

N1 : SANOUSI LABO Abdoul Madjid
N2: SIDIBÉ Moussa
Filière : Data Science
Matière : Système d'information Géographique
Professeur : M. Aniss Moumen



Livraison 3:

1. Listing and data source:

Covid19 has affected almost every country in the world, some of which were already experiencing humanitarian crises. Data about the scale, severity and duration of the outbreak is in demand as decision makers and responders seek to mitigate the impact of the virus in developing countries. Throughout our project, we were going to use two different resources:

First Ressources: Data provided by United Nations Office for the Coordination of Humanitarian Affairs (OCHA)

To make it easy to find relevant data, the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) created a dedicated page for covid19 data on their Humanitarian Data Exchange (HDX) platform which brings together data from hundreds of partners. The covid19 pandemic page includes dozens of datasets in machine-readable formats, all of which are also available through the HDX API. Some of the most notable contributions include the INFORM covid19 Risk Index which ranks countries vulnerable to health and humanitarian impacts of covid19; WHO's covid19 case data for China and the rest of the world; UNESCO's global school closures data; and WFP's covid19 global travel restrictions and airline Information. More details about the data : <https://data.humdata.org/organization/inform>

Second Ressources: Use an R resources package

Developed by Rami Krispin, the {coronavirus} package provides a tidy format dataset of the 2019 Novel Coronavirus covid19 epidemic. Pulled from the dataset of John Hopkins, the R package gives a daily summary of the Coronavirus cases by state / province. The data set contains various variables such as confirmed cases, death, and recovered across different countries and states.

Visite this link for more details about the package :

<https://cran.r-project.org/web/packages/coronavirus/index.html>

2. Conceptual data model (CDM):

For a clear study of our data use in this project, it is essential to set up a model of concetion of these data. Via the merise method, we were going to present these data by a data dictionary, a functional dependency graph and an association entity model.

a. Data dictionary:

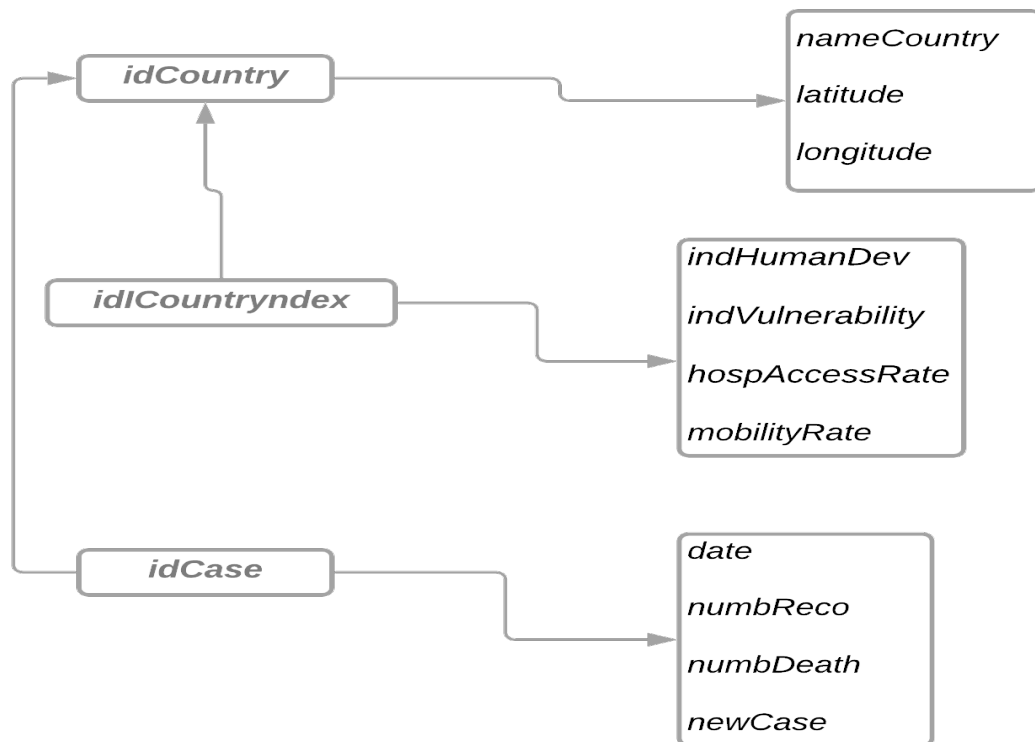
A table which lists all the key words that will intervene in the data. This gives an overview of the variables used. It contains the names of variables that will intervene in the study, each one accompanied by its description and also its type.

Name	Description	Type
idCountry	Country identifier	Int
nameCountry	Country name	Varchar
latitude	The latitude of country	float
longitude	The longitude of country	float
idCase	Cases identifier, by period(day)	Int
date	Date of new cases	Date
numbReco	Number of recover	Int
numbDeath	Number of death	Int
newCase	New case number	Int
idCountryIndex	The identifier of country index's	int
indHumanDev	Index of Human developpement	float
indVulnerability	Index of vulnerability of country	float
hospAccessRate	The hospital access rate	float
mobilityRate	The mobility rate in a country	float

We noticed that country contains even provinces.

b. Fonctional dependency graph :

To provide a good relationship between ours variables, typically between ours primary keys and other non-key attributes within our data conception, we need à Fonctional dependency graph.



c. Diagramme entity association :

Based on functional dependency graph, 4 classes were candidates in the association entity model.

Country: Who gives the information related to all country in the world, like his name, his latitude and longitude ;

Case: Which provides information on the various cases identified for a specific date;

Index: which provides some index information of each country. There are four index: indHumanDev, indVulnerability, hospAccessRate, mobilityRate. Basing on those, we are going to make some analysis that combine differentes cases raised in specific country and all of the index.

The relationships and cardinalities between these differents tables are as follows:

raised to: which links the raised cases to a country. A case is assigned to a single country, but a country can have at least one square.

Characterize by: means that a country is characterize by one single index provided. And conversely, indexcan be linked at maximum one country.

