Energy production and consumption in the United States

Tristan Kindig and Sudarsan Sadasivuni

Background and motivation

Part of my personal interest in this topic stems from my participation in the 2018 MCM competition. I chose a problem performing a feasibility analysis for an interstate renewable energy compact. The topic piqued my interest, but the competition was very short so I didn't have the time that I needed to explore it fully. I see this project as an opportunity to conduct that exploration and satisfy my curiosity.

My fascination with this topic stems from the fact that energy issues constitute some of the most critical challenges facing our society. Our transition from fossil fuels to renewable energy in the face of climate change is shaping up to be one of the defining struggles of 21st-century United States. Moreover, this is at its core a global problem and our success would have global implications because successful strategies could be applied in other nations.

But there are many topics which I find personally interesting and which are closely related to critical societal challenges. This topic in particular is a strong candidate for this project because of the overwhelming amount of energy and climate data that has been collected and made publicly available.

Finally, choosing this project compels me to improve my familiarity with geospatial data and the libraries for processing/visualizing them, which I believe is a useful skillset beyond the energy/climate domain. And it satisfies my personal interest in network science / graph theory. We aren't simply displaying a series of histograms or scatterplots but are instead producing visualizations that are superimposed on a skeleton with a meaningful topology (in this case, spatial in the literal sense).

Objectives

The domain-specific objectives are to better understand energy patterns in the United States and to motivate this understanding by visualizing the deleterious effects of climate change. If energy data holds our answer to the question "What is this," climate data is the answer to "Why does it matter?" There is also an economic component of both topics, so this is a strongly interdisciplinary project.

We hope to learn more about geospatial data processing/visualization tools such as GeoJSON, shapefiles, ArcGIS, and the geospatial elements of D3. We'd also like to become more comfortable using time-series analysis/visualization tools and techniques. There is an inherent temporal element to much (perhaps most) of the data we're interested in (e.g. prices over time, power over time, rainfall over time). It will also be an interesting challenge to handle both current system state and system trends (e.g. emphasizing most recent energy distribution while also presenting historical trends).

Integrating data from multiple sources into a single graph (for example, datasets I, O, and P) presents data cleaning challenges more demanding than doing them each separately. Hopefully the pay-off is the reward of condensing more data, intuitively and cleanly, into a single graph.

Data

We have identified several databases and quite a few datasets, which we present below:

I. Databases
   A. Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-Dive)
   B. Data.gov
   C. Homeland Infrastructure Foundation-Level Data (HIFLD)
      1. ArcGIS shapefiles
      2. Full catalog
   D. Energy Information Administration's (EIA) yearly national data
      1. Has sector-specific and resource-specific breakdowns, but only on a national level

II. Datasets
   A. Long-Term Daily and Monthly Climate Records from Stations Across the Contiguous United States
      1. Notes: Daily data include observations of maximum and minimum temperature, precipitation amount, snowfall amount, and snow depth; monthly data consist of monthly-averaged maximum, minimum, and mean temperature and total monthly precipitation across 1200 weather stations
      2. How would we graph this though? Construct Voronoi diagram
   B. Solar resources by county
      1. Has data on viability of solar power in each county
   C. Battery degradation
      1. Has raw data on actual remaining voltage in batteries after repeated discharge cycles
   D. Battery degradation
      1. Has secondary data on electric vehicle use
   E. Nationwide energy source price data
      1. Annual data on natural gas, gasoline, diesel, and residential electricity prices
   F. Uranium location
   G. Non-gasoline fueling stations
   H. Electric power planning areas
      1. Has information on peak power demands
      2. Could be useful for exploring monopolies
   I. Historical data on tropical storm tracks
   J. Power plant locations
      1. Very granular information on fuel source
   K. Oil and natural gas fields
   L. Natural gas storage facilities
   M. Coal fields
   N. Mobile home parks
      1. Most vulnerable population wrt floods
   O. Storm events

P. Flood zones

Data processing

We need to examine the data collected to make sure they are in D3 (JavaScript) accessible mode. Most of the data we obtained are in .csv tabular format (spreadsheet), they need no additional clean up processing. But we obtained some data in the layer format, so we need to organize it such that we can obtain different .csv files for necessary combinations that we want to work on. Also, the data we obtained is huge so we need to use only some part of data (mentioned near every data set considered) that is essential for project, so data cleanup is essential. This process would require **3 days** of time, because the available collection of data is pretty huge.

Must-have features

Our primary goal is to develop an understanding of energy production and consumption patterns, so related visualizations will constitute our must-have features. Several critical visualizations are listed below:

- State-level, resource-specific breakdown of energy production, over time
- County-level breakdown of energy consumption, over time
- State-level, sector-specific breakdown of energy consumption (e.g. industrial, transportation, residential), over time
- State-level, resource-specific breakdown of energy consumption (e.g. coal, oil, LNG), over time
- Regression analysis of government subsidization and private sector investment predicting renewable energy production
- Trends in energy price over time, by resource

Optional features

We are also proposing several optional features:

- Voronoi diagram of United States by closest power plant
  - Weight distances by power plant production levels?
- Examine connection between undirected graphs of power plants / cities networks
  - How do they evolve over time as switch from coal to renewable energy or LNG is made
- Impact of alternative energy development on energy security
- Display of county-level climate trends (temperature, precipitation)
- Examine how energy infrastructure might change wrt improvements in battery technology
- Apply Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin to Mississippi Delta
- Compare trends in US energy development to those in a developing nation

Project schedule

The first week we will be performing data processing and practicing GeoJSON commands.

The second week will be allocated to visualizing relationship between renewable energy subsidiaries and their production.

The third week will be allocated to visualize climatic data (including storms, floods data) obtained from weather stations across United States.

The next two weeks we plan to work on state-level, county-level visualization of data to obtain knowledge about breakdown of energy consumption, production over time.

Visualization design

The main part of the project is to display all the created visualizations in a realistic and easy to decipher way, so we are planning to display all this data on a single graph, where we are planning to divide outputs into different tabs, and allowing the user to select different factors to compare. We are planning to display our data as follows:

- To demonstrate the climatic conditions from weather stations, we use heatmaps (gives the denser areas clearly).
- To demonstrate the visualizations related to energy price, fossil fuel consumptions, production of fuels we make use of time series graphs (because we are getting data over time, a time series graph would be apt.) We can also consider bar graph but visualization wouldn't be so clean and neat.
- To demonstrate the visualizations of data related to county-level, state-level resource production/consumption over time, we make use of geospatial information, which would enable us to understand clearly the information we can obtain over different longitudes and latitudes of United States.