Università
Ca'Foscari
Venezia

Master's degree programme in Computer Science

"Data Management and Analytics"

Final Thesis

Presenting environmental data on social media: a set of guidelines for small screens and a case study

Supervisor
Ch. Prof. Fabio Pittarello

Graduand
Giovanni Pivato
Matriculation Number 862621

Academic Year
2022/2023

Acknowledgements

Before proceeding with the discussion of the topics analysed in this master thesis, I would like to express my gratitude to my supervisor, Professor Fabio Pittarello, and Professor Fabio Pranovi, who assisted and guided me throughout this project.

I would also thank my colleagues, Manuel and Lorenzo, for sharing this journey with me.

I would like to thank my parents, who supported me throughout these years, and my sister Ilaria, who always encourages me to do my best and pursue my ambitions, no matter how difficult it may be.

I would like to thank Giulia, for all the love and support she has given me over these months, even when I thought I could not make it.

Finally, I would like to thank my friends, Riccardo, Damiano, and Giulia for always sticking by my side.

Abstract

The aim of this thesis is to determine if communicating environmental data on social media, taking advantage of the methodologies used in information visualization, can have a significant impact on the awareness of environmental sustainability.

Despite the potential of social media for disseminating scientific knowledge, a proper communication methodology is required, both for distinguishing this communication from fake news and pseudo-science and for taking advantage of the point of strengths and weaknesses of small screens.

This thesis contributes to this research area with an initial assessment focused on the relationships between data and social media as perceived by the young generations and then by proposing a set of guidelines for small screens that are applied and then evaluated in a case study related to the gathering of environmental data, among which carbon dioxide data, in the city of Venice.

The thesis also includes an experimental activity that has provided real-world environmental data needed for testing the guidelines. A low-cost platform was designed and implemented for collecting these data, which were then processed with different data visualization tools for being delivered through social media. A final user study was organized for understanding the impact in terms of user engagement.

Contents

Chapter 1

Introduction

1.1 General introduction

This thesis is focused on the potential of Information visualization for representing and presenting data on social media platforms, with a particular reference to small-screen devices.

This work was carried out as a follow-up to a devised and implemented by the Ca' Foscari University of Venice. This project aims to monitor a selection of environmental variables related to the city of Venice and to disseminate the information gathered using various types of visual representations that would allow expert and non-expert audiences to become aware of the environmental conditions of the city. The ultimate goal is to implement a process of collecting and disseminating environmental data to raise awareness of sustainability issues among a large public.

The thesis is a study for analyzing the level of awareness of people aged 20-30 about using social media for accessing data and using visual representations for presenting such data on small-screen devices (such as smartphones and tablets) and social media. A related goal of the study was to indicate the appropriateness and relevance of social media to disseminate scientific knowledge and the impact of this process on awareness of socio-scientific issues.

In the following discussion, we will give further details related to the motivations behind this work, the research questions, and the methodology used to answer these questions.

1.2 Motivations

The motivations behind the project are related to the four main topics: environmental awareness, information visualization, small-screen devices, and socio-scientific data dissemination through social media.

1.2.1 Environmental awareness motivations

Global warming is one of the biggest problems our society is facing in recent decades.

The consequences of the increase of greenhouse gasses in the atmosphere due to means of transport and industries are one of the most discussed issues by the scientific community and governments worldwide. Global average temperatures rise each year, causing catastrophic damage to several ecosystems across the planet.

All over the world, there exist several Observatories that monitor the so-called "breath of the Earth," that is, the concentration of CO_2 in the atmosphere as ppm [1] and update the *Keeling Curve*. This graph shows the increase of the levels of ppm of CO_2 in the atmosphere (shown in Fig. 1.1) initially produced from the results of the research made by chemist *David Keeling* at the Mauna Loa Observatory, in Hawaii [**81**].

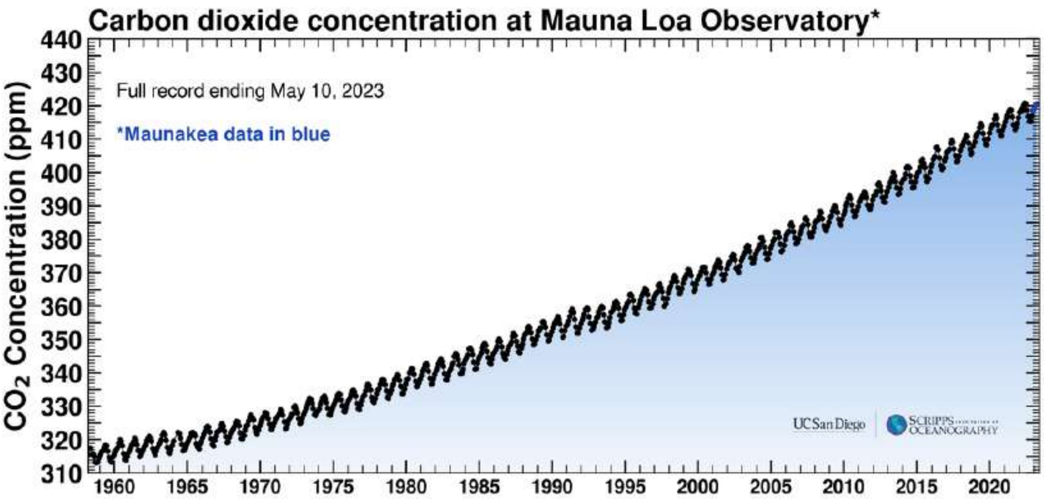


Figure 1.1: Full Record Keeling Curve up to 10 May 2023, [119]

According to the data recorded by the Observatory, since 1960, the concentration of CO_2 has increased from an average value just below 320 ppm to an average value around **419 ppm** in 2022 [**8**]. In particular, on April 15, 2023, an average value of **424.30 ppm** was recorded by the Observatory. The values registered that day correspond to the first time the value 423 ppm was exceeded in the Keeling Curve [**29**].

In recent years, the consequences of atmospheric pollution and global warming are starting to become more and more evident, awareness regarding environmental sustainability, which is "the responsibility to conserve natural resources and protect global ecosystems to support health and well-being, now and in the future" [**60**], increased. Civil society demands immediate actions to stop climate change before its impact becomes irreparable, particularly

[1] parts per millions (ppm) is a dimensionless unit of measure indicating a ratio of homogeneous measured quantities of one million to one. In the case of CO2, it suggests that for every million parts of air, one part consists of carbon dioxide.

regarding the reduction of CO_2 emission.

Even if CO_2 is neither a pollutant nor has dramatic effects on human health, its excessive production and consequent release in the atmosphere causes the temperature of the planet to rise as it prevents the transfer of energy from the earth to outer space, creating what is commonly called the "greenhouse effect." **[8]**. Moreover, it endangers the existence of Ozone, a gaseous layer in the atmosphere that protects Earth from the harmful action of UV ultraviolet rays coming from the Sun.

However, even if awareness about these topics is increasing, it is not yet sufficient. Large groups of the global population deny global warming or need to be correctly informed about climate change and environmental sustainability. Public engagement is a "very strong mean to raise awareness and stimulate responsible environmental behavior" **[21]**.

This thesis aims to spread correct information and awareness related to environmental sustainability in the city of Venice.

In spite of the fact that the island is a pedestrian zone, the continuous passage of ferries and other boats in the navigable canals, as well as of large cruise ships in the area immediately adjacent to the lagoon, cause a continuous wave motion, also raises CO_2 emissions and noise pollution in the city.

Collecting these data represents the first step in improving the knowledge about the environmental conditions of the island and a necessary preliminary action for improving the awareness of citizens. For this reason, this work includes the discussion and implementation of a data-gathering system composed of in situ sensor stations, with the final goal of creating representations that could help to raise awareness in the population regarding such conditions and push people to demand action before the ecosystem, which the island of Venice hosts, is hopelessly corrupt.

1.2.2 Data visualization

In addition to the reasons described above, which concern the domain of environmental sustainability, this work was carried out to deepen the knowledge in the discipline of Information visualization.

Information visualization is, in fact, an ever-growing discipline that can and should be used more to convey even the most complex data to the largest audience possible. Thanks to transparency and the easy fruition of data, people would be much more aware of key issues that afflict our society and would therefore be more prompted to act on them.

At the time of writing this thesis, more and more data are being produced worldwide, and their complexity is growing simultaneously. As such data are being produced, there is a constantly increasing demand for representation and efficient communication of this information. Therefore, data visualization has become fundamental to processing this info in ways that can convey its complexity and match the context they refer to. As a result, the representations have to be as complex, dynamic, and aesthetically pleasing as the context requires.

The role of visualization is so widespread nowadays that it is starting to expand beyond the confines of pure representations. Visual Representations are a tool for "validating arguments, initiating debates, raising awareness, and stimulating interest and engagement on a topic" **[57]**. It is no longer a question of simply communicating data through representations to an audience of experts but instead of creating a discipline that allows data to be communicated to citizens to stimulate the observer's interest in thinking about what he is looking at.

Three fundamental aspects have to be considered when dealing with data visualization: *context*, *aims* of the work, and *audience* **[57]**.

In fact, as the context and the intentions of the work that requires representations change, the scope of the visualization itself changes. There is quite a difference between communicating data without the need for engagement and creating a visualization aiming to raise awareness regarding a specific topic. Moreover, the audience we are trying to communicate with plays a crucial role in determining the complexity and the level of engagement we want our representation to reach.

Therefore, when dealing with data visualization, it is fundamental to define the context in which the representations will be used, the intentions of the work that will be realized, and the audience we are trying to communicate with. These three aspects are decided in the early stages of the process so that it will be possible to create the most suitable representations for the context and focus on how well the visualizations achieve the aims and engage the chosen audience.

Overall, information visualization represents a powerful tool that deserves further indication for exploiting its potential to raise citizens' awareness about different topics. This thesis explores such potential for environmental issues.

1.2.3 Data dissemination through social media platforms

As stated before, raising awareness regarding the topics of environmental sustainability is fundamental to combat climate change and taking action to slow down its damage to the world. Therefore, this thesis aims to find innovative ways to increase awareness of the aforementioned topics.

As stated in section 1.1, the context in which the data dissemination process will be carried out is social media and small-screen devices. Nowadays, social media platforms represent a relevant opportunity to communicate information, data, news, and scientific knowledge to the general public. Thanks to the algorithms with which they are developed, the process of disseminating and sharing information is very fast. It allows the user to reach a large and heterogeneous audience **[30]**. If used consistently and wisely, social media can become a powerful tool for raising awareness about important issues for man and society, as well as for initiating debates and carrying out valuable exchanges of views.

However, social media also present quite a challenge when dealing with the diffusion of information. The number of fake news and data shared on social media is constantly growing and poses a serious problem to society. Even if the moderators of these platforms are starting to use fact-checking for articles and data shared to regulate the spreading of misinformation, a large part of the population is not able to distinguish accurate data from fake ones. According to a survey conducted by Ipsos **[61]** in 2022, it turns out that young people aged 18-30 and the most educated tend to "analyze the reliability of online information and, therefore, protect themselves from disinformation" **[61]** more frequently than older generations.

Moreover, the heterogeneity of social media users can be a disadvantage. Even if social media allows it to reach a wider audience of people of various age groups and levels of education, the representations might only be understandable by some users. This fact must be considered when developing the visualizations, as the level of data literacy of users can vary significantly based on different factors (age, level of education, country of origin, etc.).

There are different questions worth to be investigated. About the relation between data and visualization and social media:

- What is the level of information regarding the reliability of data shared on social media on the target population?
- Is there any literature regarding the use of data visualization to process environmental data and disseminate them through social media and/or in any context regarding small screens (smartphones, smartwatches, tablets)?
- How can I use the literature regarding data visualization in other domains linked to social media and small-screen devices (smartphones, smartwatches, tablets) to produce effective, simple, clear, and comprehensive representations of the environmental data gathered during the project?
- According to the literature, which is the most suitable social media platform to disseminate data?
- By distributing processed environmental data on social media, will I obtain meaningful feedback?
- If so, how can I use such feedback to improve the process and prompt the audience to act regarding environmental sustainability topics?

This thesis is meant to contribute to giving an answer to these questions, as it will be detailed in the following Chapters.

1.2.4 Visualization for small screen devices

Another essential aspect that must be considered is the fact that the majority of social media users access social platforms using smartphones and tablets.

This is an important factor when developing data representations, as they must be optimized and suitable for small screen sizes. Therefore, it is necessary to perform a literature analysis to understand if there are guidelines for creating representations optimized for small screens.

Small screen devices typically refer to electronic devices with a small display screen, such as smartphones, smartwatches, digital cameras, handheld gaming devices, and some tablets. These devices are designed to be portable and compact, and their screens are usually smaller than those found on laptops and desktop computers. The screen size can vary depending on the device itself, but they are generally less than 10 inches in diagonal length. Small-screen devices are typically used for communication, entertainment, and accessing the internet worldwide.

This aspect is fundamental to the process that will be implemented within this project and must be investigated in parallel with the analysis of the relation between data and social media.

As the analysis of the literature carried out in the next chapters will demonstrate, since the introduction of small-screen mobile devices (mobile phones, personal digital assistants and more recently smartphones, smartwatches, and tablets) on the market, in the scientific and IT community there has been a growing interest in developing guidelines for representations deployed on these devices, especially considering navigation **[32]** and in more recent years images and data visualizations **[18, 20, 24]**.

Therefore, this thesis work aims to analyze the literature regarding such guidelines and tries to answer the following questions:

- Is there a set of clear, unique, and efficient rules for creating representations suitable for small screens, particularly regarding social media?

- If the answer is no, is it possible to create such a set of guidelines focusing on social media platforms?

As I said before, the analysis and implementation process used to answer these questions will be described in the following Chapters.

1.3 Methodology

This section is devoted to the explanation of the methodology followed to carry out the project.

As mentioned in previous sections, the first phase of the project consisted of an analysis of the literature regarding the main topics taken into consideration, i.e. environmental awareness and low-cost monitoring stations, information visualization, socio-scientific knowledge dissemination through social media platforms and guidelines for the development of representations suitable for small screens.

The analysis of the literature is a key aspect of the project as it allows us to understand the state-of-the-art regarding the fields of applications cited above and detect weaknesses or areas of improvement useful for determining which aspects to focus the research on.

From the point of view of the data gathering process, a low-cost data gathering system was developed, starting from the hardware components used for a previous Ca' Foscari environmental sustainability project, called Life Redune [**131**]. The units the system is composed of are located in six different sites of the Ca' Foscari University of Venice, from which they gather data regarding environmental and pollution factors. The deployment of these stations was started during the autumn of 2022 and ended in February 2023.

Data are sent in real-time to a repository developed and constantly improved by the working team, where the data were stored.

The successive phase of the project, as stated before, was focused on the initial assessment of the relationship between data and social media as perceived by the population. To this end, it was chosen the target population: the choice fell on people aged 20-30. According to a demographic study conducted in 2021, the age group between 18 and 30 is the one that uses social media the most [**11**] and is the group that claims to be able to recognize fake news more easily [**61**]. Nonetheless, another research states that 41% of the people in the age group 18-34 fall into the trap of fake news [**80**]. Therefore, it is an interesting target audience for the assessment survey and the research in general.

The assessment process started by assessing the level of information of the target population in relation to the reliability of data on social media, the presentation of data on social media, and the use of visual representations to present such data in the context of social media platforms. For this purpose, an evaluation survey was developed to determine how people in this age group learn about and use data related to socio-scientific topics, in particular in relation to social media and the presentation of data as visual representations.

After this initial assessment phase, the following work was focused on the process of development of visual representation created on the data collected by the monitoring system and the publication of such representations on social media platforms. According to the literature review, *Twitter* is the most popular social media platform for scientists for sharing their research findings, conference updates, and other scientific news [**78**]. It is an ideal platform for sharing bite-sized pieces of information and engaging in real-time conversations with other researchers. However, since the data are to be communicated through visual representations and, according to the results gathered by the initial assessment survey, ***Instagram*** is the most used social media by the target audience.

In the final phase of the project, a sample of the users who interacted with the representations was asked to fill in a series of surveys for having feedback about the user interaction experience, the impact of visual representations on the socio-scientific awareness of the volunteers, and the possible improvements to the system.

Moreover, these questionnaires were used to gather useful insights verifying a series of guidelines for creating and improving representations on small display devices.

1.4 Contributions

The work carried out in this thesis provides a series of contributions to different domains.

In particular, such contributions are related to the fields of Information Visualization, Social media, and Data Visualization in the context of applications for small-screen devices, and environmental low-cost real-time monitoring systems.

Such contributions can be summarised in the following way:

- In the context of Information visualization and Environmental Sustainability, a series of visual representations is proposed, whose purpose is to communicate environmental data. These representations will be developed with the intention of engaging a large and diversified audience and increasing the awareness of the users regarding environmental sustainability issues.

- In the context of socio-scientific knowledge through social media, the representations will be developed and published on the social media platform case study. Furthermore, the initial assessment survey and the engagement survey will respectively show the level of information of people in the age group between 20 and 30 years in relation to the presentation of data on social media, the use of visual representations to present them on social media platforms and the impact of the proposed representations on the users' awareness.

- In the context of data visualization on small screens, it will be provided a series of guidelines for developing representations suitable for this devices, with particular attention to social media platforms. Current literature provides some proposals of guidelines about how such visualizations should be created and the scientific community urges to perform more researches on the matter; however, there is little literature about the specific case of data visualizations for social media and this thesis aims to fill this gap.

- In the context of CO_2 monitoring system, the work carried out aims to upgrade the already existing "low-cost off-grid CO_2 monitoring station" **[21]** developed by the Life Redune project in several ways. In particular,

  – by upgrading obsolete hardware with more advanced one, reducing the size of each station;

  – because the new stations were meant for deployment in university buildings, changing the power supply from solar panels to a electric source;

  – connecting the stations to the network and to work with real-time solutions;

  – providing a real-time system for data collection and persistence using Google Sheets and the servers of the Ca' Foscari University of Venice. In this way, data are always available from anywhere in the world, provided the user has access to the repository where they are stored.

  Overall, the resulting stations were more compact, connected, and suitable to work with large amounts of data.

1.5 Thesis Outline

This thesis is organized into 7 chapters, including the current one. The remaining six chapters discuss the following topic:

- Chapter 2 describes all related works related to data visualization; social media as a tool for socio-scientific knowledge dissemination with particular attention on the case study, environmental sustainability; low-cost CO_2 monitoring stations and environmental data analysis; data visualization for small screen devices and applications. This chapter has particular relevance as the bibliographic research allows us to understand the state-of-the-art scientific fields cited above and determine the actions to be taken to implement innovative features.

  In particular, there is a detailed discussion related to the Life Redune project, because this thesis uses the results of such a project as starting point. Furthermore, this analysis provides interesting insights related to the features to be developed, which at the time of the Life Redune project were too complex or long to implement.

- In Chapter 3 there is an overview of the data-gathering system, implemented by a team of students with the contribution of the author of this thesis. This chapter will first illustrate the project requirements and, the work organization. It will then pass on the description of the upgrading of the Life Redune data-gathering system, illustrating the new architecture of the system, the hardware used, and the limitations encountered in the assembly process. The chapter includes then a description of the data-gathering process and the persistence system developed to store the data, going into detail about security and accessibility issues and the data flow from the stations to the repository. Finally, the locations of the collection units are described.

- Chapter 4 describes the initial assessment of the relationship between data and social media, as perceived by young people aged 20-30; the chapter explains how the survey was drafted, the target audience gathered for the survey, the statistical analysis of the information collected and finally the discussion stemming from it.

- Chapter 5 describes in detail the methodology used to devise a set of possible guidelines to create and improve representations of data for small screen devices. Such guidelines will consider information gathered in the initial assessment survey and the literature analysis described in Chapter 2.

- Chapter 6 describes in detail the process of creation and dissemination of visual representations of the data gathered with the monitoring system described in Chapter 3 using the guidelines presented in Chapter 5. The chapter will analyze the motivation for choosing Instagram as the designated social platform for data dissemination, which are the best representation types that can be used in this context according to the guidelines devised in the previous Chapter. After realizing the representations and explaining the motivation for each visualization choice, the Chapter describes the publishing process on Instagram. The final section of the chapter will be devoted to analyzing the feedback obtained from users, extrapolated first from a survey focused on assessing the level of engagement, comprehension, simplicity of the representations, and their appropriateness for small screens, and then from the insights produced by Instagram for each published representation.

- Chapter 7 summarises all the work done, then draws conclusions and describes possible future directions of work.

  In particular, the level of goodness of the guidelines produced in Chapter 5 will be indicated, based on the feedback obtained from the validation process described at the end of Chapter 6.

Chapter 2

Related works

This chapter is devoted to the presentation and description of scientific papers, articles, and other works that were analyzed and used as references as they were related to the topics investigated in this thesis work.

All sources cited are listed in the ”Bibliography” section at the end of the paper. As stated in Chapter 1, the Life Redune Project was a fundamental starting point for the work described in this dissertation, so its analysis is reported in a more detailed fashion.

As previously mentioned, the main topic of this master's thesis is data visualization and its application in the context of social media and small-screen devices, with a focus on environmental data visualization. To this end, it is important to underline that Data visualization is the discipline that regards "the visual representation and presentation of data to facilitate understanding"[**65**]. With the exponential increase in digital data production in recent years, the ability to effectively communicate complex information through visual representations to a wide audience has become increasingly valuable. Therefore, such discipline is widely used by the scientific community and it is covered in various papers and articles in the field.

The works presented in this chapter are related mainly to the following topics:

- Low-cost environmental data-gathering systems
- Environmental data analysis and citizen science
- Data visualization in the context of small-screen devices and applications, with a focus on possible guidelines on how to create such representations
- Social media and their role in the diffusion of data regarding scientific and socio-scientific issues

As a matter of fact, the research and analysis of the state-of-the-art works, papers, and articles that consider the topics investigated within the project is a fundamental part of this thesis, as it allows us to understand the current results obtained in these fields and try to find new ideas to improve those results and produce new knowledge that could be used in the future to inspire further improvements.

2.1 Low-cost environmental data collection systems

This section is devoted to the analysis of the works conducted on the field of low-cost environmental data-gathering systems.

Over the years, our society has become more and more dependent on fossil fuels, which are used in various industries and to power transportation means of various kinds. Although there is a growing trend toward the use of renewable energies (solar, wind, etc.), the damage caused by the increased concentration of greenhouse gases in the atmosphere is becoming more and more evident as the air we breathe is becoming more polluted, as demonstrated by the data gathered by the various observatories all over the world such as the Mauna Loa Observatory at Hawaii [**81**] and the Ottavio Vittori Observatory at Monte Cimone (Italy) [**62**].

Obviously, such observatories are equipped with state-of-the-art technologies as they need to register data in the most accurate and precise possible way. However, increased interest in the topic resulted in the need for the development of low-cost monitoring systems more diffused on the territory for research, education, and citizen awareness.

Over the years, the use of these technologies has increased as these systems are very useful for gathering different types of environmental information, such as air quality (AQ), in both indoor and outdoor environments, particularly in urban or densely populated settings where fine particulate matter and smog levels are constantly rising.

In addition, with the advent of the SARS-CoV-2 virus and subsequent Covid-19 disease, indoor air analysis has become a priority for maintaining clean and sanitized environments. Therefore, it was necessary to develop systems that were both low-cost and quick to produce, and efficient with a high level of accuracy.

In the following sections, will be described some examples of these systems, with particular attention to the prototype developed in the context of the Life Redune Project.

2.1.1 Keeling studies and his legacy

As mentioned in Chapter 1, in the 1950s chemist David Keeling carried out studies [**8**] that led to the discovery of the variability of CO_2 concentration in the atmosphere during circadian cycles, due to the photosynthesis and respiration processes of plants, which consume CO_2 during the day while producing it at night. To perform experiments without the interference of plant processes, he established the Mauna Loa Observatory at the Hawaii volcano of the same name in 1958, where he began to collect carbon dioxide concentration data.

Over the years, Keeling reached two crucial results that are still of parandum key today in the analysis of environmental data:

- the concentration of carbon dioxide in the atmosphere fluctuates annually due to plant respiration processes; it reaches its maximum in the spring, before plants begin the summer sequestration process, and hits a minimum in the fall, before rising again due to leaf fall and subsequent termination of the respiration process **[8]**.

- Year after year, the average concentration of CO_2 in the atmosphere increases and does so at a faster rate. He attributed this phenomenon to man's rapid and massive exploitation of fossil fuels since the industrial age, which has resulted in a massive release of hydrocarbons into the atmosphere.

Therefore, his conclusion from the analysis of the two trends, shown in Fig. 2.1, was that Seasonal fluctuations are a natural phenomenon, while the annual growth of carbon dioxide concentration is a process of anthropic origin that has nothing to do with the Earth's climatic and environmental variability but on the contrary, contributes to climate change.

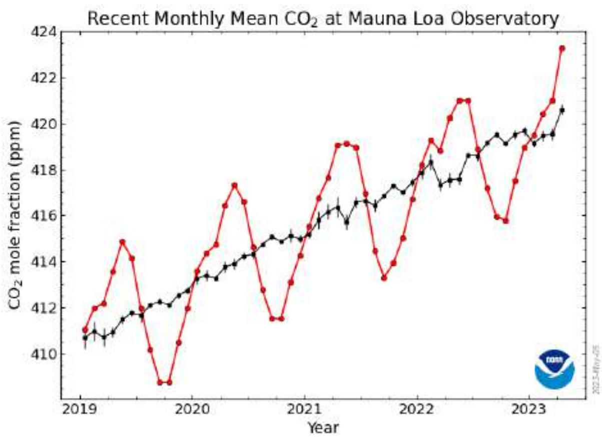


Figure 2.1: The trend of CO_2 concentration in recent years. The black curve shows the steady average increase, while the red one shows the annual swings already identified by Keeling in the 1950s. [92]

His results are still fundamental today, as the Keeling curve is a central parameter in the analysis of carbon dioxide levels in the atmosphere. Furthermore, the Mauna Loa Observatory is still active today and continues to monitor the concentration of CO_2 and update the Curve, also sharing data online, as many other observatories and institutions do **[29, 38, 62, 92, 109]**, to increase people's awareness of a process that involves the entire planet.

2.1.2 Low-cost environmental stations

As mentioned earlier, the demand for low-cost environmental monitoring technologies has grown over the years, especially in the academic field. Such emergence was also recognized in the United States by US EPA [**138**] and later on in Europe and was recommended to be included in the next Air Quality Directive [**17**].

Over the years, the deployment on the market of low-cost sensors, technologies, and Single-Board Controllers (SBC, such as *Arduino* and *Raspberry Pi* controllers, which represent respectively the 37.5% and 35% of low-cost environmental monitoring systems [**116**]) has enabled "a revolutionary shift in air pollution monitoring and assessment" [**90**]. Due to their small size, ease of programming, high availability of documentation, adequate computing power, and scalability, they have quickly become the preferred tool for these types of systems, so much so that the manufacturing companies themselves have begun to produce ready-made kits of environmental sensors in order to easily assemble small weather stations [**108, 130**] and indoor air quality systems [**49**].

Since the publication of the paper of Snyder [**124**], which recognized the role that low-cost sensor technologies will have in the future regarding air monitoring, many applications and systems using these low-cost technologies have been released and reviewed by both the scientific community and crowd-founded initiative. As shown, Raspberry Pi appears to be to this day the most used technology used to control low-cost environmental data-gathering stations. Reviewing the literature of the field, it was possible to see that it was used for many different applications.

An interesting application relates to the use of these technologies to analyze the environmental impact of road traffic in urban areas. One example is the system tested by Carl Skoglund and Dennis Sundin of Linköping University (Sweden) [**122**], which uses outdoor data collection stations assembled by combining a Raspberry Pi 4 Model B and an Arduino Uno. This system was deployed on a street in the city of Norrköping to collect and process data about humidity, temperature, particulate levels and sound to understand if was possible to estimate traffic using urban sound and levels of particles. For this purpose, they were tested in two scenarios: trafficked street and pedestrian area. However, the use of low-cost technologies showed limitations: the power bank only powered the station for six hours; the Raspberry Pi had connection issues, resulting in missing data. In addition, the sound data collected detected interference due to the crossing of cables inside the unit. However, the data obtained were satisfactory and their analysis yielded two key results: first, the sound data collected could be used to estimate traffic; second there is no correlation between the levels of collected particles and traffic.

Other works regarding traffic estimation and noise pollution were developed over the years using low-cost technology. In 2015, Dekoninck et al. [**34**] showed the presence of a correlation in the spatial and temporal variability between noise and air pollution exposure but they noted that data about this topic are principally recorded at high-density roads on peak hours, whereas data quality at low-density hours is low, affecting such environmental analysis. In 2016, the results of this research were then used by Noriega-Linares et al. **[95]** to test a low-cost Raspberry Pi-based acoustic system for noise monitoring, which had the aim of analyzing the data gathered to promote solutions for reducing noise pollution.

Another field of application of low-cost technologies is air indoor quality monitoring. The Environmental Protection Agencies and the World Health Organization (WHO) have acknowledged the impact of air quality on public health and therefore produced a series of standards and policies with the aim of regulating and possibly improving air quality both indoor **[89, 154]** and outdoor **[153]**. Therefore, many studies have been conducted by researchers on indoor air quality, especially using low-cost sensor networks due to their affordability. One interesting example is the work of Zhang et al. **[158]**, which tested a low-cost Raspbery-Pi-based air quality sensor module for comparing indoor monitoring between residential units and an educational office. The results showed that the mean concentration of $PM_{2.5}$[2] and PM_{10}[3] was higher in the residential unit and that the system managed to detect most of the critical air pollutants.

Many other studies have been carried out with low-cost technologies other than Arduino and Raspberry Pi. Examples of such systems are to be found in the applications of AirSensEUR sensor systems for particulate detection **[118]** and IoT technologies **[79]**, which are becoming increasingly prominent in the industrial sector due to their prices.

In 2014, Ferdoush and Xinrong **[106]** designed an affordable and scalable indoor air quality monitoring system based on Arduino and Raspberry Pi. Such a wireless sensor network system was designed "to provide real-time information for assisted living" **[106]** but could be used in several environmental monitoring applications. The results showed that the concentration of CO_2 was in the range of 400 ppm (clean air) when there was no occupant in the selected room, it slowly increased when people were accommodated and there was neither a door nor a window opened and it rapidly decreased when occupants left the room with doors and windows opened.

In 2016, Sonali and Venkatasubramanian **[40]** proposed an IoT system based on Raspberry Pi to monitor the concentration of several environmental parameters, ranging from carbon monoxide and dioxide to temperature and humidity, and upload all the data to the cloud. That same year, Ali et al. **[4]** produced a low-cost Arduino-based platform for long-term indoor environmental data gathering, which performed extremely well in both collecting data and logging them on the open-source platform OSBSS.

Such results were then used in 2018 by Alkandari and Moein **[5]** to support their experi-

[2] $PM_{2.5}$ are coarse particles with diameter $< 2.5\mu m$

[3] PM_{10} are coarse particles with diameter $< 10\mu m$

mental study on a low-cost system based on Raspberry Pi to measure air quality in domestic situations.

The advent of the Covid-19 pandemic has given a further boost to the use of these technologies. Lolli et al. **[75]** demonstrated that the virus that causes Covid-19 spreads faster in dry and cool environmental conditions, as well as polluted air, i.e. unfiltered air-conditioned indoor environments. As a response, the world's health institutions required to maintain social distancing and mask use to prevent the spread of infection and advised to keep rooms often open to circulate clean air.

Air quality control in public, residential, and business buildings has already been the focus of many studies **[96, 125, 128]** but all of these experiments exploited expensive equipment. A different study on indoor AQ monitoring with low-cost technologies, which could be affordable even to low-income households, was conducted by Pastor-Fernàndez et al. **[103]**. This research analyzed four different scenarios for indoor AQ based on the level of ventilation, which were tested using a system based on three different microcontrollers: Raspberry Pi model B+, Arduino MEGA 2560 R3, and ELEGOO MEGA 2560 R3. The results showed that, despite the limitations of the low-cost hardware, the systems performed very well in terms of accuracy and highlighted the increase in the concentration of CO_2 as the analysis shifted from a "cross-ventilation" scenario to a "no ventilation" one. The reference table for the value of the concentration of CO_2 and the results obtained are shown in Fig. 2.2 and 2.3.

CO_2 (ppm)	AQ
340–600	Good
601–1000	Moderate
1001–1500	Unhealthy
1501–5000	Hazardous

Figure 2.2: AQ related to CO_2 used as reference by Pastor-Fernàndez et al. [103]

Scenario	Device	Mean	Minimum	Maximum	Standard Deviation
Cross-ventilation	KIMO	450.88	400	504	27.00
	Prototype	452.49	401	503	28.98
Outdoor-ventilation	KIMO	494.07	405	585	51.55
	Prototype	494.45	407	582	50.66
Indoor-ventilation	KIMO	590.57	402	734	79.77
	Prototype	587.39	408	741	89.36
No-ventilation	KIMO	1087.69	401	1638	374.10
	Prototype	1091.24	420	1664	368.80

Note: The unit of measurement for CO_2 concentrations is ppm.

Figure 2.3: Comparison between the results obtained by the low-cost prototype developed by Pastor-Fernàndez et al. and the reference calibrated system KIMO [103]

There has been also much interest in the field of *Citizen Science*, scientific research conducted with the participation of non-expert audiences. In order to improve people's awareness about air pollution and increase of CO_2 concentration in the air, many researchers have conducted studies on air quality gathering data with the help of citizens. Low-cost sensors are used in many of these researches with the aim of improving awareness regarding data quality issues and reducing the number of wrong analyses and consequent wrong conclusions in these settings [**3, 23, 143**].

Eco-feedback is a branch of environmental psychology connected to Citizen Science. Eco-feedback exploits low-cost technologies to collect data and "provide feedback on individual or group behaviors with a goal of reducing environmental impact" [**44**]. Often, eco-feedback technologies leverage the disciplines of Information Visualization and HCI to disseminate collected data in a creative and innovative way to increase people's awareness on the topic of environmental sustainability and impact [**54**]. An interesting example is the website "Real World Visuals", whose objective is to "bring data to life" through visualizations and videos, especially related to climate change and carbon dioxide emissions [**107**].

Research conducted by Morawska et al. [**90**] tried to understand the actual efficiency of such systems. After analyzing a selection of 17 projects that used low-cost technologies in the context of air quality monitoring and environmental awareness, Morawska et al. [**90**] state that there is no clarity on the actual suitability of these systems for the intended purposes because several parameters regarding specifications and performance of these systems are lacking. However, the results obtained from many pieces of research carried out with these technologies, even at the level of commercial or crowd-founded projects, have proven to be satisfactory for the purposes and conditions intended, in the sense that the systems were fit for environmental analysis and improvement of awareness. Important factors that weigh in on the performance of such systems are papers and guides that allow one to understand how to perform calibration, quality control procedures, and improving processes on the hardware of such systems [**104**].

Another interesting work that analyzed the correlation between traffic and particles concentration in the air was the Ecomobility project coordinated by the Ca' Foscari University of Venice in association with ISAC-CNR [**139**]. This work was conducted using the resources and technologies of the ARPAV site located at Sacca Fisola, on the island of Venice, and therefore cannot be considered as related to the field of low-cost systems. However, the results are interesting for the analysis of environmental data. In particular, this study has analyzed the impact of maritime traffic and ports on the atmospheric environment. According to the data gathered in 2018, the impact of naval traffic on nanoparticles (ultrafine particles, with diameter $< 0.1\mu m$, indicated as $PM_{0.1}$) is significantly greater than that on fine (diameter between 0.1 and $1\mu m$) and coarse (diameter greater than $1\mu m$) particles and the contribute of shipping emission is between 7% and 11%. Moreover, it is possible to observe that even if the Particle Number Concentration (PNC, which is the total number of particles per unit volume of air) in 2018 is slightly increased with respect to the one of 2012, the impact of $PM_{2.5}$ (coarse particles with diameter $< 2.5\mu m$) has decreased when compared to the value of the same particles in 2012, as shown in Fig. 2.4.

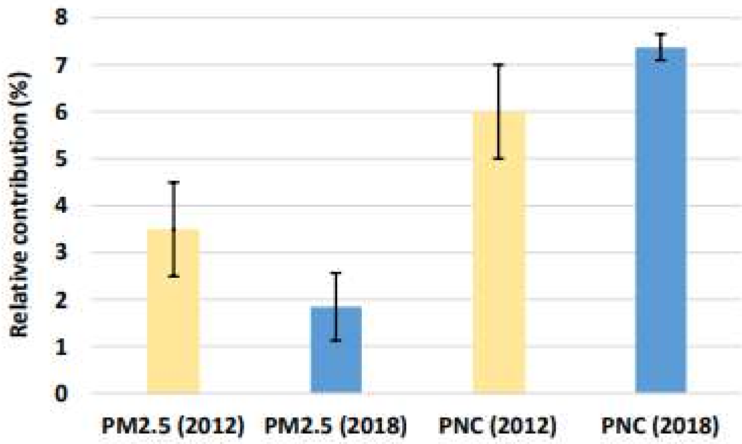


Figure 2.4: Relative traffic contribution naval at the numerical concentration of particles and $PM_{2.5}$ in 2012 and 2018 in Venice [139].

These results were then compared with other studies conducted in several sites in Europe [**140**]; the values recorded by these researches were shown to be in agreement with the results of Ecomobility for what concerns PM_{10} (coarse particles with diameter $< 10\mu m$), $PM_{2.5}$ and PM_1 (particles with diameter $< 1\mu m$. Whereas, this study was the first to be conducted in Europe on the concentration of $PM_{0.1}$, which contained the heavy metal Vanadium.

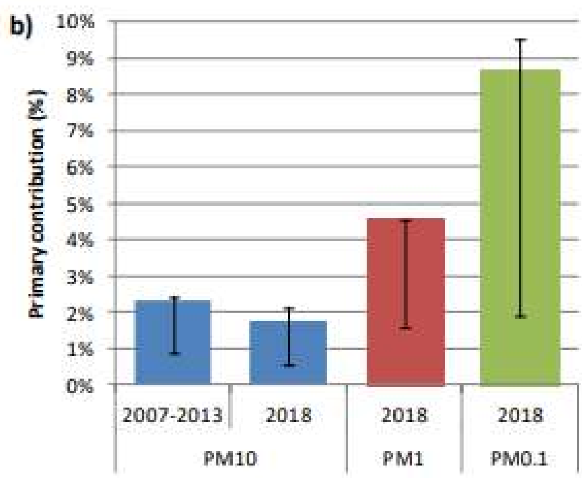


Figure 2.5: Comparison between the primary contribution of maritime traffic to atmospheric particulate matter, calculated using vanadium as a tracer, in 2018 and in previous years [139]

.

2.1.3 The Life Redune Project

The aim of this section is to describe the Life Redune Project, initiated by the Ca' Foscari University of Venice, and in particular the related activities of data gathering and analysis.

As the work described in this thesis rips off some aspects of such a project, it is important to truly understand the scope and the results of Life Redune. A Master Thesis Dissertation **[21]** describes how these activities have contributed to the field of low-cost environmental data collection systems combined with Information Visualization.

Figure 2.6: Life Redune Project Logo

Life Redune is an environmental sustainability initiative, whose purpose is "to restore and maintain the ecological integrity of 5 dune habitats and of the populations of *Stipa veneta* in 4 *Natura 2000* sites along the Adriatic coast" **[131]**.

Dune habitats play a central role in maintaining the equilibrium of coastal ecosystems, "protecting beaches from wind and water erosion, offering a suitable habitat for autochthonous animal and plant species" **[21]**. However, the presence and activity of the human population in these places seriously endanger these fragile habitats, which risk being ruined or even destroyed. Moreover, due to global warming, if those areas are damaged, such damages may be permanent unless actions are taken in this very moment to protect the natural ecosystems and the autochthonous species that live and sustain them at the same time.

The Ca' Foscari University of Venice stands out as the coordinator of the various organisations involved in the Life Redune Project, which is also co-financed by Regione Veneto and the European Commission. The Life Redune work team designed and implemented a series of actions with the aim of "replanting autochthonous species, like the Stipa Veneta, and [...] minimising damages done by humans (e.g. by installing new pathways to prevent people stomping on dune flora)" **[21]** in four sites belonging to Natura 2000. The ecosystems of these sites were monitored for the duration of the project life cycle, in order to understand if the actions of restoration taken by the Life Redune Team were effective and assess the impact of such process on the status of the ecosystems.

A fundamental aspect that the Life Redune Project aimed to understand and regulate was related to Carbon Dioxide sequestration and the impact of the dune habitats in such a process. As indicated in Chapter 1, global warming is a growing threat to our planet and is directly linked to the increase in the levels of carbon dioxide present in the atmosphere, caused by greenhouse gas emissions. For this reason, the governments of the world have drew up agreements with the aim of mitigating, and ultimately stopping, climate change. Initiatives include the Paris Climate Agreements ratified in 2015 by the member states of the European Union, whose key objective is the reduction of greenhouse gas emissions by at least 55% by 2030 compared to 1990 levels.

The U.S. Department of the Interior defines *carbon sequestration* as "the process of capturing and storing atmospheric carbon dioxide. It is one method of reducing the amount of carbon dioxide in the atmosphere with the goal of reducing global climate change" [**152**]. According to this definition, natural ecosystems and plants carry out a key role in the regulation of climate and in carbon dioxide sequestration, as they absorb CO_2 to perform the photosynthesis process. Therefore, the series of actions put in place in the context of the Life Redune Project fully are compliant with the objectives of the Paris Climate Agreement.

Life Redune Data-gathering system

As said in the previous section, regulation of Carbon Dioxide and monitoring of the CO_2 sequestration process is a fundamental aspect of the Life Redune Project. Gathering data related to levels of carbon dioxide and examining how such levels are influenced by temperature, humidity, and other environmental variables allows to understand the impact of the restoration process put in place by the Life Redune Team.

To this end, in the context of the Life Redune Project, a data-gathering system was created to collect all the information regarding environmental levels of the four sites of *Natura 2000* where the restoration process was carried out. Moreover, a tangible interface was designed and assembled to display the data gathered by the system. The interface was meant for educational purposes and tested in a primary school of the Veneto Region for testing its effectiveness [**105**]. The goal of this interface was to increase public engagement regarding environmental sustainability topics and to increase awareness of the conditions of the coastal ecosystems of Veneto.

The units

The gathering system was "highly decoupled and composed by in situ stations, data-gathering module, a persistence engine and data visualization tool" [**21**]. It was therefore assembled as a composition of off-grid low-cost stations, each of which contained modules and sensor to gather and store data. The prototype was completely designed and implemented by Lorenzo Brutti and the Team of the Environmental Sciences, Informatics, and

Figure 2.7: External view of a station of the Life Redune data-gathering system

Statistics Department of the Ca' Foscari University of Venice with "custom-Do-it-yourself solutions"[**21**] and open-source technologies. Each station was assembled using the *Grove Weather Pi kit* and additional components. In particular:

- an ***Arduino Nano microcontroller***, which runs the code used to gather and store all the data the station collects through the sensors (1 in Fig. 2.8);
- a ***K30 10,000ppm CO2 Sensor***, used to gather data regarding the Carbon Dioxide levels, expressed in ppm (2 in Fig. 2.8);
- a ***AM2315 - Encased I2C Temperature/Humidity Sensor*** used to gather data regarding temperature, expressed in Celsius, and Humidity, expressed in percentage (3 in Fig. 2.8);
- a ***GPS*** to store GPS coordinates for GIS-based analysis and keep the real-time clock of the processor (4 in Fig. 2.8);
- Four GroveWeatherPi ***solar panels*** (5 in Fig. 3.4);
- a ***SunAirPlus Solar Power Controller***, a 3rd Generation Solar Charging and Sun Tracking Board designed specifically for Arduino and Raspberry Pi systems, which takes the power from the solar panels and supplies it via USB plug to the Arduino Nano (Fig. 2.7);
- a ***battery pack*** as secondary power, in case of faulty solar panels (7 in Fig. 2.8);

- a module containing a ***SD card*** where all data were stored as .csv files (8 in Fig. 2.8).

All these components were assembled inside a box for electronic usage, which was weather-resistant, as shown in Fig. 2.8. Therefore, each station was completely self-sufficient and capable of enduring adverse climate conditions (rain, wind, hailstorm). Human interaction was needed only to extract the data from the SD cards and perform some maintenance.

Figure 2.8: Internal view of a station of the Life Redune data-gathering system

2.2 Guidelines for data visualization on small screens

The evolution of data visualization on small-screen devices has seen significant advancements over the years. With the increasing use of smartphones and tablets, data visualization can be an interesting tool for presenting information in a meaningful way. Therefore, the need to create a set of guidelines for visual representations on these types of devices has emerged as such tools have become more widely deployed in the market.

Over the years, researchers have come forward to try to create useful guidelines for representing data on mobile devices. The first attempts to create guidelines for data visualization on small-screen devices can be traced back to the early 2000s when mobile devices such as PDAs and early smartphones became popular. However, it wasn't until the advent of smartphones, tablets, smartwatches, and more sophisticated technologies that the need for specific guidelines for small-screen data visualization became more pressing.

In this context, it also arose the necessity to clearly distinguish what could be understood as mobile data visualization, as the concept of mobility began to take on an increasingly broad and fuzzy meaning [**71**].

2.2.1 Early attempts

One of the earliest efforts to establish guidelines was the work of Michael Friendly, an American statistician, which in 1999 published a paper titled "Visualizing Categorical Data" [**43**], where Friendly "outlines a general framework for data visualization methods in terms of communication goal, display goal, and the psychological and graphical design principles which graphical methods for different purposes should adhere to" [**43**].

First, Friendly asserts that in order to produce an effective graph, you need to understand its purpose and its communication goal. This is a basic guideline, but it's a fundamental rule of data representation in all types of devices even today. Another fundamental guideline expressed in this paper states that "effect ordering for data display" [**43**]. According to this principle, it is necessary to "sort the data according to the effect to be observed", as it allows to find more consistent patterns and to find outliers and catch exceptions more quickly. It also makes data visualization more enjoyable and effective.

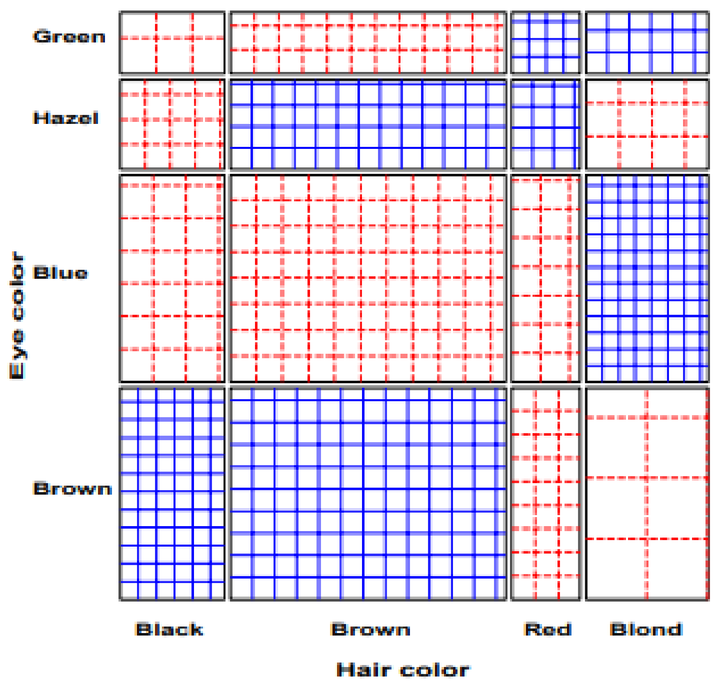


Figure 2.9: An example of visualization of categorical data extracted from Friendly's paper [43]. It shows a Sieve diagram for hair-color, eye-color data.

Even if Friendly's research was not specifically related to small-screen devices, his contributions were a fundamental starting point for that field.

In the early 2000s, the data presentation problem was emerging more and more prominently, given the exponential increase in the amount of data produced and the trend towards smaller screens of digital devices. The mobile devices available at the time were mainly PDAs and early types of cell phones; in both cases, the resolution of the screens was low (generally

240 x 320 pixels in the case of PDAs), colors were not always available and the low network bandwidth allowed only to display data stored in the device.

Therefore, the next step in the evolution of data visualization on small screens was the development of basic charts and graphs that could be displayed on these mobile devices. These charts were often simple and lacked the level of detail and complexity found in larger visualizations. Nonetheless, several tasks could be operated on such visualizations, such as overview, zoom, filter, and details exploration [**121**].

An interesting work about information visualization on small-screen devices was done by Heimonen in 2002 [**53**]. In his thesis, after providing a definition of the small-screen devices of the time, Heimonen presented a series of information visualization techniques as a solution to the presentation problem and he also presented an evaluation framework "to assess the suitability of the information visualization techniques for use on the devices in question" [**53**], highlighting the constraints imposed by the size of the display, by the visualization technique and interaction possibilities and by the implementation.

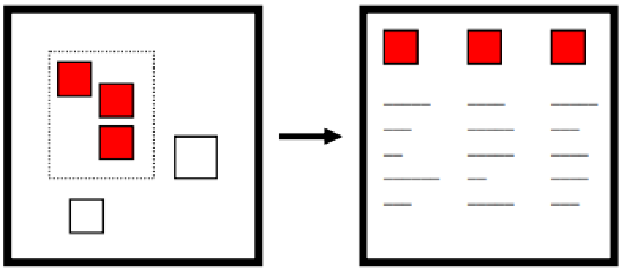


Figure 2.10: Illustration of one of the techniques devised by Heimonen for visualizing details-on-demand on small screens [53].

In 2005, Noirhomme-Fruiture et al. [**94**] presented a paper in which they proposed a series of general recommendations for small-screen devices and applied them to time series visualizations in PDAs and mobile devices. The guidelines proposed the optimization of the screen space, the use of contrast boundaries for making a shape stand out in a visualization, and even the proposal of a principle of interactivity. Even in this paper, the main representation types presented were bar charts, and some variants were proposed to improve the interactivity and optimize the screen space. An example of a possible visualization for small screens devised by Noirhomme-Fruiture et al. [**94**] can be seen in Fig. 2.11.

As mobile devices became more powerful and capable, developers began to experiment with new ways of displaying data. This led to the development of interactive visualizations that allowed users to explore data in new and exciting ways. These visualizations often included features such as zooming, panning, and the ability to drill down into specific data points. This feature was mostly used in the framework of visualization of web pages on small-screen devices, when testing web usability [**12, 32**].

Figure 2.11: Example of Alternate Bar Chart used by Noirhomme-Fruiture [94] to present an alternative to the already existing bar chart for computers.

In 2006, Luca Chittaro of the University of Udine outlined a set of factors related to visualizations on mobile devices **[27]**, taking into account the environment and the device itself. They can be summarised in six points: *visual mapping* of the data, *selection* of the relevant data among the ones available, *presentation* of the data using a limited screen space, *human factors* such as human perception and cognitive capabilities of the final user and finally *evaluation* of the effectiveness of the visualization on the users. These factors were then considered in a case study, namely an application for the visualization of POIs in a map, in order to identify the best solution to the visualization problem.

Figure 2.12: Example of three possible visualizations of POIs on a map application for small screens analyzed by Chittaro [27]

.

2.2.2 Touch-based interfaces

Another significant advancement in the evolution of data visualization on small screens was the development of touch-based interfaces. Touch-based interfaces allowed users to interact with visualizations using gestures such as swiping and tapping. This made it easier for users to explore data and provided a more intuitive way of interacting with visualizations. The advent of touch-based interfaces happened in parallel with the introduction of small-screen devices with larger displays than keyboard-equipped devices.

What is Mobile Data visualization?

In 2021, Ricardo Langner et al. published a book focused on Mobile Data Visualization [**71**], which analyses in detail all aspects regarding visualizations on several types of mobile devices. This book begins by trying to define the confines of mobile data visualization, exploring various typologies of representations and applications presented in the literature.

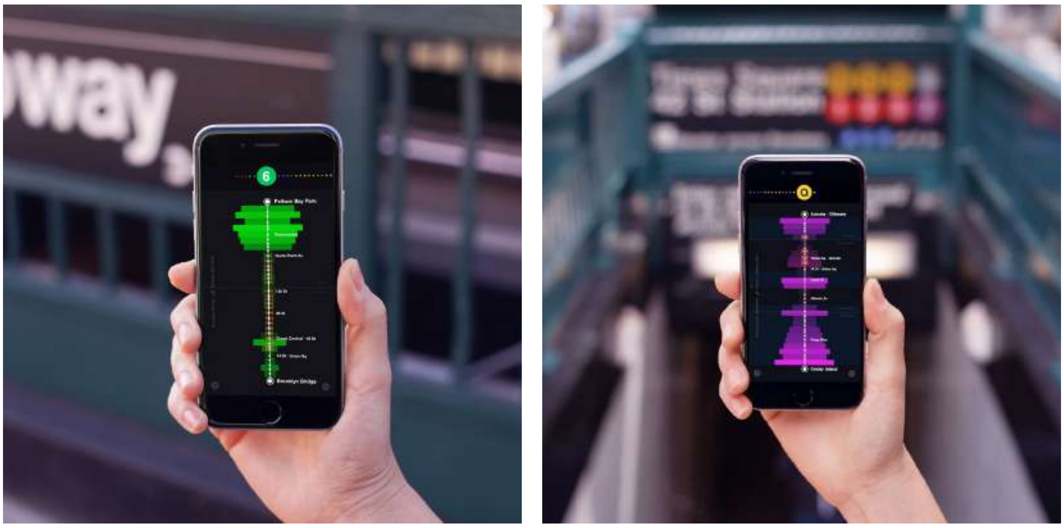
In the first chapter of the book [**68**], the authors try to define what is mobile data visualization, by presenting a series of "descriptive dimensions" [**68**] that allows to classify data visualizations. The seven dimensions consider the following aspects:

- Physical data display size: the display size of the device considered in terms of pixels (smartwatch, smartphone, tablets).
- Data display mobility: it captures the movement of the display visualizing the data.
- Data source: whether data are pre-loaded into the device, captured in real time by sensors, or a combination of both.
- Reaction of visualization to display movement: considers whether the visualization changes due to movement of the display.
- Intended viewing timespan: it considers the time available for viewing the visualization, which ranges from a sub-second glance to hours.
- Visualization interaction complexity: which level of interaction the visualization supports, ranging from none or minimal interaction to high interaction complexity.
- Intended sharing: whether the visualization is intended for being viewed by a single person at a time (generally very small displays such as smartwatches) or by multiple users (for example with a smartphone).

Each dimension is composed of several categories, defined according to specific functional parameters. Using the above dimensions, Langner et al. tried to classify the different types of devices and the visualizations themselves, presenting a series of examples of data visualizations, including both outdated (implemented for PDAs such as DateLens [**13**]) and current visualizations.

A key result of this analysis concerns the affirmation of the ambiguity of the concept of mobility: mobile data visualizations, in fact, comprehended both visualizations on devices that are movable or portable (such as smartphones, smartwatches, tablets) and visualizations that are "meant to react to viewers that are themselves mobile across devices and screens, or in space" [**68**] (such as AR and VR sets). Therefore, Langner et al. [**68**] support the thesis that the line between what is to be considered mobile data visualization is extremely blurred and that "everything is mobile depending on your frame of reference" [**68**]. To substantiate this hypothesis, in addition to examples of visualizations on common devices such as smartphones, tablets, and smartwatches, they also present some borderline cases, from "self-propelled visualizations" [**68**], which are visualizations that can bring themselves to the users rather than the users approaching the visualization, to "large movable displays". All of these examples transcend the boundaries of mobility in that they are visualizations that move in a frame of reference different from ordinary small-screen devices. They consider visualization that combines multiple tablets and vertical screens [**66**] or even considering visualizations that can move towards a static user [**73, 157**].

Related to hand-held devices such as smartphones, a key example concerns the Subspotting app [**47**], which shows the mobile phone reception along the subway of New York City. Using a previously gathered dataset of integrated and transformed data, the app returns to the user a simplified map of the track highlighting the points of the path where the network coverage is more stable, as shown in Fig. 2.13. The need for a simplified view is given by the frame of reference: the app is used on a smartphone, a small-screen device for personal use, with average interaction complexity (zooming and panning operations are allowed), and the viewing time span varies between a few seconds and several minutes.



(a) Track from Pelham Bay Park to Brooklyn Bridge. (b) Track from Astoria to Coney Island.

Figure 2.13: Examples of visualization of the two phone reception on two different tracks of the NYC Subway generated by the Subspotting app by Goddemeyer and Baur [47].

However, the key result presented is the definition of the seven dimensions shown above and their sub-categories. The possibility to categorize visualizations and devices according to the parameters defined in the analysis is fundamental in a world where technology is constantly changing and so does the frame of reference the users and the developers of visualization have to deal with when it comes to data and their presentations.

Responsive Design

One of the biggest breakthroughs in the evolution of data visualization on small screens has been the development of responsive design. The term *responsive* is generally used "to describe aspects of visualization that adapt automatically to various factors, such as changed device characteristics, environment, usage context, data, or user requirements" **[72]**.

Responsive design also allows developers to create visualizations that adapt to the *size* and *orientation* of the device being used. This meant that visualizations can be optimised for both portrait and landscape orientations, as well as for different screen sizes. As UX designer, Friedman declares guidelines for responsive web design: "As the user switches from their laptop to iPad, the website should automatically switch to accommodate for resolution, image size and scripting abilities" **[42]**.

Nowadays, responsive design is widely used in the context of data visualizations, as users move from one mobile device to another with extreme speed and frequency. Therefore, developing adaptive representations and graphs is becoming a real necessity. In addition, the development of these types of representations, combined with the integration of interactive features within the graphs themselves, increases user interest and engagement, which are two key aspects to consider in the context of data visualization.

However, responsive design still presents several challenges as mobile devices are heterogeneous in size and capabilities even in the same class (usually smartphones have an aspect ratio of 16:9 but there exist also smartphones with different aspect ratios such as 3:2 or 21:9). Therefore it is difficult to design one-size-fits-all visualization solutions.

These challenges are thoroughly described by Horak et al. **[72]**. First of all, the analysis focuses on the factors that must be considered when developing responsive data visualizations for mobile devices, which can be summarised in the following categories: *device*, *usage*, *environmental*, *data* and *human*. All these factors an be inferred by the devices using internal sensors such as cameras, environmental, motion and position sensors, clock and connection standards (Wi-Fi, Bluetooth, etc.).

A number of strategies are then presented to transform desktop visualizations into representations suitable for mobile devices. These strategies, also shown in Fig. 2.14 concern:

- The scale of the content.
- The aspect ratio of the device (portrait or landscape mode).

- The layout of the content (e.g., image + text, etc.).
- The level of detail.
- The amount of data.
- The removal of annotations and guides such as axes, grids, text, and marks and adding dynamic guides, interactive features, animations, and the possibility of streaming data.
- The possibility of changing the entire visual encoding of the data.

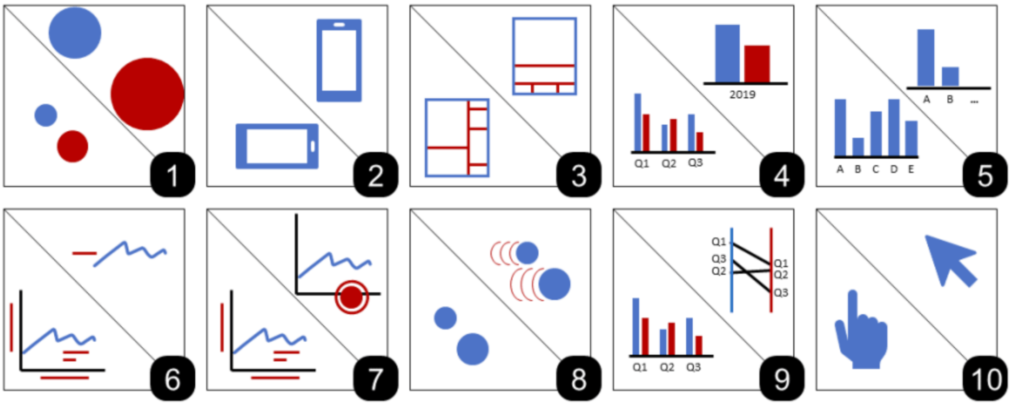


Figure 2.14: The ten strategies for responsive design proposed by Tom Horak and Tableu's Matt Brehmer in Chapter 2 of the book Mobile Data Visualization [72]

Therefore, Horak et al. attempt to provide a set of guidelines for mobile visualizations by considering all types of functional and non-functional parameters existing for these devices and the environment they come in contact with. However, these guidelines are devoted to the passage of content already created for desktop devices to small-screen devices.

The following discussion by Brehemer et al. **[19]** focuses on user interaction with visualization in mobile devices, exploring the different possibilities (haptic, spatial, and vocal), and the related challenges and opportunities. The book provides a series of examples for each modality.

Similar guidelines were also presented on the "Visual Cinnamon" website **[141]**, where is proposed a set of strategies to shift desktop-based representations to a mobile version. As shown in Fig. 2.15, an example is given for each guideline, in which the desktop version is compared to the mobile one. The analysis goes deep even in presenting the technical details of how charts are created in Javascript.

The guidelines presented by Visual Cinnamon amount to:

- Do nothing;
- Replace the visual function that creates the chart with a loaded image;

- Make the position of the elements arbitrary;
- Scale down the entire chart;
- Fit to the available width, maintaining the same height;
- Stack vertically the data on the screen;
- Consider using zoomable visuals and the range for starting zooms;
- Reposition the data in different locations;
- Change charts completely.

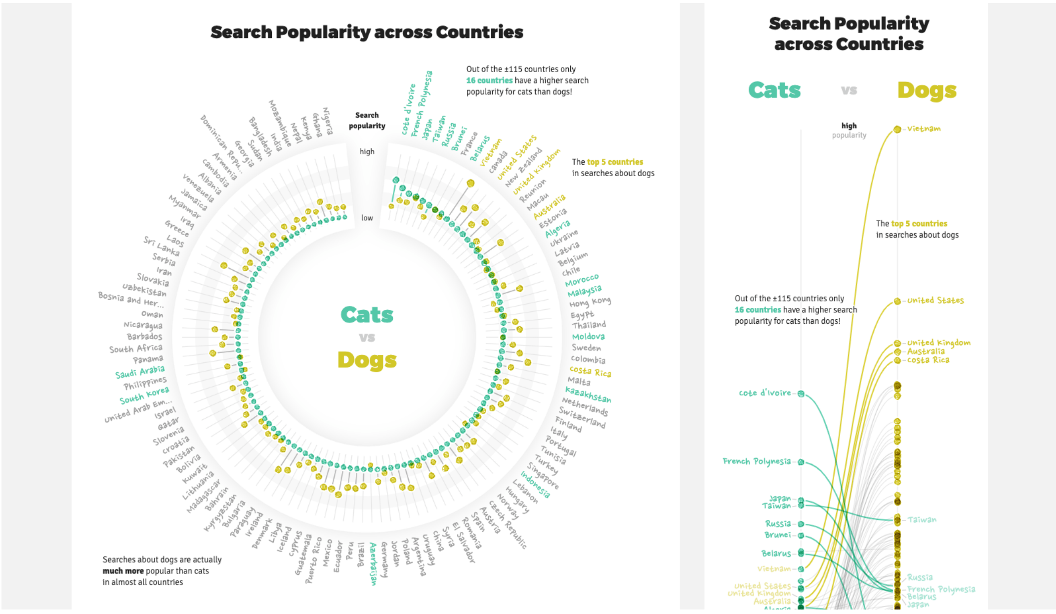


Figure 2.15: An example of a completely changed chart, when passing from desktop to mobile [141]

The team behind the Visual Cinnamon website emphasizes that there is no one set way of going about the "mobile vs. desktop" data visualization debate. A number of scientific conferences and articles are starting to propose their strategies and guidelines [**2, 83**]. There are open-source projects about patterns in data visualizations for small screens that allow users to post their personal opinions about specific graphs [**111**].

Another example is undoubtedly the checklist of guidelines proposed by the European Environment Agency (EEA), which presented a series of dos and don'ts of data visualization, summarised in the checklist in Fig. 2.16 below.

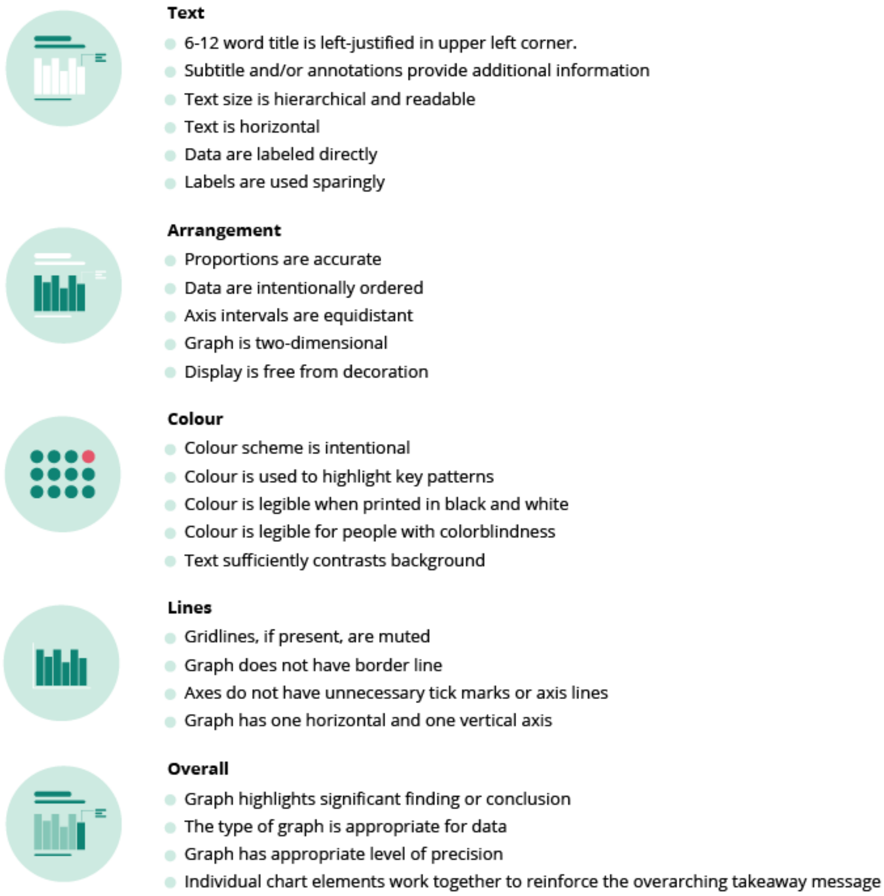


Figure 2.16: Checklist of guidelines for data visualizations proposed by EEA [37]

Archie Tse, a graphics editor of The New York Times, presented **[136]** several reasons for the low usage of interactive visualizations and responsive design, especially in journalism, and illustrated three tips for compelling visual storytelling and creation of data representations. These three guidelines can be summarised as the absolute necessity to show important information, not hide them behind a tooltip, prioritize scrolling over clicks, and limit interactivity because it is expensive to get it to work on all platforms. In fact, according to this presentation, **most visualizations used in journalism or websites are static**.

Meg Miller also addressed the problem with interactive graphics in a Fast Company article [**88**]; she highlights how 85% of New York Times readers ignore interactive graphics. Therefore, when using visualization to reach a mass audience, interactivity should be implemented so that users can extrapolate the data they are most interested in. In this way, interaction transforms a general visualization into a personal one.

Other works in the field of guidelines for mobile data visualization focus on specific cases, ranging from cartographic visualization [**120**] to large time series data [**25**]. Even big companies such as Apple are starting to provide guidelines for creating charts and graphs for their devices [35].

Wang Baldonado et al. [142] proposed a set of eight guidelines for using multiple views in Information visualization, where *multiple views* concern the investigation of a single conceptual entity using two or more views. Analyzing multiple-view systems such as CAD, the authors provided the following guidelines, which can be easily generalized for small-screen devices: *diversity*, *complementarity*, *decomposition*, *parsimony*, *space/time resource optimization*, *self-evidence through perceptual cues*, consistency, *attention management*.

An interesting work was proposed by Brehmer et al. [18], where they analyze possible visualization techniques for data ranging over time on mobile devices. According to Brehmer et al., data over time are the most visualized data by users of small displays, but there is a lack of guidance on how to represent them efficiently and clearly. Therefore, they conducted a task-based crowd-sourced experiment, presenting the audience with two ways of representing ranges of times, i.e., *linear* and *radial*. As a result, they noticed that people involved performed tasks more quickly with linear visualizations, but the error rate between the two types was similar. From a design perspective, the authors conclude that users are more confident with linear layouts and that the scientific community should move to produce guidance about visualizing data on mobile phones by performing more structured crowd-sourced experiments.

Finally, the book "How charts lie" by Cairo [22] provides a set of steps to read a graph, while also highlighting what makes a good chart. According to the author, a good chart is a visual argument and therefore, it should be seen *and* read. The features surrounding the visual encoding of the data are as important as the content of the graph itself, as they might explain important information about the data and how to read the chart itself. The steps identified are in order:

- read the title, the source, and the captions on the graph; this additional context is usually crucial for understanding the visual content;
- read Measurements, units, scales, and legends; as above, they contain important information on how to read the visualized data;
- check the methods of visual encoding. such as colors, position of the elements, and so on;
- read the annotations, which are generally added to emphasize the most relevant points of the chart;
- try to spot patterns, relationships, and trends in the data.

Even if his book is not directly correlated to small-screen devices, it provides many insightful guidelines for producing and reading effective charts that can be applied to the case considered, highlighting as a final takeaway that designers should not produce charts more complicated than they should.

Data visualization and Social media

Some attention was also drawn to the specific case of social media, as they have a potential for communicating data, in particular in visual-based platforms such as Instagram.

In 2011, Flaminia Gallo [**45**] presented an interesting analysis of the correlation between data visualization and social networks, showing how these platforms were fundamental to describing modern society, which is fast and changing. According to her study, data visualization is a key tool for representing the enormous quantities of data that social media produce every day and could be used as an educational tool.

In a post of early 2023 [**137**], Flourish Studio posed the question of whether data visualization has a potential for social media, as images, GIFs, and videos posted on the platforms lack interactivity and that such apps are mostly used with small screen devices.

Before presenting guidelines for data visualization on social media, their analysis considers the interest of users for this type of representation. Social media are the expression of our fast-paced society, where the span of attention of the user is very low and it is necessary to be able to tell a story as quickly as possible. When talking about social media data visualization, we enter in the realm of *glanceable visualization*, widely described by Blascheck et al. in Chapter 5 of Mobile Data Visualization Book [**14**]. As the user's eye might stop for just a few seconds on the image it is necessary to create representations that can be easily understood while on the go and that can simultaneously trigger some behavior changes in the viewer. Data visualization is the perfect tool for this goal as, following some directives, it can effectively transform complex data into easily readable graphs. Therefore, sharing data as visual representations can be an efficient way for informing people about relevant social issues.

The article proposes a set of guidelines that emphasize the need for simplicity and visual clues in representation, avoiding abounding with text and visual elements and ensuring that those available are recognizable and original. The article also emphasizes how the use of GIFs or videos showing composed data is much more effective than static images.

In 2020, an article [**6**] by the finance journal Economist presented the set of guidelines used by the editorial team to create the charts and graphs used in their Instagram page. Again, the focus is on simplicity and the use of colors, paying attention to the target audience. Interesting is the suggestion to limit the use of graphics when not needed and to give importance to the text.

The same indications were proposed by an article on the website Whatagraph in 2021 [**145**]. The article tried to go even deeper into the analysis started by The Economist, integrating users' brain processes, such as reading from left to right, as important factors to consider when creating data representations.

In 2022, data communicator Andy Cotgrave published an article presenting six tips for creating charts on social media [**33**]. Such tips highlight that titles of graphs should be *descriptive or provocative*, underlining the need for *simplicity*, the role of *animation* through GIFs and videos but also the need for *originality*. However, the most important contribution is the indication for ***accessibility***, by underlining the importance of ALT text and solutions for visually impaired people, a concept also analyzed by DataJournalism.com [**113**]. Another key aspect presented was related to *trust*, highlighting how it is always necessary to add the data source, in order to increase trust and avoid the stigma of fake news or message distortion when posts are shared.

In conclusion, the field of data visualization on social media is still in its prime but it is becoming increasingly overlooked by the scientific literature, as social media are becoming a more important tool for information visualization diffusion every day. The main problem stands in the reliability of such platforms, as fake news and data are diffused every day and mine the trust of users towards legit scientific publications.

Challenges of everyday use

An important aspect of mobile data visualization concerns daily use and the challenges it brings. Epstein et al. [**36**] explore such aspects, analyzing different cases of daily usage and providing design solutions to the challenges encountered.

As the authors state, the average user often comes into contact with data and news via mobile devices, but this does not always mean that they have the right level of literacy to understand it or the time to delve into it. On the contrary, very often non-expert users can be misled by data representations created without considering their skills or find themselves overloaded with information.

According to Epstein et al., designers should first consider *which information is worth visualizing* and the *level of interaction* needed, as small screens are generally used while in motion or in other settings other than sitting on a desk. Sometimes, passive and static visualizations are more suitable than active ones in these settings.

Moreover, designers should not assume users have particular skills and high level of literacy, leaning towards simple and easy-readable representations of data.

The authors also consider the ethical challenges posed by data visualization. A low level of literacy and the use of mobile devices as the only source of information can lead to the rapid spread of false information if data are presented in a misleading fashion. It is therefore the duty of designers to develop visual representations that prevent the spread of misinformation. Two blatant examples of this phenomenon can be found in the spread of false news or misconceptions about climate change and the Covid-19 pandemic; the impact of the spread of misleading or false data via social media has contributed to the creation of an erroneous and truth-blind narrative.

A cause can be found in Mobile devices; due to their small screens, designers are forced to leave some data details out of representations, which may lead to missing contexts and incomplete information, and uncertainty. Epstein et al. [**36**] state that as diffusion of scientific knowledge and presentation of data may lead to behaviour changes, mobile devices, and data visualization are rising to become fundamental tools in education and scientific research. For example, they can have a strong impact on important socio-scientific issues such as environmental sustainability and climate change [**44**] or diffusion of pandemic data.

For example, in 2020, Trajkova et al. [**135**] analyzed 5409 tweets including visualizations regarding the Covid-19 pandemic data, in order to understand the factors that lead to the retweet of such visualizations. They concluded that the sharing of visual information is determined by a series of factors: source of the data, creators of the chart, type of visualization, and variables shown on the chart.

Therefore, designers have to create a visualization in a responsible way, to "challenge unethical use of data and data visualizations" [**31**], improving education about the discipline and bringing the world to a democratic use of data, while balancing societal needs with people's privacy and freedom.

2.2.3 Conclusion

Overall, the evolution of data visualization on small screens is a fascinating journey that has seen significant advancements and applications over the years and it is now expanding beyond the realm of experts exploiting social media and at the same time enriching the classical definition of mobile visualization with types of representations beyond small-screen devices.

An interesting advancement is surely the usage of Machine Learning and Deep Neural Networks as a way to classify visualizations by type, identify colour mappings, visualizing large datasets, solve issues in mobile web representations, and much more [**52, 155**].

However, no clear and unique set of strategies can be identified. In particular, in relation to the guidelines for small screen devices and responsive design, the literature is just beginning to expand to cover a number of different use cases and contexts, although it is still very much geared towards expert users. There is a lot of interest in workshops whose target audience is experts and researchers [**28, 70, 144**], the purpose of which is to bring together experienced users to carry out experiments and studies to outline possible guidelines and test them on the spot. However, there is very little literature dedicated to non-expert users.

As the domain and demand for data visualization continue to evolve, we can expect to see even more exciting developments in the field of Information Visualization on small screens and possibly an expansion of literature regarding guidelines for this type of representation, as it is requested more and more every day.

2.3 Social media and scientific knowledge dissemination

As stated in the previous section, in recent years social media have become a fundamental platform for communicating scientific discoveries and data to a general audience. This is due to the algorithms they are based on that enable fast sharing of news, data and articles in real-time.

In particular, younger generations take advantage of this ubiquitous media to be informed about what is happening in the world around them rather than reading scientific articles or printed journals **[1]**. However, according to this study, a large percentage of young users who use social media as an information tool use it as their only information source (42.8%). The percentage of cross-media users, who use also on other platforms is slightly lower (35.9%).

This leads to two major problems, namely the emergence of eco-chambers[4] and the spread of *fake news*, by which, however, young people are not the most affected; in fact, according to the Ipsos Report of 2022 **[61]**, the younger generations and people with higher levels of education tend to control the reliability of online information more frequently than older people. Moreover, there is a surprisingly high degree of scepticism regarding scientists and science: according to the same report cited above, Italians have a positive view of science but more than 25% of the population believes that science is not objective and is not oriented toward the common good of society but is more prone to support elite interests.

Therefore, communicating information and data on social media is a risky operation if researchers are not able to convey data in a clear and easy-to-understand way. Some studies have been conducted on the efficiency and reliability of social media as a tool for scientific and socio-scientific knowledge diffusion. Most of the work conducted on data diffusion on social media is related to social network analysis, which is a powerful method to understand how social media algorithms work and how users contribute to the diffusion of data and information on these platforms **[74]**, how such data can improve the awareness about different fields such as urban sustainability **[58]** and understand people's sentiment on different socio-environmental topics **[76]**.

Analysing the existing literature is possible to affirm that there is a limited discussion that specifically addresses the type and role of social media in propagating scientific knowledge, especially environmental sustainability awareness. However, over the last years, the interest

[4]In news media and social media, an echo chamber is an environment or ecosystem in which participants encounter beliefs that amplify or reinforce their preexisting beliefs by communication and repetition inside a closed system and insulated from rebuttal [148].

is rising and there is some interesting research about how the different platforms can factor into the diffusion of socio-scientific topics.

An interesting paper was published in 2011, which focuses on the potential of blogs and social media in shortening the peer review process, which is generally a long and exhausting work [**78**]. The possibility to present a scientific study to the general public on social media is a double-bladed knife as non-expert people could oppose scientific works without even knowing the topic they are discussing and the tone of the conversation is quite different, more intimidating than the one used in the scientific community. The need for moderators and norms is therefore a key requirement for this process to work but the scientific community is divided on this issue.

In 2017, an interesting study was conducted by Hamid et al. n relation to the use of social media for increasing environmental awareness in higher education settings [**51**]. According to the findings, social media can be an incredibly powerful tool for extending knowledge and awareness of environmental sustainability and promoting responsible behavior policies to both staff and students, in order to transform into fully green universities. Their literature review confirms the lack of studies on the usage of social media as a platform for raising awareness on sustainability issues and states that this is probably due to the fact that social media lack "legitimacy" and are viewed as a platform for "informal communication" [**7, 67, 129**]. However, the authors urge a change of this viewpoint and exploit the potential of social media.

In 2018, some researchers studied the impact of the tweets containing information and data regarding the meeting in Paris for the Conference of the Parties (COP21), where representatives of nations set and discussed the objective of reducing GHG emissions to zero by mid-century. The study underlines that using the platform for diffusing data and news helped draw attention to climate change issues [**55**].

There is much interest in using social media to diffuse health information to increase health research. The work of Bardus et al. [**10**] conducted in 2020 aimed to find a correlation between the use of social media to disseminate health research and the increase of their bibliometric values, but it showed inconclusive results, calling for more testing in the field. Such lack of correlation was found in a later study by Yasemin Özkent [**102**], which analyzed the exposure of 10 scientific articles after their publication on social media and found that their visibility was increased concerning articles not published on social media.

With the advent of the Covid-19 pandemic, the speed at which information is shared with the public drastically increased, and the same happened for the diffusion of fake news on social media. People worldwide wanted real-time news and data regarding health policies and the impact of the virus and vaccines. Social media became the first information outlet due to the rapid spread of information they allowed, leaving traditional outlets behind. An interesting work by Akiko Iwasaki **[63]** showed how to promote science on Twitter in the most efficient way to gain reliability.

Another exciting work was conducted by Imperiale and Casadevall of mSphere[5] in 2022 **[112]**, where they considered advantages and disadvantages of the use of social media for scientific dissemination. They argued that social media may not be suited for scientific discourse due to the heterogeneity of users who may be engaged. On the other hand, they recognized the power social media hold in amplifying the impact of scientific information and the target audience. Therefore, the scientific community must find a way to use it responsibly and effectively.

Moreover, many online journals ad magazines use social media to communicate data and news. In particular, Instagram is also widely used for the dissemination of data in visual form by major news outlets, such as the Italian *La Repubblica*, *il Post*, *Will Media*, and many others. This is due to the fact that Instagram is an image-based platform, and posting simple graphics about social, economic, and scientific data proves to be very effective in increasing user awareness and attracting people **[2, 6]**

The data gathered by the Mauna Loa Observatory **[81]** itself about the concentration of CO_2 in the atmosphere and its steady increase over the years are diffused on a Twitter account **[93]** managed by a Canadian agency in a daily fashion, also publishing structured data so that people realize the change that is happening to our planet's climate.

Many other studies could be discussed on the use of social media for diffusing scientific knowledge **[15, 77, 101, 151]**. However, we can conclude that the literature's analysis on the topic states that social media is an extremely powerful tool for diffusing scientific information, as it allows to reach wider audiences and increases the visibility of data shared on the platform. However, the phenomenon of fake news means that social media lacks legitimacy and is seen by the scientific community as a risky platform that needs to be learned well how to use effectively.

[5] mSphere® is a multidisciplinary open-access journal that publishes high-quality work that makes fundamental contributions to any of the immense range of fields within the microbial sciences.

Chapter 3

Project Description

This chapter describes the work carried out within the project, with particular attention on the requirements and goals set, the division of labor among group members, and the specific work done in the area of the stationary environmental data gathering stations.

3.1 Introduction

As stated in Chapter 1, the work described in this thesis is part of a project set out by the DAIS of the Ca' Foscari University of Venice in 2022.

The task assigned to the work team was to assemble an environmental data collection system that could be deployed within the city of Venice to collect data regarding the concentration of carbon dioxide in the atmosphere and carry out various types of analysis on such data.

The starting point of this project was the Life Redune project, a "5-year-long project co-financed by Regione Veneto and European Commission whose goal is the restoration of selected dune habitats of the Veneto coast" **[21]**. To meet the objectives set up by Life Redune, an environmental data collection system was developed, tested, and deployed in the coastal habitat of Veneto region [6].

Since such a system was still functioning when the new project was initiated, and it met several requirements of the new task, it was used as a starting prototype for assembling the new system and devising improvements aligned with the new context and purpose.

An important requirement of the new project is the collection of environmental data through two similar yet different systems:

- The first analysis systems consisted of a series of stationary data-gathering units, a slightly upgraded version of the stations of the Life Redune Project. Such stations

[6] See section "Life Redune Project" in Chapter 2 for the description and technical requirements of such project

were deployed in selected locations owned by the Ca' Foscari University across Venice.

- The second data collection system was designed based on the Life Redune data collection system architecture but has been designed to collect data as they are transported around the city.

Following the assembly, testing, and deployment phase of the two systems, the collected data were analyzed in different ways by working group members in the context of the discipline known as *Information Visualization*. The ultimate goal was to disseminate the data collected and the analysis performed on them through visual representations with different application modalities that appealed to diverse audiences. Furthermore, the scope of this process was to raise awareness regarding environmental topics and the conditions of the city of Venice in non-expert audiences.

The work described in this thesis focuses only on the stationary system and the data collected by the static units located on the island of Venice. The portable system will be used as a reference, but its discussion does not concern the analysis described in this dissertation.

3.1.1 Case Study: The city of Venice

As stated above, the first step in the project context was identifying a new location of interest for the environmental analysis process.

The case study chosen was the city of Venice, located in northern-east Italy, specifically in the Veneto region. The reasons for this choice are the following:

- First of all, Venice is the city that hosts the Ca' Foscari University as well as the capital of our province and our region and therefore has symbolic importance.
- The lagoon of Venice hosts a rich and varied ecosystem, constantly threatened by human presence and activities. The increase in concentration levels CO_2 in the atmosphere and consequent rise in temperatures pose a risk to the survival of autochthonous plants and animals in the lagoon.
- A large part of the environmental pollution in Venice is due to the presence of ferries and boats that continuously pass through the lagoon. The environmental monitoring process implemented by the work team would allow us to determine how influential this phenomenon is on the ecological condition of the lagoon's atmosphere.
- The dissemination of the data obtained from the environmental analysis conducted by the work team could significantly impact the perception that both the local and external populations have regarding the pollution of the city of Venice. It could therefore be a starting point for implementing protocols to safeguard the town and the fragile but fundamental ecosystem it hosts.

Therefore, the monitoring systems must be implemented in a way that respects the environment in which it is located and adapts to the conditions it poses.

3.2 Requirements

Since the work described in this thesis focuses only on the data collected by the stationary stations located on the island of Venice, the requirements discussed in this section will only concern the stationary system.

The new task, initiated from the equipment, the premises, and the results of the Life Redune project, involved meeting several new requirements established early in the process to make the system and the data analysis compliant with the new context and purpose. Therefore, the requirements described below interest are both functional[7] and non-functional[8].

The functional requirements are the following:

- The system has to gather environmental data regarding levels of *concentration of* CO_2, *humidity*, *temperature*.
- Each data entry collected must have a timestamp attached to identify the day and time it was collected and be geo-located.
- The data must be gathered in real-time and stored on an online server to be accessed remotely.
- Data should be collected at regular time intervals (e.g., every 15 minutes).
- Data must be collected throughout the day, so the system must always be active.
- The online server and the storage process must be *secure*: data must not be susceptible to interception while sending from the system to the server, and the online server must be accessible only to those with authorized credentials.
- The system must overcome possible exceptions and errors, applying reboot strategies to ensure the system will not block or go into an overload/overflow state.
- The system must communicate its functioning daily by sending an e-mail to the members of the working team and the stakeholders.

[7]Functional requirements are presented as lists of features or services that the system must provide; they describe the behaviour of the system in the face of particular inputs and how it should react in certain situations.

[8]Non-functional requirements represent the constraints and properties/characteristics related to the system, such as time constraints, constraints on the development process and on the standards to be adopted, including the method used to develop the system.

As for the non-functional requirements, they concern the constraints imposed by the system and by the context in which it is deployed. For example, considering that the locations of deployment of the system are six seats of the Ca' Foscari University located on the island of Venice and the urban city of Mestre, the requirements must consider factors such as weather conditions, tide level in the city of Venice and passage of unauthorized people, such as students and tourists near the locations.

Therefore, the non-functional requirements are the following:

- **Accessibility**: the system must be protected from unauthorized access. Therefore, the chosen locations must prevent anyone besides the working group from accessing the units. The same is true for the storage space, which must be accessible only to authorized people.
- **Performance**: The system must be able to handle the required number of data samples for all established variables.
- **Availability**: The system must be available when needed, specified by the functional requirements. Therefore, it must be powered in such a way that ensures connectivity and power at all times.
- **Maintenance**: The system must be easy to maintain and update. Therefore, the units should be placed in accessible locations for the working team to retrieve, maintain and replace them, if necessary. Moreover, the software itself should be maintainable in an easy way.
- **Reliability**: The system must be reliable and meet the established requirements.
- **Usability**: The system must be easy to use and understand so that others can use it once the working team's contribution is terminated.
- **Compliance**: The system must comply with all applicable laws and regulations.

3.3 Work organization

To create a system that met the above requirements and produced a consistent and usable data set for satisfactory and scientifically consistent environmental analysis, it was necessary to divide the work among the working group members.

As stated in Section 3.1, the project is divided into two fundamental parts. The first part is focused on gathering environmental data through stationary data-collection stations located at specific places in the city of Venice. The second part aims to carry out the same process using similar data collection stations but designed to collect data as they are carried around the city.

This thesis focuses exclusively on the first part of the project, detailing all the stages of work and implementation of the hardware and software of the data collection system.

In addition, the discussion will then be shifted to the analysis of the collected environmental data and the dissemination process of such data in the chosen context, namely through visual representations suitable for small screen devices, specifically for social media.

The work done in the context of this thesis to meet the requirements described above concerns:

- Analysis of the architecture of the stationary data-gathering system.
- Analysis and development of software dedicated to operating stationary environmental data collection stations using Arduino Nano and a set of environmental sensors.
- Software development and testing for sending and receiving data from Arduino Nano to Google Sheets.
- Software development for sending daily messages notifying the operating status of each collection station.

3.4 Architecture and functioning of the system

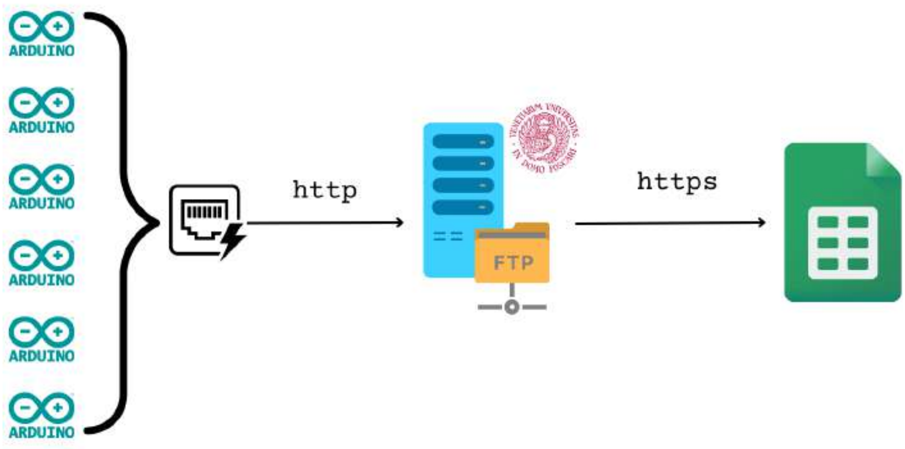


Figure 3.1: Architecture of the data collection system and data workflow

As stated in the previous sections, the system is an upgraded version of the one developed for the Life Redune project, which was modified to meet the new requirements. In particular, it was necessary to add an endless and reliable power source and a consistent and reliable network connection to transmit data in real-time, features that the previous design lacked.

The choice fell on the **Power over Ethernet (PoE)** technology, which enables Ethernet cables to provide power to network devices via the existing data connection[9].

Therefore, taking into account the requirements described in Section 3.2, the working team designed and implemented the system architecture and the data workflow in the following way (see Fig. 3.1):

- The sensor nodes consist of six data collection units controlled by an **Arduino Nano microcontroller** and composed of several environmental sensors and other types of modules (See Section 3.4.1 for technical details). The sensors collect the environmental data, and the controller sends a request to the server using HTTP protocol to send the data.

- Data are sent through an HTTP request to a proprietary **server of the Ca' Foscari University**. Using a PHP script (uploaded to the server using sFTP[10]), the server receives the data and verifies their validity. After this operation, it authenticates to Google and sends the HTTPs request to the Google Sheets endpoint, where the data will be stored.

- Data are stored on **Google Sheets** categorized by gathering station. A new sheet is created each month to make the data easily readable and usable. From here, authorized users can access the data and download it in the format most appropriate for the analysis they will need to perform.

3.4.1 Hardware

The collection units that compose the system were designed and assembled starting from the Life Redune units. Considering the requirements to be met and the architecture described above, it was necessary to upgrade the hardware, removing the components that were not useful anymore and adding new ones. In particular, it was required to:

- reduce the size of the unit while maintaining the structural integrity of the box, which would be deployed in an environment exposed to adverse weather conditions;

- change the power source from solar panels to PoE technology to ensure constant and reliable connectivity and power.

After selecting the new components in cooperation with the researchers of the Environmental Sciences, Informatics, and Statistics Department of the Ca' Foscari University of

[9] PoE is a networking feature defined by the IEEE 802.3af and 802.3at standards that pass electric power along with data on twisted-pair Ethernet cabling. [91]

[10] File Transfer Protocol (FTP) is a network protocol for transmitting files between computers over TCP/IP. sFTP is a subset of SSH protocol and provides a mechanism within SSH for secure file transfer. [64]

Venice and considering the requirements above, the proposed solution was designed using the following hardware components:

- an **Arduino Nano V3 ATP328 Mega USB CH340**, a physical programmable circuit board that contains the code that allows the system to work;

- a **K30 10,000ppm $\mathbf{CO_2}$ Sensor**, which gathers data regarding the *concentration of Carbon Dioxide* in the atmosphere;

- a **AM2315 - Encased I2C Temperature/Humidity Sensor**, which gathers data regarding *temperature* and *humidity*;

- a **ENC28J60 Ethernet Shield** allows the microcontroller to connect to the network via a wired Ethernet connection.

- an **active PoE splitter** compliant with IEEE802.3af, with an output voltage of 5.2 V and maximum output current of 2.4 A. This component allows the PoE signal from the wired connection to be transformed into electricity and an Ethernet connection to power the Arduino and enable connectivity via the shield.

The final solution was assembled as shown in Fig. 3.2.

(a) Exterior appearance of the final data collection unit.

(b) Interior appearance of the final data collection unit.

Figure 3.2: The final and approved prototype of the data collection unit.

Each unit is then connected to the network via an Ethernet cable, the ends of which are respectively connected to the PoE splitter inside the station and the PoE outlet inside the building on which it is placed.

3.4.2 Limitations and security

As discussed in Chapter 2, low-cost data collection systems present a series of limitations. These are principally related to the maintenance and calibration of the sensors assembled, the communication protocols, and their power consumption [**56**].

Because the system described here is based on low-cost components, it suffers from the same limitations but with some mitigating factors. First, since the sensors had already been used during the Life Redune project, they were already calibrated to the expected standards. Therefore, it was sufficient to conduct tests to confirm the validity of the measurements.

Third-party server for HTTPs requests

One limitation that needed to be overcome was the inability of Arduino to send HTTPs requests. This technology only supports HTTP calls, but to meet the security requirement and prevent possible cyber-attacks in transmitting data between Arduino (client) and Google Sheets (server), it was necessary to add a third party that would allow the HTTP call to be converted to HTTPs.

After evaluating several external solutions, the choice fell to using a Ca' Foscari University proprietary server as an intermediary server. In this way, the data travels exclusively within the University's network: Arduino makes the HTTP request to the Ca' Foscari server, which, as explained earlier, checks the validity of the data and the request itself and then sends it via HTTPs to the Google Sheets file once authenticated to the service with a University account (which in turn is guaranteed to be secure through the use of two-factor authentication).

Ethernet Shield limitations

Other adjustments had to be made due to the limitations of the ENC28J60 Ethernet shield. In addition, this component has two manufacturing defects: The first is a hardware issue, as the D12 pin was erroneously grounded, preventing electricity from properly powering the device. The second is a software defect that results in the component entering a deadlock status and being unable to send data. Finally, this is due to a library error, as the internal transmit logic is not reset when it should, leaving the system hanging.

Both limitations were overcome; the first required manual adjustment of all defective pins. The software problem was solved by adding a portion of code that resets all controller registers daily, returning them to a clean state. In this way, each day, the system reboots, preventing the deadlock state from occurring or terminating it in the exceptional case that it does happen. The reset procedure was also included to handle system exceptions.

3.5 Gathering process and dataflow

As stated in the previous sections, the data flow from the collection station to Google Sheets goes through several controls and is governed by several scripts, as shown in Fig. 3.3.

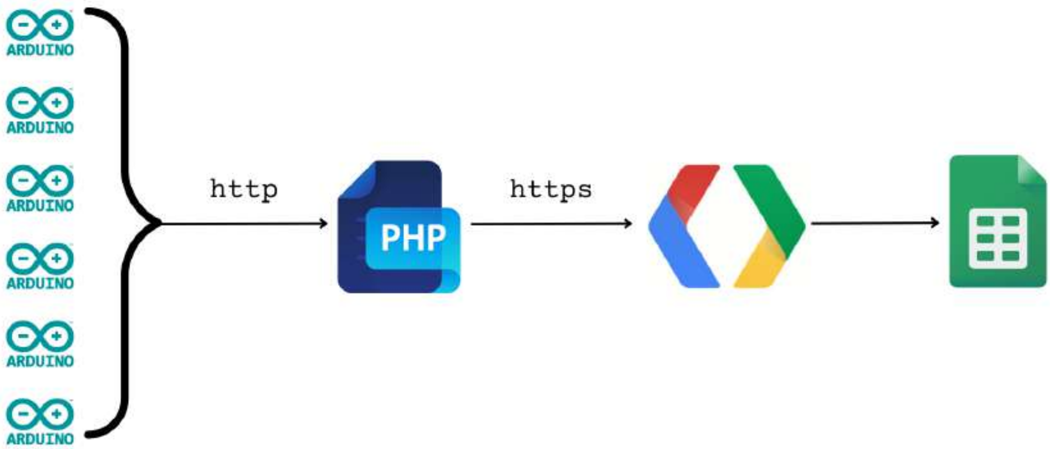


Figure 3.3: Flow of the data in the system

The system was developed with each station working as a single master module. There is neither communication between stations nor sharing of resources. Therefore, the failure of one station does not imply anything to the others, which can carry on with their operations.

Therefore, each Arduino Nano microcontroller runs the same script; the only difference between the scripts can be found in the following three variables:

- the **IP address**: each unit is identified by a different IP address belonging to the same sub-network (which belongs to the Ca' Foscari University of Venice);
- the **MAC address**: each unit is identified by a different MAC address to distinguish the devices in the network;
- the **site** and correlated **coordinates** that identify the unit based on the placement.

The *native Arduino script* first performs the setup of the sensors and Ethernet shield, creating a socket with the network parameters (MAC address, IP, DNS, and Gateway), after which it starts a loop in which it records incoming data from the sensors and tries to send it via HTTP request to the University server. The data is sent if the connection is successfully established, and the program waits a predefined number of minutes (i.e., 15 minutes in the current setup) before performing a new measurement. The system will retry the procedure if the link is not established correctly. If the request is not sent after a predefined number of attempts (i.e., seven attempts in the current setup), the program resets the system by cleaning all the registers, and the process starts over. As an additional safety measure, this reset procedure is performed every 24 hours to ensure the controller, Ethernet shield, and sensors do not overflow due to prolonged usage.

Each station sends to the server the data measured as shown in Table 3.1.

Parameter	**Unit of measurement**
Timestamp	YYYY/MM/DD, hh:mm:ss
CO_2 concentration	ppm
Relative humidity	%
Temperature	°C
Site	-
Latitude	°
Longitude	°

Table 3.1: Parameters collected and respective unit of measurement.

When the HTTP requests arrive at the server, they are handled by a *PHP script* that contains a list of all authorized IPs, which correspond to the ones of the six units, and it checks if the received requests come from one of these IPs. If so, it filters the data, checks their validity, creates the Google Client, and uses the Google API to authenticate the university user. Finally, it sends the data to the Google endpoint using an HTTPs request.

Finally, the data are received by an *Google App Script* application, which opens the current active spreadsheet on Google Sheets, parses the data (also adding a timestamp to the measurement), and stores them in a new row at the end of the sheet inside the spreadsheet that corresponds to the unit collection of origin. The application creates a new spreadsheet subdivided into six sheets on the first day of each month. A minor limitation of this data flow is the difference between the measurement time and the arrival time of data at the App Script. In fact, due to delays in the connection and data passing through the various servers, the stored timestamp will correspond to the arrival time.

Appendix B contains the three scripts (Arduino, PHP, and App Script).

3.6 Locations

A fundamental part of the project regards placing the six stationary data collection units. Since the project was founded and carried out by the Ca' Foscari University of Venice, the choice of the locations fell on six seats of the University itself.

Five units were deployed in Venice, in the Dorsoduro district (see Fig. 3.4a). The last unit was placed at the Ca' Foscari Scientific Campus in Mestre (VE) (see Fig. 3.4b). The chosen locations had to meet two essential requirements: exposure to outdoor air and access to an Ethernet/PoE network outlet.

(a) The locations of the units placed in the island of Venice.

(b) The location of the unit placed in Mestre.

Figure 3.4: The location of the data-gathering system.

3.6.1 Scientific Campus - Mestre

The location chosen for the unit placed in the Scientific Campus was the roof of the GAMMA Building. GAMMA hosts university research laboratories with restricted access; however, it was possible to access the roof as placing the unit in that place would not undermine any activity inside the building.

The roof is easily accessible only by authorized personnel, and it was provided with easy access to an Ethernet point to guarantee connection to the unit (see Fig. 3.5).

This unit is interesting as it is the only unit of the project located on the mainland.

(a) Aerial view of the GAMMA building at the Scientific Campus.

(b) The unit on the roof of GAMMA building.

Figure 3.5: The data-gathering station placed in Ca' Foscari Scientific Campus.

3.6.2 Ca' Foscari - Venice

The site chosen for the unit to be placed at Ca' Foscari is a window on the small outdoor terrace by the "Porta d'Acqua" of the Exhibition Spaces (see Fig. 3.6).

This terrace is not fully exposed to the open air. However, it has a large grated door facing the Grand Canal that provides constant and consistent air circulation, making it perfect for taking environmental measurements without exposing the unit directly to the elements.

In addition, the terrace is accessible only by authorized personnel through a glass door, preventing contact with people not involved in the project.

Finally, the terrace is directly exposed to the Grand Canal, where many maritime transports transit throughout the day.

(a) External view of Ca' Foscari.

(b) The unit at the Exhibition Spaces.

Figure 3.6: The data-gathering station placed in Ca' Foscari.

3.6.3 Auditorium Santa Margherita - Venice

The unit to be deployed at the Santa Margherita Auditorium was placed on a grate of an exterior doorway overlooking the Square of the same name (see Fig. 3.7).

Santa Margherita Square is one of the main squares in Venice and a gathering place for young people and students, which was an interesting parameter to consider regarding the concentration of CO_2. However, this factor also poses a security risk to the data collection unit, as many unauthorised persons could come into contact with it. Therefore, it was necessary to place the unit in a location exposed to the square but not accessible to external users, as shown in Fig. 3.7b. Naturally, this location reduced accessibility for the working team, but it is an appropriate compromise to guarantee the system's security.

(a) View of Santa Margherita Square.

(b) The placement of the unit.

Figure 3.7: The data-gathering station at Santa Margherita Auditorium.

3.6.4 San Basilio - Venice

The unit deployed at the Educational Pole of San Basilio was placed on a window balcony in the second building (see Fig. 3.8). The Educational Pole s located near the port area of Venice and is one of the few venues reachable by car. Thus, it was an area of interest to assess the impact of both maritime and land vehicles on CO_2 concentration.

The chosen window is above human height, and its balcony is deep enough to protect against weather conditions, such as rain and strong winds, and non-authorized people, guaranteeing access to the working team. Moreover, the Ethernet outlet is inside the building, so the wire poses no safety risk.

(a) External view of Polo Didattico San Basilio.

(b) The unit placed on the window outside.

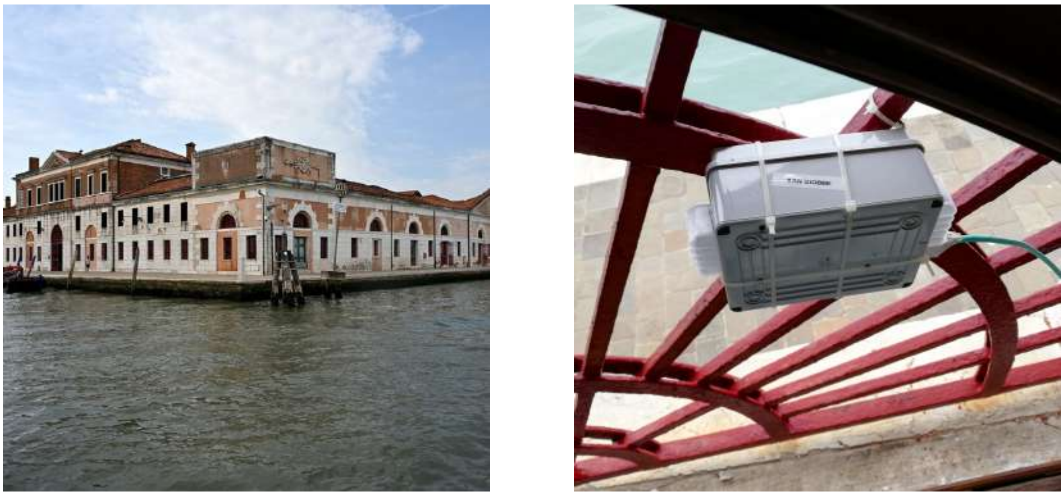
Figure 3.8: The data-gathering station placed at Polo Didattico San Basilio.

3.6.5 Campus San Giobbe - Venice

The unit deployed at the San Giobbe Economic Campus was attached to a grate of the Meeting Room 3 - Economy window (see Fig. 3.9b). This window overlooks the Venice lagoon and thus is the ideal placement to obtain interesting insights on the concentration of carbon dioxide in this location. Moreover, it overlooks a pedestrian route used by students and university personnel, meaning the levels of CO_2 in the zone might also change due to the passage of people.

In addition, the San Giobbe Economic Campus is located near the train station. Therefore, even if there is some distance between the two places, it is expected that the passages of trains, along with the maritime transports in the zone, contribute to the alteration of the carbon dioxide concentration.

The only limitation is due to the accessibility of the unit: since the Meeting Room 3 - Economy is widely used by University staff, the station is exposed to unauthorised access.



(a) External View of San Giobbe Campus and the window of Meeting Room 3 - Economy.

(b) The unit hanging from the window of Meeting Room 3 - Economy at San Giobbe Campus.

Figure 3.9: The data-gathering station placed at San Giobbe Economic Campus.

3.6.6 Briati - Venice

Palazzina Briati represents the final location in the Dorsoduro District. This building hosts some offices and classrooms, meaning the crowding level is low. Inside the structure it can be found a big garden, which overlooks a nearby kindergarten.

The unit was deployed on the terrace railing facing the garden and the nearby kindergarten (see Fig. 3.10b).

(a) External view of Palazzina Briati.

(b) The unit placed on the railing of the balcony.

Figure 3.10: The data-gathering unit placed at Palazzina Briati.

Chapter 4

Initial assessment on the reliability of data

This Chapter describes the initial assessment study about the relation between data and social media, as perceived by young people aged 20-30. It is a fundamental step to understand the successive steps taken in this thesis. The survey was realized and submitted using Google Forms and can be found in Appendix A of the thesis.

4.1 The debate about social media and data

Chapters 1 and 2 state that our society is based on data and information. Everything around us consists of data, and in recent years there has been a tendency to transpose all possible types of information into data [**109**].

Therefore, data science is becoming a fundamental tool for navigating our world, as it allows us to analyze data that are stored in multiple formats and can be collected via different approaches [**46**].

Scientific studies strongly rely on data, and researchers need ways to present their findings and related data to the scientific community and non-expert audiences.

In recent years, social media are gaining the attention of the scientific community as a viable platform for present scientific research and discoveries [**15, 77, 101, 151**]. In addition, their ability to reach a broad and diverse audience could make them an optimal tool for disseminating data on scientific, social, and other topics. According to a study [**78**], the scientific community is also interested in the potential of social media as a possible quick peer review platform.

However, social media are the expression of our fast-paced society, where the span of attention of the user is very low, and it is necessary to be able to tell a story that engages the user as quickly as possible.

Moreover, the rising problem of spreading fake news on social media and the tendency to distrust science undermine the credibility of information and data published on these platforms. Even if the moderators of these platforms are starting to use fact-checking for articles and data shared to regulate the spreading of misinformation, portions of the world population need help distinguishing accurate data from fake ones.

Therefore, the scientific community uses great caution in approaching this world, as social media are currently seen as a non-reliable tool for the diffusion of scientific data. In addition to the diffusion of fake news, there is a problem with the kind of discussion that the social media environment often entails: publishing scientific findings or data may entail having to deal with people who, without any knowledge of the facts, even respond violently or offensively to the author or publisher. This means that even if data receive a positive response from other researchers on the platform, a number of non-expert people might try to undermine such discoveries. This causes mistrust towards social media as a means of scientific dissemination.

Several studies have been conducted to understand the impact of social media on society and how they influence people based on age, education, and origin.

A study published by the Pew Research Center **[146]** surveyed people in 19 different countries to understand how social media influence people about socio-political issues and their impact on people's awareness. According to their results, 57% of the surveyed countries believe social media are good for democracy, but 70% of them think the spreading of fake news is a significant threat. Furthermore, older people are less likely to believe in the goodness of social media. In contrast, young adults think they are suitable for democracy, as they make people more informed about both their own country and international news.

However, young people and the more educated are more likely to think that even if social media are a valuable tool for raising awareness about societal issues and can significantly change people's minds regarding important topics, they also make people easier to manipulate. Moreover, the majority thinks social media enlarge divisions between people with different political views and make people less tolerating, probably due to the increasingly violent debate on the platforms. Such results were compared with the results of a 2018 study **[123]** regarding the same topics, and they found out the responses were very similar. This study also showed that in the U.S., most people get news often (17%) or sometimes (33%) from social media, even if they regularly see misleading and false content on the platforms.

The disparity between the usage of social media and the spread of misinformation by young people and older generations, as well as more and less educated users, is also confirmed by other studies **[9, 50, 61, 80]**. In general, young people use social media the most. They, therefore, are more likely to control the sources of news they see and recognize the potential of social media to raise awareness on socio-scientific issues while also being aware of the harm they pose. On the other hand, older people are more skeptical regarding social media's contribution to making people more informed. Still, they are most likely to share fake news and spread disinformation as they are less familiar with these platforms.

This thesis enters into the landscape of studies concerning the relationship between social media and news in that it analyses the relationship between social media and data in people **aged 20-30**, with particular reference to how the target audience evaluates the sharing of data about socio-scientific topics in the form of visual representations.

This study was conducted by designing and submitting an assessment survey about the topic to a sample of 32 people from the target audience and analyzing the results to understand how they respond to the submitted issues.

4.2 Initial assessment survey

As stated in the previous section, the initial survey was created to assess various information. First, to understand the level of awareness of young people aged 20-30 in relation to the use of social media for accessing data and the use of visual representations for presenting such data on small-screen devices. The survey then attempts to evaluate which social media are most used to obtain information and assess data regarding socio-scientific[11] topics and share such data. Finally, the survey seeks to evaluate the level of knowledge regarding the potential of the discipline known as data visualization and its position in the context of data dissemination regarding socio-scientific topics.

The survey was created using as references other surveys and guides **[26, 98, 100]** on the topic. In particular, an extensive guide on possible questions for a survey on social media **[41]** was used as a starting point for that part.

Therefore, it was designed in four sections:

- *Personal information*, which is necessary to establish the character of the target sample, such as age, gender, occupation, and level of education;
- *Information and media* is used to evaluate how the target audience keeps itself informed about news and scientific articles, which modalities and device people prefer to use to read news and articles, and how they evaluate the reliability of information and data.
- *Social media platforms*, which is used to assess which is the most used social media by the target audience, for which purpose they use them, and how much they think

[11] Socio-scientific issues are controversial, socially relevant, real-world problems that are informed by science and often include an ethical component [115]. Examples include fish farming, genetic testing, global warming, and captive breeding in zoos. "Socio-scientific issues are usually value-laden, and the juxtaposition of science and ethics can be uncomfortable for scientists, teachers, and students who decline science in terms of objectivity" [114].

social media are a reliable source of information and data, also trying to understand if people think social media can be a vector for diffusion of scientific knowledge.

- *Representing data with images and graphs*, which is used to assess the level of data visualisation literacy of the target audience, how they respond to visual representations of data (by presenting some examples), and if they think small screens are suitable for presenting data in the form of images and graphs. This is a crucial section as it provides valuable insights that will be used to draw the set of guidelines for small screens and social media visualisations.

4.2.1 The target audience

As stated above, young adults use social media more than any other age group. According to a 2023 study **[69]**, people aged 27-42 are the most significant users of social media, as 30.3% of US users are millenials[12]. They are followed by Gen Z[13] users, which in the US account for 24.6% of social media users.

The target audience for the initial assessment survey was identified as somewhere between these two age groups, namely people aged 20-30. However, this small range was chosen for a clear view within this age group, as the sample consisted of 32 people.

The first section of the survey was designed to collect information about the target sample. Such information is critical to understanding the factors influencing people's opinions about the topics discussed. To ensure the privacy of the respondents, the survey was conducted anonymously.

According to the data extracted by the first part of the survey, the sample of people consisted of **17 men** (53.1%) and **15 women** (46.9%), all coming from Italy. From an age point of view, the sample was divided as shown in Fig. 4.1.

In addition, the people surveyed were asked to disclose information about their occupation at the time of submission of the survey. The responses showed that half of the target sample is still a student (**50%**), whereas a third is currently working (**31,3%**). A small part of the target sample is working and studying, whereas only one person was unemployed, as shown in Fig. 4.2.

An essential piece of information is the *level of education* of the sample. The survey showed that the majority of people have a **Bachelor's degree** (**56.3%**), followed by more than a quarter of the interviewed that has a **Master's degree** (**28.1%**). In contrast, only one person had a higher education level (Postgraduate master's degree, **3,1%**). The remaining **12.5%** stated they only have the **High School Diploma**, as shown in Fig. 4.3.

[12] Th term "millennials" refers to a demographic cohort also known as Generation Y, which includes people born indicatively between 1981 and 1996 [150]

[13] The term "Gen Z" refers to the demographic cohort, also known as Generation Z or "digital natives," which includes people born indicatively between 1997 and 2012 [149].

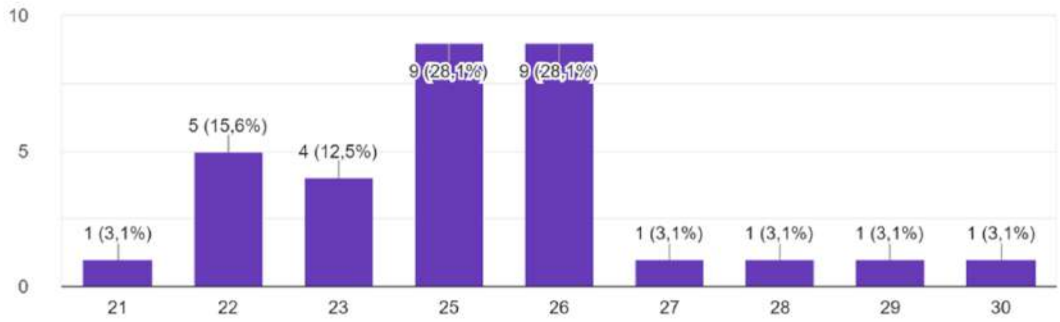


Figure 4.1: Graph showing how the target sample is divided according to their age at the time of the survey.

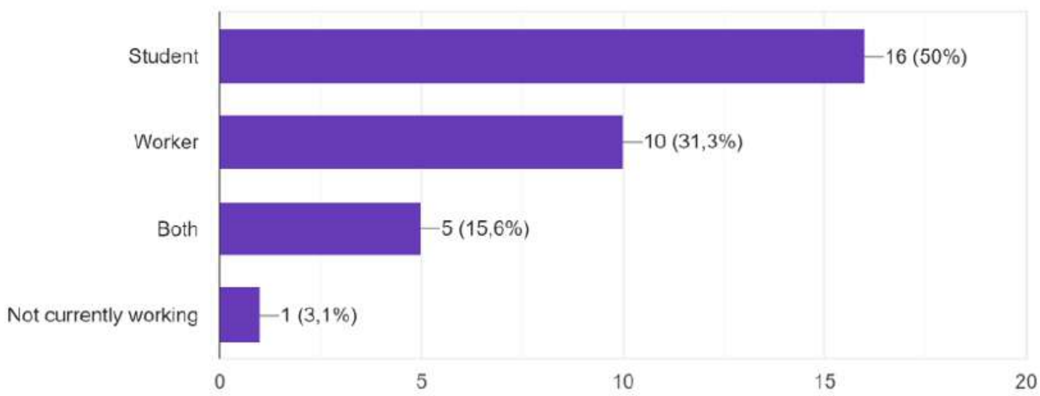


Figure 4.2: Graph showing how the target sample is divided according to their occupation at the time of the survey.

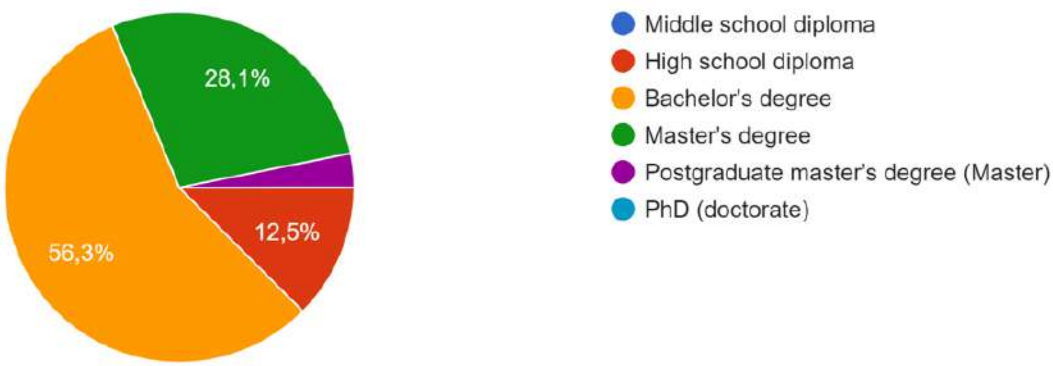


Figure 4.3: Division of the sample according to their level of education.

As shown by Table 4.1, the workers interviewed were employed in different *types of job*. This information can be useful to analyze how different positions influence people's opinions, especially distinguishing the ones who work and study (33.3% of the workers) from those who only work (66.7%).

Occupation	**Student and worker**
Designer	NO
Digital Marketing	NO
Electric technician	NO
Engineer	NO
Insurer	**YES**
IT Technician	**YES**
Procurement employee	NO
Psychologist	NO
Software Developer	NO
Teacher	NO
Teacher	NO
Teacher	NO
Technical employee	**YES**
Waitress and sommelier	**YES**
Winemaker	NO

Table 4.1: Occupation of the employed people in the target sample and their status.

4.2.2 Information and media

The second section focuses on assessing how the target sample relates to information and news, their preferred ways of keeping informed, and their stance on the credibility of the data and sources publishing them.

According to the responses, the majority of the target sample thinks it is fundamental to have access to news and information (**96.9%**) and reads news or socio-scientific articles very frequently (**68.8%**), and a small part does it every day (**12.5%**).

As shown in Fig. 4.4, the preferred outlets for reading news are *online newspapers and magazines* (**80%**) and *social media* (**67.7%**), followed by *scientific journals* (**38.7%**) and *news app*, such as Internazionale, BBC, ilPost (**35.5%**). This result is important as it underlines the predominance of the web and, in particular, of social media as a way to keep informed young people; however, print media had a high preference, **32.3%**.

As for the preferred devices for reading news, the target sample showed the highest preferences for *smartphones*, used by **90.3%** of the target, and laptops, used by **67.7%** of the interviewed.

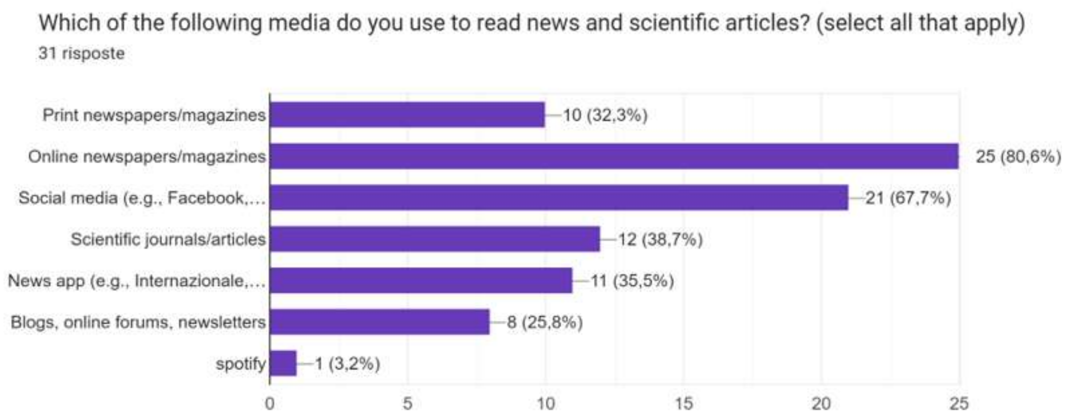


Figure 4.4: Graph showing the outlets used by the target sample to read news and scientific articles.

When asked about which are the factors considered to evaluate the reliability of the data and news, the target sample responded that the two most important parameters are *reputation of the source* (**73.3%**) and *source's expertise in the subject matter* (**70%**). Half of the sample looks for *corroboration of the information* from other sources, and only a quarter evaluates the sources based on their *political bias* (**23.3%**). However, none of the interviewees trust the sources without question. Regarding the sharing of news and data, most of the target sample tends not to share information, while only a small part claims to do so frequently (16.1%).

4.2.3 Social media platforms

The third section of the survey was essential, as it was designed to understand the relationship between social media data and fake news according to the target sample.

Of the 32 interviewees, 30 are registered on social media, while two are not. Therefore, the answers to the questions in this survey section refer to a sample of 30 people.

The **86,7%** was registered to *Instagram*, **66.7%** to *Facebook* and **56.7%** to *Twitter*. Other platforms mentioned, but with a lower percentage of registered users, are in descending order Linkedin, TikTok, Pinterest, Reddit, and YouTube, as shown in Fig. 4.5.

Almost all users (**96.7%**) use social media *to entertain themselves*; **70%** of the target sample uses social media *to keep themselves informed*, **60%** for *learning* and half of the target uses them also for *networking*. On the other hand, only a small part (**16.7%**) uses social media for *business purposes*.

Of these 30 people, the majority stated that they frequently come across data on social media, with a high percentage affirming they encounter data daily (**36.7%**). Moreover, almost all people in the sample stated that they are *very concerned* (**90%**) about the *impact of false or misleading data on social media.*

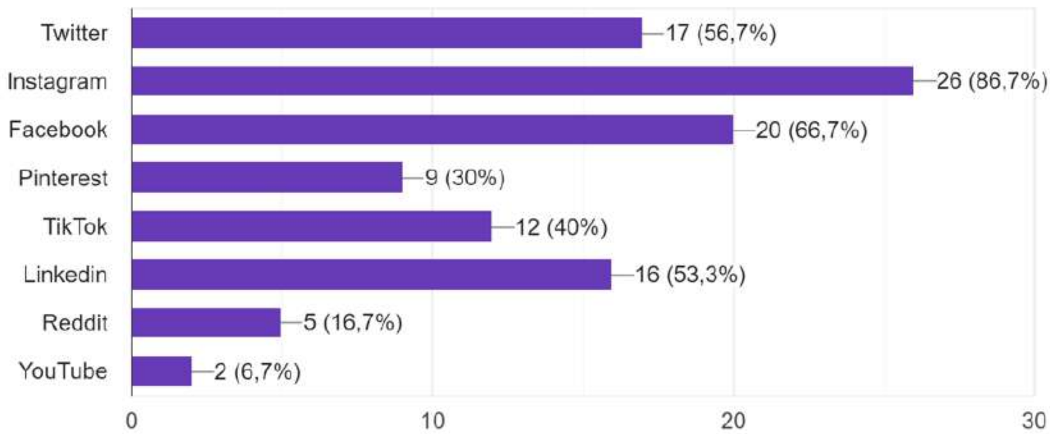


Figure 4.5: Graph showing the social media used by the target sample.

When asked how much confident they are about the *accuracy of the data shared on social media*, the majority of the sample stated that it is *depends on the source* (**36.7%**). The remaining part of the sample tends to be *unconfident* (**33.4%**), and just a tiny amount is optimistic about it, as shown in Fig. 4.6. According to a statistical analysis conducted using the R software[14], young people and students tend to rely principally on the source to decide whether or not they are confident about the data.

Moreover, people in the sample affirmed they are not very familiar with *fact-checking lables*[15], but the majority of the sample is likely to believe them (**80.6%**).

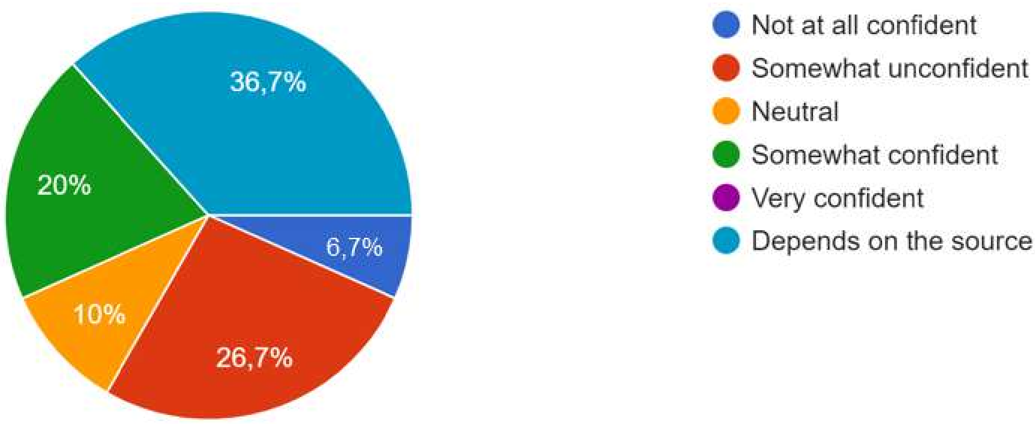


Figure 4.6: Graph showing the responses of the target sample to the question: "How confident are you in the accuracy of the data shared on social media?".

[14] R is an integrated software suite for data manipulation, calculation, and graphical display [132].

[15] Fact-checkers are third parties, independent from platforms and certified through the non-partisan International Fact-checking Network (IFCN), which fights misinformation. Fact-checkers review content and rate its accuracy. Based on the ratings, platforms decide to signal or ban the content[85].

The majority of the target sample affirmed to *verify the accuracy* of the data before sharing it, as shown in Fig. 4.7. According to the statistical analysis, *young students*, especially those who have a Master's Degree, *women* and people with *low level of confidence* on the accuracy of data on social media, tend to verify the accuracy the most.

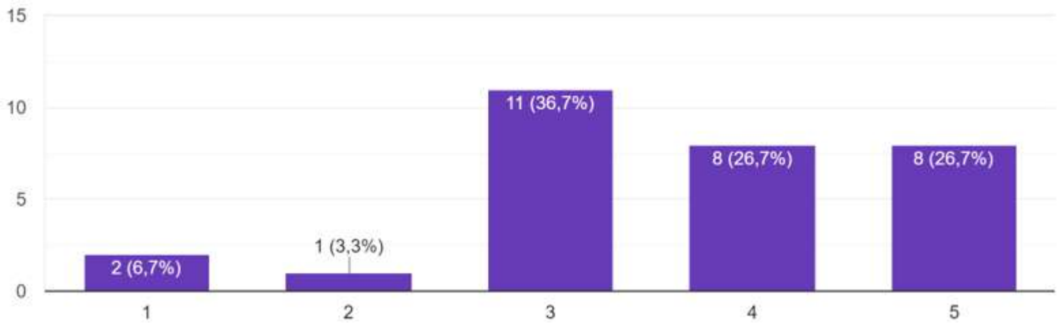


Figure 4.7: Graph showing the responses of the target sample to the question: "When you see data on social media, how often do you verify its accuracy before sharing it?" The scale goes from "1: Never" to "5: Always".

When asked if they have ever *shared data that was later found out false*, **76.7%** of the sample affirmed it *never* happened to them, whereas the remaining part of the sample proved it happened *at least once* (**23.3%**).

As for *reporting data and news*, more than half of the sample stated they reported false data at least once (**56.7%**), as shown in Fig. 4.8.

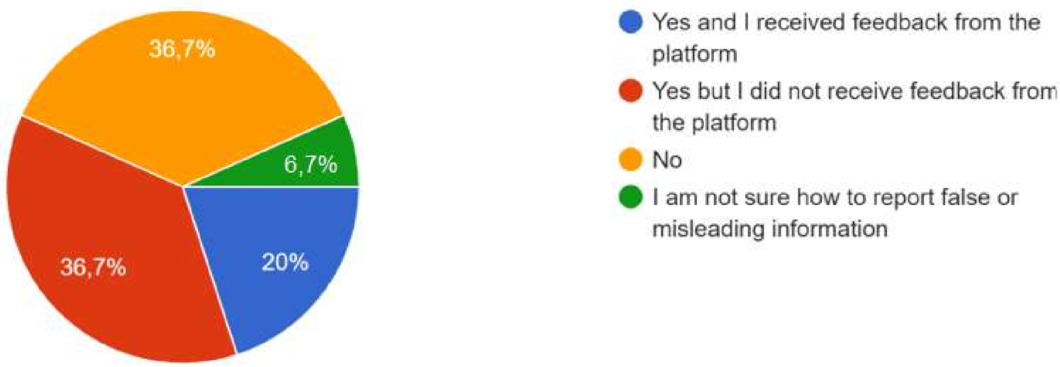


Figure 4.8: "Have you ever reported false or misleading data on social media?".

The last question of the section asked the interviewed to give their personal opinion about *how social media could improve data and information shared by users on their platforms.*

Of the 30 interviewees, only 15 responded. The answers are reported in Table 4.2 and show that the users think that to improve data and fight misinformation, social media should require *citing the sources* of data and use *more advanced fact-checking processes.*

How do you think social media can improve the news, data, and information shared by users on their platforms?
Always citing the source. The citation must be easily accessible.
By always reporting the sources and adding fact-checking labels that corroborate the data or reject them based on well-trusted sources.
By creating a dedicated section for information, where news and data must be verified before being transmitted. Something like 'Instagram News' or 'Youtube News,' in which only content of this type is present. This would bring order to the sea of different content on social networks and prevent the random posting of random things. Users could still obviously do it, but they shouldn't be taken seriously if they don't do it in the news section.
Doing more fact-checking on shared data.
Promoting profiles with a good reputation based on users' feedback.
Socials can invest more in their checking processes of shared information.
Checking the authenticity of this much information is no easy task, and I have no idea how to improve their methods.
Cross-validation of incoming/shared information.
Everyone should take a greater interest in fake news, reporting and contributing to the best of their capabilities.
Highlight reliable sources
By making it more accessible
No, for several reasons: 1) Short-form content limits how nuanced and in-depth the information can be presented. 2) Restricting content makes discussing certain topics difficult (such as terrorism and war). 3) It encourages "clickbaity" titles or extreme/not nuanced takes to attract an audience, which can warp the perception of an event.
News control.
More checks should be made before a piece of content is posted. It would also be important to make users even more aware of their responsibility in posting something.
With a verified filter able to separate fake news from legit news

Table 4.2: Answers to the question: "How do you think social media can improve the news, data, and information shared by users on their platforms?".

4.2.4 Representing data as images and graphs

The fourth and last section of the survey focused on Data Visualisation and representation of data as images and graphs.

More than half of the sample stated they encounter data presented as images almost every day (**62.5%**), whereas the remaining part said they meet them only sometimes (**37.5%**). The settings in which they find data presented as images and graphs are *social media post* (**84.4%**), *news articles* (**65.6%**) and *scientific papers* (**59.4%**).

According to the majority of the sample, data presented as images and graphs are definitely more appealing (**71.9%**), whereas the remaining part does not have strong feelings about them. Moreover, most of the sample tends always to read the text accompanying the representation if it is present (**81.3%**). However, **75%** of the sample stated that they have encountered *misleading* or *confusing* representations of data. Table 4.3 reports the factors that make images and graphs confusing and misleading.

What made the graphs you encountered confusing or misleading?
Incompatible abscissa and ordinate, you had to do an equivalence to understand it.
Wrong evaluation and elaborations of the data.
A confused index.
A not proper communication.
Too much information in a single image.
The colours and the reference level used.
Missing or wrong scale, too many details.
Too much information and colours.
The presentation of the data and the scales.
No labels, wrong scale, bad colour choice, bad graph choice to fit the data.
Data was represented incorrectly; the sources did not correctly elaborate data, taking, for example, bad samples that are not representative enough for the study.
The quantity of data presented at once.
Improper use of scales and graph types, unappealing and counter-intuitive visuals.
The data and the visualisation were at some point different and, misleading, inaccurate.
The description underneath was too short/the contact was not clear.
More complex graphs or clearer explanations.
Unusual type of graph.
Too much information and an unclear caption.

Table 4.3: Answers to the question: "What made the graphs you encountered confusing or misleading?".

Therefore, disadvantages may lie in: Wrong use of *scales* and *captions/labels*; putting *too much information*; using *wrong or complex graphs* for the type of data being represented.

In Table 4.4 are reported the sample answers regarding the advantages of data presented as images and graphs in their opinion.

In your opinion, what are the advantages of presenting data in the form of images and graphs?
Understanding data presented with graphs is more immediate and user-friendly.
Easier to communicate for both writer and reader.
It catches the reader's attention more easily than just the text.
They are useful for creating a generic idea.
You can make your idea.
Immediacy.
You get a deeper understanding of the relationships between variables. With less effort.
They are more intuitive and faster.
They're simpler and more immediate, allowing non-expert users to understand complex data.
Images and graphs are remembered better than some plain text
The clearness of the data
Immediate visualization - favouring comparison between numbers.
Having a graphical and a picture representation can make the topic clearer and more understandable, providing practical examples.
Better engagement and easier understanding of the data.
Data comparisons are easier to make and understand.
Immediate understanding and easier to comprehend. Eye-catching also.
Good DataViz is useful as it can immediately provide a global understanding of a fact.
They are more impactful and attract the attention of more people.
They catch people's attention and result in more people reading them.
They can include a lot of information in just one picture.
Very immediate for many people.
You get a better understanding of the data.
It makes comparing data that changes over time or according to other factors easier.
More intuitive and appealing.
Graphs and images can immediately give users an idea about data/a specific topic, often allowing them to skip the written text. Colours and numbers also help draw users' attention to specific content.
Images help improve understanding through schemes.

Table 4.4: Answers to the question: "In your opinion, what are the advantages of presenting data in the form of images and graphs?".

As shown by the answers, the main advantages of graphs and visual representation of data are: *immediacy*, *intuitiveness*, *appeal*, *simplicity* and *conciseness*.

The sample was asked to give their opinion about what can be done to improve the representation of data as images and graphs. The answers of the 15 respondents are listed in Table 4.5.

In your opinion, what steps can be taken to improve the quality of data presented in the form of images and graphs?
Use simple graphics and different colours; always accompany immediate and brief legends, which make the topic clear and how to read the graph.
Learn methods of data elaboration.
Deeper descriptions.
A deep knowledge of the topic you are talking about.
Create easily accessible tools to produce easy-to-read and standard plots.
Improving the use of colours to indicate differences in data and use simpler infographics.
KEEP IT AS SIMPLE AS YOU CAN: 1) Use colours with a specific purpose 2) Use texts only where it counts and to offer important information 3) Giving the references 4) Leave "white space" in the visualisation.
Doing clear and simple graphs with some text to explain the graph and some impactful visualisation to catch the user's attention.
Provide guidelines and examples. Easy-to-use software may also help.
Use more reliable sources of information and consider the right samples in line with the topic we want to communicate.
Strong data analysis and study of multiple/novel data visualisation tools/techniques.
Good explanation and previous familiarity and preparation.
Entrusting the creation of graphics to experts.
As many people find it increasingly difficult to keep their attention, images, and graphs should be more clear and sparse as possible.
Teach people how to create catchy images with the most important content.

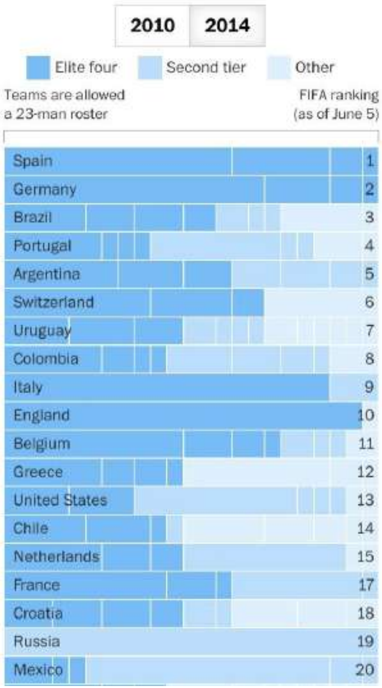
Table 4.5: Answers to the question: "In your opinion, what steps can be taken to improve the quality of data presented in the form of images and graphs?".

According to the answers, the most important steps that can be taken to improve the quality of data representations lie on *use of colors*, the pursuit of *simplicity*, use of *clear and concise captions and descriptions* and *draft guidelines* for creating graphs and representations.

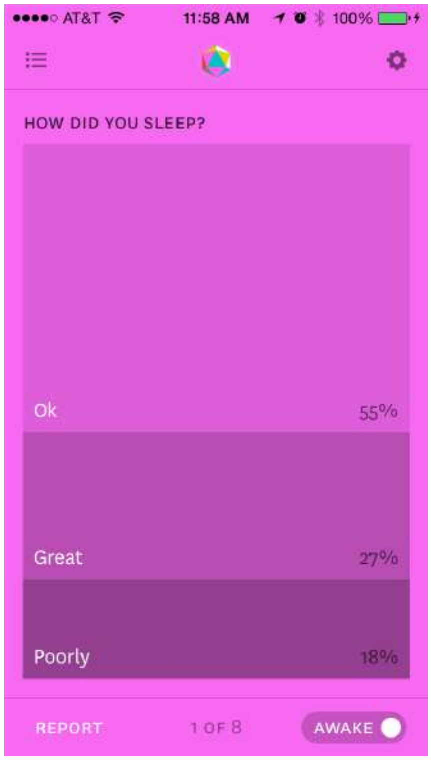
The last part of the survey focused on representations for small screens. First, the sample was asked if the small screen size of smartphones and devices is a limit for representing data as graphs and images. According to **54.8%** of the sample, the small size is NOT a limitation. However, **45.2%** of the sample thinks the small size presents several disadvantages, which can amount to:

- inability to display many elements in a single screen;
- characters and text might be too small to be readable;
- oversimplification might mislead or confuse;
- presenting a general visualization with no details about the context;
- small space means important details such as legends, texts, captions, and labels might be misplaced or removed completely, leading to confusion or illegibility of the chart;
- difficult to have a view of the whole visualization as it may exceed the screen space;
- use of a type of graph that fits into the small space but fails to present the data and correlations between them correctly;
- the need to redirect the user to another page or device to obtain a full view of the representation, the data, and the additional context;
- lack of interactivity;

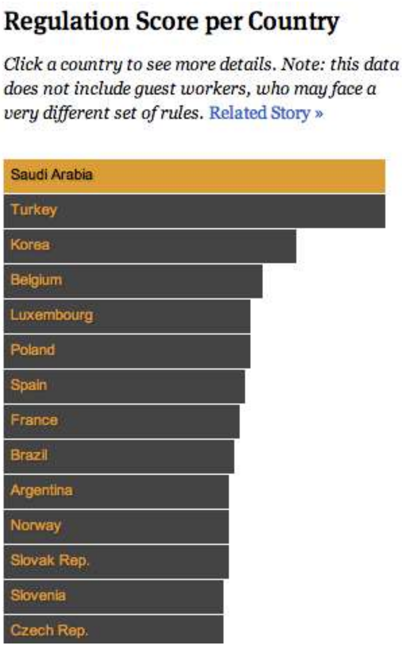
To further understand the opinion of the sample about data visualizations, especially in the context of small screens, they were presented with a series of examples taken from the online repository **MobileVis [110]** and selected based on the topic they referred to and the type of visualization they presented. The interviewed people were asked to indicate which were the point of strength and the flaws in the representation that was submitted to them, which are shown in Fig. 4.9 and Fig. 4.10. The representations were submitted to the sample, accompanying them with a text that gave them some context about what they represented.



(a) Leagues Cup of the World and their roster [117]



(b) Sleep quality in the Reporter App for iPhone [39]



(c) Temp Worker Regulations Around the World [117]

(d) Food Report in the Reporter App for iPhone [48]

Figure 4.9: The first four representations submitted to the users in the Initial Survey .

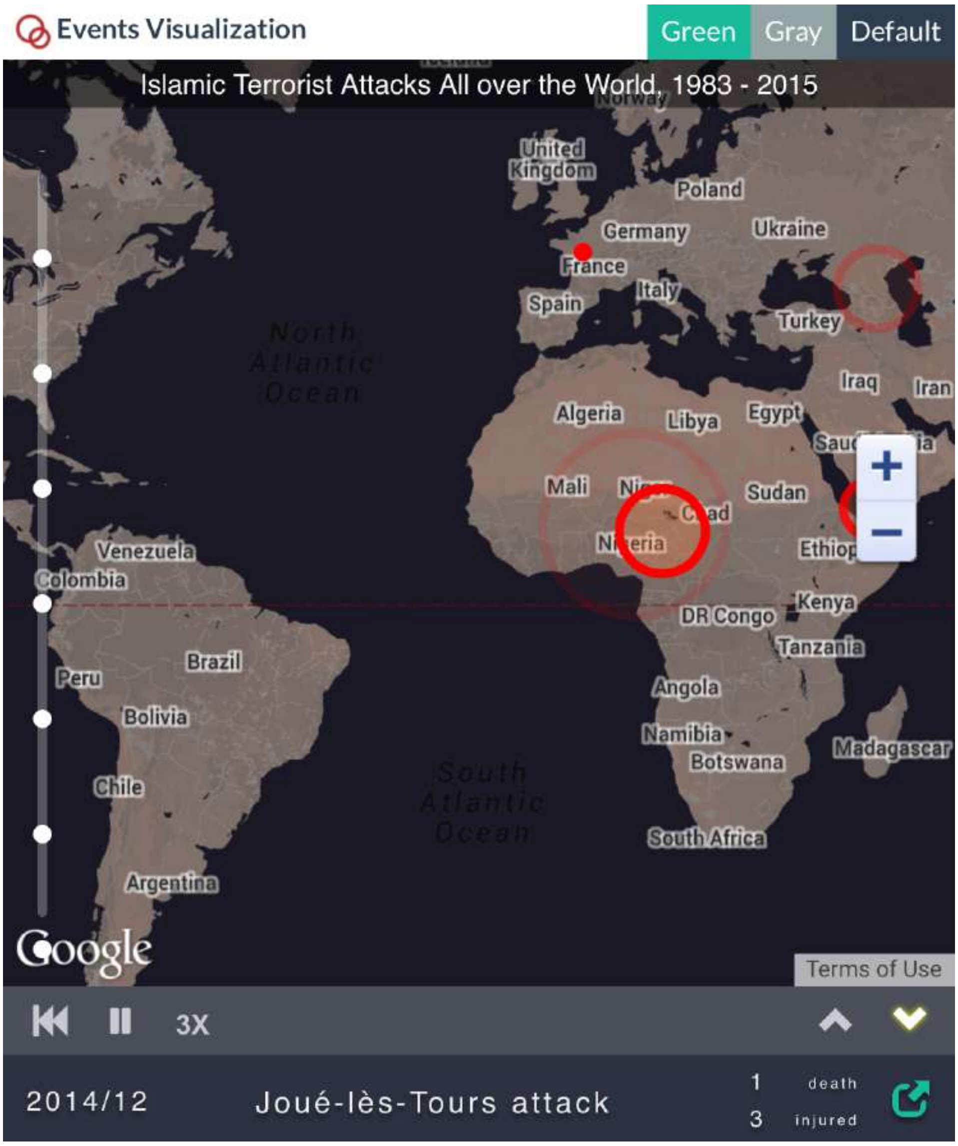


Figure 4.10: The last image presented in the survey, showing "Islamic Terrorist Attacks All over the World, 1983 - 2015" [156].

When presented with those practical examples, the users highlighted the main problems of the representations, and their answers were coherent with the responses given to the previous questions about the disadvantages of visual representations and small screens. Table 4.6 summarizes the specific comments about the representations.

They also highlighted the points of strength, indicating simplicity, immediacy, and clearness of the representations as their strong suits, especially for Fig. 4.9b, 4.9c, and 4.9d.

Example	Comment
Fig. 4.9a	1) Colors too much alike 2) Confusing unit of measurement; 3) Missing information about players; 4) The legend does not explain the char; 5) Data are not immediately understandable; 6) Overlapping data and meaningless rectangles.
Fig. 4.9b	1) The shape of the chart is not appropriate; 2) Missing correlation with other variables; 3) Too little information displayed; 4) Bad use of colors; the background is too similar to the data; 5) The graph should not occupy the entire screen size; 6) The categorical levels are not in order of "goodness"; 7) Some texts are not readable.
Fig. 4.9c	1) Missing unit of measurement; 2) No correlation between score of regulations and their effectiveness; 3) Missing axis; 4) Missing information about the types of regulations; 5) Representation seems interactive but isn't; 6) Missing information on the number of regulations.
Fig. 4.9d	1) Type of graph inappropriate; 2) The graph suggests you should complete an unrealistic goal; 3) Numbers are not readable; 4) The scale is unrealistic; 5) Too bright colors; 6) There is no categorization of food types;
Fig. 4.10	1) To many elements; 2) Colors used are not very engaging; 3) Missing scale about the red circles; 4) Overlapping elements and texts; 5) Some labels are not visible (the ones on the sea)

Table 4.6: Summary of the main defects identified by users in the example representations presented in the initial assessment survey.

4.3 Discussion

Overall, the initial assessment survey was very useful in gathering valuable data and insights for the aim of this thesis.

In particular, the data collected by the sections on information levels and the use and reliability of social media have shown that people aged 20 to 30 tend to use the Internet to keep themselves informed but are very careful to inform themselves from sources certified and experienced or confirmed by other equally legitimate sources.

Moreover, it is noted that most of the sample use social media, especially on their mobile devices, as their primary information tool. According to the data, they are aware of both the potential and the flaws of these platforms as ways for disseminating socio-scientific knowledge, showing they are highly concerned about the impact of fake news and misleading data. Young students are more likely to recognize this dichotomy because they use social media more. Therefore, they pay attention to what they see on social media, verifying the accuracy of the data and reporting fake news when they spot them.

The answers to the survey section related to data visualization provided helpful insights and users' advice for creating a set of guidelines for small screens, mainly social media, that can be user-friendly and coherent. Outlining the main benefits of graphics and data-based images provides a valuable basis for drafting guidelines for developing visual representations. Moreover, by pointing out the flaws in the examples of visualizations submitted to them and stating the general disadvantages of representing data as images, the sample gave a lot of indications on what not to do when creating data visualizations.

Chapter 5

Guidelines for visualization on small screens

As shown by previous Chapters, small screens are heavily used by people worldwide for different purposes that range from entertainment to keeping informed and accessing news and data.

As stated in Section 2.2, the use of data visualization to represent data on small screens has grown rapidly over the past decade since smartphones and other types of small-screen devices such as smartwatches and tablets began to spread massively. As a consequence, there has been a growing interest in defining possible guidelines for creating visual representations of data for these devices.

The scientific community urges to conduct studies for the development of such guidelines, as these devices are now a fundamental part of everyone's digital life, and the amount and different types of data produced that you want to visualise grows every day.

This Chapter is devoted to the creation of a possible set of guidelines for representing data on small screen devices, with particular attention to defining a set of rules that can be used for the case study of this thesis, namely social media.

5.1 Methodology

In order to establish a possible set of guidelines for small screens, two sources of information were taken into account: the *analysis of the literature* and the *initial assessment survey.*

5.1.1 Analysis of the literature

Since the end of the 1990s and the introduction of PDAs, many researchers conducted studies to try and define some rules for visualising data on small screens. Nowadays, there are hundreds of different papers and researchers that propose their set of guidelines for these types of devices.

Section 2.2 of Chapter 2 of this thesis explored in detail many different works and studies conducted in this domain, analysing the different contexts of use and factors that might play a part in defining the rules. Fundamental works that gave clear indications about this domain are the book "Mobile Data Visualization" by Bongshin Lee et al. [**71**], which gives an in-depth view of visualisation on mobile devices in all its aspects; the work by Archie Tse [**136**] on the advantages of static visualisations; the patterns in graphs for mobiles gathered by Irene Ros and Bocoup [**111**]; the presentation of Gregor Aisch on the relation between Data visualisation and the news [**2**]; the guide on data visualisations on mobile and desktop proposed by the Visual Cinnamon Team [**141**]; the guide for creating graphs on social media of Tselova and Parker [**137**]; the guide for creating charts for social media by Andy Cotgrave [**33**]; the book "How charts lie" by Cairo [**22**] on how to read a chart correctly.

From the analysis of these works, some fundamental notions have been extracted and will be used for the drafting of the set of guidelines in this thesis:

- The *seven descriptive dimensions* [**68**] that allows classifying data visualisations (physical data display size, data display mobility, data source, the reaction of visualisation to display movement, visualisation interaction complexity, intended sharing);
- The strategies for transforming desktop-based visualisations into mobile-based [**72**], which give useful hints on what and what not to put in mobile visualisations (consider the scale of content, the aspect ratio of the device, the layout of multi-element content, level of detail, removing unnecessary elements, change the visual encoding);
- If there are no other choices, consider zoomable visuals for visualisations with many details [**141**];
- Consider using static visualisations as they are more diffused than interactive ones, as the latter are more expensive to adapt to all platforms [**136**];
- Interactivity is engaging but can be misleading if it is not illustrated: if you use interactivity, indicate clearly how to trigger interactions [**24**];
- additional context is necessary, as visual illiteracy is a great problem in our society; therefore, adding texts, captions and legends is key to avoid misleading and confusion [**22**];
- Complex interactions are ignored by users: prioritise scrolling and zoom over clicks and tooltips [**136**];
- What is animated on a desktop is usually not animated in mobile [**117**];

- Colour must be used intelligently: colour schemes have to be intentional and must highlight key patterns [**37**];
- Do not use too much text in the visualisations [**37, 137**];
- Key factors to consider when developing visualisations, especially in the context of social media, are *simplicity*, *attention to the target audience*, *originality* and *accessibility* for colourblind people [**22, 33, 113, 133, 137, 145**]
- Regarding social media, consider using GIFs and videos to substitute interactivity [**33, 137**];
- Citing the source of the data is important as it allows to gain the trust of the viewers [**36**];
- Ethical implications of misleading, incomplete or confusing visualisations should be considered, especially when representing data related to socio-scientific or political topics [**36**].

All these insights and guidelines will be used as a basis for the work of this thesis.

5.1.2 Initial Assessment Survey

The results collected from the initial assessment survey, presented in Chapter 4, provided valuable insights that can be used to draft the set of guidelines for small screens. Furthermore, considering the direct suggestions from non-expert users regarding mobile and social media views, it highlights what interests end users most, considered the main factor in the personal understanding of what they have before their eyes.

According to the responses, the factors users value the most when it comes to data-based images and graphs are:

- citing the source of the data, especially in the social media context; the source of the data is crucial in gaining users' trust;
- immediacy and conciseness;
- visual appeal;
- simplicity;
- use of colours;
- understandable and concise captions/labels/descriptive text.
- putting the right amount of information on a single graph, avoiding filling it with elements also given the small screen size;

- choosing the right graph for the types of data available.

Once again, it is shown that *simplicity*, *appeal*, and *immediacy* are critical factors in the development of visual representations on the small screen and social media.

5.2 Guidelines

Having collected information, insights, and examples of other possible guidelines from the literature review and the Initial Evaluation Survey, it now becomes possible to draft guidelines for small-screen devices.

Even if there are some studies on the matter that are very detailed, such as the analysis conducted in the "Mobile Data Visualization" book **[71]**, which heavily inspired the rules drawn up in this thesis, the set of guidelines proposed here tries to be as *concise* as it is *comprehensive* to cover all possible cases without using complex language so that it can be used by people even those who are not strictly experts in the subject matter.

The guidelines are divided into sections: *key concepts*, *display size*, *data source*, *intended viewing timespan*, *interactive vs static visualisations*, *type of graph*, *colours*, *additional context*, *accessibility.*

5.2.1 Key Concepts

As previously stated, the key concepts to be considered when developing visual representations based on data are:

- **Simplicity**: A data visualization should be simple, free of complexity, essential, and accessible to anyone;

- **Immediacy**: A data visualization must be comprehensible quickly, as the context involves very short lingering times.

- **Conciseness**: A data visualization should be essential, yet complete and effective; it should not contain unnecessary elements, but only the most critical information;

- **Consistency**: if several data representations are used for presenting the same topic in the same context, the visualizations should belong to the same class and explain the concept in the same way;

- **Appeal**: A data visualization should be pleasing to the eye and should engage the viewers so that they reflect on the subject matter and linger on the representation.

- **Trust**: a data visualization should be reliable and trustworthy; it should not confuse or mislead the viewer and provide access to the data source to gain more credibility.

- **Know your audience**: if you are creating the representation for a specific audience, you must know the audience you are referring to create a visualization that suits them.

5.2.2 Display Size

The device's size is critical when developing data visualizations for small-screen devices. Small screens comprehend different devices with different display sizes, influencing the representations. Moreover, the distance between the viewer and the display affects the data display.

In particular, you could have:

- **Smartwatches**: rectangular or circular screen, 3-4 cm wide. The representations cannot convey much data; therefore, they should have few elements and not contain small and difficult-to-read text. Interactivity is available via touch screens and buttons.

- **Smartphones**: rectangular screen, generally 15 cm on the diagonal, high resolution. These devices can display more information and elements, but smartwatches require the viewer to be close to them. Interactivity is available via touch screens and buttons.

- **Tablets**: rectangular screen, ranging from 17 cm to 25 cm on the diagonal. They can display many elements, thus supporting more complex and layered representations. Interactivity is available via touch screens, pens, or buttons.

The representation should be different based on the type of device/display available.

5.2.3 Data source

Another important aspect is the source of the data being visualized. All types of small-screen devices can work with both **pre-loaded** and **real-time** data coming from in-built sensors or external sources using a connection. Based on the source of the data, the representation will be different.

- **Pre-loaded data**: data have been previously loaded into the device. Thus, data are static but can be used to create both static and interactive representations. Furthermore, they are neither affected by the external or device's environment nor time critical.

- **Real-time data**: data are coming either dynamically over connections to an external server (WiFi, Bluetooth, Internet Service) or are captured by in-built sensors (GPS, gyroscope, compass, watch, etc.). Updates to the data might dynamically change the representations.

- **Combined source**: data might come from a combination of pre-loaded and real-time data; this is useful when the device cannot elaborate or capture some types of data or (temporarily) lacks connectivity.

The source of the data should always be cited to gain the user's trust and increase the representation's credibility.

5.2.4 Intended viewing timespan

The amount of time the viewer is supposed to spend on the visualization substantially influences the type of representation to be created and its complexity. The more complex the representation, the more time a person should linger on it to understand it.

The viewing timespan depends on the context of use: small-screen devices are generally used for tasks that require a limited attention span, ranging from **sub-second** up to **minutes**. A small timespan requires simple and immediate visualizations, thus familiar chart types with essential elements highlighted. The longer the time, the more complex the representation can be.

5.2.5 Interactive vs. Static visualizations

One of the most divisive factors is the choice between **static** or **interactive** representations. The preference for one or the other type of representation depends on many factors, including the *context of use*, the *story* you want to tell with the visualization, and, more importantly, the *target audience*.

Static Visualisations

Static visualizations focus on a *single data relationship*. They display simple data that represent a specific story.

Since they are focused on a single viewpoint, their essential features are *immediacy* and *engagement*: they have to catch the viewer's attention quickly and convey the primary information in a few seconds. Moreover, since you cannot hide information behind tooltips or a second page, all the essential information should be displayed, paying attention not to overfill the representation.

Since there is no possible interaction, the viewer will not spend much time on the representation, as it is not pushed to explore forward with actions on the image. The visualization should engage the user with the specific story it tells so he decides to deepen the topic outside the context of the representation itself. Static visualizations are cheaper to create and adapt to different environments.

In general, static visualizations have a single-page layout, so there is no need to explore more than the first view of the representation. An example of these types of representations is *infographics*.

Interactive Visualisations

Interactivity is an exciting feature as it provides more engagement and makes the viewer active in the representation. Interactive representations are generally used in web applications and software [**16, 87**], used in different contexts, and directed to different target audiences. They are more expensive than static ones to create and adapt to different environments, as something interactive in one environment may be static in another.

There are different types of interactions with additional complexity. Therefore, the more complex the possible interactions, the more additional knowledge should be provided to the users a priori. Some examples of interactivity are filters, zooms, animations, and transitions; some triggers of interactivity might be sliders, buttons, and tooltips. More complex interaction concerns movements, such as reaction and adaptation of the representation to display movement (such as shifting from landscape to portrait orientation) and sounds.

When dealing with real-time data sources such as the ones cited in Section 5.2.3, data updates might change the representation: this is an example of indirect interaction.

To respect the key concepts defined in Section 5.2.1, some useful tips can be followed:

- When using *scrolling* as interaction, prefer vertical scrolling rather than horizontal one.
- *Zoom* and *Scroll* are better than clicks, as they are more intuitive.
- Try not to hide information behind tooltips; when using tooltips, it is better to show them in a fixed area of the screen rather than have them appear in a random spot.

Interactivity is useful when:

- You want to make trends understandable quickly.
- You have multiple data relationships and therefore want to tell a complex story with multiple viewpoints that change over time.
- You have complex data that you want to simplify.
- You want to engage the user so that he spends more time exploring the representation.
- You want the user to be an active part of the process, allowing them to manipulate, explore and customize the representation by selecting the information they want to see using interactions.

- You want to highlight specific data concerning other ones or make comparisons.

Interactivity should be used wisely to achieve specific goals, and it is necessary to ensure it is intuitive, consistent, and user-friendly.

5.2.6 Type of Graph

The choice of the graph is fundamental and relies on the data you have to display and the purpose of the representation, as shown by Kirk [**65**] and the **CHRTS** families of chart types: **C**ategorical, **H**ierarchical, **R**elational, **T**emporal, and **S**patial charts. The purposes can be:

- Comparing values (Bar, bullet, waterfall, radar, polar charts, pictograms, connected dot plots, word clouds, heat map, histogram, matrix, density plots, infographics, etc.)
- Finding relationships and hierarchies; (Pie, waffle, staked bar, Marimekko, diverging bar sunburst charts, treemaps, dendograms, Venn diagrams)
- Exploring correlations and connections inside the data (scatter, bubble plots, network, chord, and Sankey diagram)
- Plotting data over time (Line, bump, slope, area charts, connected scatter plot, stream graph, Gantt chart)
- Plotting spatial patterns (choropleth map, isarithmic map, prism map, dot map, flow map, grid map).

Even in this situation, knowing the audience is essential: more expert people can understand complex graphs (such as bump charts, stream graphs, etc.), whereas non-expert audiences might find more straightforward graphs more effective.

5.2.7 Colours

Colors are a fundamental feature of visualizations, as they are a powerful visual stimulus. Colors are vital in catching the user's eye and making him look at the representation. Moreover, using colors and variations in the color scheme can highlight relationships or differences between the data.

The choice of the color scheme is key. Such a decision should be made based on the *purpose* of the representation, the *type of graph* you want to use, and the *data* you wish to represent.

In general, when we want to compare data of different types, it is advantageous to use various color schemes. For example, if you want to make one particular piece of data stand out, you should ensure that the color chosen is brighter and more visible than the others.

Likewise, if you want to show a scale of values in a particular order (from worst to best, etc.), using the same color spectrum that becomes more or less intense according to the scale itself will be helpful.

In any case, it is helpful to inform oneself and follow the rules of color composition to associate different colors correctly (e.g., complementary colors). Another important tip is to ensure sufficient contrast between the colors of the background and those of the main elements and between the elements themselves.

5.2.8 Additional Context

When creating graphs, especially interactive ones, it is often necessary to include additional context **[22]**. This is good practice to ensure that the graph is readable and contains all the information the user needs to understand what is in front of them. The lack of such elements can often lead to confusion or misunderstanding. This features include *titles*, *captions*, *descriptive text*, *legends*, and *labels*.

Providing a **scale of measurement** is necessary to understand the value when dealing with quantitative data. The unit of measure must be expressed clearly and in plain sight, and the scale must be consistent with concise and meaningful labels. If the value variation causes a change within the representation, e.g., about color, this must be indicated by a brief yet complete **legend**. Furthermore, in the case of very large numbers, in the hundreds of thousands or higher, if you decide to use a small scale (x out of 10 to indicate millions of people), you must provide an easy way to obtain the total value.

Labels should be used to indicate non-quantitative data. The length of these labels might influence the graph itself: for example, in the case of a bar chart, the orientation of the graph might be chosen based on the length of the labels, as long labels might be better displayed in horizontal bar charts than vertical ones.

Sometimes, it is also necessary to add **descriptive text** to understand the visualizations clearly. Such text should be concise, simple, and accessible to all users inside the target audience. People's attention or the viewing timespan might be too brief to read a long text. Thus, the representation should not solely rely on the text to be understood.

Use easily readable fonts with a large font for all types of text.

5.2.9 Accessibility

Data visualizations on small-screen devices should be accessible to all users, even the ones with a disability.

In particular, since visual representations highly rely on colors, it is necessary to consider *colorblind people*, namely people with a Color Vision Deficency (CVD). CVD is a medical condition that causes a decrease in the ability to see colors or differences in colors **[147]**.

Therefore, it is helpful to use different *colorblind* filters that allow some CVD people to see differences in colors clearly or try to avoid putting colors such as red and green together.

Moreover, when possible, add ALT text to the images [**113**], which is an "alternative text" that can be visualized interacting with the image and that explains the content of the visualization, also giving some context. It is advantageous in situations of low connectivity on the web when images do not load correctly, as well as for blind people.

5.3 Specific Guidelines for Social media

In social media, it is possible to share images and videos and, thus, are perfect vehicles for the diffusion of data visualizations. However, the representations shared on these platforms have to meet specific requirements.

First, social media do not support interactive visualization; they only allow static images, GIFs, and videos. The only interactions possible with the pictures are zooming (which does not change the image's resolution) and other touch-based actions that express a "sentiment" concerning the content shared (but don't change the visualization). Moreover, representations do not change when passing from one environment to another, such as from smartphones to computers; social media do not support scaling of the images according to the size of the display; the app's interface changes, but the visual content does not.

Therefore, all the rules in the previous section regarding static visualizations apply to the representations shared on social media.

This also implies that composite content does not change layout neither when passing from smartphone to tablets or computers nor when the orientation of the mobile device is changed from landscape to portrait; therefore, it is best to produce content with either one or a few elements suitable for all types of devices.

In general, social media are used for tasks that happen in a few seconds timespan; as a consequence, all visualizations should be developed to highlight the primary information immediately and use colors and graphs that catch the viewer's attention in a fast way, and this may push the reader to explore the discussed topics with a link pointing outside the social platform.

Images shared on social media should be accompanied by a caption, a brief text that gives additional context to the visualization. However, even in this case, the text should be concise because the viewing timespan on social media is very short.

As for the other guidelines defined in the previous sections, they apply to social media representations as well. However, particular attention should be given to the context, and the clarity of the data shared, as social media are affected by the diffusion between fake news and data; It is, therefore, necessary to ensure that an ethical and transparent use is made of data and representations and to avoid confusing or misleading content.

Chapter 6

Case study

6.1 Introduction

Social media are not apps strictly dedicated to data visualizations. Therefore, they are not designed to be environments to host interactive representations or scientific platforms. However, their potential is unique as they are widely diffused in the civilized world and used by millions of people from different cultures, ages, and political ideologies. Consequently, they can reach large audiences and, if used correctly, can be very useful in disseminating knowledge even for people who are not experts.

This chapter is dedicated to the analysis of the case study chosen for the application of the guidelines defined in Chapter 5. The guidelines are used here for the creation of representations of the data collected by the low-cost data collection system described in Chapter 3.

Chapters 1 and 3 indicate that the data collected concerned Venice's environmental parameters (temperature, humidity, and carbon dioxide concentration levels).

Therefore, the case study refers to the manipulation and transformation into visual representations of these environmental data collected in Venice through the guidelines for small screen devices to be published on a social media platform, specifically **Instagram**.

After the representation and publication phases, it was necessary to start evaluating the visualizations to assess their level of engagement and, above all, whether and how they affected the environmental awareness of the users who came into contact with them. This process was conducted using two evaluation systems: firstly, using the **insight tools [84]** provided by the social media platform itself and then by submitting an **engagement survey** to the people who viewed the representations on social media and those who had been interviewed for the initial assessment survey.

6.2 Why Instagram

The choice of **Instagram** as the social media platform used as a case study was based on several motivations.

Instagram is a widely diffused social networking service for photo and video sharing owned by Meta Platforms [**86**] (the same company that, among other things, owns the social networking platform Facebook and the messaging application service WhatsApp). The application allows users to upload multimedia content that can be edited with filters and organized by hashtags and geographic tags. It differs from other social platforms in that it is focused on visual content: each post must contain an image and a video that a text can accompany. This leads to a high level of engagement, as images are more attractive than text [**134**]. Instagram is a reasonable choice as the project aims to create and publish data visualizations.

Furthermore, according to WeAreSocial's Digital Report published in January 2023 [**126**], Instagram ranks fourth as the most used social media in the world (behind Facebook, Youtube, and WhatsApp), with 2 billion monthly active users. In contrast, it ranks second (behind WhatsApp) when it comes to favorite social media, especially by people ages 16 to 24.

In Italy, Instagram ranks third as the most used social media (behind WhatsApp and Facebook), whereas it ranks second as the favorite platform (behind WhatsApp) [**127**].

Such data were corroborated by the results of the initial assessment survey: as indicated in Chapter 4, 26 out of the 30 people interviewed (**86.7%**) stated that they were registered to Instagram, which was the highest percentage of active users concerning the other platforms.

Finally, as discussed in Chapter 2, Instagram is used by several media outlets for publishing news, data, and data visualizations that help raise awareness on important topics or disseminate information to the general audience.

Therefore, Instagram is the right platform for publishing visual representations of environmental data, as it can reach broad audiences and is based on visual content.

6.3 Data Analysis

The first step before creating the representations is understanding the collected data. However, a complete statistical analysis of the collected data is beyond the scope of this thesis, which instead focuses on providing tools to facilitate the analysis and representation of the data, i.e., the guidelines presented in Chapter 5. Therefore, the study conducted here is restricted to the carbon dioxide values collected by the collection system, which are the main values that will be disseminated through the representation process to raise awareness about environmental issues.

Chapter 3 described in detail how the low-cost data-gathering system in Venice works, and in particular, Table 3.1 showed the types of environmental data collected. As the name and the coordinates of the locations of the units are fixed, the dynamic parameters are the ones collected by the system:

- *Date*, indicated as **YYYY/MM/DD, hh:mm:ss**;
- *CO_2 concentration*, measured in **ppm**.
- *temperature*, measured in Celsius degrees (**°C**).
- *humidity*, measured in percentage (**%**).

See Fig. 6.1 for an extract of the data from one of the sheets.

Date	Temperature	Humidity	CO2	Site	Latitude	Longitude
2023/03/01, 00:00:11	3.40	63.90	388	CampusS	45.47827626	12.25536381
2023/03/01, 00:15:40	3.40	64.50	387	CampusS	45.47827626	12.25536381
2023/03/01, 00:31:02	3.50	65.00	383	CampusS	45.47827626	12.25536381
2023/03/01, 00:46:09	3.60	65.30	384	CampusS	45.47827626	12.25536381
2023/03/01, 01:01:29	3.70	65.60	380	CampusS	45.47827626	12.25536381

Figure 6.1: Extract of the repository on google sheets containing the data collected by the system.

The key point was to evaluate the data to understand the variance in the concentration levels of CO_2. To do so, it was helpful to analyze the mean values for each station from March to the beginning of June, reported in Table 6.1. The values are compared to the mean values observed at Mauna Loa Observatory [**81**].

Unit	**March**	**April**	**May**	**June**
Briati	479	457	462	401
Ca' Foscari	525	422	423	426
Campus S.	432	428	427	431
S. Basilio	438	431	435	(N.A.)
S. Giobbe	522	475	545	387
S. Margherita	453	459	430	422
Mauna Loa Obs.	**421**	**423**	**424**	**424**

Table 6.1: Mean values at the six stations during the collection period compared to the Mauna Loa Observatory means data in the same period.

Almost all values are slightly above the mean values observed at Mauna Loa, but it is possible to observe a somewhat stable or descending pattern as we pass from March to

June; as stated in Chapters 1 and 2, the descending pattern is caused by the carbon dioxide segregation processes that start in late spring/early summer. However, some exceptionally high mean values stand out, which correspond to March at Ca'Foscari and March, April, and May at San Giobbe. In fact, by analyzing the complete monthly datasets, it is possible to identify periods within these months when CO_2 values are inexplicably high.

It is important to note that all six units were unable to record observations from May 9^{th} to May 19^{th} due to a prolonged failure in the network connection. Therefore, the mean values calculated for May are higher than the previous ones as they lack a third of the observations. The unit at San Basilio is still not working at the time of writing.

The particular cases of inexplicably high values of carbon dioxide adjusted by the collection units are presented in the following subsections. Even in this case, producing hypotheses to explain the causes of these events is beyond the scope of this thesis, except in the case of proven technical problems with the instrumentation. To provide possible explanations beyond the variance of temperature and humidity values, it would be necessary to add sensors (such as wind, rain, and particle detectors) and other instrumentation to capture external events or human factors.

Ca' Foscari - Particular cases

The unit at Ca' Foscari recorded several values higher than the Mauna Loa Observatory's mean values, especially at the first time of the month. Analyzing the mean values for each day of March at this location, it was possible to see that there were 22 mean values that range from 500 to 609 ppm.

San Giobbe - Particular cases

A somewhat different situation concerns the St Job's unit. The unit recorded observations with significantly higher values than the mean observed at Mauna Loa **[81]**. Such values can be observed for prolonged periods, even several days in a row, for then drastically dropping to a low value near the Mauna Loa mean.

The four periods in which this happened are:

- from March 19^{th} to March 21^{st}: the values started slowly raising at midday of March 19^{th}, stabilized between 900 and 1000 ppm, with a peak value at **1080 ppm** and then dropped drastically at 430 ppm at the end of March 21^{st};
- from April 29^{th} to April 30^{th}: the values rapidly increased around midday of April 29^{th}, stabilized between 900 and 1100 ppm, with a peak value at **1121 ppm** and then dropped drastically at 530 ppm at the end of April 30^{th};
- from May 5^{th} to May 8^{th}: the values rapidly increased early in the morning of May

5^{th}, stabilized between 1000 and 1200 ppm, with a peak value at **1177 ppm** and then dropped and stabilized between 800 and 900 ppm drastically at 454 ppm in the early morning of May 8^{th};

- from May 19^{th} to May 23^{nd}: the values started to increase slowly during May 18^{th}, stabilized between 700 and 900 ppm, with a peak value at **875 ppm** and then dropped suddenly at 418 ppm early in the morning of May 23^{nd}.

Auditorium Santa Margherita - Particular cases

Another particular situation occurred at the unit of Auditorium Santa Margherita. In correspondence of 5 specific days of 2023, namely March 3^{rd}, March 29^{th}, March 31^{st}, April 20^{th} and April 21^{st}, peaks of high values ranging from 800 ppm to 1000 ppm can be observed. These values are very different from the observed values (tending to be between 400 and 500 ppm).

Scientific Campus - Particular cases

A particular case concerns some data registered by the Scientific Campus'unit. As the unit is placed in a location directly exposed to the Sun, the values for temperature recorded in peak hours are significantly higher than the mean temperature in the area (30 to 50 °C during February); It has been established that this event is due to an instrumentation problem, specifically the sensor that detects temperature and humidity, which heats up more than it should in direct contact with sunlight. As a result, such data are not entirely reliable.

6.4 Representations

The next phase consisted of creating representations suitable for small screens and social media to convey the collected environmental data.

Two different types of graphs, *radar chart* and *histograms*, were selected for the representations.

6.4.1 Radar chart

Since we want to display **quantitative data** and compare the values collected by the different stations, the radar chart was an interesting choice of the graph.

"A radar chart plots values across multiple quantitative variables for one or several categorical items to enable general pattern forming" **[65]**. As shown in Fig. 6.2, it consists of a radial layout with concentric wheels that represent the value bands; on the outer circle, the categorical items are put, and the quantitative values are plotted along each scale under the corresponding item. The points are joined using lines; the final result will be a geometric shape inside the radial layout.

There are many variants of this type of chart, as it can be used to display a single dataset or several ones to make comparisons between one another.

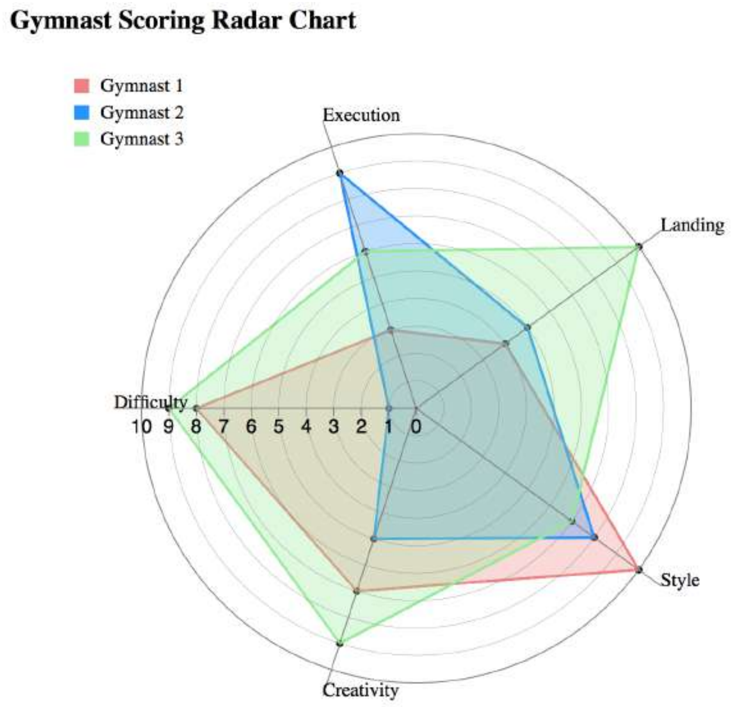


Figure 6.2: Example of a radar chart [97].

Consequently, plotting the carbon dioxide data is a great choice, as it is simple and direct. Moreover, it is possible to manipulate the lines to add color to distinguish the various entities.

The data frame to represent in the chart are:

- **carbon dioxide** data, which serves as quantitative data to put inside the radial layout;
- the **timeframe**, which serves as categorical items to put in the outer circle;

Since we have to follow the key concepts of **simplicity**, **immediacy**, and **conciseness**, the timeframe had to be chosen so that it could allow us to understand the data at hand. Therefore, it was selected to represent **a month** of data for each graph; consequently, the categories on the upper circle will be described by the days of the indicated month.

To respect the concept of **consistency** and allow for better comparisons, the scale of the quantitative values was adjusted to cover all the possible values observed in the indicated month by any of the stations.

To respect the concept of **appeal**, colors were used to distinguish the various elements in the graph; the color green was chosen to represent the values of carbon dioxide, which were plotted as a continuous line in the chart (except some blank portions amounting for the missing values).

To comply with the **trust** concept, the Ca' Foscari University logo was added to the upper right corner of each representation.

As the representation was developed for small screens and social media, it was necessary to use big fonts for the texts and the scales.

The graph was realized using the graphical features of the Excel Spreadsheet and the additional features of Kutools for advanced analysis and graphic design. The prototype of the representation can be seen in Fig. 6.3 below.

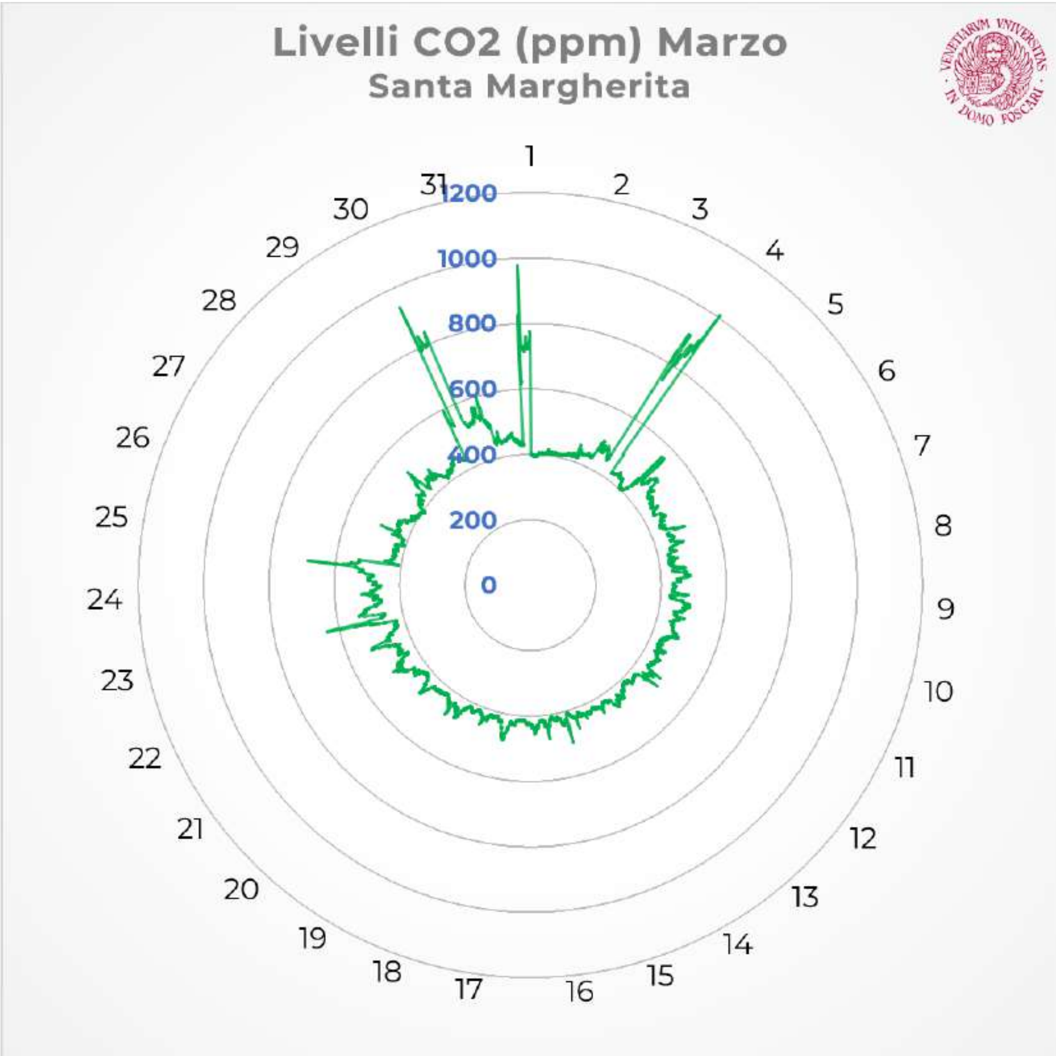


Figure 6.3: First representation of carbon dioxide data gathered at Auditorium Santa Margherita during March 2023.

This prototype was simple and immediate but lacked some context, left a lot of white space on the side, and was not so visually appealing. The second version of the representation tries to fix these issues, by adding also another variable. The choice fell on **meteorological**

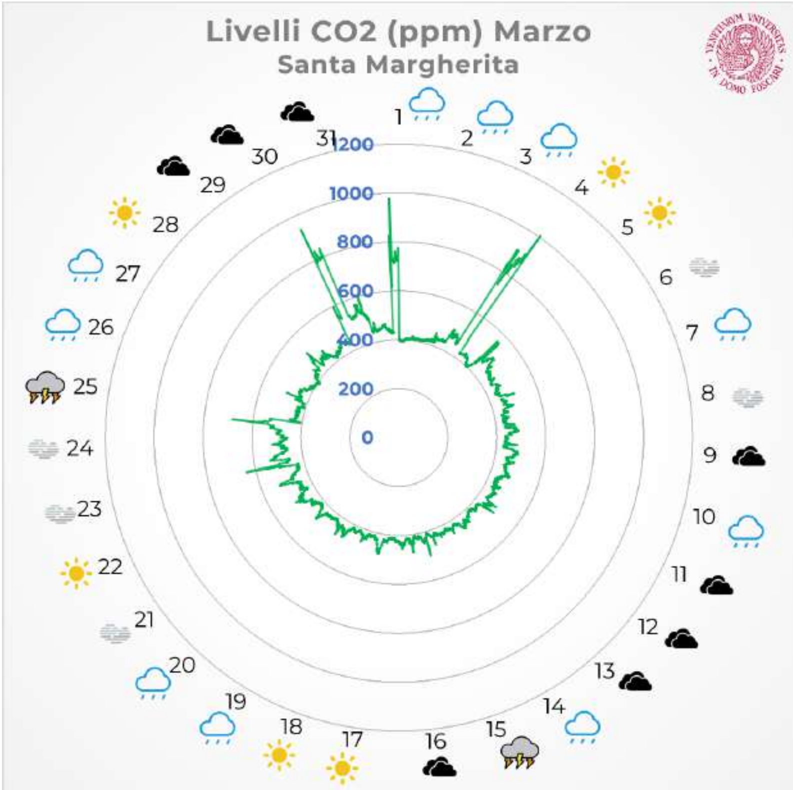
conditions of the days represented, as they could give additional insights regarding the cause of the variables concentration of CO_2. Those data were extrapolated from the historical repository of ilMeteo.it [**59**]. To make the representation readable and more engaging, such data were added using icons of weather conditions. It was necessary to calibrate the size of such elements to neither overload the representation by making them too big nor add futile small details that cannot be seen on a small screen.

Two examples of the final design for the representation of the environmental data can be seen in Fig. 6.4. These representations comply with the guidelines and fundamental concepts defined in Chapter 5.

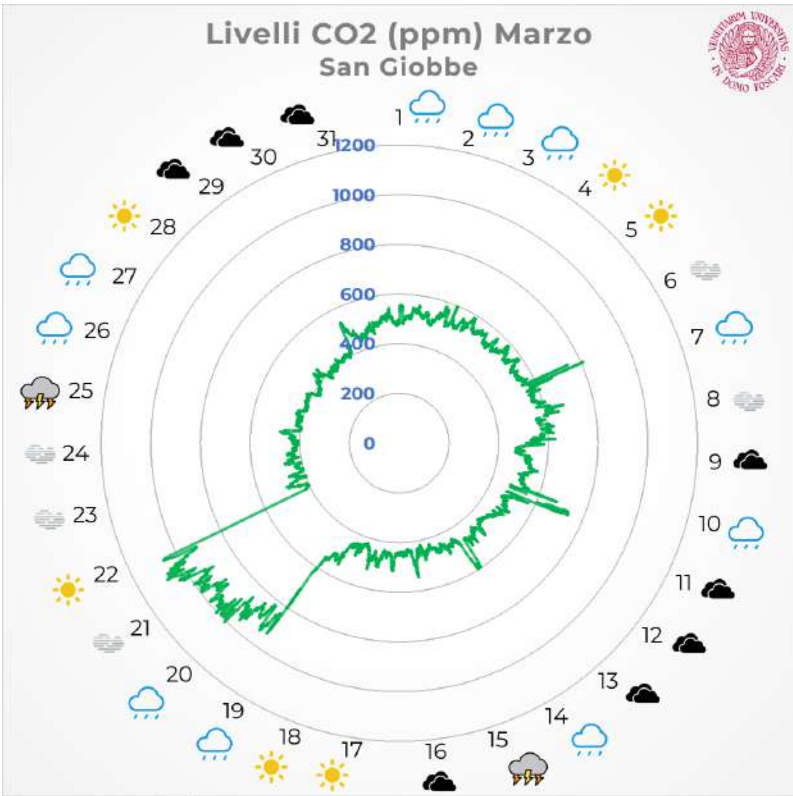
It was then interesting to produce another type of radar chart, which maintained the same format as the first one but changed the data to represent. Instead of showing all the data collected by a station during the selected month, the representation shows the mean for each day of the month as a point in the graph under the corresponding item/day.

The points were colored using a different color for each band, starting from green in the centre up to 400 ppm, then passing to light green for values between 400 and 500 ppm, then yellow between 500 and 600 ppm, orange between 600 and 800 ppm and shades of darker reds for values greater than 800 ppm (as can be seen in Fig. 6.5b). This choice was made for helping the users to identify more quickly the different intervals of carbon dioxide concentration.

In ore to ease comparisons among the different gathering points, the same scale was used for all the representations. Two examples of the final design for this type of graph can be seen in Fig. 6.5.

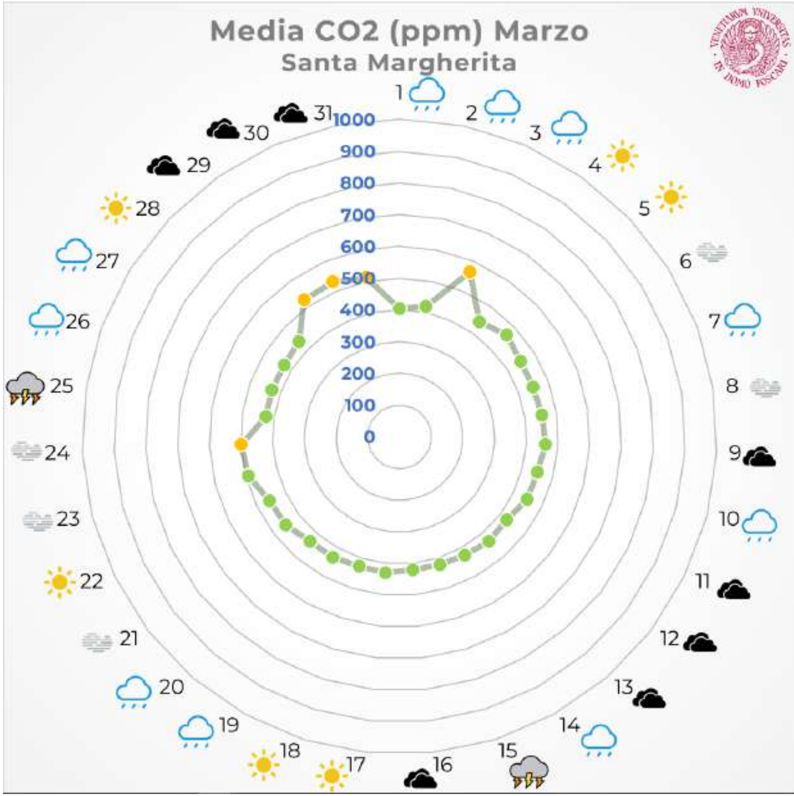


(a) Ex. 1: CO_2 levels at Auditorium S. Margherita during March 2023

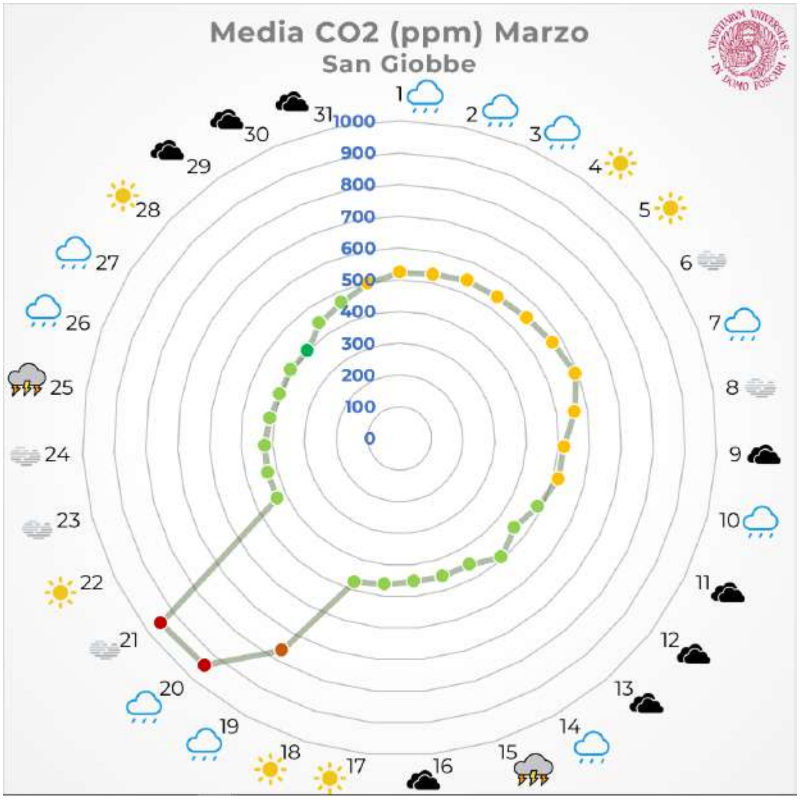


(b) Ex. 2: CO_2 levels at S. Giobbe during March 2023

Figure 6.4: Examples of the final Design Radar Chart showing the total CO_2 values across a month.



(a) Ex. 1: CO_2 mean levels at Auditorium S. Margherita during March 2023



(b) Ex. 2: CO_2 mean levels at S. Giobbe during March 2023

Figure 6.5: Examples of the final Design Radar Chart showing the mean CO_2 values across a month.

6.4.2 Histograms

Histogram, a chart that "displays the frequency and distribution of quantitative measurements across grouped values of data items." **[65]**, were selected as an alternative to radar charts. It consists of discrete bins that form ordinal value groupings. It is plotted on a Cartesian plane, so the girds corresponding to the two axes can be plotted or excluded from the graph based on which elements we want to include in the representation.

Even here, the same parameters as before are applied to produce the visualization. The x-axis represents the timeframe, selected as a weekly interval, whereas the y-axis represents the frequency of the ordinal value, namely the levels of Carbon Dioxide.

Even in this graph, the scale of the quantitative values was adjusted to cover all the possible values observed in the indicated month by any of the stations to respect the concept of **consistency** and allow for comparisons.

To respect the guideline related to **appeal**, colors were used to distinguish the various air quality bands in the graph; the color green was chosen to represent the values of carbon dioxide up to 600 ppm, yellow to represent the values between 601 and 900 ppm and orange to represent values over 900 ppm. Moreover, to distinguish the station from which the data came, a semi-transparent image of the unit's location was selected as the background of the representation.

To ensure **conciseness** and **immediacy**, a legend explaining the representations' colors was put in the left-bottom corner of the chart.

To comply with the **trust** concept, the Ca' Foscari University logo was added to the upper right corner of each representation.

As the representation was developed for small screens and social media, it was necessary to use big fonts for the texts and the scales.

The graph was realized using the graphical features of the Excel Spreadsheet and the additional features of Kutools for advanced analysis and graphic design. Examples of the final design of the histograms are shown in Fig. 6.6.

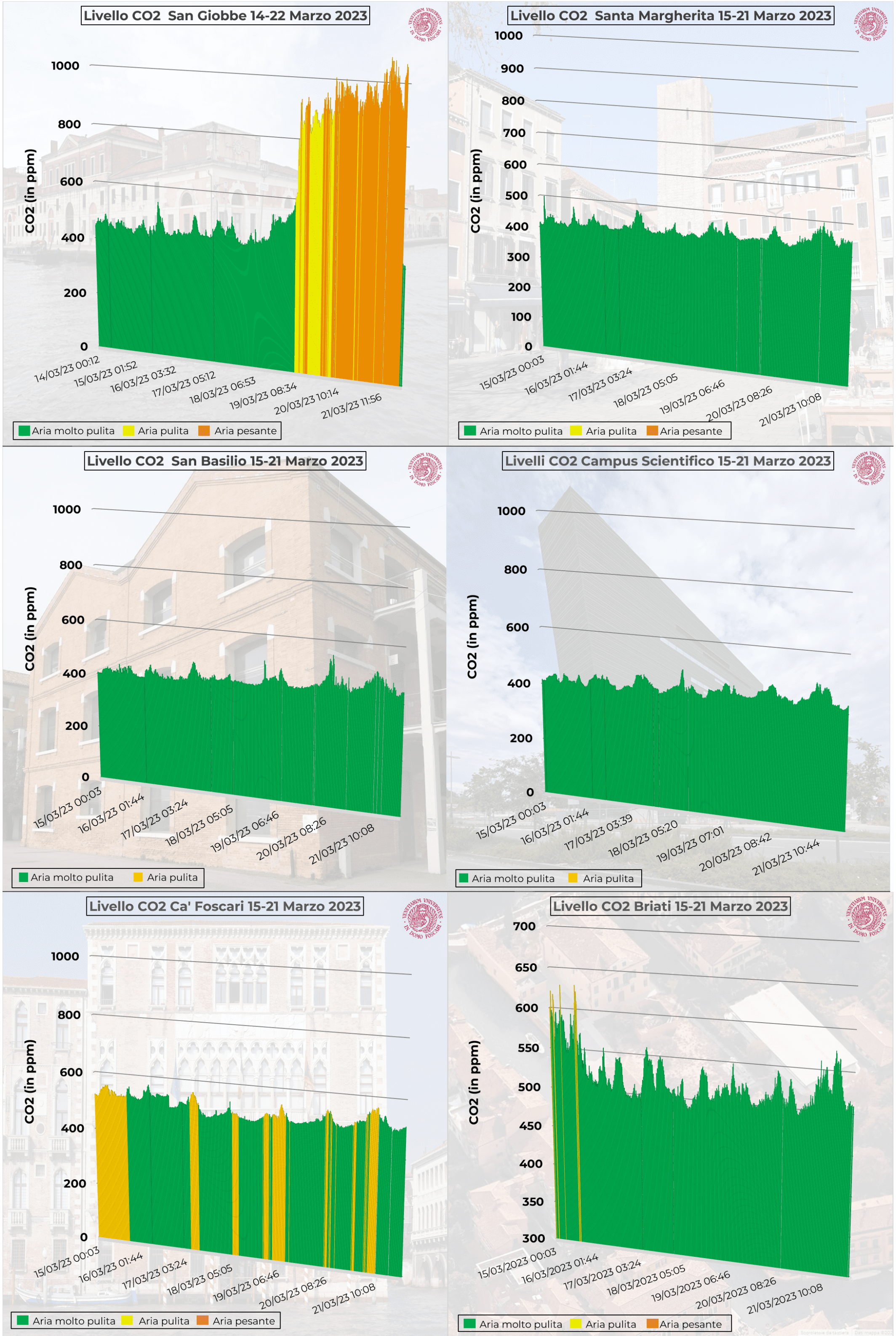
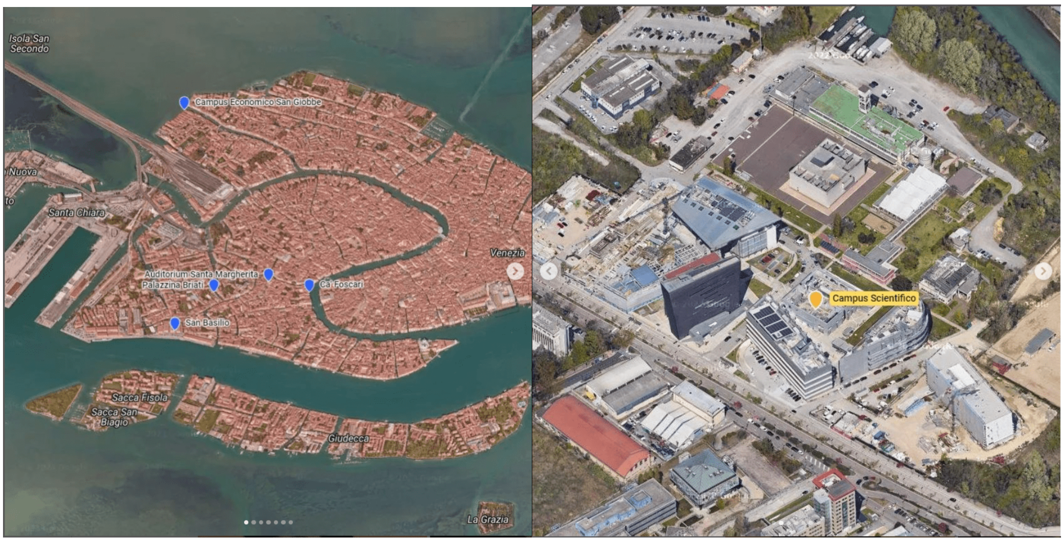


Figure 6.6: Final Design Histograms: CO_2 levels at the six units during the third week of March 2023

6.4.3 Additional representations

An additional representation was developed to give social media users more context when entering the publishing phase. Such visualization consists of two simple square maps showing the location of the data collection system, with pins on the actual coordinates of each unit. They are simple as they offer an inclined satellite view of the island without indicating roads or unnecessary POIs. For the prototypical representation we took advantage of Google Maps images. They are shown in Fig. 6.7.



(a) Venice map. (b) Scientific Campus Map

Figure 6.7: Maps designed for social media.

6.5 Publishing phase

The publishing phase consisted in diffusing the representations created on the social media case study Instagram. In order to test the results in a real situation, the personal profile of the author of this thesis was temporarily used. However, the data are intended to be published elsewhere (e.g., the official profiles of the Ca' Foscari University of Venice) to obtain results for a wider audience.

The representations were published in this order:

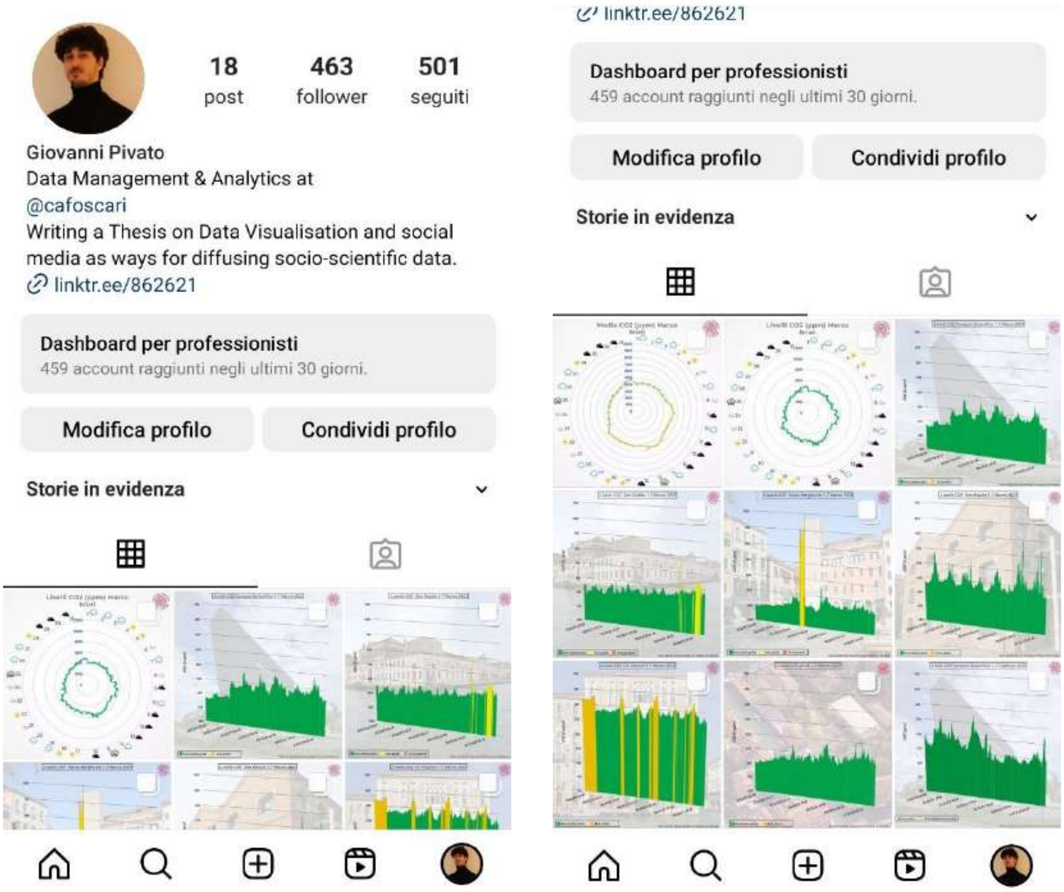
- the two map representations using a carousel post, accompanied by a text explaining the scope of the project and what the visualizations represented;
- the histograms representations using a separate carousel post for each of the stations; each carousel contained eight representations, one for each week ranging from March

to April, accompanied by a text that explained a bit of context of the building hosting the unit and then explaining the data displayed.

- The radar chart with weather indications of the complete data gathered during March 2023 using a carousel post containing all six charts of the corresponding stations, accompanied by a text explaining the visualization.
- The radar chart with weather indications of the mean values gathered during March 2023 using a carousel post containing all six charts of the corresponding stations, accompanied by a text explaining the visualization.

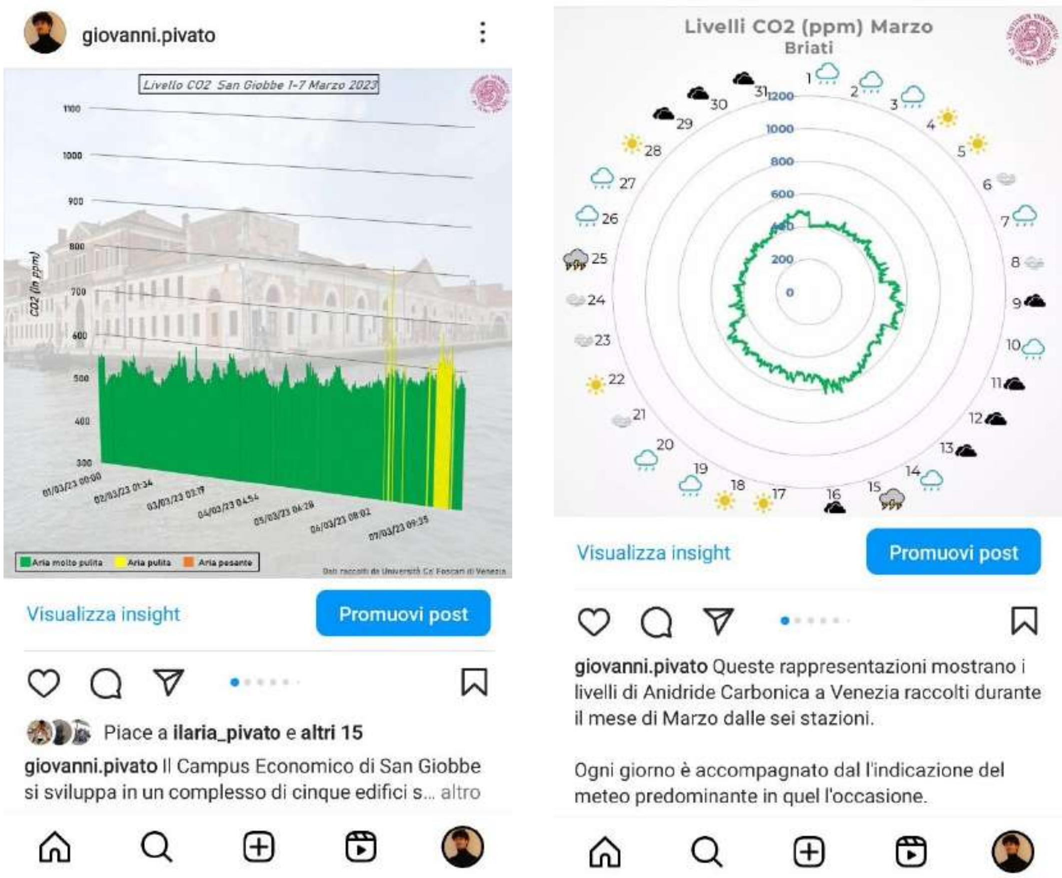
To ensure trust and accessibility, the link to the data repository was added to the profile information. The data were shared only with a visualization permit, excluding the possibility that someone manipulates the data without consent.

Each post contained **hashtags** related to the environment, Ca' Foscari, and data visualization to widen the diffusion of the representations on the platform.



(a) Instagram Profile Information. (b) Instagram Profile Grid

Figure 6.8: Instagram Profile where data were published.



(a) Histogram of San Giobbe during the first week of March.

(b) Radar Graph of total value during March

Figure 6.9: Post view of the representations.

6.6 Validation phase

The project's final phase consisted of evaluating the representations using feedback from users that came into contact with it. This process was conducted in two ways: first, by submitting an engagement survey to a sample of the people who saw the representations and then by analyzing the insights of the post on Instagram gathered using the tools for a professional account.

6.6.1 Feedback from Engagement survey

The Final User's Engagement Survey was developed using Google Forms. It was designed to evaluate the level of engagement, clearness, and adequacy of social media of the data visualizations shown in the previous sections.

The survey assesses the level of engagement and appropriateness of these representations to the mentioned context.

This survey was conducted by analyzing the work of O'Brien et al. [**99**], who defined a set of 31 statements divided into four categories capturing the six dimensions of engagement, namely **aesthetic appeal**, **focused attention**, **novelty**, **perceived usability**, **felt involvement** and **endurability**. O'Brien et al. [**99**] proposed a short version of these statements, which included 12 items referring to the four categories: **Focused Attention** (FA), **Perceived Usability** (PU), **Aesthetic Appeal** (AE) and **Rewarding Factor** (RW); the latter one summarizes felt involvement, endurability, and novelty. All the items are meant to be graded using a Likert Scale [**82**] ranging from 1 to 5, as shown in Table 6.2:

Strongly Disagree	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
1	2	3	4	5

Table 6.2: Likert scale used to grade items of the User Engagement Scale (UES) [99].

Fig. 6.10 shows the short form of the User Engagement Scale (UES-SF). As we can see, the scale refers to the engagement of an application. Still, it could be easily adapted to the visual representations created for Instagram and their level of engagement.

FA-S.1	I lost myself in this experience.
FA-S.2	The time I spent using Application X just slipped away.
FA-S.3	I was absorbed in this experience.
PU-S.1	I felt frustrated while using this Application X.
PU-S.2	I found this Application X confusing to use.
PU-S.3	Using this Application X was taxing.
AE-S.1	This Application X was attractive.
AE-S.2	This Application X was aesthetically appealing.
AE-S.3	This Application X appealed to my senses.
RW-S.1	Using Application X was worthwhile.
RW-S.2	My experience was rewarding.
RW-S.3	I felt interested in this experience.

Figure 6.10: UES-SF items devised by O'Brien et al. [99]

The O'Brien UES-SF **[99]** was adapted to the case study, i.e. visual representations of CO_2 data and used to evaluate the representations. In addition to the 12 items of this survey, two more questions were added to assess the appropriateness of the representations to the social media context, as per the guidelines defined in Chapter 5.

The 12 items **[99]** were graded by the users who took the validations three times, one for each of the three types of representations they were asked to examine. The three representations that users were asked to grade are those shown in Fig. 6.4a (Total Radar Chart of March at Santa Margherita), 6.5a (Means Radar Chart of March at Santa Margherita) and 6.6 (Histograms).

The additional two items included a multiple-choice question, which allowed the user to choose from five possible options concerning those items, and an open-ended question asking the user to explain their choice. The users had to give their feedback only once.

The twelve items of the O'Brien UES-SF **[99]** can be found in Table 6.3, slightly adjusted to serve the case study; the wording "Representation X" is replaced by the number indicating the representation, i.e. "1" for the histograms, "2" for the total monthly radar chart and "3" for the monthly averages radar chart. The two additional items are reported in Table 6.4. The survey was submitted to 22 social media users that came into contact with the representations. These users are a subset of the group to whom the initial assessment questionnaire was submitted. Therefore, they are aged 20-30; 14 of them (**63,64%**) are students, whereas the other 8 (**36,36%**) are workers; 13 of the respondent are men (**59,1%**), whereas 9 are women (**40,9%**).

Following the instructions given by O'Brien et al. **[99]**, the responses given by each

Category	Items
FA-S.1	I lost myself in this experience.
FA-S.2	The time spent viewing Representation X just slipped away.
FA-S.3	I was absorbed in this experience.
PU-S.1	I felt frustrated while assessing Representation X.
PU-S.2	I found Representation X confusing.
PU-S.3	Viewing Representation X was taxing.
AE-S.1	Representation X was attractive
AE-S.2	Representation X was aesthetically pleasing.
AE-S.3	Representation X appealed to my senses.
RW-S.1	Viewing Representation X was worthwhile.
RW-S.2	My experience was rewarding.
RW-S.3	I felt interested in this experience.

Table 6.3: UES-SF [99] adapted to the case study and submitted for the three selected examples.

Additional items	Answer format
Which representation is more suitable for social media?	1) Representation 1; 2) Representation 2; 3) Representation 3; 4) All of them; 5) None of them.
Explain your answer.	Open question.

Table 6.4: Additional items to the survey used to evaluate the appropriateness of the representation to social media.

respondent for each item in each category (Focused attention, Perceived Usability, Aesthetic Appeal, and Rewarding Factor) were averaged; in this way, it was possible to obtain a unique distribution of values for each category. This operation was done for each of the three representations tested. The results were then plotted using *Tukey Box Plots*, which display "the distribution and shape of a series of quantitative values for different categories" **[65]**; each box corresponds to a different category and it shows several statistical values (first quartile, median, third quartile, mean, max, and min). The plots are shown in Fig. 6.11, 6.12, and 6.13.

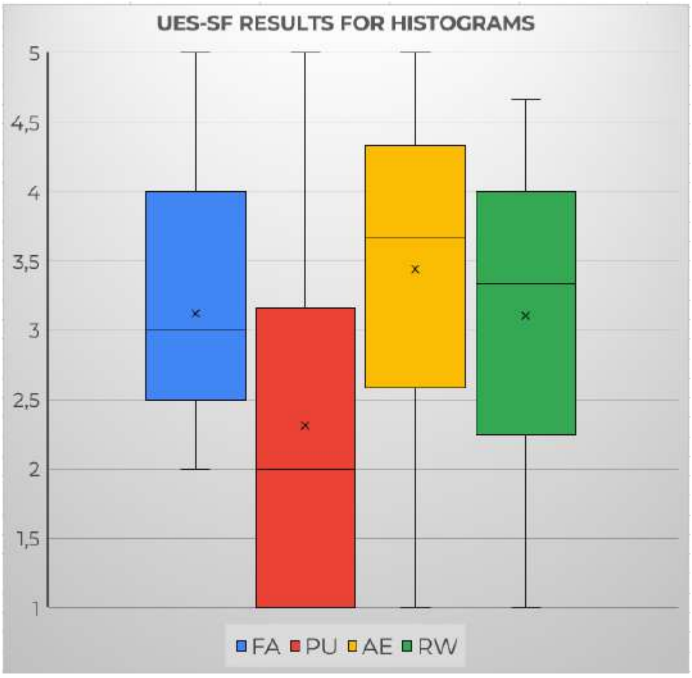


Figure 6.11: Tukey box plot displaying the distribution of mean values for the categories evaluated for the Histograms (FA, Focused Attention; PU, Perceived Usability; AE, Aesthetic Appeal; RW, Rewarding factor)

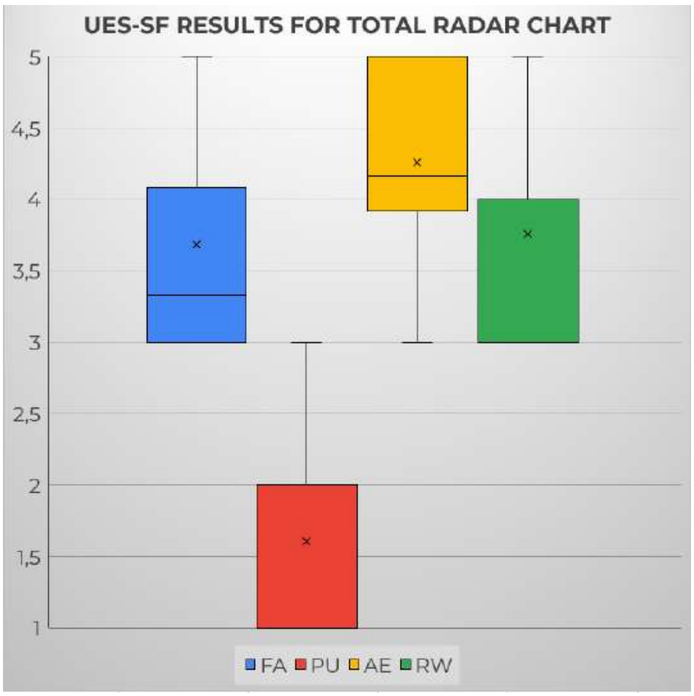


Figure 6.12: Tukey box plot displaying the distribution of mean values for the categories evaluated for the Total Radar Chart (FA, Focused Attention; PU, Perceived Usability; AE, Aesthetic Appeal; RW, Rewarding factor)

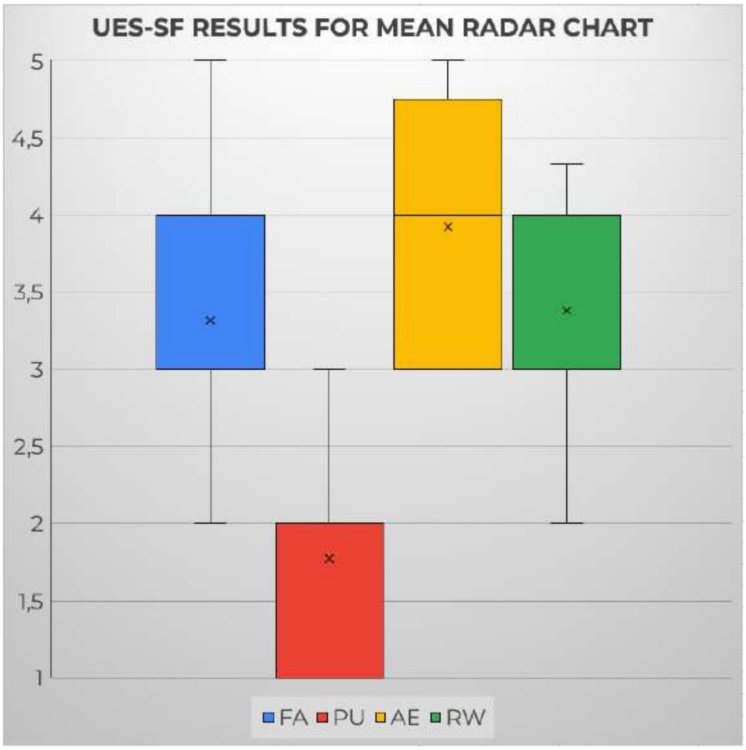


Figure 6.13: Tukey box plot displaying the distribution of mean values for the categories evaluated for the Means Radar Chart (FA, Focused Attention; PU, Perceived Usability; AE, Aesthetic Appeal; RW, Rewarding factor)

According to the analysis, the histogram chart was perceived as the less usable, as the mean and median for its PU are the highest.

The total radar chart with weather indications is the one that performs better in three of the four categories, as its mean (the ”x” inside the box) and median (horizontal line inside the box) are the highest for PU, AE, and RW. In particular, the total radar chart performs very well in the PU and AE categories and it is the only chart that reaches ”5” as the max value for the RW category. Focused Attention (FA) obtained similar scores for all the three representations.

The means radar chart performs very similarly to the previous one, being just below in AE and RW. Its mean for PU is the lowest (since it is measured using negative items, a low median/average indicates positively perceived usability).

Overall the total radar chart was perceived as the most engaging representation, aesthetically pleasing yet easy to understand, novel, and capable of involving emotionally the users even in the long term.

The second part of the survey evaluated the appropriateness of the three charts to the social media context; According to the results, shown in Fig. 6.14, most of the users think the total radar chart is the most appropriate for social media (**45,5%**), followed by the means radar chart, and finally the histograms. Only a small percentage of the users think all of the representations are suitable for social media. Users were asked to explain the reason for their choice; the answers are shown in Table 6.5.

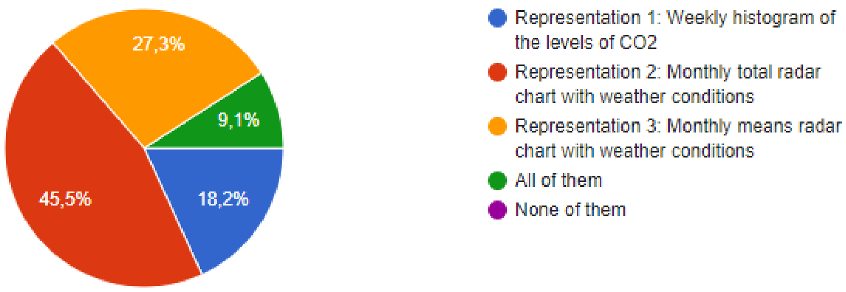


Figure 6.14: Pie chart showing which is the most suitable chart for social media.

Answer	Motivation
Histograms	The graph is very visible and in social media will attract the user. It is the one that catches the eye better. It gives more context thanks to the background of Venice. The background and the use of green-yellow-orange.
Total radar chart	Because it is simple even if it has many elements. It is more interesting as it has weather; the background is simple. The total radar chart is more simple and more direct. It presents the data better, displaying the daily peaks clearly. It is the most attractive and simple. It is the most appealing. I think the green line and the weather icons will attract viewers. It is the less confusing. It is clear and simple and can be understood by anyone. It is the one more understandable on social media.
Means radar chart	Best representation to display on mobile devices. It uses color in a more clever way. The mean data are the clearest, thanks also to the colors of the dots. The colored dots explain the difference in the risk. It is more clear and uses different colors to distinguish the levels. I think is the one that explains the data better.
All of them	They are immediately understandable. Easy to understand, both in terms of colors and symbols.

Table 6.5: The reason for the choice of the most suitable chart for social media.

Overall the answers to the final two questions confirmed the preference for the radar chart representations, and add to the previous analysis the importance of providing additional data, like the weather conditions, for attracting users and stimulating reflections.

6.6.2 Feedback from Instagram insights

Other interesting data regarding engagement could be collected by analyzing the **Insights** [**84**] provided by Instagram for each post. These tools measure how many accounts have seen the content on the platform, how many have interacted with them, and how many new followers have been added after viewing such content.

The most important insight regards the **coverage of the post**, which is divided into two elements:

- **Reached Accounts**: Number of unique accounts that have seen the content **at least once**;
- **Impressions**: Number of times the content was seen, including multiple views by the same accounts.

The insights for the various representations published on Instagram can be seen in Table 6.6 in order of publishing. The data are updated to June 16^{th} 2023.

Post	**Account reached**	**Impressions**	**Posted**
Hist. Campus February	241	332	May 9^{th} 2023
Hist. Briati March	211	293	May 16^{th} 2023
Hist. Ca' Foscari March	206	285	May 18^{th} 2023
Hist. S. Basilio March	197	277	May 19^{th} 2023
Hist. Auditorium March	187	270	May 20^{th} 2023
Hist. S. Giobbe March	177	250	May 24^{th} 2023
Hist. Campus March	139	209	May 25^{th} 2023
Radar Total March	208	279	June 7^{th} 2023
Radar Mean March	130	200	June 7^{th} 2023

Table 6.6: Instagram Insights for the published representations.

As we can see, the representations have been seen by many accounts; they perform very well in terms of reached accounts and impressions, considering that the account used for the publishing is not a certified account of the Ca' Foscari University of Venice.

In spite of the fact that numbers are higher for histograms, it is useful to mention that radar charts have been published later and that they have reached the same numbers of histograms in less time. Overall the results seem to confirm the results of the engagement survey.

The representations will be constantly monitored to see if the values of reached accounts and impressions will grow further in the coming weeks.

6.7 Conclusions

Overall, the representations performed well as a vehicle for disseminating environmental knowledge, raising awareness on social media, and as engaging examples of visual representations of scientific data. The results of the validation phase indicated that users were attracted by the data communicated as graphical representations and that the selected representations are suitable for social media. Therefore, this is an indication of the appropriateness of the guidelines for data visualizations on small screens defined in Chapter 5.

Chapter 7

Conclusions and future work

In conclusion, the project discussed here first aimed to create a low-cost environmental data-gathering system in Venice to collect data regarding temperature, humidity, and, most importantly, the concentration of Carbon Dioxide in the atmosphere. Such a system was created by upgrading an existing system, substituting its old components with new ones, and adding new features, such as the connection to a PoE grid and Internet to perform real-time data transmission to Google Sheets that served as a repository.

Secondly, the project envisaged using this data to create innovative visual representations that would increase the viewer's environmental awareness of the city of Venice.

Specifically, this thesis focused on creating visual representations suitable for small-screen devices, mainly social media, using a series of guidelines explicitly designed for the case under analysis.

To do this, four different stages were carried out. The first was to analyze the literature on the four main topics covered in the thesis: data visualization, environmental awareness, guidelines for visualizations on small screens, and social media as platforms for scientific data dissemination.

Following this, an initial assessment survey was created to understand the level of information of people between 20 and 30 years (who are the most significant users of social media), their views on the reliability of social media as a tool for the dissemination of data, and information and opinions regarding the representation of data using graphs and images. The survey showed that young people are inclined to use social media to disseminate socio-scientific data but are aware that these platforms are only partially reliable due to the problem of fake news. They also agree that using simple visual representations to present complex data is a great way to overcome the knowledge barrier for non-expert people in the socio-scientific sector.

Both the literature analysis and the results of the initial assessment questionnaire were used to develop a set of guidelines for the representation of data views on small screens, with particular attention to case studies and social media requirements for such terms.

Finally, a case study was presented; three types of representations of the levels of carbon dioxide have been realized, distinguished for month and gathering station: a weekly histogram, a radar chart containing the levels of all days of the month, and another radar chart showing the average levels of CO_2. Weather information was added to the two radar charts to give users additional context. The representations were then published on a social media platform, namely Instagram.

At the end of the publication process, these representations were evaluated by combining the results extracted from an engagement survey, created by combining an already existing User Engagement Scale **[99]** and two additional questions used to evaluate the appropriateness of the representations to social media. The survey was submitted to a sample of users of social networks who interacted with the representations. This analysis showed that one of the three types of representation, the radar chart showing the total CO_2 concentration over the month, is the most usable, the most aesthetically appealing, and most rewarding in terms of experience Overall, the most suitable in terms of engagement for social media. Histograms received the lower yet good scores in all the categories tested. Overall, all three chart types performed well in the context of social media, showing that the guidelines that were defined in Chapter 5 and used for their implementations are effective.

Possible future developments mainly concern the improvement of the data collection system, to add other parameters useful to understand the variance in carbon dioxide levels, such as wind sensors, fine dust detection, sound sensors, cameras, or other sensors to monitor external events that may affect CO_2 levels, such as the presence of people, road or maritime traffic.

From a representations point of view, it would be useful to realize and evaluate representations showing in a more direct way the danger that the constant increase of the concentration of carbon dioxide in the atmosphere may bring to the city of Venice and the world.

Overall, the work done represent also a first effort of rationalizing the work done in the literature about the creation of guidelines for communicating data on small screens, and it is offered to the Information Visualization for further analyses, testing, and improvements.

Appendix A

Initial Assessment Survey

This survey was created for various research purposes. First, to understand the level of information of young people aged 20-30. The survey then attempts to evaluate which social media are most used to obtain information, evaluate data regarding socio-scientific1 topics, and share such data. Finally, the survey seeks to evaluate the level of knowledge regarding the potential of the discipline known as data visualization and its position in the context of data dissemination regarding socio-scientific topics[16]

Disclaimer regarding privacy

We respect your privacy and appreciate your participation in this scientific survey. The information you provide will be used solely for research purposes and will be kept confidential. Your responses will not be disclosed to any third party, except where required by law. However, please note that the Internet is not completely secure, and there is a risk that your information could be intercepted or accessed by unauthorized third parties during transmission. By voluntarily participating in this survey, you agree that your information may be collected, stored, and used as described. If you have any concerns about your privacy or the use of your information, please contact us before completing the survey.

[16]Socio-scientific issues are controversial, socially relevant, real-world problems that are informed by science and often include an ethical component [115]. Examples include fish farming, genetic testing, global warming, and captive breeding in zoos. "Socio-scientific issues are usually value-laden, and the juxtaposition of science and ethics can be uncomfortable for scientists, teachers, and students who decline science in terms of objectivity" [114].

A.1 Section 1: Personal Info

Questions	**Answer format**
Age	Short answer
Gender	1) Woman; 2) Man; 3) Transgender; 4) Non-binary; 5) Prefer not to share;
Country of residence	Short answer
Status	1) Single; 2) In a relationship; 3) Married; 4) Prefer not to share;
Occupation *(select all that apply)*	1) Student; 2) Worker; 3) Not currently working;
If you are a worker, which job do you do?	Short answer
Grade of education	1) Middle school diploma; 2) High school diploma; 3) Bachelor's degree; 4) Master's degree; 5) Postgraduate master; 6) Ph.D. (doctorate);
Which languages do you speak? *(select all that apply)*	1) Italian; 2) English; 3) Other (*please specify*);

Table A.1: Questions and possible answers of Section 1 of the initial assessment survey.

A.2 Section 2: Information and media

Questions	**Answer format**
How often do you read news or socio-scientific articles?	Likert Scale 1 to 5; 1 = Never; 5 = Daily;
Which of the following media do you youse to read news and scientific articles? *(select all that apply)*	1) Print media; 2) Online newspapers; 3) Social media; 4) Scientific journals/articles; 5) News app; 6) Blogs, online forums, newsletters; 7) Others *(please specify)*;
On which device do you typically read news and scientific articles?	1) Smartphone; 2) Tablet; 3) Computer; 4) Print media; 5) Others *(please specify)*;
How do you prefer to consume news, data and scientific articles? *(select all that apply)*	1) Reading text; 2) Watching videos; 3) Listening to podcasts; 4) Viewing images; 5) Others *(please specify)*;
How important is it for you to have access to news and scientific articles?	Likert Scale 1 to 5; 1 = Not at all important; 5 = Extremely important;

Table A.2: Questions and possible answers of Section 2 Part.1 of the initial assessment survey.

Questions	**Answer format**
How do you evaluate the reliability of the news, data and scientific articles you read? *(select all that apply)*	1) I trust the sources I read without question. ; 2) I evaluate the sources based on their reputation; 3) I evaluate the sources based on their political bias; 4) I evaluate the sources based on their expertise in the subject matter; 5) I look for corroboration of the information from other sources; 6) Others *(please specify)*;
How often do you share news, data and scientific articles with others?	Likert Scale 1 to 5; 1 = Never; 5 = Daily;
What kind of news, data and scientific articles do you prefer to read? *(select all that apply)*	1) Local/national news; 2) National news; 3) Breaking news; 4) Opinion pieces; 5) Scientific research findings; 6) In-depth features; 7) Others *(please specify)*;

Table A.3: Questions and possible answers of Section 2 Part.2 of the initial assessment survey.

A.3 Section 3: Social media platforms

Questions	**Answer format**
Are you registered to any social media platform?	1) Yes; 2) No;
If so, which of these platforms are you registered for? *(select all that apply)*	1) Twitter; 2) Facebook; 3) Instagram; 4) Reddit; 5) Linkedin; 6) Pinterest; 7) Youtube; 8) Others *(please specify)*;
How frequently do you access social media?	Likert Scale 1 to 5; 1 = Monthly; 5 = Hourly;
What do you use social media for? *(select all that apply)*	1) Networking; 2) Business; 3) Learning; 4) Fun; 5) Keep yourself informed about worldwide news; 6) Others *(please specify)*;
How concerned are you about the impact of false or misleading data shared on social media?	Likert Scale 1 to 5; 1 = Not at all concerned; 5 = Very concerned;

Table A.4: Questions and possible answers of Section 3 Part.1 of the initial assessment survey.

Questions	**Answer format**
How confident are you in the accuracy of the data shared on social media?	1) Not at all confident; 2) Somewhat unconfident; 3) Neutral; 4) Somewhat confident; 5) Very confident; 6) Depends on the source;
How often do you see fact-checking labels or warnings on social media posts??	Likert Scale 1 to 5; 1 = Never; 5 = Always;
How likely are you to believe a fact-checking label or warning on social media?	Likert Scale 1 to 5; 1 = Extremely unlikely; 5 = Extremely likely;
When you see data on social media, how often do you verify its accuracy before sharing it?	Likert Scale 1 to 5; 1 = Never; 5 = Always;
Have you ever shared data on social media that you later found out to be false or misleading?	1) Yes, often; 2) Yes, occasionally; 3) Yes, once; 4) No, never;
Have you ever reported false or misleading data on social media?	1) Yes and I received feedback; 2) Yes but I did not receive feedback; 3) No; 4) I am not sure how to report false information;
How do you think social media can improve the news, data and information shared by users on their platforms?	Long answer

Table A.5: Questions and possible answers of Section 3 Part.2 of the initial assessment survey.

A.4 Section 4: Representing data with images and graphs

Questions	**Answer format**
How often do you encounter data presented in the form of images and graphs in your daily life?	Likert Scale 1 to 5; 1 = Never; 5 = Every day;
In which of the following settings have you seen data presented in the form of images and graphs? *(select all that apply)*	1) News and reports; 2) Social media posts; 3) Scientific papers; 4) Business reports; 5) Others *(please specify)*;
Do you find data presented in the form of images and graphs more appealing and easier to understand than plain text data?	Likert Scale 1 to 5; 1 = Strongly disagree; 5 = Strongly agree;
When presented with data in the form of graphs, how often do you read the accompanying text?	Likert Scale 1 to 5; 1 = Never; 5 = Always;
In your opinion, what are the advantages of presenting data in the form of graphs?	Long answer
Have you ever encountered data presented in the form of images and graphs that you found confusing or misleading?	1) Yes; 2) No;
If yes, what made the data confusing or misleading?	Long answer
Do you think there should be more efforts to teach people how to create and interpret data presented in the form images and graphs?	Likert Scale 1 to 5; 1 = Strongly disagree; 5 = Strongly agree;
In your opinion, what steps can be taken to improve the quality of data presented in the form of images and graphs?	Long answer
Do you think the small size of the display of a smartphone is a limit for visualizing data and visual representation of data?	1) Yes; 2) No;

Table A.6: Questions and possible answers of Section 4 Part.1 of the initial assessment survey.

Questions	**Answer format**
If yes, why do you think it is a limit?	Long answer
Fig. 4.9a shows from which professional soccer league the players on each country's 2014 World Cup soccer team come. Do you think this representation shows data in a clear way?	1) Yes; 2) No;
Referring to Fig. 4.9a, if no, what do you think is the problem?	Long answer
Fig. 4.9b comes from the Reporter App developed for iPhone. It shows the sleep quality of the user for the last week. What do you think are the strengths and the flaws of this visualization?	Long answer
Fig. 4.9c shows the score of each world country in terms of regulations for limiting temporary jobs. The higher the score, the higher the limitations. What do you think it is missing from this visualization?	Long answer
Fig. 4.9d comes from the Reporter App developed for iPhone. It shows the type and the quantity of food the user consumed in the last week. What do you think are the relevant elements in this representation?	Long answer
Do you think there are too many details in the map depicted in Fig. 4.10?	1) Yes; 2) No;

Table A.7: Questions and possible answers of Section 4 Part.2 of the initial assessment survey.

Appendix B

Code Scripts

Arduino Code

```
#include <SPI.h>
#include <UIPEthernet.h>
#include <SoftwareSerial.h>
#include <DS3231.h>
#include <Wire.h>
#include <Adafruit_AM2315.h>
#include <avr/wdt.h>
#include <avr/sleep.h>

byte mac[] = { 0x00, 0x00, 0x00, 0x00, 0x00, 0x00 };
char server[] = "third.party.server";

IPAddress myIP(., ., ., .);
IPAddress myMask = {., ., ., .};
IPAddress myGatewayIP = {., ., ., .};
IPAddress dnsIP(., ., ., .);

EthernetClient client;

Adafruit_AM2315 am2315;
float tempData;
float humData;
unsigned long cooData;

SoftwareSerial K30(8, 7);
byte readCO2[] = {0xFE, 0X44, 0X00, 0X08, 0X02, 0X9F, 0X25};
byte CO2response[] = {0, 0, 0, 0, 0, 0, 0};

int CO2valMultiplier = 1;

String site="Sitename,latitude,longitude";
```

```

boolean signalflag = false;
boolean callDone = false;


void(* resetFunc)(void) = 0;


void setup() {

  Serial.begin(9600);
  while(!Serial);
  Wire.begin();
  K30.begin(9600);

  Serial.println();

  int countLink = 0;

  if (!am2315.begin()) {
    Serial.println(F("Temp/hum Sensor not found, check wiring & pullups!"));
  }

    Ethernet.begin(mac,myIP, dnsIP, myGatewayIP, myMask);

    if (Ethernet.hardwareStatus() == EthernetNoHardware) {
      Serial.println(F("Ethernet shield was not found. "));
      while (true) {
        delay(1);
      }
    }
    if (Ethernet.linkStatus() == LinkOFF) {
      Serial.println(F("Ethernet cable is not connected."));
    }

    while(Ethernet.linkStatus() == LinkOFF){
      if(countLink <=10){
        Serial.println(F("Waiting for link."));
        countLink = countLink + 1;
        delay(1000);
      }
      else {
        resetFunc();
      }
    }

    if (Ethernet.linkStatus() == LinkON) {
```

```
      Serial.println(F("Ethernet cable is  now connected."));
    }
    Ethernet.begin(mac,myIP, dnsIP, myGatewayIP, myMask);
    Serial.print(F(" Assigned IP "));
    Serial.println(Ethernet.localIP());

  delay(1000);
}


void loop() {

    if ( millis()  >= 86400000){
      Serial.end();
      resetFunc();
    }
    sendRequestToK30(readCO2);
    cooData = getValueFromK30(CO2response);

    int countTemp = 0;
    int countConnect = 0;

    while (! am2315.readTemperatureAndHumidity(&tempData, &humData) &&
    countTemp < 5) {
      delay(1000);
      tempData = -1000;
      humData = -1000;
      countTemp = countTemp + 1;
    }

    Serial.println(F("Connecting1..."));
    while(callDone==false){
      signalflag=false;
      while(signalflag==false){
        if(client.connect(server,80)){
          sendData();
          signalflag = true;
          Serial.println(F("Connected"));
        }else{
            Serial.println(F("Fail to Connect..."));
        }
      }
      while(signalflag){
          if(client.available()){
              char c = client.read();
              Serial.print(c);
              callDone = true;
```

```

          }
          if(!client.connected()){
              Serial.println();
              Serial.println(F("Not Linking..."));
              Serial.println(F("Temperature Sent:"));
              Serial.println(tempData);
              Serial.println(F("Humidity Sent:"));
              Serial.println(humData);
              Serial.println(F("CO2 Sent:"));
              Serial.println(cooData);
              client.stop();
              signalflag = false;

              countConnect = countConnect + 1;

              if ( millis()  >= 86400000 || countConnect > 7){
                Serial.println(F("+++++++++++++++ RESET +++++++++++++++"));
                Serial.end();
                resetFunc();
              }

          }

        }
    }

    callDone=false;

    delay(900000);
}


void sendData(){

  Serial.println(F("Connecting2..."));
  client.print(F("GET //co2data/test.php?tempData="));
  client.print(tempData);
  client.print(F("&humData="));
  client.print(humData);
  client.print(F("&cooData="));
  client.print(cooData);
  client.print(F("&site="));
  client.print(site);
  client.println(F(" HTTP/1.1"));
  client.println(F("Host: mizar.unive.it"));
  client.println(F("Connection: close"));
```

```
  client.println();

}


void sendRequestToK30(byte packet[])
{
  int timeout1 = 0;
  while (!K30.available()&&timeout1<10)
  {
    K30.write(readCO2, 7);
    timeout1++;
    delay(50);

  }

  int timeout = 0;
  while (K30.available() < 7 )
  {
    timeout++;
    if (timeout > 10)
    {
      while (K30.available())
        K30.read();

        break;
    }
    delay(50);
  }

  for (int i = 0; i < 7; i++)
  {
    CO2response[i] = K30.read();
  }
}


unsigned long getValueFromK30(byte packet[])
{
  int high = packet[3];
  int low = packet[4];

  unsigned long val = high * 256 + low;
  return val * CO2valMultiplier;
}
```

Listing B.1: Arduino code of a single unit

php code

```php
<?php

$IPCaller = $_SERVER['REMOTE_ADDR'];
echo $IPCaller."\n";

if ($IPCaller!="0.0.0.0"&&$IPCaller!="0.0.0.0"&&$IPCaller!="0.0.0.0"){
    echo 'Non sei autorizzato';
}
else{

$error = '';
$tempData = $_GET['tempData'];
$tempData = floatval(filter_var($tempData, FILTER_SANITIZE_STRING));

if ($tempData==-1000 || filter_var($tempData, FILTER_VALIDATE_FLOAT) == false) {
    $tempData = 'ErrorTempData';
}

$humData = $_GET['humData'];
$humData = floatval(filter_var($humData, FILTER_SANITIZE_STRING));

if ($humData==-1000 || filter_var($humData, FILTER_VALIDATE_FLOAT) == false) {
    $humData = 'ErrorHumData';
}

$cooData = $_GET['cooData'];
$cooData = intval(filter_var($cooData, FILTER_SANITIZE_STRING));

if ($cooData==0 || $cooData==4294967295 || filter_var($cooData, FILTER_VALIDATE_INT) == false) {
    $cooData = 'ErrorcooData';
}

$site = $_GET['site'];
$site = filter_var($site, FILTER_SANITIZE_STRING);

if ($site=='') {
    $site = 'Errorsite';
}

require_once './google-api-php-client--PHP5.6/vendor/autoload.php';

$client = new Google\Client();

```

```
putenv('GOOGLE_APPLICATION_CREDENTIALS=./login.json');
$client->setApprovalPrompt('force');
$client->useApplicationDefaultCredentials();

// returns a Guzzle HTTP Client
$httpClient = $client->authorize();

$response = $httpClient->get('https://script.google.com/macros/s/'
    SPREADSHEET_ID'/exec?tempData='.$tempData.'&humData='.$humData.'&cooData='
    .$cooData.'&site='.$site.'');
$res=$response->getBody()->getContents();

echo $res;
}
?>

<!DOCTYPE html>
<html>

<head>
    <style>
    </style>
</head>

<body>
        <h1>Test</h1>

        <h2>
            <?php
            echo 'TempData='.$tempData.' HumData='.$humData.' CooData='.
    $cooData.' Site='.$site.''
            ?>
        </h2>
</body>

</html>
```

Listing B.2: php code for converting HTTP request into HTTPs request

Google AppScript Code

```
function doGet(e) {
  Logger.log( JSON.stringify(e) );  // view parameters

  var result = 'Ok'; // assume success

  if (e.parameter == undefined) {
    result = 'No Parameters';
  }
  else {
    var id = 'Spreadsheet_ID';

    var rowData = [];

    var date = new Date();
    rowData[0] = Utilities.formatDate(date, 'Europe/Rome', 'yyyy/MM/dd, HH:mm:ss');
    for (var param in e.parameter) {
      Logger.log('In for loop, param='+param);
      var value = stripQuotes(e.parameter[param]);

      switch (param) {
        case 'tempData':      // Parameter
          rowData[1] = value; // Value in column B
          break;
        case 'humData':
          rowData[2] = value;
          break;
        case 'cooData':
          rowData[3] = value;
          break;
        case 'site':
          var array = [{}];
                array = value.split(",");
                rowData[4] = array[0];
                rowData[5] = array[1];
                rowData[6] = array[2];
          break;
        default:
          result = "unsupported parameter";
      }
    }
    Logger.log(JSON.stringify(rowData));

    var sheet = SpreadsheetApp.getActiveSpreadsheet().getSheetByName(rowData[4]);
```

```
    if (sheet != null) {
      Logger.log(sheet.getIndex());
    }
    var newRow = sheet.getLastRow() + 1;

    // Write new row below
    var newRange = sheet.getRange(newRow, 1, 1, rowData.length);
    newRange.setValues([rowData]);
  }

  // Return result of operation
  return ContentService.createTextOutput(result);
}

/**
* Remove leading and trailing single or double quotes
*/
function stripQuotes( value ) {
  return value.replace(/^["']|['"]$/g, "");
}
```

Listing B.3: Google App Script code for parsing the data and adding it to the correct spreadsheet

Abbreviations

A Ampere. 45

AE Aesthetic Appeal. 98, 101, 102, 131

ALT text Alternative text. 34, 82

API Application programming interface. 48

AQ air quality. 12, 16

AR Augmented Reality. 27

ARPAV Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto. 17

COP21 Paris for the Conference of the Parties. 37

CVD Color Vision Deficency. 81, 82

DAIS Dipartimento di Scienze Ambientali, Informatica e Statistica. 39

DNS Domain Name System. 47

EEA European Environment Agency. 30

FA Focused Attention. 98, 101, 102, 131

FTP File Transfer Protocol. 44

GHG Greenhouse gasses. 37

GIFs Graphics Interchange Format. 33, 34, 82

HCI Human Computer Interaction. 17

HTTP HyperText Transfer Protocol. 44, 46–48

HTTPs HyperText Transfer Protocol over Secure Socket Layer. 44, 46, 48

IoT Internet of Things. 15

IP Internet Protocol. 47

IPs Internet Protocol addresses. 48

ISAC-CNR Istituto di Scienze dell'Atmosfera e del Clima del Consiglio Nazionale delle Ricerche. 17

MAC address Media Access Control address. 47

PDAs Personal Digital Assistants. 22–24, 26

PNC Particle Number Concentration. 18

PoE Power over Ethernet. 44, 45, 48, 107

POIs Point Of Interests. 25, 95

ppm parts per millions. 2, 15, 48, 85–87, 90, 93

PU Perceived Usability. 98, 101, 102, 131

RW Rewarding Factor. 98, 101, 102, 131

SBC Single Board Computer. 14

sFTP Secure File Transfer Protocol. 44

SSH Secure Shell. 44

TCP/IP Transmission Control Protocol/Internet Protocol. 44

UES User Engagement Scale. 98, 108, 133

UES-SF User Engagement Scale - Short Form. 98–100, 131, 133

US EPA United States Protection Agency. 14

UV Ultraviolet. 3

V Volt. 45

VR Virtual Reality. 27

WHO World Health Organization. 15

List of Figures

List of Tables

Bibliography

[1] AGCOM. *AGCOM: report su giovani e informazione*. 2020. Available at: https://confindustriaradiotv.it/agcom-report-su-giovani-e-informazione/ [Last access: May 13, 2023].

[2] Gregor Aisch. *Data Visualization and the news*. Information+ 2016 Conference. 2016. Available at: https://vimeo.com/182590214 [Last access: May 16, 2023].

[3] Leonardo Alfonso, Mohammad Gharesifard, and Uta Wehn. *Analysing the value of environmental citizen-generated data: Complementarity and cost per observation*. In: *Journal of Environmental Management* 303 (2022). Elsevier Publication, p. 114157. ISSN: 0301-4797. DOI: https://doi.org/10.1016/j.jenvman.2021.114157. Available at: https://www.sciencedirect.com/science/article/pii/S0301479721022192 [Last access: May 10, 2023].

[4] Akram Syed Ali, Zachary Zanzinger, Deion Debose, and Brent Stephens. *Open Source Building Science Sensors (OSBSS): A low-cost Arduino-based platform for long-term indoor environmental data collection*. In: *Building and Environment* 100 (2016). Elsevier Publication, pp. 114–126. ISSN: 0360-1323. DOI: https://doi.org/10.1016/j.buildenv.2016.02.010. Available at: https://www.sciencedirect.com/science/article/pii/S0360132316300476 [Last access: May 15, 2023].

[5] Abdulrahman Alkandari and Samer Moein. *Implementation of Monitoring System for Air Quality using Raspberry PI: Experimental Study*. In: *Indonesian Journal of Electrical Engineering and Computer Science* 10 (April 2018). Institute of Advanced Engineering and Science (IAES), pp. 43–49. DOI: 10.11591/ijeecs.v10.i1.pp43-49. [Last access: May 12, 2023].

[6] Helen Atkinson. *Charting new territory: How The Economist designs charts for Instagram*. 2020. Available at: https://medium.economist.com/charting-new-territory-7f5afb293270 [Last access: May 16, 2023].

[7] Pekka Aula. *Social media, reputation risk and ambient publicity management*. In: *Strategy and Leadership* 38 (November 2010). Emerald Publishing. DOI: 10.1108/10878571011088069. [Last access: May 12, 2023].

[8] Giovanni Baccolo. *Curva di Keeling: l'aumento di anidride carbonica (CO2) in atmosfera.* 2022. Available at: https://storieminerali.it/curva-keeling/ [Last access: May 12, 2023].

[9] João Pedro Baptista and Anabela Gradim. *Who Believes in Fake News? Identification of Political (A)Symmetries.* In: *Social Sciences* 11.10 (2022). MDPI. ISSN: 2076-0760. DOI: 10.3390/socsci11100460. Available at: https://www.mdpi.com/2076-0760/11/10/460 [Last access: May 24, 2023].

[10] Marco Bardus, Rola El Rassi, Mohamad Chahrour, Elie Akl, Abdul Sattar Raslan, Lokman Meho, and Elie Akl. *The Use of Social Media to Increase the Impact of Health Research: Systematic Review.* In: *Journal of medical Internet research* 22 (July 2020), e15607. DOI: 10.2196/15607. [Last access: June 16, 2023].

[11] Brent Barnhart. *Dati demografici dei social media per elaborare la strategia del tuo brand nel 2021.* 2021. Available at: https://sproutsocial.com/it/insights/new-social-media-demographics-it_it/ [Last access: May 13, 2023].

[12] Patrick Baudisch, Xing Xie, Chong Wang, and Wei-Ying Ma. *Collapse-to-Zoom: Viewing Web Pages on Small Screen Devices by Interactively Removing Irrelevant Content.* In: *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology.* UIST '04. Santa Fe, NM, USA: Association for Computing Machinery, 2004, pp. 91–94. ISBN: 1581139578. DOI: 10.1145/1029632.1029647. Available at: https://doi.org/10.1145/1029632.1029647 [Last access: June 16, 2023].

[13] Benjamin B. Bederson, Aaron Clamage, Mary P. Czerwinski, and George G. Robertson. *DateLens: A Fisheye Calendar Interface for PDAs.* In: *ACM Trans. Comput.-Hum. Interact.* 11.1 (2004), pp. 90–119. ISSN: 1073-0516. DOI: 10.1145/972648.972652. Available at: https://doi.org/10.1145/972648.972652 [Last access: June 16, 2023].

[14] Tanja Blaschec, Frank Bentley, Eun Kyoung Choe, Tom Horak, and Petra Isenberg. *Characterizing Glanceable Visualizations: From Perception to Behavior Change.* In: *Mobile Data Visualization (1st ed.)* November 2021, pp. 151–175. ISBN: 9781003090823. DOI: 10.1201/9781003090823-5.

[15] Mausi Bokoh, Bello Miss, and Idowu Miss. *Use of Social Media Platforms for Dissemination of Information and Creating Awareness about Library Resources and Services among Students in Lagos State University, Nigeria.* January 2022.

[16] Boost Labs. *Static vs Interactive Visualizations – How to Choose.* 2017. Available at: https://boostlabs.com/static-vs-interactive-visualizations-choose/ [Last access: June 4, 2023].

[17] C. Borrego, M. Coutinho, A.M. Costa, J. Ginja, C. Ribeiro, A. Monteiro, I. Ribeiro, J. Valente, J.H. Amorim, H. Martins, D. Lopes, and A.I. Miranda. *Challenges for a New Air Quality Directive: The role of monitoring and modelling techniques*. In: *Urban Climate* 14 (2015). Elsevier Publication, pp. 328–341. ISSN: 2212-0955. DOI: https://doi.org/10.1016/j.uclim.2014.06.007. Available at: https://www.sciencedirect.com/science/article/pii/S2212095514000479 [Last access: May 10, 2023].

[18] Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. *Visualizing Ranges over Time on Mobile Phones: A Task-Based Crowdsourced Evaluation*. In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019). IEEE, pp. 619–629. DOI: 10.1109/TVCG.2018.2865234. [Last access: May 16, 2023].

[19] Matthew Brehmer, Bongshin Lee, John Stasko, and Christian Tominski. *Interacting with Visualization on Mobile Devices*. In: *Mobile Data Visualization (1st ed.)* November 2021, pp. 67–110. ISBN: 9781003090823. DOI: 10.1201/9781003090823-3.

[20] Nadieh Bremer. *Techniques for Data Visualization on both Mobile and Desktop*. 2019. Available at: https://www.visualcinnamon.com/2019/04/mobile-vs-desktop-dataviz/ [Last access: May 16, 2023].

[21] Lorenzo Brutti. *Life Redune Project: data acquisition and dissemination. Design and implementation of an off-grid sensors network and design of a public playful tangible user interface for data presentation.* MA thesis. Università Ca' Foscari di Venezia, 2017.

[22] Alberto Cairo. *How Charts Lie: Getting Smarter about Visual Information*. W. W. Norton, 2019. ISBN: 9781324001577. Available at: https://books.google.it/books?id=qP2KDwAAQBAJ [Last access: June 14, 2023].

[23] Guillem Camprodon, Óscar González, Víctor Barberán, Máximo Pérez, Viktor Smári, Miguel Ángel de Heras, and Alejandro Bizzotto. *Smart Citizen Kit and Station: An open environmental monitoring system for citizen participation and scientific experimentation*. In: *HardwareX* 6 (2019). Springer. ISSN: 2468-0672. DOI: https://doi.org/10.1016/j.ohx.2019.e00070. Available at: https://www.sciencedirect.com/science/article/pii/S2468067219300203 [Last access: April 6, 2023].

[24] Kay Chang. *Go Small or Go Home: What to consider when transforming visualization from large to small screens*. 2021. Available at: https://medium.com/visumd/responsive-visualization-transitioning-from-large-screen-to-small-screen-4854b6cd22b [Last access: May 16, 2023].

[25] Yang Chen. *Visualizing Large Time-series Data on Very Small Screens*. In: *Proceedings of EuroVis 2017*. Ed. by Barbora Kozlikova, Tobias Schreck, and Thomas Wischgoll. The Eurographics Association, 2017. ISBN: 978-3-03868-043-7. DOI: 10.2312/eurovisshort.20171130. [Last access: May 16, 2023].

[26] CHI KT Platform. *How to Create Effective Research Surveys*. 2019. Available at: https://medium.com/knowledgenudge/how-to-create-effective-research-surveys-c282f134beb3 [Last access: May 30, 2023].

[27] Luca Chittaro. *Visualizing information on mobile devices*. In: *Computer* 39.3 (2006). IEEE Xplore, pp. 40–45. DOI: 10.1109/MC.2006.109. [Last access: May 16, 2023].

[28] Eun Kyoung Choe, Raimund Dachselt, Petra Isenberg, and Bongshin Lee. *Mobile Data Visualization (Dagstuhl Seminar 19292)*. In: *Dagstuhl Reports*. Ed. by Eun Kyoung Choe, Raimund Dachselt, Petra Isenberg, and Bongshin Lee. Vol. 9. 7. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, pp. 78–93. DOI: 10.4230/DagRep.9.7.78. Available at: https://drops.dagstuhl.de/opus/volltexte/2019/11636 [Last access: May 16, 2023].

[29] CO2-Earth. *Co2-Earth - Numbers for living Earth*. 2023. Available at: https://www.co2.earth/daily-co2 [Last access: May 12, 2023].

[30] ConductScience. *The Impact of Social Media on Knowledge Dissemination*. 2019. Available at: https://conductscience.com/the-impact-of-social-media-on-knowledge-dissemination/#:~:text=Twitter%5C%20is%5C%20arguably%5C%20the%5C%20best,interaction%5C%20between%5C%20posters%5C%20and%5C%20followers.

[31] Michael Correll. *Ethical Dimensions of Visualization Research*. In: *Proceedings of CHI '19*. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–13. ISBN: 9781450359702. DOI: 10.1145/3290605.3300418. Available at: https://doi.org/10.1145/3290605.3300418 [Last access: May 3, 2023].

[32] Carlos J. Costa, José Silva, and Manuela Aparìcio. *Evaluating Web Usability Using Small Display Devices*. In: *Proceedings of SIGDOC '07*. El Paso, Texas, USA: Association for Computing Machinery, 2007, pp. 263–268. ISBN: 9781595935885. DOI: 10.1145/1297144.1297202. Available at: https://doi.org/10.1145/1297144.1297202 [Last access: May 16, 2023].

[33] Andy Cotgrave. *6 tips for creating charts for social media*. 2022. Available at: https://www.tableau.com/blog/6-tips-to-design-charts-for-social-media [Last access: May 16, 2023].

[34] Luc Dekoninck, Dick Botteldooren, and Luc Int Panis. *Sound sensor network based assessment of traffic, noise, and air pollution.* In: *Proceedings of the European Congress and Exposition on Noise Control Engineering (Euronoise)*. Maastricht, Netherlands, June 2015, pp. 2321–2326.

[35] Apple Developer. *Charting data.* 2022. Available at: `https://developer.apple.com/design/human-interface-guidelines/patterns/charting-data/` [Last access: May 16, 2023].

[36] Daniel Epstein, Tanja Blascheck, Sheelagh Carpendale, Raimund Dachselt, and Jo Vermeulen. *Challenges in Everyday Use of Mobile Visualizations.* In: *Mobile Data Visualization (1st ed.)* November 2021, pp. 209–240. ISBN: 9781003090823. DOI: `10.1201/9781003090823-7`.

[37] European Environment Agency. *Dos and don'ts of data visualisation — European Environment Agency.* 2021. Available at: `https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts/#last-but-not-least` [Last access: May 16, 2023].

[38] European Environment Agency. *Trends in atmospheric concentrations of CO2 (ppm), CH4 (ppb) and N2O (ppb), between 1800 and 2017.* 2019. Available at: `https://www.eea.europa.eu/data-and-maps/daviz/atmospheric-concentration-of-carbon-dioxide-5` [Last access: May 22, 2023].

[39] Nicholas Felton. *Reporter App for iPhone.* Available at: `http://reporter-app.com/` [Last access: May 29, 2023].

[40] Sheikh Ferdoush and Xinrong Li. *Wireless Sensor Network System Design Using Raspberry Pi and Arduino for Environmental Monitoring Applications.* In: *Procedia Computer Science* 34 (December 2014). Elsevier Publication, pp. 103–110. DOI: `10.1016/j.procs.2014.07.059`. [Last access: May 12, 2023].

[41] FormPlus Blog. *33 Social Media Survey Questionnaires.* Available at: `https://www.formpl.us/blog/social-media-survey-questionnaire` [Last access: May 30, 2023].

[42] Vitaly Friedman. *Responsive Web Design: What It Is And How To Use It.* 2018. Available at: `https://www.smashingmagazine.com/2011/01/guidelines-for-responsive-web-design/` [Last access: May 16, 2023].

[43] Michael Friendly. *Visualizing Categorical Data.* In: *Cognition and Survey Research.* Ed. by M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau. New York: John Wiley and Sons, 1999. Chap. 20, pp. 319–438. Available at: `http://datavis.ca/papers/casm/casm.pdf` [Last access: May 16, 2023].

[44] Jon Froehlich, Leah Findlater, and James Landay. *The Design of Eco-Feedback Technology*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10*. Atlanta, Georgia, USA: Association for Computing Machinery, 2010, pp. 1999–2008. ISBN: 9781605589299. DOI: 10.1145/1753326.1753629. Available at: https://doi.org/10.1145/1753326.1753629 [Last access: May 11, 2023].

[45] Flaminia Gallo. *I data visualization e i social network*. In: (2011). Available at: https://www.academia.edu/40223967/I_data_visualization_e_i_social_network.

[46] Mubarak Ganiyu. *Everything is Data: Why we need data to improve the world?* Towards Data Science. 2019. Available at: https://towardsdatascience.com/everything-is-data-f08e291c4b11 [Last access: May 23, 2023].

[47] D. Gddemeyer and D. Baur. *Subspotting*. 2016. Available at: http://subspotting.nyc/ [Last access: May 16, 2023].

[48] Michael Grabell and Lena Groeger. *Temp Worker Regulations Around the World*. 2014. Available at: https://projects.propublica.org/graphics/temps-around-the-world [Last access: May 29, 2023].

[49] Andrew Gregory. *Monitor air quality with a Raspberry Pi*. 2019. Available at: https://www.raspberrypi.com/news/monitor-air-quality-with-a-raspberry-pi/ [Last access: May 10, 2023].

[50] Andrew Guess, Jonathan Nagler, and Joshua Tucker. *Less than you think: Prevalence and predictors of fake news dissemination on Facebook*. In: *Science Advances* 5 (2019). DOI: 10.1126/sciadv.aau4586. Available at: https://www.science.org/doi/abs/10.1126/sciadv.aau4586 [Last access: May 24, 2023].

[51] Suraya Hamid, Mohamad Ijab, Hidayah Sulaiman, Rina Md Anwar, and Azah Norman. *Social media for environmental sustainability awareness in higher education*. In: *International Journal of Sustainability in Higher Education* 18 (May 2017). Emerald Publishing, pp. 474–491. DOI: 10.1108/IJSHE-01-2015-0010. [Last access: May 12, 2023].

[52] Ali Hammad. *Visualization of large data set on Small screens*. MA thesis. University of Stavanger, Norway, 2020.

[53] Tomi Heimonen. *Information Visualization on Small Display Devices*. MA thesis. Department of Computer Sciences, University of Tampere, 2002. Available at: https://trepo.tuni.fi/handle/10024/90184 [Last access: June 13, 2023].

[54] Tiffany Grace Holmes. *Eco-Visualization: Combining Art and Technology to Reduce Energy Consumption*. In: *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*. C&C '07. Washington, DC, USA: Association for Computing Machinery, 2007, pp. 153–162. ISBN: 9781595937124. DOI: 10.1145/1254960.1254982.

Available at: https://doi.org/10.1145/1254960.1254982 [Last access: May 10, 2023].

[55] Jill Hopke and Luis Hestres. *Visualizing the Paris Climate Talks on Twitter: Media and Climate Stakeholder Visual Social Media During COP21*. In: *Social Media + Society* 4 (July 2018). SAGE Journals. DOI: 10.1177/2056305118782687. [Last access: May 12, 2023].

[56] Chih-Neng Hsu and Yun-Lin Tsai. *Experimental Measurement and Computational Simulation Analysis of Indoor Air Quality in Office—Integration of Voltage Adsorption Dust Collection Device and Energy Recovery Ventilator*. In: *Sensor Materials* 32.12 (2020). MDPI, pp. 4299–4321. DOI: 10.18494/SAM.2020.3143. Available at: https://sensors.myu-group.co.jp/article.php?ss=3143 [Last access: May 16, 2023].

[57] David Hunter. *Data Walking*. In: *Proceedings of the 16th Participatory Design Conference 2020 - Participation(s) Otherwise - Volume 2*. PDC '20. Manizales, Colombia: Association for Computing Machinery, 2020, pp. 188–191. ISBN: 9781450376068. DOI: 10.1145/3384772.3385167. Available at: https://doi.org/10.1145/3384772.3385167 [Last access: April 12, 2023].

[58] Rositsa Ilieva and Timon McPhearson. *Social-media data for urban sustainability*. In: *Nature Sustainability* 1 (October 2018). Nature Publishing Group. DOI: 10.1038/s41893-018-0153-6. [Last access: May 12, 2023].

[59] ilMeteo.it. *Che meteo faceva a Venezia a Marzo 2023?* Available at: https://www.ilmeteo.it/portale/archivio-meteo/Venezia/2023/Marzo [Last access: June 7, 2023].

[60] Sphera Solutions Inc. *What is Environmental Sustainability*. 2020. Available at: https://sphera.com/glossary/what-is-environmental-sustainability/.

[61] Ipsos. *Media e Fake news: Opinioni e attitudini degli italiani nei confronti dell'informazione*. 2022. Available at: https://www.idmo.it/wp-content/uploads/2022/03/PRESENTAZIONE-IPSOS-24-MARZO_PUBBLICO.pdf [Last access: May 16, 2023].

[62] ISAC. *Mount Cimone GAW Global Station - Italian Climate Observatory "Ottavio Vittori" at Monte Cimone*. Available at: https://cimone.isac.cnr.it/realtime [Last access: May 13, 2023].

[63] Akiko Iwasaki. *Using social media to promote science*. In: *Nature Immunology* 23 (June 2022). Nature Publishing Group, pp. 1–1. DOI: 10.1038/s41590-022-01243-w. [Last access: May 12, 2023].

[64] Sean Michael Kerner and John Burke. *File-Transfer-Protocol-FTP*. 2021. Available at: https://www.techtarget.com/searchnetworking/definition/File-Transfer-Protocol-FTP [Last access: May 15, 2023].

[65] Andy Kirk. *Data Visualisation: A handbook for Data Driven Design*. 2nd Edition. SAGE Publications Ltd, 2019. ISBN: 978-1526468925.

[66] Ulrike Kister, Konstantin Klamka, Christian Tominski, and Raimund Dachselt. *GraSp: Combining Spatially-aware Mobile Devices and a Display Wall for Graph Visualization and Interaction*. In: *Computer Graphics Forum* 36.3 (2017). Eurographics, pp. 503–514. DOI: https://doi.org/10.1111/cgf.13206. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13206 [Last access: May 16, 2023].

[67] David Landsbergen. *Government as Part of the Revolution: Using Social Media to Achieve Public Goals*. In: *Electronic Journal of e-Government* 8 (January 2011). Academic Publishing.

[68] Ricardo Langner, Lonni Besançon, Christopher Collins, Tim Dwyer, Petra Isenberg, Tobias Isenberg, Bongshin Lee, Charles Perin, and Christian Tominski. *An Introduction to Mobile Data Visualization*. In: *Mobile Data Visualization (1st ed.)* November 2021, pp. 1–32. ISBN: 9781003090823. DOI: 10.1201/9781003090823-1.

[69] Sara Lebow. *5 charts that show how to (and why you should) reach millennials online*. eMarketer. 2023. Available at: https://www.insiderintelligence.com/content/5-charts-that-show-how-and-why-you-should-reach-millennials-online [Last access: May 24, 2023].

[70] Bongshin Lee, Matthew Brehmer, Petra Isenberg, Eun Kyoung Choe, Ricardo Langner, and Raimund Dachselt. *Data Visualization on Mobile Devices*. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–8. ISBN: 9781450356213. DOI: 10.11453170427.3170631. Available at: httpsdoi.org10.11453170427.3170631 [Last access: May 16, 2023].

[71] Bongshin Lee, Raimund Dachselt, Petra Isenberg, and Eun Kyoung Choe. *Mobile Data Visualization (1st ed.)* Chapman and Hall/CRC., 2021. DOI: https://doi.org/10.1201/9781003090823.

[72] Bongshin Lee, Raimund Dachselt, Petra Isenberg, and Eun Kyoung Choe. *Responsive Visualization Design for Mobile Devices*. In: *Mobile Data Visualization (1st ed.)* November 2021, pp. 33–65. ISBN: 9781003090823. DOI: 10.1201/9781003090823-2.

[73] Roni Levit. *Traveling Datavis Game*. 2018. Available at: https://www.ronilevit.com/traveling-dataviz-game [Last access: May 16, 2023].

[74] Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. *A Survey on Information Diffusion in Online Social Networks: Models and Methods.* In: *Information* 8.4 (2017). MDPI. ISSN: 2078-2489. DOI: `10.3390/info8040118`. Available at: `https://www.mdpi.com/2078-2489/8/4/118` [Last access: May 12, 2023].

[75] Simone Lolli, Ying-Chieh Chen, Sheng-Hsiang Wang, and Gemine Vivone. *Impact of meteorological conditions and air pollution on Covid-19 pandemic transmission in Lombardy region, Northern Italy.* In: *Scientific Reports* 10.1 (July 2020). DOI: `10.21203/rs.3.pex-1045/v1`. Available at: `https://protocolexchange.researchsquare.com/article/pex-1045/v1` [Last access: May 12, 2023].

[76] Bianca Lopez, Nicholas Magliocca, and Andrew Crooks. *Challenges and Opportunities of Social Media Data for Socio-Environmental Systems Research.* In: *Land* 8 (July 2019). MDPI. DOI: `10.3390/land8070107`. [Last access: May 13, 2023].

[77] Lourdes López-Pérez and María Dolores Olvera-Lobo. *Chapter 9. Social Media as Channels for the Public Communication of Science: The Case of Spanish Research Centers and Public Universities.* In: *Use and Users.* Ed. by Kathrin Knautz and Katsiaryna S. Baran. Berlin, Boston: De Gruyter Saur, 2016, pp. 241–264. ISBN: 9783110418163. DOI: `doi:10.1515/9783110418163-010`. Available at: `https://doi.org/10.1515/9783110418163-010` [Last access: May 12, 2023].

[78] Apoorva Mandavilli. *Peer review: Trial by Twitter. Nature 469, 286-287.* In: *Nature* 469 (January 2011). MDPI, pp. 286–7. DOI: `10.1038/469286a`. [Last access: May 15, 2023].

[79] Gonçalo Marques, Nuno Miranda, Akash Kumar Bhoi, Begonya Garcia-Zapirain, Sofiane Hamrioui, and Isabel de la Torre Díez. *Internet of Things and Enhanced Living Environments: Measuring and Mapping Air Quality Using Cyber-physical Systems and Mobile Computing Technologies.* In: *Sensors* 20.3 (2020). MDPI. ISSN: 1424-8220. DOI: `10.3390/s20030720`. Available at: `https://www.mdpi.com/1424-8220/20/3/720` [Last access: May 12, 2023].

[80] Ted Marzilli. *Old and young US adults most susceptible to fake news.* 2018. Available at: `https://today.yougov.com/topics/entertainment/articles-reports/2018/06/13/old-and-young-us-adults-most-susceptible-fake-news` [Last access: May 13, 2023].

[81] Mauna Loa Observatory. *Mauna Loa Observatory.* Available at: `https://gml.noaa.gov/obop/mlo/` [Last access: June 12, 2023].

[82] Saul Mcleod. *Likert Scale Definition, Examples And Analysis.* 2023. Available at: `https://www.simplypsychology.org/likert-scale.html` [Last access: June 8, 2023].

[83] mediatemplate.net. *4 Tips for Making Data Visualizations Look Good on Mobile.* 2021. Available at: https://mediatemple.net/blog/design-creative/4-tips-for-making-data-visualizations-look-good-on-mobile/ [Last access: May 16, 2023].

[84] Meta. *About Instagram Insights.* Available at: https://www.facebook.com/business/help/441651653251838?id=419087378825961 [Last access: June 16, 2023].

[85] Meta. *How fact-checking works.* Available at: https://transparency.fb.com/en-gb/features/how-fact-checking-works/ [Last access: May 28, 2023].

[86] Meta. *Instagram.* Available at: https://about.meta.com/technologies/instagram/ [Last access: June 16, 2023].

[87] Sara Metwalli. *What Is Data Visualization?* 2023. Available at: https://builtin.com/data-science/data-visualization [Last access: June 4, 2023].

[88] Meg Miller. *The problem with interactive graphics.* 2017. Available at: https://www.fastcompany.com/3069008/the-problem-with-interactive-graphics [Last access: May 16, 2023].

[89] Minnesota Department of Health. *Carbon Dioxide (CO_2).* 2022. Available at: https://www.health.state.mn.us/communities/environment/air/toxins/co2.html [Last access: May 10, 2023].

[90] Lidia Morawska, Phong K. Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, Fahe Chai, Bryce Christensen, Matthew Dunbabin, Jian Gao, Gayle S.W. Hagler, Rohan Jayaratne, Prashant Kumar, Alexis K.H. Lau, Peter K.K. Louie, Mandana Mazaheri, Zhi Ning, Nunzio Motta, Ben Mullins, Md Mahmudur Rahman, Zoran Ristovski, Mahnaz Shafiei, Dian Tjondronegoro, Dane Westerdahl, and Ron Williams. *Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?* In: *Environment International* 116 (2018). Elsevier Publication, pp. 286–299. ISSN: 0160-4120. DOI: https://doi.org/10.1016/j.envint.2018.04.018. Available at: https://www.sciencedirect.com/science/article/pii/S0160412018302460 [Last access: May 11, 2023].

[91] NETGEAR. *Che cos'è la tecnologia PoE? (Power over Ethernet).* 2021. Available at: https://kb.netgear.com/it/209/Che-cos-%5C%C3%5C%A8-la-tecnologia-PoE-Power-over-Ethernet?language=it [Last access: May 15, 2023].

[92] NOAA. Available at: https://www.noaa.gov/ [Last access: May 12, 2023].

[93] NOAA. *CO2_Earth Twitter account.* Available at: https://twitter.com/CO2_earth [Last access: May 12, 2023].

[94] Monique Noirhomme-Fraiture, Frédéric Randolet, Luca Chittaro, and Grégory Custinne. *Data visualizations on small and very small screens.* In: *Proceedings of ASMDA*. January 2005.

[95] Juan Emilio Noriega-Linares and Juan Miguel Navarro Ruiz. *On the Application of the Raspberry Pi as an Advanced Acoustic Sensor Network for Noise Monitoring.* In: *Electronics* 5.4 (2016). MDPI. ISSN: 2079-9292. DOI: 10.3390/electronics5040074. Available at: https://www.mdpi.com/2079-9292/5/4/74 [Last access: May 10, 2023].

[96] Petra Novakova and Michal Kraus. *Carbon Dioxide Concentration in the Bedroom for Various Natural Ventilation Modes.* In: *IOP Conference Series: Materials Science and Engineering* 603.5 (2019). IOP Publishing, p. 052100. DOI: 10.1088/1757-899X/603/5/052100. Available at: https://dx.doi.org/10.1088/1757-899X/603/5/052100 [Last access: May 12, 2023].

[97] Hanna Nowicki and Carter Merenstein. *Radar*. 2016. Available at: https://www.cs.middlebury.edu/~candrews/showcase/infovis_techniques_s16/radar_chart/ [Last access: June 7, 2023].

[98] NSTA. *Writing a scientific survey.* Available at: https://static.nsta.org/ecybermission-files/helpdocs/Writing%5C%20a%5C%20Scientific%5C%20Survey.pdf [Last access: May 30, 2023].

[99] Heather O'Brien, Paul Cairns, and Mark Hall. *A Practical Approach to Measuring User Engagement with the Refined User Engagement Scale (UES) and New UES Short Form.* In: *International Journal of Human-Computer Studies* 112 (April 2018). Taylor and Francis Ltd. DOI: 10.1016/j.ijhcs.2018.01.004. [Last access: June 8, 2023].

[100] Office of astronomy for development. *General Guidelines on Designing a Questionniare.* Available at: https://www.astro4dev.org/surveys-assessments/ [Last access: May 30, 2023].

[101] Sofia S. Oliveira, Bárbara Barros, Joana L. Pereira, Paulo T. Santos, and Ruth Pereira. *Social Media Use by Citizen Science Projects: Characterization and Recommendations.* In: *Frontiers in Environmental Science* 9 (2021). Frontiers Media S.A. ISSN: 2296-665X. DOI: 10.3389/fenvs.2021.715319. Available at: https://www.frontiersin.org/articles/10.3389/fenvs.2021.715319 [Last access: May 12, 2023].

[102] Yasemin Özkent. *Social media usage to share information in communication journals: An analysis of social media activity and article citations.* In: *PLOS ONE* 17 (February

2022). Public Library of Science, e0263725. DOI: 10.1371/journal.pone.0263725. [Last access: May 12, 2023].

[103] Andrés Pastor-Fernández, Alberto Cerezo-Narváez, Paz Montero-Gutiérrez, Pablo Ballesteros-Pérez, and Manuel Otero-Mateo. *Use of Low-Cost Devices for the Control and Monitoring of CO2 Concentration in Existing Buildings after the COVID Era*. In: *Applied Sciences* 12.8 (2022). MDPI. ISSN: 2076-3417. DOI: 10.3390/app12083927. Available at: https://www.mdpi.com/2076-3417/12/8/3927 [Last access: May 12, 2023].

[104] José Pecly. *Strategies to obtain a better quality of environmental data gathered by low cost systems*. In: *Environmental Monitoring and Assessment* 195 (January 2023). Springer. DOI: 10.1007/s10661-022-10805-2. [Last access: May 10, 2023].

[105] Fabio Pittarello and Lorenzo Brutti. *A Public Tangible Interface for Engaging and Informing Young Citizens about Climate Change*. In: *Proceedings of the 8th ACM International Symposium on Pervasive Displays*. PerDis '19. Palermo, Italy: Association for Computing Machinery, 2019. ISBN: 9781450367516. DOI: 10.1145/3321335.3324938. Available at: https://doi.org/10.1145/3321335.3324938.

[106] D.M.G. Preethichandra. *Design of a smart indoor air quality monitoring wireless sensor network for assisted living*. In: *Proceedings of 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013, pp. 1306–1310. DOI: 10.1109/I2MTC.2013.6555624. [Last access: May 12, 2023].

[107] Real World Visuals. *Real World Visuals - See the world's carbon emissions (in several ways)*. 2022. Available at: https://www.realworldvisuals.com/blog-1/see-the-worlds-carbon-emissions-in-several-ways [Last access: May 13, 2023].

[108] National Association of realtors and CRT Labs. *#PiAQ*. 2016. Available at: https://github.com/NationalAssociationOfRealtors/PiAQ [Last access: May 10, 2023].

[109] Hannah Ritchie and Max Roser. *Our World in Data - Italy: CO2 Country Profile*. Available at: https://ourworldindata.org/co2/country/italy [Last access: May 13, 2023].

[110] Irene Ros and Bocoup. *MobileVis*. Available at: https://mobilev.is/ [Last access: May 29, 2023].

[111] Irene Ros and Bocoup. *MobileVis Patterns*. Available at: http://patterns.mobilev.is/ [Last access: May 29, 2023].

[112] Aislinn D. Rowan-Nash, Benjamin J. Korry, Eleftherios Mylonakis, and Peter Belenky. *Cross-Domain and Viral Interactions in the Microbiome*. In: *Microbiology and Molecular Biology Reviews* 83.1 (2019). American Society for Microbiology, e00044–

18. DOI: 10.1128/MMBR.00044-18. Available at: https://journals.asm.org/doi/abs/10.1128/MMBR.00044-18 [Last access: May 12, 2023].

[113] Emilia Ruzicka. *How to create data visualisations that serve the public*. 2022. Available at: https://datajournalism.com/read/longreads/accessible-data-visualisation [Last access: June 14, 2023].

[114] Troy Sadler, Aidin Amirshokoohi, Mahsa Kazempour, and Kathleen Allspaw. *Socioscience and ethics in science classrooms: Teacher perspectives and strategies*. In: *Journal of Research in Science Teaching* 43 (April 2006). John Wiley & Sons, pp. 353–376. DOI: 10.1002/tea.20142. [Last access: May 25, 2023].

[115] Troy Sadler, Sasha Barab, and Brianna Scott. *What Do Students Gain by Engaging in Socioscientific Inquiry?* In: *Res. Sci. Ed.* 37 (August 2007). Springer, pp. 371–391. DOI: 10.1007/s11165-006-9030-9. [Last access: May 25, 2023].

[116] Jagriti Saini, Maitreyee Dutta, and Gonçalo Marques. *Indoor Air Quality Monitoring Systems Based on Internet of Things: A Systematic Review*. In: *International Journal of Environmental Research and Public Health* 17.14 (2020). MDPI. ISSN: 1660-4601. DOI: 10.3390/ijerph17144942. Available at: https://www.mdpi.com/1660-4601/17/14/4942 [Last access: May 12, 2023].

[117] Kevin Schaul and Bonnie Berkowitz. *Leagues of the World Cup*. 2014. Available at: https://www.washingtonpost.com/wp-srv/special/sports/leagues-of-the-world-cup/ [Last access: April 30, 2023].

[118] Philipp Schneider, Matthias Vogt, Rolf Haugen, Amirhossein Hassani, Nuria Castell, Franck R. Dauge, and Alena Bartonova. *Deployment and Evaluation of a Network of Open Low-Cost Air Quality Sensor Systems*. In: *Atmosphere* 14.3 (2023). MDPI. ISSN: 2073-4433. DOI: 10.3390/atmos14030540. Available at: https://www.mdpi.com/2073-4433/14/3/540 [Last access: May 11, 2023].

[119] SCRIPPS Institution of Oceanography. *The Keeling Curve*. 2023. Available at: https://keelingcurve.ucsd.edu/ [Last access: May 12, 2023].

[120] Monika Sester and Claus Brenner. *Continuous Generalization for fast and smooth Visualization on small Displays*. In: *International Archives of Photogrammetry and Remote Sensing*. Vol. 2004. Copernicus Publications, September 2004.

[121] B. Shneiderman. *The eyes have it: a task by data type taxonomy for information visualizations*. In: *Proceedings 1996 IEEE Symposium on Visual Languages*. 1996, pp. 336–343. DOI: 10.1109/VL.1996.545307. [Last access: May 16, 2023].

[122] Carl Skoglund and Dennis Sundin. *Data analysis towards environmental and traffic information from testbed deployments of low-cost IoT sensors*. MA thesis. Linköping University, Faculty of Science & Engineering, 2021.

[123] Aaron Smith, Laura Silver, Courtney Johnson, and Jingjing Jiang. *Publics in Emerging Economies Worry Social Media Sow Division, Even as They Offer New Chances for Political Engagement*. Pew Research Center. 2018. Available at: https://www.pewresearch.org/internet/2019/05/13/publics-in-emerging-economies-worry-social-media-sow-division-even-as-they-offer-new-chances-for-political-engagement/ [Last access: May 24, 2023].

[124] Emily G Snyder, Timothy H Watkins, Paul A Solomon, Eben D Thoma, Ronald W Williams, Gayle SW Hagler, David Shelow, David A Hindin, Vasu J Kilaru, and Peter W Preuss. *The changing paradigm of air pollution monitoring*. In: *Environmental science & technology* 47.20 (2013). ACS Publications, pp. 11369–11377. [Last access: May 11, 2023].

[125] Luca Stabile, Giorgio Buonanno, Andrea Frattolillo, and Marco Dell'Isola. *The effect of the ventilation retrofit in a school on CO2, airborne particles, and energy consumptions*. In: *Building and Environment* 156 (2019), pp. 1–11. ISSN: 0360-1323. DOI: https://doi.org/10.1016/j.buildenv.2019.04.001. Available at: https://www.sciencedirect.com/science/article/pii/S0360132319302331 [Last access: May 10, 2023].

[126] Matteo Starri. *DIGITAL 2023 – I dati globali*. We Are Social. 2023. Available at: https://wearesocial.com/it/blog/2023/01/digital-2023-i-dati-globali/ [Last access: June 5, 2023].

[127] Matteo Starri. *DIGITAL 2023 – I dati italiani*. We Are Social. 2023. Available at: https://wearesocial.com/it/blog/2023/02/digital-2023-i-dati-italiani/ [Last access: June 5, 2023].

[128] Erica Stockwell, Emma L. Ballard, Peter O'Rourke, Luke D. Knibbs, Lidia Morawska, and Scott C. Bell. *Indoor hospital air and the impact of ventilation on bioaerosols: a systematic review*. In: *Journal of Hospital Infection* 103.2 (2019). Elsevier Publication, pp. 175–184. ISSN: 0195-6701. DOI: https://doi.org/10.1016/j.jhin.2019.06.016. Available at: https://www.sciencedirect.com/science/article/pii/S0195670119302786 [Last access: May 12, 2023].

[129] Jeannette Sutton, Leysia Palen, and Irina Shklovski. *Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires*. In: *Proceedings of the 5th International ISCRAM Conference* (April 2008).

[130] SwitchDocLabs. *GroveWeatherPi - Solar Raspberry Pi Based Weather Station - No Soldering Required*. 2016. Available at: https://www.instructables.com/GroveWeatherPi-Raspberry-Pi-Based-Weather-Station-/.

[131] Life Redune Team. *Life Redune*. 2020. Available at: https://liferedune.it/.

[132] The R Foundation. *R: What is R?* Available at: `https://www.r-project.org/about.html` [Last access: May 28, 2023].

[133] ThoughtSpot. *"12 Ways to improve your data visualization design"*. 2022. Available at: `https://www.thoughtspot.com/data-trends/data-visualization/ways-to-improve-data-visualization-design` [Last access: May 16, 2023].

[134] Kai Tomboc. *Text vs. Images: Which Content Format is Effective?* 2019. Available at: `https://www.easel.ly/blog/text-vs-images-which-content-format-effective/` [Last access: June 16, 2023].

[135] Milka Trajkova, A'aeshah Alhakamy, Francesco Cafaro, Sanika Vedak, Rashmi Mallappa, and Sreekanth R. Kankara. *Exploring Casual COVID-19 Data Visualizations on Twitter: Topics and Challenges*. In: *Informatics* 7.3 (2020). MDPI. ISSN: 2227-9709. DOI: `10.3390/informatics7030035`. Available at: `https://www.mdpi.com/2227-9709/7/3/35` [Last access: May 16, 2023].

[136] Archie Tse. *Why we are doing fewer interactives*. 2016. Available at: `https://github.com/archietse/malofiej-2016/blob/master/tse-malofiej-2016-slides.pdf` [Last access: May 16, 2023].

[137] Simona Tselova and Annie Parker. *A guide to creating charts and graphs for social media*. 2023. Available at: `https://flourish.studio/blog/social-media-charts-and-graphs/` [Last access: May 16, 2023].

[138] U.S. Environmental Protection Agency. *Draft roadmap for next generation air monitoring*. In: *Environmental Protection Agency: Office of Research and Development* 2 (2013). Ed. by Kurt Toppert. Available at: `https://www.epa.gov/sites/default/files/2014-09/documents/roadmap-20130308.pdf` [Last access: June 16, 2023].

[139] Università Ca' Foscari di Venezia & ISAC-CNR. *Contributo del traffico marittimo alla concentrazione di particolato atmosferico: Estratto dei risultati del progetto Ecomobility a Venezia*. 2018. Available at: `https://www.italy-croatia.eu/documents/118101/167289/ECOMOBILITY+results+Venice+-+summary.pdf/62700404-13aa-6fa5-e913-8ed3296417f0?t=1586435331501` [Last access: May 11, 2023].

[140] Mar Viana, Pieter Hammingh, Augustin Colette, Xavier Querol, Bart Degraeuwe, Ina de Vlieger, and John van Aardenne. *Impact of maritime transport emissions on coastal air quality in Europe*. In: *Atmospheric Environment* 90 (2014). Elsevier Publication. DOI: `https://doi.org/10.1016/j.atmosenv.2014.03.046`.

[141] Visual Cinnamon Team. *Techniques for Data Visualization on both Mobile & Desktop*. 2022. Available at: `https://www.visualcinnamon.com/2019/04/mobile-vs-desktop-dataviz/` [Last access: May 16, 2023].

[142] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. *Guidelines for Using Multiple Views in Information Visualization*. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '00. Palermo, Italy: Association for Computing Machinery, 2000, pp. 110–119. ISBN: 1581132522. DOI: 10.1145/345513.345271. Available at: https://doi.org/10.1145/345513.345271 [Last access: May 16, 2023].

[143] Ågot K. Watne, Jenny Linden, Jens Willhelmsson, Håkan Fridén, Malin Gustafsson, and Nuria Castell. *Tackling Data Quality When Using Low-Cost Air Quality Sensors in Citizen Science Projects*. In: *Frontiers in Environmental Science* 9 (2021). Frontiers Media S.A. ISSN: 2296-665X. DOI: 10.3389/fenvs.2021.733634. Available at: https://www.frontiersin.org/articles/10.3389/fenvs.2021.733634 [Last access: May 12, 2023].

[144] Benjamin Watson and Vidya Setlur. *Emerging Research in Mobile Visualization*. In: August 2015, pp. 883–887. DOI: 10.1145/2786567.2786571. [Last access: May 16, 2023].

[145] Whatagraph Team. *Social Media Data Visualization - The Ultimate Guide*. 2021. Available at: https://whatagraph.com/blog/articles/social-media-data-visualization [Last access: May 16, 2023].

[146] Richard Wike, Laura Silver, Janell Fetterolf, Christine Huang, Sarah Austin, Laura Clancy, and Sneha Gubbala. *Views of social media and its impacts on society*. Pew Research Center. 2022. Available at: https://www.pewresearch.org/global/2022/12/06/views-of-social-media-and-its-impacts-on-society-in-advanced-economies-2022/ [Last access: May 24, 2023].

[147] Wikipedia. *Color blindess*. Available at: https://en.wikipedia.org/wiki/Color_blindness [Last access: June 4, 2023].

[148] Wikipedia. *Echo chamber (media)*. Available at: https://en.wikipedia.org/wiki/Echo_chamber_(media) [Last access: May 13, 2023].

[149] Wikipedia. *Generation Z*. 2023. Available at: https://en.wikipedia.org/wiki/Generation_Z [Last access: May 24, 2023].

[150] Wikipedia. *Millennials*. 2023. Available at: https://en.wikipedia.org/wiki/Millennials [Last access: May 24, 2023].

[151] Clare Wilkinson and Emma Weitkamp. *A Case Study in Serendipity: Environmental Researchers Use of Traditional and Social Media for Dissemination*. In: *PLOS ONE* 8.12 (December 2013). Public Library of Science. DOI: 10.1371/journal.pone.0084339. Available at: https://doi.org/10.1371/journal.pone.0084339 [Last access: May 12, 2023].

[152] USGS: Science for a changing world. *What is carbon sequestration?* 2018. Available at: https://www.usgs.gov/faqs/what-carbon-sequestration.

[153] World Health Organisation. *Ambient (outdoor) air pollution.* 2022. Available at: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health [Last access: May 13, 2023].

[154] World Health Organization. Regional Office for Europe. *WHO guidelines for indoor air quality: selected pollutants.* Geneva: World Health Organisation, 2010, 454 p.

[155] Aoyu Wu, Wai Tong, Tim Dwyer, Bongshin Lee, Petra Isenberg, and Huamin Qu. *MobileVisFixer: Tailoring Web Visualizations for Mobile Phones Leveraging an Explainable Reinforcement Learning Framework.* In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2021). IEEE, pp. 464–474. DOI: 10.1109/TVCG.2020.3030423. [Last access: May 16, 2023].

[156] Kirby Wu. *Islamic Terrorist Attacks All over the World, 1983 - 2015.* Available at: https://0media.tw/t/geoevent/ [Last access: May 29, 2023].

[157] Wataru Yamada, Kazuhiro Yamada, Hiroyuki Manabe, and Daizo Ikeda. *ISphere: Self-Luminous Spherical Drone Display.* In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.* UIST '17. Québec City, QC, Canada: Association for Computing Machinery, 2017, pp. 635–643. ISBN: 9781450349819. DOI: 10.1145/3126594.3126631. Available at: https://doi.org/10.1145/3126594.3126631 [Last access: May 16, 2023].

[158] He Zhang, Ravi Srinivasan, and Vikram Ganesan. *Low Cost, Multi-Pollutant Sensing System Using Raspberry Pi for Indoor Air Quality Monitoring.* In: *Sustainability* 13.1 (2021). MDPI. ISSN: 2071-1050. DOI: 10.3390/su13010370. Available at: https://www.mdpi.com/2071-1050/13/1/370 [Last access: May 12, 2023].