

Final Project for GEOL590 Reproducible Data

Kate Fullerton and Katie Sipes

March 22, 2017

Overview

Energy consumption throughout the United States is a carefully watched statistic by the U.S Energy Information Association. Energy uses in biofuel, gasoline, wind power and even electricity trends are diligently annotated in consensus data ¹. The efficiency of energy usage is a big area of research as other forms of energy have been more readily used, such as wind and solar power. Energy demands across the United States vary largely on population ².

Plan

Given our experience with coding in R, we intend to begin with writing a data analysis project that we will venture to adapt into a general package. Our project is targeted at renewable energy plant data, such as that for hydroelectric power plants in the Western United States, to analyze parameters such as production efficiency, cost-production ratios, and eventually compare these values to energy consumption and population census data. By combining this energy data with the population density in specific regions we hope to conclude which areas are the most efficient in energy use and production. In order to generate our analyses, the intention will likely be for us to divide and conquer the different sub-analyses we want to perform on the datasets. Kate will focus on the ?efficiency? analyses while Katie will develop the cost-production analyses with us coming together for any census data analysis.

Division of Labor and Workflow

For development, we will likely utilized a Centralized Workflow, where we each have our own local repositories that is synchronized to the main hub. Data analysis using dplyr tools will help to streamline individuation functions given the statistical manipulations

¹Consumption and Efficiency. US Energy Information Administration. Feb 24th 2017. <https://www.eia.gov/consumption/>

²Chapter 1. World energy demand and economic outlook. <https://www.eia.gov/outlooks/ieo/world.cfm>

and analyses we intend to perform. We intend to base our analyses based on 2 different data sets from the Department of the Interior related to hydropower in the United states. The first is a dataset represents the monthly hydropower generation data by facility over the past 10 years in the US to be used for our efficiency calculations ³, the second represents the hydropower potential in the Western US to be used for cost-production analyses ⁴.

There are a variety of different analyses that can be performed with the data sets we are analyzing. The first and most simple tool will be one for unit conversion, in order to ensure analyses are consistent between datasets even if reports are in differing units. Given the capacity of a plant we can determine the hypothetical maximum production and compare annual production to that maximum to determine the capacity factor (the ratio of actual output vs maximum) of various plants. Changes in capacity factor over time can be plotted to look for trends at different plants. With reported data on initial start-up and monthly production costs of plants, we can perform a variety of cost-benefit analyses comparing the costs to run certain plans as compared to the net production of the plant amongst other things.

With the proper population census data, we should also be able to determine energy consumption of particular regions affiliated with particular energy plants. From this data, we could determine the percent of energy demand met by these renewable plants and link that to cost analyses to determine whether these kinds of plants are economically feasible (though this may be a lofty goal for us).

³Hydropower data can be found open source: <https://catalog.data.gov/dataset/monthly-hydropower-generation-data-by-facility-us-bureau-of-reclamation>

⁴Cost production analysis data can be found open source: <https://catalog.data.gov/dataset/hydropower-potential-in-the-western-us>