**“US Greenhouse Gas Emissions Relative to Shift in Energy Sources – Data Analysis from 1990 to 2017”**

**Team Members**

Stacy Giauque

Brian Walker

**Outline**

Data extraction, transformation, and load of data, in order to compare US CO2 (greenhouse gas emissions) vs changes in energy source types.

**Data Sources** - \*\*Extract\*\*

We sourced US CO2/Greenhouse Gas Emissions, Energy Production by Type, and Global Population data from the following websites, downloading in csv format.

* **US CO2/Greenhouse Gas Emissions 1990 to 2017 – As Reported by the Organization For Economic Cooperation and Development (OECD)**

<https://cfpub.epa.gov/ghgdata/inventoryexplorer/#iallsectors/allgas/inventsect/all>

* **US Energy Production By Type 1949 to 2018 – As Reported by the US Department of Energy (EIA)**

<https://www.eia.gov/totalenergy/data/browser/?tbl=T01.02#/?f=M>

* **Global Population Data 1960 to 2018 – As Reported by the World Bank (IBRD/IDA)**

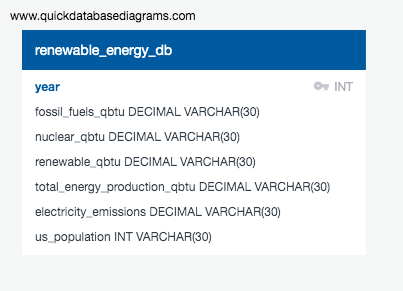
<https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US>

**Data Cleanup** - \*\*Transform\*\*

After sourcing data in csv format, we segmented US CO2/Greenhouse Gas Emissions data to reflect only Total Electric Production Emissions in millions of metric tons, eliminating emissions by various, remaining economic subsectors. For US Energy Production By Type, we consolidated figures for Total Energy sources in quadrillion BTU’s with Fossil Fuels, Nuclear, and Renewable sources as subsets. Finally, we parsed Global Population data reporting for 264 countries, to reflect only US population figures.

**Database Consolidation** - \*\*Load\*\*

In the interest of scalability given the vast amount of energy and emissions data available for analysis, we chose to consolidate in PostgreSQL, a relational database. Below is a schema of the data which was joined, consolidated in Python, instead of PostgreSQL due to flexibility.

****