HD-Vis: A visual analysis of global Human Development

Lorenzo Bianchi - Jessica Brandi - Gianluca Circelli

February 20, 2023

1 Introduction

In this project we want to give a visual analysis on the global distribution of human development. For this purpose, we serve ourselves of some of the most widely used indices for this kind of analysis: the *human development indices*. They have been developed from the *Human Development Programme* [1] and are often used as a quantitative measure for the quality of life of a country. In particular, in this project we will often refer to the *Human Development Index* (HDI). The HDI is a widely accepted and applied indicator of progress and well-being. The key factor of choosing HDI as an indicator of development, is that it is able to not completely focus on the economic growth, but rather to include other important aspects of human life, such as health and education. For this reason, it can also be an accurate estimation of the quality of life of a country. The HDI is the result of a composition of four main sub-indicators: *Life Expectancy at Birth, Expected years of schooling, Mean years of schooling* and *Gross National Income per capita*. These sub-indicators resemble three main dimensions: Health, Education and Economy. The main indicators that contribute to compose the index are shown in Fig.1.

Using *HD_Vis*, our visual analytics system for this study, can be an effective tool to extract interesting insights on the distribution of human development. We start our analysis by identifying a possible target user for our system and asking ourselves how he can benefit from our tool. In our mind, the user of *HD_Vis* can possibly be someone who is planning to move abroad and is looking for which countries better suit its needs in terms of quality of life, from an high level perspective. For this task, HDI provides a useful snapshot of a countrys economic and social development. Moreover, we can identify another category of users, they may be researchers or development organizations. In fact, the HDI provides a composite measure for human development, which is useful in monitoring progress and comparing development across countries. useful snapshot of a countrys economic and social development.

2 Related Works

In the scientific community, there is great field of research that studies, at global scale, the impact of HDI and other composite indices with respect to various socio-demographic factors.

In the field of human development research, one of the most active organizations is the *United Nations Development Programme* (UNDP). The UNDP has published various papers and reports analyzing many aspects of human development across the world. By reviewing a selection of these papers, we can gain important insights into the key themes and priorities in this field. UNDP publish every year a global report on the study of Human Development, in particular we focused on *"Human Development Indices and Indicators: 2018 Statistical Update"* by the UNDP [2]. This paper provides an general overview of the human development trends and statistics across the world. It includes updated rankings and scores for the Human Development Index (HDI), which measures countries progress in achieving high levels of education, income, and health. Overall, the UNDP papers provide valuable insights into the key priorities and challenges in the field of human development. They demonstrate the importance of considering multiple dimensions of human development beside the traditional indicators such as

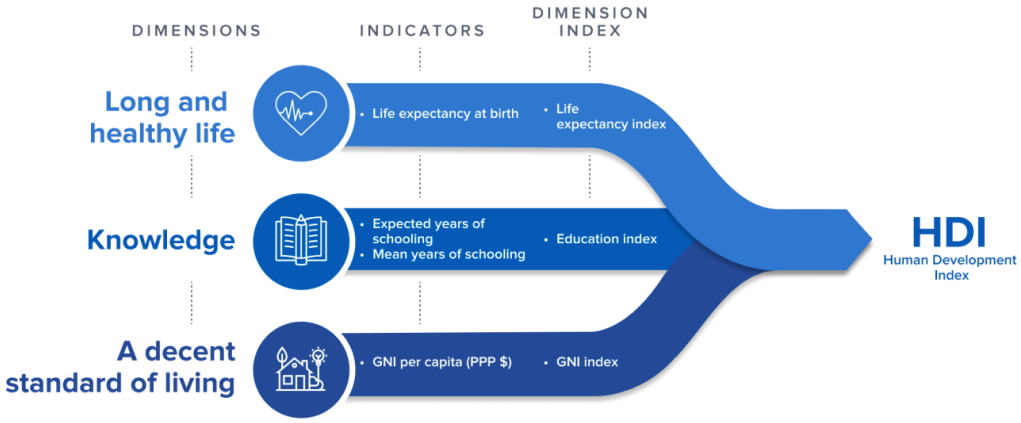

Figure 1: HDI Dimensions and Indicators

economic incomes.

Our analysis, do not limits to the HDI index, in fact in the field of human development research, another the key concern is the issue of inequality. The HDI itself does not take into account the distribution of wealth and resources within a country. To address this issue, the *United Nations Development Programme* (UNDP) has developed the *Inequality-Adjusted Human Development Index* (IHDI). One paper that provides insights into the IHDI is *"The inequality-adjusted human development index: A constructive proposal"* [3]. This paper provides includes updated rankings and scores for both the HDI and the IHDI, that we will use for our work. The IHDI adjusts the HDI to account for the distribution of education, income, and health within a country. By accounting for inequality, the IHDI provides a more accurate measure of a country's development than the HDI alone. The 2019 update shows that while many countries have high HDI scores, the IHDI reveals significant disparities in access to education, income, and health within those countries. By using the IHDI, we can identify areas where policy interventions could be most effective in promoting a more equitable and sustainable future.

Finally, we compute an extra analysis over the *Gender Development Index* (GDI), it is a key tool for measuring gender gaps in human development outcomes. One paper that provides importante insights into the GDI is *"The Uses and Misuses of the Gender related Development Index and Gender Empowerment Measure: A Review of the Literature"* [4]. This paper presents the latest GDI scores for countries around the world, as well as other gender-related indicators such as the Gender Inequality Index (GII). The GDI is a valuable tool for identifying areas where gender disparities persist, and for tracking progress over time towards a more equitable and inclusive society.

3 Dataset

The dataset we use for the system is taken directly from the official website of UNDP [1] (https://hdr.undp.org/).

Before feeding our system with the data, we needed to do some pre-processing over the data. Starting with refining the attributes name in order to manipulate them easily and arrange the data based on their referring year. Moreover we needed to deal with null values, we opted for a fair solution that is to assign to each null entry the mean value of its corresponding column.

The complete data can be summarized in three macro-categories referring to the following composite indexes:

- `Human Development Index (HDI)`
- `Inequality-Adjusted HDI (IHDI)`
- `Gender Development Index (GDI)`

As shown in Fig.1, these composite indexes are computed starting from other indicators. IHDI is composed in a similar way to HDI

ISO3	Country	Human_Development_Groups	UNDP_Developing_Regions	HDI_Rank	GII_Rank	Human_Development_Index	Life_Expectancy_at_Birth	Expected_Years_of_Schooling
AFG	Afghanistan	Low	SA	180.0	167.0	456	61.4191	9.278809547
AGO	Angola	Medium	SSA	148.0	136.0	526	57.5961	9.560079575
ALB	Albania	High	ECA	67.0	39.0	766	78.0919	13.74845028
AND	Andorra	Very High		40.0	85.37647058823529	849	82.8975	11.67192
ARE	United Arab Emirates	Very High	AS	26.0	11.0	0.84	78.5168	12.82703864
ARG	Argentina	Very High	LAC	47.0	69.0	841	76.124	17.17383003
ARM	Armenia	High	ECA	85.0	53.0	0.75	73.3045	13.12232971
ATG	Antigua and Barbuda	High	LAC	71.0	85.37647058823529	783	77.1038	14.69311047
AUS	Australia	Very High		5.0	19.0	926	82.1336	22.47087818
AUT	Austria	Very High		25.0	12.0	905	80.8285	16.0272007

Figure 2: First rows and columns of the dataset

but considering also the inequality condition in the indexes. Same reasoning for GDI but with gender distinction. All the sub-indexes that contributes to our analysis are:

```
'Life_Expectancy_at_Birth'
'Expected_Years_of_Schooling'
'Mean_Years_of_Schooling'
'Gross_National_Income_Per_Capita'
'HDI_female'
'HDI_male'
'Coefficient_of_human_inequality'
'Overall_loss_%'
'Inequality_in_life_expectancy'
'Inequality_in_eduation'
'Inequality_in_income'
```

These indexes are associated to most of the world countries, in particular we consider in our analysis 195 different of them, from every continent. We have gathered data about the index we mentioned associated to the countries by conducting also a temporal analysis over them. Thus, we found such data for every year since 1990. However, this resulted in a dataset to big to handle and to compute at runtime in our system, thus we decided to drop some less relevant indices and reduce the time range from 2011 to 2021.

Finally, in order to add some geographical context to the data we decided to use another dataset that contains all the world countries with the correspondent continent and other demographic data like population. We will use these extra information for finding some insight about the relation between the distribution of countries development in the various continents.

In the end, the complete dataset will contain 171 columns and 196 rows, thus resulting with a final $AS\ Index = 33516$.

4 Visualizations

4.1 Parallel Coordinates

Figure 4: Parallel Coordinates

Parallel coordinates is a technique for visualizing multidimensional data through parallel axes in a 2D chart. In our case, each axis could represent an index contained in our dataset, while the lines correspond to the countries intersecting each axis at its corresponding value. The user can interact with this visualization by brushing one or more columns at the same time. This results in the brushed lines being selected and changing color. This technique is particularly useful for identifying patterns, clusters, or outliers within the data, and can help to highlight any correlations or relationships that may exist.

4.2 Multidimensional Scaling

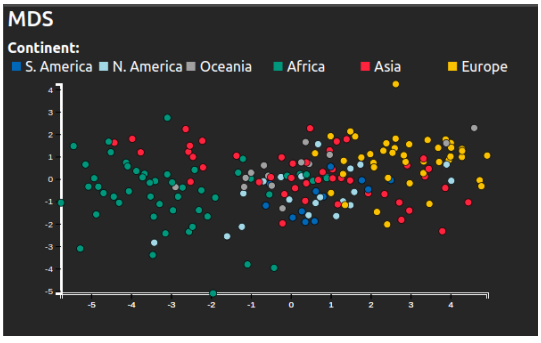

Figure 5: MDS

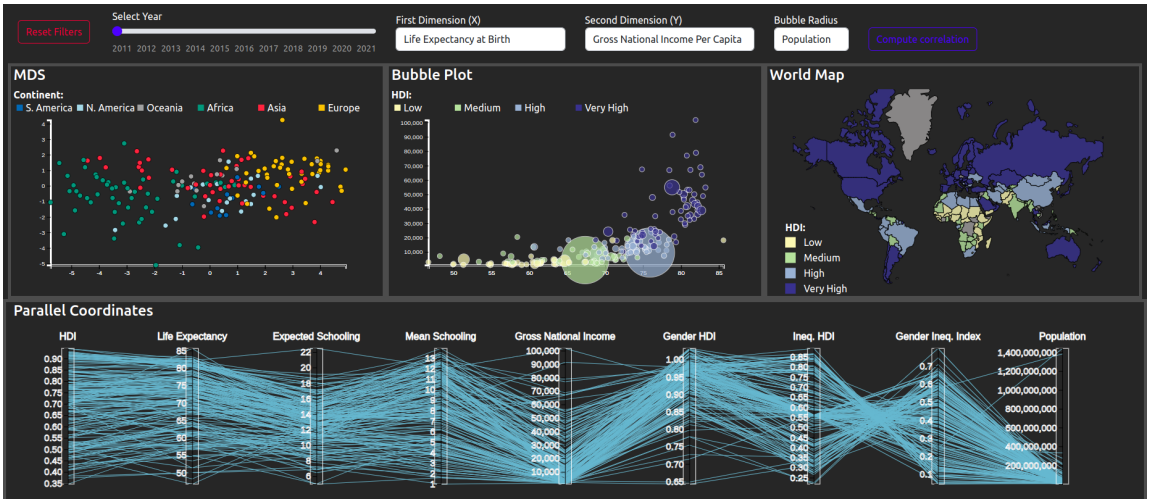

Figure 3: Complete Dashboard

We use MDS for dimensionality reduction, the results are precomputed locally starting from a subset of the attributes and are finally reported in the scatterplot. The dots are encoded and colored with respect to the continents in order to perform an analysis of similarity and dissimilarity between countries in different continents based on the human development status. The scatter plot is useful to find cluster in the data or possible outliers.

4.3 Bubble plot

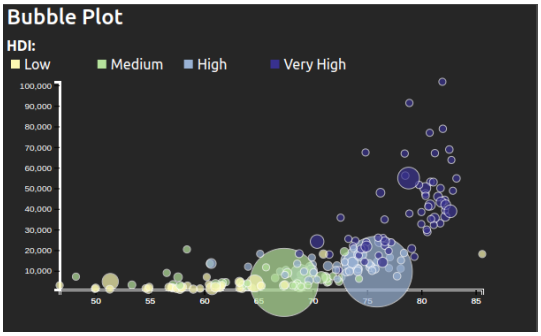

Figure 6: Bubble plot

In the bubble plot we can visualize nearly every attribute present in the dataset. Its possible to choose the attribute to use on both the X and Y axes, as well as the attribute used to determinate the size of the bubble. By default we choose to associate the size of the bubble to the country population, in order to emphasize the relationship between the country demography and the indexes of human development. The colors represent the category of human development groups they belong as shown in the legend.

4.4 World Map

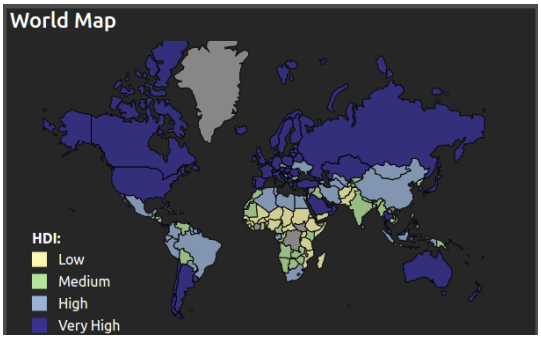

Figure 7: World Map

Provides a valuable geographical context to the data, allowing you to see how human development is distributed across various countries around the world. This visualization can help you identify any geographic trends or patterns in the data and provide additional insights into the underlying factors that contribute to human development.

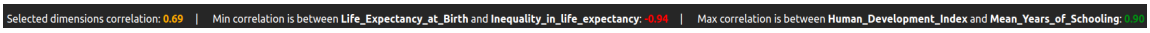

Figure 8: Computed Correlations example

5 Analytics

For the analytic part, we allow the user to do a real time computation on the correlations between the various indices or indicators present in the dataset, for example he/she can compute the *Pearson correlation coefficent* between mean year of schooling and inequality in education, the results is in the range [-1,1]. The higher the more correlated the two indices are. The user can also filter the dataset and based on the filtered data compute the correlation coefficent. Moreover, are also reported the minimum and maximum correlation coefficents, in the current data, from all possible pairs of attributes. The computation start by calculating the covariance between the two attributes using another reduce() method to sum the product of the differences between each element and the corresponding mean. The result is then divided by the length of the array to get the sample covariance. These values are also divided by the length of the array to get the sample standard deviation. Finally, the correlation coefficient is calculated by dividing the covariance by the product of the standard deviations. The correlation coefficient is a measure of the linear relationship between two variables. It ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation.

As other visualization tools, the user can interact with all the system visualizations. For example can use the brushing feature in the parallel coordinates or in the bubble plot to select different different groups of countries. Moreover, the user can select a subset of the data points in one visualization and have those data points highlighted in the other visualizations by brushing and linking the visualizations together. Finally, there is also the possibility to choose the year in which the indexes were computed.

6 Insights

6.1 Correlations

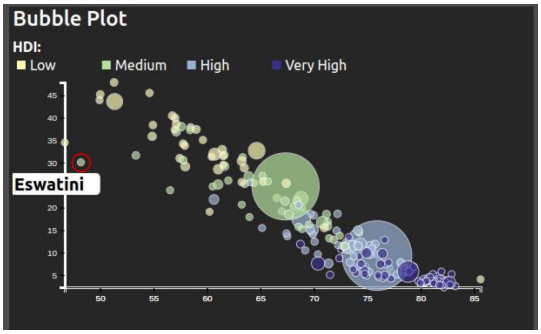

Figure 9: Estawini example

The first insight comes directly from the analytic part, where by computing the correlations coefficients we can extract very interesting insights about the relationship between the various indexes in our data. Very interesting are the minimum and maximum correlation coefficients, that we report after the user clicks the *Compute* button in the top bar. We can immediately see that the minimum correlation is between Life Expectancy at Birth and Inequality in life expectancy with a coefficient of -0.94. suggests that in countries where there is a high level of inequality in life expectancy, certain groups are likely to experience lower life expectancies than others.This means that in countries where there is a high level of inequality in educational opportunities, people are likely to receive lower levels of education on average. As we can see in Fig. 9, most of the countries with an high inequality in life expectancy have also low HDI, as we interact with the bubble plot we can notice that most of them are African countries, however we notice that from 2011 to 2014 Eswatini and Lesotho are the only two countries that do not follow the pattern of the others, in fact as they have the lowest life expectancy at birth their inequality index on this field is not very high, this means that this countries are not so affected by inequality in this field. However this trend change over the years as they get more

closer to the mean distribution, but as the life expectancy increase by almost 10 year from 2011 to 2021, on average the inequality conditions do not improve.

Another example is that Inequality in Education is negatively correlated with Mean Years of Schooling. This indicates that as the level of inequality in education increases, the average years of schooling decrease. Here we noticed that in recent years Egypt is affected more by Inequality in Education with respect to other countries that have the same Mean Year of Schooling like Portugal for example.

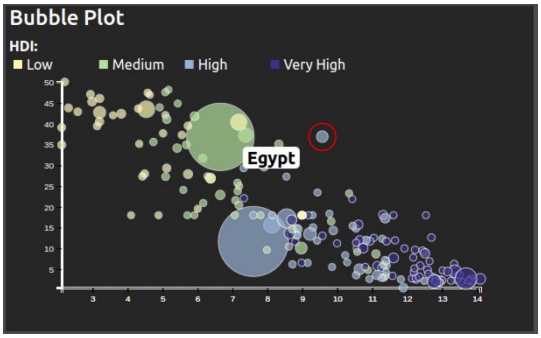

Figure 10: Egypt example

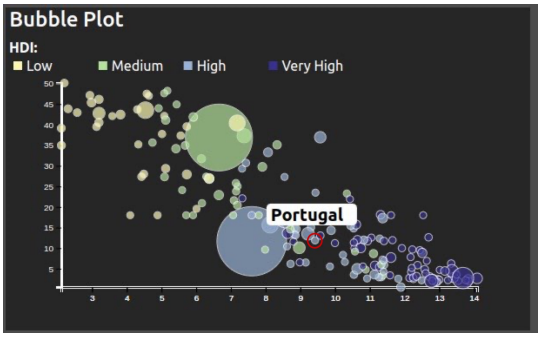

Figure 11: Portugal example

6.2 HDI and Countries population

From the parallel coordinates, we noticed that as we filter countries by lower population, on average, the corresponding HDI gets higher, also the opposite is true. This can also be observed in the bubble plot, as the radius indicates the countries population.

6.3 Continent Analysis

In the MDS plot we choose a color encoding based on the continent of the various countries, this is useful to see how countries of different continents aggregates in clusters, or if there are some outliers. By filtering on the parallel coordinates countries with high HDI, we noticed that there is a big cluster of European countries in the plot. More interesting we notice that in 2021, the only country in Europe that is very similarly correlated to Singapore and Hong Kong is Luxembourg. On the other hand, if we filter by low HDI, countries are mostly Africa, with the excpetion of Yemen, Afghanistan and Pakistan that are very similar to countries like Senegal and Gambia.

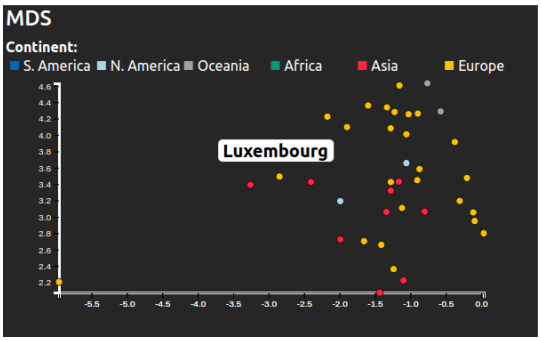

Figure 12: Luxembourg MDS example

6.4 HDI Male vs HDI female

As we have also HDI data with gender distinction, we also find that in the bubble plot the level of correlation is very high, but in Iran we can see that there is an higher development on the Male side with respect to the Female size. But this fact, is unfortunatly well-known, but with this visualization technique is even more clear.

6.5 More insights

There can be plenty of other hidden insights in the data, it is up to the user to inspect the visual tool to study them all!

7 Conclusions

In conclusion, in this project, we have presented a visual analysis of the global distribution of human development using the Human Development Index (HDI), which measures a country's progress in achieving high levels of education, income, and health. We have also explored related works, such as the United Nations Development Programme reports, which provide valuable insights into the key priorities and challenges in the field of human development. Additionally, we have discussed the importance of considering multiple dimensions of human development, including the issue of inequality and the need to account for gender disparities in measuring human development outcomes. Our visual analytics system, *HD_Vis*, can be an effective tool for extracting interesting insights on the distribution of human development and can be useful for different categories of users, including those who are planning to move abroad, researchers, and development organizations. Overall, this project contributes to the ongoing efforts to monitor progress and compare development across countries, and to identify areas where policy interventions could be most effective in promoting a more equitable and sustainable future.

References

[1] UNDP website: *https://hdr.undp.org/*

[2] *"Human Development Indices and Indicators: 2021 Statistical Update"*: This paper provides a recent general overview of the human development trends and statistics across the world.

[3] *The inequality-adjusted human development index: A constructive proposal*: proposes the Inequality-Adjusted Human Development Index (IHDI) as a more comprehensive measure of human development. The HDI fails to consider the distribution of these achievements within a country by not considering the inequality conditions.

[4] *The Uses and Misuses of the Genderrelated Development Index and Gender Empowerment Measure: A Review of the Literature*: The GDI is a valuable tool for identifying areas where gender disparities persist, and for tracking progress over time towards a more equitable and inclusive society.