Berkeley School of Information

datascience@berkeley

The Inputs and Outputs that Power America

A Final Proposal for Data Visualization

Chet Gutwein, Chuqing He, Max Lee

Professor John Guerra Gomez

Data Visualization – w209

October 4, 2017

Overview of Interest Area

Over 350 million people live in the United States of America, the world's biggest consumer of energy on a per capita basis. The mix of energy sources is changing at a rapid rate for a variety of reasons: renewable power is becoming more affordable, domestic natural gas production is booming, and climate change is evident. With all of these things becoming part of many Americans' every day conversations, we wonder - how much does the average American know about where we get our power? For this visualization we aim to connect users to an in-depth understanding of where our power comes from, how it affects the environment, and how the daily lifestyle of America's citizens contribute to our "carbon footprint". We believe that this topic has enough breadth to allow for enough scope as a final project while encouraging depth on certain focus areas that will emerge as the project progresses.

The Data

We have identified the following sources as having easily accessible data that can contribute to our visualization:

- ***Emissions & Generation Resource Integrated Database*:** Data provided on power output and emissions for almost all power generated in the United States. Summary data available for 1996-2014, allowing us to introduce a change-over-time factor.

- ***US Energy Information Administration***: Data provided on electricity production, sales, across all grids. Majority of data on state level, but some on more granular data at a plant level.

- ***US Geological Survey***: Data provided on a spectrum of environmental measures that can be used in estimating the secondary correlated effects outside of simple gas emissions.

- ***Environmental Protection Agency***: Data provided across the board on ecological, power, and other topics that can be used to augment analysis and visualizations.

- ***Energy Velocity (potential)***: A private company that has large selection set of proprietary data sets for analysis. Possible opportunity to leverage our team's connections to acquire an academic sampling of data.

The Users

We anticipate our visualization to be used primarily as a general resource by the public for inquiries into power utilization and its effects. Secondary users might be users hoping to do research on similar topics that table or plot output might be useful for identifying the trends for various types of energy productions/consumptions across different time periods, different locations/regions, and perhaps establish correlation between the energy usage and its geological effects.

The Tasks

This visualization will help users more effectively learn about energy usage and its effect on the environment. It will be a tool that will help users effectively communicate about a complicated subject by presenting data in an interactive and user-friendly way with enough detail to gain in-depth knowledge in the subject area.

The Gameplan

We believe the following are high-level checkpoints to realize our visualization:

- **Clean** data sets of interest into usable form. This includes filtering relevant data sets and planning what we want to visualize. Preliminary EDA on the datasets and pre-process the desired dataset, such as removing/replacing any NULL or extreme values.

- **Load** the cleaned data into environment for manipulation and eventual display. Determine the best form of visualization for the different types variables.

- **Code** the necessary components for desired visualization as determined in cleaning stage. Including the titles, legends, proper units of measurements on the axis, and perhaps interactive features for the users to drill down the granularity.

- **Organize** outputs and analysis in an easily understandable final product. Host the final product in the Berkeley server and make it publicly accessible.

Contact Info

Chet Gutwein
E-mail: chet_gutwein@berkeley.edu
Slack: @cgutwein

Chuqing He
E-mail: chqngh@ischool.berkeley.edu
Slack: @chuqing.he

Maxwell Lee
E-mail: maxstlee@gmail.com
Slack: @max_lee