**ETL Process Report**

By Danielle Weeks and Rachel Poulos

***Extraction:***

-Extracted data from two different csv files (sources – last page of this document) using pandas read\_csv

***Transformation:***

*Suicide Data*

-Copied columns from suicide data that were relevant – Country, Year, and Suicide Count

-Dropped rows with null values using pandas ‘dropna’

-Used ‘loc’ method to extract relevant years

-Calculated sum of suicides by country using sum aggregate function and groupby to organize by country

-Merged results from all years to one table

-Used pandas ‘melt’ method to unpivot table and normalize data

*GDP Data*

-Copied relevant columns from GDP data

-Used loc to extract GDP as indicator from world bank data

-Used pandas ‘melt’ method to unpivot table and normalize data

***Loading***

-Created database connection to postgres using SQLAlchemy ‘create\_engine’

-Confirmed table names from ETLProject\_db database created in postgres (see separate sql documentation for scheme set up)

-Loaded final pandas dataframes using ‘to\_sql’ method and connection created by SQLAlchemy dependency

***Descriptions and Source of Datasets:***

**Suicide Statistics by Country Source:**

<https://www.kaggle.com/szamil/who-suicide-statistics/download>

**Suicide Statistics by Country Description:**

Basic aggregate numbers covering 1979-2016, by country, year, age groups and sex. There is only one file, with only a few columns. I made this file using the [WHO Mortality Database online tool](http://apps.who.int/healthinfo/statistics/mortality/whodpms/).

**GDP Dataset Source:**

<https://datacatalog.worldbank.org/dataset/wealth-accounting/resource/f3a8d0aa-621c-43a1-a59d-58365a62ae51>

**Description of Total Wealth Per Capita (labeled GDP in our dataset variables):**

<https://datacatalog.worldbank.org/total-wealth-capita-constant-2014-us>

**Total Wealth Per Capita (Constant 2014 US$)**

Total wealth is calculated as the sum of produced capital, natural capital, human capital, and net foreign assets. Values are measured at market exchange rates in constant 2014 US dollars, using a country-specific GDP deflator.

**Indicator Id:**

NW.TOW.PC

**Indicator source:**

Wealth Accounts

**Dataset Id:**

59

**Indicator source:**

Lange, G.M., Q. Wodon, and K. Carey, editors, The Changing Wealth of Nations 2018: Building a Sustainable Future, Washington, DC: The World Bank

**Type:**

[Time Series](https://datacatalog.worldbank.org/data-type/time-series)

**Rating:**

0

No votes yet

**Databank ID:**

**Primary Dataset:**

Yes

**Indicator License Flag:**

NA