**ETL Group Project**

Group members: Seeh , Kurt, Betsy

Our group collected data on carbon emissions by country for the year 2016 and compared it to deaths attributed to air pollutants in that year for the same 161 countries.

**Extract:**

Data Sources include:

World Health Organization – Json file (<http://apps.who.int/gho/data/node.main.BODAMBIENTAIRDTHS?lang=en>)

Our World in Data – CVS file

(<https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions>)

**Transform:**

Our transformation process was centered on transforming the data by dropping rows and columns, joining tables, renaming and splitting columns.

More specifically, for the WHO data, we got rid of years other than 2016. We also got rid of any null values, split the death column [3, [35-55] into death and death variance columns, and finally we renamed and resorted the remaining columns. For the emissions data, we deleted all dates that weren’t 2016 and all null values. We also deleted the code column.

Finally, we then did an inner join between both of these tables on country. After this, we sorted values ascending alphabetically on country after the merge.

**Load:**

We created a database in mySQL workbench. We declared the column names and variable types so that it would mirror the table made in Python. We then created an engine, connected to the database, and loaded the data into the mySQL database.