COVID-19 Italian Dashboard

Marcello Piccinini
Engineering in Computer Science
Sapienza University of Rome
Rome, Italy
piccinini.1759133@studenti.uniroma1.it

***Abstract*—The pandemic was an unexpected event that totally changed the life of each of us, from different points of view, social, economic and healthy.**
Since the beginning of the pandemic, a lot of data related to the virus has been collected around the world and has been used to try to counter this dramatic situation. Being able to create a dashboard to view the data in a simple and clear way is of vital importance, not only to understand what are the best actions to perform in the future but also to understand if the decisions taken in the past to eradicate the covid have had a positive or negative effect on the epidemiological situation of that particular period.

I. INTRODUCTION

COVID-19 was first identified on December 31, 2019, by the Chinese authorities in Wuhan, causing more than 500,000,000 cases and 6,000,000 deaths in just over two years.
The large amount of data collected from the beginning of the pandemic to today has been the main weapon of any government to counter the advance of the virus as every decision made was based on epidemiological data that was analyzed by experts in the field.
To help countries in their decisions and citizens to know what was happening, different global dashboards were created which were constantly updated day after day with the arrival of new data.
The dashboard developed for this project displays the pandemic data of Italy by making a division by region and is not limited to the simple display of current data but is based on the idea of being able to view data from the past in any time interval chosen by the user so that an accurate analysis of the chosen period can be carried out.

II. RELATED WORKS

Since the beginning of the pandemic, numerous dashboards and papers have been published and made available to help fight the virus. The works related to this project are dashboards from Italian Civil Protection [1], World Health Organization [2], Johns Hopkins University [3], European Center for Disease Prevention and Control (ECDC) [4] and the paper “An Interactive Dashboard for Monitoring the Spread of COVID-19 in Sudan” [5].
Despite this, it is possible to highlight numerous differences. First, the primary objective. All the works cited above show the total of cases, for the various metrics chosen, up to now or up to a date chosen by the user. However, it is not possible to understand what the increase in cases has been in a period of time that is not necessarily since the beginning of the pandemic given that there is no range slider in any solution (except for the paper [5] where anyway the integration with the other visualizations is very low).
In the dashboard developed in this project, however, it is possible to show this information depending on the use of the time slider. If the date that is modified is

only the final one, the increase in cases corresponds to the total number of cases, thus showing information consistent with the other dashboards. If, on the other hand, the initial date is also modified, then only the number of new cases is displayed as if the pandemic had started on that precise day. Furthermore, by superimposing the initial date with the final one, it is possible to know the data of a single day.
A common aspect to all the related works is the use of a map to visualize the different values on the territory. In some solutions, however, the choice made is to use a bubble map. In this way the number of cases is indicated by the diameter of the bubble that represents a territory.
However, this choice presents several problems compared to a choropleth map as it is more difficult for a user to determine the quantitative value of a metric using the area of a circle rather than with the appropriate use of colors. Additionally, the plot may become less clear to understand due to overlapping.
In this project, as in paper [5], it was preferred to focus on a single nation rather than carry out a global analysis in order to be able to have a more precise and specific analysis with respect to the available database.
As for the dataset, the same source is used both in the Italian Civil Protection [1] and Johns Hopkins University [3] dashboard, while the other works use different sources.
Finally, none of the solutions mentioned use a dimensional reduction technique such as MDS or PCA, which is instead implemented in this work, displaying the result through a scatterplot.

III. DATASET

A. *Data Collection*

The system implementation requires two different types of data.
The first is a json file used to display the map of Italy with the borders of the regions [6].
The second dataset, on the other hand, is provided by the Ministry of Health and is processed and managed by the Department of Civil Protection [7]
The original dataset contains 17934 tuples and 30 dimensions, so the AngeliniSantucci index is equal to 538.020.
After preprocessing, the dataset consists of 17.080 rows and 11 columns with a AS index equal to 187.880.
The attributes that are considered are:

- *date*: Date of information.
- *region*: Name of the region.
- *hospitalized_with_symptoms*: Hospitalized patients with symptoms.
- *intensive_care*: Patients in intensive care.
- *total_hospitalized*: Total hospitalized patients.
- *home_isolation*: People in quarantine.
- *total_positives*: Number of positive patients (hospitalized patients + home isolation).
- *discharged_healed*: Recovered
- *deceased*: Number of death people.
- *total_cases*: Number of cases (total positives + discharged_healed + deceased).
- *swabs*: Tests performed.

B. *Preprocessing*

After selecting the most relevant attributes of the dataset, to create the desired dashboard it was necessary to carry out a preprocessing work.
The dataset showed a division of the *Trentino-Alto Adige* region into the two autonomous provinces of *Trento* and *Bolzano*. To obtain the same correspondence with the Italian map, the data of the two provinces were merged.
Another aspect to be solved was to express the data provided in such a way that they satisfied the goal of the project.
In fact, the developed dashboard has the purpose of expressing the number of new "cases" that have occurred for each chosen metric in a specific time interval,

highlighting only the positive increase in data without considering the negative increase or the total number of cases since the beginning of the pandemic in Italy or in a specified region.
To achieve this, each column of the dataset was preprocessed in such a way that for each day the number of new increments was indicated, thus having a value equal to zero when there is no variation in the cases or a negative one.

IV. VISUALIZATION AND INTERACTION

The COVID-19 Italy Dashboard is composed by a set of four interactive visualization coordinated by a temporal range slider. When the slider is used to choose a time interval, the plots on the dashboard are updated to show the new data.

A. *Choropleth Map*

A choropleth map is a thematic map that provides an easy way to visualize how a variable changes across a geographic area.
In this dashboard this visualization shows the Italian map with a division by regions and the analyzed variable is the number of total new cases in each region.
The choice of the color scale used to express the values was made using *ColorBrewer 2.0* [8].
Each color corresponds to a specific range of values which is expressed in the upper left using a legend.
Interaction:

- *On mouse over:* the region is highlighted in green, and the name of the region is shown with the specific number of new cases total.
  The region is also highlighted in green in the scatterplot and in the parallel coordinate plot.
- *On click*: the region that is clicked is highlighted in red on the map and in the scatter and parallel coordinate plot.
  When one or more portions of the map are selected, the boxplots are updated with the analysis of data relating exclusively to the regions that have been clicked.

B. *Parallel coordinates plot*

Parallel coordinates plots are ideal for plotting multivariate numerical data and for comparing many variables together.
In the project this visualization shows all the nine metrics used in the database end each region is represented by a line.
Interaction:

- *On mouse over*: The region is highlighted in green in the current plot, in the map and in the scatterplot. The region name is shown and a window appears to the right of the legend containing the values of each metric.
- *On click*: The path is highlighted in red as the region in the map and the point in the scatterplot. This action triggers the update of the boxplots for the selected regions.
- *On brush*: Is possible to perform a brush operation for each axis to select only the regions with a specific range of values for one or more metrics. The regions selected in the brush selection are selected also in the map and in the scatterplot and the boxplots are updated.

C. *Scatterplot (MDS)*

Multidimensional scaling is a powerful technique used to perform dimensionality reduction. The project used MDS to visualize regions in a 2D scatterplot such that if two regions are similar considering all the metrics, than they should be similar even in a 2D space. The reduction performed is from R^9 to R^2. The multidimensional scaling takes in input a symmetric matrix M of dissimilarities, the bigger the value M[x][y], the bigger is the dissimilarity between the region x and y.
The Euclidean distance is used to compute the dissimilarity value.

Once the reduction is done the regions are visualized as points in a scatterplot.
The data are not precomputed, but the operation is done each time the slider changes. The only limit is that the scatterplot is not automatically updated but it is necessary to press the appropriate blue button next to the slider.
Interaction:

- *On mouse over*: the point in the scatterplot, the path in the parallel coordinates and the region in the map are highlighted in green and the name of the region appears.
- *On click*: The point is highlighted in red as in the other visualization and the boxplots are updated.
- *On brush*: When one or more points are selected using the brush function the regions are selected also in the map and in the parallel coordinates plot and the data about the regions selected are used for the boxplots.

V. *ANALYTICS*

The dashboard provides analytics using the boxplots.
Three metrics were considered: swabs, deceased and total cases.
Using the boxplots is possible to visualize some information that is not present in the dataset as the *maximum*, *minimum*, *median*, *first quartile* and the *third quartile* of each metric in a single day of the given period of time. It doesn't consider the cumulative value such as in the other plots of the dashboard so that the user can have a more powerful knowledge of the situation.
The information is not precomputed, but the computation of the analysis is triggered by the user using two different visual interaction:

1. the user can change the time period using the slider.
2. The user can select/deselect one or more regions to trigger the analysis only for the selected regions in the period chosen before (if no regions are selected the system consider all of them).
   This operation can be done in all the three plots described before in a coordinated way.

The dashboard also provides an interaction with the boxplots such that when the mouse is on the green rectangular all the information related to the given matrix appears. In this window there is also another type of information that is computed by the analytics and can be useful to the user, the *mean*.

VI. *DISCOVERED INSIGHTS*

The purpose of creating the dashboard is not only visualizing the data, but using it appropriately, it is possible to discover some insights that otherwise would be much less intuitive to find.
The first insight that it is possible to discover is certainly the difference between the number of cases among regions in specific periods of the year.
Although the *Lombardia* appears to be the region with the largest number of total cases to now, in some pandemic periods this has not been the case. An example is the summer period in which tourism tends to concentrate more in the southern regions. If we take as a reference period the one that goes from 01/06/2021 to 01/09/2021, the region with the greatest number of new total cases turns out to be Sicily with about 50,000 cases.
Another important application of the dashboard is to analyze the decisions made by the Italian government during the pandemic to understand which decisions were the most effective or not so as to be prepared if similar situations were to be found again in the future.
An example is to take a DPCM and analyze the number of total cases before and after. To have an analysis that is as consistent as possible with reality, the periods considered are three months before and after.

Take for example 4 November 2020 as the reference date.
The entire national territory came from a period of great restrictions (closing of bars, restaurants and gyms, distance learning and smart working).
The government decided to issue a new DCPM in which Italy was divided into 3 colored zones: white, orange and red, each with different restrictions.
Using the dashboard, however, it is possible to understand that this decision did not have a positive effect on the spread of the virus.
In fact, it goes from a total of 540,000 cases in the period between 04/08/2020 - 04/11/2020 to 1,800,000 in the following quarter, 04/11/2020 - 04/02/2021.
Only later, on 02/26/2021, the restrictions in the various colored areas were updated and also the criteria for passing from one area to another became more restrictive. This change led to an improvement in the situation, in the period 04/02/2021 - 04/05/2021 the number of total new cases in Italy dropped to 1,470,000.
This analysis was carried out considering only the DPCM issued by the government, but for a more complete result it would be necessary to consider all the factors that influenced the spread of covid such as variants, vaccines etc..
In any case, the dashboard, as demonstrated, is an effective tool for this type of study.

VII. *CONCLUSION AND FUTURE WORKS*

In conclusion, the dashboard provides a complete and clear view of the Italian situation from the pandemic to now.
It is able to show various information, not identifiable by other solutions through effective and easy to interpret graphs.
The project can also be used for different purposes, adapting the dataset chosen.
For future work it is possible to integrate the dashboard with an additional metric, fundamental for the study of the epidemiological situation such as vaccines.

REFERENCES

[1] https://opendatadpc.maps.arcgis.com/apps/dashboards/b0c68bce2cce478eaac82fe38d4138b1

[2] https://covid19.who.int

[3] https://coronavirus.jhu.edu/map.html

[4] https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html

[5] A. M. O. Abdelsamad and A. Z. Karrar, "An Interactive Dashboard for Monitoring the Spread of COVID-19 in Sudan," 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 2021, pp. 1-6, doi: 10.1109/ICCCEEE49695.2021.9429561.

[6] https://github.com/openpolis/geojson-italy

[7] https://github.com/pcm-dpc/COVID-19

[8] https://colorbrewer2.org/#type=sequential&scheme=OrRd&n=5

Picture Dashboard