CECS 450 Final Project

Dylan Huynh
College of Engineering
CSULB
Long Beach, USA
Dylan.Huynh@student.csulb.edu

Sebastian Tadeo
College of Engineering
CSULB
Long Beach, USA
Sebastian.Tadeo@student.csulb.edu

Nicholas Yamamoto
College of Engineering
CSULB
Long Beach, USA
Nicholas.Yamamoto@student.csulb.edu

***Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. **CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.* (*Abstract*)**

Keywords—component, formatting, style, styling, insert (*key words****)***

I. Our Idea and how we handled our data

Our initial thought process was to see if how much energy we use, like gas, electricity, or biogases, had any effect on anything in our society. We thought of things like if there was a correlation between gas usage and global warming, gas prices affecting electricity usage, or electricity prices affecting gas usage, and many other possible ideas, and some were just so out there in terms of surface level correlation. We ultimately settled on seeing if there was a correlation between gas and electricity prices, and their respective powered cars, gas cars and electric cars.

We got most of our data on car prices from Kaggle. Kaggle had quite a bit of selection on car data, whether that was car prices, most popular cars by sales, to even cars that harmed the environment the most, so we thought Kaggle was a good source for data pertaining to cars. We got one of our big energy datasets from the United Nations Global Energy trade and production database. The United Nations, as far as we know, is a very reliable source so we trust that the huge dataset that they provided had good information. We got another dataset on gas and electricity from the US Government's database. Both of these datasets came from government agencies, and we think that the government should have the most records and hopefully the most accurate records. Internet usage data was from OurWorldInData's database, which is also a trusted source. We got electricity cost data from Statista and Computer parts/GPU data from Kaggle.

A lot of our data sets were massive with some having as much as about 20,000 data entries. Obviously, we did not use every single data entry. We did some research, and we found an R library called sqldf that allows you to use SQL queries to mold our data. We enter an SQL query as a string and sqldf makes the translation to R code to make new data frames with our desired specifications. When we didn't use the sqldf library, we just used the dplyr library to filter data. The figure below is an example of how we were able to use the sqldf library. "Result" would be a new data frame with the results from the SQL query.

```
sql <- "SELECT Year, Percent
        FROM cpu
        WHERE Year BETWEEN 1997 AND 2000"
result <- sqldf(sql)
```

When making the GPU graph, we had to filter out some empty string/null values from the "max power" column, and we had to remove certain entries where their "release date" column had "Unknown release date" in it. The data in the "max power" column was saved as a string since it was entered as "X watts". We had to get rid of watts and make the data a numerical data type in order to accurately plot the data. We also changed the data in the "release date" column to date objects using the anytime library. We also did the same thing for the date object in the CPU graph.

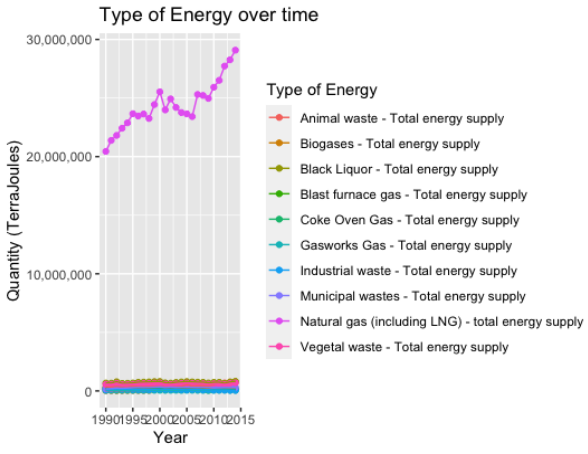
II. Issues

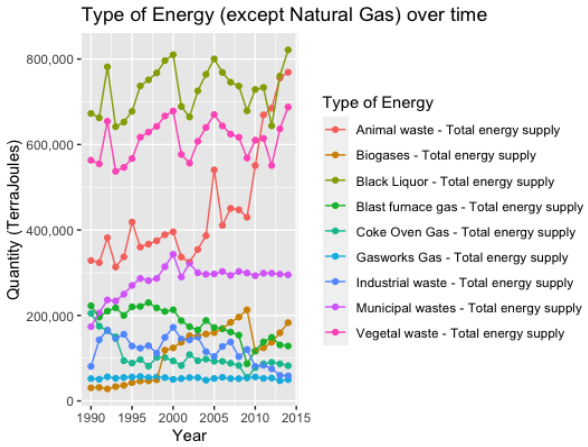
Some problems that we had to get through were limiting the time frame we visualized our data in. A lot of our graphs had different time periods that it recorded, so we had to mold the data so that the time periods matched to keep data consistent. Another problem we had was figuring out how to get the right sql query. Although we were familiar with SQL, the reason why we decided to use this library, it still took us some extra thought in writing the SQL. An additional problem we faced was dealing with null values and data type conversion. When reading the GPU data table, the max power and release date columns were interpreted as a string data type. To remove null values for both max power and release date columns, we used the library dplyr filter() function. The filter() function would remove blank string entry for the max power column and remove 'Unknown Release Date' string for the release date column. To remove the word 'Watts' we used the stringr package str_remove_all() and dplyr mutate() to edit the max power column. We converted the max power column as numeric and used the anytime package to convert the release date strings to date object. For the chip data table, we filtered only CPU and convert the release date strings to date object using the anytime package. Another problem we got through was learning how to fully customize our graphs visually. We did use some pre-set themes for some of the graphs, but for other graphs we fully customized them to fit our background color, line/bar color, and font needs. We wanted full control over the aesthetics of the graphs so that it would be more visually appealing and so that we could utilize some of the "persuasive" techniques we learned in class. For example, the graphs we made for computer usage and internet

usage, we used a specific colorway that made it feel like we were looking at a terminal. Another example would be in our electricity price comparison, we used the color blue for the lower price, and we used the color red for the higher price. Generally blue has a more positive connotation, the lower price, and red generally has a more negative connotation, the higher price. There were other aesthetics that we put thought into, but these are just some examples.

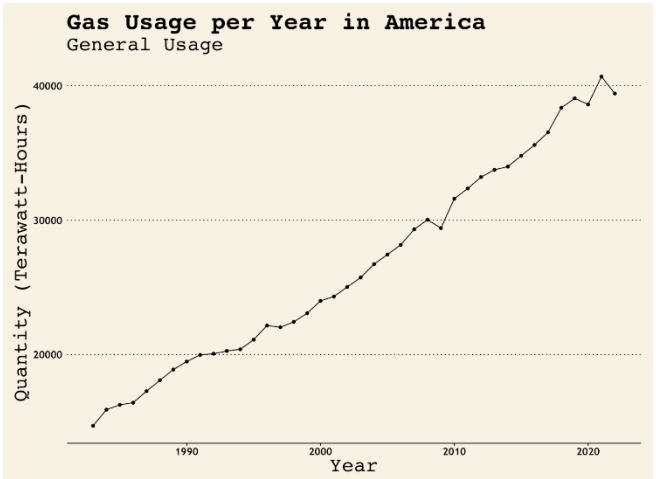
III. Results

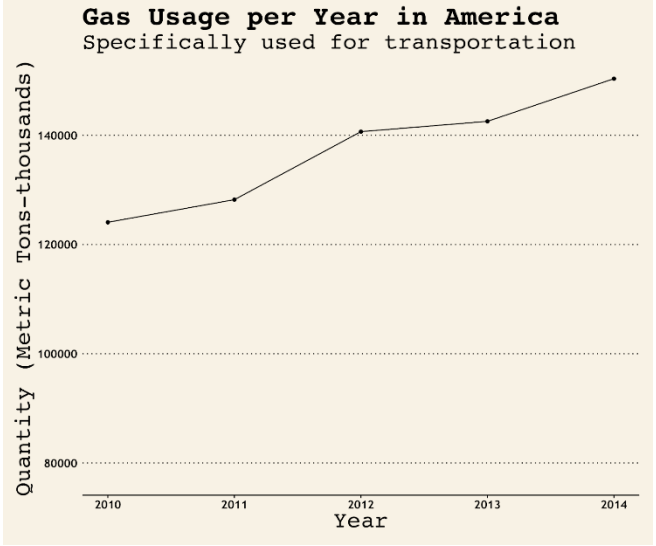
We first have a dataset that has all different kinds of energy, and its usage over time to see what stands out to us.





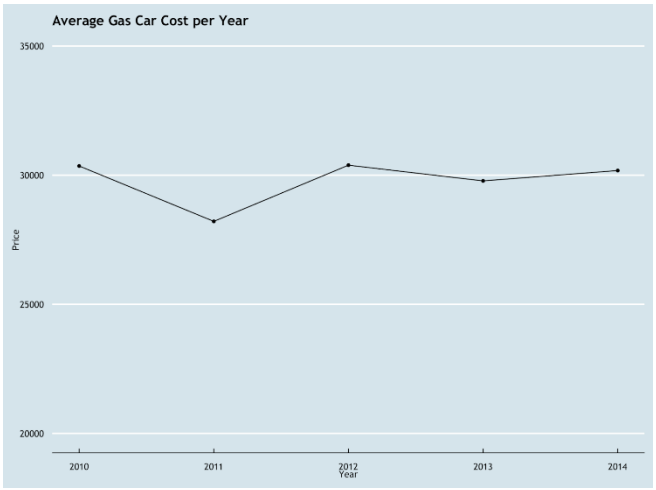
We noticed that natural gas just blew the competition away. It was clear that gas was the most used energy source, and that makes sense to us since gas is fairly cheap to produce. We wanted to see how the other types of energy compared to one another, so we graphed all of the energy types excluding natural gas. We found that Black Liquor was the highest used and averaged out, Gaswork Gas was the least used. Gas was clearly the most used, so we decided to look more into gas usage in the US. We were able to find gas usage overall, and specifically how much gas is used for transportation. We noticed the upward trend as time went on, and as we said before that makes sense since gas is the most affordable energy source to produce.



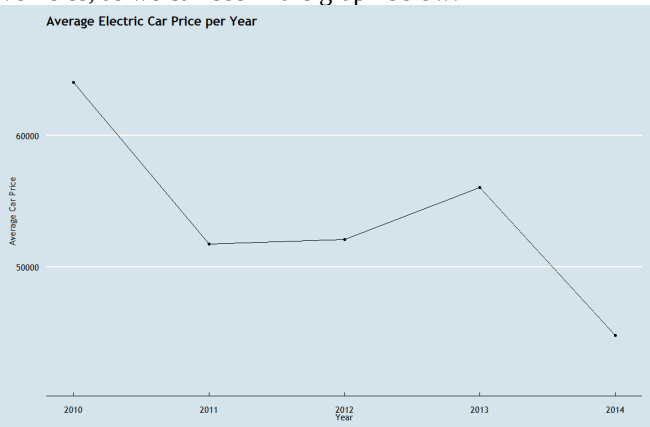


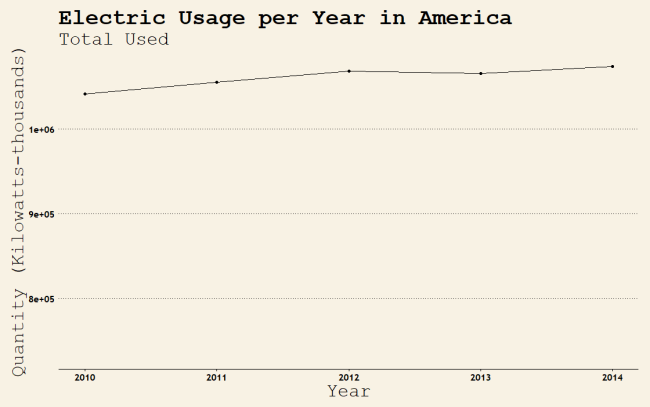
The Gas usage specifically for transportation is the dataset that limited our time frame, since the only data we were able to find on that was from 2010-2014. This time frame is used for comparison in all the gas price/gas car and electric price/electric car comparisons.

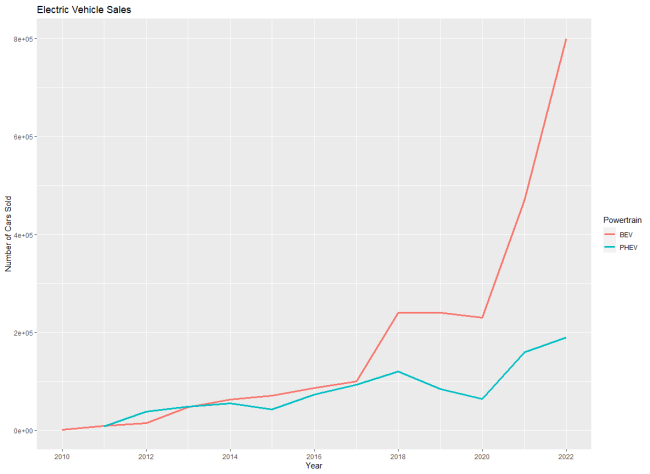
We have the two datasets side by side visually, and we want to see if there is any correlation. The gas usage as we mentioned before is on the rise every year, but the gas car prices fluctuate, with little to no correlation between the car cost and gas price.



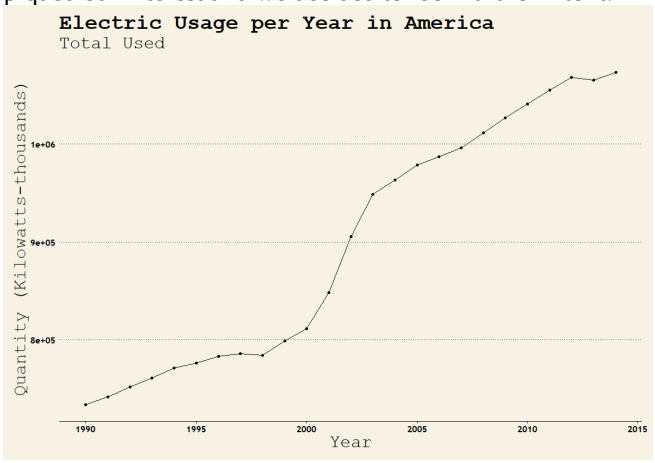
Much like the gas price and gas car correlation, there was not much of a correlation with electricity. We did notice a trend downward in electric vehicle prices, so we researched a possible reason. What may have led to this downward trend in prices was the announcement that Tesla was going to start manufacturing luxury electric vehicles, which caused other car companies to speed up their production of electric vehicles. That would lead to an upward trend in electric vehicles, as we can see in the graph below.



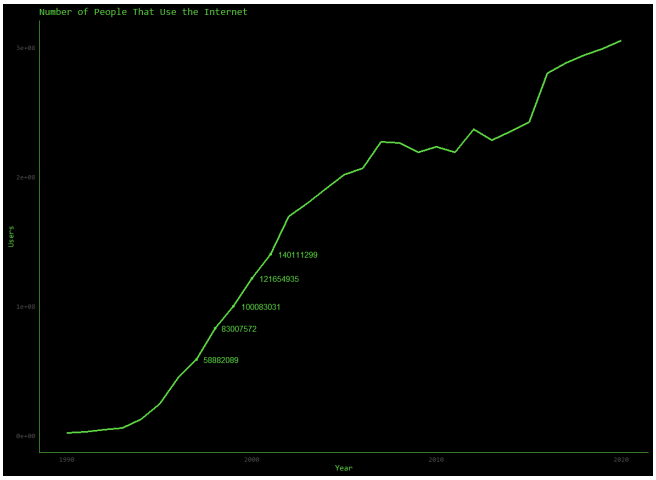


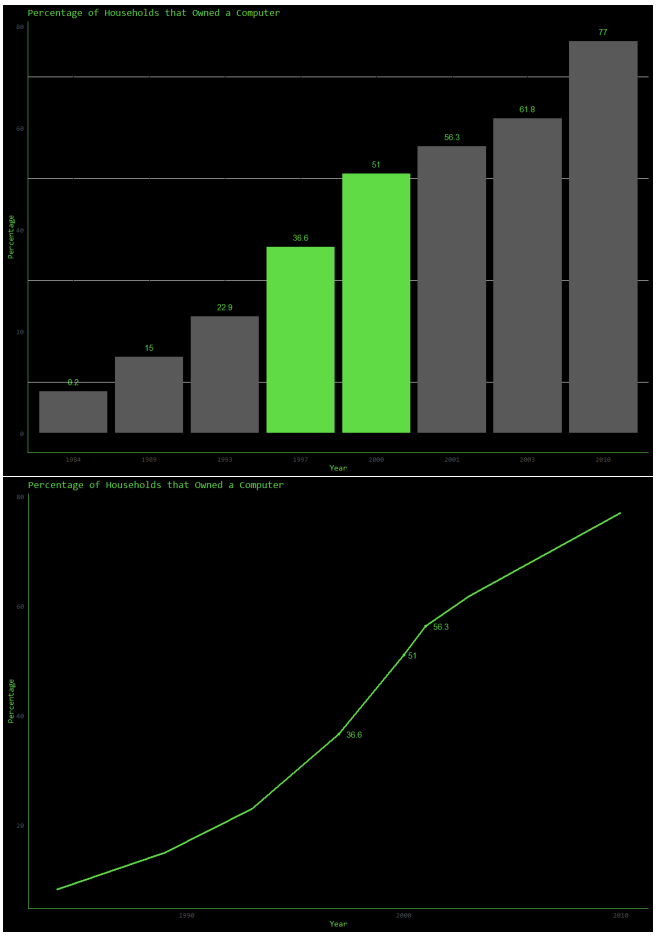


We did not find too much correlation between gas usage and gas-powered car prices, or electricity usage and electric vehicle prices, however when we looked at the overall usage of electricity in America, we noticed there was a huge spike in electricity usage around 1997-2003. This piqued our interest and we decided to look further into it.

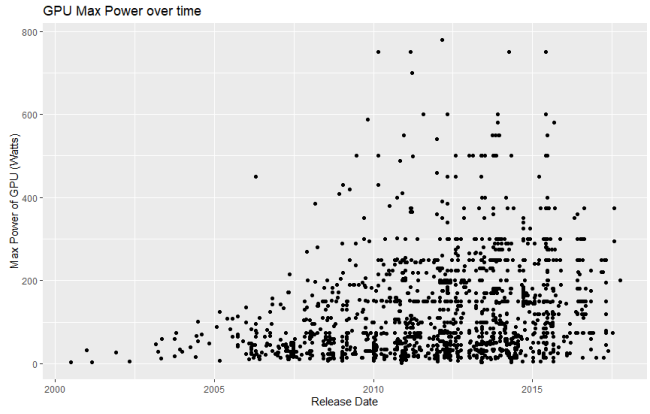


We found that one of the main reasons for this spike is the internet's release to the public in 1993. It was actually the public release of the World Wide Web in 1993, but this made it easy for anyone to access the internet, so for all intents and purposes, 1993 was the launch of the internet. By 1995 more than 24 million people spent an average of 5 hours a week on the internet, and the number of people would only increase. Easier accessibility to computers alongside the launch of the internet are some of the big reasons why there was a huge spike in electricity usage. Computers were getting more affordable for households to buy, and we see the steepest increase in the number of households that have computers is around the same time as the internet spike, 1997-2000.

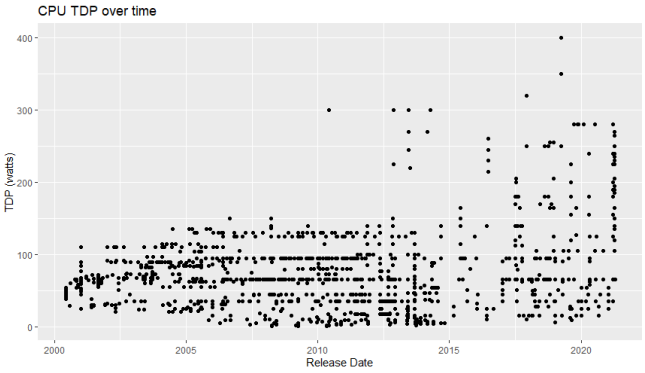




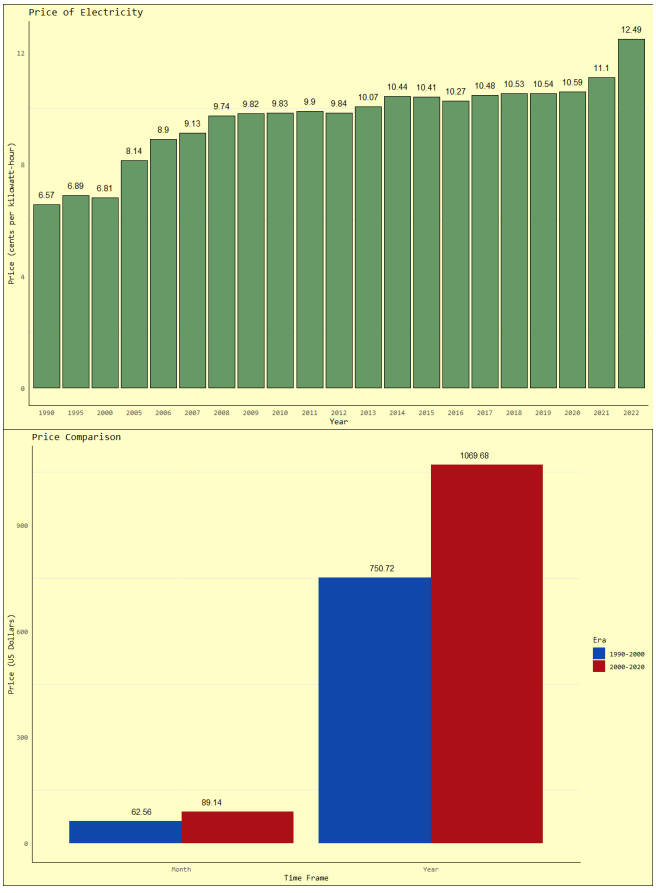
While we were on this path of computers and the internet, we decided to look at how some parts of the computer evolved over time. We looked at a GPU's max power over time and a computer's thermal design power (TDP). A computer's thermal design power is the maximum amount of heat a computer chip or component can produce while still allowing the cooling system to dissipate the heat. In terms of GPU max power, there were not too many GPUs being produced in the early 2000s, which makes sense, there wasn't too much of a demand for them during that time. None of the GPUs created before around 2007 had a max power higher than 200 watts. Once the 2010s hit they started to be produced more frequently, where the highest max power was a little under 800 watts, while the majority of the GPUs created in the 2010s are around the 200-300 watt range.



Computers throughout the 2010s, with the exception of one, all had a TDP under 200 watts, and then after 2010, there were a few computers that branched out above 200 watts with some reaching up to 400 watts. For the most part, however, most computers still stayed under 200 watts even through 2020.



We were also interested to see the difference in price before 2000 and after 2000. As expected, there was a jump in price right around the same time as the jump in electricity, computer, and internet usage. It went from about $6.81 to $8.14, and we calculated how much of an impact that makes. We averaged out all the prices before 2000, and averaged all the prices after 2000 and based it on a 898 kWh USA average. The monthly price was not too much of a difference with $62.56 pre 2000 and $89.14 post 2000. However, the difference is small because the time frame is not too large. The yearly difference is much bigger with $750.72 pre 2000 and $1069.68 post 2000. That is a significant difference.



IV. CONCLUSION

We did not find any correlation between our original idea in gas and electricity usage and gas and electric car prices, but our research led us down a different path when we were interested in the huge electricity usage spike around 2000. It led us to see different data points with the progression of technology, like how electricity usage was affected by the release of the internet and how much computers were being used. It also led us to different data points with GPU and CPU TDP.

REFERENCES

[1] "Percentage of U.S. households with a computer 1984-2010 | Survey," *Statista*, 2010. https://www.statista.com/statistics/184685/percentage-of-households-with-computer-in-the-united-states-since-1984/

[2] J. Murphy, M. Roser, and E. Ortiz-Ospina, "Internet," *Our World in Data*, 2015. https://ourworldindata.org/internet

[3] "Average retail electricity price in the United States 2018," *Statista*. https://www.statista.com/statistics/183700/us-average-retail-electricity-price-since-1990/

[4] R. Matulka, "The History of the Electric Car," *Energy.gov*, Sep. 15, 2014. https://www.energy.gov/articles/history-electric-car

[5] "Global Energy Substitution from 1983 to 2022," *www.kaggle.com*. https://www.kaggle.com/datasets/mohamedyosef101/global-energy-substitution-from-1983-to-2022.

[6] "Energy Prices, Dollars per Million Btu: Beginning 1970," *Data.gov*, Feb. 03, 2020. https://catalog.data.gov/dataset/energy-prices-dollars-per-million-btu-beginning-1970.

[7] "Computer Parts (CPUs and GPUs)," *www.kaggle.com*. https://www.kaggle.com/datasets/iliassekkaf/computerparts/
.

[8] "CPU and GPU Performances Dataset," *www.kaggle.com*. https://www.kaggle.com/datasets/michaelbryantds/cpu-and-gpu-product-data

[9] J. Ring, "30 years ago, one decision altered the course of our connected world," *NPR*, Apr. 30, 2023. https://www.npr.org/2023/04/30/1172276538/world-wide-web-internet-anniversary#:~:text=Embedded-