**API**

World Bank API was used in order to get GDP, Population and Unemployment rates among countries.

Please refer below to see the breakdown of the World Bank API

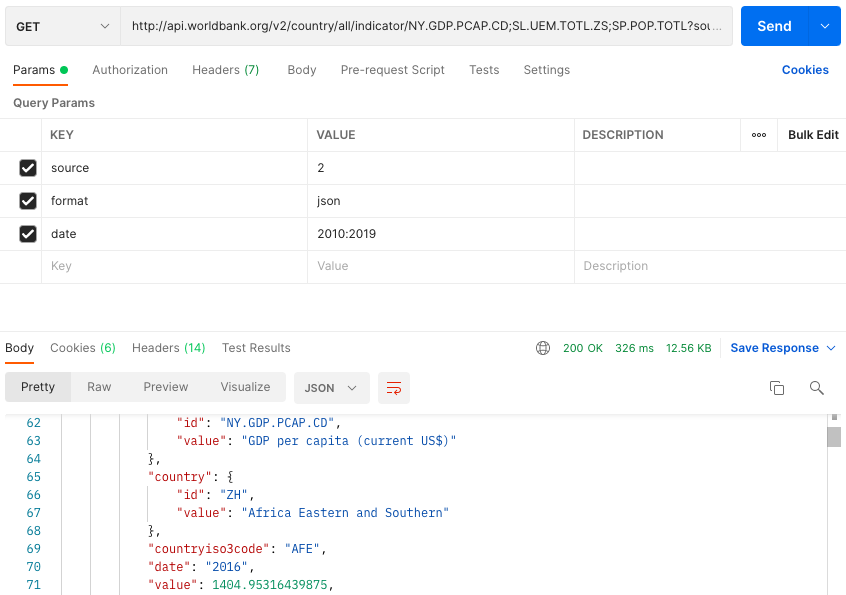
domain country *(all countries)* dataset sourcedate range

<http://api.worldbank.org/v2/country/all/indicator/NY.GDP.PCAP.CD;SL.UEM.TOTL.ZS;SP.POP.TOTL?source=2&format=json&date=2010:2019>

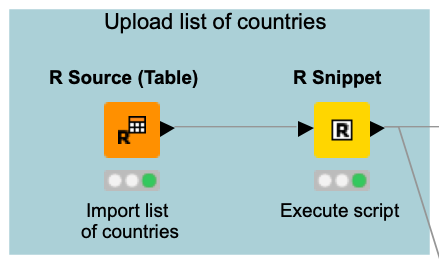
scheme version 2 of indicators *(GDP, unemployment,* format *(JSON)*

the Indicators API *population)*

We used Postman to check the validity of request. Please refer below for Postman result:



In order to build our data pipeline in KNIME we firstly loaded list of countries using R script (WDI package). Grouped rows, such as European Union, Arab World, East Asia and others were excluded from the list of countries using R script. After importing countries to KNIME with R Source node, R Snippet node was used in order to execute that R script.



Further we wrote the URL for World Bank API using String Manipulation, which is join("http://api.worldbank.org/v2/country/",$knime.in.code$,"/indicator/NY.GDP.PCAP.CD;SL.UEM.TOTL.ZS;SP.POP.TOTL?source=2&format=json&date=2010:2019").

The URLs for each country were added in new column, we send a get request, extracted values of indicators using JSON Path and converted JSON into table. We filtered out unnecessary columns so we have only Country Name, Country Code and data values.

The table is in wide format so as a next step, we transformed it into long format using R Script. A necessary step for joining with suicide data since the latter is in long format.

SQL:

We used suicide data collected by WHO <https://www.who.int/data/gho/data/indicators/indicator-details/GHO/suicide-mortality-rate-(per-100-000-population)> . Data was downloaded in .csv format then transformed into .sql so the project can be easily reproduced on any computer without being sabotaged by **secure**-**file**-**priv** issues. Unnecessary columns were dropped, variables renamed in a meaningful name. The data cleaning process is executed in the Knime workflow, so this step is simply creating the database and the main table.

Please kindly change username and password according to local instances in the Knime Workflow after running the .sql script.

In Knime, we first connect to the database then read the table.

Join and Data Cleaning:

Suicide and WDI data were joined with inner join by using year and country. The resulting table was subject to further cleaning:

* We dropped every row where suicide rate was missing. We also considered zero as missing value. Node used: Rule-based Row Filter
* In many cases where the suicide rate estimates were exact values, the lower and upper bound estimates were missing. We replaced the missing values with the exact values so the analysis can use both lower both upper bound estimates. Node used: Rule Engine
* Age group variable was messy, the Totals were as missing values, we imputed “Total” with Rule Engine Node. No data by age groups before 2019, therefore we filtered for Total only.

For the 2 main research questions, the process is split into 2. In one table we filtered out missing GDP per capita values. In the other table the missing unemployment data was filtered out so meaningful scatter plots an regressions can be constructed.