SMU Data Science Review

Volume 2 | Number 2 Article 3

2019

Visualizing United States Energy Production Data

Bruce P. Kimbark
SMU, bkimbark@smu.edu

Melissa Luzardo
mluzardo@smu.edu

Charles South
SMU, csouth@smu.edu

James Taber
jtaber@wmp.com

Recommended Citation

Kimbark, Bruce P.; Luzardo, Melissa; South, Charles; and Taber, James (2019) "Visualizing United States Energy Production Data," *SMU Data Science Review*: Vol. 2: No. 2, Article 3.
Available at: https://scholar.smu.edu/datasciencereview/vol2/iss2/3

Visualizing United States Energy Production Data

Bruce Kimbark, Melissa Luzardo, Charles South, James Taber

Master of Science in Data Science
Southern Methodist University
Dallas, Texas USA
bkimbark@smu.edu, mluzardo@smu.edu, csouth@smu.edu, jtaber@wmp.com

Abstract. The energy production, fuel load, financials and environmental impact of plants in the United States is publicly available from a variety of sources such as the Energy Information Administration, the Environmental Protection Agency or Lazard among others. The general public is interested in US energy production and its potential environmental impact but the available information is complex and difficult to properly understand and not shared in ways that are accessible. Our objective was to gather this data and create different interactive visualizations that make it consumable. Each of the five visualizations was designed to illustrate a specific aspect of energy and together can provide a solid but general understanding for the users. We then conducted a survey on the visuals to test our effectiveness on visual appeal and comprehension. In this paper, we discuss how we compiled and organized the relevant data, created the visualizations, and analyze them independently including the survey responses. We found our images can be an effective tool to gain insights and learn about the subject, regardless of the degree of prior knowledge in power plants and electricity generation.

Acknowledgment

We thank Dr. Daniel Engels (Southern Methodist University) for useful conversations, guidance and support throughout the completion of this work.

1 Introduction

While doing our research, we discovered an abundance of publicly available data regarding energy production but it was hard to digest or fully understand. Most of the data is found in lengthy tables. We began primarily interested in the energy production by fuel source. The EIA provides a lot of data in a very comprehensible visualizations on their website. The net generation by sector allowed us to understand at a high level the information we were seeking in the Electricity Data Browser[1]. Another great visual was published by the Washington

[1] https://www.eia.gov/electricity/data/browser/

Post in March, 2017. The maps[2] presented an interactive feature with the share of electricity per state by fuel source resulting in a visual providing significant insights. Both sets of visualizations allow the reader to see how natural gas is the most prevalent fuel source today.[8] We did however discover a lack of effective visuals available for an in-depth look at electricity production by power plants.

To remedy this, we intended to create visualizations that can provide these deeper insights and educate the user. Interactive art allows for data exchange and exploration by the user that is not possible with static images. Interactivity can be a deeper experience for the user and creates feedback between the user and program. [10] In this paper, we will discuss our three sources of data, a general background on electricity production itself, our collection and manipulation of the data, the creation of each visual, and analysis for each.

1.1 U.S. Energy Information Administration

Established in 1977, the EIA[14] the independent statistical arm of the U.S. Department of Energy. Its primary purpose is to collect data regarding the U.S. and global energy sector, perform analysis, and make publicly available its findings. These analyses include both micro and macro understanding of U.S. energy production, distribution, and markets with both short-term and long-term forecasting. In addition, much of the collected data is now publicly available on their website, www.eia.gov.

1.2 U.S. Environmental Protection Agency and eGRID

The EPA was founded in 1970 with the aim to help coordinate and inform response to environmental issues. The agency performs varied research into and continually monitors environmental impacts and risks to human health. These findings inform policy and limiting of pollutants from industry and the population. As part of its functions, the EMP monitors the environmental emissions produced while generating electricity and make it available to the public. This data is published under in the Emissions and Generation Resource Integrated Database (eGRID) [2]. Combined with some data from the EIA, it provides emission amounts and rates of air pollution and sub-categorizes it into nitrogen oxides, sulfur dioxide, carbon dioxide, methane, and nitrous oxide. The most recent version captures data for 2016 and was released in February of 2018.

1.3 Lazard's Levelized Cost of Energy Analysis

Lazard is a financial advisory and asset management firm that performs several research studies. One of their annual publications is the Levelized Cost of Energy Analysis (LCOE). It offers an industry-wide standard U.S. Dollar cost per megawatt hour depending on the type of fuel and generator used to produce the

[2] `https://www.washingtonpost.com/graphics/national/power-plants/?noredirect=on&utm_term=.975cdd36512e`

electricity. This report estimates a low-end, high-end, and average cost based upon capital cost of building the type of power plant, fixed and variable costs, operational and management costs, and price of fuel. [9]

1.4 Tableau

Tableau is a data visualization tool that allows users to create workbooks and dashboards with drag and drop features and coding not required for most purposes. We used Tableau to create one of our visuals. One of the most used visualization platforms by companies in the world, it provides analytics tools to allows businesses to create insights. Among its available features, is the ability to create a map from various data points. When a user scrolls over a data point on this map, a tooltip appears. The creator can edit this tooltip to present additional data when scrolled over. The platform is available for businesses or individual use via license. [15]

1.5 Processing

Processing 3 is an open-source data visualization program created in 2001. It is written in javascript and allows users to create custom images through code. We created four visuals with Processing. The program was originally created with the intent to be a tool for new programmers but has since evolved into an alternative to tools that require licenses. Since its release, it has been further developed and built upon by the community of users. While the use of Java, and Processing as a result, has been supplanted by new languages, Processing has explanded into uses in mobile phones and remains popular within education. It has recently been used within devices such as alarm clocks, wearable devices, and programmable toys. We used several resources in the creation of our images. [11][7][6][10][12]

2 Background and Tutorial

2.1 Electricity

Electricity has been defined as the form of energy resulting from charged particles that either accumulate as charge or flow as current. Electricity is not only fundamental in nature but also the most widely used form of energy. However, electricity is a secondary energy source that requires a primary source such as heat source (including fuels) or a kinetic or potential energy for its production, usually throughout generators. The most common generators are electromagnetic and convert kinetic energy into electricity. These generators present a rotary electromagnetic shaft surrounded by a stationary cylinder formed by a series of insulated wire coils. The rotation of the electromagnetic shaft induces a small electric current the coils. The insulated coils of the cylinder act as separated conductors that eventually combine into one large current. This resulting current is the what we generally refer to as electricity. [14]

2.2 Electricity Generation in the United States

Most of the electricity in the United States is generated in power plants, and most of these power plants rely on turbines or similar equipment for their generators. Turbines are responsible for converting potential and kinetic energy of moving fluids (liquid or gas) to the mechanical energy and moving the shaft of generators and thus converting that mechanical energy into electricity. There are several different types of turbines according to their application and the source of energy. Among them, steam turbines, combustion gas turbines, hydroelectric turbines, and wind turbines are the most common. In the steam turbines, the steam drives that powers the generator is created heating water using a boiler (with fuel) or a heat exchanger (with a heat source). Since the steam turbines can either be fed by fuel or a heat source, biomass, coal, geothermal energy, petroleum fuels, natural gas, nuclear energy, and solar thermal energy are commonly transformed into electricity using this type of turbines. Next are combustion gas turbines, in them fuels are burned to produce hot gases which turn the blades of the turbines. Internal combustion engines (i.e. diesel engines) are closely related but the combustion in gas turbines occurs continuously, as opposed to the internal combustion engines, in which combustion occurs intermittently. Diesel-engine and smaller internal combustion engine generators are widely used in remote villages, at construction sites and even homes for emergency or backup power supply. Both combustion gas turbines and internal engine generators can use a variety of fuels including petroleum diesel, biodiesel, natural gas, biogas, propane and even gasoline or natural gas. One most efficient systems are the cogenerators or the combined-heat-and-power (CHP) plants. These are the plants that optimize the conversion and use of the energy source. Any heat or energy source not converted into electricity will find another purpose like space heating. The other efficient system is the combined cycle in which the excess energy from one process or turbine will feed the next. For example, excess heat from a gas turbine is used as the heat source of a downstream steam turbine. In this system two separate generators use a single fuel source. Finally, the Hydroelectric turbines use water to move turbine blades, and wind turbines use wind. There are other types of generators that can produce electricity without turbines, among those the most well-known are solar photovoltaic cells and fuel cells. Solar photovoltaic cells transform sunlight into electricity. And fuel cells generate electricity from fuel, usually hydrogen, via chemical process.[5][14]

The distribution of the most common types electricity generators in the U.S. in 2018 was as follows:

· Steam turbines – 64%
· Combustion turbines – 21%
· Hydroelectric turbines – 7%
· Wind turbines – 6%
· Solar photovoltaic systems – 1%
· Internal combustion engines – less than 1%

2.3 Energy Sources for Electricity in the United States

The United State uses a variety of primary energy sources to generate electricity. The sources and the technologies that utilize those resources have progressed over time and continue to evolve as technologies improve. There are three main primary categories of energy: fossil fuels, nuclear energy and renewable energy. Historically, fossil fuels have been the largest source of energy for electricity generation and although renewable energy have increased their share in recent years, fossil fuels continue to the most important source (over 60%). Fossil fuels mainly include natural gas and coal and in a significantly lesser degree, petroleum. Natural gas was used to generate about 32% of the U.S electricity in 2018, powering steam and combustion turbine generators. Coal to generate about 30%, powering mainly steam turbine generators. Petroleum and derivatives to generate less than 1%, powering steam turbines, diesel-engine generators and gas turbines.

Nuclear Energy sourced about 20% of the electricity in 2018, were steam turbines aid nuclear fission.

Finally, the remaining 17% of the primary energy source to generate the U.S electricity in 2018 came from renewable sources. Renewable sources are quite diverse and include hydroelectric plants, wind energy, biomass, solar energy and geothermal. Hydroelectric plant contributed about 7% with hydroelectric turbines. Wind energy about 6% with wind turbines. Biomass was the source of approximately 2% of the electricity generation and it was used as fuel for steam generators, gas turbines, or internal combustion engine generators. Solar energy represented about 1% of the total source in its two forms solar photovoltaic system and solar-thermal power. The solar-thermal power uses steam turbines. And geothermal power plants with steam turbines had less than 1% of the share.[14]

2.4 United State Electricity Environmental Effects

In the United States, approximately 64% of electricity generated in 2018 was produced burning fossil fuels, biomass, or waste. The byproducts that occur during the electricity generation include carbon dioxide (CO_2), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NOx), particulate matter (PM), and heavy metals such as mercury. These pollutants have detrimental health and environmental effects. CO_2 is a greenhouse gas, and an important contributor to global warming. SO_2 is responsible for acid rain and multiple respiratory illnesses and heart diseases. NOx affects ground-level air quality by increasing ozone, irritating and damaging the lungs. PM also affects air quality, and fine PM, might be responsible for emphysema and even lung cancer. Heavy metals such as mercury are hazardous to human and animal health.[1]

Other byproducts of power generation include liquid and solid waste. The most common solid byproduct is ash and its different forms large particle from the bottoms of boilers, fly ash as the smaller and lighter particles and ash sludge or ash mixed with water to capture volatile components and stored in retention ponds. These ashes are usually hazardous and pollution control is needed to handle them.

Nuclear power plants produce radioactive waste. Radioactive waste is categorized in two levels: low-level waste and high-level waste. The low-level waste is store at the facilities until the radioactivity drops to safe levels or sent to low-level site disposal. The highly radioactive nuclear fuel assemblies need to be stored in special facilities.

In the U.S, the Clean Air Act is the law that has been in plate to regulate air pollutant emissions. The U.S. Environmental Protection Agency (EPA) is the entity in charge of administering and enforcing the Clean Air Act by settings emissions standards for power plants and a variety of programs that include the Acid Rain Program. [3][4]

The electricity generation is one of the main sources of CO_2 emissions in the U.S. Electricity and was responsible for approximately 34% of total U.S. CO_2 emissions in 2018.[14][4]

3 Solution Approach for EIA Data

3.1 Data Description

Our data sourced from the U.S. Energy Information Administration is granular data for every power plant in the United States[13]. Categorized by the type of engine and the type of fuel used, the monthly amount of power generated and fuel consumed for each plant is available. Based upon the type of fuel used, the amount is summarized by total barrels, mcf, tons, or not applicable. The power generation is summarized by total megawatt hours (Mwh) produced.

For example, the Cayuga power plant in Indiana uses two different engine types - Steam Turbine and Internal Combustion (Diesel, Piston) Engine - and two different fuels - Bituminous Coal and Distillate Fuel Oil. The internal combustion engine uses only the fuel oil whereas their steam engines employ both. For the month of December in 2017, the amount of fuel used and power generated was:

· The Steam Engine using 9,476 tons of Bituminous Coal produced 19,953 MWh of electricity.
· The Steam Engine using 1,087 barrels of Distillate Fuel Oil produced 524 MWh of electricity.
· The Internal Combustion Engine using 46 barrels of Distillate Fuel Oil produced 26 Mwh of electricity.

3.2 Data Collection

The data was collected by running a separate API for each Plant Code, Plant Type, and Fuel Type combination for the time period January 2009 through December 2018. This data can be found online at the EIA's Electricity website[3] by selecting "Plant level data" under the "Change data set" dropdown [13].

[3] https://www.eia.gov/electricity/data/browser

A list appears with all power plants and indicates their prime mover engine type and fuel source. We copied the list and created a separate API request for all engine and fuel combinations. If a power plant had four engine types and two fuel types, we would run all eight possible APIs. As expected, the first run included many API requests that were unsuccessful because of missing data for those combinations or not applicable request. Fuel consumption datasets for fuel sources such as wind, geothermal, and nuclear were not created. All APIs that did not pull any information were deleted from the list and a second iteration with 15,852 total requests successfully captured the data.

The data we were able to collect and use for each plant is presented below:

· Series Identification – Power generation or fuel consumption series
· Plant Code — A unique code for each power plant
· Plant Name
· State
· Sector Name -– A description of the type of utility the plant belongs to
· Plant Type -– The type of engine used to create the electricity
· Fuel Type — The type of energy source used
· Latitude
· Longitude
· Series ID — The EIA.gov identifier for the specific data request
· Month/Year Period
· Monthly Power Generated
· Monthly Fuel Consumed

3.3 Data Preparation

The first step we took in cleaning the data was to separate the nested JSON lists and dictionaries apart to get to the underlying plant level information. Each row in the data from the API, is a series of these lists and dictionaries. To get to the underlying information, each row was separated apart by its corresponding key and value pairs. The keys were discarded at each stage as they were only identifiers. The values were then separated again by the key and value, leaving only the individual plant data desired. The latitude, longitude, plant names, series identifier, and data were extracted and each appended to separate lists. This was repeated for all rows, and then added back to the API list dataframe.

At this point, the data for each month for both the generated and the consumed electricity were still contained within lists. For each month in the list, a new row was created copying the remaining data to the new row. Each row of the initial data set created 120 new rows of data. A unique identifier was created by combining the values for plant code, engine, fuel, and date. Based upon this identifier, fuel consumption data was merged to the power generation data as a new column for each period. For power plants where fuel consumption was a null value such as solar, zeros were imputed.

Upon examining the data, several design decisions were made. Power plant 15003 produced no power for any period in our study and was removed. A key issue arose with plants that employed both Combined-Cycle - Steam Part engines (CA) and Combined-Cycle Combustion Turbine Part engines (CT). For these plants, one fuel source is used to power the CT turbine first and the steam then used to power the second CA engine. In the EIA data, fuel consumption data was only presented on the CT row while CA was zero. Due to this, these rows were aggregated together under a combined plant type 'CA, CT'. After further review, several other instances similar to this occurred with other plant and fuel types. The process was repeated and those plant rows also aggregated.

Additional columns were added to assist with our visualizations. We added columns for Fuel Category where each fuel type was assigned to one of nine types – Coal, Geothermal, Hydroelectric, Natural Gas, Nuclear, Petroleum, Solar, Wind, and Other. Months and Years columns were added to make filtering easier. The EIA full descriptions for Plant Type, Fuel Type and the unit of measurement of the fuel were added. A subset of the data for the years 2015-2018 was created to be combined with the EPA and Lazard information.

4 Solution Approach for EPA Data

4.1 Data Description

The U.S. Environmental Protection Agency gathers a comprehensive data set capturing plant-specific environmental characteristics of the facilities that provide power to the electric grid and report to the U.S. government. The data set is published as the Emission & Generation Resource Integrated Dataset (eGRID) [2]. The data reported include:

· Renewable and nonrenewable generation
· Air emissions for nitrogen oxides (NOx), sulfur dioxide (SO_2), carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_20).
· Emission rates for CO_2, NOx, SO_2, CH_4, N_20 .
· Heat input.
· Fuel-based and non-baseload emission.
· Net electric generation.
· Others.

The information is published every two years with a year worth of data with the primary objective of providing the necessary information to generate and update:

· Greenhouse gas registries and inventories.
· Carbon footprints.
· Consumer information disclosure.
· Emission inventories and standards.
· Power market changes.
· Avoided emission estimates.

4.2 Data Collection

EGRID reports are available as Excel Workbooks accompanied by Technical Support Documents with different levels of aggregation and can be directly downloaded from the EPA website [2]. We selected the time frame from 2009 to the most recent publication released on February 2018 that presents data for 2016[1]. The Technical Support Documents describe the data collection, the methodology for segregation, and calculations used in the development of the spread sheets.

4.3 Data Preparation

The eGRID presents two Excel workbooks per publication covering metric and English units respectively with multiple levels of segregation (plant, state, balancing authority, eGRID subregion, NERC region, and United States). From each workbook, we isolated the plant data. This data was merged with the EIA data using plant codes and engine and fuel types. The EPA study was prepared and published using data for the year 2016 and extrapolated to previous years. We imputed the missing years with the average data of the contiguous years to allow improve our analysis and provide the visualizations with the proper timelapse look. While rates may have a slight skewness, we assumed a linear rate.

5 Solution Approach for Lazard Data

5.1 Data Description

Lazard publishes the annual Levelized Cost of Energy Analysis to their website. For this study, the two tables of interest are Levelized Cost of Energy Components—Low End and Levelized Cost of Energy Components—High End. These estimate the lowest and highest possible costs to produce electricity based upon the plant type. By taking these two sets of points, we arrive at the estimated average.

5.2 Data Collection and Preparation

We collected data from the past four iterations of the report, years 2015-2018. The data from the website was retrieved by manually entering the datapoints into a spreadsheet. We used three pieces of information for each plant - Capital Cost, Operational & Management Cost, and Fuel Cost. Once completed, the data was manipulated so to be in only two columns. The first column a unique identifier with fuel type and year. The second column the associated cost.

This data was merged with the larger dataset by the unique identifier created adding new columns for each low, high, and average cost. The rows with fuel types which were not presented on the Lazard study were removed for this piece. Our final cost dataset now contains the years 2015-2018 with all the data captured from all three sources.

6 Visualizations and Survey Results

6.1 Dashboard For All Plants and Energy Production

The number of power plants operational and the energy mix is continually changing. We wanted a way to present this information in an easy to comprehend method but also a way to dive deeper into the data. We chose to create a Tableau dashboard to achieve this as it gives the user has the ability to move around an interactive map. The below figure (Figure 1) shows the dashboard. We added several filters to be able to edit the visualization and gain deeper insights into the data and points of interest.

The circles on the map are created by originating the point at each power plant location by their latitude and longitude. The size of the circle is the amount of electricity generated relative to all the other plants. The color changes based upon the type of fuel used. If a power plant uses two separate forms of fuel, there will be two points created. Due to this and to allow smaller plants to appear on the map and not be overtaken by larger peers, we chose a 77% opacity. When scrolling over a datapoint, we chose to show Plant Name, Plant Type Description, Fuel Type Description, Electricity Produced, and Fuel Consumed.

There are six filters the user can select and apply. These are:

· Fuel Category
· Plant Type
· Fuel Type
· State
· Month & Year
· Total MWh Generated

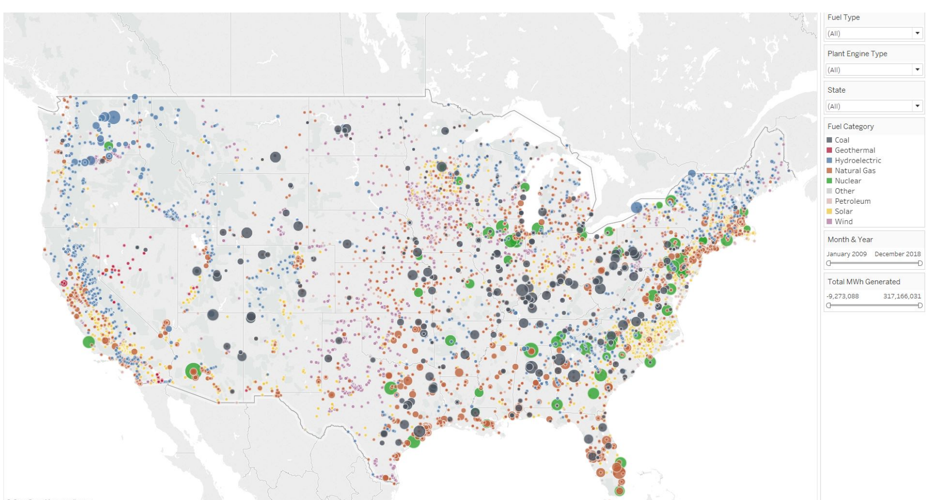

Fig. 1. U.S. Power Plant and Energy Production Dashboard

6.2 US States Fuel Mix for 2017

Of interest to our reader, will likely be the mix of energy production by state. To accomplish this utilized Processing 3.0. Each square in the below figure (Figure 2) represents each of the fifty states. The colored area inside each corresponding square corresponds to the percentage of energy produced for each fuel source in 2017. The same color schema employed in the dashboard has been used. It is a concise way of comparing individual states and to judge the overall U.S. energy mix. The colors for Coal, Natural Gas, and Nuclear encompass most of the image.

The EIA data was subset for only power generated in 2017, then subtotaling by the fuel category for all states and Washington DC and finding the total power generated for each state as well. The data was sorted by State and Fuel Category. A downward and upward cumulative percentage was calculated for each category in comparison to total MWh generated for each state. This was first used to identify the correct category at the 50% threshold. Every state square was then filled with the corresponding color. All other values were then subset into values less and more than that category.

In our visual, each square is 100x100 pixels. The upper left X,Y value pair and the lower right value pair were added to the data. The area of the next colors is calculated from these points as the center point of a ninety degree isosceles triangle. The remaining colors are then added sequentially using Pythagorean's Theorem for area of a triangle. By layering the triangles on top of each other correctly, the image shows the correct percentage for all categories.

When the image is opened, it defaults to the percentages for Total US Fuel Mix. A function was added to track the movement of the mouse. When the mouse hovers over the square for a particular state, the name and percentages will update on the right to match the image.

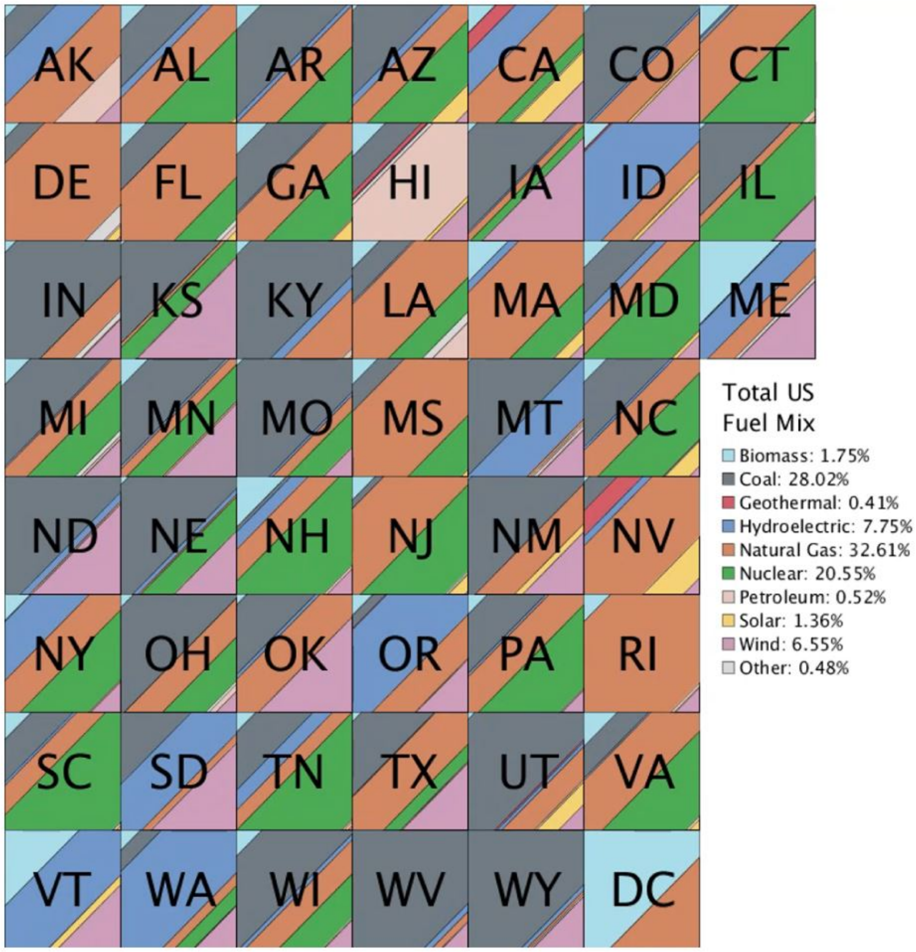

Fig. 2. 2018 Energy Production Mix for Each State

6.3 Power Generation Capacity by Year Built

This visual, shown in Figure 3, captures the maximum capacity of all current and soon to be operational power plants. It can be filtered by scrolling over fuel type option at the right. When filtered, only the selection maintains color while all the other fuel types gray out. This allows the user to see the US still has plants operational from 120 years ago. It also demonstrates the change in plant types over time and sudden influx of wind and solar to the energy mix. By allowing the filtering, it is also possible to compare each fuel type against the others and totals.

This image was created from the EPA eGRID data. All power plants that are inactive and built before 2016 were removed. We then added the corresponding fuel category. The remaining dataset contained the Fuel Category, Year Built, and MW Faceplate Capacity. This was subtotaled by the fuel type creating our final dataset.

The image was created using Processing 3.0. The fuel categories are shown in the same order as the previous two images and shown in the image on the left in descending order and the visual in ascending. To create the bars, the height value for Biomass rectangles were first calculated and placed. The second rectangle of Coal then begins at the height of the Biomass bar and so on. The legend to the left updates the image when a user scrolls over one of the squares. When the mouse scrolls over any of the squares, it triggers the image to update. /hl A grey rectangle is first drawn that is that matches the complete image. The fuel category selected is then draft from the origin in its proper color and bar height.

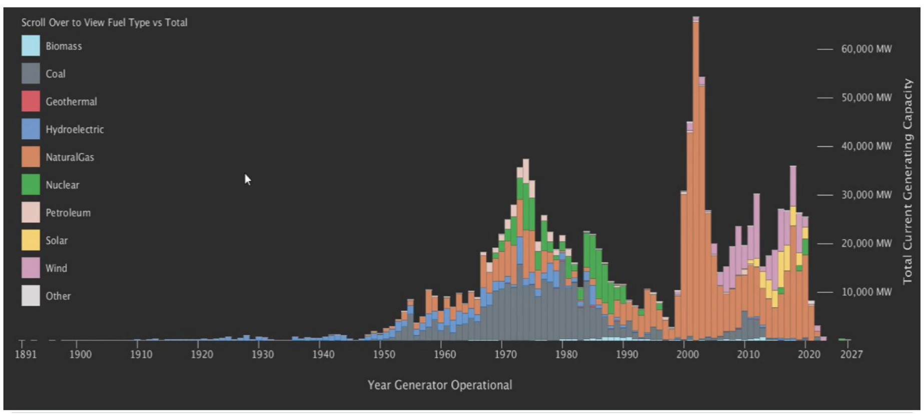

Fig. 3. Current MW Capacity by Year Built

6.4 Create Your A New Power Plant

This visualization was designed to provide a general overview of complicated power plants metrics. Understanding the costs of building and operating a power plant is relevant but difficult to engage a person in. By taking this novel approach, the user can gain knowledge about the financial impact of power plants while discovering the electricity generated from each. By allowing the user to directly compare how many homes can be powered by different plants and its costs, it can be understood that renewable energy has its limits. The user is asked to make 5 selections regarding plant type, faceplate capacity, efficiency level, whether there is a state subsidy and the percentage of public funding. Based on the selections, the display autopopulates with relevant metrics including power information, cost and revenue, cost to build and number of dwellings that could potentially be powered. The initial image of the visual is presented in Figure 4.

The data used for this visual is a mixture of data from eGRID and Lazard. We used the eGRID data to capture plant info such as standard efficiency and how much power a typical plant type generates based upon total MW capacity. Lazard's LCOE provided us the financial aspects of cost to build and operate. This was generated using Processing 3.0. The power plant selections were create from modified free cliparts downloaded from KissClipart[4]. The remaining options were drawn within processing. A mousePressed function embedded with a mouse location loop, allows the system to know when options are selected. Each changes the corresponding value to 1, that updates the image accordingly.

The user must first choose a power plant and MW capacity to begin. These two selections are required to process the calculations that creates the remainder of the image. As a different plant or MW is selected, the calculations are updated accordingly. The fuel efficiency option allows the user to change the MWh created and resulting image by -40%, -20%, 0%, 20%, and 40%. The multiplier defaults to 0%. The subsidy option is only available for Solar, Fuel Cell, Geothermal, and Wind. It calculates 10% of the total cost of the plant. If one of these plants is chosen and "Yes" is selected, the total cost values and bar will update. The Public Financing options also change the total costs.

Of particular concern for this visual are the assumptions that were made in creating it. The pricing metrics were the averages of the Lazard high and low ranges for LCOE. This assumes that Lazard's information is correct and that the mean value can be approximated from this. We also do not show the potential range in cost of the plant. Since we show a single value based upon the plant and MW capacity selected, the price could vary by billions of dollars. We chose to show the mean as it can accurately estimate the cost and provide the user with financial information. The second significant assumption is the efficiency calculations. To find the industry standard, we found the average capacity used by power plants for all fuel categories versus their maximum capacity in 2016. We deemed this calculation to be the industry standard. This estimate could change or be skewed by outliers. We feel it is accurate and a good measurement but it must be noted. The final assumption for this image, is 1,200 MWh can power a house for a full year. This number changes depending on the source. It also changes dramatically based upon the size of the home and electronics used within it. A different value would change what has been presented.

[4] www.kissclipart.com

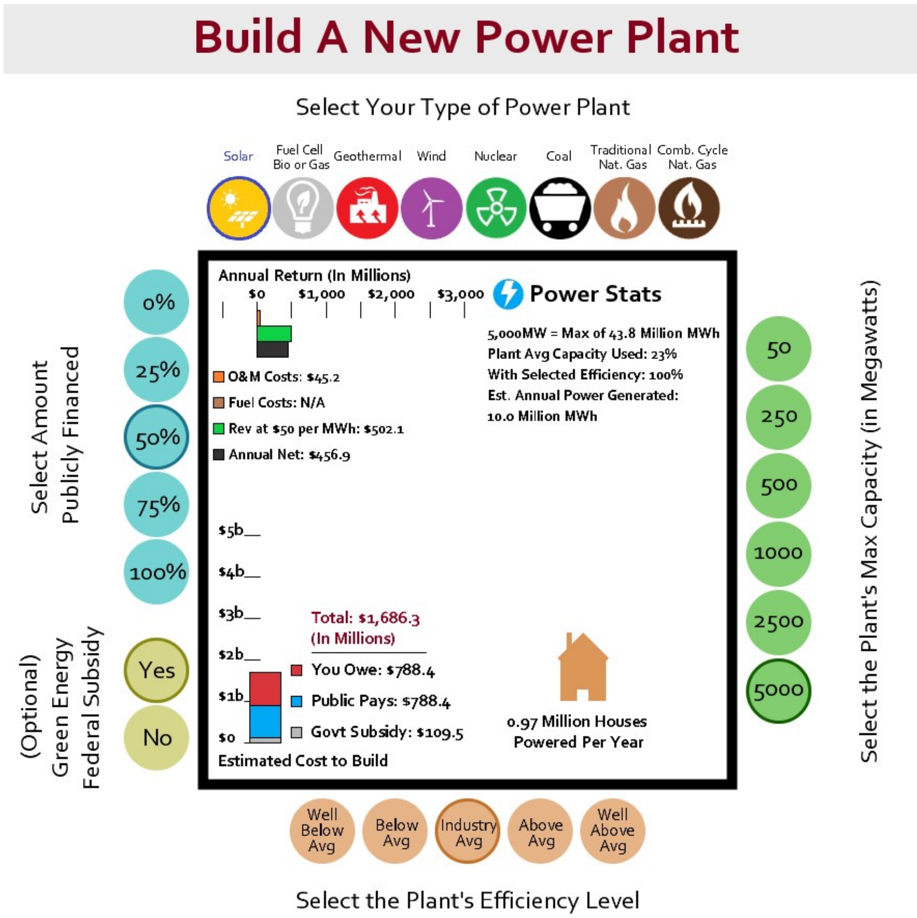

Fig. 4. Build A New Power Plant With Selections

6.5 TimeLapses of US Power Plant Emission by State

Visual present timelapses of the US power plant emissions per state. Emissions are presented as total emissions or emission rates and are broken down into the different emissions components or as an overall total. The user is asked to make a selection of desired emission type and a timeline of the emission from 2009 to 2016 is displayed. The original data was downloaded from the EPA EGRID which presents biannual publications. The missing years where calculated as the average of the contiguous years. This visual was generated using Processing 3.5.3. We used an available svg US map file from Simple Maps[5]. This visual was generated to convey state emission information over the past 10 years. We chose a timelapse visual since they are an effective way to illustrate changes over time.

[5] `https://simplemaps.com/resources/svg-us`

We chose both total emissions and emission rates in order to not only focus on which states generated the most emission but also to show that some states are more efficient with their production and even if their totals are higher due to the amount of electricity generated not an issue with efficiency and vice versa. Figure 5 shows the final image of the CO_2 emission rates.

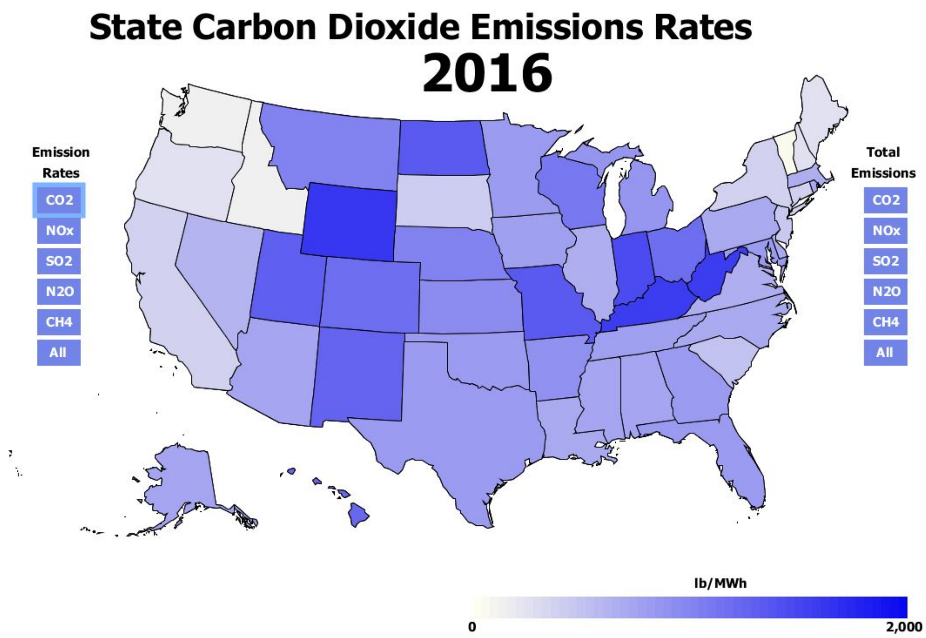

Fig. 5. Timelapse visual showing CO_2 Emission Rates

7 Analysis

7.1 Survey

To assess the effectiveness of the visuals, we conducted a survey through the SurveyMonkey[6] platform. Question designs were based on surveys used for marketing strategies. We were unable to provide the interactive visuals themselves on the survey due to the limitations of the platform. Instead, we captured short videos of interactive visual in use and published these into the survey engine for review. We asked for three responses per visual - two multiple choice and one open-ended text field. For the multiple choice questions, the respondent had the options to choose from Strongly Disagree, Somewhat Disagree, Neither Agree nor Disagree, Somewhat Agree, and Strongly Agree. The questions where:

[6] `https://www.surveymonkey.com/r/BKMLCAP`

· Multiple Choice: Is the above image visually appealing and engaging?
· Multiple Choice: Does the image conveyed its intended message clearly?
· Open-Ended: What do you think is the main purpose of this visualization?

We intentionally did not provide a title or background information on the survey to explain what each image is meant to convey. This allowed us to review the responses and judge how accurate each is in conveying our intended purpose. To arrive at the survey average, we multiplied each of the responses by the corresponding value 1-5, with 1 being Strongly Disagree and 5 being Strongly Agree. This response mean is used with the corresponding percentages we can mathematically determine effectiveness.

We added four more questions at the end to attempt to gauge the visualizations ability to educate and garner interest. These were multiple choice as well. The questions and the available responses were:

· Before taking this survey, how interested were you in US Electricity Production?
· Not at all interested, Not so interested, Somewhat interested, Very interested, Extremely interested
· After viewing the visualizations, has your interest in US Electricity Production increased?
· None at all, A little, A moderate amount, A lot, A great deal
· Before taking this survey, how knowledgeable were you regarding US Electricity Production?
· Not at all familiar, Not so familiar, Somewhat familiar, Very familiar, Extremely familiar
· After viewing the visualizations, has your knowledge regarding US Electricity Production increased?
· None at all, A little, A moderate amount, A lot, A great deal

We received a total of 135 responses. Table 1 summarizes the result of the multiple choice for the visuals. Table 2 summarizes findings of their impact. The results and analysis for each follow below.

7.2 Dashboard For All US Plants and Their Energy Production

Regarding this visual, over 70% founded appealing and engaging and close to 70% thought the message was clear for averages scores of 3.96 and 3.78 respectively. Most respondents understood the visual and its intended message regarding the US power plants, their spacial distribution, the different types of fuel and plant across the US, and power generation and consumption. Few respondents did not know the intended message (about 6%) or found the information confusing, too crowded or overwhelming (about 3%).

7.3 US States Fuel Mix for 2017

Regarding this visual, about 65% found it appealing and engaging, and about 70% thought the message was clear for averages scores of 3.60 and 3.73 respectively. Most respondents understood the visual and its intended message regarding fuel mix usage per state. Few respondents did not understand the image (about 5%) or found information confusing (about 4%). Several iterations of this visual were presented in the survey as the visual was improved to incorporate suggestions or address issues. As negative responses and comments arrived, it was edited based upon the feedback.

The first version was a static image as currently shown but we arranged the 50 states in a 5x10 state rectangle. It was too tall to view without scrolling and was not understood by the user. The average survey responses were approximately 2.5 for both appeal and comprehension. The second iteration adjusted the sizing to be viewed on the screen all at once including Washington DC but performed only marginally better. The final and current version added the shown percentages and the interactive features. The responses from this change were overwhelmingly positive.

7.4 Power Generation Capacity by Year Built

Regarding this visual, about 70% found it appealing and engaging, and close to 70% thought the message was clear for averages scores of 3.75 and 3.81 respectively. Most respondents understood the visual and its intended message regarding the power generation capacity over time by fuel type. Few respondents did not understand (about 3%) or found information confusing (about 2%).

7.5 Create Your Own Power Plant

Regarding this visual, close to 75% found it appealing and engaging, and over to 70% thought the message was clear for averages scores of 3.93 and 3.89 respectively. Most respondents understood the visual and its intended message regarding power plant metrics. Few respondents did not (about 3%) or found the information confusing, too crowded or overwhelming (about 15%).

7.6 TimeLapses of US Power Plant Emission by State

Regarding this visual, close to 90% found it appealing and engaging, and over to 85% thought the message was clear for averages scores of 4.42 and 4.27 respectively. Most respondents understood the visual and its intended message regarding total emission and emission rates per state over the past 10 year.

Table 1. Visualization average survey scores

Visualization	Average Appeal Score	Average Message Conveyed
Power Plant Dashboard	3.96	3.78
2017 Fuel Mix	3.6	3.73
Power Generation Capacity	3.75	3.89
Create Power Plant	3.93	3.89
Emission Timelapses	4.42	4.27

Tables shows the average score of survey responses per each visualization

7.7 Overall

Regarding our survey, close 40% of the surveyors reported a previous moderate interest in the US Electricity production with the remaining almost equally divided between not interested or highly interested. The average interest in the US electricity prior to taking the survey was averaged at 3.09. After completing the survey, close to 50% reported a moderate or significant interest increase for an average interest increase score of 2.76. Regarding knowledge, only about 13% of the surveyors reported significant previous knowledge, while more than 60% reported little to no previous knowledge of the US Electric production for an average prior knowledge of 2.76. After viewing the visualization about 60% reported at least a moderate knowledge increase for an average knowledge increase of 2.87. Most surveyors found the visuals useful and informative and reported their interest in interacting with the visuals themselves.

Table 2. Visualization Results

	Interest Score	Knowledge Conveyed
Average Prior to survey	3.09	3.78
Average Increase After survey	2.76	2.87

Average scores reflect surveyors prior knowledge and interest and the impact and effect of the visualizations

7.8 Ethics

There are several ethical issues to consider in our analysis. One is the implications of the takeaways of the people reviewing our visuals. Our intent is to inform and educate without providing any bias toward a specific conclusion. This may be difficult though, since there is an inherent negative connotation to emissions. By providing emissions data, we may bias the user towards an unintended conclusion. We have taken steps to avoid this by trying to simply represent the data. We also have only provided a very small piece of the energy production industry.

There is much more to learn and investigate. We would strongly advise anyone who views these images to research further into the topic. With these visuals and some additional research, it is possible for the user to be properly aware of the subject and make decisions regarding policy preferences.

Given the public discourse regarding energy sources, we must consider the potential influence of our images to public policy decision and how they could affect customers. Inadequate or poorly planned changes in energy source might result in energy shortages due to insufficient energy storage, or improper transmission, thus compromising the electricity supply to the customers. Switching too many or to quickly from fossil fuels to renewable sources could cause deficient energy service to a general area. When blackouts are already a concern, this could unintentionally exacerbate the problem. In trying to increase efficiency in power plants and thereby lowering emissions, detrimental influence on the power production and transmission itself must be considered.

It is also important not to overlook how energy source replacement could affect the workforce associated with the power plants and fuel sources. The coal industry has lost 50,000 jobs over the past five years and its market share continues to decline. A massive migration to a different power source would have an immediate impact on the workers there and downstream effects on the workforce producing the fuel itself. In its place, has been a significant increase for jobs in natural gas and renewable. The families relying on coal mining have had to relocate and some permanently unemployed. If coal were eliminated or natural gas changing direction and losing market share, tens of thousands of workers and their families would be affected. While no direct solution to this issue, our concern remains.

Our last significant concern is the impact to population centers and geography. While a power plant may be inefficient, if it is the primary service of a nearby major US city, the benefit of replacing or refurbishing it may be outweighed by the detriment to the city. Likewise, when considering to build a new plant, it must be built where there aren't homes already in the vicinity. Options are thereby limited to sparsely populated areas or where the existing population can accommodate a plant. This would likely mean some houses would need to be purchased and torn down. It also would require clearing out land area and building access roads and other required buildings. In a mostly developed country, it may be difficult to establish locations that are more suitable than the existing plants. Environmental destruction for a new power plant is not desirable either. Because of this, whenever possible, recommendations for refurbishing or replacement of plants on existing locations may be ideal.

8 Conclusions

We found that some visuals were more intuitive, user friendly and descriptive than others leaving room for visual improvement. Novel approaches like "generating a new plant" can convey deeper and larger amounts of information if presented well. The scores received for that visual were much higher than average.

Finally, our best performing visual was the time lapse of emission production. We believe time lapses can be used to communicate the effects over time very efficiently. With these visualizations we were able to improve the accessibility the complicated US energy production data showing that visuals can educate and create interest.

Our intended goals were to provide visualizations that people enjoy using and to increase topic knowledge. The energy production data the visuals are created from is both complicated and inaccessible despite being publicly available. The survey responses for the five visuals received a combined average score of 3.93 with respect to appeal and 3.89 for effectiveness in message conveyed. With these scores, we can cautiously conclude our visualizations are accomplishing our stated goals. In addition, we were interested in the background of our survey respondents. We asked for prior knowledge and interest on the subject and if either have increased as a result of viewing what we created. We were able to increase the respondents interest in the US electricity generation to a moderate degree with a average score of 2.76 with an average surveyor prior knowledge of 3.09. Similarly, we were able to moderately educate the surveyors with an average knowledge increase of 2.87 for respondents that presented some prior knowledge averaged at 2.41. This indicates the topic of energy production is not something the public is aware of nor have much interest. These scores do suggest our images allow insights and learning of the subject, despite the lack of prior knowledge.

For future work, we would distribute our survey and visualizations to a wider audience to gauge additional response data. While we improved visuals as responses arrived, additional responses and iterations could refine the visuals further. We also would explore additional avenues for more visuals on the topic. While we focused on five different topics for each visual, there are a number of simple charts and graphs that would be effective in displaying one piece of information. These would bolster the visuals we have and would simplying some aspects to a single data point. Finally, we would expand upong the Tableau report. The chart, filter, and tooltip display numerous data. We explored adding the year each plant was built, emissions data, and estimated operating costs. This seemed to be too much data for one dashboard. We would explore options to expanding the report and building additional visuals.

References

1. United States Environmental Protection Agency. The emissions and generation resource integrated database - technical support document for egrid with year 2016 data. United States - Environmental Protection Agency, 2018. `https://www.epa.gov/sites/production/files/2018-02/documents/egrid2016_technicalsupportdocument_0.pdf`.
2. United States Environmental Protection Agency. The emissions and generation resource integrated database (egrid). United States - Environmental Protection Agency, 2018. `https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid`.

3. United States Environmental Protection Agency. Fact sheet - proposed ace rule - co_2 emission trends. United States - Environmental Protection Agency, 2018. https://www.epa.gov/sites/production/files/2018-08/documents/ace_trends.pdf.
4. United States Environmental Protection Agency. Draft inventory of u.s. greenhouse gas emissions and sinks: 1990-2017. United States - Environmental Protection Agency, 2019. https://www.epa.gov/sites/production/files/2019-02/documents/us-ghg-inventory-2019-main-text.pdf.
5. Jean-Luc Bessède. Eco-friendly innovations in electricity transmission and distribution networks. Elsevier, 2014.
6. Colubri. Processing for andriod: Create mobile, sensor-aware, and vr applications using processing. Apress, 2017.
7. Kumar Greenberg, Xu. Processing. Friendof, an Apress Company, 2013.
8. Dan Keating John Muyskens and Samuel Granados. Mapping how the united states generates its electricity. The Washington Post, 2017. https://www.washingtonpost.com/graphics/national/power-plants/?noredirect=on&utm_term=.b4c84724cf78.
9. Lazard. Lazard's levelized cost of energy analysis —version 12.0, 2018. https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2018/.
10. Noble. Programming interactivity, second edition. O'Reilly Media, Inc., 2012.
11. Processing.org. Processing overview. https://processing.org/overview/.
12. Shiffman. Learning processing. Morgan Kaufmann, 2008, 2015.
13. EIA Independent Statistic and Analysis U.S Energy Information Administration. Electricity data browser. EIA U.S. Department of Energy, (accessed February 8, 2019). https://www.eia.gov/electricity/data/browser/.
14. EIA Independent Statistic and Analysis U.S Energy Information Administration. Energy explained - your guide to understanding energy. EIA U.S. Department of Energy, (accessed February 8, 2019). https://www.eia.gov/energyexplained/index.php.
15. Tableau. Tableau products and solutions. https://www.tableau.com.