Summary Report

By Vishal, Pradham, Komal

Precondition

1. *Learning*

   The US Department of energy has tasked our group with identifying patterns within the energy sector to generate policy recommendations based on past data. Taking this into consideration, we have decided to conduct a comparative analysis of state wide energy consumption to discern which states are utilizing the most energy - broken down by both sector and source. We will also focus on the expenditures of each state and compare states that use renewable vs non-renewable energy in order to form policy recommendations that encourage healthier energy usage. This is an important analysis as using non-renewable energy has been proven to have negative effects on our environment and if we are able to show, through data, that renewable energy is not only beneficial for the environment but also for state expenditures/consumption, then we will have a more compelling argument for our policy recommendations.

2. *Winnowing*

   For our analysis, we will be using the dataset from the US Energy Information Administration. From this dataset, we will use the following columns:

   - State
   - Consumption
   - Expenditure

   Consumption and Expenditure are both broken down by Sectors and Sources. The sectors being Commercial, Electric Power, Industrial, Refinery, Residential, and Transportation and the sources being Coal, Distillate Fuel Oil, Geothermal, Hydropower, Kerosene, Petroleum, Natural Gas, Solar, Wind, and Wood. However, the dataset does miss expenditure data for all the renewable energy sources: Geothermal, Hydropower, Solar, and Wind. This will make our data analysis a little bit more difficult as we cannot compare expenditures when trying to prove a shift to renewable energy. In addition to this, several values in the dataset are assigned the value of 0 if it is not reported, so our analysis is impacted by this as well.

   Regardless of the shortcomings of the data, we will still be able to conduct a comprehensive breakdown of energy consumption and reflect on how expenditures may compare to those states using more renewable energy sources. Since we are also only focusing on current data, we are not concerning ourselves with the Year column, we are simply going to use the most recent year in the dataset as our means of comparison, which is 2019.

Core

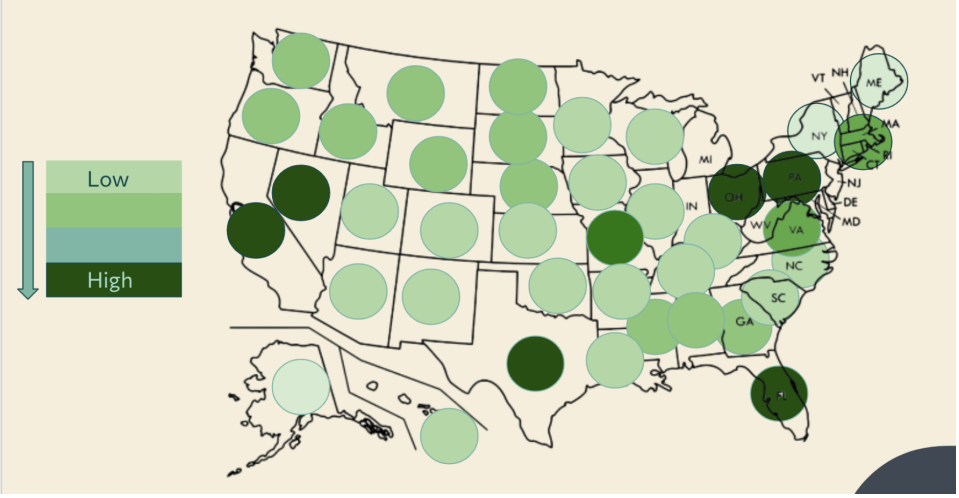
3. *Discover*

   With a think-aloud process, our group came up with two initial tasks that we wanted to conduct with our data. We discussed how we would like to explore data surrounding energy consumption and concluded that a comparative analysis across states would be the best approach. Therefore, our first task was to *develop a map with an overview of energy consumption by state* and our second task was to *provide further breakdown of energy consumption by sector and source upon clicking a state*.

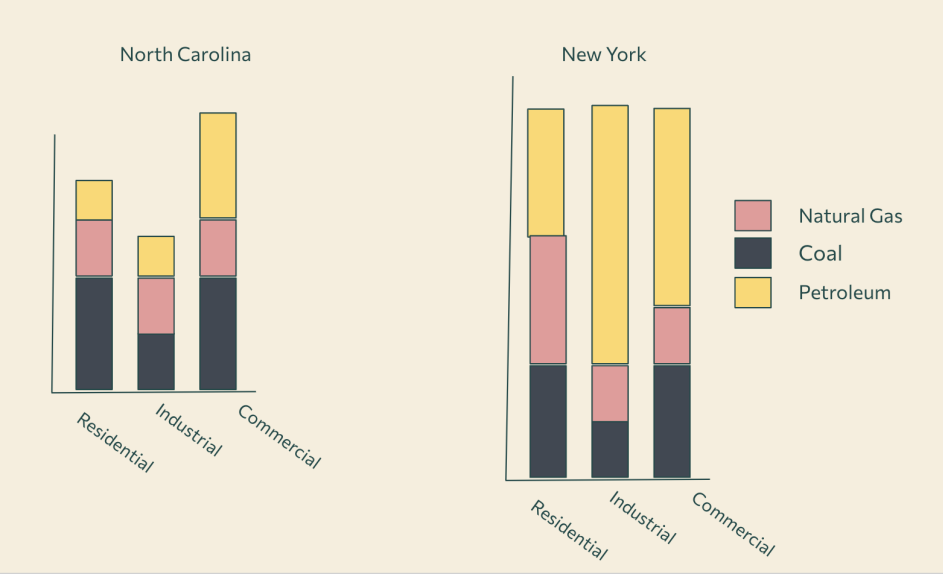
   The 5Ws of the first task are as follows:
   Who is executing the task: Those analyzing the data to derive insights, in this case, us.
   What does a task seek to learn about the data: How does energy consumption vary by state.
   Where does the task operate: The task utilizes data provided by the US Department of Energy to summarize energy consumption.
   When is the task performed: As part of an initial exploration into US energy data for creation of policies.
   Why is a task pursued: To show a comparison between state consumption and emphasize the most energy usage.
   How is a task performed: By aggregating all energy consumption data by state.

   The 5Ws of the second task are as follows:
   Who is executing the task: Those analyzing the data to derive insights, in this case, us.
   What does a task seek to learn about the data: The distribution of energy consumption across different sectors like residential and industrial as well as across different sources.
   Where does the task operate: The task utilizes data provided by the US Department of Energy to summarize energy consumption.
   When is the task performed: As part of a deeper exploration into US energy data for creation of policies.
   Why is a task pursued: To highlight the different areas of energy usage within the different states.
   How is a task performed: By grouping consumption data by its sector and by its energy source.

4. *Design*



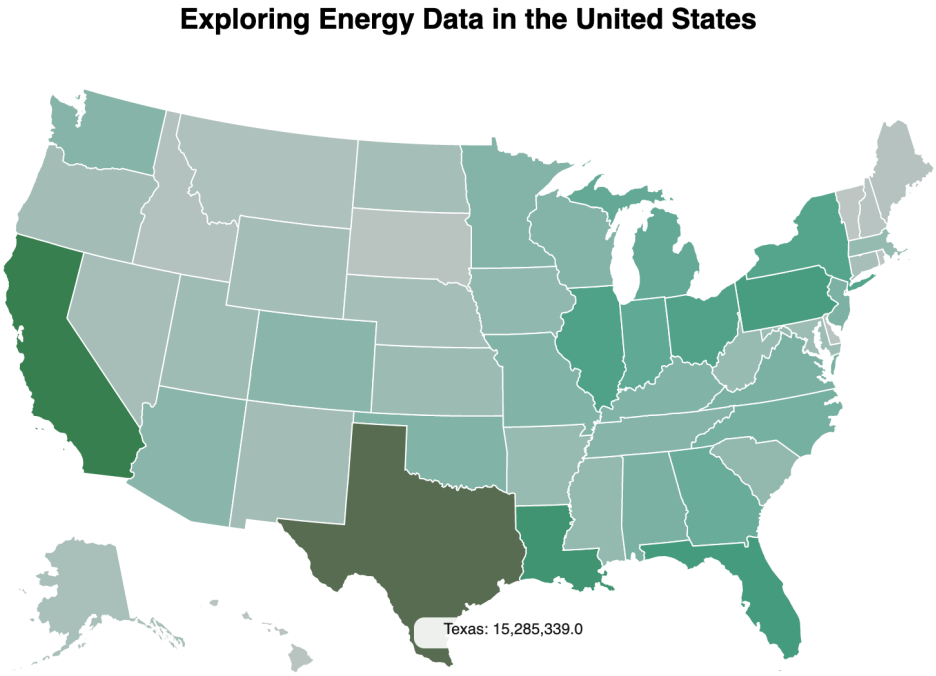
The image above represents a low fidelity prototype of a visualization for Task 1. This addresses the comparison of overall consumption at a state level through the use of a bubble chart across the United States map. The intensity of the color represents the amount of total energy consumption. Because of the colors and gradient, any user of the tool can identify which states consume the most energy by taking a quick glance at it. However, the size of the bubbles can be misleading and can imply other results, therefore, it is best to use these bubble charts with at least 3 dimensions.



The image above is a low fidelity prototype of a visualization for Task 2. Upon the click of a state on the map (from task 1), the user will see a stacked bar chart that breaks down the consumption of energy across sectors and sources. We hope to, in the actual vis tool, allow a user to see these stacked bar charts for up to 3 states so they can compare the breakdowns of energy consumption in an easier manner. The color coding helps users understand the different categories, and these colors will not be a gradient as there is a

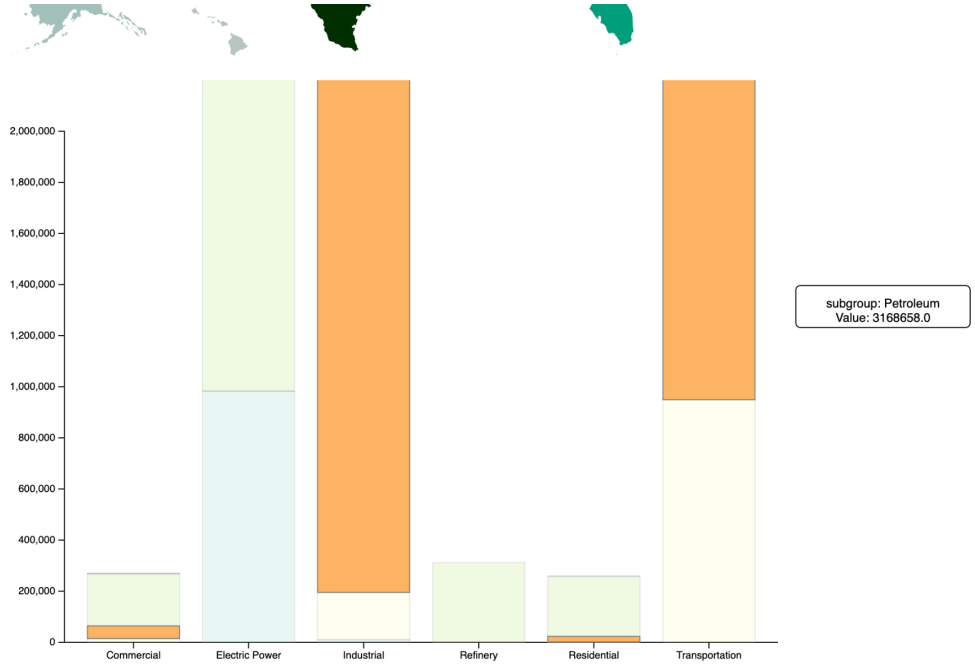
categorical difference between these different values, not gradual like the map from task 1. We do realize that stacked bar charts can be more difficult to read, and when you want to compare each segment to another it is more difficult, as they're not aligned on a common baseline, however a stacked bar chart will allow us to compare the sectors and sources in a more organized and easier way.

5. *Implement*
   We implemented our design using D3.js. Instead of our initial prototype of a bubble chart on the United States map, we decided to implement a choropleth map as it allows users to see states with higher energy consumptions without the size of any bubbles confusing them. Upon hovering over a state, you can view the state name and its total energy consumption across all sectors and sources.



In addition to this, from our prototype of task 2, we created a stacked bar chart, with the x axes representing all the different sectors and the y axis representing the total consumption of energy, split into individual bars by energy source. A user is able to click on a state and view its respective stacked bar chart in order to compare the use of energy. You can see this in the example below. The state Texas was clicked on the map above, and a stacked bar chart with its energy consumption across different sectors and sources shows up. You can hover over individual parts of the bars, which are color coded by energy source, and view how much of that sector exactly uses a specific energy source.

For example, in this picture below, we can see that Industrial and Transportation sectors use a lot of Petroleum.



We made our designs like these because they fulfill our task requirements and allow us to form some thoughtful insights when it comes to policy-making. The choropleth map allows us to immediately identify the states with the highest level of energy consumption, which gives us a general idea of any trends across states. However, to further analyze this data, we decided it was necessary to have a full breakdown of consumption by both sectors and sources of energy, which is exactly what our second task and design fulfills. It allows us to identify the dominant sector and source of energy for specific states. This design also allows us to compare states to each other and their consumption patterns, which can reveal several other factors of energy consumption. We can use these visualizations to make predictions and suggest what sources of energy may be better for consumption for specific sectors.

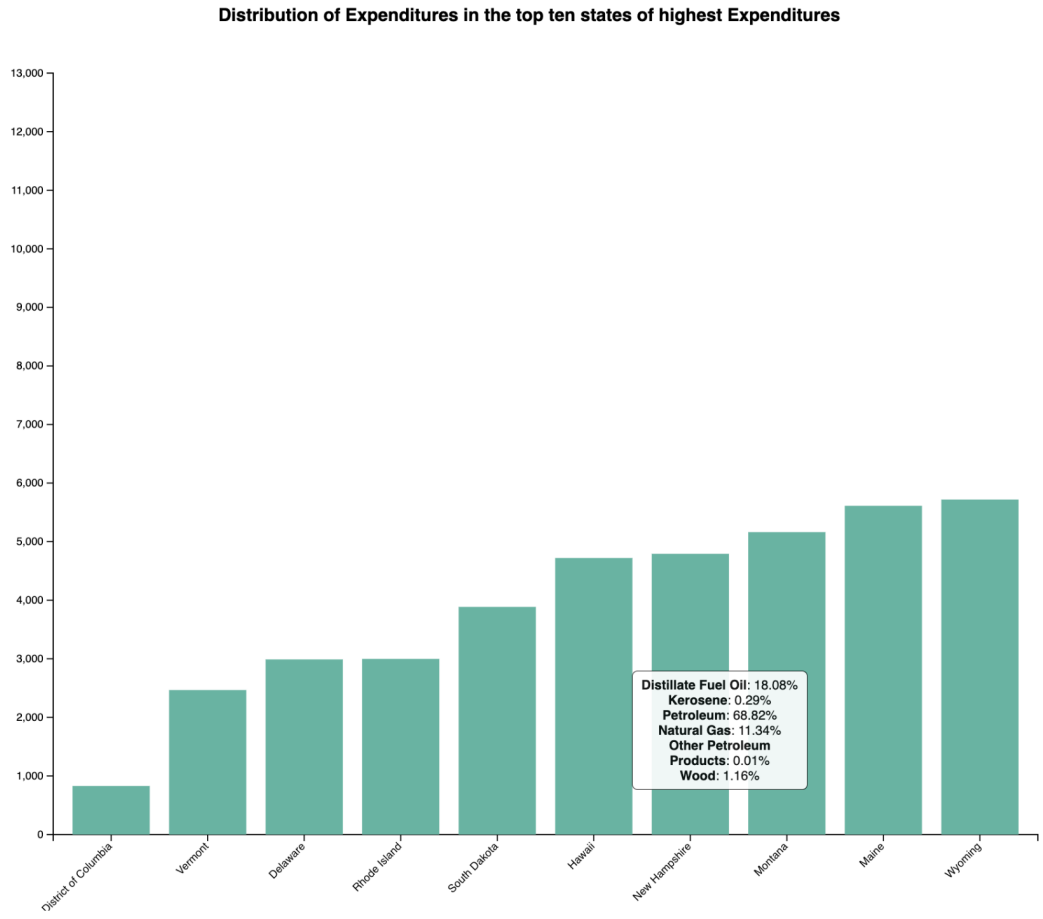
6. *Deploy*

Upon implementation of these visualizations, we were able to understand the trends in the energy consumption data. With task 1, we're able to see which states have the highest energy consumption at first glance just by the intensity of the colors, the darker the color, the higher the consumption. And with this, we are able to see that Texas, California, Florida, New York, and Pennsylvania are some of the states with the highest energy consumption. For further understanding of these details, our tool for task 2 provides even more details. By viewing the energy consumption breakdown by sector and source

through a stacked bar chart, we are able to understand where most of the energy is being consumed, and through further insight, potentially how it can be reduced as well. For example, when we look at California on the map, we can see that there is a lot of energy being consumed there, but we don't know how/where. With this stacked bar chart, we can see that its dominant source of energy is Petroleum, and the sector that uses most of this is Transportation, which can help us further analyze the Transportation industry in not only California, but other states as well. In fact, when you look at other states' bar charts, it's evident that Petroleum is the most commonly used energy in the Transportation sector, so with policy recommendations to change that, we can maybe reduce our carbon footprint.

7. *Iterate*

   Upon developing visualizations for the first two tasks, we also realized that it may be helpful to view additional data on the states that spend the most money on energy, therefore we developed a new task, *task 3: determine the states with the highest expenditures and view their distribution across different energy sources*. To combat this task, we developed a bar chart of the top 10 states with the highest expenditures, and upon hovering over it, you can view what percentage of those expenditures are spread out throughout the different energy sources. And not surprisingly, Petroleum reigned as the dominant source of energy across all expenditures. Take, for example, New Hampshire, in this picture below, it is the fourth highest expenditure state of which almost 70% consists of Petroleum consumption. But when we look at it on the map, it doesn't seem to have as high of an energy consumption overall, which means despite its low consumption of energy, it has high expenditures. This can be addressed by changing energy sources to something that is more energy efficient and affordable in the long term such as Solar, Wind, Geothermal, or other renewable energy sources.



It would also be interesting, for later iterations, to implement a visualization that updates the rest of the visualizations upon some sort of interaction, this can make all these different types of visualizations connect to each other and allow users to make more sense of the information being presented to them. Currently, it is more of a dashboard with a variety of data exploration tools, but creating something more connected to each other may make the solutions we suggest more visible and clear to any user.

Analysis

8. *Reflect, Pt 1*

   The visualizations that we have implemented paint a very clear, yet worrying picture showing the dependence of states on non-renewable energy sources. Not only is the impermanence of these sources a matter of concern but the side effects of processes used to extract them are a more immediate threat to our environment. Industrial processes like offshore oil drilling and fracking run the risk of leaking/spilling harmful gasses and substances into our environment thereby polluting our oceans and climates. Judging by the data and visualizations, there is an urgent need to shape future policies to encourage

the hastened adoption of renewable sources of energy which don't inflict as much damage to our world.

9. *Reflect, Pt2*

   The visualizations were created in such a manner as to engage viewers and ensure to express the necessary information to allow the DOE to make new policies. When we constructed the heat map we saw a way to display information in an easy to understand manner that would allow the viewer to quickly understand which states were consuming the most amount of energy. The interactivity of this visualization allowed for the viewer to dive deeper into the data and understand what kinds of energy was being used by each state. Finally, we wanted to ensure that the actual monetary cost of energy usage was understood by the department of energy, so the top 10 spenders on energy were highlighted and the percentage split was also shown.

   Throughout all of these designs, we also had to ensure they were accessible and easy to understand, which is why we chose to add tooltips and color palettes that are aesthetically pleasing, because while the data is important, the way it is presented is also an important factor.