media Verbale	min Verbale	mediana Verbale	max Verbale
028 - UNIVERSITÀ DEGLI STUDI DI ROMA "SAPIENZA" - DIPARTIMENTO DI MEDICINA TRASLAZIONALE E DI PRECISIONE - U.O.C. EMATOLOGIA	ROMA	18	18	17	197	9	110	645

Figure 30

The last part of the regulatory timing is dedicated to studies and centers together to better monitor the situation for each study in each center. The information found this time is aggregated and it states only the average days that it takes to complete one of the specific events for that study in that center; the following figure (Fig. 31) is an example of the study AML2120 conducted at the Università degli Studi di Roma Sapienza which for example states that it takes 95 days to complete a Report for the study in Rome.

Tempistiche Regolatoorie centri e studi

21 aprile 2024

Show 50 **entries** **Search:** sapienza

Centro Gimema	Citta	studio	Tempo verbale CE in giorni
028 - UNIVERSITÀ DEGLI STUDI DI ROMA "SAPIENZA" - DIPARTIMENTO DI MEDICINA TRASLAZIONALE E DI PRECISIONE - U.O.C. EMATOLOGIA	ROMA	AML2120	95

3.4.7 Publications

The last section of the dashboard is dedicated to Publications. This section is different from all the others since the code that creates the different visualizations is not based on Gimema's Database but on the API (*Application Programming Interface*, a way for two or more computer programs or components to communicate with each other) offered by Pubmed, the biggest database for publications on life sciences and biomedical topics. The API allowed me to directly connect to their database and with a simple SQL query retrieve only the publications that I wanted to display, thus those that have Affiliation, Gimema or Italian Group for Adult Hematologic Diseases, and those that have Gimema as Title or present in the Abstract of the publication. Thanks to this quick access to their database, I created a new table that shows all the publications and other visualizations related to the number of publications and their topics.

Like the sections dedicated to charts and Regulatory timings, the publication section has its home tab (Fig. 31) split into three main views, Pubmed Gimema, Pubblicazioni per anno (Annual publications), and Abstracts.

Figure 31

The first section, *Pubmed Gimema*, is dedicated to a table (Fig. 32) in which the user can find all the publications, as defined before. In this table there are much of information dedicated to each publication, such as the name of the Authors, the journal in which it was published, the Keywords of the publication, and then two very important codes that are *PMID* and *DOI* two identifiers for the publication, the first one is dedicated only to PubMed citations (PubMed Unique Identifier) and thus it allows the users to find that publication un PubMed. The second one, DOI, is known as Digital Object Identifier and it allows us to find publications all over the internet. In my Dashboard, the user can click on the PMID and they will be directly redirected to the PubMed page dedicated to that Abstract.

Pubmed Gimema

26 aprile 2024

Show 10 entries

Search:

Titolo	Autori	Giornale	Parole Chiave	pmid
Key lifestyles and interim health outcomes for effective interventions in general populations: A network analysis of a large international observational study.	Li, Fong, Lok, Wong, Ho, Choi, Pandian, Davidson, Duan, Tarrant, Lee, Lin, Akingbade, Alabdulwahhab, Ahmad, Alboraie, Alzahrani, Bilimale, Boonpatcharanon, Byiringiro, Hasan, Schettini, Corzo, De Leon, De Leon, Deek, Efficace, El Nayal, El-Raey, Ensaldo-Carrasco, Escotorin, Fadodun, Fawole, Goh, Irawan, Khan, Koirala, Krishna, Kwok, Le, Leal, Lezana-Fernández, Manirambona, Mantoani, Meneses-González, Mohamed, Mukeshimana, Nguyen, Nguyen, Nguyen, Nguyen, Nurumal, Nzabonimana, Omer, Ogungbe, Poon, Reséndiz-Rodriguez, Puang-Ngern, Sagun, Shaik, Shankar, Sommer, Toro, Tran, Urgel, Uwiringiyimana, Vanichbuncha, Youssef	J Glob Health	Male; Adult; Female; Humans; Quality of Life; Cross-Sectional Studies; Life Style; Exercise; Outcome Assessment, Health Care	37861130

Figure 32

The next section ‘*Pubblicazioni per Anno’* has two main parts, one with a chart (Fig. 33) showing the number of publications per year published by Gimema’s affiliates, and the second one which is a table describing the chart above, meaning that it has two columns, year, and Number with the number of publications per year.

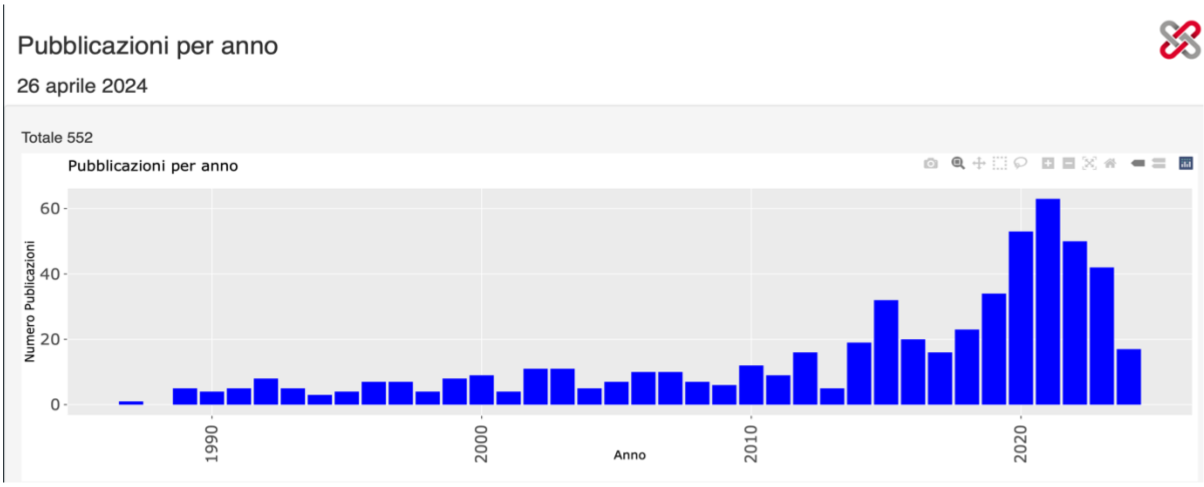


Figure 33

In the chart above (Fig. 33) the user can quickly check by hovering with the mouse, the number of publications per year published on PubMed by a Gimema affiliate or with the Italian Group for Adult Hematologic Diseases. Since it may be useful to check the numbers in a table, I also decided to insert a new data table that shows the data plotted above. In the chart below (Fig. 34) I ordered the publications for the number published per year, but the user can order them as he wants, both per year or per number of publications per year.

Lista Pubblicazioni per anno

Show 100 entries

Search:

Anno	Numero
2021	63
2020	53
2022	50

Figure 34

The last section that a user can find in the parts dedicated to Publications is 'Abstracts'. In this section, the user has different information for the abstracts of the publications. There are three sections, the first (Fig. 35) has three tables showing a split for the abstracts based on Year, Patology, or type of study.

Abstracts per Anno

Anno	Numero
2022	22
2023	40
2024	19

Patologia	Anno	Numero
ALL	2022	4
ALL	2023	5
ALL	2024	3
AML	2022	9
AML	2023	20
AML	2024	6
CLL	2022	1
CLL	2023	2
CLL	2024	2
CML	2022	4

Tipo	Anno	Numero
Clinico-biologico	2022	17
Clinico-biologico	2023	29
Clinico-biologico	2024	14
QoL	2022	5
QoL	2023	11
QoL	2024	5

Figure 35

The second section has a chart (Fig. 36) showing the number of abstracts per year split based on the Congress for which the publication was published. The chart (Fig. 36) is a barplot that has again three columns for the last three years and each bar is color-coded based on the Congress. As with all the other charts in my dashboard also this one is interactive and by hovering the mouse over any column, the users can quickly check the numbers shown. For example in this case we can see that in 2023 the number of abstracts publicated is higher than both 2022 and 2024 and that the biggest congress is ASH.

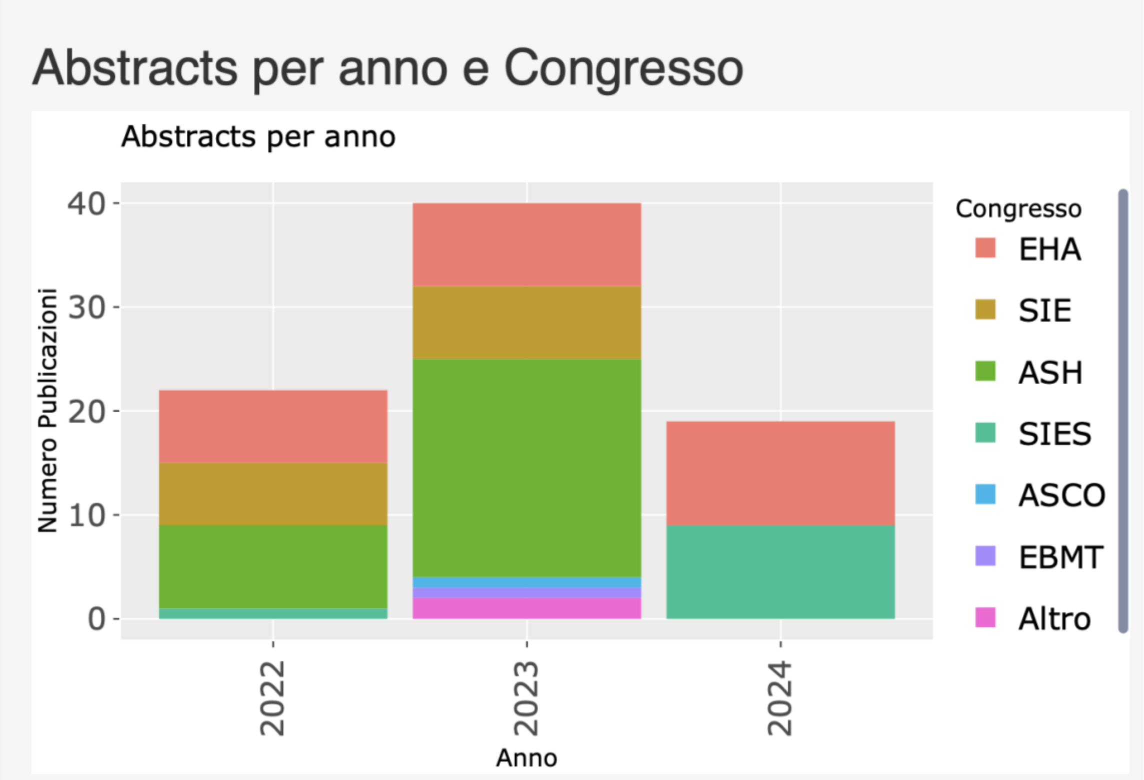


Figure 36

The last section of Abstracts consists of a table (with a list of all the abstracts with information for each abstract like First author, the study with all the specifications on the type of study and the pathology, the congress for which it was published, and the names with which the Publication is saved internally so to quickly access it. Each publication is saved in one or more formats (PDF, PPTX, and Poster) in the internal Database of Gimema, from my dashboard the user can easily access each abstract and download it by clicking on a link above the table that redirects to the page where all the abstracts have been stored and download them. This part can be improved by allowing the user to directly download the abstract format from the dashboard but as of the day I'm writing my thesis, because of how the abstracts are saved internally. The names of the files are most of the time equal from one abstract to another and they are saved by the author of the abstract who is not always an employee of Gimema but only one affiliated with them or who wants to publish something for them.

The way to solve this would be to create a standard for saving the abstract but it would require also changing all the older publications' file names.

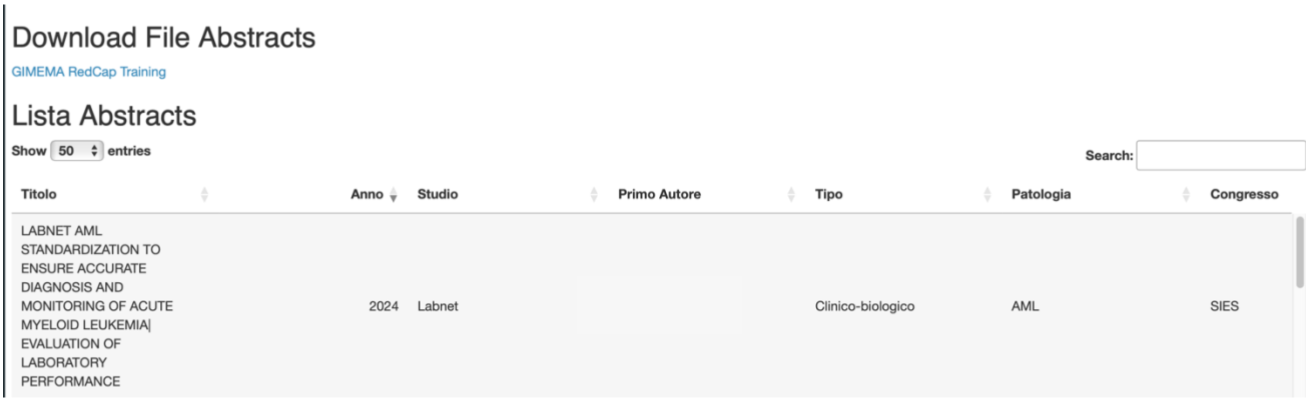


Figure 37

Privacy issues that we have found while publishing it; both for the data related to patients and for the names of the Doctors, the dashboard should be only published in a private server, and thus not available to the public, but being bulti with R-shiny we are forced to use the servers offered by Posit. As of today, we are still working to find a way around so to make this dashboard available for every employee in Gimema and improve their efficiency and effectiveness while working.

4. Conclusion

In this thesis, we explored the field of data visualization, first by examining the historical developments in this field, such as the Florence Nightingale case. We also looked at more recent examples of data visualization, such as COVID-19, and saw how beneficial it was, but also how dangerous it could be to present data incorrectly because it can easily skew people's perceptions.

Additionally, we witnessed the use of data science and visualization in the healthcare sector, highlighting the significance of a sound analysis and the fact that, in the absence of an appropriate chart, data cannot reveal much about the observed event.

In the last section, I wanted to show how data visualization is used in practice by showing a Dashboard that I created for Gimema. In this last part of the thesis, it was shown how to create a dashboard using R and its many libraries, but also how a dashboard should be structured to give the most effectiveness and efficiency to a company that should use it every day.

When I started creating my dashboard, I realized that my knowledge of HTML and CSS was limited. Coming from a Data Science and Management degree I did not have many skills in the web development area, thus I had to step out of my comfort zone and dive into this world so best present my data in the most possible appealing and user-friendly way. For example, creating a well-structured UI may not have been within my initial skill set as a data scientist, but it was a challenge that I embraced enthusiastically. Through dedication, self-learning, and a willingness to push beyond my comfort zone, I overcame the obstacles and developed a dashboard that effectively communicates insights from data and provides an engaging user experience.

Another important milestone in the creation of the thesis was working with Gimema. While there I learned how data analysis and visualizing the data might be lifesaving if you deal with patients and hematology. The healthcare industry uses data not only to improve itself but also to save people’s lives and thanks to the many different technologies we have today this has been made much easier and faster but it’s still a very hard challenge. The experience at Gimema taught me the importance of caring about the data itself and how to treat it respectfully since it is not just a number but it might be representing a person or a disease that has to be thoroughly analyzed and studied.

Gimema constantly uses data analytics to increase the life expectance of many patients with hematological diseases and thanks to data visualization, and my dashboard, they can quickly look at some of the main statistics and thus be more efficient with their everyday work.

This experience has not only expanded my technical skill set but also deepened my appreciation for the interdisciplinary nature of data science and the importance of effective communication through visualization. Thanks to this experience I learned that data science can be applied in many different aspects of life and that thanks to its interdisciplinary nature it allows us to gain many different knowledge in different fields of work.

In conclusion, this thesis journey delves into the complex world of data visualization. Through this investigation, it became clear that, while data visualization can be a powerful tool for conveying insights, its effectiveness is dependent on correct depiction and careful interpretation. The building of a dashboard for Gimema provided a realistic demonstration of these principles, necessitating a combination of data science expertise and my newly acquired web development skills. The work with Gimema showed the transformative power of data analytics in healthcare, where timely insights can directly improve patient outcomes. Lastly, this experience demonstrated the interdisciplinary character of data science, underlining the significance of effective communication and ethical issues while handling data.

Bibliography

1. **“Charles Joseph Minard.”** *Wikipedia*, Wikimedia Foundation, 17 Jan. 2024, en.wikipedia.org/wiki/Charles_Joseph_Minard.

2. **Frerichs, Ralph** R. *Maps of the 1854 Broad Street Pump Outbreak*, https://www.ph.ucla.edu/epi/snow/mapsbroadstreet.html

3. **Unwin, Anthony**. “Why is data visualization important? what is important in data visualization?” *2.1*, 31 Jan. 2020, https://doi.org/10.1162/99608f92.8ae4d525.

4. **Writer, Rock Content, and Chris Dornellas.** “45 Ways to Communicate Two Quantities.” *Rock Content*, 12 Feb. 2021, rockcontent.com/blog/45-ways-to-communicate-two-quantities/.

5. https://raw.githubusercontent.com/ft-interactive/chart-doctor/master/visual-vocabulary/poster.png

6. **Krylov, Andrii.** “Healthcare Data Visualization: Importance, Benefit, Examples.” *Kodjin*, 1 Nov. 2023, kodjin.com/blog/healthcare-data-visualization-importance-benefits/.

7. **Iftikhar, Aleeha, et al.** “Role of dashboards in improving decision making in Healthcare: Review of the literature.” *Proceedings of the 31st European Conference on Cognitive Ergonomics*, 10 Sept. 2019, https://doi.org/10.1145/3335082.3335109.

8. **Zhang, Guangyue, et al.** “Continuous monitoring with machine learning and Interactive Data Visualization: An application to a healthcare payroll process.” *International Journal of Accounting Information Systems*, vol. 46, Sept. 2022, p. 100570, https://doi.org/10.1016/j.accinf.2022.100570.

9. **Stadler, Jennifer G., et al**. “Improving the efficiency and ease of healthcare analysis through the use of data visualization dashboards.” *Big Data*, vol. 4, no. 2, June 2016, pp. 129–135, https://doi.org/10.1089/big.2015.0059.

10. **Rabiei, Reza, and Sohrab Almasi**. “Requirements and challenges of hospital dashboards: A systematic literature review.” *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, 8 Nov. 2022, https://doi.org/10.1186/s12911-022-02037-8.

11. **O'Connor, Siobhan, et al.** "Data Visualization in health care: The Florence effect." *Journal of Advanced Nursing*, vol. 76, no. 7, 8 Mar. 2020, pp. 1488–1490, https://doi.org/10.1111/jan.14334.

12. **Siegel, Corey A., et al.** "Real-time tool to display the predicted disease course and treatment response for children with Crohn's disease." *Inflammatory Bowel Diseases*, vol. 17, no. 1, Jan. 2011, pp. 30–38, https://doi.org/10.1002/ibd.21386.

13. *Application layout guide*. Shiny. (2024, January 10). https://shiny.posit.co/r/articles/build/layout-guide/

14. **Posit**. (2024, February 29). https://posit.co/