

## COMP 590 173

### Module 3 Write-Up

Aayush Mehra, Vidyuth Jagadeesan, Pranav Vinaayak

The public link to our repository:

<https://github.com/COMP790-InfoVis/module-three-design-study-lite-aayush110>

1.

The main problem that we will be focusing on is the consumption of commercial petroleum within all fifty states of the U.S. and Washington D.C., from 1960 to 2020. According to the U.S. Energy Information Administration, approximately 134.55 billion gallons (or about 3.20 billion barrels) of finished motor gasoline were consumed in the United States in 2022, which averages to about 369 million gallons per day (or about 8.78 million barrels per day).<sup>1</sup> Petroleum is devastating to the environment mainly due to its contribution to air pollution. Gasoline, in particular, is a significant source of air pollution because when it is burned, it emits carbon monoxide, nitrogen oxides, particulate matter, and unburned hydrocarbons, which are known to be major air pollutants<sup>2</sup>. Therefore, our data will focus on gaining a better understanding of petroleum consumption over time across the United States. We will also compare

---

<sup>1</sup> This data was found from the below source:

<https://www.google.com/url?q=https://www.eia.gov/tools/faqs/faq.php?id%3D23%26t%3D10%23%26text%3DIn%25202022%2520C%2520about%2520134.55%2520billion.8.78%2520million%2520barrels%2520per%2520day&sa=D&source=docs&ust=1682086780395198&usg=AOvVaw1IOtZl-Ft7q4k8EZSVidgk>

<sup>2</sup> This data was found from the below source:

<https://www.google.com/url?q=https://www.eia.gov/energyexplained/gasoline/gasoline-and-the-environment.php%23%26text%3DGasoline%2520use%2520contributes%2520to%2520air%2520pollution%26text%3DThe%2520vapors%2520given%2520off%2520when.carbon%2520dioxide%2520C%2520a%2520greenhouse%2520gas&sa=D&source=docs&ust=1682086780396482&usg=AOvVaw00eqh3T8tdhD-Tc4ZdlRiQ>

petroleum consumption data with data on solar energy, which is a much more environmentally friendly source of power.

2.

The data we will be using to explore the problem of increasing petroleum usage will primarily be provided to us by the US Department of Energy. The data is useful in understanding our problem since it contains historical data on the industrial and commercial consumption of petroleum from 1960 to 2019 for all fifty states and Washington D.C. Additionally, the dataset includes data on the industrial and commercial consumption of solar power, which we will compare with the data on petroleum consumption. The specific columns we will be using in the "energy.csv" file are "State" (categorical variable), "Year" (sequential variable), "Consumption.Commercial.Petroleum" (quantitative variable), "Consumption.Industrial.Petroleum" (quantitative variable), "Consumption.Commercial.Solar" (quantitative variable) and "Consumption.Industrial.Solar" (quantitative variable). These columns will allow us to determine the consumption of petroleum and solar power across the U.S. and compare these two power sources.

3.

The first task that we wish to conduct with our data is to understand the industrial and commercial consumption of petroleum across the United States, from 1960 to 2019.

The main reason why we wish to do this task is because we are aiming to explore the specific data points (for each year) that already exist on petroleum consumption, in order to produce informative visualizations that highlight how petroleum consumption has changed over a period of almost sixty years.

Through our task analysis for this task, we will identify all petroleum products consumed by the industrial and commercial sectors in billion BTU, for each year and for every U.S state in the given time period.

After this navigation portion, we will organize our data and categorize it by year, and name of the state. Furthermore, we will also categorize the data based on whether it was from the industrial sector or the commercial sector.

After this, we will make connections within our data and will observe how petroleum consumption has changed over the time period of 59 years, based on whether the consumption has happened within industry, or commercially.

What this task seeks to learn about the data is whether there is an overall upward trend in the consumption of petroleum (industrially and commercially) from 1960 to 2019. Furthermore, the task also seeks to understand if there is a difference in trends for industrial petrol consumption compared to commercial petroleum consumption over time.

The target dataset we are using revolves around industrial and commercial petrol consumption (in billion BTU) for the years 1960 - 2019 for all fifty U.S. states and Washington D.C.

In terms of the workflow we used to generate these tasks, we first wanted to figure out a metric that we could use to understand the consumption of fossil fuels over a period of time across the United States. After viewing the dataset from the US Department of Energy, we tried to understand what columns of data we should focus on, keeping in mind that we wanted to focus on a fossil fuel. We ultimately chose to work with petroleum because of how ubiquitous of a fossil fuel it is - from propelling cars and planes to heating a building, petroleum is everywhere! In terms of focusing on the specific columns of data we were going to use, we narrowed down to industrial and commercial petroleum consumption. These two categories specifically stood out to us, because they are known to be notorious contributors to air pollution caused by greenhouse gases (like petroleum).

As “the subject matter experts” in this case, we will be conducting this task of understanding the industrial and commercial consumption of petroleum across the United States, from 1960 to 2019. As experts, we will be designing visualizations, which we will then use to compare and understand the general trends seen in gasoline consumption (in billion BTU<sup>3</sup>). In this case, our audience is the US Department of Energy which will ultimately use our data to generate data-driven policy recommendations.

---

<sup>3</sup> BTU means British Thermal Unit (a measure of the heat content of fuels or energy sources)

The second task that we wish to conduct is to compare the petroleum products consumed by the commercial and industrial sectors in billion BTU with the solar energy consumed by the industrial and commercial sectors in billion BTU.

The main reason why we wish to do this task is because we are aiming to explore the specific data points (for each year) that already exist on petroleum consumption and compare these data points with the consumption of solar thermal energy across all fifty U.S. states. We wish to explore and compare this data, in order to produce informative visualizations that highlight the differences between the consumption of petroleum (which is a non-renewable energy source) and solar thermal energy (which is a renewable energy source).

Through our task analysis for this task, we will identify all petroleum products consumed by the industrial and commercial sectors in billion BTU, for each year and for every U.S. state in the given time period. Furthermore, we will also identify the industrial and commercial solar thermal energy consumption (in billion BTU) for each U.S. state across the fifty-nine-year long time period.

After this navigation portion, we will organize our data and categorize it by year, and name of the state. Furthermore, we will also categorize the data based on whether it was from the industrial sector or the commercial sector, for both petroleum and solar thermal energy.

After this, we will make connections within our data and compare differences between petroleum consumption and solar thermal energy consumption for every year and every U.S. state over the time period of 59 years.

What this task seeks to learn about the data is understanding how the difference between petroleum consumption and solar energy consumption (industrially and commercially) has changed over time, from 1960 to 2019. The task seeks to understand if the difference between using gasoline as an energy source, and solar energy as an energy source, has increased or decreased over time.

The target dataset we are using revolves around industrial/commercial petrol consumption and solar thermal energy consumption (in billion BTU) for the years 1960 to 2019 for all fifty U.S. states and Washington D.C.

In terms of the workflow we used to generate these tasks, we wanted to figure out metrics that we could use to understand the consumption of a non-renewable energy source, compared to the consumption of a renewable energy source over a period of time across the United States. After viewing the dataset from the US Department of Energy, we tried to understand what columns of data we should focus on, keeping in mind that we wanted to focus on a fossil fuel, and also a renewable source of energy. We ultimately chose to work with petroleum and with solar energy because of its

growing rise in popularity as an alternative source of power. In terms of focusing on the specific columns of data we were going to use, we narrowed down to industrial and commercial petroleum consumption, as well as industrial and commercial solar energy consumption.

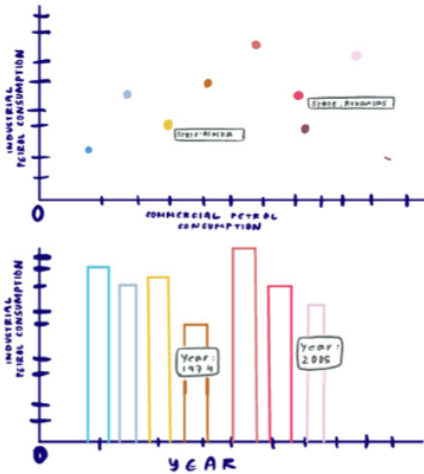
As “the subject matter experts” in this case, we will be conducting this task of comparing the industrial and commercial consumption of petroleum with the industrial and commercial consumption of solar energy, from 1960 to 2019, across the United States. In this case, our audience is the US Department of Energy which will ultimately use our data to generate data-driven policy recommendations.

4. On the next page there is an image of our low fidelity prototype (this image is also in our repository and it is titled “LOFI PROTOTYPE FOR MODULE 3 Copy (2).pdf”).

## COMMERCIAL/INDUSTRIAL PETROLEUM & SOLAR THERMAL ENERGY CONSUMPTION (in billion BTU)

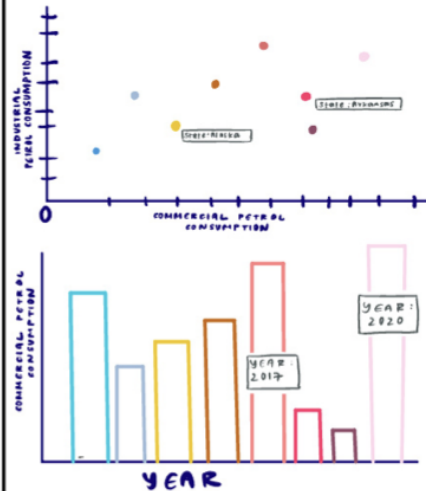
### Visualization A

Visualization A shows a scatterplot with the Industrial Petrol Consumption on the Y axis and the Commercial Petrol Consumption on the X axis. Each "State" datapoint on the scatterplot is coordinated with a histogram where each bar represents a year, and the Y axis has the amount of industrial petrol consumed.



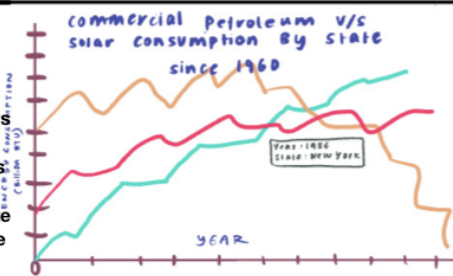
### Visualization B

Visualization B shows a scatterplot with the Industrial Petrol Consumption on the Y axis and the Commercial Petrol Consumption on the X axis. Each "State" datapoint on the scatterplot is coordinated with a histogram where each bar represents a year, and the Y axis has the amount of commercial petrol consumed.



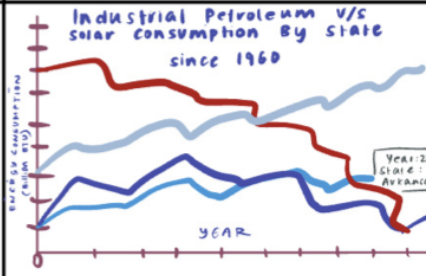
### Visualization C

Visualization C is a line graph that shows commercial petrol consumption versus commercial solar consumption by state for every state in the U.S.



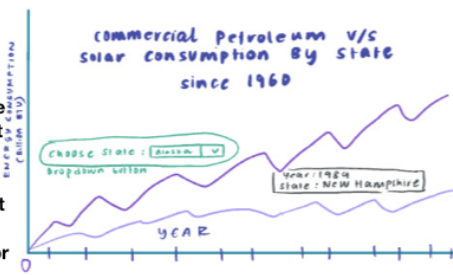
### Visualization D

Visualization D is a line graph that shows industrial petrol consumption versus industrial solar consumption by state for every state in the U.S.



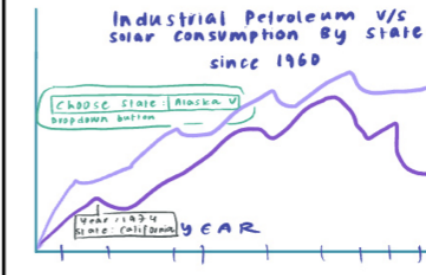
### Visualization E

Visualization E is essentially the same graph as above, but with a dropdown button that allows users to select what state they want to see the line graph for



### Visualization F

Visualization F is essentially the same graph as above, but with a dropdown button that allows users to select what state they want to see the line graph for



In the prototype, we have only decided to highlight commercial/industrial petroleum consumption and solar energy consumption, instead of other variables because we wanted our prototype to focus solely on the trends related to one non-renewable energy source and one renewable energy source.

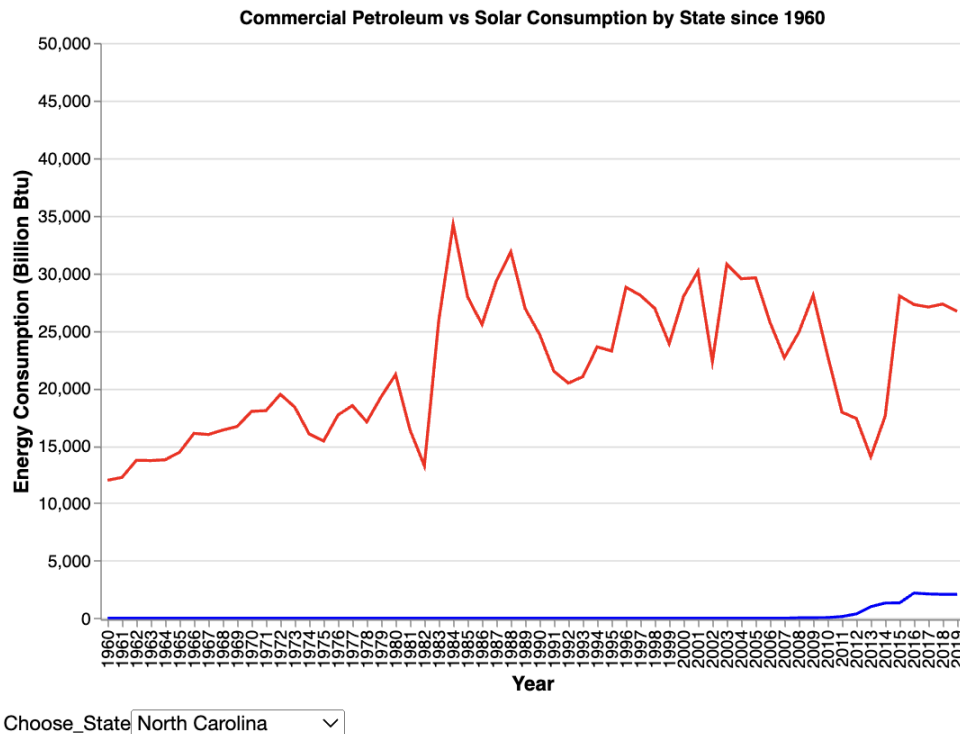
A design related choice we chose to make was to include one pair of graphs that show the industrial and commercial petroleum versus solar consumption levels for all states in one view (visualizations C and D). We then included a version of these graphs where one can use a dropdown menu to choose to see data for a specific state (visualizations E and F).



5. Based on the prototype above, we implemented our design as a more refined dashboard of visualizations. You can find this prototype as “CommercialIndustrial Petroleum & Solar Thermal Energy Consumption (in billion BTU).pdf” in our GitHub repository. Essentially we used Altair to design an assortment of visualizations that will help us conduct our two tasks. Visualizations A and B are view-coordinated visualizations that help us with our first task which is centered around understanding the trends of industrial and commercial petroleum consumption across the United States over the fifty-nine-year time period. While each scatterplot shows the trends over time for industrial and commercial petroleum consumption, we chose to coordinate these scatterplots with histograms that provide specific state-wise data on either commercial petrol consumption or industrial. To help us conduct our tasks more efficiently we added this functionality of allowing users to hover over a datapoint in each scatterplot in order for them to view the state that it corresponds to. Furthermore, we also have an added functionality in our histogram where one can hover each bar in order to view the year it corresponds to. Another design choice we included was the functionality of allowing users to click on a bar in each histogram to view corresponding data points in the scatterplot above the histogram. We utilized these design choices in order to more efficiently carry out our first task of identifying the trends in industrial and commercial consumption of petroleum (in billion BTU) across the United States, from 1960 to 2019. To carry out our second task of comparing petroleum consumption with solar thermal energy, we designed visualizations C, D, E and F. We created visualizations C and D (as line graphs) to understand the comparison between the usage of industrial/commercial petroleum versus solar for the time period in the dataset. These

visualizations show the data for every state at the same time so that we can carry out our task of comparing solar versus petroleum consumption over time, for all states. Meanwhile, visualizations E and F also help to carry out the task but have a dropdown button so that individuals can focus and understand the trend for one state at a time. One design choice we made to carry out our tasks effectively, using visualizations E and F, was to allow users to pan and zoom in on these graphs to either see a zoomed-in perspective or a zoomed-out one. These visualizations will aid in the Department of Energy's policy designing because we can use them to understand if the United States is on the general path of consuming more petroleum and/or more solar thermal energy, given historical trends. One can view the fully interactive visualizations in the Python notebooks "Code for Visualizations A and B.ipynb" and "Vis-Code.ipynb.gz" in our GitHub repository that is linked at the top of this document.

An image of visualization E (from the final prototype) for the state of North Carolina is provided on the next page. The red line corresponds to petroleum consumed while the blue line corresponds to solar thermal energy consumed.



6. Our full prototype is in the GitHub repository and it is titled as “CommercialIndustrial Petroleum & Solar Thermal Energy Consumption (in billion BTU).pdf”.

We used this prototype to carry out our first task of understanding whether there is an overall upward trend in the consumption of petroleum (industrially and commercially) from 1960 to 2019. Visualization A<sup>4</sup> helped us understand that overall there has been an increase in the overall consumption of petroleum industrially. However, visualization B helped us understand that overall there has been a decrease in the overall commercial consumption of petroleum.

Our prototype also helped us do our second task of comparing commercial/industrial petroleum consumption levels with solar thermal energy consumption levels.

Visualizations C, D, E, and F all indicate that overall the industrial and commercial

<sup>4</sup> These visualization names are in our low fidelity prototype which is in this very document, as an inserted image

consumption of petroleum has always been higher than the industrial and commercial consumption of solar thermal energy for a given state and year.

7. One new task that we would like to conduct is to compare industrial petroleum consumption with industrial hydropower consumption. To incorporate this new task and iteration, we will tweak our prototype in a manner that incorporates a line graph that compares industrial petroleum consumption with industrial hydropower consumption from 1960 to 2019.

8. Our prototype visually shows how there is an increase in the commercial consumption of petroleum from 1960 to 2019, while there is a general decreasing trend in the industrial consumption of petroleum. Based on this general upward trend in petroleum consumption within the commercial sector, we believe that future policy directions should focus more on reducing petroleum consumption within large entities that make up the commercial sector in the United States. Furthermore, our prototype also highlights how the U.S. Department of Energy should focus more on increasing the consumption of solar thermal energy since its levels of consumption are consistently lower than that of petroleum, despite it being a cleaner source of power. As per a climate report by the U.N's World Meteorological Organization, the year 2020 (a year after the scope of this dataset) was one of the warmest years on record<sup>5</sup>. This is yet another reason why the Department of Energy should propagate and put into place

---

<sup>5</sup> The source for this information is below:

[https://www.google.com/url?q=https://www.un.org/en/climatechange/reports%23%3A~:text=The%2520State%2520of%2520the%2520Global%2520Climate%25202020%2520finds%2520the%2520year,\(1850%252D1900\)%2520level&sa=D&source=docs&ust=1682111407294597&usq=AOvVaw0BTP7tUPOEvzKvcJKegvoJ](https://www.google.com/url?q=https://www.un.org/en/climatechange/reports%23%3A~:text=The%2520State%2520of%2520the%2520Global%2520Climate%25202020%2520finds%2520the%2520year,(1850%252D1900)%2520level&sa=D&source=docs&ust=1682111407294597&usq=AOvVaw0BTP7tUPOEvzKvcJKegvoJ)

specific policies that increase the industrial and commercial consumption of solar thermal energy. Solar thermal energy is cleaner and much less environmentally degrading than petroleum.

9. Putting our designer hat on, our visualizations made us realize how effective it can be to view-coordinate one graph with another. In a specific situation like ours where we are trying to view industrial/commercial petroleum consumption for every year and state, it made sense to create a visualization with two separate and specific views tied to each other.

**A note on the bells and whistles part of the assignment**

1. In terms of deployment, our repository is publically available at the link provided on the first page of this document

2. In terms of view coordination, visualizations A and B are coordinated with each other where clicking on a bar on the histogram makes specific data appear on the scatterplot above

3. In terms of iteration, you can find all the relevant information regarding it down below!

### Information on our iterated design study:

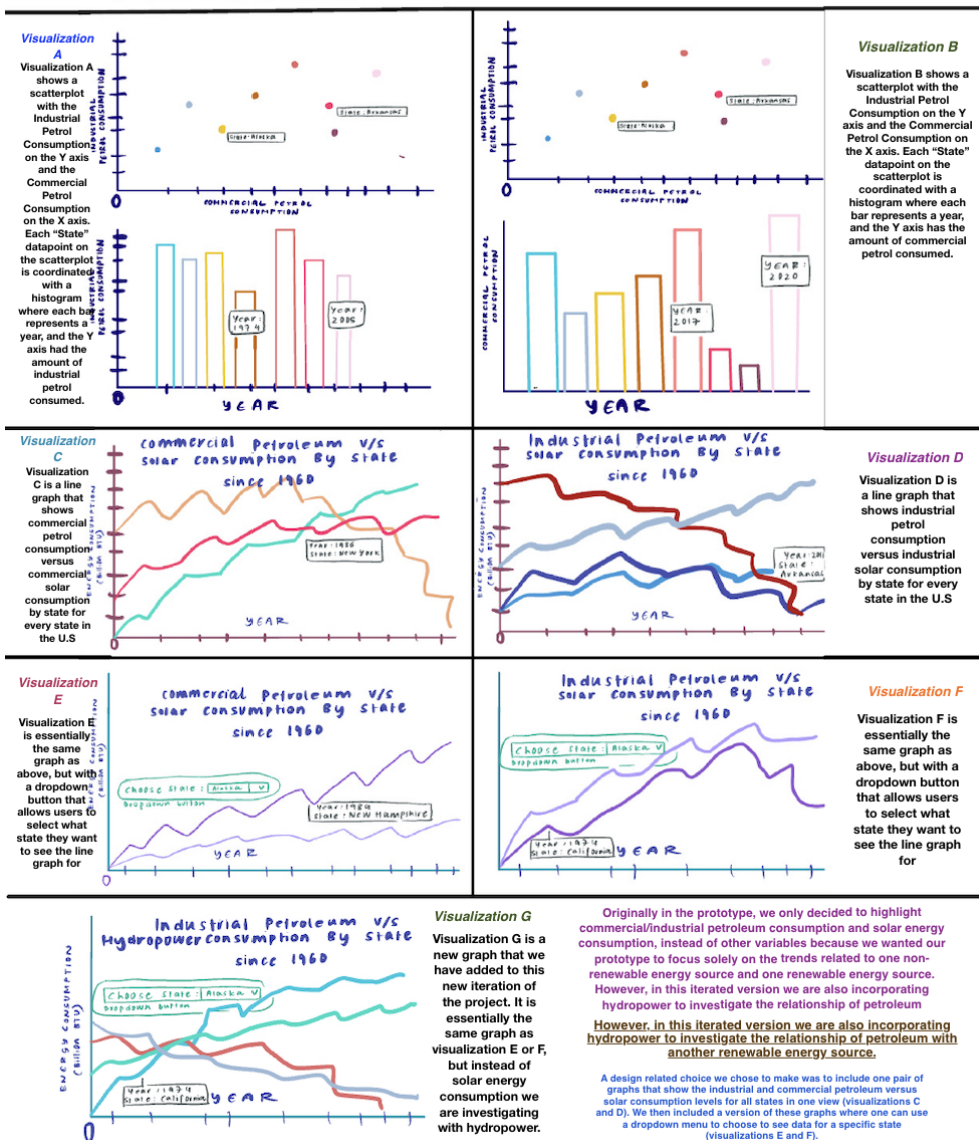
The original discussion: *One new task that we would like to conduct is to compare industrial petroleum consumption with industrial hydropower consumption. To incorporate this new task and iteration, we will tweak our prototype in a manner that incorporates a line graph that compares industrial petroleum consumption with industrial hydropower consumption from 1960 to 2019.*

Description of our reimplementation: The reason why we would like to compare industrial petroleum consumption with industrial hydropower consumption, is to understand if the large differences that we saw between industrial solar energy and petroleum consumption (in visualization F) are also present when comparing petroleum with another renewable energy source - hydropower.

### Steps 4 - 6, but with the iteration:

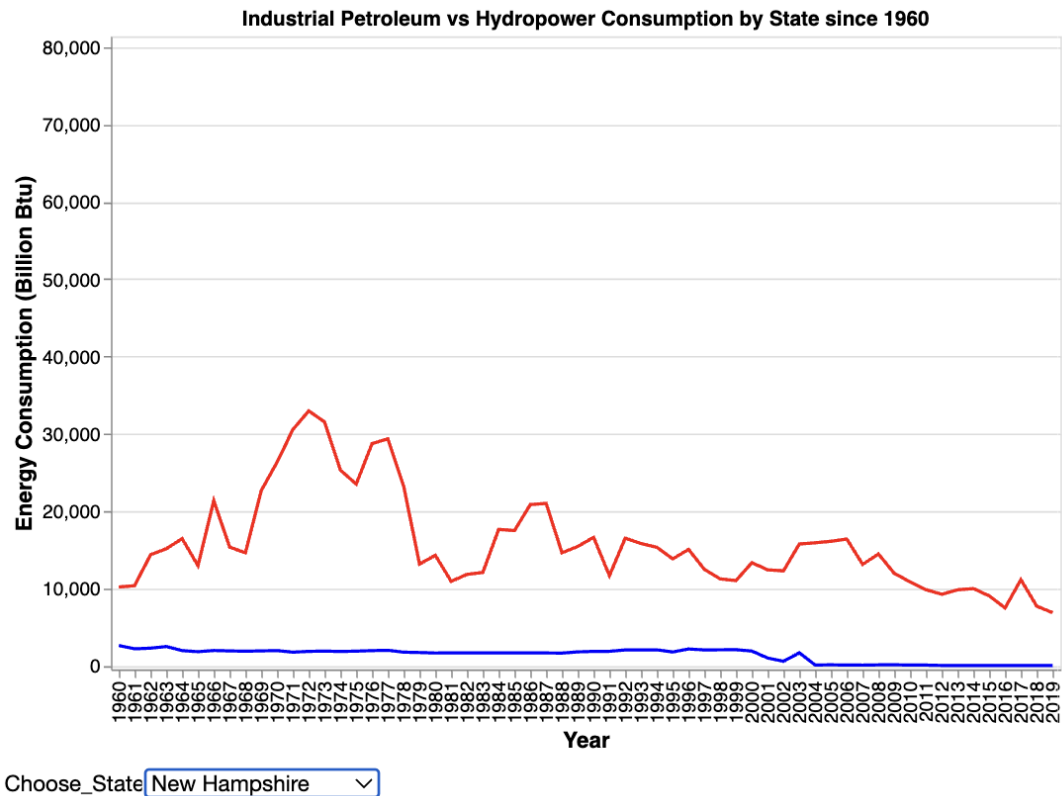
4. Below is an image of our low fidelity prototype with the iteration (this image is also in our repository and it is titled "ITERATED LOFI PROTOTYPE FOR MODULE 3.pdf").

## COMMERCIAL/INDUSTRIAL PETROLEUM & SOLAR THERMAL ENERGY CONSUMPTION (in billion BTU)



5. Based on the prototype above, we implemented our design as a more refined dashboard of visualizations with the iterated visualization included as well. You can find this prototype as "Iterated-CommercialIndustrial Petroleum & Solar Thermal Energy Consumption (in billion BTU).pdf" in our GitHub repository. We chose to incorporate a line graph showing industrial petroleum versus hydropower consumption so that we could achieve our task of understanding if there is a large difference between

hydropower consumption and petroleum consumption for every U.S. state. An image of the iterated visualization that we added is shown below. The red line corresponds to petroleum consumed while the blue line corresponds to hydropower consumed.



6. The iterated visualization that we added (as shown above) highlights how the level of hydropower consumption within the industrial sector in the United States (from 1960 to 2019) is consistently lower than the consumption of petroleum. We can ascertain this because, for every state that we selected in our graph, we saw that the red line for petroleum consumption was higher than the blue line for hydropower consumption.