UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

Trabajo Fin de Grado en Ingeniería Informática

Desarrollo de un Dashboard para la Evolución de la COVID-19
Development of a Dashboard for the Evolution of COVID-19

Dirigido por:
Sonia Estévez Martín

Adrián García Sánchez
Harold Luis Pascua Cajucom
Dale Francis Valencia Calicdan

Curso académico 2020-21

Convocatoria Junio

Acknowledgements

First of all, we want to give thanks to our project director, Sonia Estévez Martin, who has always helped us when we have needed it and has given us guidelines to develop the project in the best possible way. We also want to mention Victoria López, who has also helped us in planning the work and choosing adequate metrics for the correct preparation.

We also want to acknowledge all the efforts made on the part of the partner group, composed of Iván Jesús Saugar Peinado, Sergio Sánchez Ortiz, and Álvaro San José Guerra, who have provided us with the datasets with which to work, a fundamental part for the development of the project and the achievement of the objectives.

Finally, we want to thank all family and friends for all the moral support given throughout this project.

Abstract

The SARS-CoV-2 coronavirus is responsible for the COVID-19 disease that has caused economic and health issues around the world. To enforce measures that curbs the growth of this disease, this pandemic must be monitored.

The objective of this project is to create an online, open source dashboard that monitors the evolution of COVID-19 by country and by its principal indicators. To do so, this project has developed a series of Scrum sprints, together with various technologies, such as RStudio, Shiny, and Google Sites.

The dashboard created serves to analyze data and to present it in a clear manner because of mastery of different tools, methodologies, and analytical techniques that have been used.

This dashboard differs from other existing dashboards such that it specializes in presenting indicators such as transmission index, danger index, and cumulative incidence in addition to raw data like positive cases, deaths, and recoveries.

As a result, a dashboard was created that can inform users about the progress of the pandemic in graphic and map forms, especially under the aforementioned indicators.

Keywords

COVID-19, Coronavirus, Dashboard, RStudio, Shiny, Data Analysis, Accumulated Incidence, R0, Danger Index

Resumen

El SARS-CoV-2 coronavirus es responsable de la COVID-19, la cual ha causado problemas económicos y sanitarios alrededor de todo el mundo. Para hacer cumplir medidas que frenen el crecimiento de la enfermedad, la pandemia debe ser monitorizada.

El objetivo de este proyecto es crear un dashboard online y de código abierto que monitorice la evolución del COVID-19 por país y por sus principales indicadores. Para lograr esto, el proyecto ha sido elaborado mediante una serie de iteraciones de Scrum, unidas a varias tecnologías como RStudio, Shiny y Google Sites.

El dashboard creado sirve para analizar los datos y para presentarlos de una manera clara, debido al dominio de las diferentes herramientas, metodologías y técnicas analíticas que han sido utilizadas.

Este dashboard difiere de otros existentes en que se especializa en presentar algunos indicadores como son el índice de transmisión, el índice de peligrosidad y la incidencia acumulada además de datos brutos como son los casos positivos, las muertes y los recuperados.

Como resultado, ha sido creado un dashboard que informa a los usuarios sobre el progreso de la pandemia mediante mapas y gráficos, especialmente sobre los indicadores mencionados anteriormente.

Palabras Clave

COVID-19, Coronavirus, Dashboard, RStudio, Shiny, Análisis de datos, Incidencia Acumulada, R0, Índice de peligrosidad

Contents

List of Figures

Chapter 1

Introduction

COVID-19 [1] is an infectious disease caused by the coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It is held responsible for the pandemic that was happening around the world declared by the World Health Organization (WHO) on March 11, 2020. An article published by WHO [2] suggested that the start of the outbreak of this virus originated from Wuhan, China at a local market around late December 2019.

Evidently, the COVID-19 pandemic brought devastating effects to everyone around the globe. Public health facilities were brought to their limits [3], [4], economies worldwide took a massive hit [5], and, most importantly, many lives have been lost. WHO had stated that, as of November 1, 2020, nearly 46 million positive cases and 1.2 million deaths have been reported from around the world [6].

As such, the global phenomenon has challenged every state and country to formulate their own countermeasures and impede the rate at which citizens are being infected. While some of these implementations have been proven to be effective, other factors such as inexperienced health workers, poor health systems, lack of social distancing, and scarce knowledge of the coronavirus make this problem all the more difficult to resolve.

1.1 Motivation

Various dashboards, both online and offline, have already been developed for analytical use, with some having been developed to visually represent the development of the COVID-19 pandemic in numerous aspects. Various features are also integrated into these dashboards to extend their functionality and, when used properly, facilitate the interpretation of the data for the users.

However, the existing dashboards mainly present raw information such as the number of infections, the number of deaths, and the number of recoveries. This kind of information is difficult to use as a basis not only for health and government officials in establishing preventive measures, but also the general public in understanding the state of the pandemic. Therefore, indicators must also be presented in dashboards to provide deeper insights and ease the understanding of the situation.

The dashboard created in this project aims to complement existing COVID-19 dashboards by providing the previously mentioned indicators. This would help assist users in understanding the pandemic, be it for making information-backed decisions, for conducting further research, or simply for staying updated on the situation.

It is also important to have a wide range of knowledge to successfully develop this project. Some of the subjects seen throughout the career have been fundamental for the proper development of the work. The Web Applications subject (*Aplicaciones Web*) covers the whole idea of the project while the process of developing the project and designing the application is tackled in the subjects of Software Engineering (*Ingeniería del Software*) and Interactive Systems Development (*Diseño de Sistemas Interactivos*). As the dashboard created involved analyzing data on the ongoing pandemic, the Probability and Statistics subject (*Probabilidad y Estadística*) played an important role. Finally, skills acquired through the various programming subjects taken throughout the degree (*Fundamentos de Programación*, *Estructura de Datos*, *Fundamentos de Algoritmos*, *Tecnología de Programación*, etc.) helped in adapting to any requirements in creating software, such as using different programming languages.

1.2 Objectives

The main goal of this project is to create a visualization system for the evolution of COVID-19 by building a well-designed, web-based dashboard that is able to present data retrieved from reliable sources. This project aims to visually represent a simplified analysis of the growth of COVID-19 via a user-friendly dashboard filled with graphs, maps and other visual aids. The project should provide different interactive features for further analysis of the different representations of gathered data. The goals of this project are enumerated as follows:

- Develop an public, open source, web-based dashboard that:
  - Allows users to follow the progression of COVID-19 in different regions/countries, as well as globally
  - Analyzes data under the perspective of different variables that compose the COVID-19 analytical model used as basis
  - Presents organized summaries and graphs for data visualization
  - Is accessible to as many users as possible in terms of usability and clarity of data
  - Competes yet complements with existing online COVID-19 dashboards in terms of visual and structural quality and user experience
- Create a database to be used in the dashboard that is derived from reliable sources of data and is publicly available (for future analysis and usage in other applications) by downloading from a repository.
- Learn the competencies required to implement the necessary technologies used in developing the dashboard, such as programming languages, and methodologies.

1.3 Work Plan

The procedure followed throughout the project is as follows: case study, initial planning, prototype designing, dashboard creation, and testing and debugging.

To start, case studies were done to gain insights on the existing COVID-19 dashboards, including their characteristics and common features. This would help establish the general expectations that the project should fulfill.

An initial planning is then done to establish the specification and requirements of the project, as well as to plan out the tasks that would need to be done (including scheduling and assignment of the said tasks). An important note in this stage is that, as this is a project created by students, the planning would be done with cost-effectiveness (and zero cost, if possible) in mind.

After creating an initial plan, the development of the dashboard began. This comprised of three main aspects: the preparation of the dataset (in conjunction with a partner group) to be used in the dashboard, the creation of prototypes for the design of the dashboard and the actual implementation of the dashboard via creation of codes and scripts mainly in R/Shiny. This stage was done in an incremental and iterative manner to adapt to changes in requirements that arose and, thus, is the most substantial part of the project.

Finally, the dashboard was tested and debugged to ensure that the dashboard functioned properly. This involved testing for errors in input validations, delays in loading data, and the overall user experience the dashboard provided.

1.4 Document Structure

As this work contains details on both the process done throughout the project and the observations and discoveries made, it is divided into seven chapters for reading convenience.

- Chapter 1 is the introduction, providing context, objectives, and rationale behind the execution of the project.
- Chapter 2 contains the State of the Art, which includes the types of COVID-19 dashboards present, the applications available for dashboard creation, and the technical requirements that dashboards are expected to meet.
- Chapter 3 describes data of interest, such as databases that may be of interest, graphs that may prove useful when displayed on a dashboard, and variables available in making thorough analyses of the evolution of COVID-19
- Chapter 4 explains the databases chosen as bases for creating the dataset used in the dashboard, as well as transformations made in this data to create what would be the final dataset of the dashboard.
- Chapter 5 tackles the details of the developments made on the project and is therefore a focal point in this work. This chapter contains explanations on decisions made in implementing the project such as the methodology used, Scrum, and the hosting service used to contain the

dashboard, which would be Shinyapps.io and Google Sites. The work done in each development sprint executed in this stage is then described along with realizations that would cause changes in the requirements of the dashboard.

- Results are covered on Chapter 6. The user guide for the dashboard and other points for consideration about the results of the project are also discussed.
- Chapter 7 contains the individual contributions made and different tasks executed in the project by the members of the group.
- Finally, Chapter 8 contains the conclusions made after developing the project and recommendations for future work.

Chapter 2

State of the Art

In this chapter, the state of the art is developed to understand the technology behind dashboards. The state of the art then aims to identify the different types of dashboards that are existent, the purposes they serve, the applications that allow the creation and configuration of such dashboards, and the technical requirements that must be satisfied in order to create them.

This state of the art serves two main purposes:

- To analyze the situation of COVID-19 dashboards
- To provide readers with the necessary information to create a dashboard that meets their needs

2.1 Types of COVID-19 Dashboards

Generally speaking, there are four types of Dashboards [7]:

- Strategic Dashboards: dashboards that focus on monitoring the progress of strategies in relation to a company's long-term goals
- Operational Dashboards: dashboards that show data on short-term operational activities by the business
- Analytical Dashboards: a mix of Strategic and Operational Dashboards, these contain data collected from analysts, giving executives a comprehensive overview of a company's growth based on the analysis of historical data

- Tactical Dashboards: dashboards that monitor and analyze processes that are under mid-level management

However, an issue with this sort of categorization is that it only applies to a business context. Should this be applied in relation to COVID-19, most, if not all of them, would fall under the Analytical Dashboard type.

Hence, a new grouping must be implemented to further differentiate the different COVID-19 dashboards that are present. From what was observed from existing examples, these types may depend on two factors: geographic scope and type of data used.

2.1.1 Types of Dashboards by Geographic Scope

The geographical scope of a dashboard may vary, depending on its intended coverage. Three categories were observed to be present: dashboards with a national scope, dashboards with a global scope, and those that are hybrids of the two.

National Scope

Dashboards of this type focus on data that covers the situation of the pandemic in a single country. Usually a national scope permits the presentation of more detailed information about a country, where a dashboard can analyze and present data categorized by municipalities or by provinces. This is because it can allot more of its resources to analyze a single country, rather than dividing it for the analysis of multiple countries as observed in dashboards of global scope.

An example of this sort of dashboards include the COVID-19 Dashboard created by Instituto de Salud Carlos III (ISCIII) [8] (see Figure 2.1), a dashboard that compiles data on the COVID-19 situation in different parts of Spain from government-published and other public sources and presents them such that they can be visualized by region.

Global Scope

On the other hand, dashboards that aim to have a more global scope present data from as many countries as possible to get a larger picture about the development of the pandemic in a worldwide context. Depending on the infrastructure used for storing the data, the data shown for each country

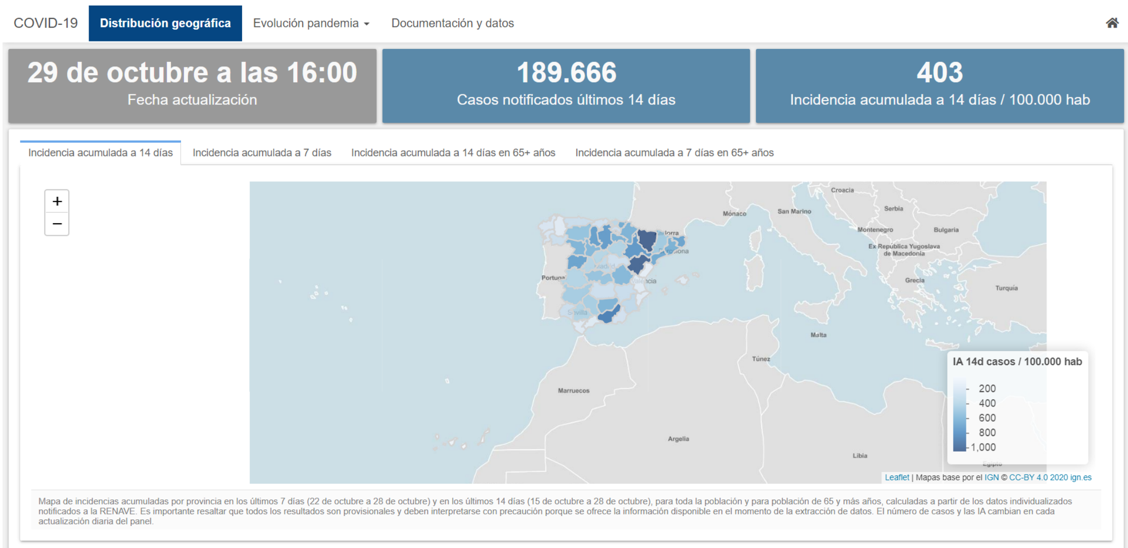

Figure 2.1: ISCIII COVID-19 Dashboard (Application Used: Shiny)

may or may not be detailed. For example, a dashboard that uses data stored in the cloud will have the capacity to present data in a degree of specificity similar to dashboards of national scope, thanks to the ability of cloud-based platforms to have a virtually limitless amount of storage [9], while dashboards that are more limited in storage may opt to show summaries of data for each country instead.

Covidly [10] (see Figure 2.2) is an example of a dashboard with a global scope, projecting data, such as cases, deaths, recoveries, and other variables related to the pandemic, summarized for each country.

Hybrid Scope

While there are dashboards that either have a national scope or a global one, there are certainly dashboards that are a mix of the two, carrying both national data of various countries and local data of certain countries, as seen in the two maps presented by the COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [11] that show both the national data of different countries and the data of the counties in the United States of America (see Figures 2.3 and 2.4).

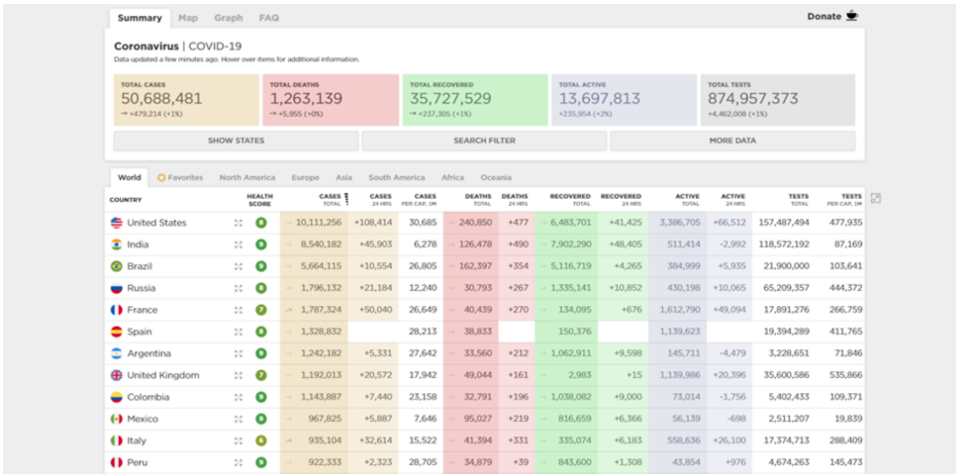

Figure 2.2: Covidly Dashboard (Application Used: Mapbox)

As with dashboards of global scope, the specificity of the data on a per-country level may depend on the capacity of the database being used for the application.

Figure 2.3: Global Map found in the COVID-19 Dashboard by JHU (Application Used: ArcGIS Dashboards)

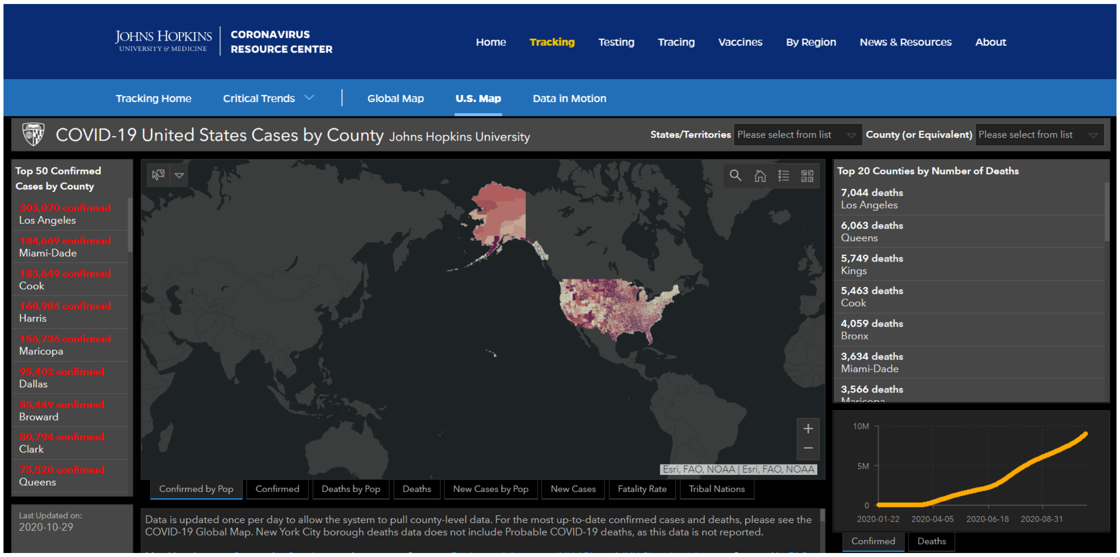

Figure 2.4: U.S. Map found in the COVID-19 Dashboard by JHU (Application Used: ArcGIS Dashboards)

2.1.2 Types of Dashboards by Type of Data Used

In terms of the data that are often analyzed and presented in existing dashboards, it can be observed that dashboards use one of the four: epidemiological data, data regarding clinical trials, data that describe the seroprevalence of the Coronavirus in a certain area, and data on the progress of vaccine development.

Epidemiological Data Tracker

This is the type of dashboard that is usually associated with the term "COVID-19 dashboard". dashboards centered around epidemiological data focus mostly on the number of confirmed cases and deaths, shown both on a daily basis and as an accumulation over a certain period of time. Other data related to the development of the pandemic, including testing rate, hospitalization count, intensive care unit (ICU) admission count, and recoveries, may also be included and presented to provide users a more in-depth analysis.

As previously mentioned, Epidemiological Data Trackers are the most common type of COVID-19 dashboards. Hence, many well-known dashboards, such as the Dashboard of JHU and the World Health Organization (WHO) COVID-19 Dashboard [12] (see Figure 2.5) are good examples.

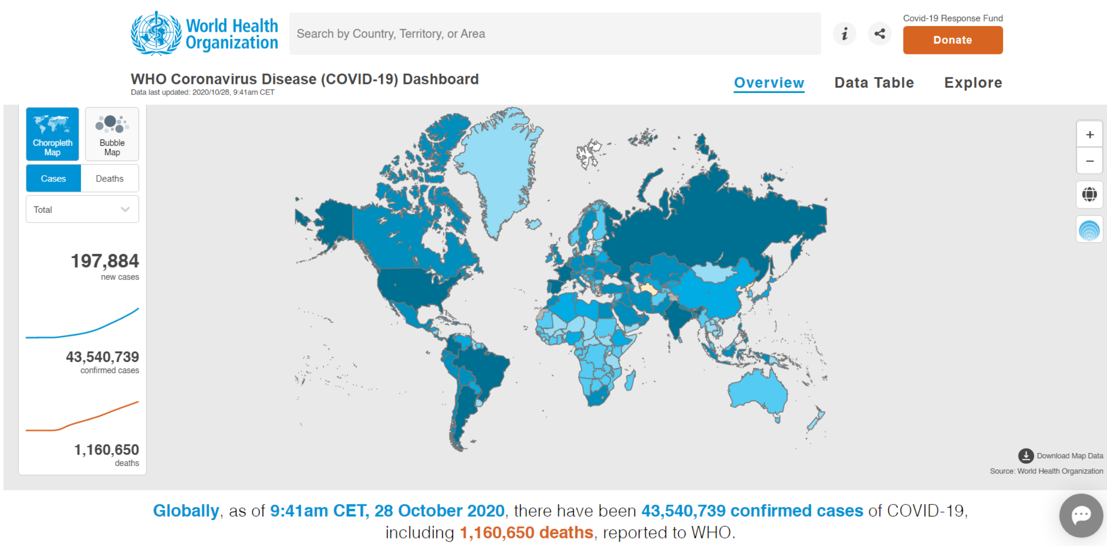

Figure 2.5: WHO COVID-19 Dashboard (Application Used: Sprinklr)

Clinical Trial Tracker

This type of dashboard aims to present users with updates regarding the clinical trial research of COVID-19 intervention methods like vaccines, antiviral drugs, and even traditional Chinese medicine. Usual presented data include trial ID, trial location, trial start date, trial completion date (if applicable), type of treatment, presented outcome and the URL to the results.

Examples include the Covid-19 TrialsTracker [13] (see Figure 2.6), a product of The DataLab at the University of Oxford. It uses data from the International Clinical Trials Registry Platform (ICTRP) to display clinical trials related to the treatment of COVID-19 through tables and graphs. Another example is Covid-Trials.org [14, 15](see Figure 2.7), whose functionality is similar, with the exception of being able to display the data in a map as well.

Seroprevalence Dashboard

Seroprevalence Dashboards, such as SeroTracker [16] (see Figure 2.8), focus on seroprevalence of COVID-19 antibodies. This means that dashboards of this type present the frequency of individuals in a certain population that contain COVID-19 antibodies.

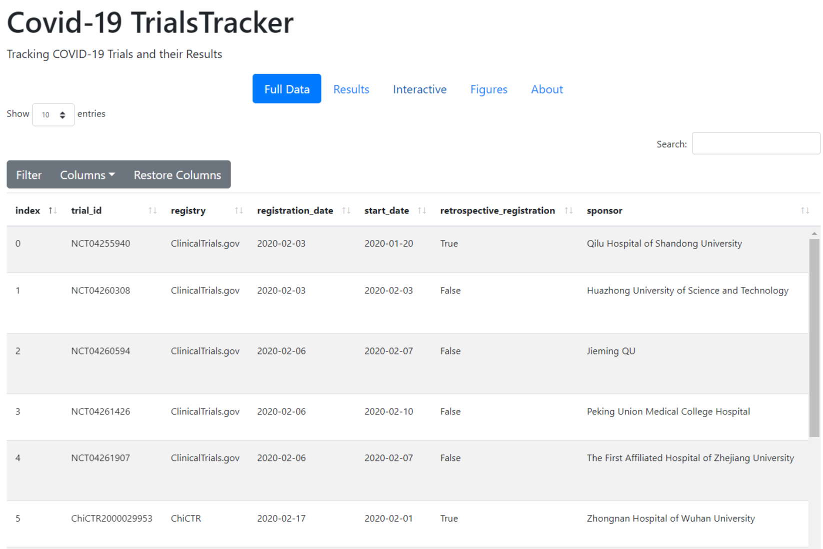

Figure 2.6: Covid-19 TrialsTracker Dashboard (Application Used: Tableau)

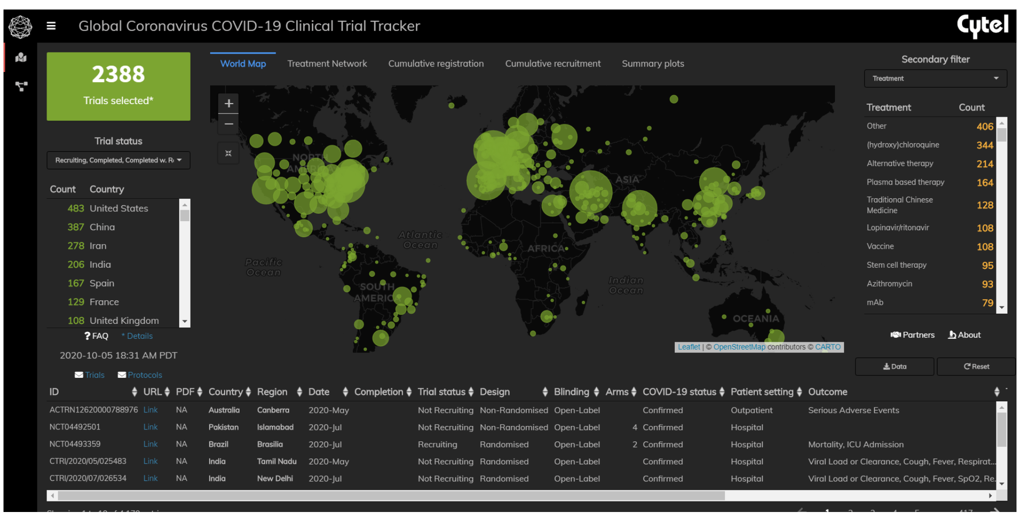

Figure 2.7: Covid-Trials.org Dashboard (Application Used: Cytel)

For health authorities, seroprevalence may be an interesting variable to observe, as it plays a role in antibody testing, a crucial factor in both monitoring the evolution of the Coronavirus and estimating the fatality rate caused by infections [17].

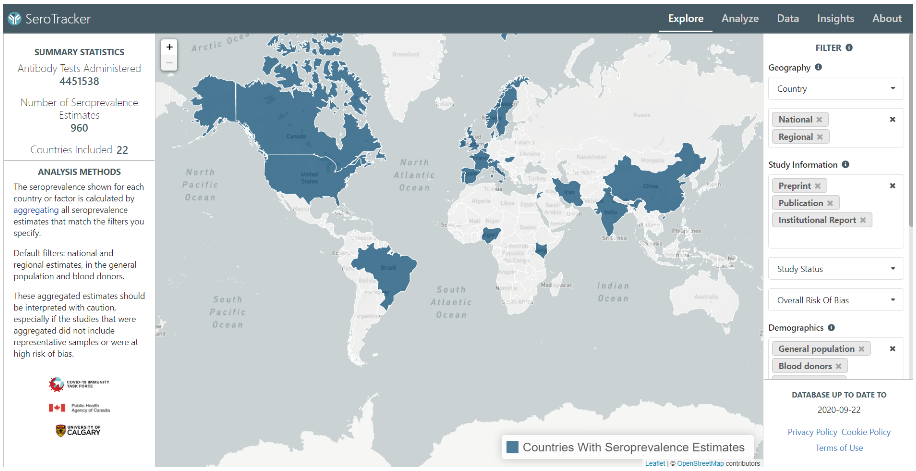

Figure 2.8: SeroTracker Dashboard (Applications Used: Recharts and Leaflet)

Vaccine Tracker

This type of dashboard tracks the progress of the development of vaccines against the Coronavirus. The name of the vaccine and the current status/phase of the vaccine in the development process is included. Additionally, vaccines can be filtered according to their current status.

A good example of this is The Vaccine Tracker [18] (see Figure 2.9).

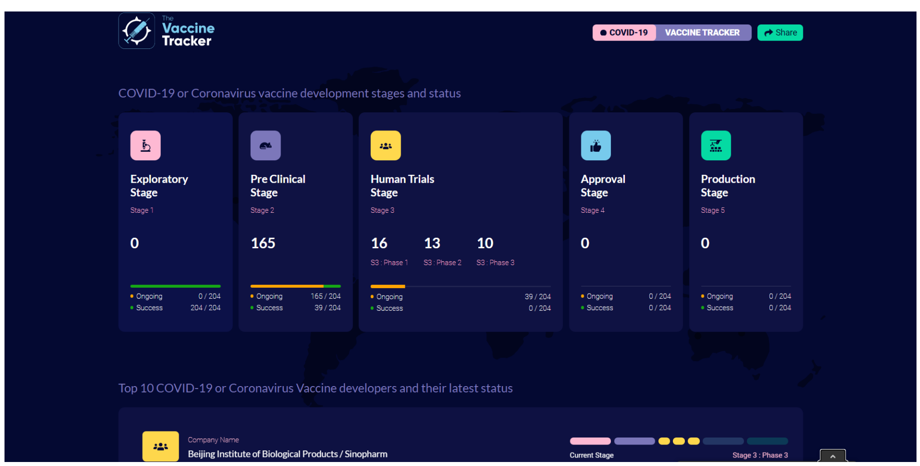

Figure 2.9: The Vaccine Tracker (Applications Used: Gatsby and React)

2.2 Applications for Dashboard Creation and Configuration

There are numerous applications available on the market that specialize on the creation and configuration of dashboards, both paid and free. For illustration purposes, six well-known applications are discussed: Shiny, Power BI, Grafana, Sisense, Domo, and ArcGIS.

Shiny

Shiny [19] is an open-source package based on R [20], a language and environment commonly used for statistical programming that allows an easy creation and configuration of interactive web-based applications, including dashboards. Shiny applications can be made using RStudio [21], the integrated development environment (IDE) of R, and they can also be extended through CSS, htmlwidgets, and JavaScript.

A huge advantage that Shiny has over other applications for dashboard creation is that it is free to use, without imposing limitations of any sort on its complete functionality. Any user can create a fully functional dashboard through Shiny without incurring any costs. Developed products can either be deployed to the Shinyapps.io cloud, with both free and paid options [22], or be integrated on-premises with an existing server.

Power BI

Power BI [23] is a business intelligence (BI) software created by Microsoft. It allows data modeling and visualization through the use of artificial intelligence (AI). In addition, Power BI provides interoperability with other Microsoft applications such as Azure and Excel, end-to-end data protection, and numerous data connectors that permit connections to data sources such as Azure SQL Database and Excel.

While Power BI is a paid service, an option of a free trial is also available for those who want to get an idea of how the application works first hand.

Grafana

Grafana [24] is a free and open-source platform that features support for over 30 databases, including Graphite, InfluxDB, and Prometheus, for use in a single dashboard and has a wide range of dashboards and plugins available in its official library. It also allows the development of dashboards via collaboration and the definition of alerts (for example, when a certain threshold value is surpassed for the data being monitored) that notifies the users.

Sisense

Sisense [25] is a cloud-native platform that gives its users the functions necessary to create powerful analytical solutions such as embedded analytics, data mashups, access controls at system, object, data, and process levels, and customized alerts. It offers Products can be deployed either via a private cloud, on commodity hardware, or by taking a hybrid approach. Sisense also offers data connectors to AWS, Snowflake, and Google BigQuery.

Being a BI platform that focuses on analytics of enterprise data, this application does not offer any free versions.

Domo

Domo [26] is a cloud-based BI platform that consists of three layers: data integration, business intelligence and analytics, and the creation of intelligent applications. It boasts of fast performance, unlimited scalability, data automation, and a large array of data connectors, from databases to even social networks, allowing users from businesses to software developers to focus on creating the product without worrying about the underlying infrastructure.

As with Power BI, Domo also has a free trial option for users who want to try its capabilities in connecting, transforming, and visualizing data.

ArcGIS Dashboards

An online and a desktop platform created by Esri, ArcGIS Dashboards [27] allows the creation of dashboards that feature location-based analytics as its main advantage over other platforms. It also has flexibility, configurability, a suite of available and ready-to-use data visualization tools like maps, lists, and charts, as well as tools that allow users to interact with dashboards created through this application.

This application usually requires to be purchased for use, but it also offers a 21-day free trial period, given that it is for non-production use only.

2.3 Technical Requirements for Creating a Dashboard

The typical architecture of a dashboard consists of four layers, as shown in Figure 2.10 [28]. As such, these will make up the technical requirements that must be complied in order to create a fully functioning dashboard.

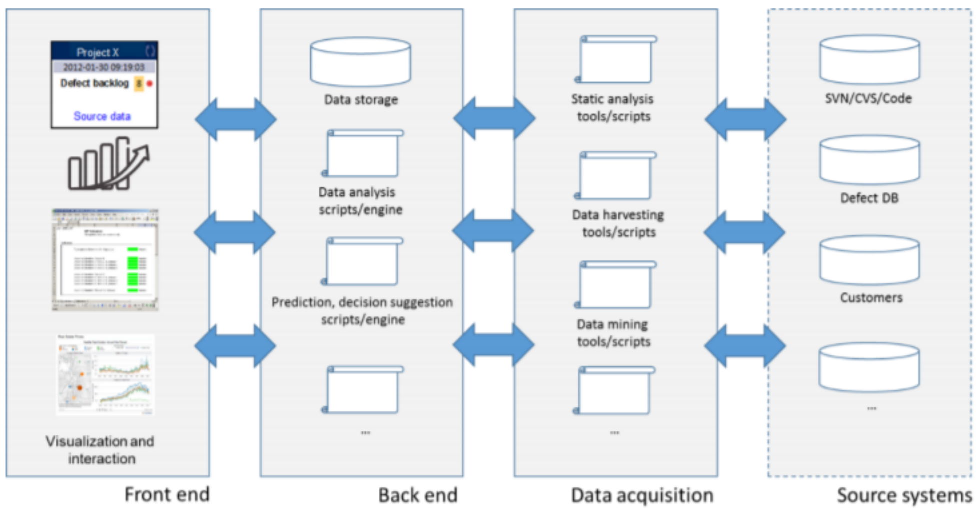

Figure 2.10: Typical architecture of a Dashboard

Front End

Being the part of the application with which the user interacts the most, it is undeniably the most important part of the dashboard. The interface implemented in this layer must be user-friendly and interactive. Careful layout planning and selection of graphs and other types of visual data play a significant role in helping the users understand and interpret clearly the data being shown. In addition, accessibility features, such as appropriately sized text and labels and graph annotations, must also be considered.

Back End

The back end is where the server performs most of its operations. It contains various scripts and source codes for performing analyses and transformations on the data collected, as well as scripts for making predictions and decisions based on the data. In the case of dashboards, the main goal of this layer is to prepare all data for visualization in the front end.

Mechanisms for data storage, such as databases used by the server, can also be found here. As such, knowledge of basic database concepts such as creation, insertion, deletion, aggregation, and projection are a must. It is also important to know the advantages and disadvantages of relational and non-relational databases to determine which type of database is best suited to the application. In the case of dashboards, relational databases, many of which are created using SQL, are ideal as this makes data analysis and presentation easier to do.

Data Acquisition

As explained by its name, the data acquisition layer consists of scripts whose objective is to obtain data from source systems and place that data as metric values in the storage found in the back end. There are multiple methods of acquiring data from these sources, from static analysis to data mining. Hence, knowledge of these processes, as well as knowledge of the structures found in the source systems, is a prerequisite for this layer to function properly.

Source Systems

While source systems are not actually part of the dashboard, they still hold a crucial part in its operations. These systems store data that may be acquired to perform analyses, transformations, and visualizations. Examples include code repositories like GitHub and even databases from other websites.

Chapter 3

Data of Interest for the Project

This chapter explores data that were of importance in implementing the dashboard, such as the different databases and methods of displaying information.

3.1 Existing Databases of Interest

Since the beginning of the pandemic, more and more data are collected and processed, helping in the proliferation of massive databases with extensive groups of data. This can be observed through databases such as MoMo (*Vigilancia de la Mortalidad Diaria*) [29].

One of the main platforms on the Internet for gathering information on various topics is Ourworldindata.org [30], which also contains a vast amount of information about this new pandemic. This database contains data from many countries, together with graphics that allows the acquisition of a more general vision of the subject (see Figure 3.1). The website also allows third parties to download their dataset in CSV format, for further analysis.

Another source that can be highlighted is the EU Open Data Portal [31] that contains, among other things, a dashboard and downloadable data on daily counts of COVID-19 cases worldwide (available in XML, XLSX, CSV, and JSON formats). For example, a table with the number of confirmed cases and deaths for the past 14 days can be found (see Figure 3.2).

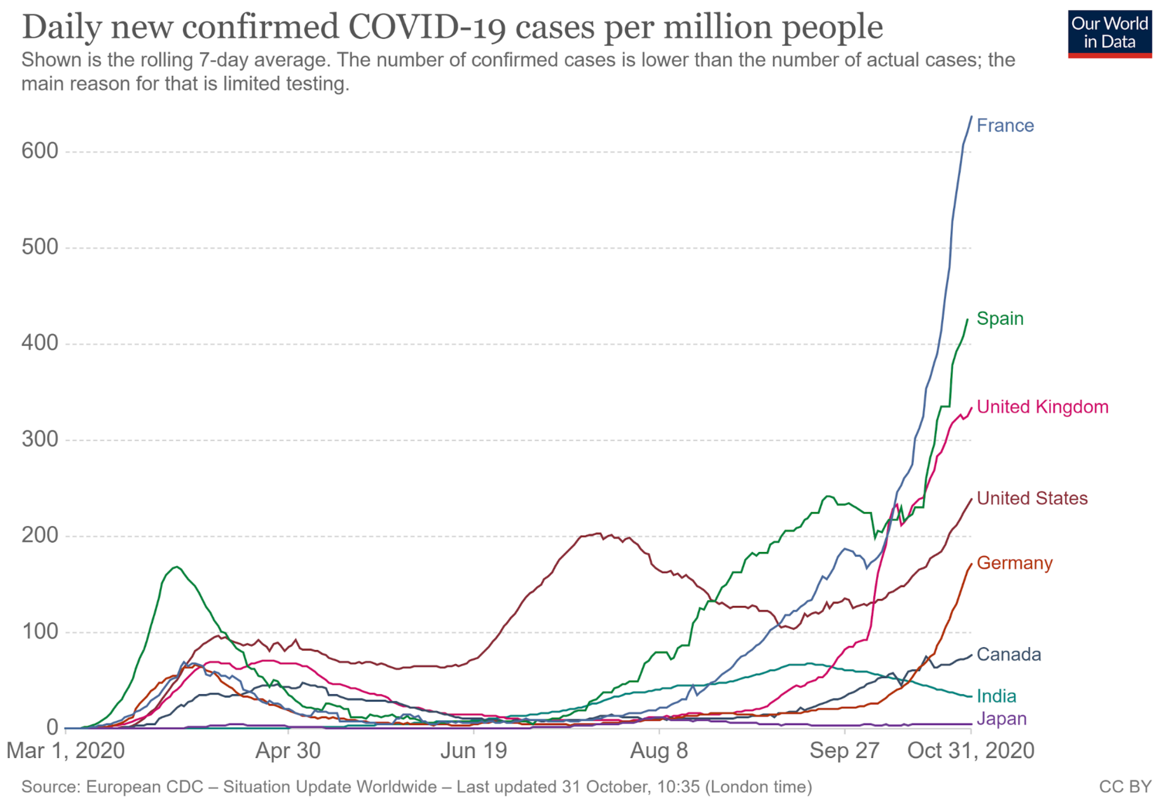

Figure 3.1: Rolling 7-day average of COVID-19 cases as of October 31, 2020

3.2 Graphs of Interest

One of the best ways to display a large amount of data is through graphics. The following types of graphs are those most commonly found in existing dashboards and, therefore, are of interest for presenting data in our own dashboard.

Line Graphs

A line graph is one of the most common and easiest ways to represent data visually and they are perfect for showing trends over a period of time. For example the evolution of the cases with respect to time, with each line representing the evolution of a country's COVID-19 situation, can be observed (see Figure 3.3).

Confirmed cases and deaths - Last 14 days

Cumulative numbers and rates for time period

Country	Cases	Deaths	Cases per 100 000 population	Deaths per 100 000 population
United States Of America	1,335,486	12,342	405.84	3.75
France	648,817	5,678	968.20	8.47
India	643,698	7,597	47.11	0.56
Italy	409,322	4,056	678.14	6.72
United Kingdom	318,213	4,148	477.46	6.22
Poland	292,737	3,434	770.91	9.04
Russian Federation	260,457	4,487	178.55	3.08
Germany	234,002	1,296	281.86	1.56
Spain	230,512	3,802	491.11	8.10
Brazil	195,897	3,972	92.82	1.88
Czechia	156,731	2,657	1,471.68	24.95

Figure 3.2: Table of confirmed cases and deaths for each country in the past 14 days

Maps

Another way of graphically representing the data is with a map. Maps may be used to show the concentration of a variable (such as positive cases or recoveries) in a certain area or its geographic distribution. Depending on the scope of the dashboard, either a world map or a map of a certain country may be used (see Figure 3.4).

Box Plots

Box plots, also known as box-and-whisker plots, are commonly used to demonstrate the concentration of numerical data collected and to determine their variance as a whole. They consist of five parts: the minimum value, the first quartile, the median, the third quartile, and the maximum value; they may also be displayed vertically or horizontally.

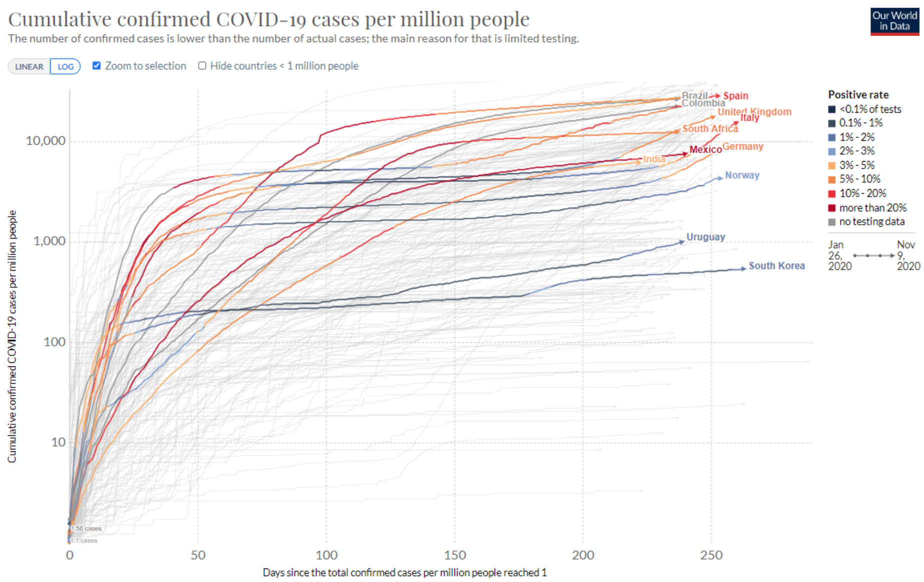

Figure 3.3: Line graph of the cumulative confirmed COVID-19 cases per million people

As explained by the name of the graph, box plots primarily consist of boxes, which represent the interquartile range (IQR), or the difference between the first and third quartiles (see Equation 3.1). These boxes also contain a line in the middle (representing the median), as well as whiskers that represent the sample minimum and maximum values and can extend up to 1.5 times the IQR from both the first and third quartiles [32]. Additionally, outlier data may also be plotted.

$$IQR = Q_3 - Q_1 \tag{3.1}$$

For the COVID-19 dashboard, box plots are useful for determining the variance of readings for variables such as daily cases and deaths [33] (see Figure 3.5).

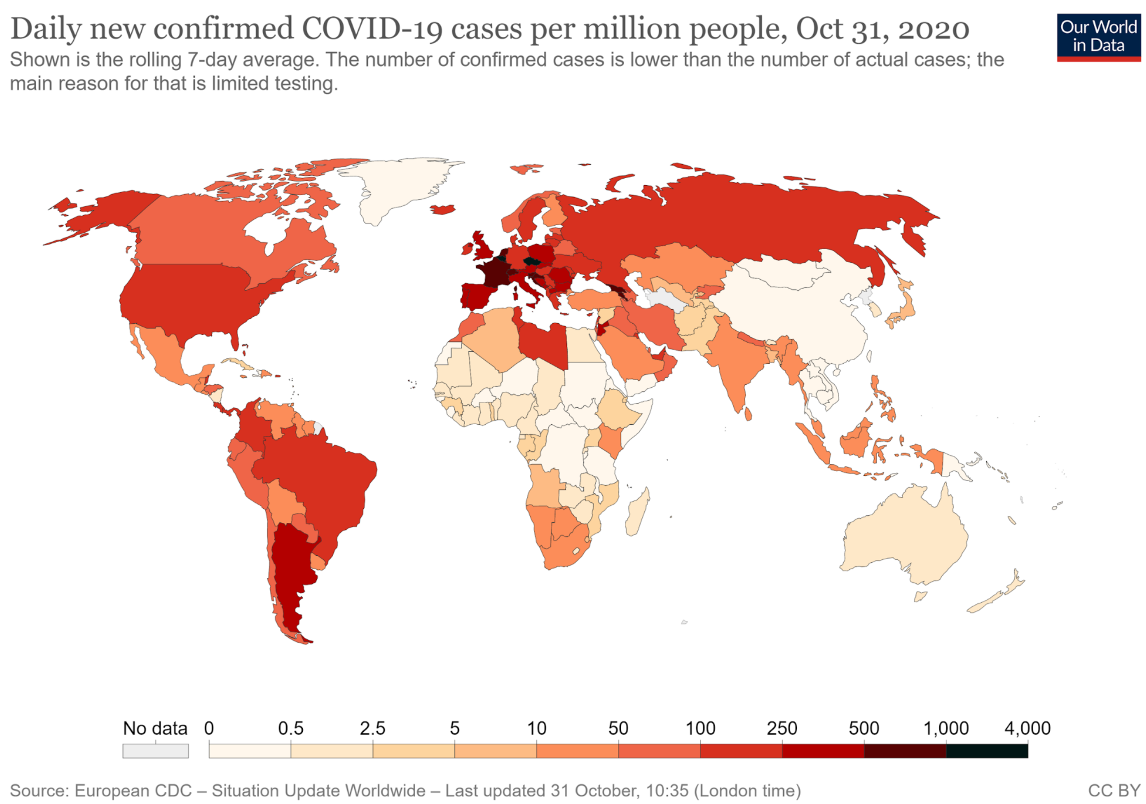

Figure 3.4: Daily new confirmed COVID-19 cases per million people, Oct 31, 2020

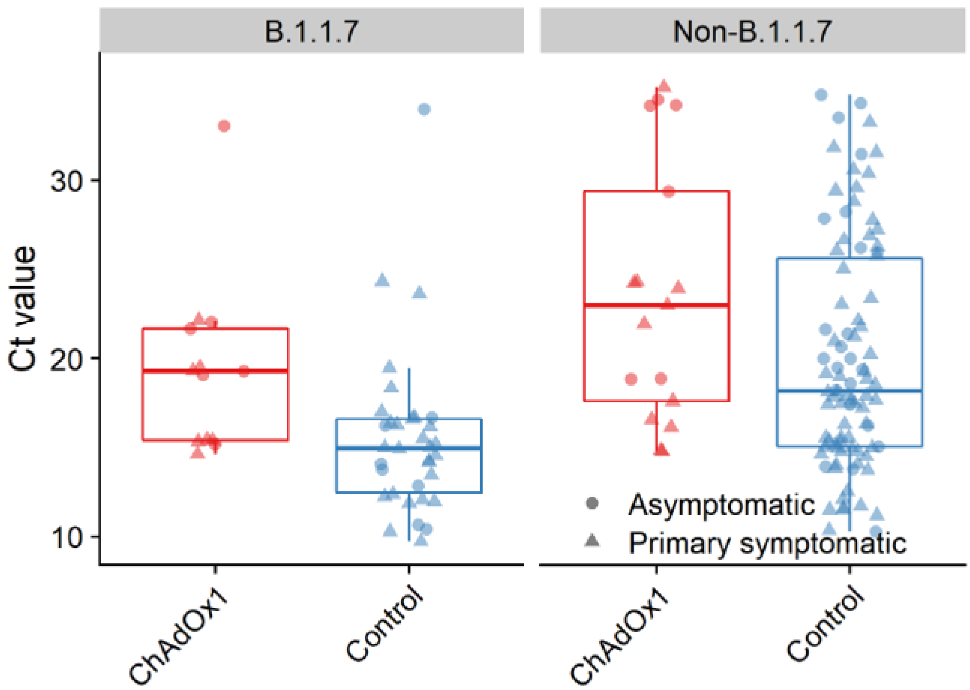

Figure 3.5: Box Plot

3.3 Variables and Indicators of Interest

Since the beginning of the COVID-19 pandemic, various models have been developed in an attempt to accurately describe and predict the evolution of the virus. These models include different variables, some of which were discovered and calculated by means of scientific investigations. As the model to be used for the dashboard is inspired by the flow network models presented by López and Cukić [34, 35], as well as Kucharski et al [36], the team would use not only variables commonly used in existing databases for COVID-19 surveillance, but also those that are related to the model and can be derived mathematically from the commonly found variables, such as accumulated cases and deaths, as well as indicators such as the transmission and danger indices and the accumulated incidence, which will be discussed shortly.

3.3.1 Variables

Daily Positive Cases

This pertains to the number of positive COVID-19 cases recorded in a 24-hour interval. Furthermore, this variable can be classified according to the the type of diagnostic test used to confirm the positive cases [37]:

- **Antigen Tests:** Using samples collected through nasal or throat swabs, antigen tests are used to detect a protein that makes up part of the coronavirus and are useful for identifying cases that are approaching states of peak infection. Compared to the other types of tests, antigen tests are cheaper and faster, but are also less accurate, being able to produce false positives and false negatives.

- **Molecular/PCR Tests:** Also using nasal and throat swabs, molecular tests focus on identifying the genetic makeup of the coronavirus using different methods, with the most famous one being the polymerase chain reaction (PCR). While some molecular tests have shown false negative results for up to 20% of the time, this type of test remains more accurate in identifying positive cases than antigen tests.

Deaths

These are the number of deaths caused by the coronavirus disease. We assume these figures are reported by the hospitals daily.

Recoveries

These are patients who have been discharged from the hospital or those that are home-quarantined and have been considered recovered by their attending physicians. This variable will remain useful when taking into account the total number of recoveries reported.

Accumulated Positive Cases

As explained in its name, this variable gathers the total number of positive cases reported so far. This is calculated by retrieving the sum of daily positive case counts reported from start to present.

Active Cases

These are positive cases that are neither counted as a death nor a recovery, meaning that these are ongoing cases of COVID-19.

As a person infected with COVID-19 may experience mild to severe symptoms, it is recommended that not all active COVID-19 cases be considered as hospitalizations, but rather be broken down into the following categories:

- **Home-Quarantined:** Patients who experience mild symptoms are not recommended to be hospitalized (to allow hospitals to accommodate cases of higher severity), so they are quarantined at home instead. They are usually monitored by their respective physicians through remote means such as examinations over the phone.

- **Hospitalized:** Patients who experience severe symptoms and are in need of hospital care fall under this category. These patients stay in this state until they recover and are deemed safe to discharge or until their condition worsens into a critical state.

- **Critical:** Critical cases involve patients who are experiencing extreme symptoms and require immediate attention by being admitted into intensive care units (ICUs).

Number of Tests Done

This consists of the accumulated number of tests performed in a certain area. This will include both tests that returned positive results as well as negative ones. To produce more comprehensible numbers, this number may be presented as a ratio of tests per 1M/100,000 inhabitants.

Number of Cases per 1M/100,000 Inhabitants

Comparing the number of cases (be it active cases, number of deaths, or number of recoveries) may be difficult to understand for the average user. This variable aims to project these numbers as a ratio of number of cases to number of inhabitants, providing a density of cases within a part of the population.

Population-Weighted Density (PWD)

PWD is one of the variables that explain the spread of the COVID-19 pandemic [38]. As this variable describes the density at which an average person lives, it can be observed that this influences the rate of deaths caused by the coronavirus. Increased social distancing results in a reduction in PWD (albeit temporarily), indicating effectiveness in measures taken by a unit of government.

3.3.2 Indicators

Transmission Index (R_0)

The transmission index [34] represents the power of virus transmission and can be inferred by calculating the ratio between the number of infections recorded at a certain day t and the number of infections recorded 14 days before t, as seen in Equation 3.2. A value of 1 or greater indicates danger in the transmission of the virus while a value of less than 1 indicates no danger.

$$\widehat{R_0}(t) = \frac{Inf(t)}{Inf(t-14)} \tag{3.2}$$

Cumulative Incidence per 100,000 Inhabitants (Inc)

As seen in Equation 3.3, this index determines the cumulative number of positive COVID-19 cases in the past 14 days per 100,000 inhabitants [34, 35].

As this variable takes into consideration the population of the country, it is useful for making fairer comparisons between the situations of different countries with regards to the pandemic. However, its disadvantage comes from the fact that this indicator only considers the number of infected people.

$$Inc^i = (\sum_{j=1}^{14} C_{i-j}) * 100000/P, \forall i \geq 15 \tag{3.3}$$

Danger Index (DI)

The danger index [34] is used to determine the level of danger a flow network presents. Using Equation 3.4, the DI obtained indicates little to no danger if its value is equal to or less than 0, while a higher DI indicates a more dangerous flow network.

While both R0 and Inc only considers the number of infected cases, DI has an advantage such that it considers the number of positive cases, the number of deaths, and the number of recoveries, each of which can be weighed to provide more adequate values depending on the situation. However, compared to the Inc indicator, the population of the country is not taken into consideration.

$$DI(t) = Inf(t) + F(t) - Rec(t) = \sum_{x \in V} f_{x,INF}(t) + \sum_{x \in V} f_{x,F}(t) - \sum_{x \in V} f_{x,R}(t) \tag{3.4}$$

3.3.3 Smoothed Variables and Indicators

In obtaining data on COVID-19, there are times when large peaks are followed by sudden drops in value, possibly resulting into noisy data [39]. To counter this, data smoothing is applied by using an algorithm to remove noise from a dataset and allow patterns to stand out more clearly.

In this case, the variables and indicators mentioned previously were smoothed by calculating the 7-day moving averages, found in Equation 3.5.

$$\forall i \geq 4, mm(x_i) = \frac{1}{7}\sum_{k=i-3}^{i+3}(x_i) \tag{3.5}$$

Chapter 4

Data Preparation

After having a list of possible sources of data for the project, the data would need to be prepared for use in the dashboard. This preparation process includes operations such as unifying the format of data, cleaning out invalid data, and placing the results together in a file that the dashboard can access and use to present the data visually.

In this part of the project, a collaboration was initiated with another group, whose goal is to search and analyze available information on the COVID-19 pandemic. They would ultimately be in charge of providing the dataset to be used by the dashboard.

4.1 Databases Used as Bases for Database Creation

As previously explained, the partner group was in charge of obtaining available information on the situations of each country. While most information can be found on the website of each country's health agency (which was the case for major countries with considerably reliably governments such as Spain, Japan, and Russia), the following repositories are highlighted due to the fact that they contain information that has already been compiled and is constantly updated for convenience in data gathering:

- https://github.com/datasets/covid-19
- https://github.com/cssegisanddata/covid-19

4.2 Dataset Creation for Project Use

4.2.1 First Dataset

In order to initialize the the development of the project, another dataset must be temporarily used for testing the functionality of the dashboard. For this purpose, a dataset provided by the project directors that contained data on the pandemic situation in 14 countries was used.

The following columns are found in this dataset:

- `COUNTRY`: country in which the data was recorded
- `FECHA`: date of recording
- `CONTAGIADOS`: daily number of positive cases of COVID-19
- `FALLECIDOS`: daily number of deaths caused by COVID-19
- `HOSPITALIZADOS`: daily number of hospitalizations due to COVID-19
- `UCIs`: daily number of COVID-19 patients admitted into ICUs

4.2.2 Second Dataset

Created by the partner group, this dataset is separated into two files: one containing data pertaining to the United States alone and another containing data for the rest of the world. Compared to the previous dataset, this contains information on more countries, as well as more columns that consider variables that were previously not present, such as the number of active cases and the danger index [34, 35]. Another difference is that, when applicable, data is also organized by the provinces of the involved countries.

This dataset consists of the following columns:

- `Country_Region`: country/region in which the data was recorded
- `Province`: province of the country in which the data was recorded, if available
- `Lat`: latitudinal coordinate of the area's location
- `Long`: longitudinal coordinate of the area's location

- `Last_Update1`: date of recording
- `Confirmed`: cumulative number of confirmed positive COVID-19 cases
- `Deaths`: cumulative number of deaths caused by COVID-19
- `Recovered`: cumulative number of COVID-19 recoveries
- `Active`: cumulative number of active COVID-19 cases
- `Daily_Confirmed`: daily number of confirmed positive COVID-19 cases
- `Daily_Deaths`: daily number of deaths caused by COVID-19
- `Daily_Recovered`: daily number of COVID-19 recoveries
- `Daily_Active`: daily number of active COVID-19 cases
- `IP`: danger index
- `R0`: transmission index
- `Daily_ConfirmedMM`: moving average of the confirmed positive COVID-19 cases
- `Daily_MuertosMM`: moving average of the deaths cause by COVID-19
- `Daily_RecuperadosMM`: moving average of the COVID-19 recoveries
- `Daily_ActivosMM`: moving average of the active COVID-19 cases
- `IPMM`: moving average of the danger index
- `R0MM`: moving average of the transmission index

Meanwhile, the dataset pertaining to the United States contained additional columns:

- `Daily_Tested`: daily number of people who took the COVID-19 test
- `Daily_Hospitalized`: daily number of people admitted to the hospital due to COVID-19
- `People_Tested`: cumulative number of people who took the COVID-19 test

- People_Hospitalized: cumulative number of people admitted to the hospital due to COVID-19

4.2.3 Final Dataset

Upon studying the previous dataset, the following issues were encountered:

- The dataset contained columns deemed necessary for the dashboard implementation (i.e. Lat, Long).
- The files containing the dataset has a total size equal to 50MB. This was a potential problem with regards to the limit in the size of the application to be uploaded in the server.
- Some of the data contain names of countries that do not match with those used in the dashboard, particularly its map functionality.
- The INC indicator, which was a variable of interest, was not included.
- The calculation of the moving average was done by region, this was a problem as the dashboard being implemented was organized by country.

The team had created an R script named dataTransform.R to solve these issues. This script parses the data from both dataset files and performs the necessary transformations (i.e. column renaming, data merging, calculation of extra variables and smoothing variables and indicators using moving averages), creating the final dataset, datasetCODA.csv. The structure of this dataset is explained in the Chapter 6.

4.3 Dataset Quality Evaluation

Creating the dataset for the dashboard involved multiple tasks that required not only time and effort, but also the required knowledge to perform such tasks and in order to redact the dataset efficiently. Not only does the dataset have to be well-defined, it also has to be reliable. One important task in this process is the evaluation of the quality of these datasets. It gives the idea of how valid the information inside a dataset can be. This can be estimated by some qualitative factors present in the data.

An important aspect about dashboards is that part of their quality and effectiveness is derived from the dataset it uses to present information. As such, it is important to evaluate at least the final dataset used.

Working with this dataset, the following observations, some of which led to modifications in the project, were made:

- Some countries do not have updated data.
- Some values, like recordings of negative deaths, do not make sense and can be attributed to errors in recording data.
- Most countries contain outlier data due to inconsistencies such as incorporating data recorded in a certain day into that of a later date.
- Data from countries such as North Korea, Antarctica, Western Sahara and Turkmenistan are unavailable and therefore could not be shown in the dashboard.

Chapter 5

Project Development

5.1 Project Planning

5.1.1 Development Methodology

Scrum [40] is an iterative and incremental agile methodology for developing software that places an emphasis on developing and inspecting software rather than on documentation. Compared to other software development methodologies, Scrum allows for adaptations to changing requirements. This is especially crucial to this project since the team was generally inexperienced with creating dashboard and, as such, the requirements were be expected to evolve over time. Figure 5.1 elaborates how the Scrum development process is realized.

Due to the fact that the requirements of the project may change during the development process, the Agile methodology Scrum was applied in developing the dashboard. This will also help ensure that the dashboard is of high quality in functionality and appearance, thanks to its iterative and incremental nature.

As the team only consists of three members and that the project was being done in parallel with other academic work, some modifications had to be made to the original methodology. The team would not apply the Scrum Master role and instead coordinate with each other throughout the duration of the sprints. Additionally, each sprint would last for 1-4 weeks (depending on the workload presented by the sprint) in consideration of the other academic requirements that the team members are completing and sprint meetings were held at the beginning and end of each sprint to track

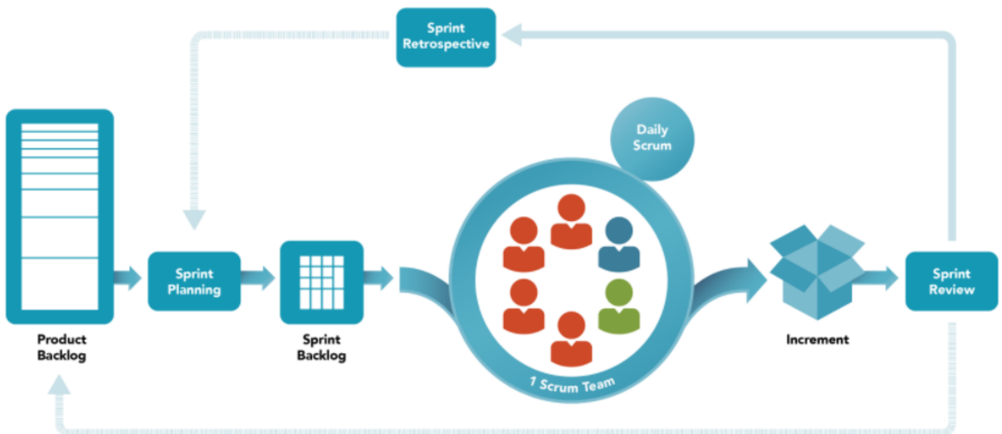

Figure 5.1: Scrum process diagram

progress.

It was also at this point that the team members decided that the dashboard would be an epidemiological data tracker of global scope made using Shiny. The dashboard would be developed mainly in the RStudio IDE and, to minimize costs, the free plan of Shinyapps.io [22] would be used for publishing the dashboard to the cloud, allowing it to be embedded in websites. As a consequence of the selected plan, the size of the application made must be at most 1 GB.

5.1.2 Google Sites

As it was also planned that the dashboard would be placed in a website to make it more accessible to the public, a hosting service was needed. Taking this into consideration, the online platform Google Sites [41] was selected, as this option facilitates the creation of websites with high quality designs with a low effort. Despite the design options being limited to different templates, Google Sites fulfills the functional requirements of hosting the online dashboard.

No.	Story	Priority
1	View a graph to analyze variables	10
2	Visualize the map view with colored countries to differentiate the severity of the situation of each country	9
3	Choose between map or graph to show in the dashboard	5
4	Choose to filter by a variable to visualize the severity of that desired variable in the Map View	7
5	Click on a country in the map view to see all the variables associated with that country in a table alongside the map	4
6	Choose to filter by country to visualize the situation of the number of cases of that country in a graph	7
7	Choose to filter by a variable to visualize that desired variable in a graph	7
8	Be able to choose more than one country filter to be able to compare situations of these countries in the graph view	6
9	Be able to choose more than one variable filter to be able to compare variables of different types in the graph view	6
10	Configure both the map and graph view to view data in function of time/date	10
11	When applicable, choose the type of graph used to represent the variables	6
12	Be able to see the global situation of the pandemic and summarized tables in both views (number of infections, deaths, etc.)	6
13	Be able to visualize the dashboard in a web page and hosted through a server	8
14	Be able to locate the dashboard's source of data for downloading	8

Table 5.1: Product backlog

5.2 Product Backlog

Based on the requirements that the final product should fulfill, a product backlog (see Table 5.1) was created. It should be noted that the functionalities listed in this backlog take the perspective of the end users. Each task was also assigned a priority on a scale of 1 to 10, with 1 being the lowest priority and 10 being the highest.

5.3 First Sprint

Sprint Planning

After establishing the product backlog, tasks that would be included in the sprint backlog in the first sprint (see Table 5.2) were selected. Some of the requirements or specifications chosen have been broken down into small tasks, which were divided to the members of the group.

Story	Task
1	A. Prepare the initial dataset B. Implement the functions to process dataset C. Implement Graph View UI D. Implement server function to visualize a line graph based on a chosen variable in function of time
2	A. Implement Map View UI B. Implement server function to visualize and colorize countries based on a variable
3	A. Design Map/Graph Tab UI B. Implement Switch Mode Tab in the UI
4	A. Map View - Implement Variable Select Box UI B. Implement Variable Select Box server function
5	A. Implement UI to view Country Table in the Map View B. Implement Update Country Table UI server function
6	A. Graph View - Country Select Box UI B. Implement Country Select Box server function
7	A. Implement server functions to fetch data by chosen variable B. Add these variables in the options in the Map and Graph View C. Graph View - Variable Select Box UI D. Implement Variable Select Box server function
8	A. Graph View - Single Variable/Country Mode Radio Button B. Graph View - Multiple Country Radio Button C. Implement Multiple Country Mode server function
9	A. Graph View - Multiple Variable Radio Button B. Implement Multiple Variable Mode server function
10	A. Graph View - Time Interval UI B. Implement Time Interval server functions

Table 5.2: First sprint backlog

Design

As part of the sprint backlog, the team initiated the design of the project. First, five initial designs (see Figure 5.2) were created individually by the team members. Taking into account the various feedback received about these initial designs, a final prototype design (see Figures 5.3, 5.4, 5.5, and 5.6) has been created based on the fourth initial design (Figure 5.2).

All prototype designs were created using Balsamiq Wireframes due to time constraints, as well as the team members' experience with the software.

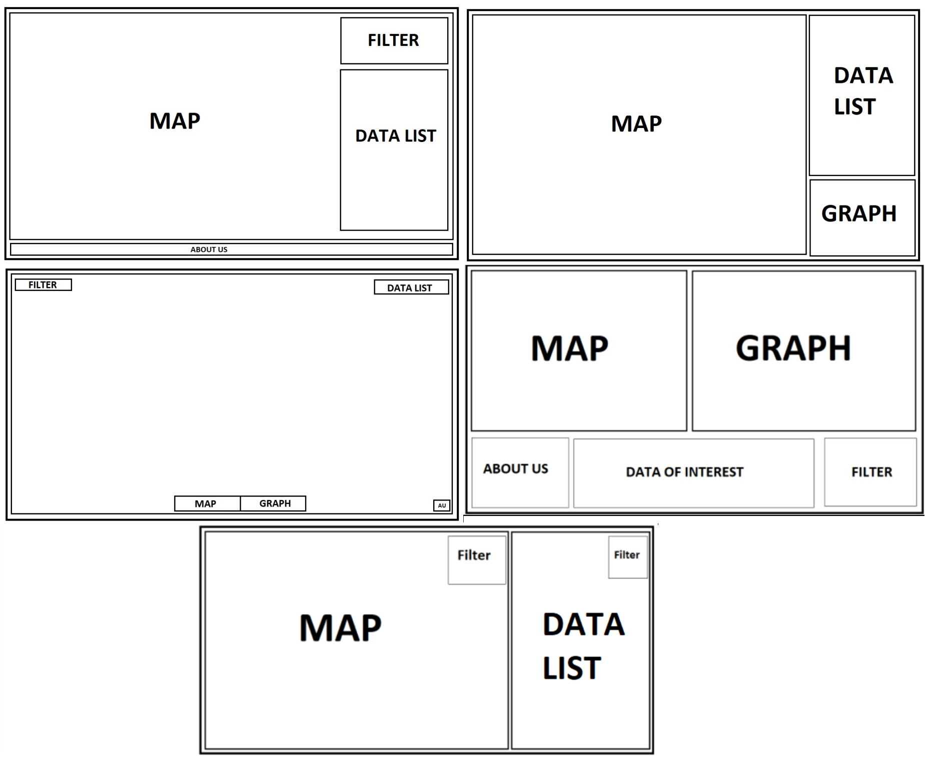

Figure 5.2: Initial designs (ordered from top to bottom, left to right)

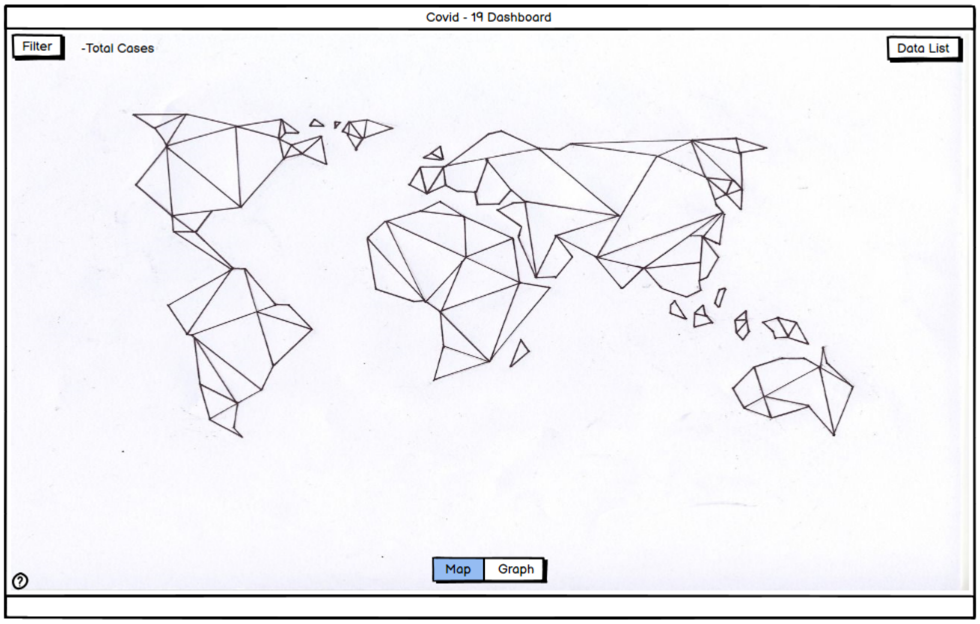

Figure 5.3: Final prototype design. Map View 1

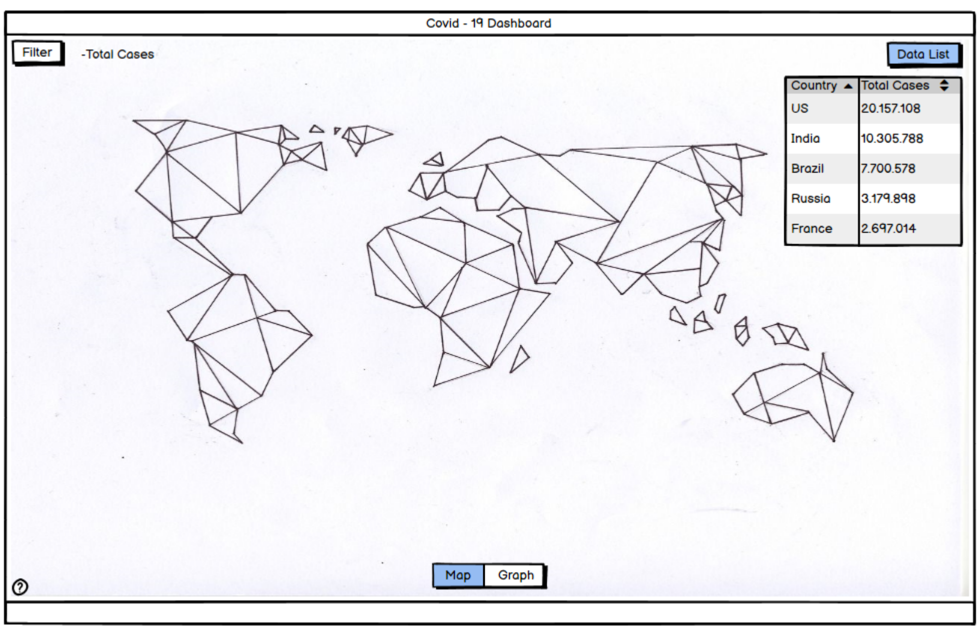

Figure 5.4: Final prototype design. Map View 2

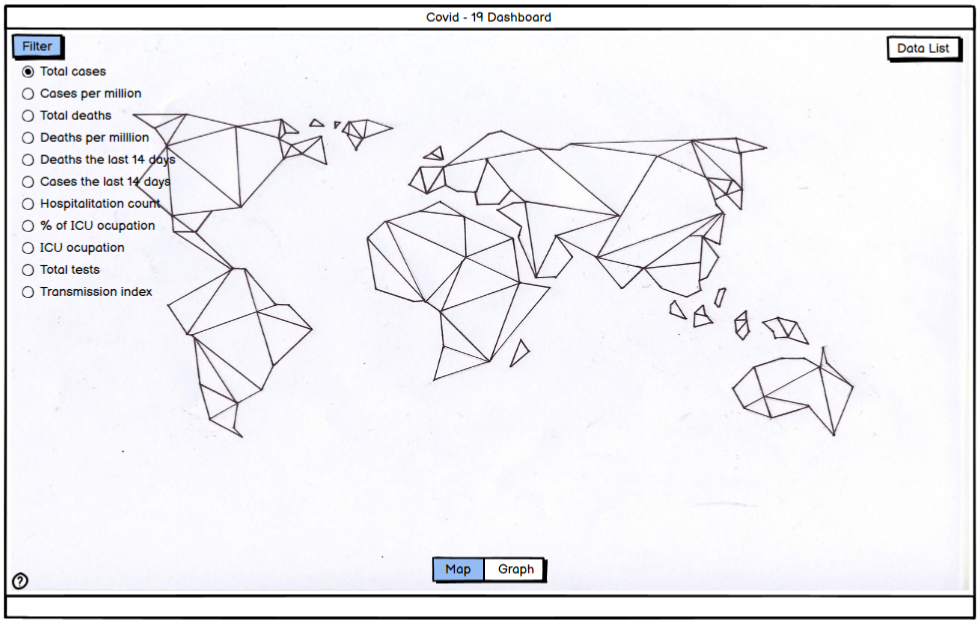

Figure 5.5: Final prototype design. Map View 3

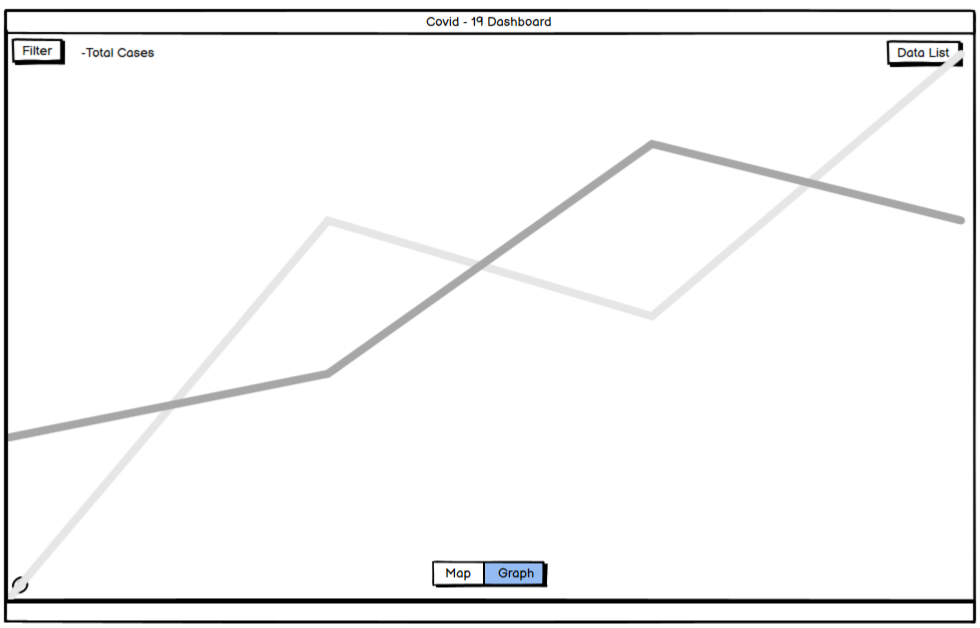

Figure 5.6: Final prototype design. Graph View

Improvements to Dashboard Implementation

As agreed in the prototype design, the dashboard was split into two screens/tabs, one for presenting users the graphs that show the evolution of the different variables in various countries (see Figure 5.7) and another one for displaying the cumulative numbers of each country in a map (see Figure 5.8). Additionally, the graph function allows users to see the evolution of the pandemic in multiple countries or multiple variables to allow comparisons.

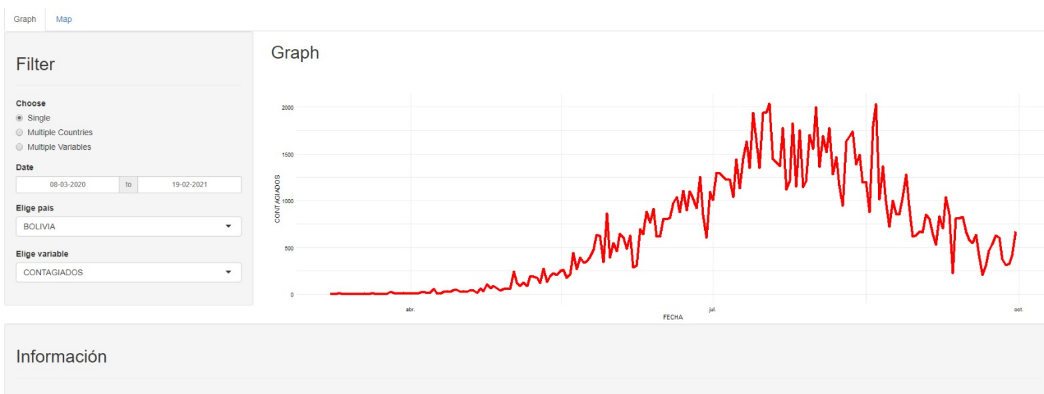

Figure 5.7: Sprint 1 result (graph view)

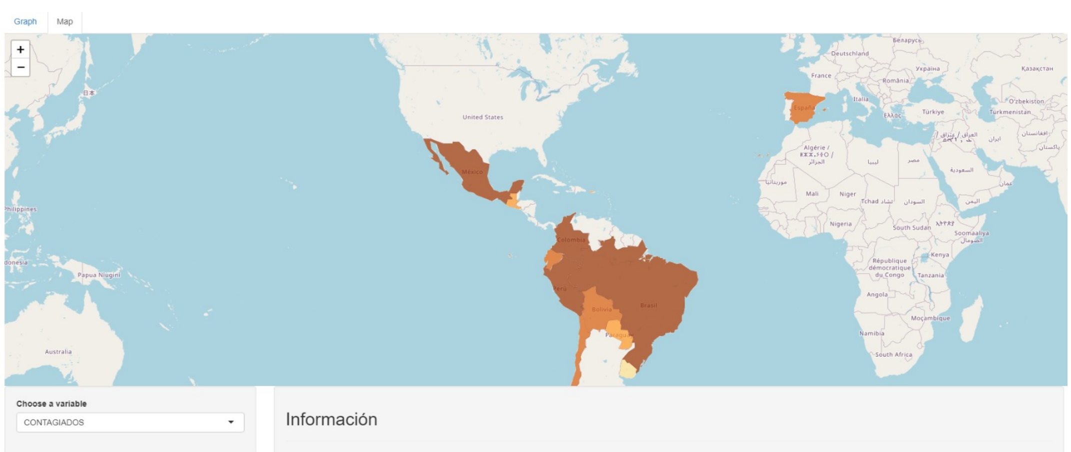

Figure 5.8: Sprint 1 result (map view)

Sprint Evaluation

At the end of the sprint, some of the tasks included in the backlog for this iteration were not implemented, such as the updating List View UI in the Map View. Instead, only an Information Panel UI that contained no data was displayed. A meeting was then held with the project director to provide updates on the progress of the project, as well as present the dashboard in its state after the sprint. Feedback was received from the director, which would then be taken into consideration upon beginning the next sprint. Some of these included changing the names of the countries in the map view to their English names, as they were originally presented as their respective names in their own languages, selecting more than just two countries or variables, including a legend for graphs, and presenting a summary of all data in the information panel by means of presenting, for example, the countries with the most accumulated number of positive cases.

5.4 Second Sprint

Sprint Planning

After conducting a meeting at the end of the first sprint, we have selected tasks to be included in the second sprint backlog as shown in Table 5.3. Also learning from our experience in the first sprint, some of the tasks have been slightly modified to accommodate the limitations presented by Shiny, such as difficulties in layering components of the dashboard UI. As previously done, the tasks have been divided into smaller ones to ensure manageability by the team members.

Design

As previously stated, the second sprint focuses on adjusting the implementation of the dashboard to the design limitations of Shiny and on refining the dashboard contents as much as possible. As such, the design prototypes have been refined, as seen in Figures 5.9 and 5.10.

Story	Task
1	A. Implement mouse hover function to show data at a specific point in the graph B. Implement legends in the graph C. Show the axis labels bigger D. Fix date axis scale to 1 week E. Fix date axis angle to 45 degrees F. Add controls to visualize graph (zoom in/out, pan, auto-scale) G. Relocate the panel for filters above the graph.
2	A. Set color division to percentages depending on the highest number of accumulated cases of the chosen variable B. Implement mouse hover function to show the country's name and the value of the variable chosen C. Change the map to show all countries in English
8	A. Limit to 5 countries to choose B. Remove Single Variable/Country Mode Radio Button and set default mode to Multiple Country
9	A. Limit to 5 variables to choose B. Establish a palette of colors to be used for every different line to be drawn in the graph.
10	A. Change UI to slider for both Map and Graph Views B. Put controls to modify both the start and end date in the Graph View
11	A. Investigate other graphs that could well describe some variables B. Implement a radio button UI to choose between some type of graphs C. Implement server functions to visualize selected graph
12	A. Implement a table showing 5 countries with the highest number of accumulated cases of the chosen variable (or the first variable chosen in multiple variables mode) in the Graph View B. Implement the same table in the Map View C. Implement another Information Panel UI in the Graph View that shows the total number of infected and death cases D. Implement server functions to calculate these variables E. Relocate the Information Panel and Filter Select Box UI in the Map View and place it over the map

Table 5.3: Second sprint backlog

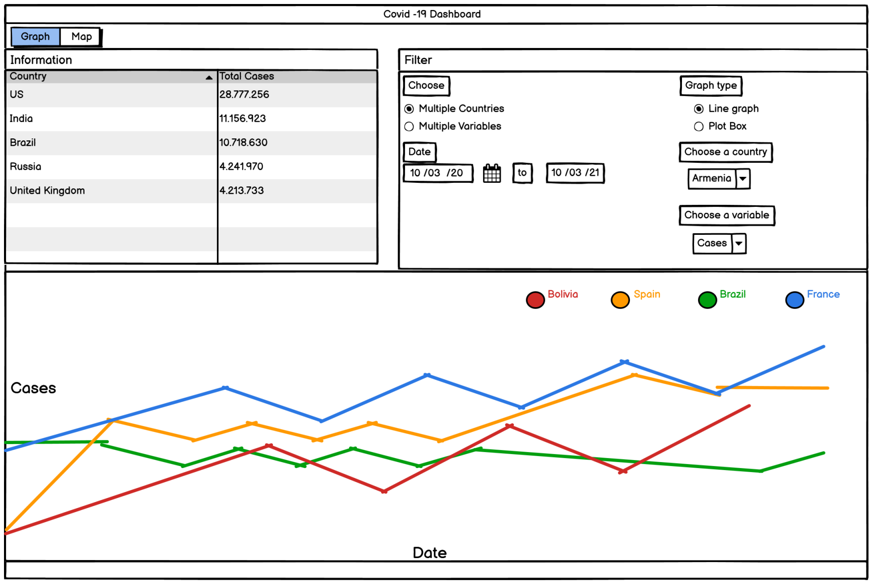

Figure 5.9: Revised prototype design (graph view)

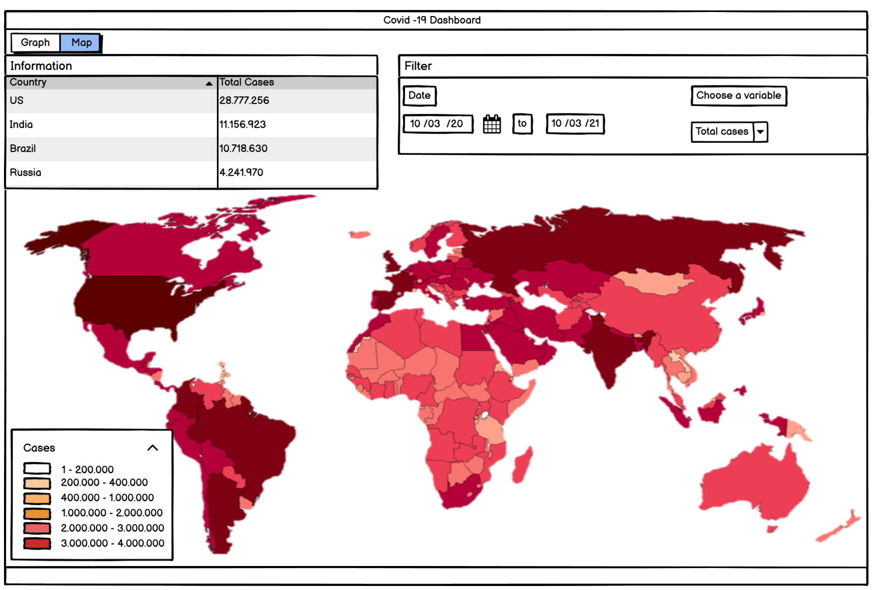

Figure 5.10: Revised prototype design (map view)

Improvements to Dashboard Implementation

One of the biggest changes introduced in the dashboard implementation of this sprint is the option to select more than 2 countries when selecting the "multiple countries" option. Users would now be allowed to choose up to 5 countries to be displayed in the graph. This limit was placed to allow users to compare different countries without possibly overcrowding the resulting graph with various lines that may become unreadable.

Another major change in the graph view of the dashboard is the inclusion of box plots. As discussed in Section 3.2, box plots are used to determine the deviation of the data and the numerical range they occupy, which users may find interesting.

The final major change in the graph view of the dashboard is that the information panel from the previous sprint was replaced with two smaller information panels that present summaries about the current pandemic situation globally. For example, users could to see the total number of cases and deaths recorded worldwide. They could also find a list of the top countries in terms of number of positive COVID-19 cases.

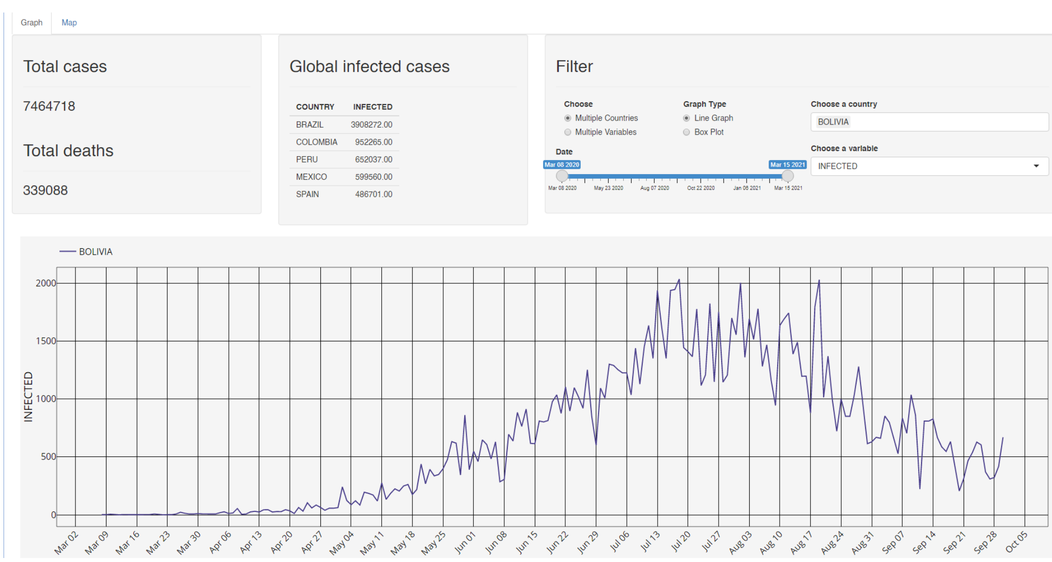

Figure 5.11: Sprint 2 result (line graph)

With respect to the map view of the dashboard, the country names were all renamed to their English names. This was done by replacing the map

Figure 5.12: Sprint 2 result (box plot)

used with the one by Esri, which also changed of the color palette to shades of blue. A legend was also placed to indicate the values a certain color in the map represents. The panel for filters together with the Information Panel were placed over the map to maximize the space to be occupied by the map itself. Finally, the information panel, implemented in the 1st iteration and that contained data, is repurposed to present the countries with the highest number of COVID-19 infections.

Sprint Evaluation

Feedback from the project director included changing the colors from shades of blue to shades of red, as this would appear more alarming and would give the impression that there is an ongoing threat, the inclusion of digit group separators in the numbers presented in the dashboard to improve readability, and changing the labels in the information panels of the Graph View (i.e. changing "Total cases" to "Global number of cases") to prevent confusion. Additionally, it would present a bigger convenience to users for the dashboard to have a function that captures the graph being shown, be it the entire graph or a portion of it.

Figure 5.13: Sprint 2 result (map view)

5.5 Third Sprint

Sprint Planning

The sprint began by selecting the tasks to be executed, elaborated in the sprint backlog found in Table 5.4

Story	Task
1	A. Process the new dataset to adapt it with the functions of the dashboard B. Add the accumulated variables in the select box UI C. Include Danger Index to the list of variables that can be shown as graphs
2	A. Change the color palette of the map to indicate intensity of numbers by shades of red
5	A. Implement UI to view Country Table in the Map View B. Implement Update Country Table UI server function C. Implement map click function to enable this feature
12	A. Improve the formatting of numbers by placing digit group separators

Table 5.4: Third sprint backlog

Design

Improving on the dashboard designs created in the previous sprint and considering the feedback received in the previous sprint evaluation, the presentation of the data and the labels found in the graph view of the dashboard were enhanced (see Figure 5.14). The colors used in the map were changed to shades of red and a panel that updates whenever the user clicks on a country in the map to see more information was included (see Figure 5.15).

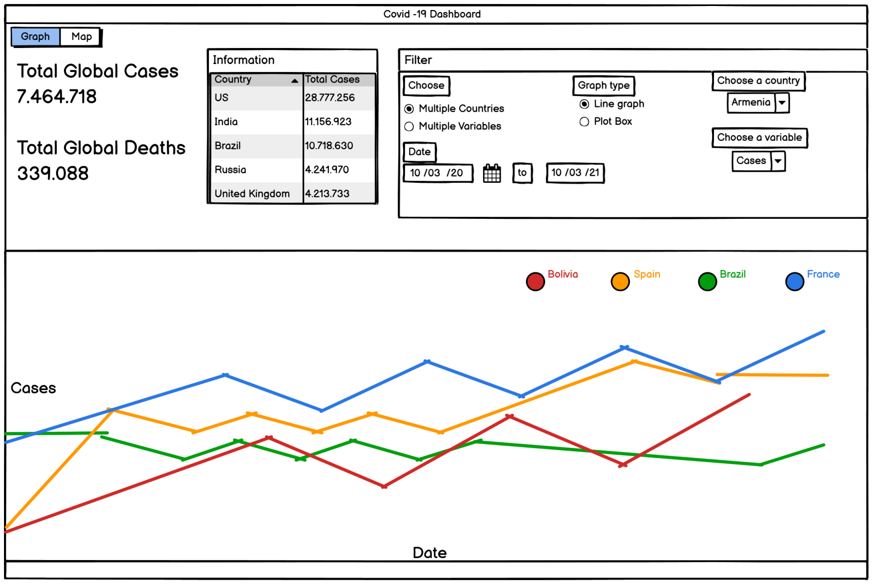

Figure 5.14: Revised prototype design (graph view)

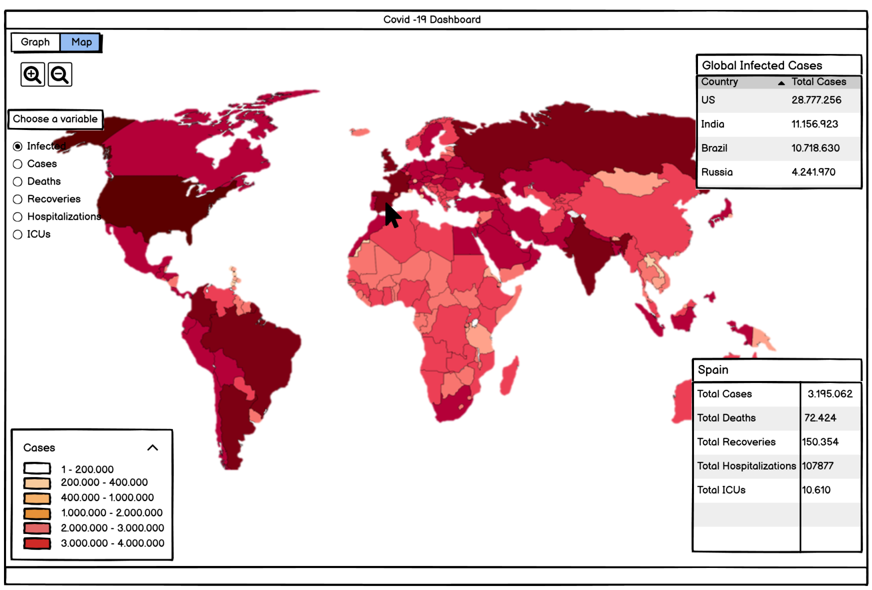

Figure 5.15: Revised prototype design (map view)

Improvements to Dashboard Implementation

The largest change implemented in this sprint is the use of a new dataset created by the partner group. Aside from having the variables we initially used as columns, this dataset also includes accumulated variables, as well as the Danger Index.

An obstacle encountered by the group in using the new dataset was that information was categorized not only by country, but by province as well. As the dashboard mainly presents data by countries, the data for each country was obtained by calculating the sum of all records of the country's provinces in a given day. Nevertheless, the dataset would be useful should the team decide to extend the dashboard by presenting information by provinces or regions as well.

Another interesting aspect about the new dataset is that it contains columns for the accumulated values of the variables over time for each country. This proved to be a convenient feature, as loading these values is a more cost-efficient option than calculating them in the dashboard's backend.

A new functionality was implemented in the map view of the dashboard where users may now click on a country to see information about it, such as its accumulated number of infections, deaths and recoveries (see Figure 5.17). The colors were also modified to warm colors to denote a sense of danger more clearly.

For ease in future usage and extensions of the project, the source code was also cleaned and made more presentable and understandable.

Additionally, the dashboard was deployed to Shinyapps.io and was embedded in a Google Sites website to test its functionality in other websites. This can be seen in the architectural structure of the dashboard in Figure 5.24.

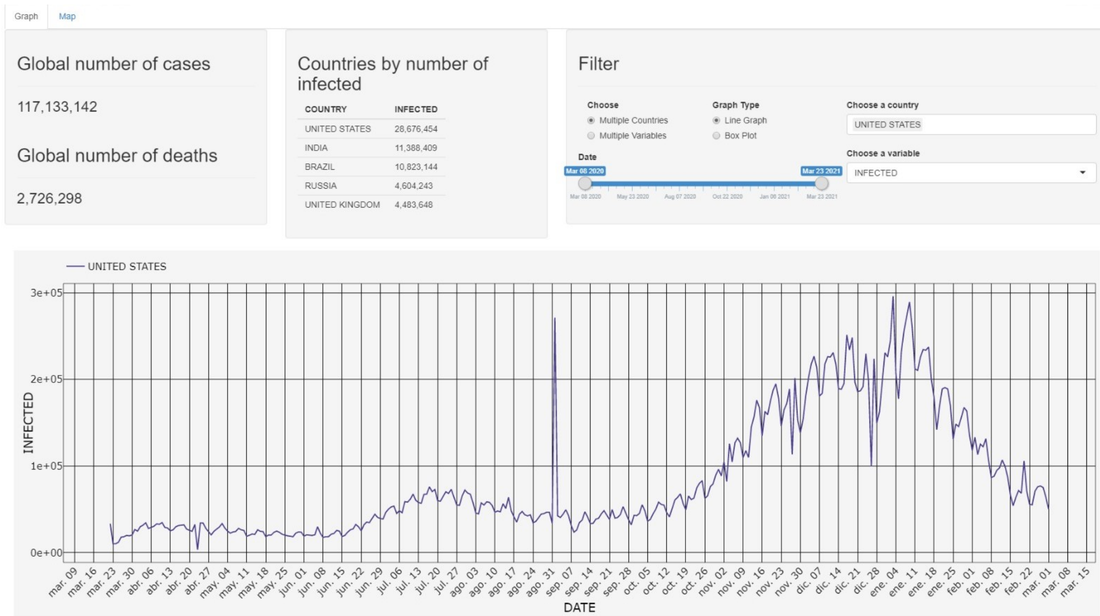

Figure 5.16: Sprint 3 result (graph view)

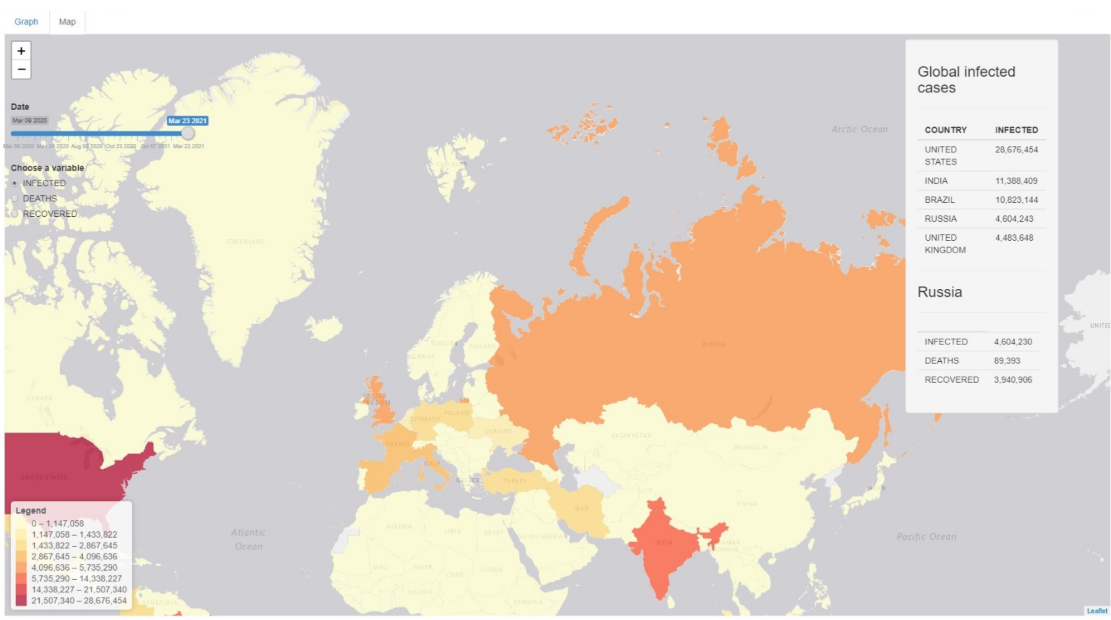

Figure 5.17: Sprint 3 result (map view)

Sprint Evaluation

In the sprint evaluation, it was found that most functional requirements have already been met by the dashboard though some minor improvements can be made, such as reordering the panels in the map view so that more emphasis is given on the data of a given country. It was also suggested that the performance should be improved, if possible.

Additionally, since the dashboard was placed onto a website at this point, work on the design of the website was possible. This design would have to include useful information such as details on the work done, a simple user guide, and a link for downloading the data used.

Finally, it was found that the dataset did not contain some expected variables, such as R0 and Inc. Therefore, this would need to be resolved in the next sprint.

5.6 Fourth Sprint

Sprint Planning

After a meeting at the end of the third sprint, tasks that were included in the fourth sprint backlog were selected (see Table 5.5), taking into account the requirements of the application and the suggestions given. One of the important modifications the team was required to do was to calculate the Inc indicator that was not included in the dataset given by the other group. The team also had to transform the given dataset into a more efficient one as explained in Section 4.2.3.

As this sprint focuses on refactoring code and refining the visual aspects of both the dashboard and the website, the changes to be done were expected to be minimal.

Story	Task
1	A. Include R0 and Inc to the list of variables that can be shown as graphs B. Download a reliable dataset containing the world population C. Implement server functions to calculate R0, Inc, and smoothed variables and indicators using 7-day moving average. D. Create a script to create a new dataset from the provided one and perform necessary transformations (rename countries, rename columns, remove unwanted data) and include these variables E. Draw the points of the graph together with the line graph.
2	A. Set the color of the countries with 0 value same as if the data was (NA) B. Swap the positions of global data table and country data on the map C. Include the variables Danger Index, R0 and Inc
4	A. Change radio buttons to select box input B. Separate accumulated variables from non-accumulated variables and add labels
13	A. Design the web page dedicated for the project B. Configure the dashboard's HTML page design to fit the designated window on the web page C. Create CSS file and improve the dashboard's accessibility and usability D. Upload the dashboard to the server E. Develop the web page and integrate the hosted dashboard F. Formulate the user manual for the dashboard and add it on the web page
14	A. Upload the datasets together with the script onto the designated repository and include the link in the web page

Table 5.5: Fourth sprint backlog

Design

The fourth sprint focuses on the finer details of design on both the dashboard and the website. With this in consideration, the design of the dashboard remains mostly the same. The only significant change in the dashboard design for this iteration is that the position of the total and particular tables are swapped in the map view (see Figures 5.18 and 5.19).

As for the design of the website (see Figures 5.20), prototypes for each page (namely, the Home page, the About Us page, and the About This Project page) were created. The Home page would include the dashboard, the About Us page would include information about the people who created the dashboard, and the About This Project page would include information on how to use the dashboard, the data used and how to download them, the objectives of the project, and such.

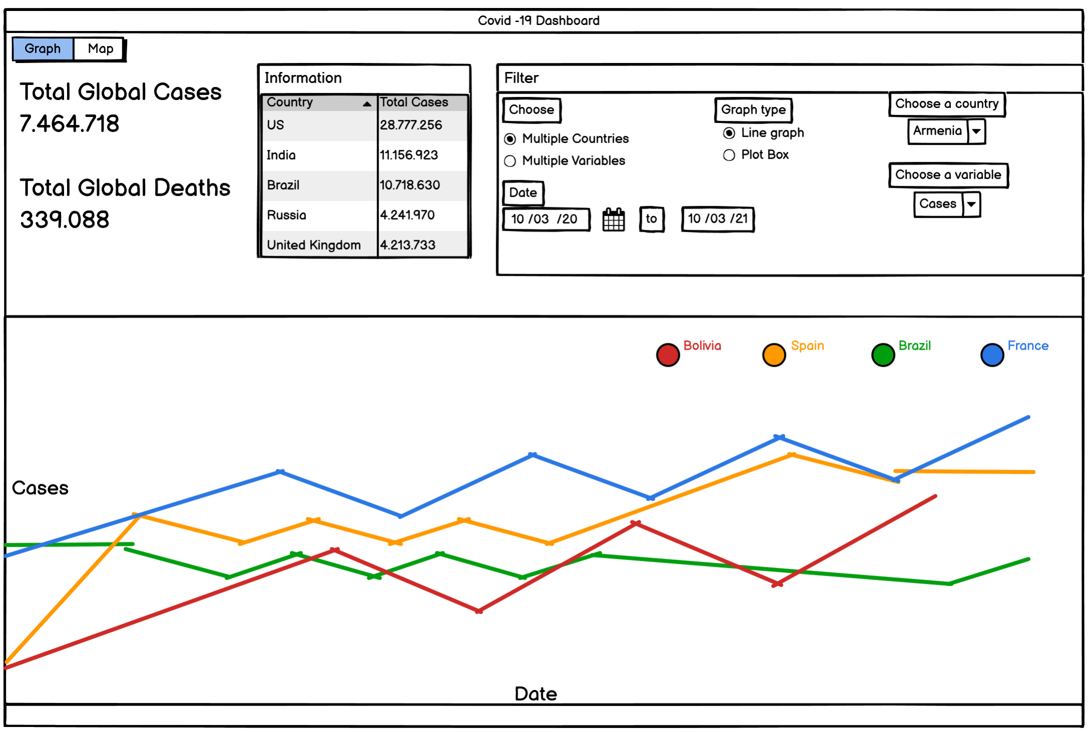

Figure 5.18: Dashboard design (graph view)

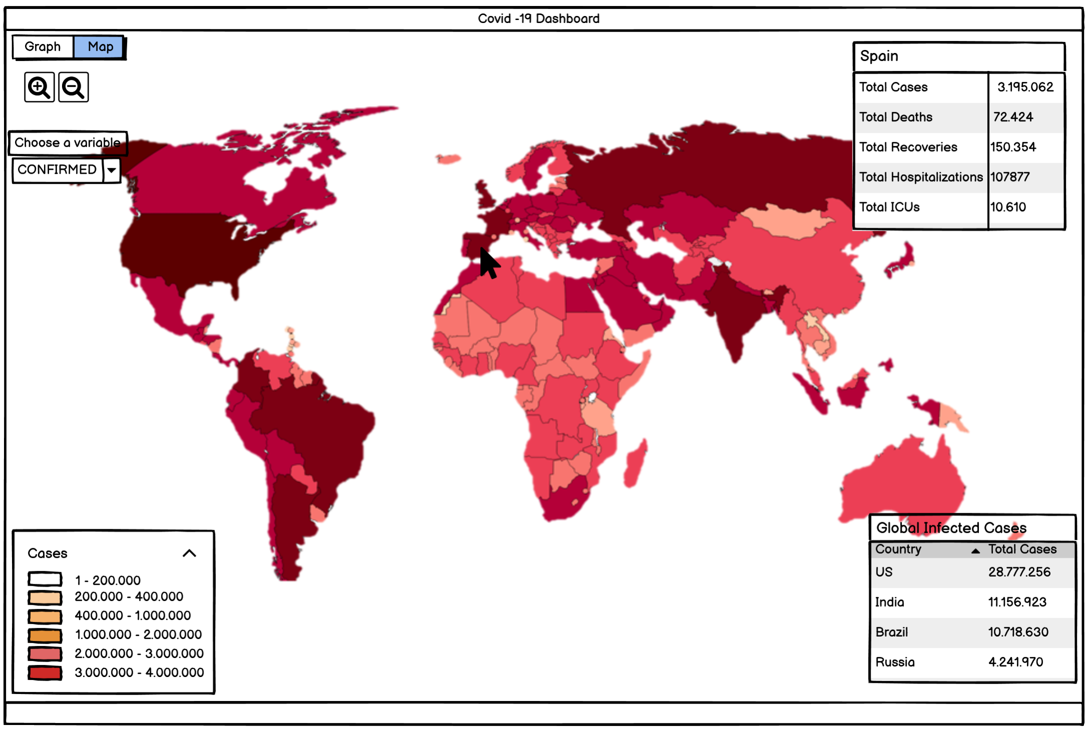

Figure 5.19: Dashboard design (map view)

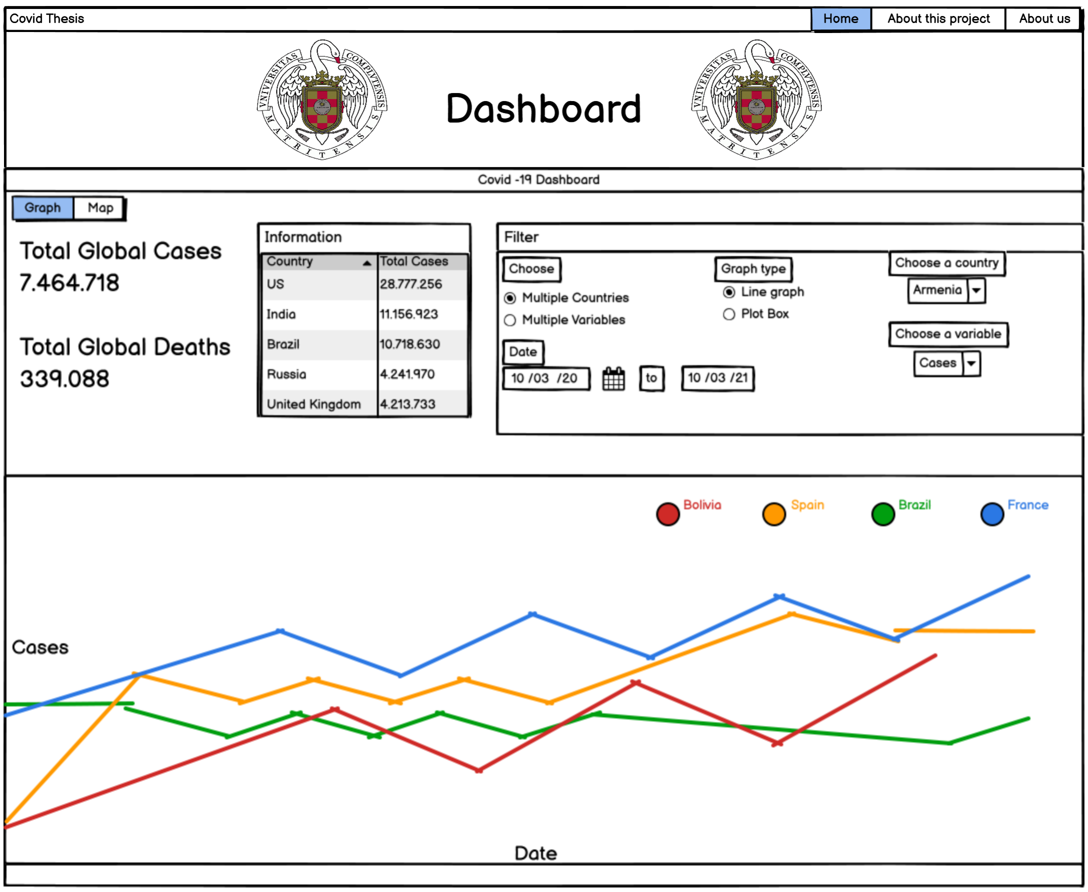

Figure 5.20: Website design (home view)

Improvements to Dashboard Implementation

The most significant implementation in this sprint is the creation of the website in Google Sites. In the website, both the completely developed dashboard and additional information on both the project itself and the team that made the dashboard were placed. By creating and using a CSS file `style.css` to indicate the style adjustments and describe how HTML elements will be displayed in the web page (margins, paddings, etc.), the design of the website was made such that it allowed the dashboard to fit within the allocated space when integrated. The CSS file is included along with the source code.

Another important change is that the source code was refactored, resulting in a more efficient output of data by the dashboard. The memory usage of the dashboard is also kept below 1 GB, allowing the team to continue using the free plan of Shinyapps.io.

Other changes made in this iteration include the addition of the Danger Index, R0 and Inc variables in the graph view of the dashboard. The values under the Danger Index, R0, Inc columns were calculated using Equations 3.5, 3.2, 3.3 found in Chapter 3. In addition, the World Populations Prospects 2019 dataset published by the United Nations [42] was utilized to obtain the populations for each country, which were used to calculate Inc. Some minor design changes, such as the swapping of total and partial tables in the map view of the dashboard and the addition of data points in the plot to allow further visualization of the graph, were also made.

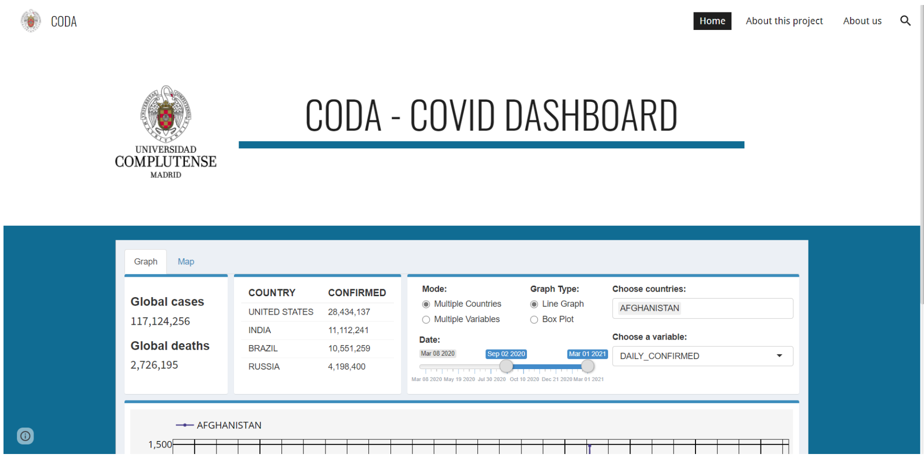

Figure 5.21: Website implementation (home view)

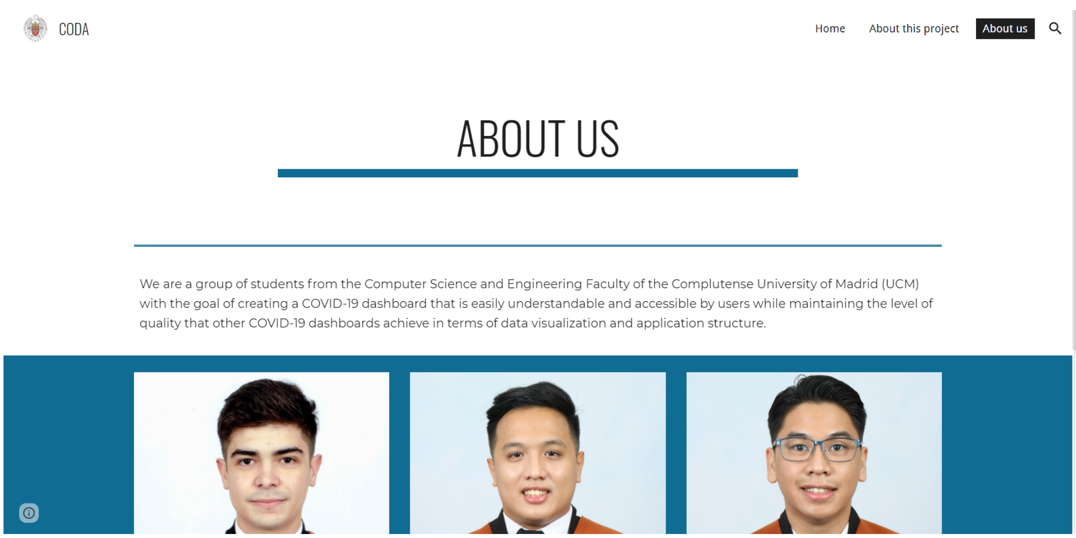

Figure 5.22: Website implementation(about us view)

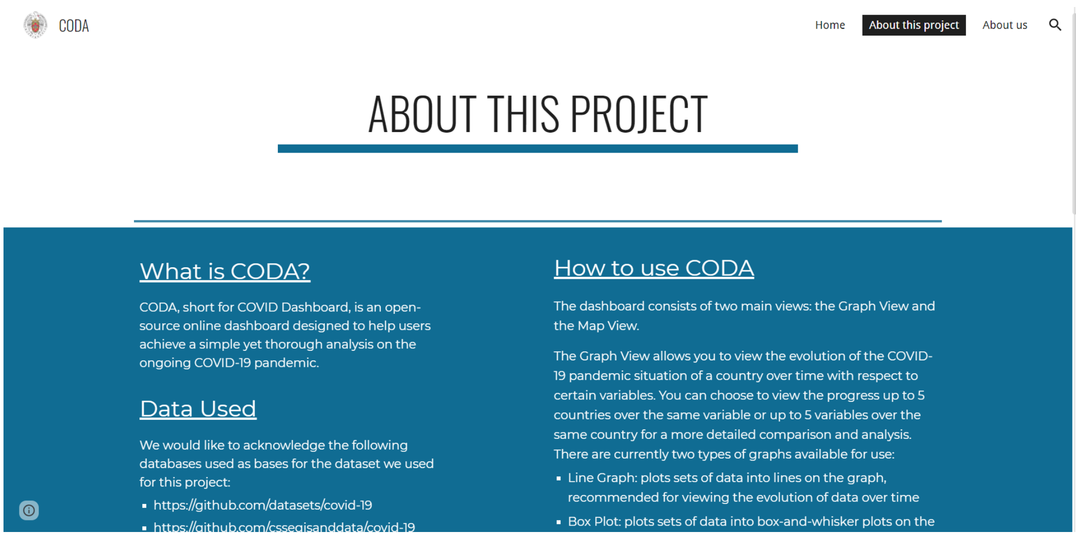

Figure 5.23: Website implementation (about this project view)

Sprint Evaluation

In this sprint evaluation, the web page's design was well-received. The latest developments done in the dashboard's implementation and its design were all accepted.

Minor adjustments, like correcting grammatical errors in labels in the dashboard and changing the radio buttons to a drop-down list for the variables in the map, would have to be made. Some additional information, such as the user manual, were also required in the About This Project page of the website. These minor corrections were then made immediately and were included in the sprint backlog of this iteration.

With all criteria listed in the Product Backlog fulfilled and evaluated, the development phase of the dashboard was concluded and deployed for assessment.

5.7 RStudio Libraries Used

Aside from the `base` RStudio library, which is used whenever the RStudio IDE is launched, the following libraries were utilized in the development of the dashboard:

- `shiny`: R framework used to build the application's UI and server interactions
- `shinydashboard`: used to design the structure of the panels of the application
- `dplyr`: used for parsing, processing and creating data frames in the Rstudio IDE
- `leaflet`: used for creating and visualizing the map
- `ggplot2`: used for creating and visualizing the graph
- `rgdal`: used for parsing a shapefile which is then needed to create and draw polygon layers on the map
- `reshape2`: used for merging data frames more effectively than `dplyr`

- `htmltools`: used for manipulating HTML elements such as the hover tool on the map
- `plotly`: used for the additional features of the graph such as the toolbar and the hover function

In summary, the dashboard contains the architecture seen in Figure 5.24.

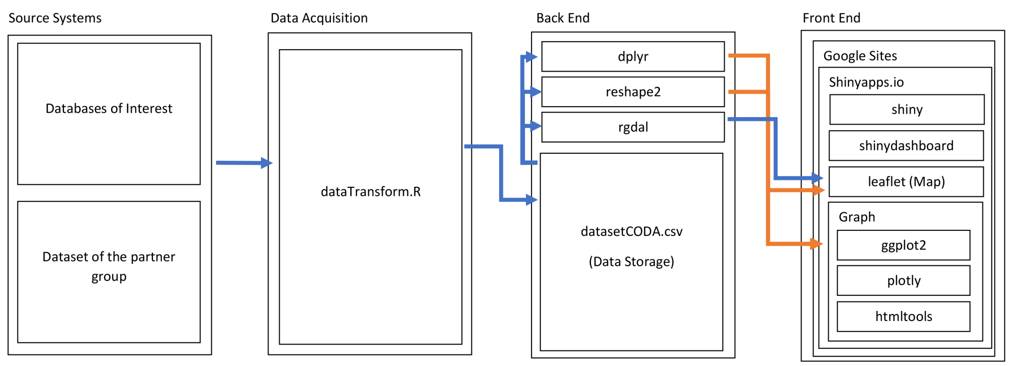

Figure 5.24: Dashboard architecture

5.8 Scrum Sprint Schedule

Figure 5.25 presents the dates in which each sprint was executed. The duration of each sprint was approximately two to three weeks.

Although the first sprint, that lasted from January 11, 2021 to February 19, 2021, seemed extensive, the effective development time on the project was 2 weeks, as this period coincided with the completion of numerous academic requirements.

Month	Monday	Tuesday	Wednesday	Thurday	Friday	Saturday	Sunday
January	4	5	6	7	8	9	10
	11	12	13	14	15	16	17
	18	19	20	21	22	23	24
	25	26	27	28	29	30	31
February	1	2	3	4	5	6	7
	8	9	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
March	1	2	3	4	5	6	7
	8	9	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30	31	1	2	3	4
April	5	6	7	8	9	10	11
	12	13	14	15	16	17	18

	First Sprint		Third Sprint
	Second Sprint		Fourth Sprint

Figure 5.25: Realized sprint dates

Chapter 6

Results

In this chapter, the results of the development of the dashboard and the corresponding web page are presented.

To properly interpret these results, some assumptions must be made. One assumption is that the data provided by each country is updated almost daily. Another assumption is that the data truthfully reflects the situation of the pandemic for each country. The final assumption is that the values recorded for each day do not contain errors that drastically affect previous data. However, in some periods of time, some countries have provided data that violate these assumptions, as they were published in such manner.

In the first section, the final dataset is explained, along with its structure. The second section presents the completed dashboard and website, along with a short guide that explains their usage. To further demonstrate the use of the dashboard and the website, the third section presents an example analysis that compares the situation of the pandemic in Spain and in the Philippines. The final section of this chapter provides a guide on integrating the dashboard to other websites.

6.1 Final Dataset

The final dataset created for the use of the dashboard was also made available in a public Google Drive folder, found at:

- `https://tinyurl.com/CODAdataset`.

This dataset contains the following columns, each of which may be used for further analyses:

- `COUNTRY`: country/region in which the data was recorded
- `DATE`: date of recording
- `DAILY_CONFIRMED`: daily number of confirmed positive COVID-19 cases
- `DAILY_DEATHS`: daily number of deaths caused by COVID-19
- `DAILY_RECOVERED`: daily number of COVID-19 recoveries
- `ACTIVE`: cumulative number of active COVID-19 cases
- `CUMULATIVE_CONFIRMED`: cumulative number of confirmed positive COVID-19 cases
- `CUMULATIVE_DEATHS`: cumulative number of deaths caused by COVID-19
- `CUMULATIVE_RECOVERED`: cumulative number of COVID-19 recoveries
- `DANGER_INDEX`: danger index as of the date of recording
- `R0`: transmission index as of the date of recording
- `INC`: cumulative incidence per 100,000 habitants as of the date of recording
- `SMOOTH_*`: smoothed variable/indicator using moving average

Additionally, the Inc values of Spain in this dataset (see Figure 6.1) was compared to the graph created using the official published data of the Spanish Ministry of Health [43] (see Figure 6.2) to verify the accuracy of the calculations done to obtain the indicator. The two sets of data cover the same range of dates, from July 2020 to May 2021. As a result, the values found in the dashboard are slightly higher than those found in the official data, which may be attributed to slightly different values used for the population of Spain. Nonetheless, the pattern created by the graph in the dashboard matches that in the official graph, meaning that the calculations were successful.

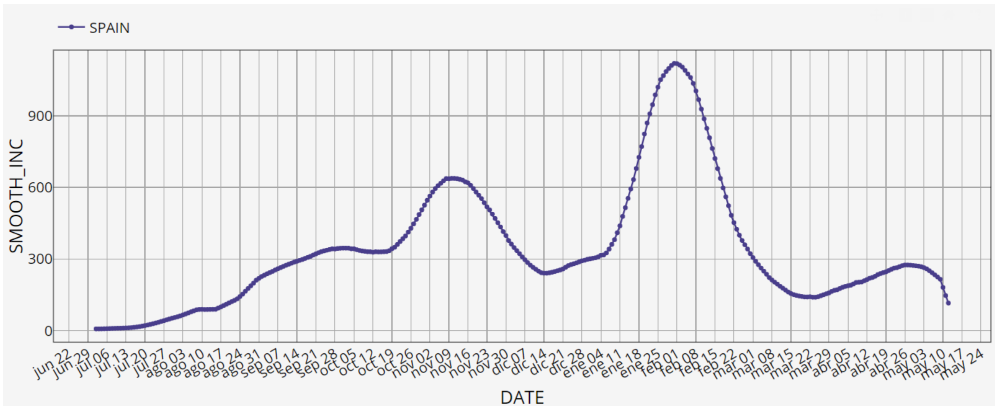

Figure 6.1: Inc of Spain from July 2020 to May 2021, as shown in Dashboard

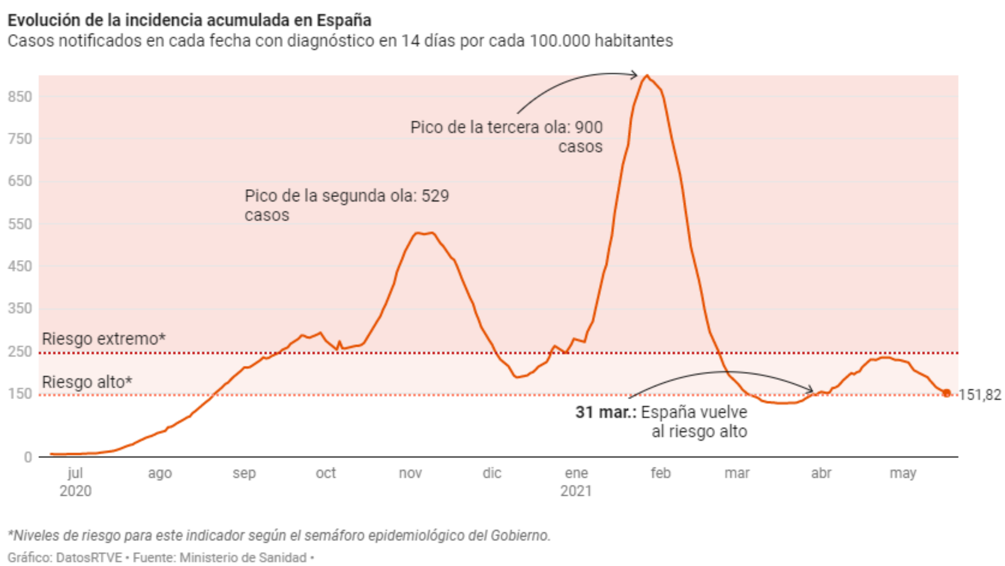

Figure 6.2: Inc of Spain from July 2020 to May 2021

6.2 Dashboard and Website

The website created throughout the duration of this project (see Figure 6.3) was published online.

The following links may be accessed to find the dashboard and website, as well as the source code of the project:

- Dashboard and website: `https://sites.google.com/ucm.es/coda`
- Source code: `https://github.com/soesteve/Covid_19_TFG/tree/main/Dashboard`

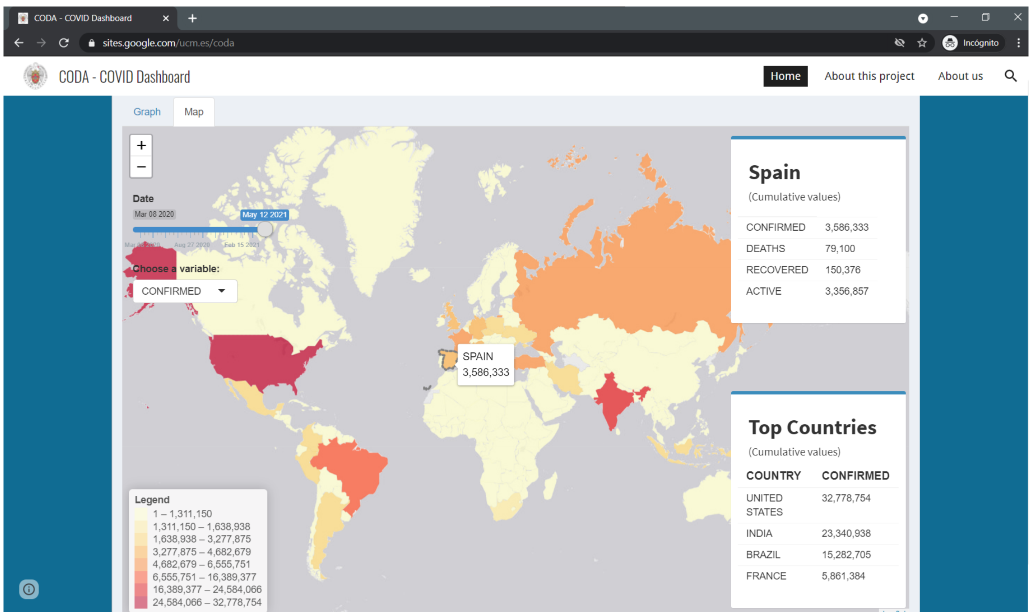

Figure 6.3: Website with integrated Dashboard

The dashboard consists of two main views: the Graph View and the Map View. Users can switch between these two views via the tabs on the upper left hand corner of the dashboard.

In the Graph View, users can opt between two modes: multiple countries, where the user can compare up to 5 countries under the same variable, or multiple variables, where the user can compare up to 5 variables under the same country. The user can also choose between displaying a line graph or a box plot and the range of dates being covered can be manipulated using a slider.

Users can hover their cursor over the points of data or over the plots to view more information about the data, such as the date of recording and its respective value. When hovering the cursor over this area of the dashboard, users can also take advantage of the functions on the right hand corner of the graph to manipulate their view of the graph as they see fit (see Figure 6.4):

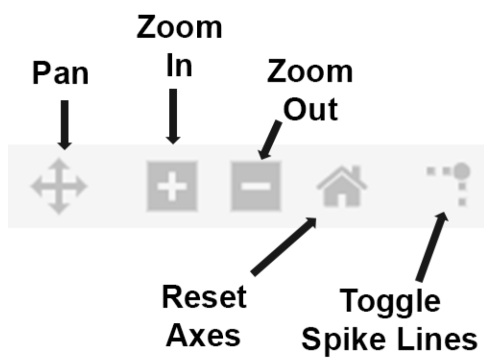

Figure 6.4: Functions available in the Graph View

- Pan: move the area being displayed in the graph
- Zoom In: zoom into the graph
- Zoom Out: zoom out of the graph
- Reset Axes: reset the scaling of the graph so that it can be viewed in its entirety
- Toggle Spike Lines: toggle the appearance of spike lines, which are dotted lines that connect points of data to their values on the X and Y axes

Additionally, by highlighting an area on the graph, users can zoom into that highlighted area.

As for the Map View, the controls are more simple. Users are able to manipulate a slider for the date and a drop-down list for the variable being viewed.

6.3 Example Analysis: Comparing Spain and the Philippines

To provide an example for the use of the dashboard, a comparison was made on the situation of two countries, Spain and the Philippines, specifically under the smoothed indicators (R0, DI, and Inc). This study covers the data recorded from March 8, 2020 to June 8, 2020, which were the first few months of the pandemic.

Beginning with the analysis under the R0 indicator (see Figures 6.5 and 6.6), it was found that, over time, the transmission power of the coronavirus in the Philippines steadily increased, even reaching an R0 value of 3.10 on June 2, 2020, and recorded an R0 value of 2.30 on June 9, 2020. As for Spain, the value for R0 slowly increased reaching an R0 value of 1.29 on April 30, 2020 and then steadily decreased and stayed below R0 value of 1 until it recorded an R0 value of 0.09 at the end of the time period.

Looking at the box plots of the data, it was observed that the Philippines contain multiple outliers compared to Spain, which had more homogeneous data and no outliers. This may suggest better gathering of data in Spain than in the Philippines. Additionally, the median of the data in the Philippines is greater than 1 while the median of the data in Spain is less than one, indicating that the transmission of the virus in the Philippines is generally stronger than in Spain.

As for the danger index curve (see Figure 6.7 and 6.8), it was found that, on May 12-18, 2020, Spain had an extremely high danger index of around 15,000. Interestingly, Spain also had a period when the danger index was negative from April 24, 2020 to May 11, 2020. The Philippines, on the other hand, had a danger index of up to 601, therefore making its curve appear flat compared to that of Spain.

Interestingly, the box plots show that both countries contain outliers (with the Philippines having slightly more outliers than Spain), which may suggest an unusual trend in the pandemic in terms of the danger index. Since the median of Spain is larger than that of the Philippines, it may be said that the flow network in Spain was more dangerous than in the Philippines. However, with a median of 189.50, the flow network of the Philippines may still be regarded as alarming.

Finally, under the indicator of accumulated incidence (see Figure 6.9 and 6.10), we see a similar case as to the previous indicator. In the case of

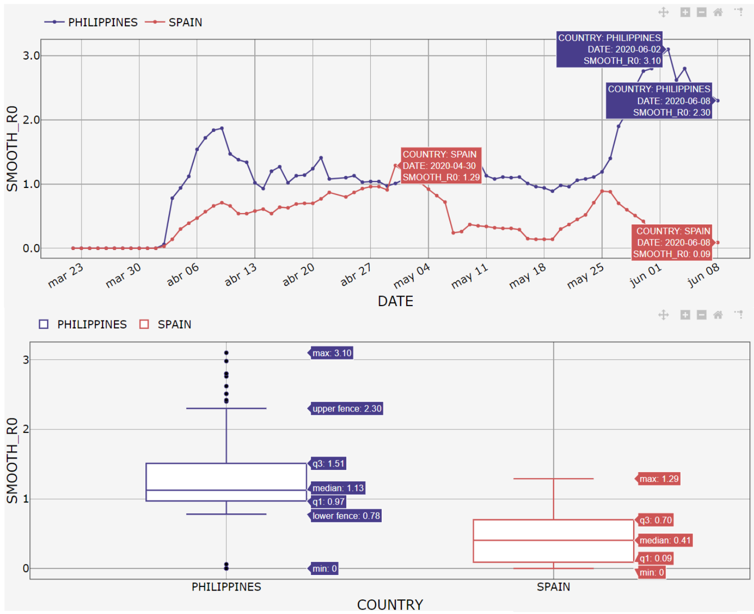

Figure 6.5: Graphs comparing Spain and the Philippines under R0

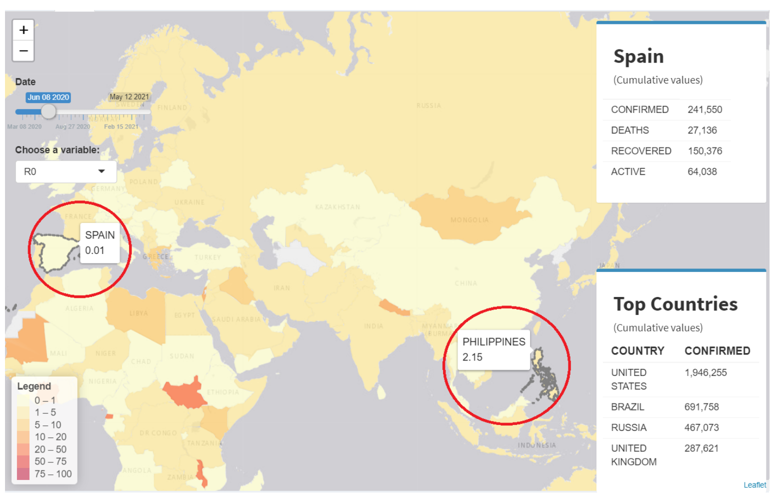

Figure 6.6: Map comparing Spain and the Philippines under R0 on June 8, 2020

Spain, the accumulated incidence increased gradually up to a value of 239.92 on April 8, 2020, and then decreased to a value of 57.30 on May 12, 2020.

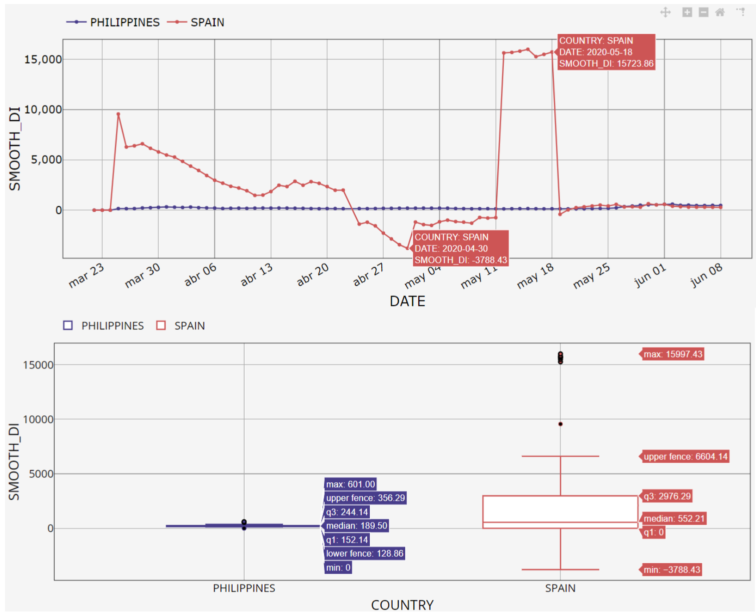

Figure 6.7: Graphs comparing Spain and the Philippines under DI

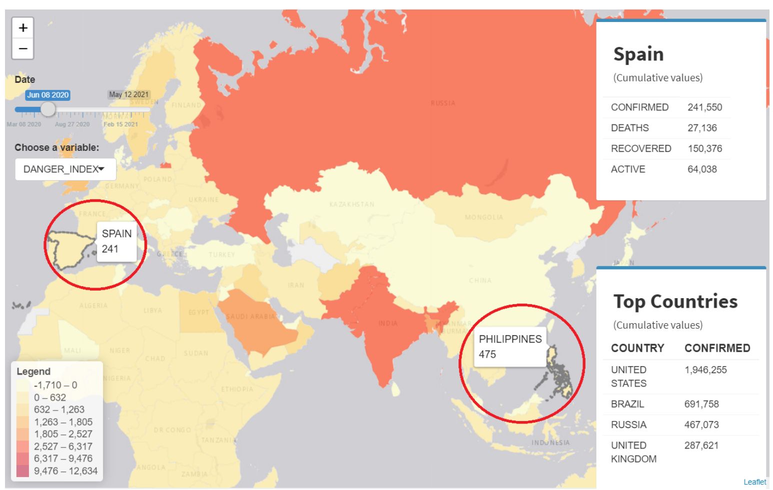

Figure 6.8: Map comparing Spain and the Philippines under DI on June 8, 2020

After that, the country, reaching an accumulated incidence of 500 to 520, saw another spike from May 19, 2020 to May 27, 2020, before decreasing again and reaching an accumulated incidence of 12.50 by June 8, 2020. The curve for the Philippines appeared flat but, upon closer inspection, steadily increased over time and reached a maximum of 7.06, which was recorded on June 8, 2020.

As with the box plots under the danger index, the box plots of the two countries under the Inc indicator shows that both countries contain multiple outliers. However, it remains clear that, in this period, both the median and the IQR Spain contained greater values of accumulated incidence than those of the Philippines. This means that, in an interval of 14 days, more positive COVID-19 cases per 100,000 inhabitants have been detected in Spain than in the Philippines.

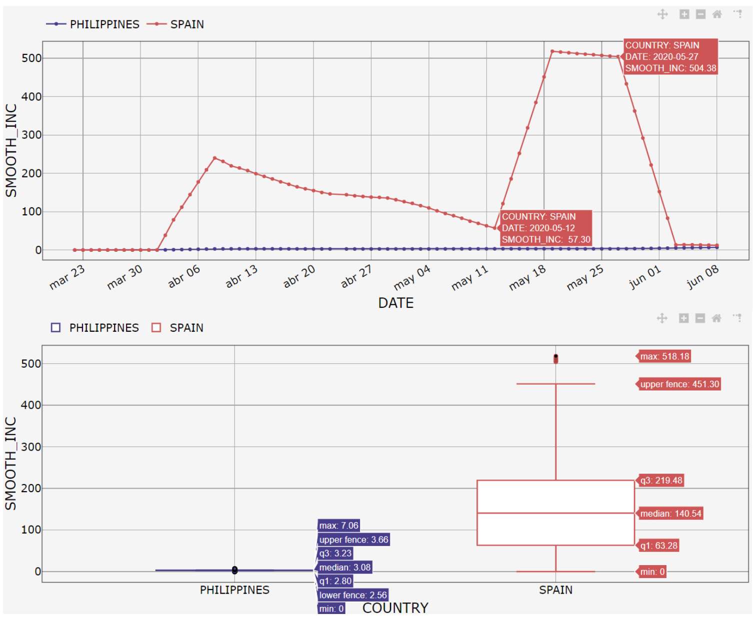

Figure 6.9: Graphs comparing Spain and the Philippines under Inc

Piecing together the information gathered from these indicators, it was found that, despite the Philippines having a more powerful transmission of the virus, Spain, which generally had drastically higher values in danger index and accumulated incidence than the Philippines, experienced a more dangerous situation overall in the given time period.

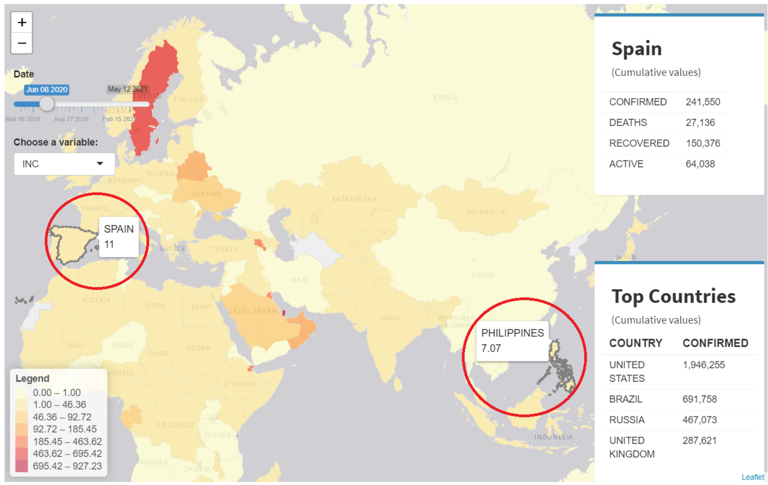

Figure 6.10: Map comparing Spain and the Philippines under Inc on June 8, 2020

6.4 Integrating the Dashboard to Other Websites

As the developed dashboard, being deployed in Shinyapps.io, is cloud-based, users can integrate the dashboard into their own websites by simply adding it as an HTML element using the following code:

```
<iframe height="720" width="100%" frameborder="no"
src="https://covid-2019-dashboard.shinyapps.io/dashboard" />
```

The advantage to this is that, using the `height` and `width` attributes, users may also change the dimensions of the dashboard to help it fit in any website layout.

Chapter 7

Individual Contributions

Throughout the entire development of the project, each member has contributed in every phase, be it through research, through implementation, or through verification. For full transparency, each team member has expounded his own contributions in this chapter.

7.1 Contributions made by Adrián García Sánchez

After the establishment of the main theme of the project, the objectives and motivation, I was in charge of establishing the work plan together with the structure of the document, which has allowed us during the whole project to be able to be organized and focused on our objectives, thus avoiding some risks related to the non-achievement of the objectives or to the disorganization and waste of time and effort.

In the research phase I focused on the technical requirements for the creation of a dashboard, such as the different layers that every functional dashboard requires. These are the backend, which is the part in charge of making all the logic of a page work, the frontend, which is the part of the software that interacts with the users, and of course the data from which the dashboard is fed. Together with Harold I was in responsible of researching about the existing databases on the internet related to this topic and the most useful types of graphics for this project, such as line graphs, maps or box plots among others. All this allowed us to get an idea of the main points of a good dashboard and those things we should avoid in our project.

In the creation of the dataset, I helped in the evaluation of the datasets, to find possible risks in their creation or related to duplicates or a bad use of space, since shiny has a 1GB limit for its free plan, which is the one we use. I did this by examining the contents of the dataset thoroughly using R and RStudio.

After the research phase, the three of us created the planning to follow, both the tools to use (R studio, Shiny, Google Sites) and the methodology to use (Scrum) and also the duration of each iteration and the division of tasks. I was in charge of distributing and planning each of the iterations, the product backlog that includes the assignment of priorities of the tasks, based on the initial planning and on the problems that arose in the development that made us adapt the planning from time to time. We chose Scrum over other agile methodologies to have this flexibility when changing the planning. I also did the dashboard design for each iteration, based on the limitations encountered in the development and above all to make an intuitive and simple tool for all the public.

In the development phase, I helped in implementing the map by studying the `leaflet` library and the layers that can be included with the map such as the legends and the polygons and shapes the define each country. I also performed testing and debugging tasks to find the bugs and things to improve in the source code. After each of the iterations we met with the director and established the conclusions and prepared the basis for the next iteration.

Regarding the website I took care of the creation and the design and organization of the pages, taking into account the many limitations that google sites has, but enough for our purpose. The site was organized focusing on the dashboard, which is the most important piece and occupies most of the main page. We added several extra pages explaining various aspects of the work, such as its user manual or data repositories. We also found ways to contact the members of the group. I also pointed out the possible modifications and adjustments concerning the aesthetics such as the font sizes, the positioning of the panels, the colors, the margins, etc. This is to improve the visualization of the dashboard in the web page.

Finally, we established together the conclusions, where we commented all the knowledge acquired during the elaboration of the project, both at the level of tools and data analysis or methodology. Then we defined some future work, where we establish possible lines of action for the continuation of the project. I was also in charge of the modification of the report, in latex and through the Overleaf platform, adjusting the format and fixing the errors.

7.2 Contributions made by Dale Francis Valencia Calicdan

At the start of the project, first thing we did was specify the main topic of this project. We defined the subject this application was about which was the coronavirus, its definition, origin, and the importance of studying its evolution. With that, I defined the objectives this whole project will encircle into and the main goals the team should achieve by the end of this project's duration.

As we go on into the investigation phase of the project, I was in charge of studying the different variables involved in studying the evolution of the coronavirus disease. I spent time reading articles and looking through the various dashboards my colleagues were investigating to identify the common variables and others that were significant. I evaluated them by which I chose the most important ones which could be of interest in developing the dashboard.

Upon the end of the investigation phase, we moved on into the development phase. In this phase, we first laid out the plan the team should follow, the methodology to use, which was Scrum methodology, the duration of each iteration, and the partition of tasks. In each iteration I was primarily focused in building the source code of the program in R language, while strictly following the sprint backlog and the design established for every iteration. This also included in researching and studying various libraries which could be of use in developing the dashboard. By the end of every iteration, I also reported the various limitations I have encountered. This led into suggestions on how to improve the dashboard for the next iteration. By the last iteration, I uploaded the dashboard into the dedicated server which included signing up an account, uploading the source code and resources, and finally, deploying the application.

At the begginning of this phase, we worked on an initial database provided by the directors of the project, which is a small dataset containing recordings of coronavirus cases of some countries. By the 3rd iteration, the first version of the proposed dataset, created by the other group, was sent to us. The format of this dataset (number of columns, column names, variables available, etc.) was different from the one we previously worked on. I restructured the code and adapted it to the format the proposed dataset used.

Later on it was seen that the data in the dataset provided by the partner group was organized by regions while the dashboard is structured to show data by country. We also learned that calculations done inside the source code takes a lot of memory. With these findings, the team decided to create another dataset by transforming the ones we were provided. I created the script which performs the necessary transformation of the dataset, as well as the calculations of some variables, to avoid unnecessary operations while the application is deployed in the server.

By the end of the development phase, my colleagues were implementing the web page for the dashboard. After discussing some modifications and possible adjustments for the HTML elements of the dashboard, for it to fit in the allocated window in the web page, I created a CSS file and adjusted the margins and paddings of the dashboard as well as some of its aesthetics to further improve its accessibility and usability.

After the development phase, I contributed in redacting the documentation of this project, assisting in the chapters I was primarily focused at. I explained the final dataset that was one of the results of this project, and proofread the entire part of Chapter 5 explaining the project's development.

7.3 Contributions made by Harold Luis Pascua Cajucom

I started by spearheading the research on the different types of dashboards available, as well as the examples for each one. I initially found that most dashboards are used in businesses and that the classifications in that context do not work with the context of monitoring COVID-19 and this led me to analyze the existing dashboards based on their common factors: geographic scope and type of data used. After that, I researched on the types of software tools available, especially for creating and configuring dashboards.

I also contributed to the research of data of interest, namely the existent databases of interest and the graphs of interest. These were databases and graphs that may be put into consideration for inclusion in the dashboard. In addition, I also researched on some variables of interest: R0, DI, and Inc.

As for the data preparation stage of the project, I focused on evaluating the datasets, checking for points that would become problematic or concerning when integrated with the dashboard to display data.

For the initial planning of the implementation of the project, I contributed in deciding the different specifications. For example, we decided to apply the Scrum methodology, as we had some experience with this methodology from our experience in learning Software Engineering. We also decided at this point to implement the dashboard using the RStudio IDE, together with Shiny and the other RStudio libraries used. Finally, we decided to embed the dashboard onto a Google Sites website, as we wanted a website that could easily be configured so we could focus more on the development of the dashboard.

Also in the planning stage, we decided on what type of COVID-19 dashboard we wanted the final product to be. Based on the information we received from the research on the different types of dashboards, we decided to mainly focus on epidemiological data with a global scope, as we wanted the dashboard to be reachable to as many people as possible.

Throughout the development phase, I contributed to all development sprints by testing and debugging the dashboard and by formulating the conclusions for each sprint. It was through this debug process that we were able to detect the anomalies found during development that would have to be resolved by the following sprints.

I also contributed in developing the dashboard by implementing the graphs, especially the box plots that was allowed further analysis of variables. By the end of this phase, I managed to refactor the source code to improve its comprehensibility. This includes defining the auxiliary functions found in the `functions.R`script, such as the creation of tables and fetching of information from the dataset.

In the presentation of results, I prepared the repository that contained the datasets used for the dashboard, making those datasets available for public use. Additionally, I created the user manual that would help orientate the users on the different available functions of the dashboard and facilitated the example analysis of the situations of Spain and the Philippines. This would help facilitate the users in knowing how to use the dashboard to get the results they need and properly interpreting those results to get insights.

Together with my colleagues, we formulated the conclusions made post-development. We mainly focused on the things we learned from the different fields of dashboard development, which includes preparing data, adapting to using different libraries and technologies, and data analysis. We also created the recommendations for future work, some of which were actions that we initially wanted to take but were unable to due to time constraints.

Finally, I was in charge of tasks related to the documentation of the project. This included the compilation of sources cited in this work into the bibliography and proofreading the work in general, making sure that it was free of any grammatical errors and that every detail about the project was well-explained.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

A dashboard was created that can inform users about the progress of the pandemic in graphic and map forms, especially under indicators such as transmission index, danger index, and cumulative incidence.

Throughout the project, the team has used R and Shiny, as well as libraries such as Plotly and Leaflet to develop the dashboard. Balsamiq Wireframes was also used to create wireframes and prototypes for the design of both the dashboard and the website. Additionally, LaTeX was learned for the creation of this work.

Applying the Scrum methodology in this project was also an important practice learned. As Scrum has become a standard in most companies, learning how to employ this Agile methodology in a practical setting proves to be useful in preparing for work under employment.

Another crucial skill taken from this project is the application of various analytical techniques. Having multiple types of graphs present in the dashboard allowed for the understanding of different aspects of the data (e.g. its evolution over time, variance in values). Exploratory data analysis also plays a part due to the nature of the dashboard of being based on a mathematical model. As the sources of data found in the dataset mainly consist of simple recorded values such as positive cases and deaths, indicators such as the R0, DI, and Inc were mathematically derived from these records.

The final conclusion is that, as important it is to analyze data thoroughly and extensively, making these data understandable and accessible to the public is equally important. The users of the dashboard will consist of people of different backgrounds and of different perspectives on the pandemic. Hence, it is essential that the same idea is objectively imparted to them by the dashboard to attain its purpose.

8.2 Future Work

Despite having successfully created a fully functioning COVID-19 dashboard, the team strongly believes that the project still contains room for improvement.

The first recommended action is the automation of the data acquisition process. Currently, the dashboard uses only one dataset, meaning that the data used could easily become obsolete in a matter of days if the dataset is not constantly updated. Manually updating this dataset by uploading a later version after a given time interval is a viable yet strenuous option in resolving this issue. Thus, automating this process will allow the dashboard to stay up to date without requiring a large amount of effort from the team in terms of maintenance.

Another recommended action is incorporating animations into the map view of the dashboard, which could amplify its data visualization capabilities. This would allow users to see the evolution of the pandemic in different countries more clearly, providing further interaction to the application.

Bibliography

[1] World Health Organization. Coronavirus disease (covid-19). https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed: Oct. 30, 2020.

[2] World Health Organization. Origin of sars-cov-2, 26 march 2020. Technical documents, 2020. Accessed: Oct. 30, 2020.

[3] Emilio de Benito. Nearly one-third of all intensive care beds in spain occupied by coronavirus patients. https://english.elpais.com/society/2020-11-04/nearly-one-third-of-all-intensive-care-beds-in-spain-occupied-by-coronavirus-patients.html, Nov 2020. Accessed: Nov. 26, 2020.

[4] Ashley Killough and Omar Jimenez. 'busting out of the seams': West texas hospitals pushed to the limit in unprecedented covid-19 surge. https://edition.cnn.com/2020/11/19/us/west-texas-covid-19-hospitalizations-surge/index.html, Nov 2020. Accessed: Nov. 26, 2020.

[5] Lora Jones, Daniele Palumbo, and David Brown. Coronavirus: How the pandemic has changed the world economy. https://www.bbc.com/news/business-51706225, Jun 2020. Accessed: Nov. 26, 2020.

[6] World Health Organization et al. Covid-19 weekly epidemiological update. 2020. Accessed: Nov. 26, 2020.

[7] Sandra Durcevic. Types of dashboards: Strategic, operational & analytical. https://www.datapine.com/blog/strategic-operational-analytical-tactical-dashboards/, Jul 2020. Accessed: Nov. 9, 2020.

[8] Instituto de Salud Carlos III. Situación y evolución de la pandemia de covid-19 en españa. https://cnecovid.isciii.es/covid19/, 2020. Accessed: Oct. 30, 2020.

[9] S Obrutsky. Cloud storage: Advantages, disadvantages and enterprise solutions for business. In *Proceedings of the Eastern Institute of Technology Conference*, page 10, 2016.

[10] Alexey Pakhomov. https://covidly.com/, Mar 2020. Accessed: Nov. 9, 2020.

[11] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.

[12] World Health Organization. Who coronavirus disease (covid-19) dashboard. https://covid19.who.int/, 2020. Accessed: Oct. 28, 2020.

[13] Nicholas DeVito et al. https://covid19.trialstracker.net/, 2020. Accessed: Oct. 28, 2020.

[14] Kristian Thorlund, Louis Dron, Jay Park, Grace Hsu, Jamie I Forrest, and Edward J Mills. A real-time dashboard of clinical trials for covid-19. *The Lancet Digital Health*, 2(6):e286–e287, 2020.

[15] Kristian Thorlund, Louis Dron, Jay Park, Grace Hsu, Jamie I Forrest, and Edward J Mills. https://www.covid-trials.org/, 2020. Accessed: Oct. 28, 2020.

[16] Rahul K Arora, Abel Joseph, Jordan Van Wyk, Simona Rocco, Austin Atmaja, Ewan May, Tingting Yan, Niklas Bobrovitz, Jonathan Chevrier, Matthew P Cheng, et al. https://serotracker.com/, 2020. Accessed: Nov. 9, 2020.

[17] Rahul K Arora, Abel Joseph, Jordan Van Wyk, Simona Rocco, Austin Atmaja, Ewan May, Tingting Yan, Niklas Bobrovitz, Jonathan Chevrier, Matthew P Cheng, et al. Serotracker: a global sars-cov-2 seroprevalence dashboard. *The Lancet. Infectious Diseases*, 2020.

[18] Rohit Bind, Akash Deep, et al. https://thevaccinetracker.com/, 2020. Accessed: Oct. 31, 2020.

[19] RStudio, PBC. `https://shiny.rstudio.com/`, 2020. Accessed: Oct. 30, 2020.

[20] The R Foundation. What is r? `https://www.r-project.org/about.html`, 2020. Accessed: Oct. 31, 2020.

[21] J.J. Allaire. Rstudio: integrated development environment for r. *Boston, MA*, 770:394, 2012.

[22] RStudio, PBC. `https://www.shinyapps.io/`, 2020. Accessed: Nov. 9, 2020.

[23] Microsoft Power BI. What is power bi | microsoft power bi. `https://powerbi.microsoft.com/en-us/what-is-power-bi/`, 2020. Accessed: Oct. 30, 2020.

[24] Grafana Labs. Grafana features. `https://grafana.com/grafana`, 2020. Accessed: Oct. 30, 2020.

[25] Sisense. Business intelligence (bi) software & analytics platform. `https://www.sisense.com/`, Oct 2020. Accessed: Oct. 30, 2020.

[26] Domo. Why domo? `https://www.domo.com/why-domo`, 2020. Accessed: Oct. 30, 2020.

[27] Esri. Arcgis dashboards. `https://www.esri.com/en-us/arcgis/products/arcgis-dashboards/overview`, 2020. Accessed: Nov. 9, 2020.

[28] Miroslaw Staron. Dashboard development guide how to build sustainable and useful dashboards to support software development and maintenance. 2015.

[29] Instituto de Salud Carlos III. Vigilancia de la mortalidad diaria (momo). `https://momo.isciii.es/public/momo/dashboard/momo_dashboard.html`, 2020. Accessed: Oct. 31, 2020.

[30] Hannah Ritchie, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, and Max Roser. Coronavirus pandemic (covid-19)-statistics and research, Oct 2020. Accessed: Oct. 31, 2020.

[31] European Union Open Data Portal. Covid-19 coronavirus data - daily (up to 14 december 2020). https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data-daily-up-to-14-december-2020, 2020. Accessed: Feb. 15, 2021.

[32] Tim Miller. What is a box plot and when to use it. https://chartio.com/resources/tutorials/what-is-a-box-plot/, 2021. Accessed: May 17, 2021.

[33] Centers for Disease Control and Prevention. Cdc library: Covid-19 science update: 2/19/2021. https://www.cdc.gov/library/covid19/02192021_covidupdate.html, Feb 2021. Accessed: Feb. 21, 2021.

[34] Victoria López and Milena Čukić. A dynamical model of sars-cov-2 based on people flow networks. *Safety Science*, 134:105034, 2021.

[35] Victoria Lopez and Milena Čukić. The comparison of trends in spain and the nederland: a dynamical compartment model of the transmission of coronavirus. *arXiv preprint arXiv:2004.09874*, 2020.

[36] Adam Kucharski, Timothy Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind Eggo, Fiona Sun, Mark Jit, James Munday, Nicholas Davies, Amy Gimma, Kevin Zandvoort, Hamish Gibbs, Joel Hellewell, Christopher Jarvis, Samuel Clifford, Billy Quilty, Nikos Bosse, and Stefan Flasche. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 20, 03 2020.

[37] UC Davis Health. Different types of covid-19 tests explained. https://health.ucdavis.edu/health-news/newsroom/different-types-of-covid-19-tests-explained/2020/11, Nov 2020. Accessed: Dec. 1, 2020.

[38] Pete Riley, Allison Riley, James Turtle, and Michal Ben-Nun. Covid-19 deaths: Which explanatory variables matter the most? *medRxiv*, Jun 2020.

[39] Georgia Rural Health Innovation Center. What is a moving average and why is it useful? https://www.georgiaruralhealth.org/blog/what-is-a-moving-average-and-why-is-it-useful/, 2021. Accessed: May 20, 2021.

[40] Scrum.org. What is scrum? `https://www.scrum.org/resources/what-is-scrum`, 2021. Accessed: May 17, 2021.

[41] Google. Google sites. `https://sites.google.com/`. Accessed: May 17, 2021.

[42] Department of Economic United Nations and Population Division Social Affairs. World population prospects - population division - united nations. `https://population.un.org/wpp/Download/Standard/CSV/`, 2019. Accessed: April 27, 2021.

[43] RTVE.es. El coronavirus: Gráficos, mapas y datos del covid-19. `https://www.rtve.es/noticias/coronavirus-graficos-mapas-datos-covid-19-espana-mundo/`, 2021. Accessed: April 28, 2021.