What Makes Us Healthy?

| Diana Lopes  Instituto Superior Técnico  Lisboa, Portugal  dianamlopes@tecnico.ulisboa.pt | João Costa  Instituto Superior Técnico  Lisboa, Portugal  joao.aparicio@tecnico.ulisboa.pt | Jorge Marques  Instituto Superior Técnico  Lisboa, Portugal  jorgemmarques@tecnico.ulisboa.pt |
| --- | --- | --- |

# INTRODUCTION

Human health quality and life expectancy have been on the rise since humans started investigating and learning more about science and our bodies. Particularly in the last 100 years our life expectancy has skyrocketed. Still there is still a noticeable between different countries.

This project was developed with the objective of discovering which habits, in each country, influence our health metrics the most. For this, we are going to focus on what can, potentially, affect our health. The study will be relative to OECD countries. In terms of definitions, we will be using several indicators relative to the percentage of the population aged 15 and older, such as overweight or obese population, smokes daily, alcohol consumption, deaths by cancer, suicide rates, life expectancy at birth and age 65.

On the other hand, the variables used to correlate to the health will be work (based on wages, employment rates and hours worked), air pollution (exposure), adult education level, country’s social spending and GDP.   
We hope to find what factors are more relevant in terms of what constitutes a healthy life. This visualization will be entitled: “What makes us healthy?”

At first, we had 6 questions we wanted our visualization to answer:

1. Does a better wage mean a healthier life or a longer life expectancy?
2. What is the optimal number of hours to work that lead to a healthier life or more life expectancy?
3. What is the relationship between, more people working and being healthier and live more?
4. How does adult education influence our health?
5. How does the Air and GHG emissions affect our health? (Based on life expectancy and suicide.)
6. Does a bigger social spending in general influences people to live more and suicide, smoke and drink less?

But finished with these 7 questions:

1. Does a better wage mean a healthier life or a longer life expectancy?
2. What is the optimal number of hours to work that lead to a healthier life or more life expectancy?
3. What is the relationship between, more people working and being healthier and live more?
4. How does adult education influence our health?
5. How does air exposure affect our health? (Based on life expectancy and suicide.)
6. Does a bigger social spending or a higher GDP in general influences people to live more and suicide, smoke and drink less?
7. How does self-reported happiness corelates to Alcohol consumption, Smoking habits and Suicide rates?

It’s possible to notice that question 5 was changed. This was due to the need to adapt our visualization to the data we had available. This is better explained section of implementation details

We also added 1 question (7) in order to have a more complete visualization.

# Related work

At first, we had quite a hard time choosing our theme. We noticed there are a lot of visualization and articles about health factors around the world in the internet. Mostly about life expectancy. However, not many try to explain or show if these metrics are related to the habits in each country.

Next, we enumerate some websites from where we got inspiration:

1. Life expectancy by Mark Roser [1]

We find this article interesting because it talks about the evolution of life expectancy around the world along the years. But like it’s said in the beginning of this section it doesn’t relate this evolution with people habits. So, we we’re inspired to do more. On top of this, it has some well put together visualization from where we got some inspiration. Mainly for our choropleth map.

1. Health Data [2]

In this website we found numerous visualizations and data about health around the world. Here we mainly got to see different ways of showing health related data.

1. OECD [3]

Finally, this was where we got most of our data. The reason why we chose to include this website in this section was because, on top of all the data available in this website, there we also found some visualizations of said data. So, we were not only getting the data but also an idea on how to represent it on our visualization. Overall, we consider this website the main contributor and source of inspiration of our work.

# The data

**Where did you get your data from?**

Our project is about health. Mainly health metrics and potential health influencers. We consider potential health influencers different habits the population of a country has. For example, we consider adult education level and average wage potential health influencers.

All this data can be found in the OECD website [3]. In this website you can find vast amounts of data and it even has 1 section dedicated to health stats and health risks.

The data comes in csv format

E.g.

(from “Cancer\_deaths.csv”)

LOCATION; INDICATOR; SUBJECT; MEASURE; FREQUENCY; TIME; Value; Flag Codes

AUS; DEATHCANCER; TOT; 100000PER; A; 1968; 230.6; B

One of the challenges we faced was that it’s much easier to work with json than csv in d3. So we had to change something that looked like the example shown above, to something like the example shown below.

E.g.

{

    "Time": 2010,

    "Var": "Cancer",

    "Average Wages": 0.511975832,

    "Hours worked": 0.264762136,

    "Education": 0.264762136,

    "Pollution": 0.474763931,

    "GDP": 0.045886541,

    "Happiness": 0.096052411,

    "Social spending": 0.206467598,

    "Employment rate": 0.422307046

}

**What challenges did you face?**

**Did you have to correlate sources? Did you have to clean it up?**

**What data did you think you’d get and ended up not finding?**

**Which compromises did you make?**

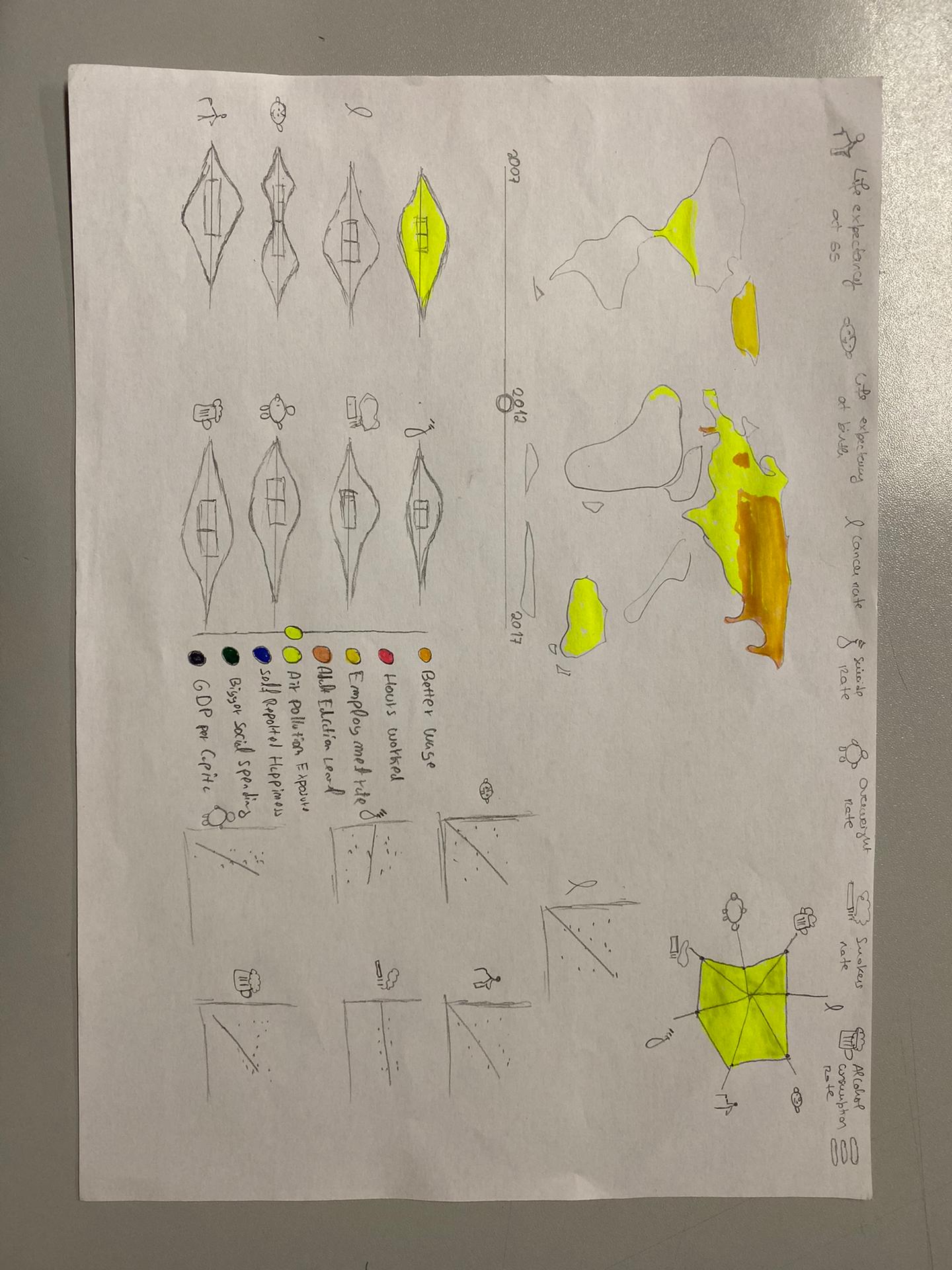
**Also, be sure to mention scalability issues**

# Visualization

## Description

Our visualization has 7 idioms. All of them have interaction between each other. 4 main visualization idioms where the user can see the data being displayed and 3 idioms which only serve as slicers. The 4 main idioms where the user can see the data also have some slicing capabilities.

In figure 1 it’s possible to see the 1st draft of the visualization.



Figure

In order, to better understand the idioms and the interactions they have between each other, in the context of this report, each idiom is numbered and described:

1. Choropleth map

This map will change the color based on the health influencers selected on the slicer described above.

The user can then use the map as a slicer, where he can click a country to highlight the data being shown to that country.

1. Scatter plot

We’ll have 1 scatter plot per each variable of health being analyzed. This means we have 7 different scatter plots as small multiples. On the x axis we have the health factor and on the y axis we have the habit selected on the slicer described above. The point of the country selected on the choropleth map will have higher luminance so the user can identify the country selected and compare it to the rest of the countries. You can also click or hover over a dot to see what country it is.

1. Star plot

In this plot we’ll have the correlation coefficient between each health variable and the variable selected on the slicer.

1. Violin plot

Here once again we’ll have a violin plot per each health variable, which means we’ll have seven different violin plots. Each violin plot will show the distribution of countries in relation with one of the health variables. We will also have an 8th violin plot to understand the evolution of the selected variable, as well as the health factors (by using year slider).

1. Health influencer slider

You can use this to select the health influencer you want to analyze.

1. Health influencer label and Pictogram caption

This serves as a label for the icons close to the plots, so the user can understand which data is in the plots. The Pictogram caption associates the pictograms with their meaning.

1. Years slider

This allows the user to select the year he wants to analyze.

## Rationale

1. Choropleth map

Map -> Color: filling each country with the color of the selected variable.

Channel -> Color -> hue: a higher value for the hue represents a higher value of the selected variable.

Channel -> Color -> lightness: if a point from a country is selected on a scatter plot, the lightness goes up on the map.

1. Scatter plot

Marks -> Point: represents a pair (health variable, health influencer)

Marks -> Line: represents the linear approximation of the scatterplot between the 2 variables.

Channel -> Position: represents the position of the point in a Cartesian axis.

Pictograph -> Represents the respective health variable being correlated with health influencer

1. Star plot

Marks -> Point: represents the correlation coefficient normalized between 0 and 1.

Channel -> Color -> hue: The color of the area between the points in the start chart represents the variable being analyzed.

Pictograph -> Represents the respective health variable being correlated with health influencer

Channel -> Color: color represents the selected variable.

1. Violin plot

Pictograph -> Represents the respective health variable

Curve -> Represents the distribution of each of the health variables and the selected health influencer

Size -> Is proportional to cardinality countries of the value interval

Channel -> Color: the color of the first violin plot represents the selected variable

1. Health influencer slider

Channel -> Position: represents the health influencer being displayed

Channel -> Color: represents the health influencer being displayed

1. Health influencer label

Channel -> Position: represents the year being displayed

Channel -> Color: represents the relation between the color with each influencer (word on its right)

1. Years slider

Channel -> Position: represents the year being displayed  
Channel -> Color: represents the relation between the color with each influencer (word on its right)

1. Health influencer caption

Channel -> Color: represents the relation between the color with each influencer (word on its right)

## Potential

# Implementation details

Completing this project was no easy task. We faced multiple challenges along the way. This section enumerates these challenges and explains how we overcame them.

1. Format the data is stored

At first all our data was in the csv format. We quickly realized it’s much easier to work with json than with csv when using d3. So, we found ourselves with a “useless” data format. This was easily fix, since we only had to transform the csv file into a json file through the use a python script.

1. Data chosen at first was not complete

When we first picked our data, back in checkpoint 1, we didn’t check if all the fields were complete. We later realized that it was not, as is usually expected. This is one of the main lessons learned from doing this project.

To address this problem, we took 2 steps:

* 1. We limited the interval of years being analyzed to the interval we had more data. In this case we limited the data to the years between 2012-2018.
  2. How did we fix missing spots?

1. Data chosen at first had almost 0 values

This problem is similar to the previous one. Why did we choose to split 1 problem into 2? Because, while some data first collected had some missing values, some had almost no values at all. From our point of view this is a completely different problem that required a different solution. So, it’s deserving of it’s on section.

In this case, the data first chosen for the values of self-reported happiness had almost 0 values. So, we had to find a solution. We opted for a simple one, and simply decided to find another source with similar kind of data. More specifically we changed the OECD data source for a …

1. Choropleth map
2. kjhg

# Conclusion

# Future work

.

# References

|  |  |
| --- | --- |
| [1] | [Online]. Available: https://ourworldindata.org/life-expectancy. [Accessed 14 12 2019]. |
| [2] | [Online]. Available: http://www.healthdata.org/results/data-visualizations. [Accessed 16 12 2019]. |
| [3] | [Online]. Available: https://data.oecd.org/. [Accessed 16 12 2019]. |