

Semantic Web (CS6315.501)

Team Project Report

"Electricity Generation, Consumption & Pollution Analysis"

(Project Type: Custom)

By:

QuadSquad

Team Members:

| | |
|----------------|-----------|
| Manan Lakhani | mxl164830 |
| Ramya Sudarsan | rxs164530 |
| Hardik Punjabi | hvp160030 |
| Gaurav Fogla | gmf160030 |

Introduction:

The aim of this project is to find out which is the most efficient energy source for electricity production and reduction of air pollution. The 3 datasets used in this project are about US electricity generation, consumption and pollution for the years 1990 – 2016 at Federal level. The datasets contain details regarding the energy sources used for the generation or production of electricity like Coal, Natural Gas, Petroleum and Geothermal. The pollution dataset consists of data about 3 different pollutants: CO₂, SO₂ and NO_x which are produced during the usage of energy source to generate electricity.

Target Audience:

- Departments dealing with environment protection
 - By getting to know which regions have massive generation of pollutants, department can invest in these areas with the aid of government and take appropriate preventive measures.
- Industrial audience
 - People working in industries that produce electricity can research based on these datasets and come up with alternative solutions to reduce pollution.
- Common people
 - People can get insights on the most suitable places to reside given the level of harmful pollutants state wise and the trends of generation and consumption of electricity.
- Statisticians and Investigators
 - The data and their visualizations can be used to by statisticians to find some correlation in the data which might aid to their research.

Description of the Data Sources:

Three datasets are used to create a relationship between energy generation, consumption and pollution with respect to different energy resources:

1) Energy Generation (https://www.eia.gov/electricity/data/state/annual_generation_state.xls)

This dataset provides net generation of electricity by State by Energy Source for the years 1990 to 2016.

Columns: Year, State, Energy source, Generation (MWH)

Number of Triples: 17020

2) Energy Consumption (https://www.eia.gov/electricity/data/state/annual_consumption_state.xls)

This dataset provides net consumption of electricity by State by Energy Source for the years 1990 to 2016.

Columns: Year, State, Energy source, Consumption (MWH)

Number of Triples: 16604

3) Annual Emissions (https://www.eia.gov/electricity/data/state/emission_annual.xls)

This dataset provides Electric Power Industry Estimated Emissions by State for the years 1990 to 2016.

Columns: Year, State, Energy source, CO₂ (MT), SO₂ (MT), NO_x (MT)

Number of Triples: 25392

Data Integration:

- The data sets are downloaded from <https://www.eia.gov/electricity/data/state/>
- We used Google Refine tool to convert CSV to RDF.
- The RDF datasets were loaded into Apache Jena Fuseki.
- The datasets were integrated on State, Year and Energy source since these were the common columns in each dataset.
- We used SPARQL queries to query across multiple datasets to combine the results to derive some valuable information.
- The derived information was displayed using Google visualization API.

Results:

Query 1:

```
select ?s ?state ?generatedwatts
FROM NAMED <http://localhost:3030/MyNewDataset/data/generation>
WHERE {
    GRAPH <http://localhost:3030/ MyNewDataset /data/generation>
    {
        ?s <https://data.cdc.gov/resource/generation/year> '2001' .
        ?s <https://data.cdc.gov/resource/generation/energysource> 'Coal' .
        ?s <https://data.cdc.gov/resource/generation/state> ?state .
        ?s <https://data.cdc.gov/resource/generation/output> ?generatedwatts
    }
}
```

Generation

Year [2001] Resource [Coal]

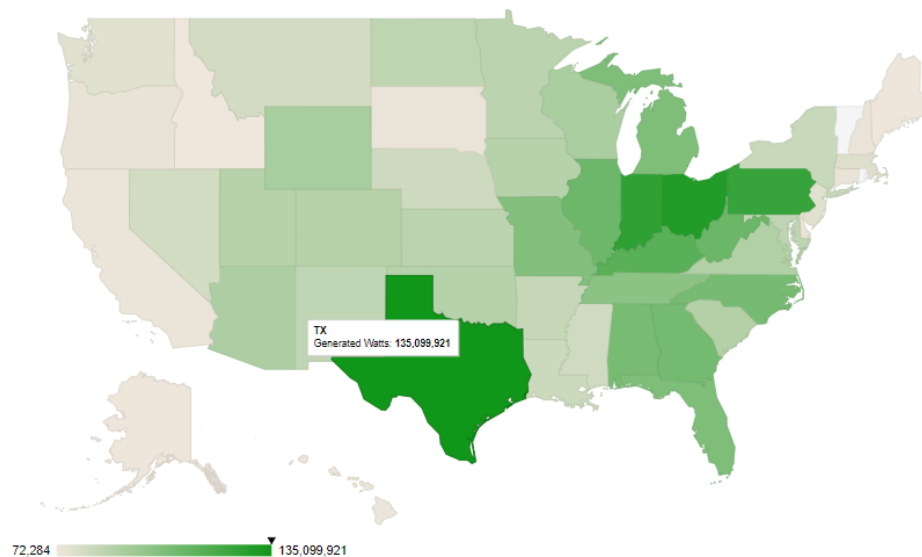


Figure 1 Electricity generated in MWH for the given year, energy resource state wise

Query 2:

```
select ?s ?state ?consumedwatts
FROM NAMED <http://localhost:3030/MyNewDataset/data/consumption>
WHERE {
  GRAPH <http://localhost:3030/MyNewDataset /data/consumption>
  {
    ?s <https://data.cdc.gov/resource/consumption/year> '2007' .
    ?s <https://data.cdc.gov/resource/consumption/energysource> 'Coal' .
    ?s <https://data.cdc.gov/resource/consumption/state> ?state .
    ?s <https://data.cdc.gov/resource/consumption/output> ?consumedwatts
  }
}
```

Consumption

Year (2007) Resource (Coal)

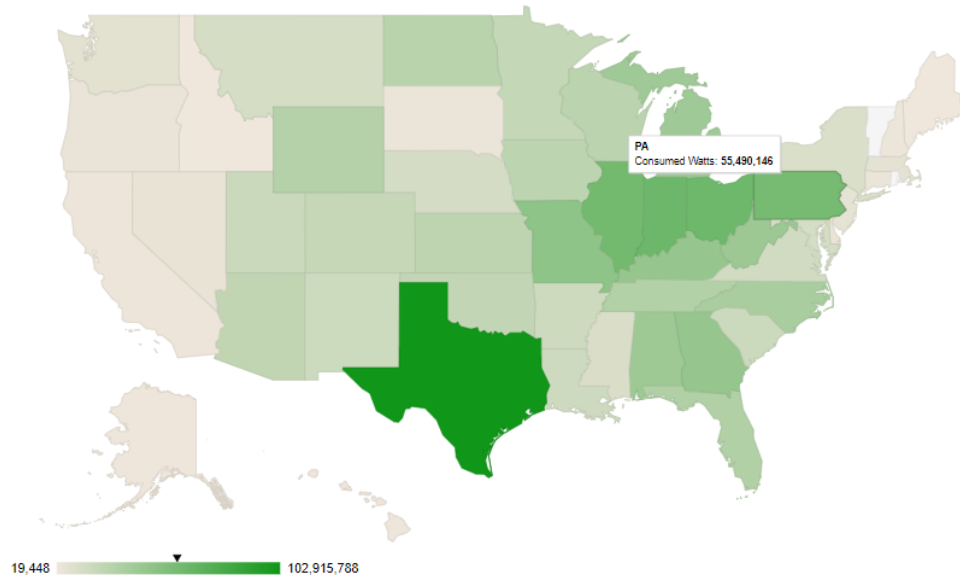


Figure 2 Electricity consumed in MWH for the given year, energy resource state wise

Query 3, 4, 5:

```
select ?year ?generatedwatts ?consumedwatts
FROM NAMED <http://localhost:3030/MyNewDataset/data/generation>
FROM NAMED <http://localhost:3030/MyNewDataset/data/consumption>
WHERE {
  GRAPH <http://localhost:3030/MyNewDataset/data/generation>
  {
    ?s1 <https://data.cdc.gov/resource/generation/year> ?year .
    ?s1 <https://data.cdc.gov/resource/generation/state> ?o FILTER (lcase(str(?o)) = "us-total") .
    ?s1 <https://data.cdc.gov/resource/generation/energysource> 'Coal' .
  }
}
```

```

?s1 <https://data.cdc.gov/resource/generation/output> ?generatedwatts
}
GRAPH <http://localhost:3030/MyNewDataset/data/consumption>
{
?s2 <https://data.cdc.gov/resource/consumption/year> ?year .
?s2 <https://data.cdc.gov/resource/consumption/state> ?o1 FILTER (lcase(str(?o1)) = "us-
total") .
?s2 <https://data.cdc.gov/resource/consumption/energysource> 'Coal' .
?s2 <https://data.cdc.gov/resource/consumption/output> ?consumedwatts
}
}

```



Figure 3 Comparison of Electricity generation and consumption year wise for the energy resource coal

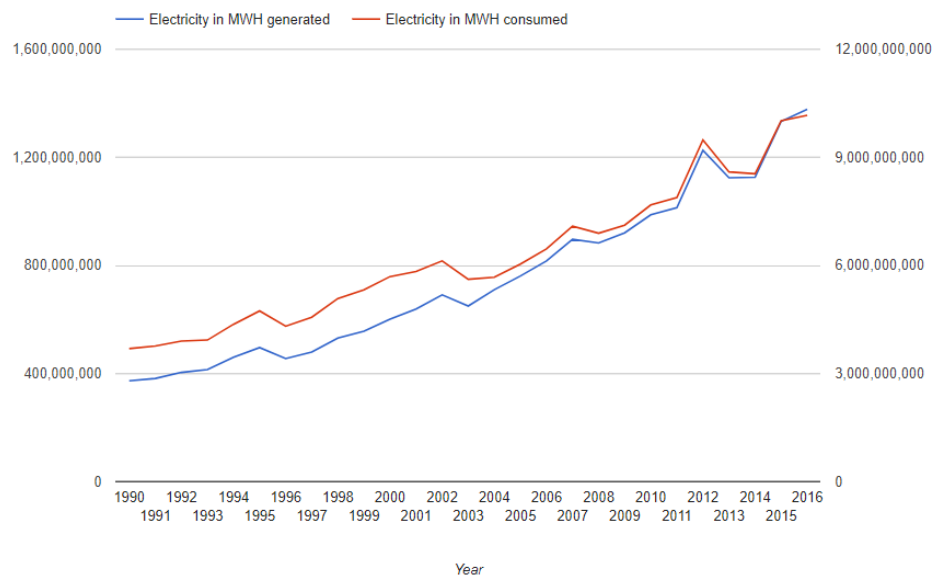


Figure 4 Comparison of Electricity generation and consumption year wise for the energy resource Natural gas



Figure 5 Comparison of Electricity generation and consumption year wise for the energy resource Petroleum

Query 6, 7, 8:

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?state ?source ?co2_count ?foutput

WHERE

{
{

select ?state ?source (sum(xsd:decimal(?co2)) as ?co2_count)

WHERE {

GRAPH <http://localhost:3030/MyNewDataset/data/pollution>

{

?s1 <https://data.cdc.gov/resource/pollution/state> ?state FILTER ((lcase(str(?state)) != "us-total") && (lcase(str(?state)) != "dc") && (lcase(str(?state)) != "hi") && (lcase(str(?state)) != "ri") && (lcase(str(?state)) != "vt")) .

?s1 <https://data.cdc.gov/resource/pollution/energysource> ?source FILTER (lcase(str(?source)) != "geothermal") .

?s1 <https://data.cdc.gov/resource/pollution/co2> ?co2

}

} GROUP BY ?state ?source

ORDER BY ?state ?source

}

{

select ?state ?source (sum(xsd:double(?wattsgenerated)) as ?foutput)

WHERE {

GRAPH <http://localhost:3030/MyNewDataset/data/generation>

{

?s2 <https://data.cdc.gov/resource/generation/state> ?state FILTER ((lcase(str(?state)) != "us-total") && (lcase(str(?state)) != "dc") && (lcase(str(?state)) != "hi") && (lcase(str(?state)) != "ri") && (lcase(str(?state)) != "vt"))).

?s2 <https://data.cdc.gov/resource/generation/energysource> ?source FILTER (lcase(str(?source)) != "geothermal") .

?s2 <https://data.cdc.gov/resource/generation/output> ?wattsgenerated

}

} GROUP BY ?state ?source

ORDER BY ?state ?source }}

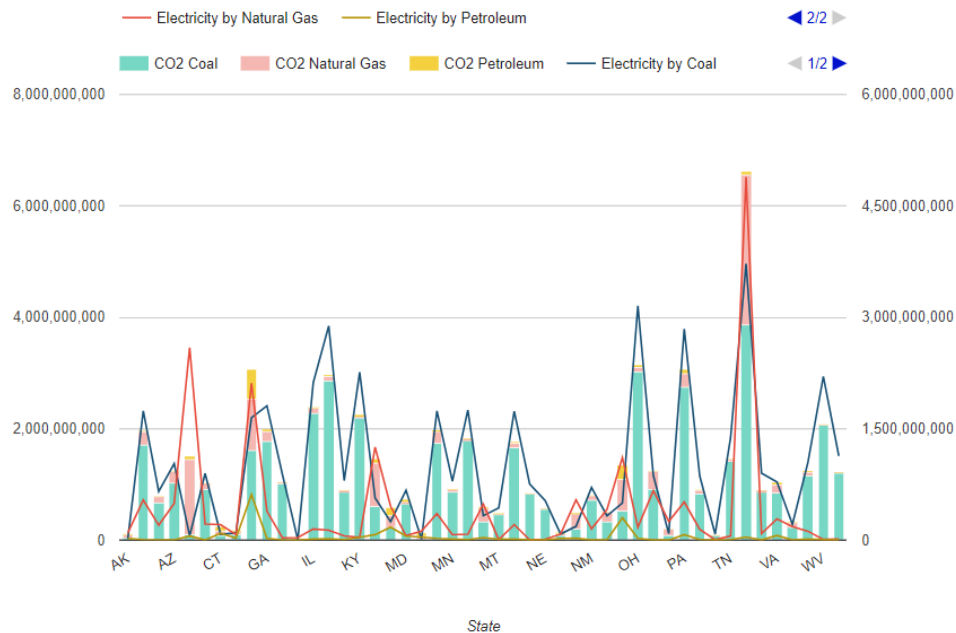


Figure 6 CO2 emission analysis for each energy sources by state

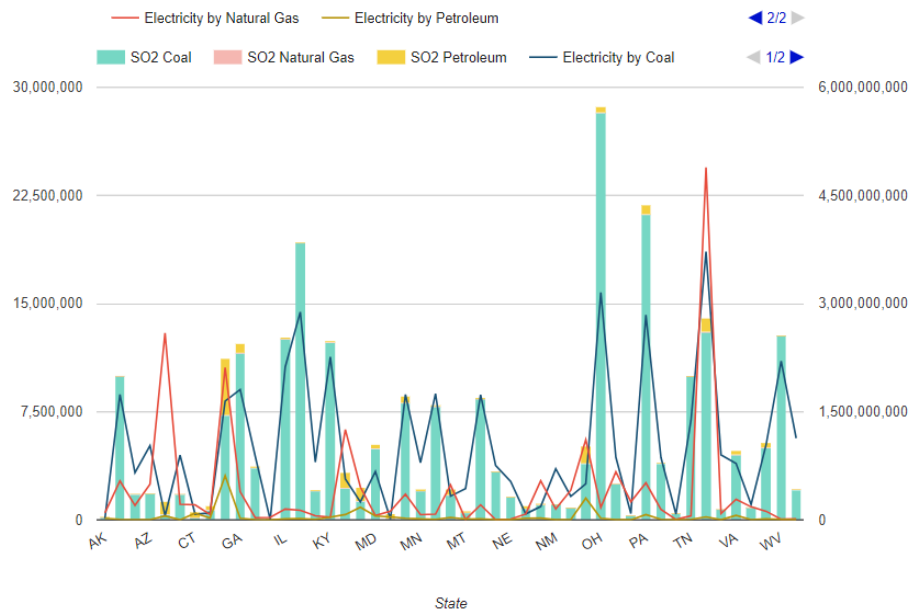


Figure 7 SO2 emission analysis for each energy sources by state

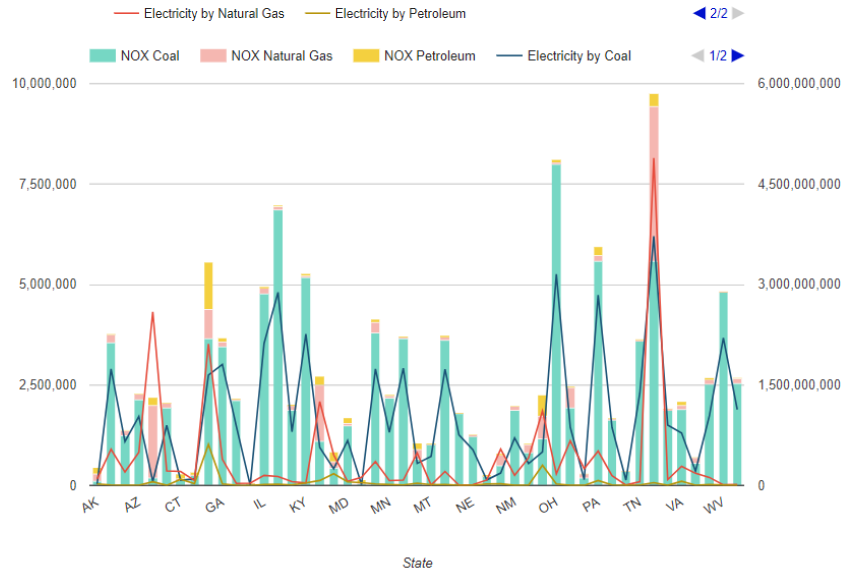


Figure 8 NOX emission analysis for each energy sources by state

Justification:

- We integrated three large datasets using FUSEKI server and created a SPARQL end-point to run our queries.
- Our project provides extensive analysis between multiple datasets
- We use Google Visualization API to represent our statistics state wise.
- Custom Projects allow us to mix and match any number of datasets also it provides creation of our own SPARQL endpoint.

Summary:

From the analysis done above, we could compare the electricity generated and consumed at the Federal level. Also, we could conclude that electricity generated using natural gas released least amount of harmful pollutants. We hope this will provide insight into data, and thus give meaningful information to help prove the extensive reach of global warming and how it can damage earth the way we know it. We present this information to help take a step towards restoring the environment at global and individual level by taking appropriate measures to control the emission of these pollutants.