Visual Analytics tools for Mental Health

Felice Massotti, Daniela Moretti
2022, June

***Abstract*—Mental disorders are health conditions involving changes in emotions, thinking or behavior (or a combination of these). Mental illnesses are associated with distress and/or problems in social, work or family activities. These are very common and each year the number of cases increases. Fortunately, they are treatable and improvement is possible. Very often these disorders are linked to the general condition of the country, so it is certainly useful to identify groups of similar countries that suffer from the same disorders and therefore need a close intervention, in order to improve the situation. A visual analytics is a good tool to support the whole health care system and political decisions.**

***Index Terms*—Mental Health, Visual Analytics, K-Means, socio-demographic**

I. Introduction

Visual analytics has a great potential to support the whole health care system, such as decision support in clinical medicine and also public health sector. For this reason in our *MentalHealthAnalyzer* project, we analyze the correlation between health care resources and some socio-demographic status to find common patterns and the behaviors of mental health disorders among the different countries of the world, so that anyone, such as pharmaceutical companies or even charities, has the possibility to intervene in a specific and accurate way.

In particular, we focus on 9 mental health disorders: Depressive, Anxiety, Bipolar, Eating, Schizophrenia, Attention, Conduct, Intellectual Disability and Autism. In order to identify some trends in mental health disorders and to have the possibility to derive useful hypotheses about them, three main categories of variables are studied, namely population ages, genre and the capital expenditure health of GDP. These are analyzed in the context of countries and in range of time from 2010 to 2019.

With regard to support public health professionals, we need to find some visual analytics solutions that are easy to be analyzed and understood. This is achieved thanks to the use of interactive and adaptive visual interfaces that enable filtering to handle the amount of information to be displayed. For this purpose, we realized six types of visualizations (Fig. 1):

- Disorders mapping: to get an initial view of the distribution of these disorders;
- Parallel coordinates: useful to find patterns and trends, analyze disorders for each country and also confront the values over time;
- Bar chart: for comparing the population categories(ages and genre);
- Line chart: for a complete overview of GDP values over the years;
- Scatter Plot: useful to find possible clusters among countries;

II. Related works

In the scientific community, there are many papers that analyze the relationship between health and socio-demographic elements of one or more countries. After reading some of them, we selected some that could fit well with our project, by taking inspiration for the categories and views to be implemented. The first one we took inspiration from was *An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health* [1]. In the latter, five categories are analyzed: chronic conditions, mental health, behavioral habits, preventive health and demographic data over several years to find a correlation among these in the different regions and states of the United States. Considering the topic of our project, we did not find enough informations to cover all categories and therefore, in our analysis, we were able to take into account only few categories. Among these, those from which we have taken inspiration for the demographic character are the population and gender. The first visualization they proposed, is the choropleth map, which shows for different diseases, which are the states that have more cases in different time ranges. Like them, we too have used this type of map, where it is possible to analyze a single disease but in different states of the world. We have extended the capabilities of their visualization by giving the possibility to select and analyze multiple diseases simultaneously through filters and select different ranges with a slider. To find important trends and patterns they looked at the distribution of chronic conditions for both gender and ethnicity. And as these differ greatly on the basis of sex, we also find a strong correlation for mental illnesses. In our case, however, we did not find all the information for each disease, but only the generic percentage of males and females who suffers from one or more diseases within the population. To represent this data, they used barplots and stacked barplots. As in the previous case, we started from this idea but trying to make it more interactive, for example by moving the mouse over a bar, the various links with all the other graphics are highlighted. In addition to them, we have decided to distinguish the population based on age. Then, we used stacked bar plots to show the different age ranges in the population, and according to them, it is possible to sort the bars. The idea of taking these data came from the paper *A framework for identifying similarities among countries to improve cross-national comparisons of health systems* [2]. In the latter, a geo-spatial analysis is carried out to find the relationships and differences that exist between the different nations through a Multivariate statistical analysis using PCA and K-Means. For the analytical part, we tried in a similar way to identify clusters using and applying the K-Means algorithm, however, to carry out dimensionality reduction, we initially tried also to use PCA, but we saw that this was not suitable for our dataset, so after several tests, we saw that t-SNE fit better. In this paper, the identified clusters are displayed on the map, we also have decided to show them on the map, not changing their color but lighting them when zoomed on a cluster, of our interest, in the scatter plot. Another paper that inspired us was *Spatial analysis of COVID-19 and socio-economic factors in Sri Lanka* [3]. This paper like the previous one suggested the idea of identifying clusters using K-Means. Moreover, the paper highlighted another important feature, namely the economy and showed how this too can affect when a socio-demographic analysis is carried out, in relation to the distribution of a disease. For the representation of this data, we got the idea from the line chart of this paper: *Visual Analytics in Effects of Gross Domestic Product to Human Immunodeficiency Virus Using Tableau* [4]. Here they analyze the relationships that exist between GDP and HIV/AIDS diseases over a range of years from 2000 to 2017, by showing how in some continents the connection between cases and the economy is correlated. In our case, we did not use the total GDP but only the one concerning health and we did not analyze the continents but every single country.

Given the high multidimensionality of the dataset, we finally decided to use a parallel coordinates and we analyzed this paper *Big Data Visual Analytics with Parallel Coordinates* [5] to color the lines of the same colors as the clusters. Inside, it is explained how this is not a type of visualization accessible to all types of users and sometimes it can be complicated to read. For this reason, a possible solution was to color the lines in this way, showing the relationships captured by dimensionality reduction and trying to understand if some variables have more influence than others.

III. Dataset

The datasets of our project can be divided into two macro-categories: the part relating to mental health data and the part relating to socio-demographic data. For the first, we took the csv called *Number with a mental or neurodevelopmental disorder by type* on the *Our World in Data* website [6]. Inside it, there are the numbers of the population of each country in the world suffering of a certain mental disorder. The categories analyzed are nine and include: `anxiety disorder`, `depressive disorder`, `developmental intellectual disability`, `attention-deficit/hyperactive disorder`, `conduct disorder`, `bipolar disorder`, `autism spectrum disorder`, `schizophrenia` and `eating disorder`. All of these are represented in a range from 1990 to 2019, where each row represents a year and a country.

Two datasets were used to represent the socio-demographic elements. The first was taken from the previous website [6] and is called *Prevalence of mental health disorders in men vs. women*. It contains the percentages of men and women with any mental health or development disability disorder for each country and in a time frame ranging from 1990 to 2019.

The second was taken by *World Bank Data* [7] in the *Health Nutrition and Population Statistic category*. The categories we took are: `Level of current health expenditure expressed as a percentage of GDP`, `total population`, `population ages 00-14`, `population ages 15-64` and `population ages 65 and above`.

Also here it was possible to select countries from all over the world and a particular range of years.

After analyzing all of them, we noticed that some categories were missing a lot of data, which make accurate analysis impossible, so we have decided to remove some of them. The first choice was on the range of years to be considered, the temporal space that went from 1990 to 2009 contained several null values, so we decided to focus on the window that goes from 2010 to 2019. The second choice was to exclude some countries that did not have enough data, and we will find them with grey colour on the map. The final dataset contains 16 columns and in each row, we find a country and a particular year, for a total of 174 countries.

For the creation of the map another dataset was used, this is a GeoJson file taken from the web for the creation of all the paths of the world map.

A small part of preprocessing was done in the analytical part, where for the calculation of the clusters, we decided to remove the column concerning the population as it was a redundant data since it is already included in the columns that represent the ages. For the computation, also the percentage representing the male and female gender was transformed into real values containing exactly the number of people suffering of mental disorders using the population value.

IV. Visualization Techniques

A. World Map

As mentioned before, a GeoJson file was used to create the map with the different paths for each country. Among all the types of maps, we used the *geoNaturalEarth1* because it seemed to us the most intuitive and it maintained the proportions of each continent in the best way.

This type of visualization is called a choropleth map where regions, states or geographical areas are coloured, using

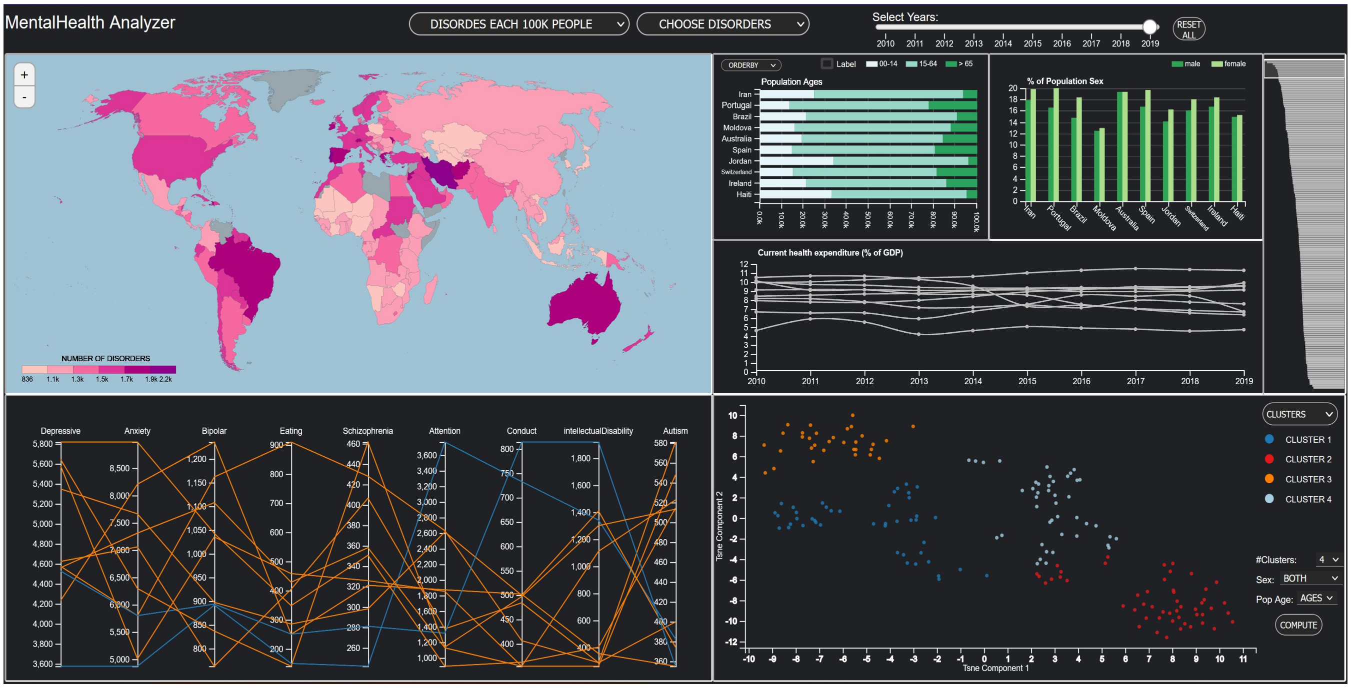

Fig. 1. Mental Health Analyzer.

different intensities, based on the aggregation of some data that fall within certain ranges of values. In our case, the data that we will take into consideration are those of mental disorders. By selecting them from the menu above and taking into account the range of years, they will be added and made an average to take the corresponding value (Fig. 2).

Many are the colours used to represent the different mental disorders, but among all, we decided to take the shade that reaches purple because this is the colour used for the World Mental Health Day. We took the range of tonalities to use from the *Colorbrewer2* website [8].

Passing over each country, it will be illuminated and a tooltip will be shown. It contains specific information about that particular country such as the continent to which it belongs, the population in millions, the percentage of GDP used in medical expenses and the total number of disorders (as mentioned before in the case of several values the average will be expressed).

At the bottom left, there is the legend that explains how the different shades of colour fall into specific ranges (and if the calculation of the data on 100k is not selected but there is the total one, we will have an additional element with a stroke yellow dashed that represents the outliers of the dataset). By passing over it, it will be possible to highlight all the countries with that given colour. The outliers have been excluded from the choropleth map domain and have a specific value reserved for them. In particular, they are countries that have a very large population and for this reason with a proportionately high rate of cases. To identify them, we have to calculate the distribution of the dataset but instead of using the average, we will use the median because it is less sensitive to these and replace the standard deviation with the median absolute distance to the median. From here, we then compare the values obtained with a specific threshold for our case and exclude them from the overall calculation.

It is also possible to zoom and pan, in particular for the first we put buttons at the top left. In this way, we will be able to enlarge some parts of the globe and see even the smallest islands.

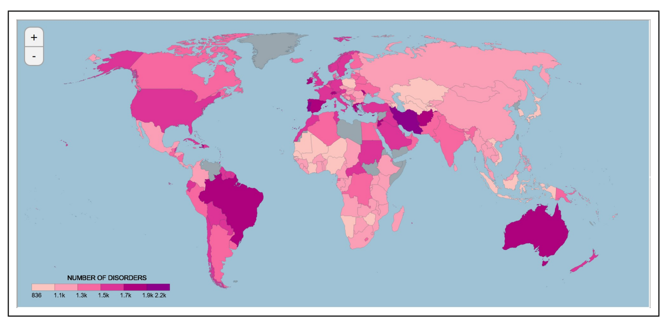

Fig. 2. World map.

B. Bar Chart

The bar charts are the two visualizations used to represent data regarding the three age ranges and the male/female percentages of the population affected by mental disorders. This type of visualization is usually used to compare different categories. In our case, we will compare a maximum of ten/eleven countries to see if we can find distinctions or common elements in the different areas mentioned above (Fig. 3).

For all the information concerning the demographic part,

we have chosen the green color (also in this case the colors were chosen from the *ColorBrewer2* site [8]), to show a consistency between these two macro-categories and a division with the previous data of mental disorders.
Horizontal stacked bar charts were used for three age ranges. Two main cases can be considered: the case in which there are the values on a population of 100k inhabitants and the case in which the whole population are considered.
In the first case, the bars fill all the space and we will be able to compare the portions representing the category from 00 to 14, from 15 to 64 and 65 upwards. Labels will be placed above the bars, and this allows to understand well the different values.
In the second case, the bars will have different lengths, and given the large differences in population among the different countries, it was not possible to keep the labels. For this reason, in order to see the values, a tooltip is implemented that will appear when we move, with the cursor, over them. But in this case, another element it is able to notice and it is the population of each country, which was not possible to see previously.
A select button has been placed at the top left of the visualization, that allow us to sort the countries according to five categories: mental disorder (default case), population, range from 00 to 14, range from 15 to 64 and the range 65 and up.
Instead, to represent the male and female percentages, *vertical multi-series bar charts* were used, again to give the possibility to compare these values for different countries by reading the corresponding values on the y axis.

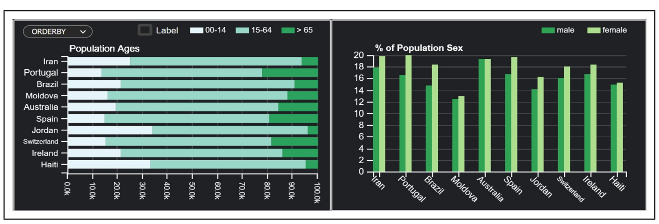

Fig. 3. Stacked Bar Chart and Multi-series Bar Chart.

C. Line Chart

A line chart is a type of visualization that is used very often to show information that changes over time. Inside, it is possible to find lines that will represent the evolution in the range from 2010 to 2019 of the level of current health expenditure expressed as a percentage of GDP for the countries that will be analysed at that particular moment (Fig. 4).
This type of graph helps to find a relationship between two sets of different values where one usually depends on the other. In this case, the independent variable is the range of years and the value of the GDP depends on it.
We choose a grey color for this visualization, in order not to exceed and to avoid confusion with the other graphics.

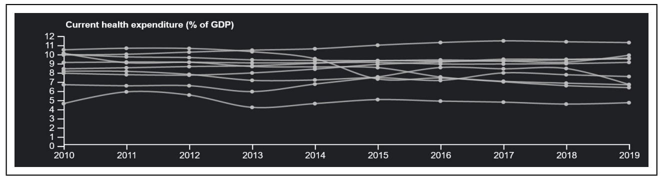

Fig. 4. Line Chart.

D. Parallel Coordinates

The parallel coordinates is used for plotting multivariate, numerical data. They are ideal for comparing many variables together and seeing the relationships between them.
In a Parallel Coordinates Plot, each variable is given its own axis and all the axes are placed in parallel to each other. Each axis can have a different scale, as each variable works off a different unit of measurement, or all the axes can be normalised to keep all the scales uniform. Values are plotted as a series of lines that are connected across all the axes. This means that each line is a collection of points placed on each axis, that have all been connected together.
We have used it to represent the disorders (Fig. 5). However, when the dataset is very large, the lines became too dense. To remedy this limitation, we have used the brushing technique. Brushing highlights allows imposing a range of value for each disorder, highlighting only the countries that respect this constriction. This permits to isolate sections of the plot we're interested in, while filtering out the noise. We choose to represent each line of the countries with the respective color of the cluster that the country belongs to. It is very useful to individuate common trends between the countries inside the same cluster. Another addition, it is the tooltip for providing the name and year of the country selected and also important, is the possibility to highlight all the lines of the same country to confront the different values for each disorder during the years, and so find some useful disorders trends.

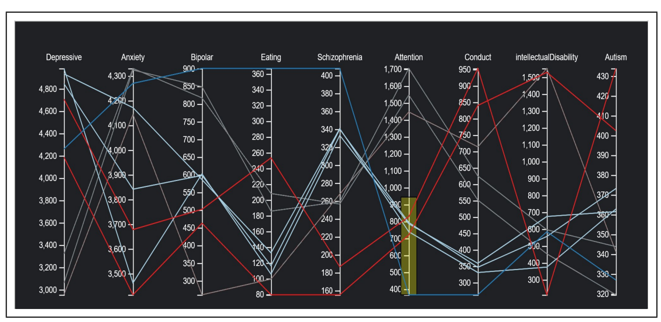

Fig. 5. Parallel Coordinates.

E. Scatter Plot

Scatterplot uses a collection of points placed using Cartesian Coordinates to display values from two variables (Fig. 6). By displaying a variable on each axis, we can detect if a relationship or correlation between the two variables exists. The strength of the correlation can be determined by how closely packed the points are to each other on the graph. We use multiple variables to cluster our data and scatter plots can only display two of them, so we decided to use data reduction method to consolidate our variables into a smaller number of factors. We used t-SNE and it worked very well. The data that we have passed to t-SNE are: Depressive, Anxiety, Bipolar, Eating, Schizophrenia, Attention, Conduct, Intellectual Disability, Autism, population(divided by age), percentage of male/female affected by disorders, Capital Health Expenditure. Then, we have passed the result to K-means algorithm. It is possible to recompute the cluster and each time some values can change, for example, when disorders, years, genre and age range change according to the type of analysis that is taking place. The scatter plot is fully linked to all the others visualizations, giving the possibility to clearly individuate the countries. Another important and useful feature is the zoom. It is possible to zoom into the different clusters, and the slider bar is updated showing only the countries of the selected cluster. At the same time, all the other visualizations change and show the first countries of the updated slide bar. It is a very useful tool because it allows analyzing each cluster in more detail.

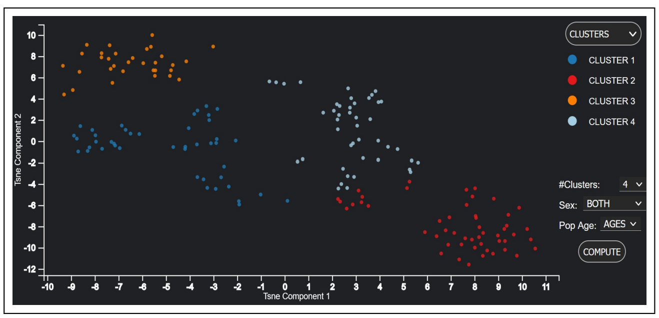

Fig. 6. Scatter Plot.

F. Visualizations coordination and interaction

We give the possibility to the user to choose, through some selection options, the filters to apply according to its kind of study. In particular, the menu is composed of two selection options and a slider. The first selection is used to choose if doing an analysis considering the number of disorders related to a sample of 100.000 citizens for each country, while the second option takes the total population. The second select is used to filter the disorders that we want to analyze, so the clusters are computed based on the checked disorders. The slight bar instead is used to choose the range of years that we want to take into consideration, also here when it changes the clusters are recalculated. Finally, there is a lateral slider bar that is a list of all countries sorted by the sum of all the checked disorders in the selected range of years. It shows 10 countries at time and with a fixed brush is possible to scroll between the countries. According to it, all the other visualization are adapted.

All the visualizations are connected, so when we pass with the mouse over one element the corresponding value will be highlighted in the others. For the map, as we said before, it is possible to pass over each country and a tooltip is shown.

By passing over the legend, it will highlight all the countries with that given colour and by clicking on the outliers rectangle, it allows to include them in the other calculations and visualizations. By clicking again is possible to remove them.

In the case that a user wants to do an analysis on specific countries, it can click on these and they are illuminated. All the other visualizations will be reset and only those countries will be shown. By clicking again, they will be deleted and excluded from the analysis.

For the stacked bars, labels are visible when the 100k analysis is selected and a tooltip in the other case. With the select button, we can order this visualization for certain values.

By passing over the different lines of the line chart it is possible to see the reference country and passing over the circles the specific value for that year of that country is shown through a tooltip.

The brush of the parallel is used for selecting some countries that satisfy a specific range of value for each disorder. In particular, when it is applied all the other visualization are adapted, showing only the resulted countries from the brush, lowing the opacity of the other.

Another important interaction is the zoom on the scatter plot. When we zoom on some specific cluster, the lateral slide bar is updated showing only the list of all the countries that belong to that cluster. As consequence, all the other visualizations show only the countries of the selected cluster and also in this case by clicking on these, a more specific analysis can be performed.

V. Analytics

For the analytical part, we have tried to create clusters that group different countries, to find similarities and differences that can help the user analysis.

In order to do this, the first important thing, given the large amount of data, is dimensionality reduction through t-SNE and then applies K-Means to them.

This whole part was created on *PythonAnywhere* a web hosting service (PaaS) which allows to develop and run a Python back-end service that is active always. The service implements RESTful API which, through POST and GET calls, returns the data elaborated. To develop it, a Flask framework was used. The URL for the calls is: `http://serfelix.pythonanywhere.com/clusters` which for the get method takes default values, while the post method takes in input these data: the number of clusters you want to create, a variable sets to 0 or 1 to calculate the values in 100k or on the entire population, a list of countries on which the algorithm will apply the clusterization, and two other lists that will contain the columns that we want to take into consideration for the diseases and for the socio-demographic aspects. The result of this call returns a json response containing for each country: the code, two values of the t-SNE and the number of the cluster to which the country belongs.

Through pandas dataframe and the data taken as input, the rows of the dataset are aggregated by year and made on it an average. It the end, before passing them to the algorithm, the columns that contained the code and the name of the country, the years and the population (data redundant which is already contained in the sum of the values of the three age ranges) are eliminated. Within the python file, the sklearn libraries are used to perform dimensionality reduction and the clustering algorithm.

For the first phase have been applied and tested two methods: the first was *principal component analysis* (PCA) and the second **T-distributed Stochastic Neighbor Embedding** (t-SNE). One of the major differences between PCA and t-SNE is that the latter preserves only local similarities whereas PCA preserves large pairwise distance maximize variance. Then, t-SNE is a non-linear method that will try to preserve the local neighbours by trying to optimize a stress function. All this, allows to have a greater separation between the different samples and to create more defined clusters, which made us prefer one over the other.

For the creation of the clusters is been used the well-known K-Means algorithm, but instead of the classic algorithm, we have used **K-Means++**. The main difference between the two lies in the initial choice of centroids around which the clustering takes place. K-Means++ removes the drawback of K-Means which is it is dependent on initialization of centroids. So, this version guarantees a more intelligent introduction of them and improves the nature of the clustering, by starting with the allocation of one cluster center randomly and then searching for other centers given the first one. Leaving the initialization of the mean points, the K-Means++ algorithm is more or less the same as the conventional K-Means.

The disadvantage of this algorithm, is that we need to previously choose the number of clusters we want to obtain. To find the optimal one, it is used the Elbow method and the result obtained is 4, which is why we set it by default. But the user can decide according to his needs to change it up to a minimum of 3 to 6 clusters.

VI. Insights

Our visualization has been designed to have multiple targets that can go from simple journalistic information, research fields, to help in political and economic choices and it can create awareness in this field, which has grown especially in recent years and was not given much importance before.

A. Pharmaceutical Company

One of the main targets of our visualization tool was the pharmaceutical companies. Suppose a pharmaceutical company wants to create and develop medicine for a particular mental disorder such as depression, anxiety, schizophrenia, and bipolarity. From the menu at the top, it can select the diseases concerned and start analyzing which countries suffer most from that disorders using the map and the slider on the right. We can see in the parallel coordinates that the countries, that have greater cases of anxiety and consequently almost directly proportional to depression, are also those that have greater differences between men/women, such as Portugal, Iran and Brazil. Especially in the latter two, it is possible to notice that they have a high rate of children and young people, so these will probably be the most affected groups. We can see that by moving the slider of the years and looking at the GDP, many countries that have increased the percentage of expenses have not had great improvements, so we can deduce that the expenses have not concentrated in this field. On the contrary of what we would expect, it is precisely the countries that have a high percentage of GDP that have the highest cases rate in the orange cluster. Based on these considerations, the company could decide to produce a medicine that is perhaps purely female, especially for cases of depression and anxiety for a youth group, choosing whether to invest in a country like Iran or Brazil in which the GDP, in the last range of years, was increasing.

B. Charitable associations

Another target was charitable associations (a general overview of this case is in the Fig. 7). Starting from the analysis of each cluster, it's possible to note that the GDP has a strong influence on how the countries are grouped, in fact countries with the lowest GDP are all grouped together such as Central Africa. According to the correlation between each disorder, it is possible to make a division among them, one of them is by selecting the group formed by attention, conduct, intellectual disability and autism. By scrolling a little through them, we note that Africa (plus Iran and India) are the countries that suffer most from problems of conduct, intellectual disability and autism. Suppose we want to invest starting from the analysis of the GDP and the poorest countries, in these another thing to see is a very high rate of young people while instead, the differences between male and female are not too high. So it can decide to invest more in these fields.

C. World or Continental organizations

A reporter wants to publish an article for the World Health Organization(WHO) to document the mental situation of women in the last year before the diffusion of the Covid, so she selects only the year 2019 and the genre female. However, she is not particularly interested in women over 65, as she believes that the women most

Fig. 7. Case B : Charitable associations.

affected are the younger ones, like girls that due to the Covid situation will attend online lessons and women not too old in age, who may have a more active and hectic private and working life than an elderly woman. So, she first checks all disorders, then in the cluster section she checks the range 0-14 and 15-64 years. In order to have a more balanced study, she wants to see the disorders over 100k citizens. At this point, she decides that the most important disorders for her are Depressive, Anxiety and Bipolar, so now she uses the selection to check only these three disorders. As result, she obtains a list of the top 10 countries where the women more suffer from these disorders, the first 5 are:

- Portugal: with Depressive: 5800, Anxiety: 9000, Bipolar: 1050, and current health expenditure of GDP is 9.53
- Iran: with Depressive: 5380, Anxiety: 7700, Bipolar: 900, and current health expenditure of GDP is 6.71
- Brazil: with Depressive: 4200, Anxiety: 6200, Bipolar: 1300, and current health expenditure of GDP is 9.5
- Switzerland: with Depressive: 4580, Anxiety: 7300, Bipolar: 1100, and current health expenditure of GDP is 11.29
- Greece: with Depressive: 6000, Anxiety: 5600, Bipolar: 1000, and current health expenditure of GDP is 7.84

Before the Covid pandemic, the country where women suffered more from these three disorders is Portugal. In addition from the cluster computation, she can observe that the countries belonging to the cluster 1 are the most affected (a general overview of this case is in the Fig. 8).

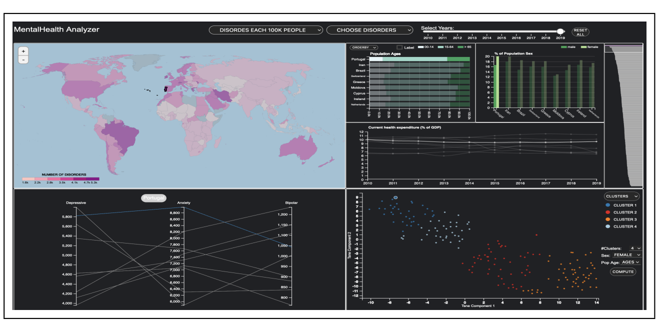

Fig. 8. Case C : World or Continental organizations.

D. Disorders trend over years

We suppose that a researcher wants to identify trends over the years of the countries of all the world, for example to help the local health authorities to focus more only on certain disorders. So, first he decides to check all the disorders, and for having a more balanced study, he selects the disorders over 100k citizens. To analyze the evolution of these disorders over time he selects the range from 2010 to 2019. At this point, four clusters are formed, which are mostly based on the number of disorders but what provides more information for this study is the parallel coordinates. As usual from the slidebar, he can see the list of countries sorted by the sum of disorders over the years. Analyzing one country a time, year by year, he notes that there are some trends, for example in cluster 2 there is the Portugal, which over the years, always maintains its bad trend with medium-high peaks for depression and anxiety. Although for example Portugal, but also Brazil, which belongs to the same cluster, are among the countries with most cases. It is possible to note that all disorders have decreased from 2010 to 2019, except for depression and anxiety. Therefore, the researcher can exclude disorders, that despite being large values, are decreasing over time and above all identify the disorders on which it would be possible to act as being related, intervention on one would also act passively on the other.

VII. Conclusion

The presented Mental Health analyzer allows thanks to various interactive visualizations to analyze the situation and evolution of the mental health disorders in the world, choosing particular case of study through the various selection option. In particular, it is possible to analyze the correlation between countries, finding some possible patterns and also see the situation in a specific country, individuating trend among the different disorders both from the spatial point of view and from the temporal one. The applications of this tool are very large, it can be lent for various purposes, for example by providing useful information to pharmaceutical companies to beneficiaries, or it is possible to do various researches.

A. Future works

For new future works, new analysis can be done considering new aspects. For example, if the site Our World Data will update the dataset with the last 2 years, it will be possible to analyze how the situation is changed after the Covid pandemic. New characteristics could also be considered: for example, the specific number of cases for each age group will certainly lead to more defined clusters and make it easier to identify a possible correlation between the countries of each cluster. Another important characteristic cited in different papers could be the educational level of the country. It has been shown that the level of instruction, which is closely related to education, also influences to some extent the different mental disorders.

References

[1] Raghupathi, Wullianallur, and Viju Raghupathi. 2018. "An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health" International Journal of Environmental Research and Public Health 15, no. 3: 431. https://doi.org/10.3390/ijerph15030431

[2] David T. Bauer, Carl F. Ameringer. 2010. "A framework for identifying similarities among countries to improve cross-national comparisons of health systems". Health Place. Volume 16. Issue 6,. ISSN 1353-8292. https://doi.org/10.1016/j.healthplace.2010.07.004.

[3] Perera, Rumali Weligampola, Harshana Marikkar, Umar Sritharan, Suren Godaliyadda, G M Roshan Ekanayake, Mervyn Parakrama Herath, Vijitha Rathnayake, Anuruddhika Dharmaratne, Samath. 2021. "Spatial analysis of COVID-19 and socio-economic factors in Sri Lanka". 444-449. arXiv:2108.05651v1

[4] Kim, Eunbi Huang, Ching-Yu. 2021. Visual Analytics in Effects of Gross Domestic Product to Human Immunodeficiency Virus Using Tableau. International Journal of Machine Learning and Computing. 11. 219-223. http://www.ijmlc.org/vol11/1038-IJMLC-604.pdf

[5] J. Heinrich and B. Broeksema. 2015. "Big Data Visual Analytics with Parallel Coordinates," 2015 Big Data Visual Analytics (BDVA). pp. 1-2. doi: 10.1109/BDVA.2015.7314286

[6] OurWorldinData.org[online].Available:https://ourworldindata.org/mental-health.

[7] WorldBankData.org[online].Available:https://data.worldbank.org/.

[8] ColorRrewer2.org[Online]. Available: https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3.