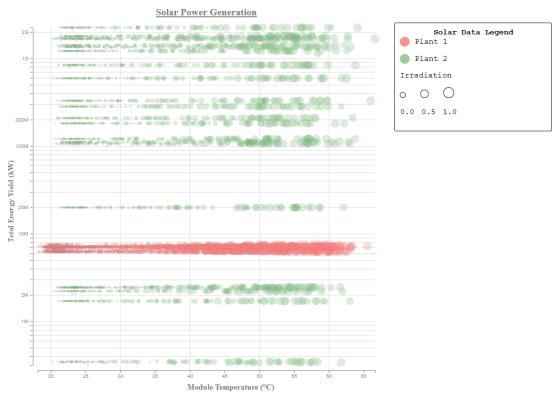# Project Report

Alexander Pirouznia, Carl Chen, Allison Yin, Haneul Song

Part A: Visual Screenshots

Part B: Data Description

The data for this project was sourced from a Kaggle dataset created in 2019 found here. Included was data collected from 2 solar power plants in India over a 34 day period. Each solar plant included generation information where AC Power, DC Power, daily yield, and total yield was collected every 15 minutes. Additionally, weather sensor data was collected for both solar plants where ambient temperatures, module temperatures, and irradiation levels were collected every 15 minutes.

For this analysis, we were interested in examining how energy yield levels were impacted by key environmental factors. In doing so, we narrowed down the datasets into modular temperatures, irradiation, and total yield. Modular temperature is the data collected directly from the module attached to the sensor panel thus seemingly having the most influence on solar energy production when it comes to temperatures. Another key component to solar energy production is irradiation which is the amount of energy from the sun reaching the surface of the panels. Finally, we chose

to focus on total yield because we wanted to display the total energy yield of both plants among the other variables.

With the datasets of Plant 1 and Plant 2 weather sensor and generation data, it was important that we implemented a Promise.all call to load all the csv files together. This was because we were using data values from all four of these csv files to create our data visualization. Promise.all let us load the following csv files together all at once.

Then, we deemed it easiest to extract the individual data values we needed and combine them into a new combined dataset for both plant 1 and 2. This would provide us a simpler foundation for the rest of the data visualization and in identifying the domain and range when using d3 features. To do this, we used the Javascript map function to extract individual columns from the data array, such as irradiation, module temperature, and total yield. Next, we wanted to convert them to numeric values by using the data conversion operator "+" so they could be parsed correctly. Then using the data.forEach function, we were able to create a new data array by iterating through the extracted values for Plant 1 and 2. For each index, we appended the extracted data values for temperature, yield, and irradiation for each of the two plants.

With this new dataset, we were able to use the method we learned in class to build scales for the Combined dataset. For yScale, we decided on a log scale instead of a linear scale because it was better for handling data with large ranges, which was apparent with the differences in data values between plant 1 and 2. This is also represented in our data visualization screenshot. Furthermore, it emphasized proportional changes between Plant 1 and 2. It was a better representation of skewed data, as we can see, plant 2 has more scattered values and a much larger range. The log scale allowed us to compress outlier values and stretch out clustered values, providing a more balanced visualization.

Lastly, with the margins in mind and using the previously constructed scales, we were able to construct the gridlines. As a refined touch, we created labels for the axes and added annotations for the gridlines.

Part C: Overview of Design Rationale
We chose to represent our data with a scatterplot, with circles as the marks. We used the Module Temperature as the x-location of each point, as this is one of the independent variables. The y-location of each point is determined by the Total Energy Yield, as this is the dependent variable in our data. We also wanted to show the irradiation of each point, and chose to do so through the radius of the circles. Lastly, in order to differentiate between the two solar plants, we colored the points using the "fill" attribute, varying the color hue depending on the plant. We used an ordinal scale with a range of 2 colors because the Plant is a categorical variable with 2 possible values.

We utilized several visual channels to observe the influence of multiple independent factors at once, in the same graph. However, this brought some challenges with the design of our graph. First, the data for solar plant 1 has a much smaller range in total energy yields than solar plant 2. By having the data for both plants on the same scale, it is difficult to see changes in the total energy yield for plant 1 due to its small range. However, we felt that the ability to compare the plants with one another outweighed this challenge. Thus, we kept the data for both of the plants on the same graph.

In addition, the values of total energy yield are clustered around certain values, with large gaps in the range of data. This leaves some y-values highly represented and others with no data points. To better observe the individual data points, we reduced the opacity of the circles to 0.25. This allows us to see any overlaps in data points, providing insight into common values for module temperature and total energy yield. To further support visibility, we added a stroke to each circle with an opacity of 0.5. The opacity of the stroke is higher than that of the circles in order to clearly distinguish between data points. However, it is reduced from its default value of 1 to avoid blocking relevant data.

Part D: The Story
Our visualization reveals the intricate relationship between environmental factors and the energy yield of two solar power plants in India. By comparing module temperature, irradiation levels, and total yield, we can see how these variables impact solar energy production.

One surprising aspect of our analysis was the extent to which module temperature affected total yield. We found that even small fluctuations in temperature had a significant impact on energy output, emphasizing the need for effective temperature management in solar installations.

The key insights we want to convey to viewers are that understanding these environmental influences is crucial for optimizing solar energy production. Our visualization demonstrates how variations in temperature and irradiation levels lead to substantial differences in energy yield, offering stakeholders valuable data to enhance operational strategies and improve the efficiency of solar power plants.

Team Distributions

Carl: First, I contributed to all the group meetings and worked accordingly on the separate milestones. For the html file itself, I worked on pre-processing the four csv files we had to upload the data. From there, I extracted data columns that we needed: such as total yield, module temperature, and irradiation and created a new dataset to compare the two solar plants. Next, I created scales using d3 features such as Extent and Scale that would assist in creating the axises. Finally I created the left and bottom axises and added left and bottom gridlines to our data visualization. For the documentation aspect of the report, I worked on the data description and my design rationale for the decisions I made within the html file.

Alex: I contributed to all of the group meetings and completed timely work for each of the milestones. Specifically, I helped initiate the project by creating and administering the GitHub repository. After creating the boilerplate HTML file and importing the datasets, I created the default SVG element and set up basic data extraction functions ensuring that the data was loaded correctly. Next, I helped team members wherever possible.I also added axis labels and worked on the description of the data explaining why we chose this dataset and chose the values that we ended up putting in the plot. Lastly, I did polishing for the final web page's presentation and added the screenshot to this report.

Allison: I contributed to all group meetings and milestones. For the code, I extracted and added the irradiation data to our created dataset. I also created a color scale for the Plant variable. I then used the data set to add the marks to our graph, plotting each data point and utilizing each variable as a visual channel. This took the majority of my time, about 1 hour. For the report, I wrote the design rationale explaining our design choices for the graph. Specifically, I documented the design choices I made in plotting the data points.

Lisa: I helped manage the team's time management by creating meeting agendas and timelines to track our project effectively, making sure to submit all milestone files on time. Additionally, I contributed to the dataset by designing a legend that provides a clear and organized representation of the data points in our visualization. I contributed to the project report by writing "The Story" section of the report. The goal of the story was to explain what our visualization reveals and insights.