

EU CoVis-19: visual analysis of Covid-19 effects in Europe

Valerio Coretti
Engineering in Computer Science
Sapienza University of Rome

Rome, Italy
coretti.1635747@studenti.uniroma1.it

Fabio Caputo
Engineering in Computer Science
Sapienza University of Rome

Rome, Italy
caputo.1695402@studenti.uniroma1.it

Weihao Peng
Engineering in Computer Science
Sapienza University of Rome

Rome, Italy
peng.1713518@studenti.uniroma1.it

Abstract—We live in the era of big data. For each topic, we have a huge amount of data that we can analyze. Powerful tools have been created over the years to manage big data. In this document we will try to use these tools to make an in-depth analysis of one of the largest pandemics the world has ever suffered. We are talking about Covid-19. In the past two years of the pandemic, a vast amount of epidemiological data has been collected. We have created a platform for visualizing this data, using the latest available Visual Analytics techniques. We have come up with a solution that can help users better understand information about COVID-19 deaths, cases and vaccines with a focus for the European countries. The repo containing all the material is accessible at the following link: <https://github.com/Covid19-visualization>

Index Terms—Visual Analytics, Covid-19, Vaccine

I. INTRODUCTION

II. RELATED WORK

III. DATASET

Before we started implementing our system, we needed a lot of information about COVID-19 and therefore we took the *Our World in Data* [1]. The Dataset is very very huge (AS index greather than 6 milion), it contains the collected for all the world. This dataset was built by collecting data from different sources:

- 1) COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)
- 2) European Centre for Disease Prevention and Control
- 3) Government sources

Dataset is composed by a total o 65 columns, which are the features, and more than 125.000 rows.

A. Preprocessing

The size is very large and for this reason we have decided to select the European countries, so the number of rows is about 40 thousand, and to select only a part of the features, the ones related to vaccines, death and cases:

- *name*: Country name
- *continent*: Continent of the geographical location
- *date*: Date of observation
- *population*: Population (latest available values).

- *population_density*: Number of people divided by land area, measured in square kilometers, most recent year available
- *median_age*: Median age of the population, UN projection for 2020
- *gdp_per_capita*: Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available
- *cardiovasc_death_rate*: Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people)
- *diabetes_prevalence*: Diabetes prevalence (% of population aged 20 to 79) in 2017
- *female_smokers*: Share of women who smoke, most recent year available
- *male_smokers*: Share of men who smoke, most recent year available
- *life_expectancy*: Life expectancy at birth in 2019
- *human_development_index*: A composite index measuring average achievement in three basic dimensions of human development a long and healthy life, knowledge and a decent standard of living.
- *total_cases*: Total confirmed cases of COVID-19
- *new_cases*: New confirmed cases of COVID-19
- *new_cases_smoothed*: New confirmed cases of COVID-19 (7-day smoothed)
- *total_cases_per_million*: Total confirmed cases of COVID-19 per 1.000.000 people
- *new_cases_per_million*: New confirmed cases of COVID-19 per 1,000,000 people
- *new_cases_smoothed_per_million*: New confirmed cases of COVID-19 (7-day smoothed) per 1,000,000 people
- *total_deaths*: Total deaths attributed to COVID-19
- *new_deaths*: New deaths attributed to COVID-19
- *new_deaths_smoothed*: New deaths attributed to COVID-19 (7-day smoothed)
- *total_deaths_per_million*: Total deaths attributed to COVID-19 per 1,000,000 people
- *new_deaths_per_million*: New deaths attributed to COVID-19 per 1,000,000 people
- *new_deaths_smoothed_per_million*: New deaths attributed to COVID-19 (7-day smoothed) per 1,000,000

people

- *people_vaccinated*: Total number of people who received at least one vaccine dose
- *people_fully_vaccinated*: Total number of people who received all doses prescribed by the vaccination protocol
- *new_vaccinations*: New COVID-19 vaccination doses administered (only calculated for consecutive days)
- *new_vaccinations_smoothed*: New COVID-19 vaccination doses administered (7-day smoothed). For countries that don't report vaccination data on a daily basis, we assume that vaccination changed equally on a daily basis over any periods in which no data was reported. This produces a complete series of daily figures, which is then averaged over a rolling 7-day window

B. Data management

Due to the fact that we have to manage a very huge amount of data, we have chosen to store them inside a non relational DB, making the accessibility easier.

C. Principal component analyses (PCA)

Dimensionality reduction, or dimension reduction, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension [2].

Specifically, for this task, we decided to use PCA (Principal component analysis), a linear technique for dimensionality reduction that performs a linear mapping of the data to a lower-dimensional space so that the variance of the data in the low-dimensional representation is maximized. Therefore, this method allows us to plot each multidimensional tuple on a bidimensional space, still maintaining all the underlying properties. The algorithm is applied on all the attributes and the results are showed inside a scatterplot.

IV. TECHNOLOGIES

Covid19 Visualizer is a platform and it is built as a proper web application with the following technologies:

A. NodeJS and MongoDB

We used NodeJS [4] to built our Back-end, where we do all the computation and where we retrieve the data from the DB. To store the data we have chosen the widely used MongoDB [5], a non relational database that is very easy to use with Node.

B. D3.js

The D3.js framework [6] have been used for the development of the visualizations that compose the service.

C. React

V. VISUALIZATIONS

EU CoVis-19 is composed by a set of different visualizations. In this chapter we want to analyze each component individually to understand what it is showing and how it interacts with the others. As mentioned before, the strength of our platform compared to the existing ones is the interaction between the various charts. In fact, each of them is connected to the others in order to make the user experience simpler and more direct. There are three main views that we are using:

- The first view shows data about the deaths (Fig. 1);
- The second view shows data about the cases (Fig. 2);
- The third view shows data about the vaccinations (Fig. 3);

A. Navbar selection

Opening the platform the first data that are showed are the one related to the Europe. They are an aggregation of the data for all the european countries. Then we hav a navbar where the user can select the views, the countries and also the time span. Furthermore, clicking the circle with the europe flag we can reset the selection.

B. Choropleth Map

C. Line chart

A line chart is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments [11].

D. Bar chart

A bar chart is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. Therefore, it is used to show comparisons among discrete categories. One axis of the chart shows the specific categories being compared, and the other axis represents a measured value [8].

E. PCA chart

A scatter plot is a type of plot or mathematical using Cartesian coordinates to display values for typically two variables for a set of data. If the points are coded (color/shape/size), one additional variable can be displayed. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis [9].

F. Radar chart

A radar chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point [10].

G. Table chart

VI. CASE OF STUDY

VII. CONCLUSION

VIII. FUTURE WORK

REFERENCES

- [1] <https://github.com/owid/covid-19-data>
- [2] https://en.wikipedia.org/wiki/Dimensionality_reduction
- [3] <https://www.javascript.com/>
- [4] <https://nodejs.org/it/>
- [5] <https://www.mongodb.com>
- [6] <https://d3js.org/>
- [7] https://en.wikipedia.org/wiki/Choropleth_map
- [8] https://en.wikipedia.org/wiki/Bar_chart
- [9] https://en.wikipedia.org/wiki/Scatter_plot
- [10] https://en.wikipedia.org/wiki/Radar_chart
- [11] https://en.wikipedia.org/wiki/Line_chart