2019

Data Story

FIGHTING CHRONIC DISEASE – COMBINING SDGS – 3,1,8

PETROV,DIMITAR D.D.

Link: Data Story – “Dimitar Petrov”

https://wg4imqffaiettqbejfzcog-on.drv.tw/Dimitar%20Petrov%20-%20DS-B-44/Data%20Story/Non-communicable%20diseases%20-%20Data%20Story/

Summary

This document describes the process of creating a data story concerning United Nations’ Sustainable Development Goals, to help us understand the broad picture and how to deal with the current world problems and reach the ambitious, however great, targets of creating a better world and ending unnecessary suffering.

Introduction

The first step to building the data story was to find and choose topic. I wanted to research a goal that I did not know a lot about and I would be interested to learn about. I was also looking for one that is of the biggest world problem, extremely important to solve. The last condition I had, in choosing a topic, was the amount of data available. After brief research on data availability I found the following picture, describing the scope of data available and how rich and helpful (colouring) the data is:

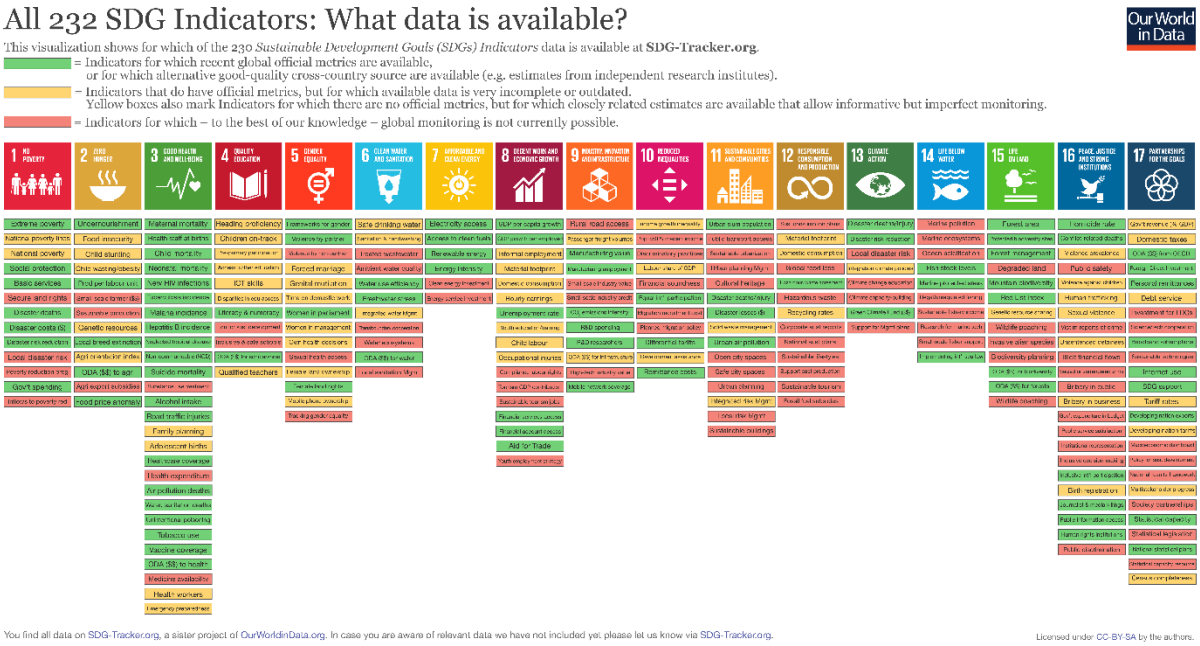


Figure 1 - https://ourworldindata.org/sdg-tracker-update

As I am very interested in health and well-being I was happy to find that there is a large set of data available for the whole spectrum of SDG Goal 3 and started researching the goal more in depth.

As optimistic I was looking at how fast we were solving Poverty during classes, I was surprised to see that things did not go exactly with plan in regards to World Health and well-being.

There are two major types of disease, Communicable (infectious – HIV, 3 types of hepatitis, measles and others) and Non-commutable (non-infectious – Diabetes, Cancer, Heart Disease, Respiratory disease). While both are not meeting the targets, Major advances have been made in combating infectious diseases. There is more than 50% decline in several of those diseases for just 15 years after 2000. And as Sub-Saharan Africa (the most affected part of the world) is being pulled out of poverty and clean sanitation and reproductive health is being improved, the facts suggest promising results.

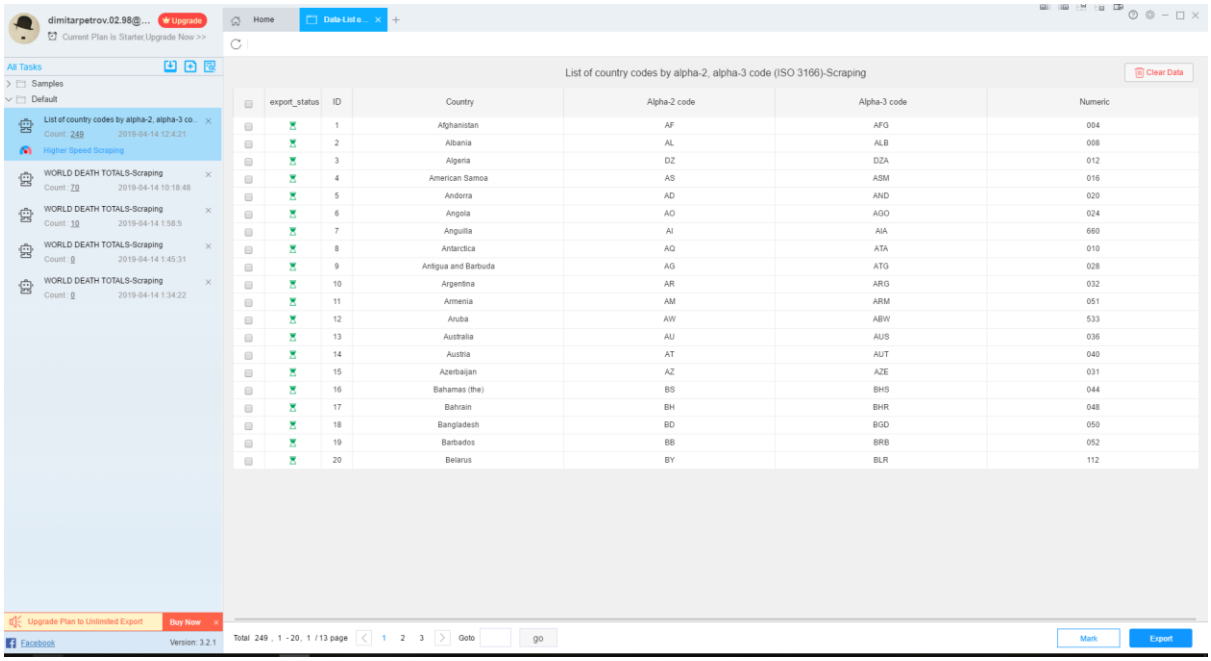
On the other side of the spectrum are Chronic Diseases (NCDs). Those, contrary to communicable diseases, have been increasing (as a total number of people affected). The target from 2000 to 2020 was to have 25% decrease worldwide and now 30% by 2030. The data, however, suggests otherwise, which I will explain in more detail in 'Exploration' paragraph.

Data sources

There were two data sources where I took the majority of data sets I used – The United Nations statistics indicators database (https://unstats.un.org/sdgs/indicators/database), having all the goals, targets and available data sources in a tree-like structure – easy to use and downloadable in excel format directly. The other is an Online Cloud API – (https://unstats.un.org/SDGAPI/swagger/). Here, you put the desired targets datasets as numbers with a comma in a text field and the API returns a JSON file with the issued data.

Another source that help me find very useful information about tobacco and alcohol usage, obesity and inactivity was OurWorldData.org (https://ourworldindata.org). I also found some very useful data about the causes of death worldwide by percentage, which helped me introduce the topic and show the problem in relative to other disease and death causes terms.

A lot of the data had the countries with only their abbreviation which made some of the visualizations not so easy to understand. For that reason, and also to enrich some of the tooltips in my visuals I used a data set having countries names with both their 2-alpha and 3-alpha codes, as well as their 3-digit unique number (https://www.iban.com/country-codes) and used the table as a mediator in many of the visualizations I created. That source did not have an API or download as csv, excel or any other data format option, so I used a very user friendly web scrapper – Scrape storm, to download the data in csv format:



Exploration and Data Manipulation

Exploring and cleaning the data consumed a large proportion of creating the data story. Strangely enough, when downloading any of the data sets in the UN statistics indicators database as a excel format (the only available), they had every field wrapped in parentheses. That made the data set unreadable for Tableu. In order to fix that, I used excel to find and replace all parentheses with empty spaces.

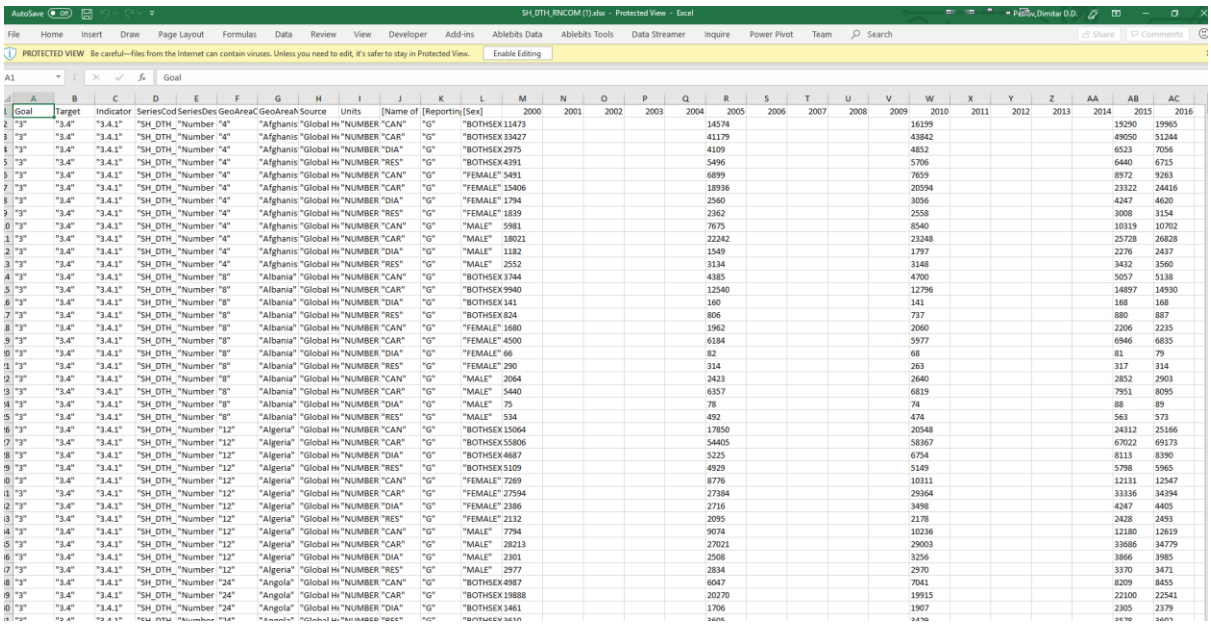


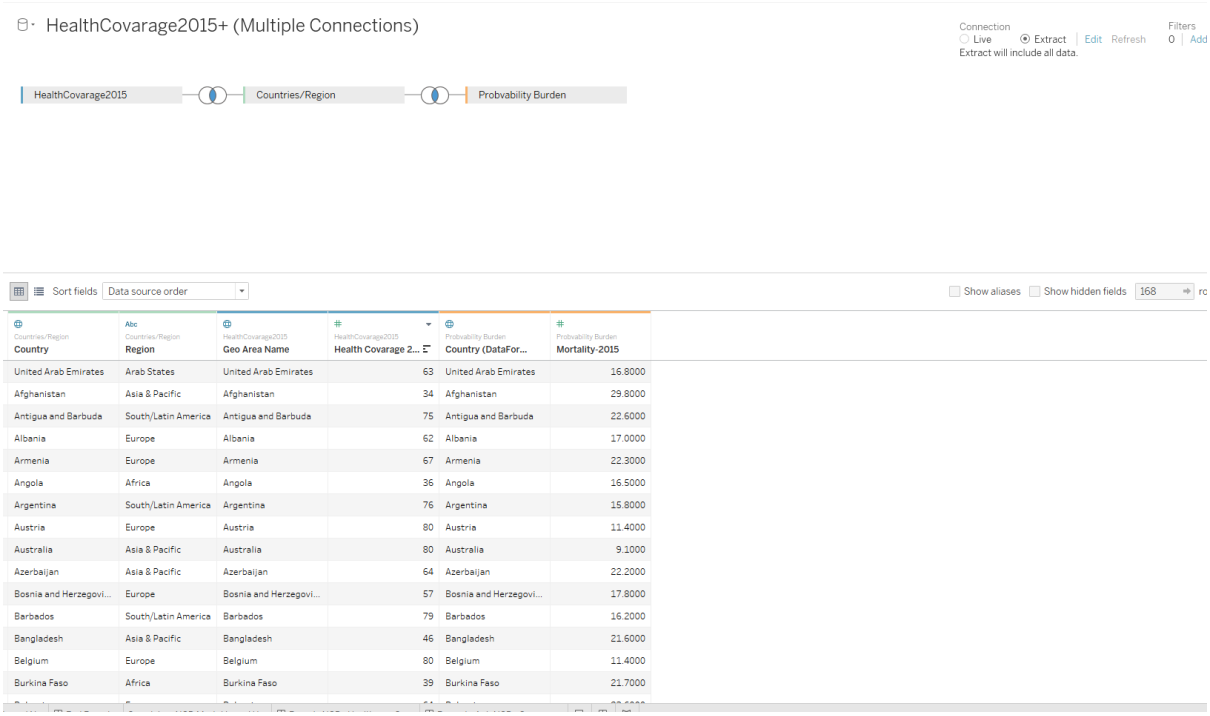
Figure 2

Then I could Import the data sets in Tableu. Now I had to change the data types, because Tableu did not recognize them correctly. The name of the countries I changed to country/region data format, which will later help me create maps and the values I changed to number or decimal format, which I used for aggregations. Many of the columns I deleted since they did not contain data or were irrelevant to what I wanted to present.

If we examine the data set in figure 2, even though there are value with different years, we cannot make a map with a timeline out of it. The reason being that timeline should be based on one column (for example – 'year') and we have all those values and years spread out. To fix that, I used the pivot command in Tableu to unpivot the the columns containing year values.

The second half of my data story presentation included a lot of correlations. That meant that I had to create connections between the Tables. Usually that meant having countries as primary keys, which also meant that I had to previously design the data sets to have 1 row by country (aggregating the data, pivoting it). Or, as in several cases I there was data for each country by sex – I had to delete the rows containing 'male' and 'female' and only keep 'both sexes'.

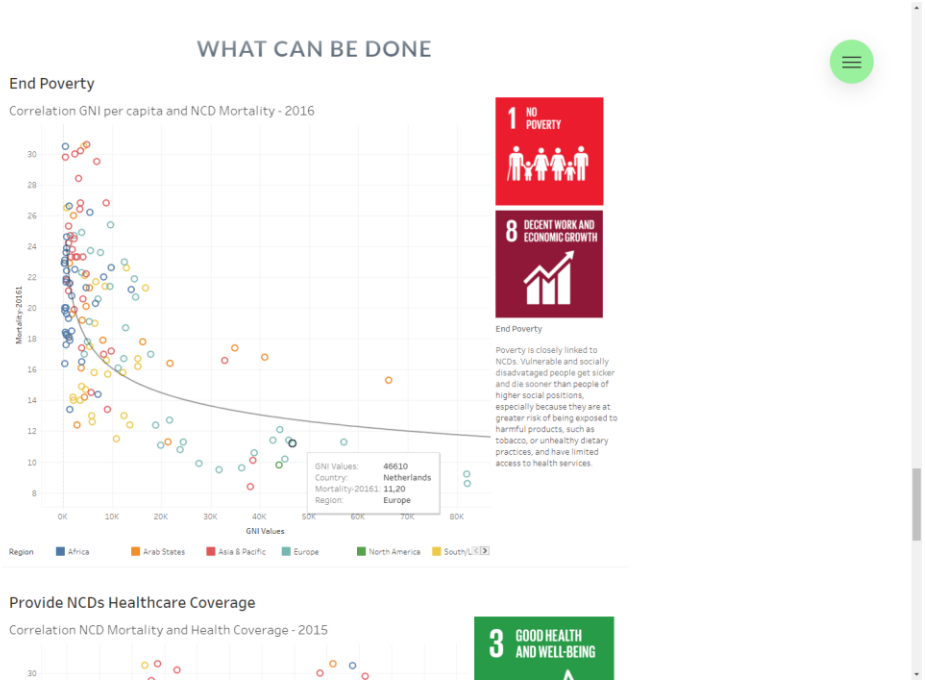
Some datasets had 3 digit or 3 letter abbreviation of the countries and others had the full names of the countries and in order to connect them I used a mediator table – factual table, which had all the Countries, with all their codes – as stated in the "Data Sources" chapter.



Country	Region	Geo Area Name	Health Covarage 2...	Country (DataFor...	Mortality-2015
United Arab Emirates	Arab States	United Arab Emirates	63	United Arab Emirates	16.8000
Afghanistan	Asia & Pacific	Afghanistan	34	Afghanistan	29.8000
Antigua and Barbuda	South/Latin America	Antigua and Barbuda	75	Antigua and Barbuda	22.6000
Albania	Europe	Albania	62	Albania	17.0000
Armenia	Europe	Armenia	67	Armenia	22.3000
Angola	Africa	Angola	36	Angola	16.5000
Argentina	South/Latin America	Argentina	76	Argentina	15.8000
Austria	Europe	Austria	80	Austria	11.4000
Australia	Asia & Pacific	Australia	80	Australia	9.1000
Azerbaijan	Asia & Pacific	Azerbaijan	64	Azerbaijan	22.2000
Bosnia and Herzegovi...	Europe	Bosnia and Herzegovi...	57	Bosnia and Herzegovi...	17.8000
Barbados	South/Latin America	Barbados	79	Barbados	16.2000
Bangladesh	Asia & Pacific	Bangladesh	46	Bangladesh	21.6000
Belgium	Europe	Belgium	80	Belgium	11.4000
Burkina Faso	Africa	Burkina Faso	39	Burkina Faso	21.7000

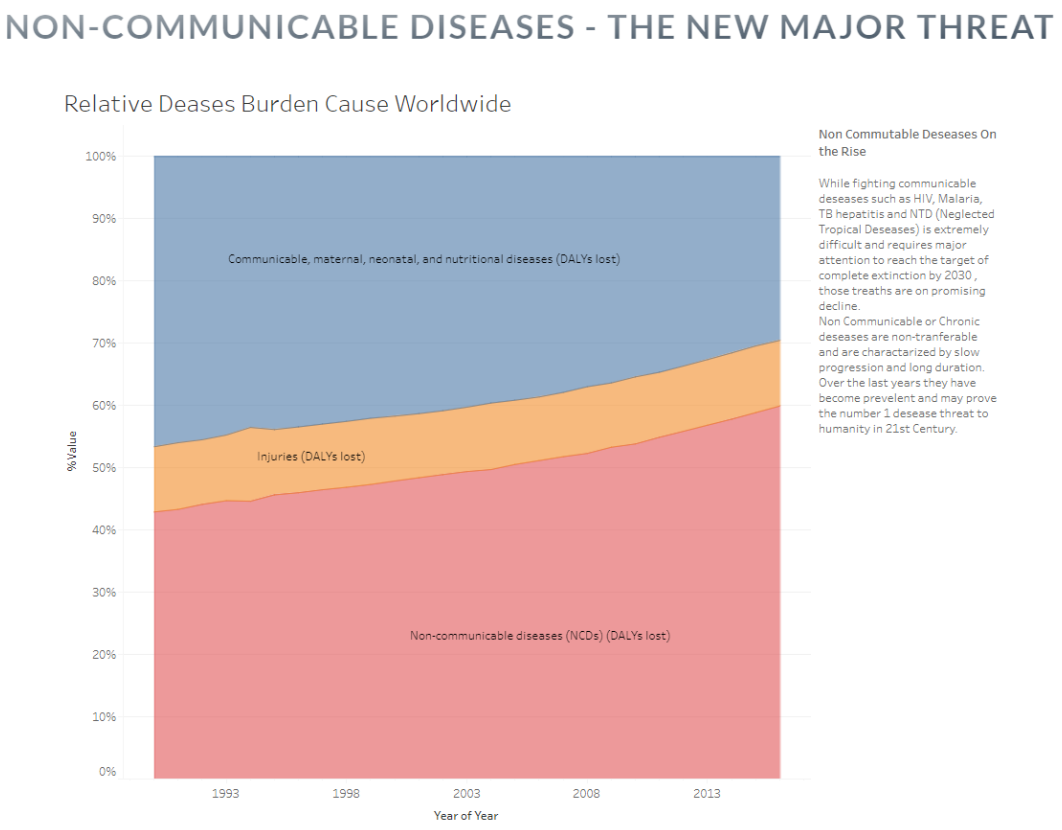
Figure 3

That also helped me include both regions and the countries in those regions, both including them in tooltip and/or painting them in different colors:



Insights

This project was extremely insightful for me. Creating the following visualizations shocked me, showing that half of diseases today are Chronic – extremely high number – on the rise (in relative terms). That was what motivated me to find the answer to why:



The data and visualizations showed me a picture and story I did not expect. To begin with I expected that Chronic disease would be most prevalent and would contribute to most deaths in so called 'rich' or developed regions (Europe and North America), since it's a widespread knowledge that those regions have one of the highest percentages of obesity relatively to other regions. This is also something I show in my story:

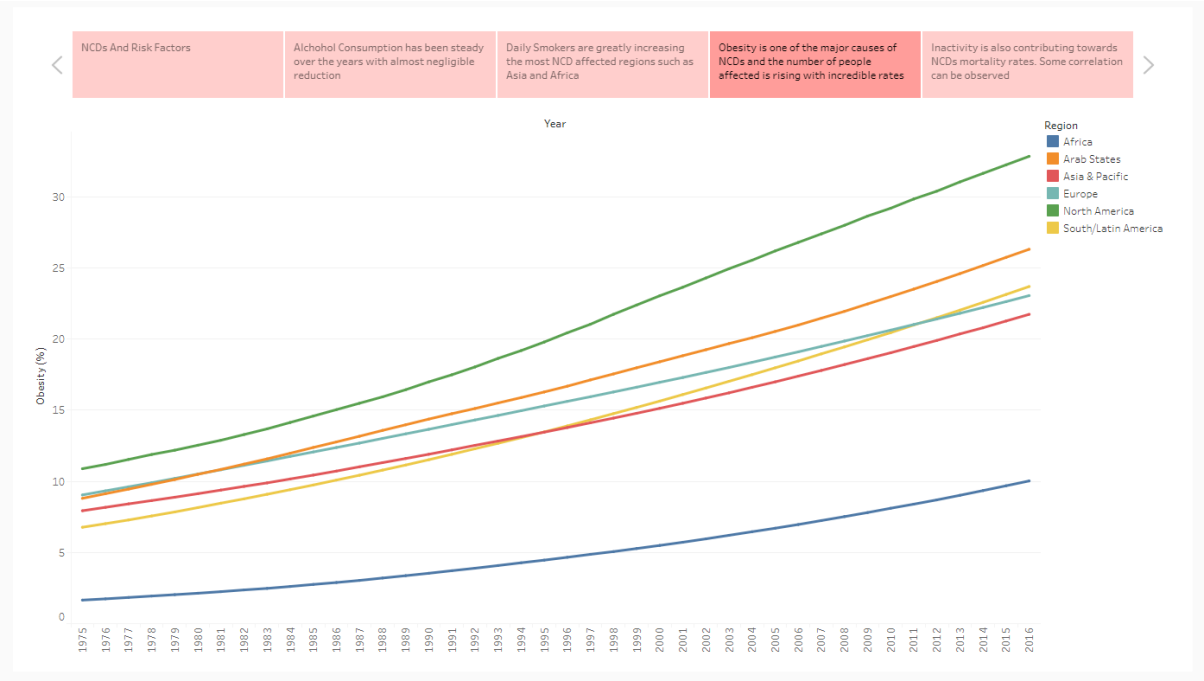
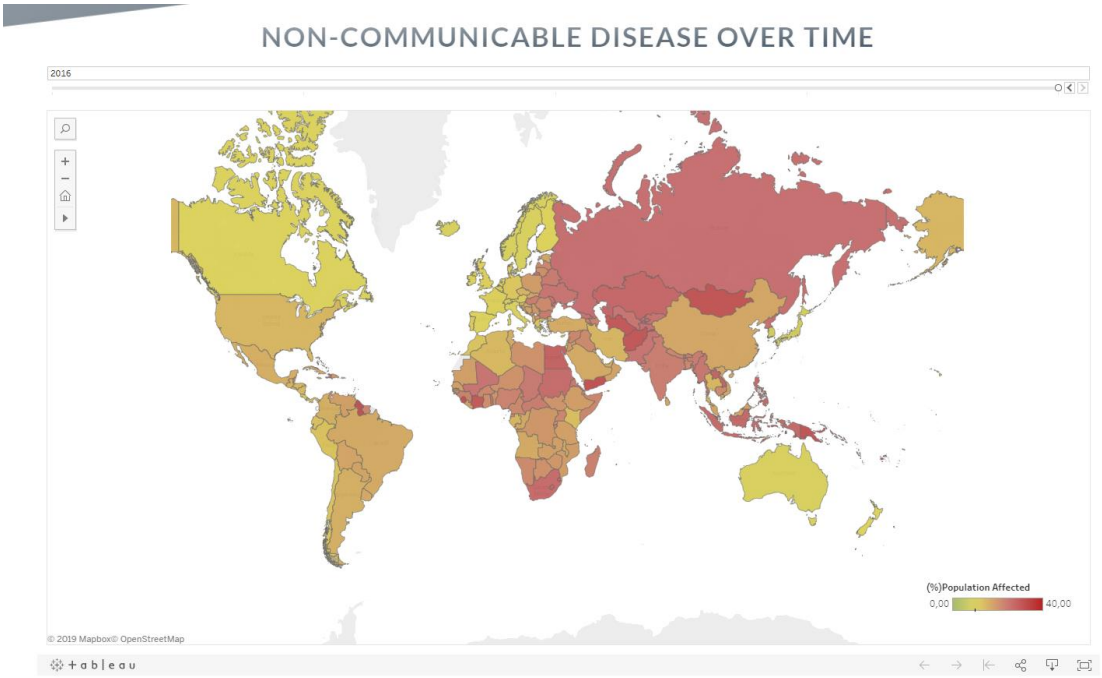


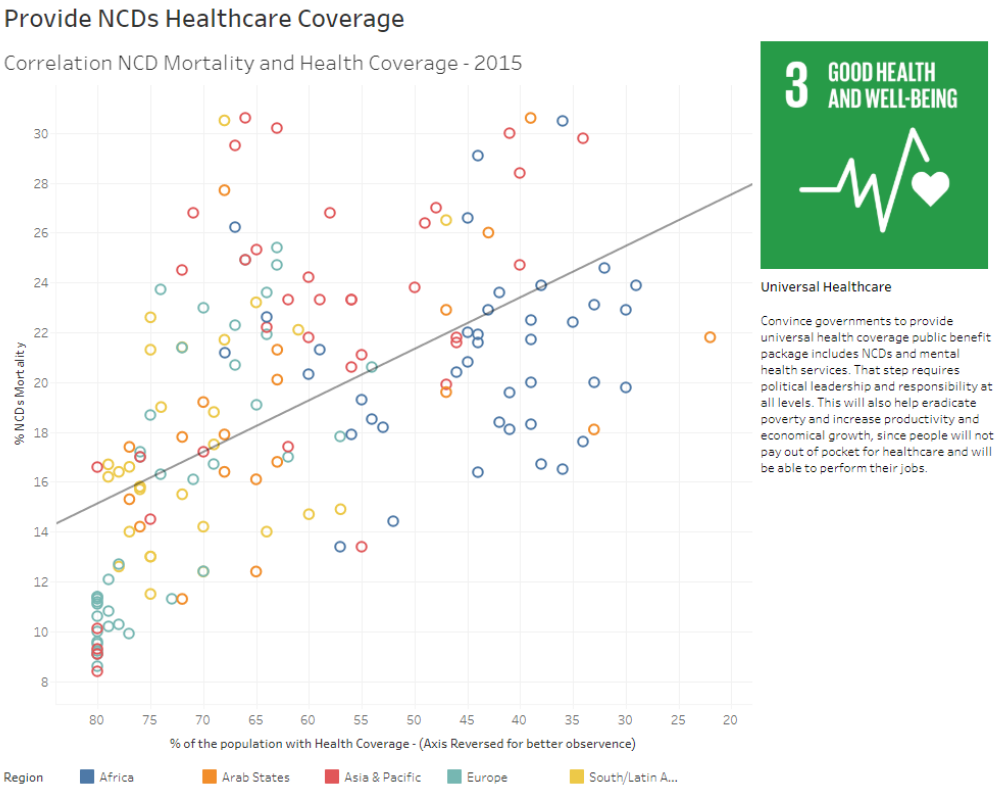
Figure 4

Looking at the world map of NCDs we see a very unexpected picture:

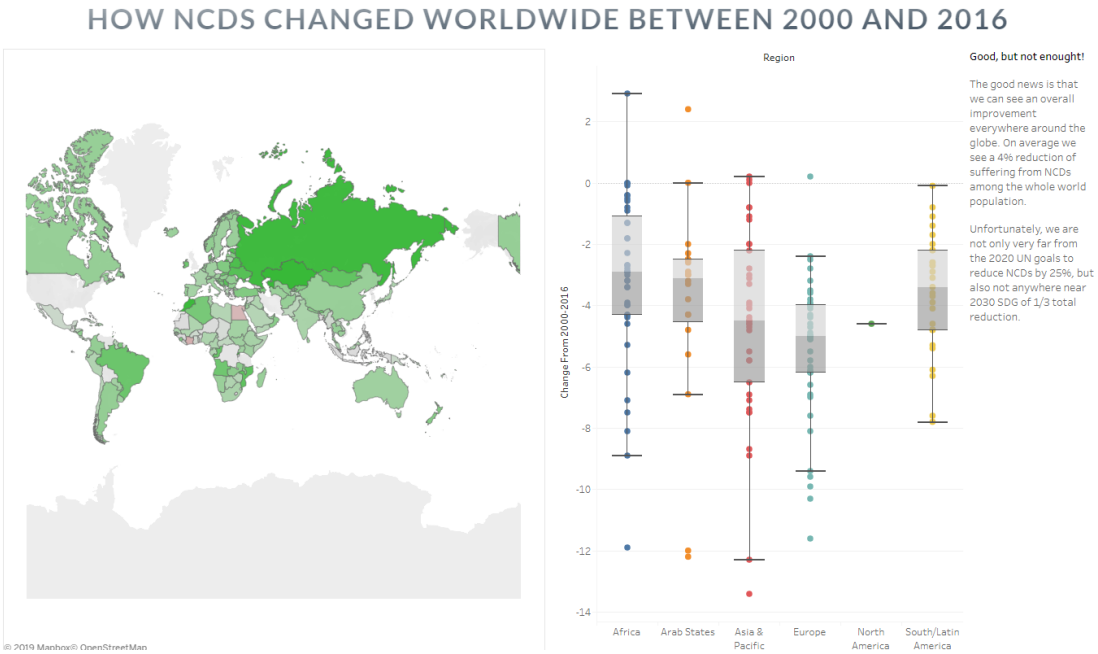


What the map shows is that actually those two regions have the lowest burden today. The most affected regions are South-Saharan Africa and Asia. In some countries going as high as 1/3 people are affected by those diseases (e.g Sierra Leone).

Some of the reason could be clearly seen at the end of my story. Poverty is closely linked to NCD'. Vulnerable and socially disadvantaged people get sicker and die sooner than people of higher social positions, especially because they are at greater risk of being exposed to harmful products, such as tobacco, or unhealthy dietary practices, and have limited access to health services. Better health is also vital for managing those diseases. Healthcare coverage has high correlation with lower mortality rates from Chronic diseases:

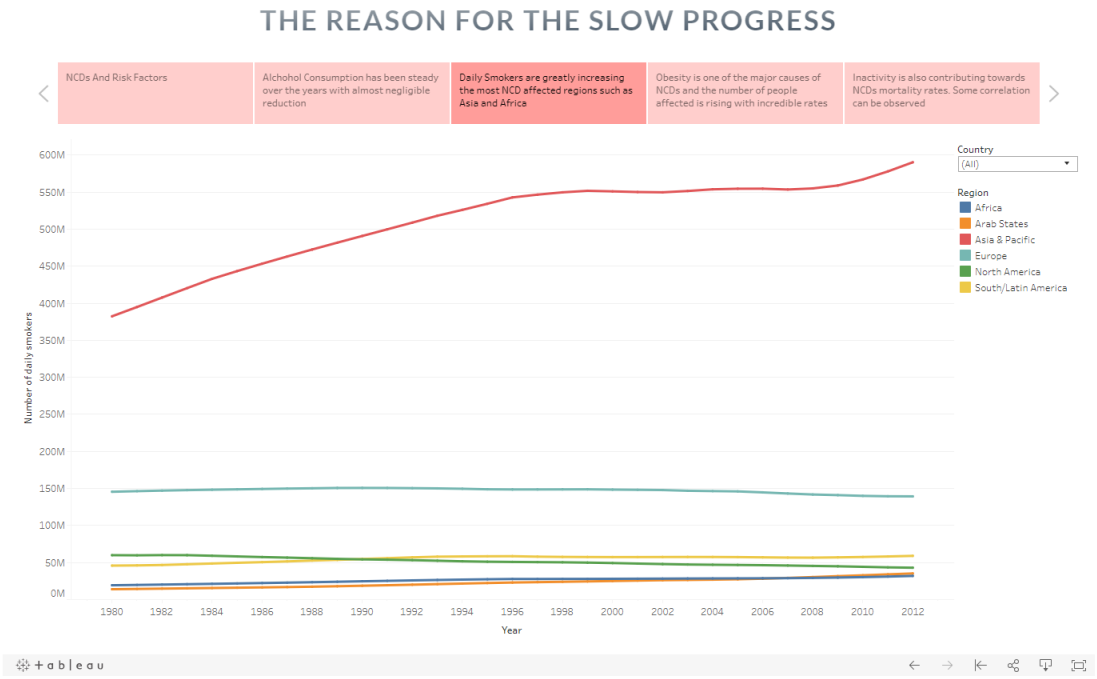


Furthermore, after creating the two visualizations (below) I found that the average decrease of Chronic diseases worldwide, in relative terms, is only 4%:

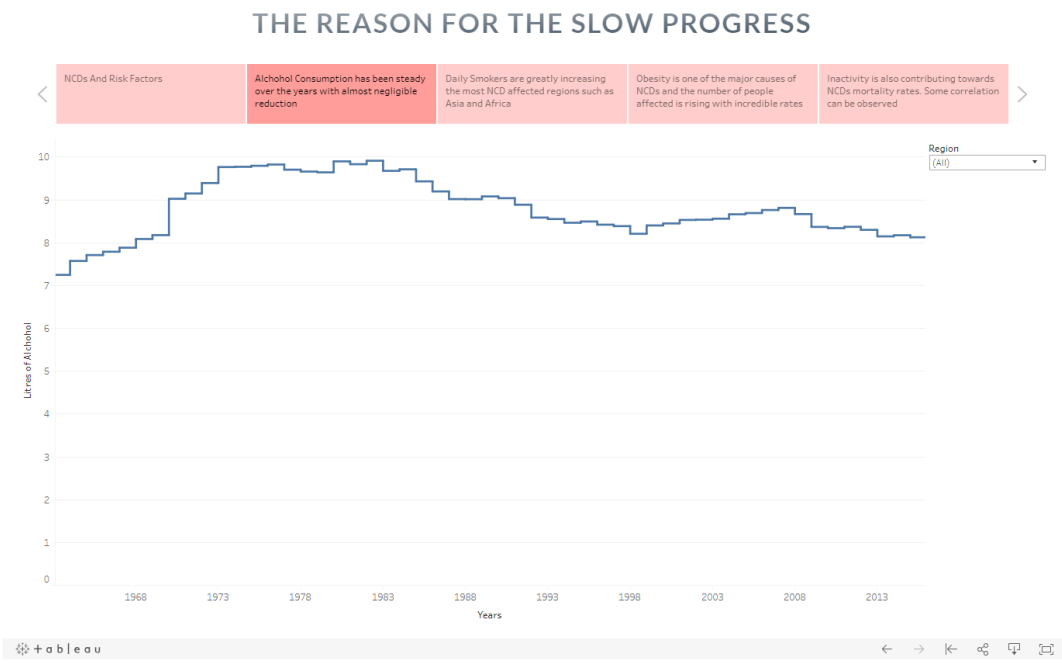


In order to understand why, I looked at all major risk factors of Chronic diseases. there were 3 other major risk factors contributing to NCDs – Alcohol consumption, Tobacco consumption and inactivity. When I looked at the statistics about Alcohol consumption and Tobacco consumption, I realized why we are so far off from the projected goals of UN.

In terms of tobacco consumption, I was stunned to find that daily smokers are increasing with rather high rate. In Asia the number has almost doubled and in the other regions it has either stayed the same or gone slightly up. Only North America and Europe show between 15-20% decrease in total number of smokers for the past 30 years:



Another unfortunate figures show that Alcohol consumption has not changed significantly over the past 50 years:



As an ending statement I propose 3 impactful actions that could significantly decrease the number of people suffering from Chronic diseases:

- End poverty
    - All those diseases are manageable if treated correctly. People in poverty, not only can't afford such medications, but are also more vulnerable and more likely to indulge in risky behaviours, such as smoking or drinking alcohol. As I show in my story – GNI (Gross National Income) is very strongly correlated with NCDs, having R squared value of 0.43. Having a steady GDP growth (connecting this to Goal 8) could greatly help fight those diseases.
- Provide NCDs Healthcare Coverage
    - On average only 10% of people on average (smallest number compared to other regions) in Western European countries carry the burden of NCDs. As I show in my visualization for this point – Health Coverage is also greatly correlated with mortality rates from Chronic diseases. Prompting Governments to design a well-working Healthcare could save millions of lives.
- Promote Anti NCDs Campaign
    - As we saw, The major risk factors for NCDs are behavioural (tobacco, alcohol consumption, obesity and lack of exercise) – meaning that promoting healthy habits by creating large educational campaigns, with the help of governments, can be the number one most important action to be taken.

Design choices

Since in order to make a point and have some conclusion I had to tell a whole story to the viewers of the story. In order to ensure that the reader of my story would actually read my story, I had to came up with a way to catch his attention.

In order to do so, I decided to create a whole one page website, having many interactive elements and interactive visualizations.

Some of the design choices I made for the website included: Vivid colours, sticky elements, Navigation menu, elements animations, mixing two in one element with linear gradient, clipping of elements (images)

Regarding the interactive visualization part, I chose Tableu. It has a very easy to use embed option, great visualizations (such as maps and box plots) and the possibility to create clusters and advanced antilithics represented in a very eye pleasing manner (colours, etc.)

Final Product

The final product was a one page Website – Data Story having the following section and visualizations:

- Navigation menu (light green circle on the top right)
    - Drop-down possibility with link every section
- Home (Topic Introduction)
    - Image (logo) of the main targeted UN SDG (goal 3)
    - Background Image of all SDGs with goal 3 in the middle
    - Heading (Main Topic)
    - Sub-Heading (Sub-Topic)
- Objectives (Sub-targets considered)
    - Goal 3 (3.4/3.5/3.8) + SDG image logo
    - Goal 8 (8.1) + SDG image logo
    - Goal 1 (1.1/1.3) + SDG image logo
- Problem (with NCDs)
    - Area chart (Type of Disease burden/population percentage)
    - Problem definition paragraph (right side)
- Types of NCDs
    - 4 interactive cards
        - Disease type
        - Facts
- NCDs over time
    - Map showing percentage of population affected by country (redder – worse)
    - Timeline Filter (above map)
- Change 2000-2016
    - Connected visualizations
    - Map showing progress (greener – better)
    - Box plots – change per region
        - Each region in different colours
    - Paragraph describing the situation (progress is slower than expected)
- Slow Progress (Story with 5 tabs)
    - Risk factors per disease (Table)
    - Alcohol consumption per region per year (Line chart)
        - Filter for region (single selection or all)
    - Total number of daily smokers per region per year (Line chart)
        - Each region in different colour
        - Filter for country (multiple selection)
    - Obesity (%) per year pear region (Line chart)
        - Each region in different colour

    - Inactivity rate / Total deaths (Scatter Plot - Correlation)
        - Each region in different colour
        - Trend line
- Conclusion (3 suggestion cards)
    - End Poverty
        - Heading (suggestion)
        - Correlation GNI /Mortality – 2016 (Scatter plot)
            - Each region in different colour
            - Trend line
        - Image SDG involved
            - Goal 1 logo
            - Goal 8 logo
        - Paragraph
            - Description
    - Provide NCDs Healthcare Coverage
        - Heading (suggestion)
        - Correlation NCD Mortality and Health Coverage – 2015 (Scatter plot)
        - Image SDG involved
            - Goal 3 logo
        - Paragraph
            - Description
    - Promote Anti NCDs Campaigns
        - Heading (suggestion)
        - Image (UN suggestions)
        - Paragraph
            - Description

Conclusion

Creating a concrete and genuine data story turned out both complex and time demanding endeavour. On the other hand, very rewarding in terms of accomplishment and knowledge. Building the full product required me to expand my field of competence in research, abstract thinking, design patterns, designing website, story flow design, using Cloud services for both accessing and scraping data, Tableu functionalities, website embedding, clustering, advanced analytics and statistics, measures and dimensions, connecting data sources, data cleaning and manipulation.