Computer Science 215 Chetima.I Maina Boukar

Introduction to data science

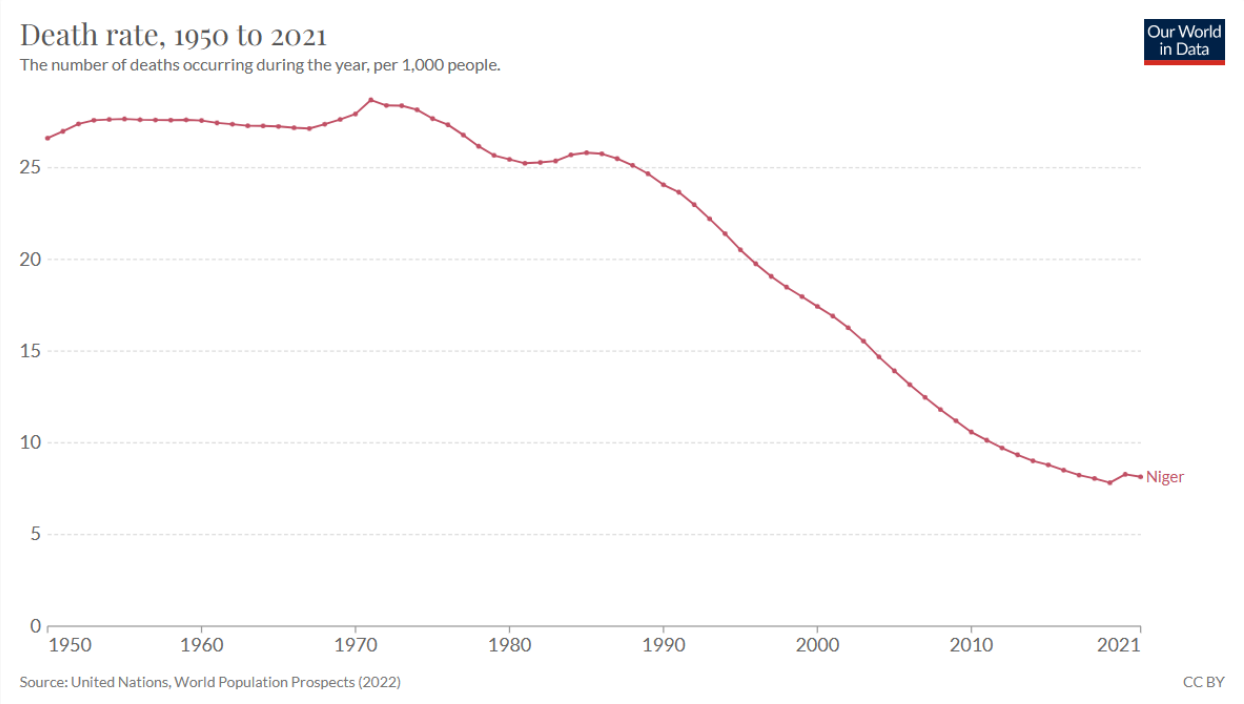
Spring 2023

Prof. Jordan Wirfs-Brock

Data manifesto Chetima:

When I think of data, what comes to mind is anything that could be used to generate information, anything that can provide understanding or meaning to human life. Data can be the number of sugar spoons in our daily breakfast tea, or even the duration of our showers. For the purpose of this paper I will be using the death rate in Niger (West Africa), in order to get my points across the table.

Niger is one the countries with the lowest human development index countries, meaning that its people go through the toughest living conditions. The data provided by "Our world in data", gives an insight upon the death rate (per 1000 people) in Niger over a period of 71 years. In this case, the death rate is our data, it is what is going to be processed in accordance to our will, to generate a specific conclusion. As a reflection of the DIKW pyramid, it can be said that the death rate in Niger collected is an information about a specific aspect of the country and its people -their standard of living, the country's level of safety etc- this information is providing an understanding of the country's state, and this knowledge can be used to take necessary measures to better human life in Niger.



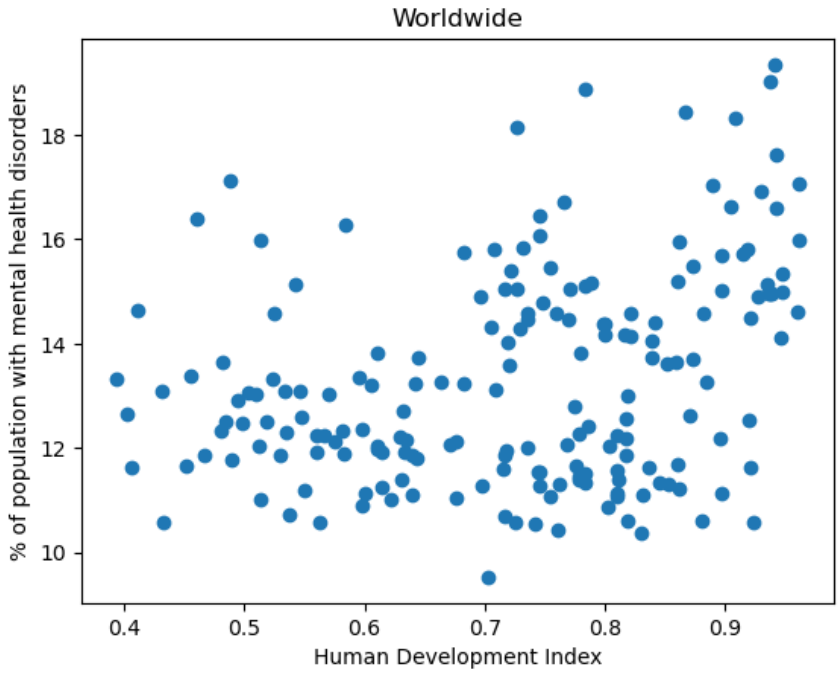
Graph.1

In Graph.1 above, is a display of the variation of death rates in Niger from 1950 to 2021. In the first 20 years, the death rates were revolving around the same range up until it reached its peak in 1971 and witnessed a constant decrease, with some inconsistencies, until 2021, the end of the data collected. What is behind this sudden and persistent drop in the death rates ? What happened during the years with increased death rates ? To what extent do we expect the death rates to diminish ? These are all questions that can be elaborated and studied by data scientists. Those who are able to locate and accurately collect data in order to better represent the truth. We cannot confidently claim that there were in 2021, 8 deaths per 100000 in Niger, for example. A salient number of people might have died and gone unnoticed, be it incidents or even those who passed away in remote areas where there is no existence of a system that thoroughly records the number of deaths. All of these are factors that are to be taken into account by data scientists while studying or working the data. Data scientists are those who are able to better determine the fitting

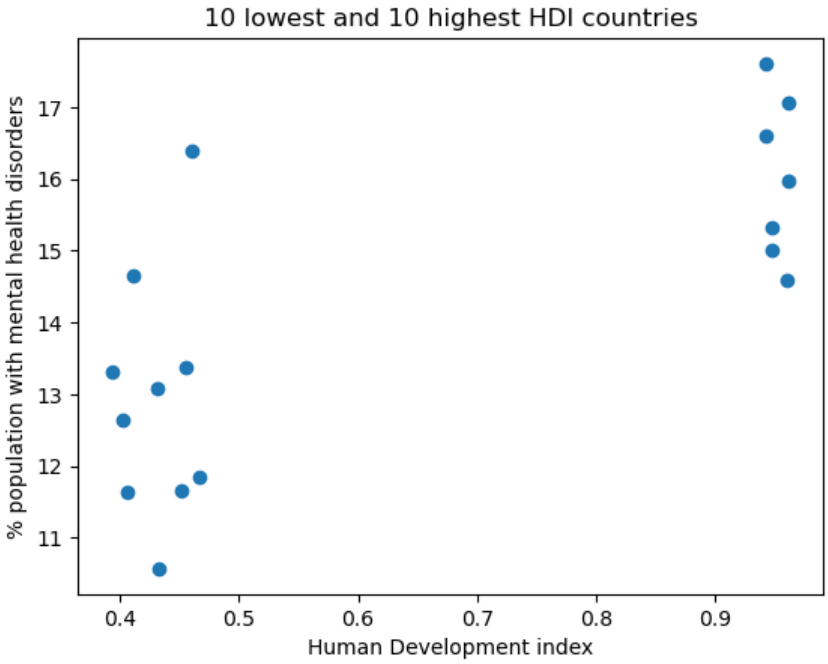
processing and visualization techniques to get the best out of the data, those who are able to extract meaning from the most random of things, in order to provide useful knowledge for the mass. I think that few skills are important in doing data work, having some background in coding can be vital when it comes to analysis and processing data, having some knowledge about how to get datasets from the web, but also knowing to navigate around to be informed about innovations in the field and the works of other data scientists. I think it is important as a data scientist to have that will to provide new knowledge and make it accessible to the world. If I were to imagine myself giving advice to future data scientists, it would tell them to be curious, innovate, try new things, make as many projects as possible because the more you practice, the more familiar you become with this world, do not be scared of venturing into new worlds but also and most importantly, try your best to benefit the world with your work.

Oftentimes the data does not reflect reality. I found on Our World in Data a study about happiness and life satisfaction across the world, in which people were questioned about, and found out that western countries, with greater economic prosperity had a higher happiness and life satisfaction index. This is something that I found strange because I have lived in countries with different economic and political situations and I did not come to the same conclusion. It is also what I want to disprove in my 9th project for this class by studying the correlation between Human Development Index(HDI) and mental health disorders in the world. My results showed that mental health disorders such as depression and anxiety tend to be more prevalent in countries with higher HDI , in other words “developed” countries, as displayed in Graph.2 and Graph.3 below. What I want to state is that it can also be problematic to try to depict reality without looking at all aspects of the matter being studied, which is something that data scientists

need to be meticulous upon. Data science can be a great tool to provide a better understanding of human life and help tackle modern day issues that we face, if utilized in the best of manners.



Graph.2



Graph.3

Sources:

[Population & Demography Data Explorer - Our World in Data]
[Project 9 Chetima - Jupyter Notebook]
[Happiness and Life Satisfaction - Our World in Data]