Noah Fehr

Documentation

March 14, 2022

CSC 324

Web/online sources: None

Help obtained: None

*I confirm that the above list of sources is complete AND that I have not talked to anyone else about the solution to this problem*

Individual Visualization: Documentation

**Basic Documentation**

Purpose

*The overarching purpose of this project is to provide users with meaningful visualizations so that they can understand the development of renewable energy production over time and across the globe. While this project will not crease an exhaustive resource for understanding renewable energy production, I do intend it to enhance viewers' background understanding and lay the foundations for further inquiry. Furthermore, I hope these visualizations can present the developments in such a way that users begin to recognize patterns which coincide with historical and current events defining the renewable energy sector over the past twenty years.*

Data Description

*These visualizations are based on data from multiple datasets, which work effectively together to help achieve the purpose defined above.*

1. *renewablesPowerGeneration97-17 contains the production of renewable energy over the twenty years between 1997 and 2017. Furthermore, it specifies how much of this energy was from hydro, solar & geothermal. Each row represents a single observation (year) and each column represents a single attribute*
2. *top20CountriesPowerGeneration contains the data from 2017 of each of the top 20 countries’ renewable energy generation. Each row is a single observation (country) and each column represents a single attribute.*
3. *Country\_Consumption\_TWH 2 has the total energy consumption for each year, broken up by country. Each row is an observation (year) and each column is a country. Thus, the cells represent the energy consumed by that country each year.*

Data Collection

*This data was found on Kaggle; however, it was originally collected by and sourced from: Nature-Inspired Optimization Algorithms for Renewable Energy Generation, Distribution and Management—A Comprehensive Review by Vamsi Krishna Reddy Aala Kalananda and Venkata Lakshmi Narayana Komanapalli from Intelligent Paradigms for Smart Grid and Renewable Energy Systems. - 2021. Springer*

Intended Users

*This visualization is intended for a general audience of university students interested in renewable energy development and policy. The information conveyed via this visualization will help to provide these students with perspective, insights, and curiosity for their studies.*

Questions to Answer

*The questions to outline are included on each section; however, they are also included here for complete documentation:*

1. *How has the composition of renewable energy production changed over time? What trends have defined these changes?*
2. *Is there a relationship between total energy consumption and renewable energy production?*

Insights

1. *Over the two decades represented by this data, total production of renewable energy grew significantly. Within this growth, we can see a significant expansion in energy sources like solar and geothermal energy which represented a tiny portion of total renewable energy in 1997 and now represent a substantial proportion of total renewable energy production.*
2. *There is a loose positive relationship between total energy consumption and renewable energy production. Obvious outliers to this rule include Brazil and Canada, who produce high amounts of renewable energy relative to their total energy consumption, and Iran, who produces less than expected.*

Improvements

*Although this visualization establishes a strong basis for understanding recent developments in renewable energy policy, there is significant room for further improvements in future work / visualizations:*

1. *This visualization fails to provide significant context for the trends in renewable energy composition. Instead, users are left to view the trends and hypothesize how these came to be. Future studies may incorporate relevant current / historical events as appropriate. Enerdata created a visualization in regard to energy consumption and the pandemic that included specific contextual points. Additional visualizations may draw inspiration from this work.*
2. *The limitations of these datasets fail to illustrate these trends over longer periods of time or in recent years. Future visualizations could incorporate more comprehensive data.*
3. *To better understand the impact of government action on renewable energy, future apps could group countries in the scatterplots by government policies / actions*
4. *Stronger UI design in more flexible frameworks, such as React, would allow for better interactivity / reactivity and thus a more presentable UI*

Sources / References

*This data was found on Kaggle; however, it was originally collected by and sourced from: Nature-Inspired Optimization Algorithms for Renewable Energy Generation, Distribution and Management—A Comprehensive Review by Vamsi Krishna Reddy Aala Kalananda and Venkata Lakshmi Narayana Komanapalli from Intelligent Paradigms for Smart Grid and Renewable Energy Systems. - 2021. Springer*

**Proper Documentation**

Process and Development

*The process and development of my application was divided into two distinct processes: one for the visualizations answering the question of “How has renewable energy production changed over time?” and another for those answering the question “Is there a relationship between total energy consumption and renewable energy production?”.*

*To understand the changes in composition over time, I began by renaming the columns of the provided dataset (yearbyyear.csv) for readability. Then, for the pie and bar chart visualizations, the year\_by\_year data was filtered to focus on a specific year (for the stacked chart, the data was filtered for all years until the current year). To tidy the data such that each row represents one observation: the renewable energy produced from one specific source in that given year. Finally, to ensure that each chart was colored in the same way, I set the types accordingly. This guarantees that ‘Biofuel’ will be colored in the same way for the pie, bar and stacked chart.*

*Once the data was prepared as above, the charts were built within the ggplot2 framework. These plots were then configured in the Shiny format and given input from the slider on the ‘Composition’ tab.*

*To explore the relationship between total energy consumption and renewable energy production, I first renamed the columns in the countrygeneration.csv dataset and selected the two columns of interest: Country and Total Renewable Energy Produced. Then, I focused on configuring the countryconsumption.csv with the intent of combining these two datasets into one. First, I filtered the consumption data to only include data from 2017, since the countrygeneration dataset only includes data from 2017. Next, I worked to create a data frame with countries and the numeric representation of their consumption (originally stored as strings). Once configured, I merged these two datasets into a tidy dataset such that each row is an observation representing a given country’s energy consumption and renewable energy production in 2017. To highlight China in the first plot, I created a no\_outliers dataset that excluded China, and plotted China over this dataset in red.*

*Like the composition charts, I then applied this data to the ggplot2 framework and created the visualizations in my R Shiny App.*

Design Decisions

|  |  |
| --- | --- |
| Idiom | *Word Cloud* |
| What: Data Abstraction | The data is a table in which the items are countries and the attribute is the amount of renewable energy produced in 2017 in TwH (Ordered, Quantitative). |
| Why: Task Abstraction | This visualization is intended to consume and present data on the amount of renewable energy produced in 2017. The primary target of this visualization is all data to identify outliers (in this case, countries with particularly large names due to their high renewable energy production). |
| How | By mapping the size of the country name and its hue by the amount of renewable energy produced in 2017, this visualization leverages users’ ability to differentiate both size and color to emphasize the different in production between countries and highlight the primary producers. |

|  |  |
| --- | --- |
| Idiom | *Bar Chart* |
| What: Data Abstraction | The data is a table in which the items are the different renewable energy sources, and the attribute is the quantity of energy from that given source (Ordered, Quantitative) in the provided year. |
| Why: Task Abstraction | This visualization is intended to consume data to present the changes in renewable energy composition to users, to empower users to discover trends in these relationships, and for general user enjoyment. The target of this data is a single attribute (quantity of renewable energy produced) with the goal of exploring the distribution. |
| How | Like all bar charts, this visualization expresses quantitative data values (renewable energy from a given source) with height in a spatial position with aligned axes. To encode the relative composition of renewable energy production, the bar charts have been aligned and mapped with color. These design decisions allow for easy differentiation and comparison between renewable energy sources. The colors chosen are intentionally different hues due to the categorical nature of the different sources of renewable energy. Finally, the data displayed has been filtered to only show the data provided in the specified year. |

|  |  |
| --- | --- |
| Idiom | *Pie Chart* |
| What: Data Abstraction | The data is a table in which the items are the different renewable energy sources, and the attribute is the quantity of energy from that given source (Ordered, Quantitative) in the provided year. |
| Why: Task Abstraction | This visualization is intended to consume data to present the changes in renewable energy composition to users, to empower users to discover trends in these relationships, and for general user enjoyment. The target of this data is a single attribute (quantity of renewable energy produced) with the goal of exploring the distribution. |
| How | The pie chart represents mapping of both: size to convey the difference in quantity produced, and hue to differentiate the different sources of renewable energy. Like the bar chart, hues were chosen to represent the exclusive, categorical nature of the different renewable energy sources, and the data was filtered to only include the specified year. |

|  |  |
| --- | --- |
| Idiom | *Stacked Chart* |
| What: Data Abstraction | The data is a table in which the items are the different renewable energy sources in a specific year, and the attribute is the quantity of energy from that given source (Ordered, Quantitative). |
| Why: Task Abstraction | This visualization is intended to consume data to present the changes in renewable energy composition to users, to empower users to discover trends in these relationships, and for general user enjoyment. The target of this data is a single attribute (quantity of renewable energy produced) with the goal of exploring the distribution as well as trends over time. |
| How | The stacked chart uses height in spatial positions to illustrate the production of renewable energy from each source. Rather than aligned axes like the bar chart, this visualization features stacked quantities, as we are interested both in the composition and the developments in overall renewable energy production as well (ie the change in the sum total is also important, thus the stacked chart). The same logic held here with using hues for different sources of renewable energy. The data is filtered to reflect all of the data up until and including the specified year. |

|  |  |
| --- | --- |
| Idiom | *Scatter Plot* |
| What: Data Abstraction | The data is a table in which each item is a country with two attributes: the total energy consumption in 2017 (Ordered, Quantitative) and renewable energy production in 2017 (Ordered, Quantitative). |
| Why: Task Abstraction | This visualization is intended to query data and identify outliers for further inquiry as well as consuming the data to discover trends in the relationship between renewable energy produced and total energy consumed. The targets of this data are both trends and outliers, which will help to answer the second question of focus and identify countries for further exploration. |
| How | The scatter plot with all of the data points is manipulated to select and highlight a specific point, China. As an outlier, China is highlighted with a red hue to emphasize that it is categorically different than the other points (it is an outlier). The points on this graph represent a given country and its renewable energy production and total energy consumption in 2017. Additionally, selection has been added with interactivity such that a user can click on a given point of interest and the app will show the name of this country. |

|  |  |
| --- | --- |
| Idiom | *Scatter Plot (No Outliers)* |
| What: Data Abstraction | The data is a table in which each item is a country with two attributes: the total energy consumption in 2017 (Ordered, Quantitative) and renewable energy production in 2017 (Ordered, Quantitative). |
| Why: Task Abstraction | This visualization is intended to query data and identify outliers for further inquiry as well as consuming the data to discover trends in the relationship between renewable energy produced and total energy consumed. The targets of this data are both trends and outliers, which will help to answer the second question of focus and identify countries for further exploration. |
| How | This visualization builds upon the scatter plot above and filters the data to exclude China, the outlier highlighted previously. This chart has the same interactivity to select a point and display the name of that country. |